SBIAgro 2017

2 a 6 de outubro de 2017

XI Congresso Brasileiro de Agroinformática

Ciência de Dados na Era da Agricultura Digital

Anais

Campinas
Outubro de 2017

Embrapa Informática Agropecuária

UNICAMP

SBIAgro 2017

Anais do
XI Congresso Brasileiro de Agroinformática

Campinas, São Paulo
2 a 6 de outubro de 2017

Organizadores:
Jayme Garcia Arnal Barbedo
Maria Fernanda Moura
Luciana Alvim Santos Romani
Thiago Teixeira Santos
Débora Pignatari Drucker

Promoção
Associação Brasileira de Agroinformática
Realização
Embrapa Informática Agropecuária
Unicamp – Universidade Estadual de Campinas

Publicado por: Unicamp – Universidade Estadual de Campinas

ISBN: 978-85-85783-75-4

Créditos:
Capa: Tuíra Santana Favarin, sob supervisão de Flávia Bussaglia Fiorini
Projeto gráfico e editoração eletrônica: Thiago Teixeira Santos
usando o pacote LaTeX's 'confproc', versão 0.8

https://www.sbiagro.org.br/sbiagro2017

Congresso Brasileiro de Agroinformática *(11. : 2017 : Campinas, SP)*
Ciência de dados na era da agricultura digital : anais do XI Congresso Brasileiro de Agroinformática, 2 e 6 de outubro de 2017 / Jayme Garcia Arnal Barbedo... [et al.], organizadores.- Campinas : Editora da Unicamp : Embrapa Informática Agropecuária, 2017.
784 p.

ISBN: 978-85-8578375-4

1. Agricultura digital. 2. Inovação. 3. Ciência de dados. 4. Internet das coisas 5. Agricultura de precisão. 6. Processamento de imagens. 7. Aplicativos móveis. I. Barbedo, Jayme Garcia Arnal. II. Universidade Estadual de Campinas. III. Embrapa Informática Agropecuária. IV. Título.

CDD (21 ed.) 630.285

1ª edição on-line 2017
©Unicamp 2017

COMISSÕES

COMISSÃO ORGANIZADORA

Presidente: **Luciana Alvim Santos Romani** (Embrapa Informática Agropecuária)
Vice-Presidente: **Jurandir Zullo Jr.** (Cepagri-Unicamp)

Adauto Luiz Mancini (Embrapa Informática Agropecuária)
Ariovaldo Luchiari Júnior (Embrapa Informática Agropecuária)
Débora Pignatari Drucker (Embrapa Informática Agropecuária)
Jayme Garcia Arnal Barbedo (Embrapa Informática Agropecuária)
Junia Rodrigues de Alencar (Embrapa Informática Agropecuária)
Kleber Xavier Sampaio de Souza (Embrapa Informática Agropecuária)
Maria Fernanda Moura (Embrapa Informática Agropecuária)
Maria Giulia Croce (Embrapa Informática Agropecuária)
Renata Ribeiro do Valle Gonçalves (Cepagri-Unicamp)
Silvia Maria Bau Idalgo (Embrapa Informática Agropecuária)
Solange Kahl (Cepagri-Unicamp)
Sônia Ternes (Embrapa Informática Agropecuária)
Thiago Teixeira Santos (Embrapa Informática Agropecuária)

COMISSÃO CIENTÍFICA

Presidente: **Jayme Garcia Arnal Barbedo** (Embrapa Informática Agropecuária)
Vice-Presidente: **Ricardo Torres** (IC-Unicamp)

Adenauer Correa Yamin (UFPel)
Alaine Margarete Guimarães (UEPG)
Ana Maria Heuminski de Avila (Cepagri-Unicamp)
Ana Paula Lüdtke Ferreira (Unipampa)
Andreiwid Sheffer Corrêa (IFSP)
Ariovaldo Luchiari Jr. (Embrapa Informática Agropecuária)
Augusto Guilherme de Araújo (IAPAR)
Balbino Antônio Evangelista (Embrapa)
Bruno de Carvalho Albertini (USP)
Bruno Magalhães Nogueira (UFMS)
Caetano Traina Jr. (USP)
Carlos Alberto Alves Meira (Embrapa Informática Agropecuária)
Carlos Eduardo Cugnasca (USP)
Catize Brandelero (UFSM)
Celina Takemura (Embrapa Monitoramento por Satélite)
Débora Pignatari Drucker (Embrapa Informática Agropecuária)
Edson Takashi Matsubara (UFMS)
Eduardo Augusto Bezerra (UFSC)
Fabiano Santos (USP)
Maria Fernanda Moura (Embrapa Informática Agropecuária)
Filipe Neves dos Santos (INESC)
Gustavo Enrique de Almeida Prado Alves Batista (USP)
Heloisa de Arruda Camargo (UFSCAR)
Jayme Garcia Arnal Barbedo (Embrapa Informática Agropecuária)
João Francisco Gonçalves Antunes (Embrapa Informática Agropecuária)

João Batista Tolentino Júnior (UFSC)
João Camargo Neto (Embrapa Informática Agropecuária)
Josenalde Oliveira (UFRN)
Juan Manuel Adan Coelho (Puccamp)
Juliana Saragiotto Silva (IFMT)
Júlio César Dalla Mora Esquerdo (Embrapa Informática Agropecuária)
Kleber Xavier Sampaio de Souza (Embrapa Informática Agropecuária)
Leonardo Ribeiro Queiros (Embrapa Informática Agropecuária)
Lúbia Vinhas (INPE)
Luciana Alvim Santos Romani (Embrapa Informática Agropecuária)
Luciano Gonda (UFMS)
Luis Gustavo Barioni (Embrapa Informática)
Luiz Camolesi Júnior (FT-Unicamp)
Luiz Henrique Rodrigues (Unicamp)
Maria do Carmo Fasiaben (Embrapa Informática Agropecuária)
Maurício Fernando Lima Pereira (UFMT)
Pedro Luiz Correa (USP)
Poliana F. Giachetto (Embrapa Informática Agropecuária)
Rafael Geraldeli Rossi (USP)
Renata Ribeiro do Valle Gonçalves (Cepagri-Unicamp)
Renato Fileto (UFSC)
Ricardo Rodrigues Ciferri (UFSCar)
Ricardo da Silva Torres (IC-Unicamp)
Roberto Higa (Embrapa Informática Agropecuária)
Rodrigo Bezerra de Araújo Gallis (UFU)
Samuel de Assis Silva (UFES)
Sandro da Silva Camargo (Universidade Federal do Pampa)
Santiago Vianna Quadra (Embrapa Informática Agropecuária)
Sergio Manuel Serra da Cruz (UFRRJ)
Silvio Roberto Medeiros Evangelista (Embrapa Informática Agropecuária)
Sônia Ternes (Embrapa Informática Agropecuária)
Thiago Teixeira Santos (Embrapa Informática Agropecuária)
Tiago Carvalho (IFSP)
Vilmar Rodrigues Moreira (PUC-PR)
Willingthon Pavan (UPF)

COMISSÃO DE INICIAÇÃO CIENTÍFICA

Presidente: **Maria Fernanda Moura** (Embrapa Informática Agropecuária)
Vice-Presidente: **Luiz Henrique A. Rodrigues** (Feagri-Unicamp)

Adauto Luiz Mancini (Embrapa Informática Agropecuária)
Adhemar Zerlotini (Embrapa Informática Agropecuária)
Alan Massaru Nakai (Embrapa Informática Agropecuária)
Alexandre Coutinho (Embrapa Informática Agropecuária)
Ana Carolina Lorena (Universidade Federal de São Paulo)
Carla Geovana Macário (Embrapa Informática Agropecuária)
Daniel Castro Victoria (Embrapa Informática Agropecuária)
Evandro Bacarin (Universidade Estadual de Londrina)
Jansle Vieira Rocha (Feagri-Unicamp)
Jaudete Daltio (Embrapa Monitoramento por Satélite)
João Camargo Neto (Embrapa Informática Agropecuária)
João Francisco Gonçalves Antunes (Embrapa Informática Agropecuária)
Kleber Xavier Sampaio de Souza (Embrapa Informática Agropecuária)
Marcos César Visoli (Embrapa Informática Agropecuária)
Maria Fernanda Moura (Embrapa Informática Agropecuária)
Mauricio Alvarenga Mudadu (Embrapa Informática Agropecuária)
Ricardo da Silva Torres (IC- UNICAMP)
Ricardo Marcondes Marcacini (Universidade Federal Mato Grosso Sul)

Ronaldo Cristiano Prati (Universidade Federal do ABC)
Stanley R. de M. Oliveira (Embrapa Informática Agropecuária)

Comissão do Concurso de Aplicativos

Presidente: **Débora Pignatari Drucker** (Embrapa Informática Agropecuária)
Vice-Presidente: **Tiago Carvalho** (IFSP)

Andreiwid S. Côrrea (IFSP)
Beatriz Pedrosa (IFSP)
Carlos Eduardo Beluzo (IFSP)
Fabio Feliciano (IFSP)
Henrique G. Salvador (IFSP)
Samuel Botter Martins (IFSP)
Virginia Costa Duarte (TIC em Foco)

Comissão do SBIAgro Conect@

Ariovaldo Luchiari Junior (Embrapa Informática Agropecuária)
Fernando Matsuura (Embrapa Produtos e Mercados)
Junia Rodrigues de Alencar (Embrapa Informática Agropecuária)
Luiz Manoel Silva Cunha (Embrapa Informática Agropecuária)

Assessoria de Comunicação

Flávia Bussaglia Fiorini (Embrapa Informática Agropecuária)
Maria Giulia Croce (Embrapa Informática Agropecuária)
Nadir Rodrigues Pereira (Embrapa Informática Agropecuária)

PALESTRANTES

Maurício Antônio Lopes
Embrapa
Brasil

Bruno Tisseyre
Montpellier SupAgro
França

Abel Packer
SciELO/FAPESP
Brasil

Fedro Zazueta
Universidade da Flórida
EUA

Filipe Neves Santos
INESC TEC
Portugal

PREFÁCIO

Estes anais contêm o texto completo dos trabalhos apresentados no XI Congresso Brasileiro de Agroinformática (SBIAgro 2017), o qual foi promovido pela Embrapa Informática Agropecuária e pela Faculdade de Engenharia Agrícola, Instituto de Computação e pelo Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura da Universidade Estadual de Campinas (Unicamp). Esta edição do evento foi realizada no Centro de Convenções e na Casa do Lago da Unicamp, localizados na cidade de Campinas (SP). O propósito do evento foi o de reunir pesquisadores, professores, estudantes, empresários e funcionários de empresas para discutir o tema da informática aplicada à agricultura, além de promover um ambiente propício para o surgimento de novos relacionamentos, projetos e negócios.

O SBIAgro 2017 recebeu 129 trabalhos completos, dos quais 71 foram aceitos para apresentações orais. A taxa de aceitação de 55% garantiu que os trabalhos selecionados fossem de alta qualidade, o que foi confirmado pelo ótimo nível das apresentações realizadas ao longo do evento. Além disso, foram submetidos 15 trabalhos de iniciação científica, dos quais 7 foram aceitos, uma taxa de 47% de aceitação. Por fim, 8 trabalhos foram submetidos e 7 foram aceitos para o Desafio de Inovação SBIAgro 2017, evento em forma de Concurso de Aplicativos Móveis, promovido como incentivo para o surgimento de novas soluções de interesse ao agronegócio. Cada submissão de artigo completo e de Iniciação Científica foi analisada por dois revisores de reconhecida experiência nas respectivas áreas, no formato *double blind*.

O programa técnico do congresso contou ainda com três painéis de discussão e com cinco palestras realizadas por pesquisadores reconhecidos internacionalmente: Maurício Antônio Lopes (Embrapa, Brasil), Bruno Tysseire (Montpellier SupAgro, França), Abel Packer (Scielo, Brasil), Fedro Zazueta (University of Florida, EUA) and Filipe Neves Santos (INESC TEC, Portugal). A participação desses palestrantes contribuiu para elevar ainda mais a qualidade do evento. Por fim, é importante mencionar a realização do Conect@, no qual cerca de 100 representantes de várias empresas ligadas à tecnologia da informação e ao agronegócio se reuniram para debater as tendências do setor e construir novos relacionamentos e negócios.

A Comissão Organizadora do SBIAgro 2017 agradece aos autores por suas submissões e participação no evento. E, também, agradece aos membros do comitê científico por sua valiosa contribuição na seleção dos trabalhos. Expressamos também nossa gratidão aos palestrantes convidados, cujas excelentes apresentações nos brindaram com novos conhecimentos a respeito de variados temas de interesse específico e geral. Estendemos um agradecimento especial a toda equipe de suporte, cuja incansável dedicação tornou possível a realização de um evento do qual todos podem se orgulhar.

Para finalizar, estendemos à comissão organizadora do SBIAgro 2019, a se realizar em Vitória da Conquista (BA), sinceros votos de sucesso, ressaltando que nossa experiência na organização do Congresso Brasileiro de Agroinformática pode ser compartilhada para contribuir na superação dos desafios que surgirão ao longo do processo.

SUMÁRIO

Dia 3 de outubro

Sessão 1 – Mineração de dados

9 *Alessandro Stachak, Fábio dos Santos, Giancarlo Rodrigues, Alaine Margarete Guimarães, Eduardo Fávero Caires, José Carlos Ferreira da Rocha, Arion de Campos Júnior*
Predição do teor alcoólico de Vinho Verde por meio de técnicas de Mineração de Dados

19 *Antônio David Viniski, Adriano Haliski, Alaine Margarete Guimarães, Eduardo Fávero Caires*
Regressão de vetores de suporte em mineração de dados para estimativa de produtividade de grãos

29 *Fernando Maia da Mota, Klauss Machareth de Souza, Renato Porfirio Ishii, Rodrigo da Costa Gomes*
BovReveals: uma plataforma OLAP e data mining para tomada de decisão na pecuária de corte

Sessão 2 – Aplicativos

41 *Deyvison Bruno Morais de Abreu, Ícaro Ramires Costa de Souza, Igo Romero Costa de Souza, Leonardo Barreto Campos, Carlos Eduardo Cugnasca*
AgrônomoBot: um ChatBot para exibição de dados de uma RSSF EKO aplicado a agricultura de precisão

51 *Marcos Filipe Alves Salame, Paulo Igor Moraes da Silva, Allex de Lima Sousa*
Ferramenta móvel para gestão cronológica das atividades de manejo das culturas do guaraná e da banana

59 *Glauber José Vaz, Daniel Rodrigo de Freitas Apolinário, Jorge Luiz Correa, Isaque Vacari, Luis Eduardo Gonzales, Débora Pignatari Drucker, Joice Machado Bariani, Sílvio Roberto Medeiros Evangelista, Luciana Alvim Santos Romani*
AgroAPI: criação de valor para a Agricultura Digital por meio de APIs

Sessão 3 – Processamento de Imagens e Visão Computacional

69 *Jayme Garcia Arnal Barbedo*
Automatic Image-Based Detection and Recognition of Plant Diseases – A Critical View

79 *Gustavo Ribeiro Teixeira, Victor Mesquita Ferreira, Juliana Martins de Mesquita Matos, Alda Mieko Rocha Kimura Vidal, Flávio de Barros Vidal*
Contagem automática de produção frutífera utilizando técnicas de visão computacional e aprendizagem de máquina

89 *Thiago Teixeira Santos, Luis Henrique Bassoi, Henrique Oldoni, Roberto Luvisutto Martins*
Automatic grape bunch detection in vineyards based on affordable 3D phenotyping using a consumer webcam

Sessão 4 – Aprendizado de máquina I

99 *Márcio Nicolau, Marcia Barrocas Moreira Pimentel, Casiane Salete Tibola, José Mauricio Cunha Fernandes, Willingthon Pavan*
Fusarium Damaged Kernels Detection Using Transfer Learning on Deep Neural Network Architecture

109 *Monique Pires Gravina Oliveira, Luiz Henrique Antunes Rodrigues*
How does atypical weather affect sugar content estimates obtained by machine learning?

119 *João Paulo da Silva, Jurandir Zullo Jr., Luciana Alvim Santos Romani*
Active learning e sua aplicação no monitoramento da cana-de-açúcar utilizando o algoritmo *SVM*

Concurso de Iniciação Científica

129 *Bruno José Alves, Rachel Scrivani, Sílvio Roberto Medeiros Evangelista, Luciana Alvim Santos Romani*
Agritempo GIS: Apoio para a tomada de decisão no campo

139 *Paulo Igor Moraes da Silva, Marcos Filipe Alves Salame*
Solução móvel de viabilidade econômica para os custos de produção do guaraná e avaliação de usabilidade com a ferramenta *Google TestLab*

149 *Luiz Rodolfo Machado, Antônio David Viniski, Pablo Henrique Ferreira, Alaine Margarete Guimarães, Eduardo Fávero Caires*
Seleção de atributos agressiva para estimativa de teor de nutrientes em soja utilizando espectroscopia NIR

159 *Amanda do Prado Mattos, Adriana Terumi Itako, João Batista Tolentino Júnior*
Determinação da severidade de *Septoria lycopersici* em folhas de tomateiro por meio de imagens digitais

169 *David Patrick Zampier, Antônio David Viniski, Gislaine Gabardo, Alessandro Stachak, Alaine Margarete Guimarães, Maristella Dalla Pria*
Influência da resolução espacial em imagens obtidas por RPA na estimativa de doenças na soja

179 *Matheus Agostini Ferraciolli, Felipe Ferreira Bocca, Luiz Henrique Antunes Rodrigues*
Efeito da autocorrelação espacial na avaliação de modelos empíricos de produtividade na cana-de-açúcar

187 *Rafaella Pironato Amaro, Felipe Ferreira Bocca, Diego Della Justina, Luiz Henrique Antunes Rodrigues*
Melhorias no processamento dos dados de um espectrômetro de baixo custo para uso agrícola

Sessão 5 – Modelagem e simulação I

197 *Urbano G. P. Abreu, Lisiane D. de Lima, Fernando H. M. A. R. de Albuquerque, Juan D. F. Souza*
Modelagem da eficiência, fator importante na rentabilidade da ovinocaprinocultura, dos produtores do Semiárido

207 *Bruno Henrique Pereira, Cristina Freitas Bazzano, Luis Gustavo Barioni, Adauto Luiz Mancini, Márcio Nicolau*
Uma Aplicação Desktop Interativa para Desenvolvedores e Usuários de Modelos de Sistemas Dinâmicos em R

217 *Cristina Freitas Bazzano, Bruno Henrique Pereira, Luis Gustavo Barioni, Adauto Luiz Mancini, Márcio Nicolau*
SDsim: Um Pacote para Modelagem e Simulação de Sistemas Dinâmicos Contínuos em R

Sessão 6 – Rede de Sensores

227 *João Felipe Resende Nacer, Irineu Sotoma*
Análise de protocolos de roteamento em Redes de Sensores sem Fio para Monitoramento Bovino

237 *Igo Romero Costa de Souza, Leonardo Barreto Campos, Carlos Eduardo Cugnasca*
Integração de Redes de Sensores Sem Fio eKo a Sistemas de Informação para Rastreabilidade do Vinho

247 *Bruno Rogora Kawano, Renata Maria Marér, Roberto Fray da Silva, Carlos Eduardo Cugnasca*
Rastreabilidade da cadeia do vinho brasileiro: um modelo baseado em tecnologia de radiofrequência e rede de sensores sem fio

Dia 4 de outubro

Sessão 7 – Aprendizado de máquina II

257 *José Leonardo Dos Santos Costa, Diego Saqui, Tiago Gonçalves Botelho, Paulo César dos Santos*
Análise de agrupamento de dados climatológicos e de plantio de alface baseada no algoritmo K-Means

267 *Stanley R. de M. Oliveira, Urbano G. P. Abreu, Maria do Carmo Ramos Fasiaben, Luis Gustavo Barioni, Helano Póvoas de Lima, Maxwell Merçon Tezolin Barros Almeida, Octávio Costa de Oliveira*
Identificação de padrões tecnológicos do sistema de pecuária de corte desenvolvido no Cerrado

277 *Henrique Lima Cará de Oliveira, Stanley Robson de Medeiros Oliveira, José Eduardo Boffino de Almeida Monteiro*
Metodologia baseada em Algoritmos de Aprendizado de Máquina para geração de séries espaço-temporais de precipitação

Sessão 8 - Modelagem e simulação II

287 *Angela Mazzonetto, Carlos Amaral Hölbig, Marcelo Trindade Rebonatto, Willingthon Pavan*
Uma abordagem paralela para a execução do modelo de simulação de cultura CSM-Cropsim: Wheat

297 *Marina Silva Gomes, Ana Paula Lüdtke Ferreira, Vinícius do Nascimento Lampert*
Inferência da estimativa de produtividade do solo a partir de dados de produção pecuária

307 *Aline de Holanda Nunes Maia, Marcelo Gonçalves Narciso, Durval Dourado Neto*
Sistema para avaliação de risco de resistência de insetos a toxinas Bt expressas em culturas transgênicas

Sessão 9 - Instrumentação

317 *José Cláudio da Silva Junior, Maria Bernadete de Morais França*
Sensor de umidade do solo por pulsos de calor em ponteira simples utilizando transistor bipolar

327 *Antônio Carlos Loureiro Lino, Marcos Valério Gebra da Silva, Edson d'Avila, Denival da Costa Silva*
Desenvolvimento de um perfilômetro laser para a geração de modelos 3D aplicados à Engenharia Agrícola

335 *Edson d'Avila, Antonio Carlos Loureiro Lino, Inácio Dal Fabbro, Kirlian Kitzinger d'Avila*
Desenvolvimento de um instrumento medidor de área foliar

Sessão 10 - Automação

343 *Gustavo Carolino Girardi, Pedro Luis de Paula Filho, Itamar Iliuk, Daniel Rodrigues Blanco, Vitor Alex Alves de Marchi, Jonathan Galdino da Silva, Danilo Luiz Cavazim Neto*
Produção de *Agaricus blazei* em função de diferentes umidades acionadas automaticamente por sensores

353 *Adriano Ricardo de Abreu Gamba, Sérgio Francisco Pichorim*
Sistema Telemétrico Via RFID em UHF para Medição da Umidade do Solo

363 *Luiz Fernando Delboni Lomba, Leandro de Jesus, Hana Karina Salles Rubinsztejn, Luciano Gonda, Pedro Paulo Pires*
Sistema para Monitoramento dos Dados de Movimentação Bovina e Aferição dos Comportamentos

---

Concurso de Aplicativos Móveis

375 *Diego de Carvalho Neves da Fontoura, Diego Mengarda, Fabiane Nunes Prates Camargo, Leandro da Silva Camargo, Leonardo Bidese de Pinho*
Javalix: um sistema para acompanhamento de malhadas de javalis

377 *Arilson José de Oliveira Júnior, Silvia Regina Lucas de Souza*
Aplicativo móvel para análise de conforto térmico de ambientes

379 *Bruno José Alves, Elvis William de Oliveira Barbieri*
ClimaZone: Sistema integrado de zoneamento climático

381 *Matheus M. Leonardo, Carlos A. B. Ribeiro, Fabio A. Faria*
Fruit Fly Classification

383 *Marcel da Silva Camargo, Ana Paula Lüdtke Ferreira, Naylor Bastiani Perez*
Avaliação do Bem-estar Animal no Transporte de Gado de Corte

385 *Renan de Paula Rosa*
Cyber Leaf

387 *Bruno Queiroz Santos, Franciscone Luiz de Almeida Junior, Guilherme da Silva Teixeira, Matheus Galvão Ferreira, Gabriel Marques de Melo*
TomBanana – Um aplicativo para previsão de colheita da bananeira cv. Tropical

---

Sessão 11 - Geotecnologias e sensoriamento remoto I

389 *Cassiane Jrayj Bariani, Nelson Mario Victoria Bariani, José Rodrigo Fernandez Caresani, Roberto Dutra de Felice*
Análise multitemporal do arroz usando NDVI de superfície

399 *Gisela Maiara Silva, Ana Cláudia dos Santos Luciano, Karina Maria Berbert Bruno, Daniel G. Duft, Matheus Phelipe Miller Dias da Silva, Guilherme Martineli Sanches*
Análise comparativa entre técnicas de fusão de imagens Landsat-8 para delimitação de áreas de cana-de-açúcar

409 *Daniel de Castro Victoria, Santiago Viana Cuadra, Aryeverton Fortes de Oliveira*
Séries temporais de índices de vegetação para identificação de quebra de safra

---

Sessão 12 - Tecnologias web

419 *Ligia Francielle Borges, Claudio Leones Bazzi, Kelyn Schenatto, Eduardo Godoy de Souza, Fernando Nakayama de Queiroz*
Aplicação web para criação de mapas temáticos em agricultura de precisão

429 *Fábio Danilo Vieira, Danilo Gomes de Moura, Diego Félix da Silva, Roberto Hiroshi Higa, Adhemar Zerlotini*
BDGF: um sistema web para recuperação de informação de genótipos e fenótipos

439 *Marcelo Gonçalves Narciso, Silvando Carlos da Silva, Fábio Fernandes Noleto, Márcia Gonzaga de Castro Oliveira, Luciene Fróes Camarano de Oliveira, Elcio Perpétuo Guimarães*
Ferramenta Web para acesso aos períodos de semeadura de arroz de sequeiro e feijão comum

Dia 5 de outubro

Sessão 13 - Visualização e Processamento de Dados

449 *Allan Silva Salles, Hugo Portocarrero, Aluísio Granato de Andrade, Rulian Rodrigues*
Realidade Aumentada como Ferramenta de Apoio a Estudos Pedológicos

459 *Edson Luis Bolfe, Marcos A. G. Pena Jr., Elisio Contini, Fabiano Mariath, Giani Silva*
Dados da agricultura brasileira: apoio ao planejamento estratégico

Sessão 14 - Geotecnologias e sensoriamento remoto II

469 *Silvia Ribeiro Mantuani, Monica Cristine Scherer Vaz, Maria Salete Marcon Gomes Vaz*
Integração de Dados Geoespaciais na Rastreabilidade de Grãos

479 *Silvia Ribeiro Mantuani, Denise do Rocio Maciel, Monica Cristine Scherer Vaz, Maria Salete Marcon Gomes Vaz*
Indicadores Geográficos para o Framework de Rastreabilidade de Grãos

489 *Ana Paula dos Santos Braatz Vieira, Rafael Stoffalette João, Luciana Alvim Santos Romani, Marcela Xavier Ribeiro*
ST-Vis: Uma ferramenta para visualização espaço-temporal em regiões de plantio de cana-de-açúcar

Sessão 15 - Imageamento aéreo I

499 *Renan Cunha, Luan Casagrande, Yuri Crotti, Rodrigo Maciel, Roderval Marcelino*
Identificação de falhas no plantio de soja através da utilização de veículos aéreos não-tripulados (VANTs)

507 *Luan Casagrande, Yuri Crotti, Renan Cunha, Roderval Marcelino, Rodrigo Maciel*
Sistema de sugestão de densidade para plantações de banana utilizando veículos aéreos não tripulados

517 *Ederval Pablo Ferreira da Cruz, Rodolfo Oliveira*
Análise de viabilidade no uso de Multi-VANTs em ambientes agrícolas através da avaliação de protocolos de roteamento

Sessão 16 - Bibliometria

527 *Sandro da Silva Camargo, Rebeca Einhardt Fiss, Fabiane Nunes Prates Camargo, Leandro Bochi da Silva Volk, José Pedro Pereira Trindade*
Um Panorama da Colaboração Científica na Pesquisa Agroinformática Brasileira

537 *Maria Fernanda Moura, Celina Maki Takemura, Ivan Lucas Cabral Silva, Luisa Miyashiro Tápias, Carolina Tavares de Oliveira, Luis Henrique Bassoi, Stanley Robson de Medeiros Oliveira*
Metodologia para a construção de portfólios tecnológicos agrícolas a partir de publicações técnico-científicas

Sessão 17 - Geotecnologias e sensoriamento remoto III

547 *César Huegel Richa, Leandro B. S. Volk, José Pedro P. Trindade, Leonardo Bidese de Pinho*
OpenSLID: Uma ferramenta capaz de tornar mais eficientes os estudos sobre a dinâmica dos campos

557 *Nelson Mario Victoria Bariani, Cassiane Jrayj de Melo Victoria Bariani*
Resolution of NDVI for phenology monitoring of irrigated rice in Southwestern Brazil

567 *Cassiane Jrayj de Melo Victoria Bariani, Nelson Mario Victoria Bariani*
Sensitivity of NDVI for remote sensing monitoring of soybean stages under irrigation in Southern Brazil

---

Sessão 18 - Imageamento aéreo II

577 *Leandro Rosendo Candido, Lúcio André de Castro Jorge, Maximiliam Luppe*
A Importância do GSD na Construção do Mosaico de Imagens Aéreas

585 *Lucas Diogo de Mendonça, Cesar G. Chaves, Luiz Eduardo Vicente, Eduardo Rodrigues de Lima*
A Low-Cost Six Band Multi-Spectral Camera Platform for In-Flight Near Real-Time Vegetation Index Computation and Delivery

595 *Alessandro Stachak, Jessyca Kawana Franquitto, Selma Regina Aranha Ribeiro, Alaine Margarete Guimarães, Antônio David Viniski, Eduardo Fávero Caires*
Amenização de ruídos de reflexão da radiação eletromagnética em imageamento aéreo

---

Sessão 19 - Ontologia e Classificação de Textos

605 *Ana Carolina Clivatti. Ferronato, André Gomes de Lima, Aline Pires Vieira de Vasconcelos, Henrique Rego Monteiro. da Hora, Rogério Atem de Carvalho*
Um modelo de Ontologia para o Agronegócio

615 *Flavio M. M. Barros, Stanley R. de M. Oliveira*
Avaliação de métodos de detecção de tópicos em pré-processamento para classificação de textos agrícolas

625 *Heitor Magaldi, Regina Braga, Wagner Arbex, Mariana Magalhães Campos, Carlos Cristiano Hasenclever Borges, José Maria N. David, Fernanda Campos*
Uso de ontologias e técnicas de visualização no apoio às pesquisas em eficiência alimentar de gado de leite

---

Sessão 20 - Geotecnologias e sensoriamento remoto IV

639 *Eduardo Antonio Speranza, Ricardo Rodrigues Ciferri*
Integração de Ferramentas de SIG e Mineração de Dados para Utilização em Atividades de Gestão Espacialmente Diferenciada Aplicada na Agricultura de Precisão

653 *Joel Zubek Rosa, Selma Regina Aranha Ribeiro, Alaine Margarete Guimarães, Carlos Hugo Rocha*
Distorções de valores altimétricos obtidos com modelos numéricos do terreno gerados com GNSS e RPA

663 *Camila de Matos Alonso, Ana Paula Lüdtke Ferreira, José Pedro Pereira Trindade, Alexandro Gularte Schafer*
Classificação do tipo de ocupação do solo por meio de valores espectrais de imagens de satélite

---

Sessão 21 - Análises e avaliações

673 *Martha Delphino Bambini, Roberto Hiroshi Higa, Maria Beatriz Machado Bonacelli*
Direcionadores tecnológicos da Pesquisa Agropecuária Intensiva em Dados: mapeamento de competências, ferramentas e infraestrutura

683 *Henike Guilherme Jordan Voss, Sergio Luiz Stevan Jr.*
Farming 4.0: Estado da Arte e Potenciais Futuros

693 *Junia Rodrigues de Alencar, Ana Paula da Silva Dias, Fernando Attique Maximo, Miguel Sardenha Francisco*
Agência Embrapa de Informação Tecnológica (Ageitec): avaliação de seus impactos econômicos

Sessão 22 - Processamento de Imagens e Visão Computacional II e Ciência de Dados

705 *Daniele Candido da Costa, Juliana Sanches, Antonio C. L. Lino, Edson D'Ávila, Patrícia Cia*
Uso da técnica de silhueta para a estimativa do volume de ameixa "Reubennel"

713 *Kleber Xavier Sampaio de Souza, Sônia Ternes, Stanley R. de M. Oliveira, Maria Fernanda Moura, Luis Gustavo Barioni, Roberto Hiroshi Higa, Maria do Carmo Ramos Fasiaben*
A prospective study on the application of Data Science in agriculture

Sessão 23 - Agronomia

723 *Cinthia Cabral da Costa*
Sistema para recomendação de adubação

733 *Laura Beatriz Alves, Francisco José Arnold, Elaine Cristina Catapani Poletti, Eduardo Tavares Costa*
Revisão de técnicas de caracterização de leite e derivados com uso de ultrassom

743 *Samira Luns Hatum de Almeida, Alexandre Rosa dos Santos, Gleissy Mary Amaral Dino Alves dos Santos, Samuel Ferreira da Silva, José Romário Carvalho*
Risco Potencial De Lixiviação De Herbicidas Na Cultura Do Café Conilon (*Coffea canephora*)

Sessão 24 - Redes sociais no agronegócio

753 *Alessandro Mancio de Camargo*
Modos de troca cognitiva nas redes sociais do agrossistema digital

763 *Leonardo Mariano Gravina Fonseca, Denis Teixeira da Rocha, Carolina Rodrigues Pereira, Vanessa Maia Aguiar de Magalhaes, Luiz Ricardo da Costa, Rita de Cássia Bastos Souza, Rosangela Zoccal*
RepiLeite: uso de rede social temática para compartilhamento de conhecimento

773 **Lista de autores**



Predição do teor alcoólico de Vinho Verde por meio de técnicas de Mineração de Dados

Alessandro Stachak[1], *Fábio dos Santos*[1], *Giancarlo Rodrigues*[1], *Alaine Margarete Guimarães*[2], *Eduardo Fávero Caires*[3], *José Carlos Ferreira da Rocha*[2], *Arion de Campos Júnior*[2]

[1] Programa de Pós-graduação em Computação Aplicada, Universidade Estadual de Ponta Grossa
Ponta Grossa, Paraná, Brasil
alessandro.sbk@gmail.com, fah_ds@live.com, giancarlops2@gmail.com

[2]Departamento de Informática, Universidade Estadual de Ponta Grossa
Ponta Grossa, Paraná, Brasil
alainemg@uepg.br, jrocha@uepg.br, arion@uepg.br

[3]Departamento de Ciência do Solo e Engenharia Agrícola, Universidade Estadual de Ponta Grossa
Ponta Grossa, Paraná, Brasil
efcaires@uepg.br

RESUMO

Verificar o teor alcoólico do vinho após o processo de fabricação possibilita averiguar a ocorrência de problemas durante sua fabricação, distinguir quanto à sua designação ou então conferir se apresenta o teor alcoólico mínimo exigido por lei. Atualmente esse procedimento é realizado por meio de métodos laboratoriais custosos ou por equipamentos específicos, os quais demandam esforço, tempo e mão-de-obra qualificada, além de apresentarem elevado custo de aquisição. Neste trabalho é apresentada uma alternativa para essa tarefa, baseada em técnicas de mineração de dados. Para esse fim, as técnicas de Regressão Linear (RL) e Máquinas de Vetores de Suporte para Regressão (SVR) foram aplicadas sobre 10 atributos físico-químicos do Vinho Verde tinto a fim de estimar seu teor alcoólico. A ferramenta SynthesisFS foi utilizada para apoiar a seleção de atributos e identificar aqueles com maior influência na determinação do teor alcoólico. Os resultados foram submetidos à estatística pertinente, a qual comprovou a dissemelhança entre o coeficiente de determinação da RL e SVR, tendo a SVR um desempenho superior em todos os casos. Por fim, os resultados foram comparados aos métodos já empregados na determinação do teor alcoólico do vinho e comprovaram que o método aqui apresentado é de igual ou superior acurácia.

PALAVRAS-CHAVE: Aprendizado de Máquina, Máquina de Vetores de Suporte para Regressão, Regressão Linear.

ABSTRACT

Checking the wine alcohol content after winemaking process allows to verify the occurrence of problems during its manufacture, distinguish as to their designation or to verify if it presents the minimum alcohol content required by the law. Nowadays this procedure is performed either by costly laboratorial methods or by specific equipments which require effort, time, skilled labor and have a high acquisition cost. This paper presents an alternative based on data mining techniques for that task. Linear Regression (LR) and Support Vector Regression (SVR) techniques were applied on 10 physical-chemical attributes of *Vinho Verde* red wine in order to estimate its alcohol content. To support the attributes selection and identify those with greater influence in alcohol content determination the SynthesisFS tool was used. The results were submitted to the pertinent statistic, which proved the dissimilarity between the LR and SVR determination coefficient results, having SVR a better performance in all cases. At last, the results were compared against already employed methods to determine the wine alcohol content proving that the proposed method offers an equals or greater accuracy.

KEYWORDS: Machine Learning, Support Vector Regression, Linear Regression.

INTRODUÇÃO

O vinho é uma bebida quimicamente complexa obtida através da fermentação alcoólica de uvas ou de seu mosto. Sua composição química se dá, de maneira geral, por água, alcoóis, açúcares, polissacarídeos, ácidos e fenóis, os quais variam de concentração conforme o fruto utilizado[1] e o processamento[2] realizado durante a fabricação (ALPUIM, 1997). Essa variação de concentração possibilita a fabricação de vinhos com diferentes sabores, texturas e aromas. Alguns métodos de medição da concentração de compostos físico-químicos são empregados durante a fabricação da bebida a fim de assegurar a qualidade do produto e o cumprimento das normas regulamentadoras locais[3] (SOLEAS et al., 1997; CORTEZ et al., 2009).

Um dos critérios avaliados nas medições é o teor alcoólico, o qual é determinado principalmente pela concentração de Etanol na bebida (JACOBSON, 2006). O Etanol é o mais importante e abundante álcool do vinho, sendo seu segundo constituinte maioritário e essencial para estabilidade, envelhecimento e propriedades sensoriais da bebida, o qual tem sua concentração determinada principalmente pelo estado de maturação das uvas utilizadas, os níveis de açúcares fermentáveis, a levedura aplicada durante a fermentação e também pela temperatura durante o

[1] A maturação e a composição química do fruto são influenciadas pela temperatura, incidência solar, nutrientes do solo e da planta, volume de água utilizado na irrigação e produtividade da planta (SOLEAS et al., 1997).

[2] Desengace, esmagamento, maceração, prensagem, fermentação, clareamento, filtragem e envase são os principais processos realizados durante a fabricação do vinho (SOLEAS et al., 1997).

[3] No Brasil, a Lei nº 7.678 de 8 de Novembro de 1988 e seus decretos dispõem sobre a produção, circulação e comercialização do vinho e derivados da uva enquanto que a Portaria nº 229 de 25 de Outubro de 1988 do Ministério da Agricultura complementa os padrões de identidade e qualidade do vinho.

processamento (ALPUIM, 1997; SOLEAS et al., 1997). Verificar a concentração de Etanol após a fabricação permite averiguar a ocorrência de problemas de diluição, distinguir os vinhos quanto à sua designação e checar se o produto final apresenta o teor alcoólico mínimo exigido (JACOBSON, 2006; KING; DUNN; HEYMANN, 2013).

A determinação do teor alcoólico pode ser feita por meio de fórmulas pré-estabelecidas, as quais utilizam os valores de outros atributos para realizar o cálculo, ou então por métodos aceitos pelas organizações regulamentadoras, como a Organização Internacional da Vinha e do Vinho (OIV) ou o Instituto Brasileiro do Vinho (IBRAVIM), sendo alguns deles (JACOBSON, 2006; NETO et al., 2015; OIV, 2017):

- Ebuliometria, o qual consiste na diminuição do ponto de ebulição da mistura de água e álcool, sendo que o ponto de ebulição da água representa o zero e o de ebulição do vinho determina o teor alcoólico.
- Cromatografia a gás, o qual baseia-se no movimento das moléculas até um detector, as quais quanto maiores ou mais pesadas demandam maior tempo para atingir o detector, possibilitando a identificação dos compostos através do tempo gasto e pelos pesos moleculares.
- Densitometria, o qual utiliza um oscilador de frequência ou uma balança hidrostática para determinar a densidade, a qual está correlacionada com o teor alcoólico da bebida.
- Espectroscopia, o qual baseia-se na lei de Lambert–Beer e determina o teor alcoólico por meio da quantidade de luz absorvida pela amostra em diferentes faixas de ondas do espectro.

Segundo Jacobson (2006) a ebuliometria é o método mais utilizado no mundo, razoavelmente fácil de executar, porém tende a ser substituído devido aos avanços tecnológicos; a cromatografia a gás permite mensurar apenas a concentração de etanol do vinho, mas é um sistema automatizado confiável para uma fábrica de médio porte, pois é necessário um controle temperatura preciso durante o processo mensuração, onde a identificação é feita de acordo com o peso molecular de cada elemento e a espectroscopia permite a obtenção precisa, rápida e fácil dos dados referentes a composição química do vinho, porém este método não é regulamentado por algumas organizações e, portanto, seu uso não está autorizado para a determinação do teor alcoólico. Já a densitometria é um processo custoso, exige esforço e desperdiça tempo quando executado manualmente, além de exigir um processo de calibração do equipamento que demanda conhecimento, já que o mesmo pode comprometer o resultado da medição, no entanto apresenta bons resultados, é o método de maior acurácia e já conta com métodos digitais que possibilitam uma menor chance de erro durante a fase de ajustes e análise dos resultados obtidos (JACOBSON, 2006; OLIVEIRA; FILHO; AFONSO, 2013; JIMÉNEZ-MÁRQUEZ et al., 2016).

Outros aspectos relacionados a tais métodos são o alto custo de aquisição de alguns dos equipamentos e os diversos procedimentos laboratoriais que precisam ser executados a fim de obter o teor alcoólico do vinho[4]. No entanto, a disponibilidade de dados referentes à composição

[4]Detalhes dos procedimentos podem ser vistos em OIV (2017).

físico-química do vinho possibilita o emprego de técnicas de Mineração de Dados (MD), as quais realizam a descoberta de conhecimento em bases de dados com o objetivo de evidenciar informações e padrões úteis que numa análise convencional seriam ignorados (TAN et al., 2006). Uma dessas técnicas é a regressão, a qual é apresentada por Tan et al. (2006) como a tarefa de aprender uma função alvo f que realiza um mapeamento de um conjunto de atributos x em uma variável de saída de valor contínuo y com um erro de predição mínimo.

Dentre as técnicas de regressão que podem ser utilizadas, destacam-se neste trabalho a Regressão Linear Multivariada (RL), a qual baseia-se na teoria estatística, e a Máquina de Vetores de Suporte para Regressão (SVR) (VAPNIK et al., 1997), a qual é uma versão adaptada da técnica de Máquina de Vetores de Suporte (SVM) (VAPNIK, 1963) utilizada em tarefas de classificação em MD. A escolha dessas técnicas se deve pela geração de modelos de predição, o que possibilita o ajuste posterior de seus parâmetros com o intuito de obter maior acurácia na predição. Assim, o objetivo deste trabalho foi investigar a aplicabilidade das técnicas RL e SVR na estimativa do teor alcoólico do *Vinho Verde* tinto a partir de dados de análise laboratorial de 10 características físico-químicas desta bebida. Além disso, essas técnicas foram comparadas estatisticamente a fim de identificar aquela com melhor desempenho de predição e também de acurácia, comparando com os métodos tradicionais já estabelecidos.

MATERIAL E MÉTODOS

Descrição da base de dados

A base de dados utilizada foi elaborada por Cortez et al. (2009)[5] e contém dados da composição físico-química do *Vinho Verde* tinto, produzido na região norte de Portugal, obtidos por meio de testes físico-químicos computadorizados em amostras do estágio final de produção, ou seja, quando já está pronto para o consumo. A base de dados original contém 1.599 instâncias e 12 atributos, porém dois atributos foram retirados, sendo eles o atributo de saída referente à qualidade do vinho foi desconsiderado por não se aplicar ao contexto deste trabalho e o teor alcoólico que é o atributo a ser estimado neste trabalho.

Os atributos apresentados na Tabela 1, junto com seus respectivos descritores estatísticos, permitem realizar uma análise pontual de suas dosagens comparando com a literatura com o intuíto de verificar possíveis problemas no processo de fabricação do vinho. Alpuim (1997) destaca a importância de cada um destes atributos para a fabricação do vinho, sendo que a acidez fixa influencia na concentração de Etanol pois gera uma percepção organoléptica diferenciada e está correlacionada com o processo de fabricação do vinho e a qualidade e variedade da uva empregada no preparo da bebida. Já a acidez volátil acrescenta um aroma mais picante e rançoso pelo fato dos ácidos acético e carboxílico da série acética estarem relacionados, sendo que a soma da quantidade destes dois atributos resulta na acidez total do vinho, e por fim o

[5]A base encontra-se disponível no repositório UCI *Machine Learning Repository* através do endereço https://archive.ics.uci.edu/ml/datasets/Wine+Quality.

ácido cítrico auxilia na conservação do vinho.

Tabela 1: Descritores estatísticos das 1.599 instâncias de características físico-químicas da base de dados de *Vinho Verde*, variante tinto.

ID	Atributo	Mínimo	Máximo	Média	D.Padrão	Variância
1	Acidez fixa (g(ácido tartárico)/dm^3)	4,6	15,900	8,32	1,741	3,031081
2	Acidez volátil (g(ácido acético)/dm^3)	0,12	1,580	0,528	0,179	0,032041
3	Acido cítrico(g/dm^3)	0	1	0,271	0,0195	0,000380
4	Açúcar residual(g/dm^3)	0,9	15,500	2,539	1,41	1,988100
5	Cloretos (g(Cloreto de sódio)/dm^3)	0,05	611	77,47	51,629	2665,554
6	Dióxido de enxofre livre(mg/dm^3)	1	72	15,875	10,46	109,4116
7	Total de dióxido de enxofre (mg/dm^3)	6	289,00	46,468	32,895	1082,081
8	Densidade (g/cm^3)	0,99	1,004	0,997	0,002	0,000004
9	pH	2,74	4,01	3,311	0,154	0,023716
10	Sulfatos (g(Sulfato de potássio)/dm^3)	0,33	2,00	0,657	0,17	0,028900
11	Teor alcoólico (vol. %)	8,4	14,90	10,423	1,066	1,136356

O açúcar possibilita a geração de álcool e dióxido de carbono, sendo que 90% da quantidade de açúcar aplicado produz estes componentes. A quantidade restante é o açúcar residual, sendo a densidade relacionada a este processo, pois alterações nos valores da quantidade de açúcar e do álcool irá gerar uma alteração no valor da densidade. A quantidade de cloretos e sulfatos varia de acordo com o solo em que a uva é plantada, pois o mesmo esta diretamente ligado a quantidade de sais minerais da área de plantio. O dióxido de enxofre é utilizado como conservante, dito como insubstituível em qualquer processo de conservação de qualquer tipo de vinho. O valor do pH tem influência direta no sabor do vinho, pois se o mesmo apresentar um valor elevado será produzido um vinho com uma sensação de paladar ”magra”e seca (ALPUIM, 1997).

Pré-processamento da base

Cortez et al. (2009) destacam que nem todos os atributos disponíveis na base de dados podem ser relevantes durante a execução das tarefas de MD. Portanto, um método de seleção de atributos se faz necessário. Esse método consiste na identificação dos atributos que mais influenciam o valor da variável de saída e a remoção daqueles irrelevantes, o que melhora o desempenho da tarefa de MD devido à redução de processamento, memória e dimensionalidade (TAN et al., 2006).

Neste trabalho a seleção de atributos foi realizada através do software *SynthesisFS*[6] (SANTOS, 2016), a qual realiza uma síntese de todos os algoritmos de seleção de atributos disponíveis na versão 3.6.12 da ferramenta *WEKA* (HALL et al., 2009). O *SynthesisFS* reconhece quais métodos de seleção de atributos são aplicáveis à base utilizada e, por meio de ranqueamento

[6]Disponível em http://www3.uepg.br/computacaoaplicada/SynthesisFS/.

com pesos entre os atributos, apresenta aqueles mais aptos a estimar a variável de saída (SANTOS, 2016). Os atributos são apresentados ordenados conforme o peso atribuído pela ferramenta na base utilizada, dos quais quanto maior o peso, mais influência o atributo tem sobre o teor alcoólico.

Por conta da base de dados apresentar apenas valores contínuos, os algoritmos *CfsSubsetEval, ClassifierSubsetEval, LatentSemanticAnalysis, PrincipalComponents, ReliefAttributeEval* e *WrapperSubsetEval* foram os disponibilizados para a seleção de atributos no software *SynthesisFS*. O *SynthesisFS* também permite priorizar o resultado da seleção de atributos de certos algoritmos, porém neste trabalho foi atribuído o peso 1 a todos os algoritmos disponibilizados, os quais foram executados utilizando conjunto de treinamento completo, ou seja, os resultados de seleção de cada algoritmo tem igual importância no computo da seleção final.

Algoritmos de regressão

As tarefas de regressão, bem como os testes estatísticos, foram executados na versão 3.3.3 da ferramenta R (R Core Team, 2014)[7] conforme o resultado fornecido pelo *SynthesisFS* quanto ao peso dos atributos, ou seja, primeiro a regressão foi executada com somente o atributo de maior peso, depois com o primeiro e o segundo, até executar com todos os atributos disponíveis já que os pesos atribuídos foram distintos (ver Tabela 2). A RL foi executada através do método *lm* e o SVM do tipo *eps-regression* (tipo de SVM para regressão) com o kernel não-linear *Gaussian RBF* (Função de Base Radial Gaussiana)(SMOLA; SCHÖLKOPF, 2004), disponível no pacote *e.1071*(DIMITRIADOU et al., 2011)[8] através do método *svm*. Ambas as técnicas fizeram uso da validação cruzada com 10 *folds*, onde a base é dividida em 10 partes, sendo uma para a geração do modelo e 9 para testes, o processo é repetido até as 10 partes serem utilizadas para criação do modelo.

Os coeficientes de determinação das técnicas de regressão foram comparados por meio do teste de hipóteses de Wilcoxon (WILCOXON, 1945), bilateral a 5% de significância, através do pacote *Modern Applied Statistics with S* (VENABLES; RIPLEY, 2013)[9], a fim de verificar se os resultados foram distintos e indicar qual técnica de MD obteve um melhor desempenho. Depois de comprovar que o SVR apresentou um resultado superior à RL, os valores do erro relativo de estimativa foram comparados por meio do teste de Tukey pós-hoc (TUKEY, 1949), univariável a 5% de significância, para verificar a existência de diferença significativa entre a quantidade de atributos utilizados na tarefa de regressão.

Por fim, a acurácia do nosso modelo foi comparada com a de outros métodos já consolidados e utilizados na determinação do teor alcoólico do vinho. O cálculo da acurácia é realizado através do somatório do módulo da subtração do valor predito pelo valor real dividido pelo número de instâncias da base, conforme ilustrado na equação 1, onde N é o número de

[7]Disponível em https://www.r-project.org/.
[8]Disponível em https://CRAN.R-project.org/package=e1071.
[9]Disponível em https://CRAN.R-project.org/package=MASS.

instâncias, y é o valor predito e x é o valor real contido na base de dados.

$$acuracia = \frac{\sum_{k=1}^{N} |y_k - x_k|}{N} \quad (1)$$

RESULTADOS E DISCUSSÃO

Os resultados obtidos através da ferramenta *SynthesisFS* em relação aos atributos mais influentes no teor alcoólico do vinho tinto são apresentados na Tabela 2. O atributo *Densidade* apresentou o maior peso na determinação do teor alcoólico do vinho tinto, concordando com o resultado apresentado por Neto et al. (2015). Isso se deve pela correlação negativa existente entre a densidade e a concentração de etanol no vinho, pois a água, quando livre de outros elementos, apresenta o valor 1,000g/cm^3 de densidade enquanto o etanol apresenta um valor aproximado de 0,720g/cm^3, então quanto menor a densidade do vinho maior será seu teor alcoólico (ALPUIM, 1997; NETO et al., 2015).

Tabela 2: Resultado da seleção de atributos executada no software *SynthesisFS* com ranqueamento em função dos pesos na base de vinho tinto. Quanto maior o valor do peso, mais influência o atributo possui na determinação do teor alcoólico.

Peso	Atributo
18,463	Densidade
11,3939	Acidez fixa
10,4478	pH
9,929	Açúcar residual
8,967	Acido cítrico
8,3508	Cloretos
7,4037	Sulfatos
6,656	Dióxido de enxofre total
2,225	Dióxido de enxofre livre

A Tabela 3 apresenta os coeficientes de correlação (r) e de Determinação (R^2), Erro Quadrático Médio ($RMSE$) e Erro Médio Absoluto (MAE) obtidos para estimativa do teor de álcool do vinho tinto por meio dos algoritmos RL e SVR, conforme o número de atributos utilizados na tarefa. O Teste de Wilcoxon aplicado sobre os valores de R^2, uma das medidas de avaliação do resultado obtido, comprovou com 95% de confiança a existência de diferença significativa[10] entre os resultados dos algoritmos avaliados, sendo superior o desempenho da SVR em todas as execuções independente do número de atributos. Tal superioridade pode ser atribuída ao fato de a SVR utilizar um *kernel* do tipo Gaussian RBF, o qual é um modelo não-linear para regressão e que tem obtido melhores resultados que o tipo de *kernel* linear, modelo este utilizado para dados lineares(TAN et al., 2006; SMOLA; SCHÖLKOPF, 2004; JAVED et

[10] *p-value* = 0,005889.

al., 2009). Em alguns casos a diferença de resultado foi expressiva entre os algoritmos, como quando utilizando apenas um atributo em que o SVR obteve um coeficiente de determinação de 0,37, aproximadamente 54% superior em relação à RL (0,24) (Tabela 3).

Tabela 3: Resultados das técnicas de regressão aplicadas nas 1.599 instâncias da base de vinho tinto com diferentes quantidades de atributos. Coeficientes de determinação rotulados com letras distintas são significantemente diferentes (Teste de Tukey pós-hoc, α = 0,05).

Quantidade	Regressão Linear (RL)				Support Vector Regression (SVR)			
de Atributos	*r*	R^2	RMSE	MAE	*r*	R^2	RMSE	MAE
1	0,494	0,244	0,928	0,747	0,614	0,377	$0{,}851_a$	0,640
2	0,613	0,375	0,840	0,674	0,716	0,513	$0{,}740_b$	0,546
3	0,710	0,505	0,747	0,594	0,777	0,603	$0{,}649_c$	0,473
4	0,787	0,620	0,654	0,502	0,852	0,727	$0{,}525_d$	0,381
5	0,795	0,631	0,644	0,497	0,857	0,735	$0{,}504_d$	0,367
6	0,794	0,631	0,643	0,498	0,863	0,745	$0{,}483_{de}$	0,351
7	0,794	0,630	0,643	0,498	0,875	0,767	$0{,}480_{de}$	0,347
8	0,811	0,658	0,619	0,473	0,874	0,765	$0{,}447_{ef}$	0,321
9	0,814	0,663	0,613	0,467	0,881	0,774	$0{,}428_f$	0,303
10	0,814	0,663	0,613	0,467	0,881	0,776	$0{,}425_f$	0,300

Conforme apresentado na Tabela 3, os resultados de correlação e erro de ambos os algoritmos melhoraram à medida que mais atributos foram adicionados, o que comprova o vínculo entre os atributos do vinho na determinação do teor alcoólico da bebida. Com o objetivo de identificar a quantidade ideal de atributos para se obter o melhor resultado possível, o teste de Tukey pós-hoc foi aplicado sobre os valores de RMSE do algoritmo SVR, por ter sido aquele que forneceu os melhores resultados, onde letras diferentes significam resultados estatisticamente distintos. Diante disso, constata-se com 95% de confiança de que não existe diferença significativa nos melhores resultados de correlação utilizando oito, nove ou dez atributos para o vinho tinto. Isso significa que utilizar apenas os oito primeiros atributos exibidos na Tabela 2 já é suficiente para obter o melhor resultado possível com uma acurácia média de $\pm$ 0,3 %v/v, pois a acurácia da ebuliometria é de $\pm$ 0,5 %v/v, da densitometria varia de 0,10 a 0,15 %v/v e da Cromatografia a Gás é de $\pm$ 0,2 %v/v (JACOBSON, 2006), resultado superior à ebuliometria, porém inferior a densitometria e cromatografia a gás.

Os resultados obtidos foram validados na base de dados de *Vinho Verde* branco, a qual é disponibilizada junto com a base de *Vinho Verde* tinto e que contém 4.898 instâncias dos mesmos atributos físico-químicos. Nessa base, utilizando 9 atributos, a SVR obteve R^2 de 0,91 e acurácia de $\pm$ 0,2 %v/v, resultado superior à ebuliometria e a cromatografia a gás, demonstrando a eficiência do método apresentado neste trabalho.

CONCLUSÕES

Este trabalho demonstrou que o emprego de tarefas de regressão em MD sobre os dados de atributos físico-químicos do *Vinho Verde* tinto permite estimar seu teor alcoólico com acurácia superior ao método da ebuliometria, porém inferior aos métodos da densitometria e cromatografia a gás. O software *SynthesisFS* utilizado durante o pré-processamento indicou que os atributos mais relevantes para determinação do teor alcoólico foram densidade, acidez fixa, pH, açúcar residual, ácido cítrico, cloretos, sulfatos e dióxido de enxofre total. Por sua vez, os testes estatísticos aplicados comprovaram que o algoritmo SVR apresentou melhor capacidade de estimativa do teor alcoólico do *Vinho Verde* tinto utilizando esses atributos. Como trabalho futuro, sugere-se que o método aqui proposto seja empregado em outras variedades de vinho buscando seu aprimoramento, visto que existem diferenças entre os processos produtivos e na qualidade das uvas conforme a região de produção, o que pode acarretar em alterações significativas nos resultados finais sobre qual atributo tem maior ou menor influência sobre o teor alcoólico.

AGRADECIMENTOS

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo suporte financeiro.

REFERÊNCIAS

ALPUIM, J. P. Aprendendo a química do vinho. *Boletim da Sociedade Portuguesa de Química, Lisboa*, n. 65, p. 13–27, 1997. Disponível em: <http://www.spq.pt/magazines/BSPQ/589>.

CORTEZ, P. et al. Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, Elsevier BV, v. 47, n. 4, p. 547–553, nov 2009.

DIMITRIADOU, E. et al. *e1071: Misc Functions of the Department of Statistics (e1071), TU Wien. R package version 1.5-27*. 2011.

HALL, M. et al. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, ACM, v. 11, n. 1, p. 10–18, 2009.

JACOBSON, J. L. *Introduction to wine laboratory practices and procedures*. [S.l.]: Springer Science & Business Media, 2006.

JAVED, F. et al. Rbf kernel based support vector regression to estimate the blood volume and heart rate responses during hemodialysis. In: IEEE. *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. [S.l.], 2009. p. 4352–4355.

JIMÉNEZ-MÁRQUEZ, F. et al. Optoelectronic sensor for measuring ethanol content during grape must fermentation using nir spectroscopy. *Microsystem Technologies*, v. 22, n. 7, p. 1799–1809, 2016. ISSN 1432-1858.

KING, E. S.; DUNN, R. L.; HEYMANN, H. The influence of alcohol on the sensory perception of red wines. *Food Quality and Preference*, v. 28, n. 1, p. 235–243, 2013.

NETO, F. S. P. P. et al. Effect of ethanol, dry extract and reducing sugars on density and viscosity of brazilian red wines. *Journal of the Science of Food and Agriculture*, John Wiley & Sons, Ltd, v. 95, n. 7, p. 1421–1427, 2015.

OIV. *Compendium of international methods of wine and must analysis*. International Organisation of Vine and Wine Paris, 2017. Vol. 1. Disponível em: <http://www.oiv.int/en/technical-standards-and-documents/methods-of-analysis/compendium-of-international-methods-of-analysis-of-wines-and-musts-2-vol>.

OLIVEIRA, B. d. M.; FILHO, J. A. M. M.; AFONSO, J. A. C. A densidade e a evolução do densímetro. *Revista Brasileira de Ensino de Física*, Scielo, v. 35, p. 1 – 10, 03 2013.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2014. Disponível em: <http://www.r-project.org>.

SANTOS, R. B. L. Método para Síntese de Técnicas de Seleção de Atributos Disponíveis no Software WEKA. *Universidade Estadual de Ponta Grossa*, p. 1–16, 2016.

SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. *Statistics and computing*, Springer, v. 14, n. 3, p. 199–222, 2004.

SOLEAS, G. J. et al. Wine as a biological fluid: history, production, and role in disease prevention. *Journal of clinical laboratory analysis*, v. 11, n. 5, p. 287–313, 1997.

TAN, P.-N. et al. *Introduction to data mining*. Boston: Pearson Education, 2006. 769 p. ISBN 9788131764633.

TUKEY, J. W. Comparing individual means in the analysis of variance. *Biometrics*, JSTOR, p. 99–114, 1949.

VAPNIK, V. Pattern recognition using generalized portrait method. *Automation and remote control*, v. 24, p. 774–780, 1963.

VAPNIK, V. et al. Support vector method for function approximation, regression estimation, and signal processing. *Advances in neural information processing systems*, Morgan Kaufmann Publishers, p. 281–287, 1997.

VENABLES, W. N.; RIPLEY, B. D. *Modern applied statistics with S-PLUS*. [S.l.]: Springer Science & Business Media, 2013.

WILCOXON, F. Individual comparisons by ranking methods. *Biometrics bulletin*, [International Biometric Society, Wiley], v. 1, n. 6, p. 80–83, 1945.

Regressão de vetores de suporte em mineração de dados para estimativa de produtividade de grãos

Antônio David Viniski[1], *Adriano Haliski*[2], *Alaine Margarete Guimarães*[3], *Eduardo Fávero Caires*[4]

[1] Mestrando do Programa de Pós-Graduação em Computação Aplicada, Universidade Estadual de Ponta Grossa, Ponta Grossa, Paraná, Brasil, davidviniski@gmail.com

[2] Doutorando do Programa de Pós-Graduação em Agronomia, Universidade Estadual de Ponta Grossa, Ponta Grossa, Paraná, Brasil, adrianohaliski@yahoo.com.br

[3] Departamento de Informática, Universidade Estadual de Ponta Grossa, Ponta Grossa, Paraná, Brasil, alainemg@uepg.br

[4] Departamento de Ciência do Solo e Engenharia Agrícola, Universidade Estadual de Ponta Grossa, Ponta Grossa, Paraná, Brasil, efcaires@uepg.br

RESUMO

Estimar safras agrícolas com precisão e antecipação é uma atividade complexa, em função da diversidade de solos, cultivares, tipos de manejo e variação climática. A Mineração de Dados (MD) vem emergindo em pesquisas agrícolas, sendo a tarefa de regressão frequentemente utilizada em diferentes estimativas. Tendo em vista que a técnica Máquina de Vetores de Suporte para Regressão (SVR) possui grande capacidade de generalização de modelos de regressão, o objetivo deste trabalho foi avaliar o desempenho da técnica SVR na mineração de dados de atributos químicos do solo para estimativa de produtividade de grãos. Para tanto, foram utilizados dados do solo das safras de milho e trigo entre os anos de 2004 a 2014, contendo os atributos: pH, teores de Al, Ca e Mg, saturação por bases (V%), saturação por Al (m%), %Ca e %Mg na capacidade de troca catiônica (CTC) a pH 7,0, e %Ca e %Mg na CTC efetiva (CTCe). Para análise do desempenho da SVR, foram também aplicados os algoritmos de rede neural artificial e regressão linear múltipla, todos disponíveis no software WEKA. A MD

dos atributos químicos do solo apresentou potencial de estimativa de produtividade, explicando até 72% de sua variação. Os percentuais de Ca e Mg na CTCe, apresentaram correlações com a produtividade maiores que os teores de Ca e Mg, melhorando a capacidade de estimativa dos modelos. A SVR se mostrou eficiente, obtendo resultados superiores aos obtidos pelas outras técnicas, alcançando correlações de 0,9 (*using training set*) e 0,85 (*cross validation*).

PALAVRAS-CHAVE: Mineração de dados, regressão, produtividade de grãos.

ABSTRACT

Estimating agricultural crops with precision and anticipation is a complex activity, due to the diversity of soil, cultivars, types of management, soil and climate. Data Mining (DM) has been emerging in agricultural research and the it's regression task has been used in different estimates. Considering that Support Vector Machine for Regression (SVR) technique has a large capacity for generalization of regression models, the goal of this work was to evaluate the SVR technique performance in mining data of soil chemical attributes to estimate grain yield. Therefore, the soil data of the corn and wheat crop (between the years 2004 and 2014) were used, containing the attributes: pH, concentrations of Al, Ca, and Mg, base saturation (V%), Al saturation (m%), %Ca and %Mg in the cation exchange capacity (CEC) at pH 7.0, and %Ca and %Mg in the effective CEC (ECEC). In addition to the SVR, artificial neural network and multiple linear regression algorithms also were applied. Mining data of soil chemical attributes was efficient in the productivity estimation, explaining up to 72% of its variation. The Ca and Mg percentages in the ECEC presented higher correlations with yield than the exchangeable Ca and Mg contents, improving the models estimation capacity. SVR was efficient, obtaining results superior to those obtained by other techniques reaching correlation of 0.9 (using training set) and 0.85 (cross validation).

KEYWORDS: Data mining, Regression, Crop yield.

INTRODUÇÃO

O conhecimento prévio e preciso das safras agrícolas é uma questão estratégica para os produtores, bem como para o país, tanto no planejamento do abastecimento interno, quanto para a orientação das ações referentes ao mercado externo (GOMES et al., 2014). Estimar safras agrícolas com precisão e antecipação é uma atividade complexa, em função da diversidade de cultivares encontradas no mercado, diferentes tipos de manejo,

tipos de solos ou clima que são encontrados no Brasil. Vários fatores referentes ao solo, à planta e à atmosfera interagem entre si determinando a produtividade das culturas agrícolas (ARAUJO et al., 2011).

Segundo Frizzone (2005), considerar todos estes fatores que influenciam na produtividade para geração de modelos de estimativa, aumenta a complexidade dos mesmos, tonando difícil a sua aplicação. Assim, é possível expressar a produtividade exclusivamente por um único fator, como por exemplo, as propriedades físico-químicas do solo, assumindo que os outros fatores inerentes à produtividade estejam fixos em nível ótimo. Diversas técnicas podem ser utilizadas para geração de modelos de regressão, sendo a mineração de dados (MD) frequentemente utilizada em diferentes áreas, devido à grande variedade de algoritmos e softwares disponíveis.

A MD, que é o processo de extração de informações úteis e importantes em grandes conjuntos de dados, vem emergindo em pesquisas relacionadas à agricultura, sendo frequentemente utilizada no estudo de características do solo, plantas e culturas (MUCHERINO et al., 2009). A regressão é uma tarefa de MD que pode ser utilizada em diversas estimativas agrícolas, como produtividade (GERKE, 2017; MERCANTE et al., 2010), índice severidade de doenças (HIKISHIMA et al., 2010; LACERDA, 2017) e teor de nutrientes no solo, grãos e folhas, dentre outras. No entanto, escolher a técnica que seja adequada para um conjunto de dados específico pode ser uma tarefa difícil.

Os métodos de regressão estatísticos tradicionais são frequentemente apresentados como os processos que derivam uma função f(x) que tenha o menor desvio entre as respostas previstas e observadas experimentalmente, para todos os exemplos de treino. A máquina de Vetores de Suporte para Regressão (SVR – do inglês *Support Vector Regression*), por sua vez, ao estimar um modelo de regressão tenta minimizar o erro generalizado ao invés do erro de treinamento observado, obtendo assim um desempenho generalizado (BASAK et al., 2007). A SVR é uma modificação da técnica de classificação conhecida como Máquina de Vetores de Suporte (SVM – do inglês *Support Vector Machine*), para resolução de problemas envolvendo regressão.

O objetivo deste trabalho foi realizar a estimativa da produtividade de grãos, por meio da mineração de dados de atributos químicos do solo, com foco específico no desempenho da SVR.

MATERIAL E MÉTODOS

Os dados deste trabalho foram coletados de um experimento manejado em sistema plantio direto de longa duração. O experimento foi realizado no período de maio de 2004 a maio de 2014, no município de Ponta Grossa, PR, na Fazenda Estância dos Pinheiros (25º 10’ S, 50º 05’ W), com altitude de 970 m e precipitação pluvial média anual de 1550 mm. O experimento foi constituído por quatro doses de calcário, quatro doses de nitrogênio e três repetições, totalizando 48 parcelas. O tamanho de cada parcela foi de 6,5 × 6,4 m. Mais detalhes a respeito da área experimental podem ser vistos na publicação de Caires et al. (2015). No presente estudo, avaliaram-se as culturas de milho, em 2004-2005 e 2008-2009, e trigo, em 2009, 2010 e 2014. Antes da semeadura das culturas de milho e trigo dessas safras, coletaram-se amostras de solo da camada de 0–05 cm para avaliação dos atributos químicos. Em virtude da variabilidade química do solo, foram coletadas 12 subamostras por parcela para constituir uma amostra composta de solo. Considerando que o experimento tinha 48 parcelas e que foram avaliadas duas safras de milho (2004-2005 e 2008-2009) e três safras de trigo (2009, 2010 e 2014), a base de dados de solo foi constituída por 96 amostras referentes à cultura do milho e 144 amostras referentes à cultura do trigo. Os atributos químicos do solo analisados neste estudo foram: pH em $CaCl_2$, teores de alumínio (Al), cálcio (Ca) e magnésio (Mg), saturação por bases (V%), saturação por alumínio (m%), saturação por Ca e por Mg na capacidade de troca de cátions (CTC) a pH 7,0 do solo e saturação por Ca e por Mg na CTC efetiva (CTCe) do solo.

A avaliação de produtividade de grãos foi realizada por meio de colheita manual e trilhagem em máquina debulhadora estacionária em uma área útil de 6,0 m² de cada subparcela. A produtividade relativa de grãos de milho e trigo foi calculada tomando-se o rendimento máximo de grãos obtido em cada subparcela como referencial de 100%, dentro de cada ano de cultivo (HALISKI, 2015). Assim torna-se possível analisar juntamente as culturas de milho e trigo, nos diferentes anos de coleta das amostras de solo.

Seleção de atributos

A base de dados foi submetida a seleções de atributos levando em consideração o conhecimento do especialista, devido ao fato de alguns atributos serem apresentados em função da concentração de outros, como, por exemplo, o atributo m%, que corresponde ao percentual de saturação por Al. A seleção se deu então por meio do conhecimento do domínio. Assim, a Tabela 1 apresenta os três subconjuntos de atributos gerados, os quais

foram submetidos separadamente à mineração de dados, sendo atributos independentes, para a estimativa da produtividade relativa da soja (atributo dependente).

Tabela 1 – Subconjuntos de atributos gerados na seleção supervisionada.

Subconjuntos	Atributos
Subconjunto 1	pH, Al, Ca, Mg e V%
Subconjunto 2	V%, m%, saturação por Ca e por Mg na CTC a pH 7,0
Subconjunto 3	pH, V%, m%, saturação por Ca e por Mg na CTCe

Mineração de dados

A etapa de MD foi realizada no software WEKA (do inglês - *Waikato Environment for Knowledge Analisys*), ambiente que dispõe de algoritmos de seleção de atributos, classificação, associação, agrupamento, regressão dentre outras técnicas de descoberta de conhecimento em bases de dados. O WEKA tem implementado os algoritmos de SVM e SVR, sendo SMO (do inglês - *sequential minimal optimization*) e *SMOreg* (SHEVADE et al., 1999), respectivamente.

Existem diferentes *kernels* da SVR disponíveis no WEKA, sendo neste trabalho utilizado o *kernel* PUK (do inglês - *Pearson VII Universal Kernel*), que é uma variação do *kernel* universal da SVR, sendo incorporada a função *Pearson VII* (UESTUEN et al., 2006). Os demais parâmetros do algoritmo *SMOreg* foram mantidos no padrão do WEKA.

Para confrontar o desempenho da SVR, foram aplicados também o algoritmo *Multilayer Perceptron* de rede neural artificial (RNA), e o *Linear Regression* de regressão linear múltipla (RLM). O algoritmo *Multilayer Perceptron* foi executado com a configuração padrão do WEKA. O algoritmo *Linear Regression* foi modificado para que considerasse todos os atributos preditores.

A validação dos modelos de regressão foi realizada por meio das opções *Using Training Set* e *Cross Validation* do software WEKA. Na opção *Use Training Set*, a validação é feita com o mesmo conjunto de dados de treinamento. Já na opção *Cross Validation*, o modelo é avaliado por validação cruzada, na qual o conjunto de testes é dividido em partes iguais e a predição é aplicada a cada parte separadamente. Neste trabalho, a validação cruzada foi feita usando dez conjuntos (*folds*) repetidos em dez simulações conforme proposto por Witten e Frank (2005).

Como critério de avaliação das técnicas de regressão, foram utilizados o coeficiente de correlação de Pearson (r) e a raiz do erro quadrático médio (RMSE – do

inglês *Root mean Square Error*). Estas medidas são fornecidas pelo software WEKA após a execução dos algoritmos de regressão.

RESULTADOS E DISCUSSÃO

A estatística descritiva dos atributos químicos do solo, com valores de mínimo, máximo, média, desvio padrão (DP), coeficiente de variação (CV), assim como a correlação de Pearson (r) destes atributos com a produtividade relativa, são apresentados na Tabela 2. A maioria dos atributos apresenta CV alto, segundo a classificação de Pimentel Gomes (1995), para valores acima de 30%, indicando grande variação dos valores desses atributos em relação às suas médias. Somente os atributos pH e Al apresentam CV médio, com valores entre 10% e 20%. Verifica-se uma grande amplitude de valores para os atributos V%, m% e %Ca na CTCe, para os quais, assim como para o Al, são apresentados os maiores valores de correlação de Pearson com a produtividade relativa (Tabela 2).

Os atributos do solo apresentaram altas correlações lineares (r) com a produtividade relativa das culturas de milho e trigo, sendo o menor valor (r = 0,67) obtido para pH e Mg, e o maior valor (r = -0,80) para Al e m% (Tabela 2). A correlação negativa com o teor de Al e m% já era esperada, pois o aumento na acidez no solo implica em aumento no teor e na saturação por Al, com consequente redução na produtividade. A correlação de 0,67 entre a produtividade relativa (milho e trigo) e o pH está em concordância com Reichert et al. (2008) que constataram correlação direta entre a produtividade de grãos de feijão e o pH, com coeficiente de correlação de 0,69.

Tabela 2 – Estatística descritiva dos atributos químicos utilizados no presente estudo e sua correlação com a produtividade relativa

Atributos do solo	Mínimo	Máximo	Média	[2]DP	[1]CV(%)	[3]r
pH	3,60	6,70	4,97	0,75	15	0,67
Al	0,00	22,00	4,25	6,02	14	-0,80
Ca	2,00	56,00	29,21	15,36	53	0,68
Mg	1,00	55,00	17,22	9,62	56	0,67
V%	3,60	79,40	43,68	21,73	50	0,71
m%	0,00	80,90	13,83	21,83	158	-0,80
%Ca na CTC a pH 7,0	1,60	50,70	25,71	14,05	55	0,69
%Mg na CTC a pH 7,0	0,80	39,70	15,02	8,33	55	0,71
%Ca na CTCe	8,50	76,90	49,75	16,13	32	0,74
%Mg na CTCe	4,20	55,70	29,36	10,62	36	0,73

[1]CV (%) = Coeficiente de variação em percentual, [2]DP = Desvio Padrão, [3]r = correlação dos atributos com a produtividade

Resultados diferentes dos obtidos no presente estudo foram relatados em outros estudos para outras culturas. Souza et al. (2010) obtiveram correlações baixas com todos os atributos químicos do solo estudados na estimativa de produtividade de cana-de-açúcar. Durigon et al. (2009) também encontraram baixas correlações entre os atributos químicos do solo e a produtividade de arroz irrigado, sendo que os maiores coeficientes de correlação (r) obtidos foram de 0,38 e -0,38 para teor de Ca e m%, respectivamente. Isso indica que as culturas de trigo e milho, estudadas no presente trabalho, refletiram de maneira mais significativa às alterações nos atributos químicos do solo.

A Tabela 3 apresenta os resultados encontrados com os algoritmos de MD utilizados nos três subconjuntos de atributos (Tabela 1) e com os diferentes métodos de avaliação. Pode-se verificar que, em comparação com os resultados apresentados na Tabela 2, a análise multivariada nos modelos de regressão possibilitou um ganho significativo nos coeficientes de correlação encontrados.

Tabela 3 – Correlação e erro obtidos na estimativa de produtividade relativa utilizando os algoritmos de MD

Subconjuntos de dados	Método de Avaliação	*Linear Regression*		*Multilayer Perceptron*		*SMOreg*	
		r	RMSE	r	RMSE	r	RMSE
Subconjunto 1	*Using Training Set*	0,82	13,27	0,84	12,80	0,87	11,24
	Cross Validation	0,81	13,62	0,76	15,52	0,82	13,33
Subconjunto 2	*Using Training Set*	0,83	12,96	0,85	11,98	0,86	11,98
	Cross Validation	0,82	13,25	0,77	14,67	0,83	12,94
Subconjunto 3	*Using Training Set*	0,85	12,09	0,87	13,12	0,90	10,27
	Cross Validation	0,84	12,39	0,80	14,37	0,85	12,20

Dalchiavon et al. (2011) obtiveram r = 0,47 (R^2 = 0,221) utilizando a regressão múltipla para estimar a produtividade de feijoeiro por meio de atributos químicos do solo. De acordo com os autores, a análise de regressão múltipla indicou que 22% da variação da produtividade de grãos do feijoeiro de inverno foram atribuídos à variação nos atributos químicos do solo [fósforo (P), matéria orgânica, pH, potássio (K), Ca, Mg, acidez potencial (H+Al), soma de bases (S), CTC a pH 7,0 e V%]. Com isso, considerando os valores de correlação obtidos pela SVR (algoritmo *SMOreg* - Tabela 3), a técnica apresentou resultados consideravelmente superiores, 0,90 (R^2 = 0,81) e 0,85 (R^2 = 0,72) para os métodos de avaliação *using training set* e *cross validation*, respectivamente. Levando em consideração que a validação cruzada possui maior rigor na avaliação dos modelos de regressão, a SVR se mostrou com maior capacidade de generalização, indicando que 72% da variação de produtividade foram atribuídos à

variação nos atributos químicos do solo: pH, V%, m%, %Ca na CTCe e %Mg na CTCe (Subconjunto 3 – Tabela 1).

Dos diferentes subconjuntos de atributos presentes na Tabela 1, o Subconjunto 3 apresentou coeficientes de correlação superiores em todos os algoritmos de MD utilizados para estimativa de produtividade. Isso indica que considerar os valores de saturação por Ca e Mg na CTCe aumenta a capacidade preditiva dos modelos de regressão. Essa informação pode ser reforçada com base nos valores de correlação de Pearson (r) presentes na Tabela 2, sendo 0,75 para %Ca na CTCe e 0,73 para %Mg na CTCe, enquanto para os teores de Ca e Mg os valores foram 0,68 e 0,67, respectivamente.

Embora os valores de correlação obtidos em ambas as técnicas utilizadas tenham sido significativos, a SVR apresentou valores de correlação superiores, tanto para os diferentes subconjuntos de atributos, quanto para os diferentes métodos de avaliação. Já a rede neural apresentou valores maiores que a regressão linear múltipla, quando utilizado o método de avaliação *using traning set*, o que indica uma capacidade inferior da rede em generalizar o modelo de estimativa. A regressão linear múltipla, por sua vez, quando utilizada com o método de avaliação C*ross Validation*, apresentou desempenho semelhante a SVR.

CONCLUSÕES

A mineração de dados dos atributos químicos do solo estudados apresentou potencial de estimativa de produtividade, tendo gerado modelos de regressão capazes de explicar até 72% da variação da produtividade das culturas de trigo e milho.

Os percentuais de Ca e Mg, quando avaliados na CTCe apresentaram correlações com a produtividade maiores que os teores de Ca e Mg no solo. Assim, quando utilizados na estimativa de produtividade, por meio da regressão multivariada, possibilitaram a melhoria na capacidade de estimativa dos modelos.

A SVR se mostrou eficiente para realização da estimativa de produtividade de grãos por meio de atributos químicos do solo, obtendo resultados superiores aos obtidos pelas técnicas de rede neural artificial e regressão linear múltipla.

AGRADECIMENTOS

A Capes, pelas bolsas de mestrado e de doutorado.

REFERÊNCIAS

ALES, V. T. O algoritmo sequential minimal optimisation para resolução do problema de Support Vector Machine: uma técnica para reconhecimento de padrões. 2008. Dissertação (Mestrado em Ciências da Computação). Universidade Federal do Paraná, Curitiba, 2008.

ASLANARGUN, A., MAMMADOV, M., YAZICI, B., & YOLACAN, S. Comparison of ARIMA, neural networks and hybrid models in time series: tourist arrival forecasting. Journal of Statistical Computation and Simulation, v.77, n. 1, p. 29–53, 2007.

ARAUJO, M. A.; SOUZA, J. L. M.; TSUKAHARA, R. Y. Modelos agro-meteorológicos na estimativa da produtividade da cultura da soja na região de Ponta Grossa, Estado do Paraná. Acta Scientiarum. Agronomy, v. 33, n. 1, p. 23-31, 2011.

BASAK, D.; PAL, S.; PATRANABIS, D. C. Support vector regression. Neural Information Processing. v. 11, n. 10, p. 203-224, 2007.

CAIRES, E. F.; HALISKI, A.; BINI, A. R.; SCHARR, D. A. Surface liming and nitrogen fertilization for crop grain production under no-till management in Brazil. European Journal of Agronomy, v. 66, p.41–53, 2015.

DALCHIAVON, F. C.; CARVALHO, M. P.; FREDDI, O. S.; ANDREOTTI, M.; MONTANARI, R. Variabilidade espacial da produtividade do feijoeiro correlacionada com atributos químicos de um Latossolo Vermelho Distroférrico sob sistema de semeadura direta. Bragantia, Campinas, v. 70, n. 4, p. 908-916, 2011.

DURIGON, R.; SCHLOSSER, J. F.; ALEXANDRE RUSSINI, A.; DORNELLES, M. E. C.; PINHEIRO; E. D. Correlações entre atributos químicos do solo e atributos da cultura e da produtividade de arroz irrigado determinadas com técnicas de manejo localizado. Ciência Rural, v.39, n.9, 2009.

FRIZZONE, J. A.; ANDRADE JÚNIOR, A. S.; SOUZA, J. L. M.; ZOCOLER, J. L. Planejamento de irrigação: análise e decisão de investimento. Brasília: Embrapa. p. 283-314. 2005.

GERKE, T. Mineração de dados de imagens obtidas com aeronave remotamente pilotada para estimativa de produtividade do trigo. 2017. Dissertação (Mestrado em Computação Aplicada) – Setor de Ciências Agrárias e de Tecnologia, Universidade Estadual de Ponta Grossa, Ponta Grossa, 2017.

GOMES, A. C. S.; ROBAINA, A. D.; PEITER, M. X.; SOARES, F. C.; PARIZI, A. R C. Modelo para estimativa da produtividade para a cultura da soja. Ciência Rural, v.44, n.1, 2014.

HALISKI, A. Atributos químicos do solo e produtividade da soja influenciados pela calagem superficial e adubação nitrogenada no sistema plantio direto. 2015. Dissertação (Mestrado em Agronomia) – Setor de Ciências Agrárias e de Tecnologia, Universidade Estadual de Ponta Grossa, Ponta Grossa, 2015.

HAYKIN, S. Neural Networks: a comprehensive foundation. Segunda Edição b. Prentice Hall, 1999.

HIKISHIMA, M.; CANTERI, M. G.; GODOY, G. V., KOGA1, L. J.; SILVA, A. J. Quantificação de danos e relações entre severidade, medidas de refletância e produtividade no patossistema ferrugem asiática da soja. Tropical Plant Pathology, vol. 35, n. 2, p. 096-103, 2010.

LACERDA, V. S. Estimativa do índice de severidade de ferrugem asiática na cultura da soja por meio de imagens obtidas com aeronave remotamente pilotada. 2017. Dissertação (Mestrado em Computação Aplicada) – Setor de Ciências Agrárias e de Tecnologia, Universidade Estadual de Ponta Grossa, Ponta Grossa, 2017.

MERCANTE, E.; LAMPARELLI, R. A. C.; URIBE-OPAZO, M. A.; ROCHA, J. V. Modelos de regressão lineares para estimativa de produtividade da soja no Oeste do Paraná, utilizando dados espectrais. Revista Engenharia Agrícola, Jaboticabal, v.30, n.3, p.504-517, 2010.

MUCHERINO, A.; PAPAJORGJI, P.; PARDALOS, P. A survey of data mining techniques applied to agriculture. Operational Research, v. 9, n. 2, p. 121–140, 2009.

PIMENTEL-GOMES, F. Curso de estatística experimental. São Paulo: Esalq, 467 p, 1985.

UESTUEN, B.; MELSSEN, W. J.; BUYDENS, L. M. C. Facilitating the application of Support Vector Regression by using a universal Pearson VII function based kernel. Chemometrics and Intelligent Laboratory Systems. V. 81, n. 1, p. 29-40. 2006.

REICHERT, J.M.; DARIVA, T.A.; REINERT, D.J.; SILVA, V.R. Variabilidade espacial de Planossolo e produtividade de soja em várzea sistematizada: análise geoestatística e análise de regressão. Ciência Rural, v.38, p.981-988, 2008.

SHEVADE, S.K.; KEERTHI, S. S.; BHATTACHARYYA, C.; MURTHY, K. R. K. Improvements to the SMO Algorithm for SVM Regression. In: IEEE Transactions on Neural Networks, v. 11, n. 5, p. 1188-1193, 1999.

SOUZA, Z. M.; CERRI, D. G. P.; COLET, M. J.; RODRIGUES, L. H. A.; MAGALHÃES, P. S. G.; MANDONI, R. J. A. Análise dos atributos do solo e da produtividade da cultura de cana-de-açúcar com o uso da geoestatística e árvore de decisão. Ciência Rural, v.40, n.4, 2010.

VISCARRA ROSSEL, R. A.; WALVOORT, D. J. J.; MCBRATNEY, A. B.; JANIH, L. J.; SKJEMSTAD, J. O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma, v. 131, n. 1-2, p. 59-75, 2006.

WITTEN, I. H., FRANK, E. Data mining: Practical machine learning tools and techniques (2nd ed.). San Francisco: Elsevier (2005).



BovReveals: uma plataforma *OLAP* e *data mining* para tomada de decisão na pecuária de corte

Fernando Maia da Mota[1], *Klauss Machareth de Souza*[1], *Renato Porfirio Ishii*[2], *Rodrigo da Costa Gomes*[3]

[1]Associação Sul-Mato-Grossense de Produtores de Novilho Precoce – ASPNP
Campo Grande, Mato Grosso do Sul, Brasil
fernando@bovbi.com.br, klaussvet@terra.com.br

[2]Faculdade de Computação, Universidade Federal de Mato Grosso do Sul
Campo Grande, Mato Grosso do Sul, Brasil
renato@facom.ufms.br

[3]Embrapa Gado de Corte
Campo Grande, Mato Grosso do Sul, Brasil
rodrigo.gomes@embrapa.br

RESUMO

Observa-se na atividade da agropecuária regional esforços voltados para a produção de carne bovina com qualidade acima da média brasileira. A indústria da carne criou programas de incentivo à produção de carcaças bovinas de características específicas com bonificação para aquelas que se enquadram e descontos para as que não se enquadram. Problemas no processo de produção e de abate nas indústrias frigoríficas costumam ser o principal motivo de desclassificação de carcaças e, consequentemente, de perdas financeiras. Para lidar com esses problemas, neste artigo é proposta uma abordagem de análise de dados que utiliza de meios computacionais como armazém de dados (*Data Warehouse*), consultas analíticas online (OLAP) e mineração de dados (*Data Mining*) integradas e em conjunto com as habilidades humanas para servir de suporte à tomada de decisão na atividades relativas a pecuária de corte. A abordagem se divide em 4 etapas: 1) responsável pela extração, transformação e carga dos dados fornecidos pela Associação Sul-Matogrossense de Produtores de Novilho Precoce (ASPNP); 2) etapa de criação do modelo multidimensional para armazenagem dos dados; 3) etapa de visualização e exploração dos dados armazenados no *Data Warehouse*, é caracterizada pela configuração e personalização das ferramentas da suíte *open source* Pentaho; e 4) a aplicação de algoritmos de *Data Mining* por meio da ferramenta *open source* `Weka` com objetivo de descobrir padrões

e relacionamentos vinculados ao grau de acabamento e ao rendimento de carcaça. Os resultados alcançados até a terceira etapa na forma de um portal de análise de dados da ASPNP encontram-se em produção e constante evolução. Na quarta etapa, há indícios de que a adoção de algoritmos de *Data Mining* fornecem uma taxa média de acerto acima de 62% em relação a predição do grau de acabamento e do rendimento de carcaça, auxiliando o produtor na tomada de decisão no momento de definição do melhor momento para o abate.
PALAVRAS-CHAVE: Acabamento, Bonificação, Carcaça, Rendimento, Predição.

ABSTRACT

The activity of regional livestocking recently produced beef with quality above the Brazilian average. The meat industry created in last years incentive programs for the production of bovine carcasses with specific characteristics, paying bonus above the meat market average for those that has these characteristics and losses for those did not have. Problems in the production inside farm and slaughtering processes in the industries is often the main reason for the declassification of carcass and, consequently, financial losses. To fix these problems, this paper proposes a data analysis approach which applies computational data warehousing, online analytical queries (OLAP) and data mining integrated and jointly with the human skills to support the strategy decision activities related to livestocking segment. The proposed approach has 4 steps: 1) responsible for the extraction, transformation and loading of data provided by the Associação Sul-Matogrossense de Produtores de Novilho Precoce (ASPNP); 2) creation step of the multidimensional model for data storage; 3) visualization step and data exploration stored in the Data Warehouse, is characterized by the configuration and personalization of the open source Pentaho suite; and 4) the application of Data Mining algorithms through the open source tool Weka, with the objective to discover patterns and relationships in about the carcass coverage and the carcass yield. The results achieved until the third step are available in a Web portal for data analysis. This portal is in production and in constant evolution. In the fourth step, there are indications that the adoption of Data Mining algorithms provide an average hit rate above 62% in the prediction of carcass coverage and the carcass yield, assisting the producer in decision making at the best time for slaughter.
KEYWORDS: Finishing, Bonification, Carcass, Yield, Prediction.

INTRODUÇÃO

O agronegócio representa, aproximadamente, 20% do PIB brasileiro (BARROS et al., 2016), o que demonstra sua importância no contexto econômico-social do país. Este resultado é obtido em parte por uma pecuária forte, que responde por 1/3 desta riqueza e insere o Brasil nos primeiros lugares entre produtores e exportadores de carne bovina do mundo (USDA; FAS, 2017). Apesar de se mostrar uma potência mundial na produção de carne bovina, o produto brasileiro é considerado como de menor qualidade ao ser comparado a outros países tais como Uruguai, Argentina, Austrália e EUA. A consequência disso é que o produto brasileiro não alcança alguns mercados mais exigentes em qualidade, que por sua vez, remuneram melhor.

Diante deste contexto, observa-se na pecuária, esforços voltados para a produção de carne bovina com qualidade acima da média brasileira. Associações de produtores, cooperativas, indústria frigorífica e varejista têm estabelecido programas que oferecem bonificações a carcaças bovinas produzidas dentro de critérios específicos que elevam a qualidade da carne. A indústria da carne brasileira criou programas de incentivo à produção de carcaças com características específicas baseadas na normativa 9 de 04/05/2004 do Ministério da Agricultura, Pecuária e Abastecimento (RODRIGUES, 2004), com bônus para carcaças que se enquadram nessas características e descontos para as que não se enquadram. Essas características específicas são descritas por Bridi (BRIDI, 2002): sexo, peso de carcaça, idade ou maturidade fisiológica e grau de acabamento.

A definição das regras de bonificação e seus valores são definidos por cada programa instituído pela indústria frigorífica, de acordo com as necessidades de seus mercados. O resultado destes programas é o aumento na oferta de carne de melhor qualidade e também incremento na remuneração do produtor acima da média de mercado. Entretanto, tais programas enfrentam desafios complexos. O nível de adequação aos critérios dos programas das carcaças entregues à indústria frigorífica é muito dependente de um complexo conjunto de fatores ligados ao processo produtivo: genética, nutrição, gestão, preços de *commodities* agrícolas e de insumos, clima, qualidade de mão-de-obra, dentre outros. As variações e as inadequações de um ou mais fatores estão ligados a falhas em atender aos critérios de qualidade estipulados pelos programas de bonificação. Outro desafio é equilibrar os benefícios das bonificações em relação aos níveis de investimentos necessários no processo produtivo, de forma que a atividade seja atrativa financeiramente.

Ao considerar o crescimento deste tipo de programa e a evidência das complexidades inerentes, uma melhor compreensão das fragilidades e das potencialidades nos seus processos produtivos pode levar a ganhos em eficiência e qualidade. Neste sentido, técnicas de análise e mineração de dados podem ser de grande valia e sua aplicação pode ser considerada uma inovação neste contexto. Vários fatores podem interferir no rendimento dependendo do modo que se pesa o animal na fazenda: pesagem em jejum, descalibragem da balança, dieta hídrica, acabamento de carcaça, suplementação, etc. Por outro lado, na indústria, vários processos industriais podem interferir também no rendimento, tais como: contusões, oscilações de peso na hora da aferição da carcaça e limpeza de carcaça excessiva. Esta última afeta diretamente o grau de acabamento que é um dos critérios que mais penaliza as carcaças, uma vez que a idade e o peso são mais facilmente controlados pelo produtor (CANCIAN et al., 2014). Atualmente, a bonificação das carcaças é realizada no instante do abate por uma análise visual humana, ou seja, de forma subjetiva, por técnicos da própria indústria, o que, normalmente, é alvo de críticas dos produtores (ALLEN; FINNERTY, 2001).

A existência de conjuntos de dados que relacionam informações de abate e que estão sob domínio de associações ligadas a estes tipos de programas podem representar uma grande oportunidade para aplicação de técnicas computacionais capazes de descobrir padrões e gerar

conhecimento a partir de grandes volumes de dados. Técnicas de análise e mineração, do ponto de vista prático, fornecem um paradigma poderoso para exploração e descoberta de informação útil, comumente referenciado na literatura como consultas analíticas Online (do Inglês *On-line Analytical Processing* – OLAP) e *Data Mining* (do Inglês *Data Mining*) respectivamente.

Neste artigo é apresentado o `BovReveals`, uma abordagem de análise de dados que utiliza ferramentas computacionais tais como armazém de dados (do Inglês *Data Warehouse* – DW), consultas OLAP e Mineração de Dados integradas e em conjunto com as habilidades humanas para servir de suporte à tomada de decisões no segmento da pecuária de corte. Esta abordagem tem por objetivo fornecer suporte às associações de produtores, cooperativas, indústria frigorífica e varejista como uma forma de apoio na tomada de decisões, auxiliando na busca e descoberta de conhecimento embutido nos dados e, consequentemente, elevação da qualidade de carcaça no grau de acabamento esperado e na consolidação do rendimento.

Neste contexto, os objetivos específicos são:

- Construção de um DW com dados de abate de gado de corte fornecido pela Associação Sul-Mato-Grossense de Produtores de Novilho Precoce – ASPNP;
- Disponibilização de um portal Web para acesso aos dados armazenados no DW de forma fácil e eficaz por meio de consultas OLAP, painéis gráficos e relatórios;
- Aplicação de algoritmos de mineração para predição do momento de abate ideal de animais, observando os pré-requisitos de qualidade da carne preestabelecidos no que se refere a grau de acabamento e rendimento de carcaça.

MATERIAL E MÉTODOS

A abordagem proposta neste artigo divide-se em quatro etapas, conforme apresentada na Figura 1.

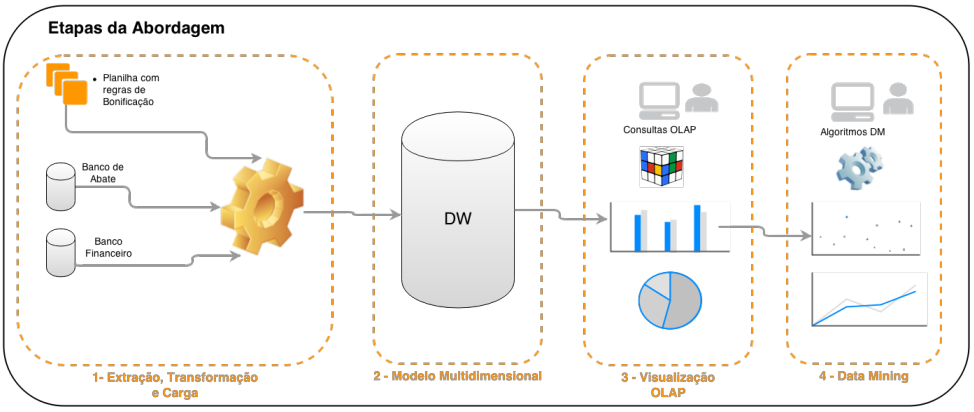


Figura 1: Etapas da abordagem `BovReveals`

Na etapa 1 de extração, tratamento e carga, são extraídos os dados das três fontes de dados utilizadas neste trabalho e armazenados no SGBD PostgreSQL: banco de dados de abate, banco de dados financeiros e planilha de regras técnicas. A partir disso são transformados, integrados,

calculados e por fim armazenados no DW. Foi utilizada a ferramenta *Pentaho Data Integration* (KETTLE, 2017), que permite ao desenvolvedor de ETL (*Extract, Transform and Load*) conectar-se a diversos bancos de dados, extrair dados, copiar, transformá-los, combiná-los, apagá-los, atualizá-los, enviá-los a diversos locais, criar "*jobs*", enviar *emails*, acessar um servidor via ssh, ftp, realizar tratamentos de erros e tantas outras tarefas necessárias a um processo de ETL. A utilização do *Data Integration* permite integrar empresas e sistemas de uma forma visual e organizada (SOUZA, 2010).

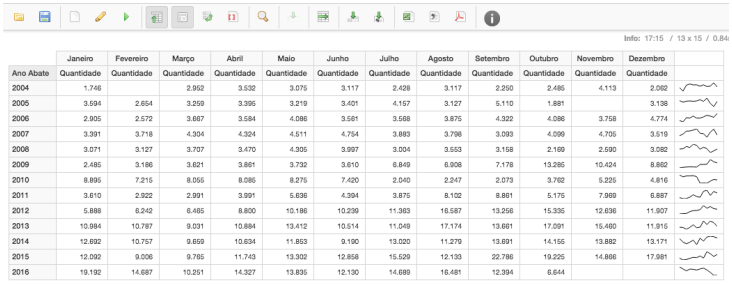
Na etapa 2 de modelagem multidimensional, é definido o modelo multidimensional que armazena os dados processados pela etapa 1. Nessa etapa são definidas dimensões, medidas e fatos, construindo assim o modelo estrela, caracterizando desta forma o DW, que também foi armazenado no SGBD PostgreSQL. Nesse modelo, foram definidas doze tabelas dimensões, doze medidas e oito dimensões degeneradas. Dimensão degenerada (ou do Inglês *Degenerate Dimension*) é uma dimensão derivada da tabela fato e, consequentemente, não tem sua própria tabela de dimensão (KIMBALL, 1996).

Na etapa 3 de visualização, as ferramentas para acesso aos dados do DW são configuradas. Nessa etapa é conduzida a configuração e a personalização do portal Web com o auxílio de ferramentas da suíte `Pentaho` para acesso e consumo dos dados armazenados no DW. As ferramentas utilizadas nesta etapa são:

- *Business Intelligence Server* (biserver), o biserver é uma aplicação Web que compõe a suíte de ferramentas `Pentaho`, disponibiliza aos usuários um ambiente Web controlado, diretório de arquivos, ferramentas de administração e acesso dos usuários a ferramentas OLAP. Neste trabalho, o biserver tem a funcionalidade de centralizar o conteúdo e, por isso, também é denominado portal;
- *Schema Workbench*, esta ferramenta fornece suporte ao processo de criação do modelo lógico criado para o DW, ou seja, é nesta ferramenta que são mapeadas as tabelas dimensões e seus atributos, tabelas fatos e suas métricas e as chaves estrangeiras da tabela fato que se relacionam com as tabelas dimensões;
- *Saiku Analytics*, é uma ferramenta de visualização OLAP com a funcionalidade de *Drag and Drop* (arrastar e soltar) que opera como um *plugin* do biserver (BARBER; STOELLBERGER, 2008). Esta ferramenta OLAP consome o arquivo XML criado pelo *Schema Workbench* a fim de executar consultas ao DW e, consequentemente, apresentar ao usuário as dimensões e métricas disponíveis para o uso. Na Figura 2 é possível visualizar um exemplo de consulta OLAP com o *Saiku Analytics* sobre o cubo de Abate. Nessa análise, as dimensões "Ano de Abate" e "Mês de Abate" são correlacionadas para uma análise geral de abates no período. Na Figura 3, as dimensões Acabamento, Maturidade, Sexo e Faixa de Arroba são apresentadas pelo volume de animais abatidos no ano de 2015.

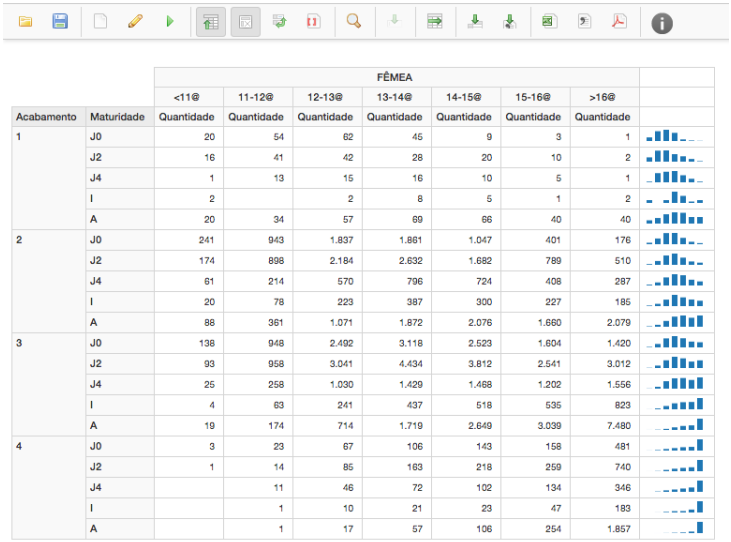
Na Figura 2 é possível visualizar um exemplo de consulta OLAP com o *Saiku Analytics* sobre o cubo de Abate. Nesta análise, as dimensões "Ano de Abate" e "Mês de Abate" são

confrontadas a fim de analisar os abates no período. Na Figura 3, as dimensões "Acabamento", "Maturidade", "Sexo" e "Faixa de Arroba" são apresentadas pelo volume de animais abatidos no ano de 2015.



	Janeiro	Fevereiro	Março	Abril	Maio	Junho	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro	
Ano Abate	Quantidade	Quantidade	Quantidade	Quantidade	Quantidade	Quantidade	Quantidade	Quantidade	Quantidade	Quantidade	Quantidade	Quantidade	
2004	1.746		2.952	3.532	3.075	3.117	2.428	3.117	2.250	2.485	4.113	2.062	
2005	3.594	2.654	3.259	3.395	3.219	3.401	4.157	3.127	5.110	1.881		3.138	
2006	2.905	2.572	3.667	3.584	4.086	3.561	3.568	3.875	4.322	4.086	3.758	4.774	
2007	3.391	3.718	4.304	4.324	4.511	4.754	3.883	3.798	3.093	4.099	4.705	3.519	
2008	3.071	3.127	3.707	3.470	4.305	3.997	3.004	3.553	3.158	2.169	2.590	3.082	
2009	2.485	3.186	3.621	3.861	3.732	3.610	6.849	6.908	7.178	13.285	10.424	8.862	
2010	8.895	7.215	8.055	8.086	8.275	7.420	2.040	2.247	2.073	3.762	5.225	4.816	
2011	3.610	2.922	2.991	3.991	5.636	4.394	3.875	8.102	8.861	5.175	7.969	6.887	
2012	5.888	6.242	6.485	8.800	10.196	10.239	11.383	16.587	13.256	15.335	12.636	11.907	
2013	10.984	10.787	9.031	10.884	13.412	10.514	11.049	17.174	13.661	17.091	15.460	11.916	
2014	12.692	10.757	9.659	10.634	11.853	9.190	13.020	11.279	13.691	14.155	13.882	13.171	
2015	12.092	9.006	9.785	11.743	13.302	12.858	15.529	12.133	22.786	19.225	14.866	17.981	
2016	19.192	14.687	10.251	14.327	13.835	12.130	14.689	16.491	12.394	6.644			

Figura 2: Análise geral de volume abatido por período



		FÊMEA						
		<11@	11-12@	12-13@	13-14@	14-15@	15-16@	>16@	
Acabamento	Maturidade	Quantidade	Quantidade	Quantidade	Quantidade	Quantidade	Quantidade	Quantidade	
1	J0	20	54	82	45	9	3	1	
	J2	16	41	42	28	20	10	2	
	J4	1	13	15	16	10	5	1	
	I	2		2	8	5	1	2	
	A	20	34	57	69	66	40	40	
2	J0	241	943	1.837	1.861	1.047	401	176	
	J2	174	898	2.184	2.632	1.682	789	510	
	J4	61	214	570	796	724	408	287	
	I	20	78	223	387	300	227	185	
	A	88	361	1.071	1.872	2.076	1.660	2.079	
3	J0	138	948	2.492	3.118	2.523	1.604	1.420	
	J2	93	958	3.041	4.434	3.812	2.541	3.012	
	J4	25	258	1.030	1.429	1.468	1.202	1.556	
	I	4	63	241	437	518	535	823	
	A	19	174	714	1.719	2.649	3.039	7.480	
4	J0	3	23	67	106	143	158	481	
	J2	1	14	85	163	218	259	740	
	J4		11	46	72	102	134	346	
	I		1	10	21	23	47	183	
	A		1	17	57	106	254	1.857	

Figura 3: Análise de volume por critério técnico no ano de 2015

O uso de painéis gráficos e de relatórios é uma prática comum em ferramentas de *Business Intelligence* (BI) e caracterizam-se por fornecer um ambiente rico visualmente. Além disso, potencializa a habilidade humana na construção de formas de visualização que enriqueçam a experiência do usuário no consumo das informações. A partir do início da utilização da ferramenta OLAP, inicia-se um período de análise dos dados, resultando em alguns indicadores de alto valor estratégico para a associação e seus associados. A Figura 4 apresenta um exemplo deste painel com dados de abate do ano de 2015.

O relatório "Performance de Qualidade", apresentado na Figura 5, tem por objetivo fornecer um resumo técnico e financeiro dos animais abatidos. Assim, por meio do indicador "Índice de Classificação", o usuário tem a informação de performance em relação aos critérios básicos de classificação de carcaças, ou seja, maturidade, acabamento e peso.

Na etapa 4 de *Data Mining*, por meio dos dados organizados, centralizados, calculados e combinados do DW são aplicados seus resultados e avaliados a partir dos algoritmos

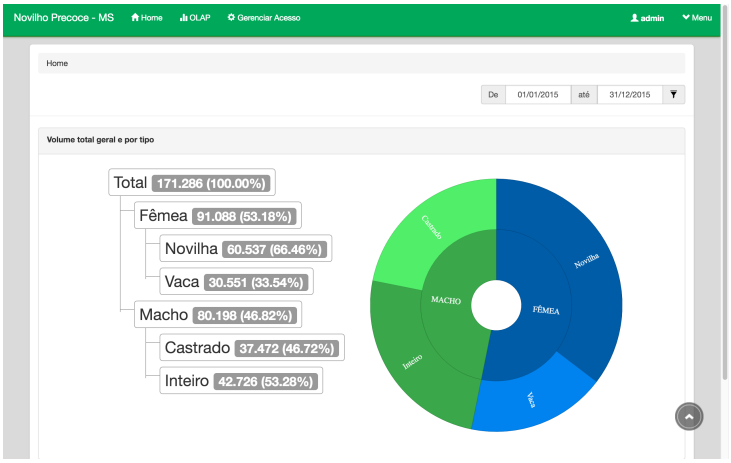


Figura 4: Painel Home com dados de abate do ano de 2015

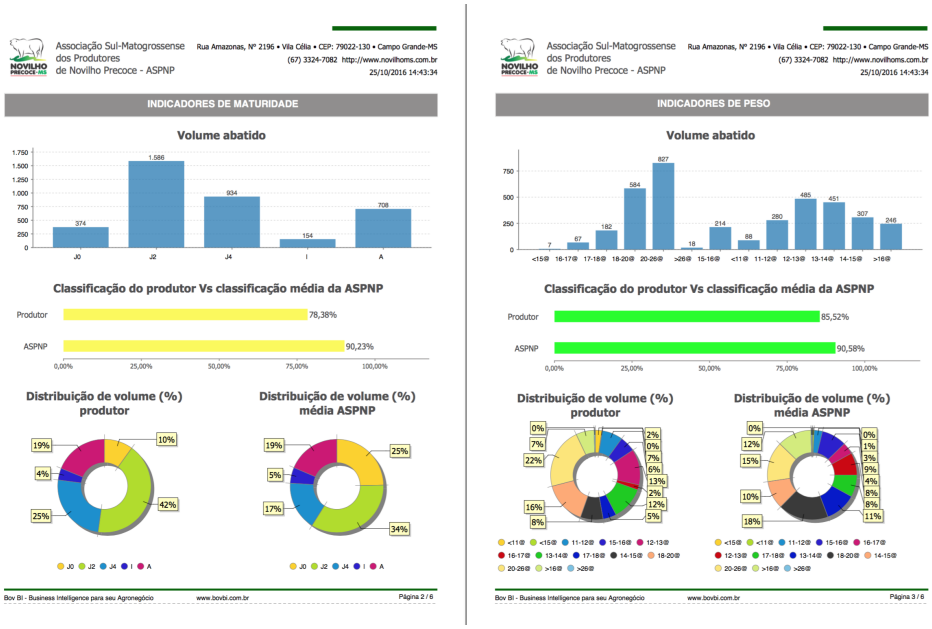


Figura 5: Indicadores técnicos do relatório Performance de Qualidade

de classificação e predição. Utilizou-se das implementações fornecidas pela ferramenta `Weka` (HALL, 2009). Duas perguntas básicas, guiaram toda a condução dos experimentos (discutidos na Seção Resultados):

- É possível identificar o acabamento de um animal antes do abate?
- É possível identificar o rendimento de carcaça de um animal antes do abate?

Os atributos utilizados foram: Sexo do animal, Peso de carcaça, Peso vivo individual, Idade do animal ou maturidade, Grau de acabamento e Raça. Além disso, as seguintes métricas de avaliação foram consideradas:

- Percentual de instâncias classificadas corretamente ou acurácia: é uma métrica que representa a taxa de exemplos positivos e negativos classificados corretamente;
- Precisão ou *Precision*: é um métrica que representa o número de verdadeiro positivos divididos pelo número de verdadeiro positivos acrescido do número de falso positivos,

conforme apresentado pela Equação 1;

$$\text{Precisão} = \frac{\text{verdadeiroPositivos}}{(\text{verdadeiroPositivos + falsoPositivos})} \quad (1)$$

- Sensitividade ou *Recall*: métrica que representa o número de verdadeiro positivos dividido pelo número de verdadeiro positivos acrescido do número de falso negativos, conforme apresentado pela Equação 2;

$$\text{Recall} = \frac{\text{verdadeiroPositivos}}{(\text{verdadeiroPositivos + falsoNegativos})} \quad (2)$$

- F-measure ou F-score: métrica que representa a harmonia entre *Precision* e *Recall*, conforme apresentado pela Equação 3;

$$\text{F-measure} = 2 * \frac{(\text{Precision * Recall})}{(\text{Precision + Recall})} \quad (3)$$

Outra característica que deve ser definida antes da apresentação dos resultados obtidos por este trabalho é a definição das classes. Portanto, para definição dos exemplos positivos e negativos com relação ao acabamento de carcaça, animais com gordura ausente (1) ou escassa (2) são considerados como exemplos negativos e animais com gordura mediana (3), uniforme (4) e excessiva (5) como exemplos positivos (SOUZA, 2014), já com relação ao rendimento de carcaça foram aplicadas as regras apresentadas pela Tabela 1, como regra para definir exemplos positivos e negativos.

Tabela 1: Regras de classe para rendimento de carcaça

Sexo	positivo (+)	negativo (-)
Macho	$\geq 54\%$	$< 54\%$
Fêmea	$\geq 51\%$	$< 51\%$

RESULTADOS E DISCUSSÃO

Na primeira estratégia de planejamento dos experimentos, os conjuntos de dados utilizam todo o histórico de abates disponível (1.142.230 registros de abate), ou seja, todos os abates de todos os associados e de todas as fazendas. Foram aplicados os algoritmos `J48` (C4.5), *Naive Bayes*, Redes Neurais e `svm`, todos em suas configurações padrões. Esta abordagem mais ampla gerou em média uma taxa de acerto de 69% para acabamento de carcaça e de 59% para rendimento de carcaça, após a execução de 100 rodadas de experimentos. Portanto, esta abordagem mais ampla não se mostrou satisfatória.

Uma outra estratégia foi considerar separadamente cada associado e suas respectivas fazendas, ou seja, contendo exclusivamente o histórico de abates referente aos seus animais abatidos. Com essa nova estratégia os experimentos apresentaram melhora nas métricas de avaliação.

Os resultados apresentados nas Tabelas 2 e 3 após a execução de 100 rodadas de experimentos, consideram dados de abate de um produtor selecionado de forma aleatória e com um total de 14.934 registros de animais abatidos. Neste cenário, a taxa de acerto média para grau de acabamento de carcaça foi de 76% e para rendimento de carcaça de 62%. Nesta estratégia, o total de animais foi de 14.934 (100%) da amostra e considera-se 30% para teste (4.480), ou seja, 3.406 animais foram classificados corretamente (segunda linha da Tabela 2).

Tabela 2: Resumo dos resultados para classificação de grau de acabamento de carcaça considerando o histórico de abates de um associado da ASPNP

Instâncias classificadas corretamente (acurácia)	
% média	76.03%
Quantidade média	3.406
Coeficiente de variação	0.05%
Precisão média	70%
Recall médio	59%
F-measure média	60%

Tabela 3: Resumo dos resultados para classificação de rendimento de carcaça considerando o histórico de abates de um associado da ASPNP

Instâncias classificadas corretamente (acurácia)	
% média	62.21%
Quantidade média	2.787
Coeficiente de variação	0%
Precisão média	63%
Recall médio	62%
F-measure média	61%

A partir dos resultados apresentados, conclui-se que segmentar os dados por produtor e fazenda é uma estratégia mais adequada em relação a estratégia que considera o uso conjunto do histórico de abates de todos os associados. Um dos motivos da melhora no uso dessa estratégia é a não existência de um sistema de produção porteira adentro uniforme entre todos os associados. Embora ocorra uma elevação na taxa de acerto utilizando a estratégia de segmentar por associado e fazenda, quando o classificador é aplicado a novos conjuntos de dados, o mesmo se mostrou inviável, ou seja, o classificador decorou os exemplos fornecidos, o que portanto, inviabiliza seu uso. O atributo mais relevante identificado nos experimentos foi o do peso vivo, que é armazenado no banco de dados de abate da ASPNP na forma de média de todos os animais de um lote abatido, ou seja, não é um dado preciso, este atributo é essencial para o cálculo de rendimento de carcaça e também influencia diretamente no grau de acabamento. Desta forma, é essencial que o peso vivo seja armazenado de forma individual por animal, assim como é realizado para os demais atributos de qualidade como grau de acabamento, maturidade e peso de carcaça.

CONCLUSÕES

O objetivo principal deste trabalho é elaborar, aplicar e avaliar uma abordagem de análise de dados para fornecer meios de extração de conhecimento útil a fim de suportar a tomada de decisão por qualquer organização envolvida no processo de produção de gado de corte no Brasil. A partir dos resultados discutidos neste artigo, nota-se a efetiva utilização de um portal Web de consumo e exploração de dados. Além disso, estão disponíveis ferramentas OLAP que possibilitam a qualquer usuário do portal, inclusive associados da ASPNP, exploração e avaliação de dados particulares, de forma simples e eficaz. Há também a disponibilização de painéis gráficos e relatórios, ou seja, indicadores de performance e qualidade que fornecem informação útil por meio de formas ricas de visualização, análise e consumo de dados relativos a abate. Por fim, há indícios de que por meio de algoritmos de *Data Mining*, com taxa média de acerto acima de 62%, é possível gerar informações que possam auxiliar o produtor na tomada de decisão no momento de definição do melhor momento para abate, no que diz respeito a rendimento de carcaça e grau de acabamento. Cabe ressaltar que os dados utilizados devem ser precisos e consistentes, ou seja, é altamente recomendado o armazenamento individual do peso vivo do animal e a utilização separada por produtor de conjuntos de dados no processo de aprendizado dos classificadores. Isso é necessário uma vez que não existe um sistema de produção uniforme entre os produtores, o que, portanto, ocasiona vários *outliers* na aplicação em conjunto no processo de aprendizagem.

AGRADECIMENTOS

Os autores agradecem o apoio da Associação Sul-Matogrossense dos Produtores de Novilho Precoce – ASPNP na disponibilização das bases de dados. Este trabalho é parcialmente apoiado financeiramente pela FUNDECT, editais Nº 11/2014 – Universal e T.O. 0219/2014.

REFERÊNCIAS

ALLEN, P.; FINNERTY, N. *Mechanical Grading of Beef Carcasses*. Oak Park, Carlow, Ireland, 2001. Disponível em: <http://www.teagasc.ie/research/reports/foodprocessing/4659/eopr-4659.pdf>.

BARBER, T.; STOELLBERGER, P. *Saiku Analytics*. 2008. Disponível em: <http://meteorite.bi/saiku>.

BARROS, G. S. C. et al. *Relatório PIBAgro–Brasil 2016*. 2016. Disponível em: <http://www.cepea.esalq.usp.br/upload/kceditor/files/Relatorio%20PIBAGRO%20Brasil_DEZEMBRO.pdf>.

BRIDI, A. M. *Normas de Avaliação, Classificação e Tipificação de Carnes e Carcaças*. [S.l.], 2002.

CANCIAN, P. H. et al. Correlations of visual scores, carcass traits, feed efficiency and retail product yield in Nellore cattle. *Scientia Agricola*, scielo, v. 71, p. 17 – 22, 02 2014. ISSN 0103-9016. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-90162014000100002&nrm=iso>.

HALL, M. The weka data mining software: An update. *SIGKDD Explorations*, v. 11, 2009.

KETTLE, P. D. I. 2017. http://community.pentaho.com/projects/data-integration. Acessado em 24/05/2017.

KIMBALL, R. *The Data Warehouse Toolkit*. [S.l.]: John Wiley & Sons, 1996.

RODRIGUES, R. *Instrução Normativa Nº – 9, de 04/05/2004*. 2004. http://www.defesaagropecuaria.sp.gov.br/www/legislacoes/popup.php?action=view&idleg=643.

SOUZA, C. M. *Integração de Ferramentas de Código Aberto (Java, Pentaho e Android) e Mapas, Aplicada a Projetos de Inteligência de Negócios*. [S.l.], 2010. Monografia de Especialiazação.

SOUZA, K. M. de. *Rendimento de Carcaça*. [S.l.], 2014.

USDA; FAS. *Livestock and Poultry: World Markets and Trade*. 2017. https://docs.google.com/viewerng/viewer?url=https://apps.fas.usda.gov/PSDOnline/Circulars/2017/05/Livestock_poultry.pdf.

AgrônomoBot: um *ChatBot* para exibição de dados de uma RSSF *EKO* aplicado em agricultura de precisão

Deyvison Bruno Morais de Abreu[1], *Ícaro Ramires Costa de Souza*[1], *Igo Romero Costa de Souza*[1], *Leonardo Barreto Campos*[1,2], *Carlos Eduardo Cugnasca*[2]

[1] Coordenação de Sistemas de Informação – Instituto Federal de Educação, Ciência e Tecnologia da Bahia (IFBA) – Vitória da Conquista, Bahia, Brasil, deyvison.abreu@ifba.edu.br, icaro.souza@ifba.edu.br, igoromero@ifba.edu.br, leonardobcampos@ifba.edu.br

[2]Departamento de Engenharia de Computação e Sistemas Digitais – Escola Politécnica da Universidade de São Paulo (USP) – São Paulo, São Paulo, Brasil, carlos.cugnasca@gmail.com

RESUMO

Com o crescimento do uso de dispositivos móveis, vários aplicativos mensageiros foram criados para tornar a experiência de trocas de mensagens cada vez mais automatizada, é o caso dos robôs capazes de conversar por texto com humanos, conhecidos por *ChatBots*. Porém, sua difusão no domínio do agronegócio ainda é sutil. Dessa forma, unindo os conceitos de aplicativos mensageiros e os dados de uma Rede de Sensores Sem Fio dispostos em um ambiente agrícola, este artigo busca apresentar a criação do *AgrônomoBot*. Esse *ChatBot* é capaz de trazer as informações coletadas pelos sensores em um parreiral a um usuário por meio de interação no aplicativo *Telegram*.

PALAVRAS-CHAVE: ChatBot, RSSF, IoT, JSON, API REST.

ABSTRACT

With the growing use of mobile devices, several messaging applications were created to make the messaging experience more and more automated, such as the human-readable robots known as *ChatBots*. However, its diffusion in the field of agribusiness is still subtle. Thus, combining the concepts of messenger applications and the data of a Network of Wireless Sensors arranged in an agricultural environment, this article seeks to present the creation of

AgrônomoBot. This ChatBot is able to bring the information collected by the sensors in a vineyard to a user through interaction in the Telegram application.

KEYWORDS: ChatBot, RSSF, IoT, JSON, API REST.

INTRODUÇÃO

A popularização da Internet e o crescimento na utilização dos dispositivos móveis transformaram significativamente a forma de comunicação entre os usuários. A New Media Consortium revela que a venda de smartphones é a que mais cresce no mundo entre os eletrônicos da área de Tecnologia da Informação e destaca que o Brasil é um dos maiores mercados do mundo nesse segmento, com 130 dispositivos para cada 100 habitantes (NMC, 2012).

De acordo com Alves et al. (2016), em um período de grande crescimento econômico e aumento da renda disponível, o celular passou a ser um dos itens de necessidade básica do brasileiro o que reflete na utilização de diversos aplicativos para atender as mais variadas necessidades: troca de mensagens instantâneas, transferência de fotos e arquivos, uso de redes sociais, finanças, localização geográfica, solicitação de transporte, entrega de alimentos, música, viagem, entretenimento, entre outros.

Segundo um estudo do Business Insider *apud* Rohan et al. (2016) as plataformas de mensagens instantâneas como o *Facebook Messenger*, o *Slack*, o *WhatsApp* e o *Telegram*, tem mais usuários ativos do que qualquer outro aplicativo da Internet, incluindo redes sociais, aplicativos de correspondência, etc. O mesmo relatório mostra que as dez plataformas de mensagens mais usadas correspondem a quase 4 bilhões de usuários. Com a comodidade e praticidade proporcionadas pelos aplicativos mensageiros, aliado a grande quantidade de usuários, fizeram com que essas ferramentas ganhassem mais significado transformando-as num grande canal de comunicação entre usuários, empresas e serviços.

Diante deste cenário, novas tecnologias vão surgindo como meio de aprimoramento de troca de mensagens tornando-as automatizadas e inteligentes, elevando a qualidade da experiência de comunicação dos usuários, como as interações por *Bots* e *ChatBots*. *Bot*, diminutivo de *robot*, também conhecido como *Internet Bot* ou *Web Robot*, é uma aplicação de software concebido para simular ações humanas repetidas vezes de maneira padrão, da mesma forma como faria um robô.

Os *ChatBots* é a implementação de um *Bot* capaz de executar tarefas pré-definidas gerenciando cada inserção de dado, fornecendo respostas automáticas em conversas por meio de perguntas e respostas. Ribas (2016) aponta uma evolução dessas tecnologias mostrando

que os *Bots* tendem a substituir boa parte dos aplicativos, pois simplifica a interação entre usuários e organizações. Em resumo, os *ChatBots* caminham para evoluir do simples reconhecimento de mensagens para a análise de diferentes aspectos cognitivos, como detecção de linguagem e tom de voz, intenções e até sentimentos.

Diante do exposto, utilizando o potencial da API *Bot* do aplicativo *Telegram*, foi desenvolvido o agente *AgrônomoBot*. Sua finalidade é promover a comunicação entre pessoas e das pessoas com dados coletados pelos sensores de uma Rede de Sensores sem Fio (RSSF) eKo, permitindo automatizar a coleta de informações ambientais, como temperatura e umidade do ar, que são parâmetros fundamentais para o monitoramento numa fazenda em estudo.

MATERIAL E MÉTODOS

Para realização do projeto, foram utilizados os dados coletados pela RSSF eKo instalada em uma vitivinícola localizada na região do Vale do São Francisco. Os dados monitorados pela RSSF são armazenados em uma base de dados e disponibilizados por meio de uma API REST. A Figura 1 representa a arquitetura da API eKo iniciando pela Camada Física que é formada pelos Nós e sensores da RSSF eKo. Esta camada é responsável pela coleta e envio dos dados dos equipamentos para *Gateway* presente na Camada de Interconexão utilizando a rede *ZigBee*.

Figura 1 – Arquitetura da RSSF.



Fonte: Autor

Já a Camada de Interconexão visa proporcionar comunicação entre Nós da RSSF, *Gateway* e o Serviço. Nesta camada são implementados os mecanismos para o sincronismo dos dados monitorados entre a camada Física e Camada de Dados e Serviços. Assim que os registros coletados pelos sensores chegam ao *Gateway*, os mesmos são organizados em uma estrutura JSON e enviados para a API através do método *POST*.

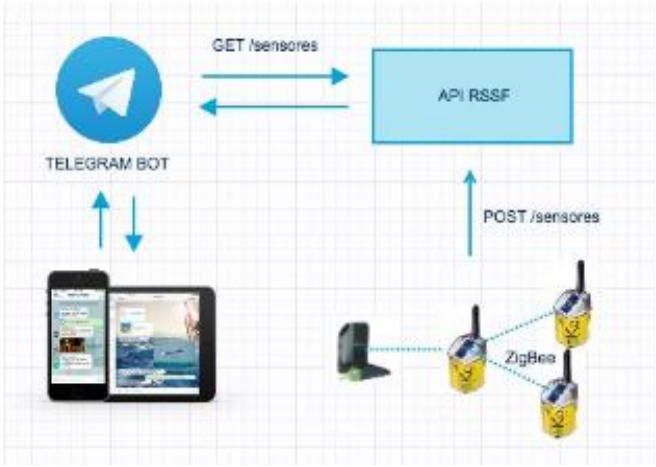
Por fim a Camada de Dados e Serviços é responsável por armazenar e disponibilizar os dados para outras aplicações através da *Web* por meio de uma API (*Application Programming Interface*). A API foi implementada seguindo o modelo *Arquitetural* REST. Neste modelo todos os recursos e informações são disponibilizados por meio de *URLs*. Toda a troca de informações entre as aplicações com esta camada é realizada por meio requisições *HTTP* implementadas pelos métodos (*GET*, *POST*, *PUT*, *DELETE*).

A API *Bot Telegram* permite a criação de *Bots* capazes de conectar ao sistema do Telegram. *Telegram Bots* são contas especiais que não exigem um número de telefone adicional para configurar. Essas contas servem como uma interface para o código em execução em algum lugar em qualquer servidor criado para este fim. Não é necessário ter conhecimento sobre o protocolo de criptografia que o *Telegram* utiliza, o *MTProto*, pois ele funciona no servidor intermediário irá lidar com toda a criptografia e comunicação com a API *Telegram*. Então, é feita uma comunicação com este servidor através de uma interface *HTTPS* simples que oferece uma versão simplificada da API do *Telegram*.

O agente *AgrônomoBot* opera em um *Servidor Web* e utiliza um *token* fornecido pela própria API do *Telegram* atuando com um identificador específico do *Bot*, separando-o dos demais além de ser a autenticação dentro da própria API. É através dele que a API sabe reconhecer o *Bot* e processar as instruções executadas. Por ser Orientado a Regra, o *AgrônomoBot* funciona através de comandos específicos ou palavras chaves.

Para promover a interação do agente com a API REST foram desenvolvidos 4 comandos pré-definidos: */start*, */help*, */no get {id}*, */sensor get {modelo}*. Ao enviar uma mensagem contendo qualquer uma destas instruções, uma requisição é enviada para API do *Telegram* "invocando" o agente, que posteriormente requisita os dados da API RSSF. A Figura 2 apresenta a Arquitetura da RSSF atrelada ao serviço do *ChatBot* acessando a API e fornecendo as informações ao usuário.

Figura 2 – Fluxo de comunicação entre o *ChatBot* e a API RSSF.



Fonte: Autor

A API do *Telegram* sempre retorna uma estrutura JSON após a execução de qualquer um dos comandos pré-definidos vistos anteriormente. Na Figura 3 é demonstrada a estrutura JSON a partir do comando */no get 3*. Ao passar o valor 3 no comando ele tem acesso ao nó específico que está na RSSF e os dados apresentados são: quem enviou, em qual janela de chat foi enviada, o conteúdo e o tipo da mensagem, data de envio dentre outros parâmetros. Essa estrutura então é processada pelo o *AgrônomoBot* que, a partir desses dados, realiza uma requisição à API da RSSF coletando as informações e retornando o resultado a quem solicitou.

Figura 3 – Estrutura JSON retornada pela API do Telegram.

```
{ message_id: 1092,
  from:
   { id: 24993309,
     first_name: 'Icaro',
     last_name: 'Ramires',
     username: 'icaroramiires' },
  chat:
   { id: 24993309,
     first_name: 'Icaro',
     last_name: 'Ramires',
     username: 'icaroramiires',
     type: 'private' },
  date: 1498904998,
  text: '/no get 3',
  entities: [ { type: 'bot_command', offset: 0, length: 3 } ] }
[ '/no get 3',
  'no',
  'get 3',
  undefined,
  index: 0,
  input: '/no get 3' ]
```

Fonte: Autor

RESULTADOS E DISCUSSÃO

Por meio de uma biblioteca de código aberto para a criação de *ChatBots* do aplicativo mensageiro *Telegram* disponibilizada no *GitHub*, foi possível implementar o *AgrônomoBot* e através do JSON acessar a API REST de uma RSSF, captar as informações oriundas dos sensores e proporcionar ao usuário informações pertinentes promovendo uma abstração de toda a estrutura interna e lógica da rede.

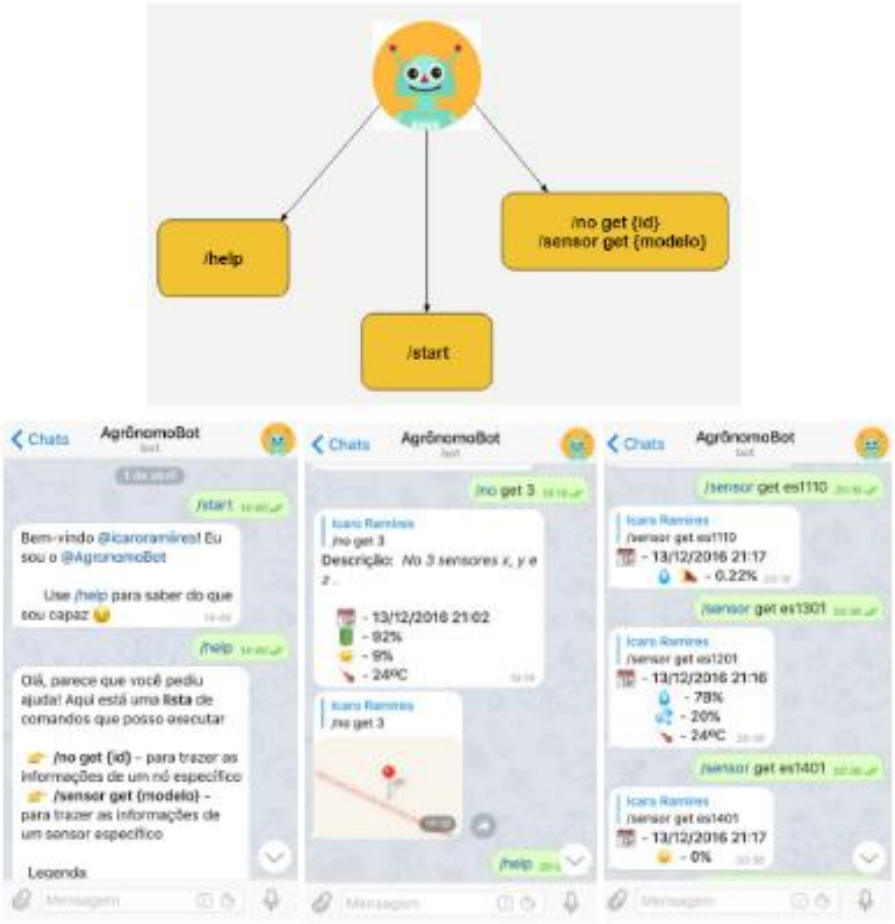
Qualquer usuário do aplicativo mensageiro *Telegram* pode iniciar uma conversa com o *AgrônomoBot* e por meio de regras definidas (apresentadas na seção anterior), interagir com o *ChatBot* e obter informações ambientais, como temperatura e umidade do ar da RSSF.

As regras enviadas ao *AgrônomoBot* via conversa no *Telegram* retornam ao usuário as informações de acordo ao que foi especificado na API *Bot*:

- O comando */start* é aplicado assim que inicia o chat e na tela principal é apresentado ao usuário quem é o *AgrônomoBot*.
- O comando */help* mostra ao usuário um conjunto de regras e o que elas retornam.
- O comando */no get {id}* é passado o número do nó (sensor) e o usuário tem as informações: data e hora, nível da bateria, carga solar, temperatura e localização.
- O comando */sensor get {modelo}* é passado a referência do sensor presente no nó para capturar a informação e retornar ao usuário. Essa referência está presente na documentação do nó.

A Figura 4 apresenta os comandos e as informações adquiridas por meio das regras pré-definidas.

Figura 4 – Comandos e informações recebidas no aplicativo *Telegram.*



Fonte: Autor

CONCLUSÕES

O *AgrônomoBot* apresenta ser uma boa ferramenta para a visualização das informações disposta numa RSSF visto que é de grande interesse o monitoramento de uma região onde se cultiva agricultura de precisão para garantir a qualidade da produção.

O JSON se encaixou bem na arquitetura da aplicação proposta, demonstrando ser rápido, simples e flexível tornando viável o encapsulamento e transporte dos dados através de todos os módulos da aplicação. Grande parte das linguagens de programação oferecem suporte ao formato JSON permitindo realizar serialização de objetos para o formato, como também manipulação de estruturas neste formato.

Como os sensores produzem uma grande quantidade de dados, pretende-se utilizar estes dados para realizar um cruzamento de informações e a partir de uma análise de dados identificar combinações de fatores climáticos que possam favorecer o desenvolvimento de possíveis doenças/fungos ou agentes causadores.

Outra funcionalidade a ser desenvolvida é capacitar o agente *ChatBot* a emitir notificações através de mensagens, possíveis falhas que possam acontecer no funcionamento da RSSF e seus sensores.

REFERÊNCIAS

NMC. HorizonReport: Panorama Tecnológico para Ensino Fundamental e Médio Brasileiro de 2012 a 2017. Texas: Sistema Firjam, 2012. Disponível em: <http://zerohora.com.br/pdf/14441735.pdf>. Acesso em: 24 mar. 2017.

ALVES, Alessandro Martins, et al. Não vivo sem ele - Quando o celular deixou de ser supérfluo. Revista PMKT, 2016. Disponível em: <http://www.revistapmkt.com.br/Portals/9/Revistas/v9n3/2%20-%20N_o%20vivo%20sem%20ele%20-%20Quando%20o%20celular%20deixou%20de%20ser%20sup%C3%A9rfluo.pdf>. Acesso em: 24 mar. 2017.

KAR, Rohan; HALDAR, Rishin. Applying Chatbots to the Internet of Things: Opportunities and Architectural Elements. 2016. Disponível em: <https://arxiv.org/ftp/arxiv/papers/1611/1611.03799.pdf>. Acesso em: 24 mar. 2017.

RIBAS, Paulo. Cedro Technologies. O que é chatbot? Tendências e aplicações. 2016.Disponível em: <http://blog.cedrotech.com/o-que-e-chatbot-tendencias-aplicacoes/>. Acesso em: 24 mar. 2017.

SHAIKH, Ayesha, et al. A Survey On Chatbot Conversational Systems. International Journal of Engineering Science 3117 (2016). Disponível em: <http://ijesc.org/upload/464758c5f7d1a1cd13085e8a584ec5f3.A%20Survey%20On%20Chatbot%20Conversational%20Systems.pdf>. Acesso em: 24 mar. 2017.

LIN, Lue; D'HARO, Luis Fernando; BANCHS, Rafael. A Web-based Platform for Collection of Human-Chatbot Interactions. Proceedings of the Fourth International Conference on Human Agent Interaction. ACM, 2016. Disponível em: <https://www.researchgate.net/profile/Luis_DHaro/publication/308830073_A_Web-based_Platform_for_Collection_of_Human-Chatbot_Interactions/links/582522c808aeebc4f89e7674.pdf>. Acesso em: 24 mar. 2017.

FIELDING, Roy Thomas. Estilos arquitetônicos e o design de arquiteturas de software baseadas em rede. Dissertação (Mestre em Ciências da Computação). Universidade da Califórnia, Irvine, 2000. Disponível em: <http://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm>. Acesso em: 24 mar. 2017.

TURING, A. Computing Machinery and Intelligence. Mind, 59, 433-560, 1950. Disponível em:

<http://repositorios.questoesemrede.uff.br/repositorios/bitstream/handle/123456789/338/GT2--238.pdf?sequence=1>. Acesso em: 24 mar. 2017..

W EIZENBAUM, Joseph. ELIZA - A Computer Program for the Study of Natural Language Communication between Man and Machine. Communications of the Association for Computing Machinery 9 (1966): 36-45. Disponível em: <http://s3.amazonaws.com/academia.edu.documents/31085335/ElizaScript.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1490394313&Signature=sJcWUAbyE8MrflbGnG7j6afh0SU%3D&response-content-disposition=inline%3B%20filename%3DELIZA_a_computer_program_for_the_study_o.pdf>. Acesso em: 24 mar. 2017.

LOUREIRO, Antônio AF, et al. Redes de sensores sem fio. Simpósio Brasileiro de Redes de Computadores. Vol. 21. 2003. Disponível em: <http://homepages.dcc.ufmg.br/~loureiro/cm/docs/sbrc03.pdf>. Acesso em: 24 mar. 2017.

TELEGRAM Bot API. Disponível em: <https://core.telegram.org/bots/api>. Acesso em: 15 mar. 2017.

APP Telegram. Disponível em: <https://telegram.org/>. Acesso em: 15 mar. 2017.

WINK, Leonardo. Internet das coisas e agricultura digital. Revista eletrônica Destaque Rural. 2016. Disponível em: <http://www.destaquerural.com.br/internet-das-coisas-criara-novas-oportunidades-de-negocios-na-agricultura-digital/>. Acesso em: 6 abr. 2017.

LEMOS, Mario. A internet das coisas criará novas oportunidades de negócios na agricultura digital. DATAGRO, 2016. Disponível em: < http://www.uagro.com.br/editorias/tecnologia/2016/09/06/internet-das-coisas-criara-novas-oportunidades-de-negocios-na-agricultura-digital.html>. Acesso em: 7 abr. 2017.

Ferramenta móvel para gestão cronológica das atividades de manejo das culturas do guaraná e da banana

Marcos Filipe Alves Salame[1], *Paulo Igor Moraes da Silva*[2] *Allex de Lima Sousa*[2]

[1]Analista de Tecnologia da Informação, Embrapa Amazônia Ocidental
Manaus, Amazonas, Brasil, marcos.salame@embrapa.br
[2]Bolsistas de Iniciação Científica FAPEAM, Embrapa Amazônia Ocidental
Manaus, Amazonas, Brasil, pauloigormoraes@gmail.com, allexlima@unn.edu.br

RESUMO

Os índices de produtividade agrícola do Amazonas são baixos quando comparados a outras regiões do país. Algumas possíveis causas são: falta de manejo adequado da terra, restrições naturais como umidade excessiva, alta temperatura, alta qualidade e baixa oferta de nutrientes no solo. A banana é uma das frutas mais consumidas no Estado e o guaraná é produzido praticamente apenas no Brasil. A produção desses dois tipos de cultivo no Amazonas não supre a demanda do Estado. Com base nesse cenário foi desenvolvido um aplicativo para dispositivos móveis com sistema operacional Android que tem como objetivo auxiliar os agricultores na gestão cronológica das atividades de manejo das culturas de guaraná e da banana, na tentativa de indiretamente, melhorar a produtividade. As ferramentas utilizadas foram Google Android Studio, Java JDK 1.8, SQLite, biblioteca JXL. O aplicativo apresenta compatibilidade com Android desde a API nível 15, versão 4.0.3 até a API nível 24, versão 7.0.

PALAVRAS-CHAVE: Aplicativo, Android, Cultivo, Amazonas

ABSTRACT

Amazon's agricultural production rates are low when compared to other regions of the country. Some possibilities are: lack of proper land management, natural restrictions such as excessive humidity, high temperature, high quality and low supply of nutrients in the soil. The banana is one of the most consumed fruits in the State and the guaraná is a product that, commercially, is produced only in Brazil. The production of these two types of cultivation in Amazonas does not supply the demand of the State. Based on this scenario an application was developed for mobile devices with Android operating system that aims to assist farmers in chronological management of guaraná and banana crop management activities, in an attempt to indirectly improve productivity. As tools used were Google Android Studio, Java JDK 1.8, SQLite, JXL library. The app Android featured compatibility from a level 15 API, version 4.0.3 to a level 24

API, version 7.0.

KEYWORDS: Software, Android, Cultivation, Amazon.

INTRODUÇÃO

A produção agropecuária na região Norte gera uma grande preocupação ambiental, pelo simples fato de abrigar a maior floresta equatorial do mundo. É importante se atentar também que o solo amazônico é, em geral, pobre em nutrientes. Desse modo, se retirada a cobertura vegetal, a área pode ficar prejudicada para plantios.

De acordo com (LUIZÃO et al., 2009), no início da década de 1960, o governo brasileiro tentou utilizar os abundantes recursos naturais da Amazônia para abastecer o crescimento econômico regional e nacional. No entanto, as tentativas iniciais de desenvolver a região através do estabelecimento de assentamentos agrícolas enfrentaram sérias dificuldades. Após o estabelecimento do crédito federal subsidiado no final dos anos 1960, centenas de projetos de agricultura e industriais foram aprovados e implementados na Amazônia, mas a maioria dos projetos falhou e foram, consequentemente, abandonados.

É provável que a principal razão para o fracasso se deu pelo processo de migração assistida e colonização ter sido rápida e intensa e milhões de hectares de terras foram entregues a recém-chegados com poucos conhecimentos para apoiar a agricultura. Pouca consideração foi dada às condições do solo, água ou bacias hidrográficas quando os locais foram escolhidos (LUIZÃO et al., 2009).

De acordo com (IBGE, 2006), a maior parte dos estabelecimentos rurais no Amazonas atendem ao critério legal de agricultura familiar. Ou seja, o cultivo de produtos alimentícios ocorre, especialmente, em propriedades de pequeno porte, com aplicação de técnicas rudimentares, pouca mecanização e tecnologias, resultando em índices baixos de produtividade.

O sistema de cultivo refere-se às práticas comuns de manejo associadas a uma determinada espécie vegetal, visando sua produção a partir da combinação lógica e ordenada de um conjunto de atividades e operações. A falta de manejo adequado da terra, juntamente com restrições naturais como umidade excessiva e alta temperatura, alta acidez e baixa oferta de nutrientes no solo na maior parte da região, são algumas causas da produtividade baixa de alimentos e fibras na Amazônia Brasileira (LUIZÃO et al., 2009).

A banana possui relevante importância econômica e social para o Amazonas, sendo uma das frutas mais consumidas. Situa-se em segundo lugar como produto agrícola, logo após a mandioca. A demanda pela fruta é expressiva, principalmente na capital, onde requer a importação de aproximadamente 50% da banana consumida, uma vez que a produção estadual não é suficiente (PEREIRA et al., 2002).

Um dos motivos que levam a uma baixa produtividade dos bananais decorre da falta de planejamento e do manejo inadequado dos plantios. Os danos durante a colheita e os cuidados pós-colheita não são considerados pela maioria dos bananicultores amazonenses, levando

a perdas de produção significativas. Isso mostra a necessidade de mais informações ao produtor, permitindo a ele aumentar sua produtividade, e assim ampliar a oferta de produtos com qualidade, tornando-os mais competitivos no mercado regional (PEREIRA et al., 2002).

Para que a crescente demanda mundial de alimentos seja atendida é necessário aumentar a produtividade agrícola. Com os avanços tecnológicos que possibilitam ganho de produtividade, as áreas de cultivo podem ser aproveitadas de modo a extrair o máximo possível de seu potencial, principalmente pelo fato de os recursos naturais serem limitados, devendo ser utilizados de forma sustentável (MANTAU et al., 2013).

Em termos comerciais, o Brasil é o único produtor de guaraná do mundo. No Estado do Amazonas, o guaranazeiro é uma cultura plantada tanto por grandes como por pequenos produtores. De acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE, 2006), o estado produziu 779 toneladas de sementes secas de guaraná em 5.178 hectares, com produtividade média de 150 kg de sementes secas por hectare, menor que a média do Brasil, que foi de 298 kg/ha (TAVARES et al., 2005).

Atualmente quase toda a produção brasileira de guaraná é consumida no mercado interno, sendo pequena a quantidade exportada para outros países. Estima-se que, da demanda nacional de sementes de guaraná, pelo menos 70% seja absorvida pelos fabricantes de refrigerantes, enquanto o restante é comercializado na forma de xarope, bastão, pó, extrato e outras formas (TAVARES et al., 2005).

Esse cenário motivou o desenvolvimento um aplicativo para dispositivos móveis com sistema operacional *Android* que tem como objetivo auxiliar os agricultores na gestão cronológica das atividades de manejo das culturas do guaraná e da banana. O aplicativo apresenta alarmes, calendário, utiliza banco de dados *SQLite* para a persistência das informações, além de dicas pertinentes e a possibilidade de recalcular as datas, caso ocorra adiantamento ou atraso em alguma atividade.

Há um aplicativo para *smartphones* com sistema operacional *Android OS* ou *iOS* desenvolvido pela Embrapa Milho e Sorgo, conhecido como Doutor Milho, este aplicativo visa auxiliar agricultores, técnicos e estudantes, em atividades cotidianas nas lavouras de milho, oferecendo informações de cunho técnico e recomendações de manejo (EMBRAPA, 2017).

Outro software que pode ser citado é o *Bonanza Gold*, sistema desenvolvido pela empresa TD *Software* em parceria com o Departamento de Economia Rural da Universidade Federal de Viçosa, (TDSOFTWARE, 2017). Este destina-se ao controle completo de todas as atividades agrícolas de uma área de produção rural, além também de gerenciar e controlar a parte financeira, administrativa e de manejo.

Esses sistemas não trabalham com as especificidades das culturas do guaraná e da banana. No primeiro, seu foco é reservado apenas para a cultura do milho. O segundo é extremamente robusto, mas é destinado apenas para computadores com sistema operacional *Microsoft Windows*, o que dificulta o acesso dos produtores rurais do interior do Amazonas a este tipo de *software*. Portanto, o aplicativo desenvolvido pode ser visto como uma ferramenta estratégica.

MATERIAL E MÉTODOS

O desenvolvimento teve início com a identificação, análise e refinamento das informações necessárias a partir de entrevistas com especialistas da área de guaraná e da banana para estabelecer as regras de negócio para o desenvolvimento do aplicativo. No geral, foram realizados estudos técnicos nas áreas de agricultura e computação.

Durante o desenvolvimento da solução, foi realizado o processo de engenharia de requisitos, onde foram coletadas informações específicas necessárias para compreender o processo de atividades cronológicas de manejo das culturas do guaraná e da banana. Vale destacar que um requisito não funcional elencado como muito importante, foi o funcionamento do aplicativo *offline*, sem necessidade de conexão com a *Internet*, devido a ausência de conectividade em áreas de produção agrícola na maior parte do Estado do Amazonas.

Ao final do processo de engenharia de requisitos, foram obtidos os diagramas de casos de uso e de classes da UML *(Unified Modeling Language)* e um documento com os requisitos funcionais e não funcionais, além de outras informações pertinentes para ajudar na etapa de codificação.

As ferramentas utilizadas na construção do aplicativo concentraram-se em: ambiente de desenvolvimento *Google Android Studio*, linguagem de programação *Java JDK* 1.8, versão mínima para execução da aplicação com *API* nível 15, banco de dados com funcionamento local e *offline* - *SQLite*, e a biblioteca livre JXL para exportação das atividades para o formato *excel* e a ferramenta de modelagem Astah Communinity.

Foram realizados diversos testes de compatibilidade com vários tipos de dispositivos celulares e várias versões do sistema operacional Android para garantir excelente execução e ausência de erros desde a API nível 15, versão 4.0.3 até a API nível 24, versão 7.0.

RESULTADOS E DISCUSSÃO

O aplicativo foi desenvolvido para auxiliar nas atividades de manejo das culturas do guaraná e banana de acordo com a data de iniciação do registro dessas atividades no aplicativo.

A Figura 1 apresenta a tela inicial do aplicativo, contendo o nome do aplicado e o menu principal.

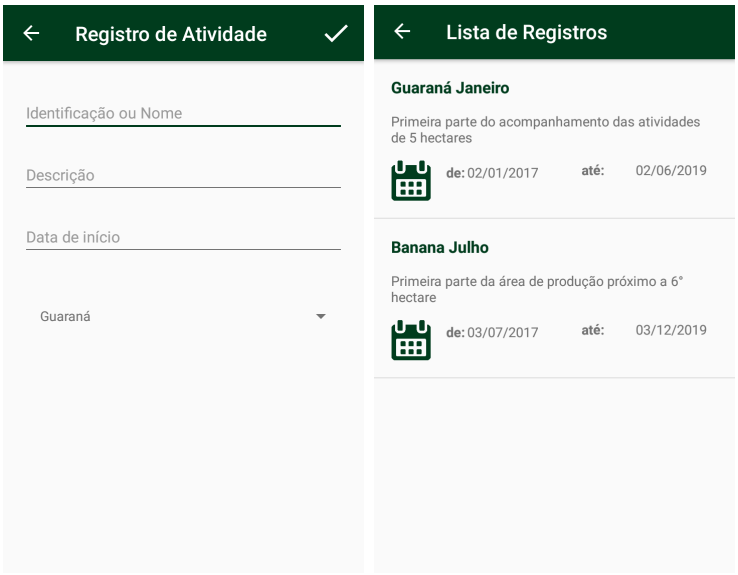
As operacionalizações dessas atividades são mensuradas a partir de uma data inicial, inserida pelo usuário, que o aplicativo usa para produzir cronologicamente um calendário de atividades, com todas as etapas a serem seguidas. A Figura 2a ilustra esta funcionalidade, onde pode-se observar os campos de identificação ou nome, para identificar os registros cronológicos cadastrados; descrição, que é um campo livre usado para descrever quaisquer informações importantes. Por exemplo, pode ser usado para armazenar quantidade de hectares e localização da área de manejo; e data inicial, que é a data específica de início das atividades agrícolas.

A Figura 2b apresenta uma lista de registros de manejos previamente cadastrados. Vale lembrar que as informações utilizadas no exemplo são fictícias.

Figura 1: Telas de menu principal do aplicativo Cronos Plantio.



Figura 2: Telas para cadastro e listagem de registros cronológicos.

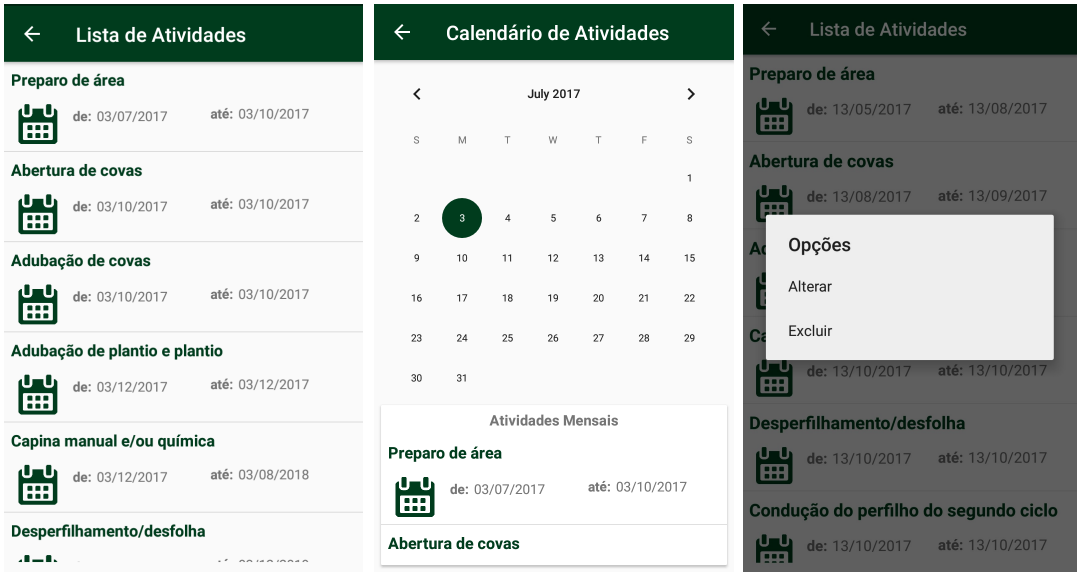


(a) Inserir dados (b) Lista de registros

Para apresentar ao usuário as atividades, referentes a um determinado registro, de forma dinâmica, foram desenvolvidas duas telas diferentes. Uma em forma de lista, conforme ilustrada na Figura 3a e outra em forma de calendário, conforme exibido na Figura 3b. Ambas as telas possuem a mesma função, que é mostrar as atividades de manejo e as datas a serem seguidas para realização dessas atividades, entretanto, cada uma delas possuem experiências de usabilidade e manuseio diferentes. A primeira apresenta visualmente todas as atividades listadas, transferindo ao usuário uma visão mais ampla de todo o processo de manejo; enquanto que na segunda é apresentada de forma mais clara e objetiva, mostrando apenas as atividades de determinado dia e/ou mês.

A funcionalidade de atualização de atividade, consiste em realizar atualização de uma determinada data de uma atividade específica, esta funcionalidade só poderá ser realizada no modo de listagem de atividade, conforme ilustrada na Figura 3c. Cada atividade possui uma data específica para início e término. Porém, caso ocorra algum atraso, adiantamento, tanto da execução de uma atividade quanto do resultado observado, como o caso do florescimento, a data pode ser modificada e os algoritmos criados se encarregarão de reconstruir um novo calendário de atividades a partir da sequência daquela atividade com data alterada.

Figura 3: Formas dinâmicas de apresentação das atividades cronológicas e tela de atualização das datas efetivas.



(a) Listagem (b) Calendário (c) Atualizar data

Outra funcionalidade importante do aplicativo é o alarme. Como as culturas do guaraná e da banana possuem um longo período de manejo e produção, poderia ter o risco do produtor rural esquecer da data de determinada tarefa. De forma a mitigar esse risco, foi introduzido no aplicativo uma forma de lembrar as datas recomendadas dessas atividades. O alarme acontece em dois momentos, em primeira instância o usuário é notificado um (01) dia, ou vinte e quatro (24) horas, antes da realização da atividade, e em segundo momento, no dia da realização da atividade.

CONCLUSÕES

O desenvolvimento de novas tecnologias podem ajudar a melhorar a eficiência nas práticas agrícolas e assim, consequentemente, trazer benefícios para a sociedade.

O aplicativo desenvolvido fornece dicas de boas práticas para um planejamento e execução das atividades de forma mais eficiente. Apresenta a sequência das atividades de manejo das culturas do guaraná e da banana com os períodos recomendados das atividades, apresentando

calendário para melhor visualização e alarme para mitigar o risco de esquecimentos. Possibilita alterar uma data específica, fazendo com que as datas das atividades subsequentes sejam recalculadas automaticamente e permite exportar diretamente para uma planilha.

Como trabalhos futuros, pode ser realizada uma validação com o público alvo para mensurar a utilidade, funcionalidades e usabilidade do aplicativo. Podem ser desenvolvidos módulos com outras culturas para ampliar o alcance dos usuários, além de outras funcionalidades, como por exemplo, um módulo que mapeia a localização geográfica das áreas utilizadas de plantio de cada registro.

AGRADECIMENTOS

Agradecimentos à Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM), pelas bolsas disponibilizadas de iniciação científica aos orientandos que contribuíram com o desenvolvimento deste trabalho e à pesquisadora Mirza Carla Normando Pereira pelo apoio com a disponibilização das informações técnicas referentes à área da agricultura.

REFERÊNCIAS

EMBRAPA. *Site da Embrapa. Doutor Milho - aplicativo para Smartphone*. 2017. Disponível em: <https://www.embrapa.br/busca-de-produtos-processos-e-servicos/-/produto-servico/3991/doutor-milho---aplicativo-para-smartphone>. Acesso em: 28 abr. 2017.

IBGE. *Censo Agropecuário*. 2006. Disponível em: <http://biblioteca.ibge.gov.br/visualizacao/periodicos/51/agro_2006.pdf>. Acesso em: 28 abr. 2017.

LUIZÃO, F. J. et al. The maintenance of soil fertility in amazonian managed systems. *Amazonia and global change*, Wiley Online Library, p. 311–336, 2009.

MANTAU, M. J. et al. Gerenciamento agrícola com suporte à mobilidade: uma ferramenta móvel para automatizar cálculos de plantio. In: *IX Congresso Brasileiro de Agroinformática*. [S.l.: s.n.], 2013.

PEREIRA, M. C. N. et al. Manejo da cultura da bananeira no estado do amazonas. *Embrapa Amazônia Ocidental. Circular Técnica*, Manaus: Embrapa Amazônia Ocidental., 2002.

TAVARES, A. et al. Cultura do guaranazeiro no amazonas. *Manaus: Embrapa Amazônia Ocidental. Sistema de Produção*, v. 4, 2005.

TDSOFTWARE. *Software Bonanza Gold*. 2017. Disponível em: <https://www.agropecuaria.inf.br/produtos/bonanza-gold>. Acesso em: 28 abr. 2017.

AgroAPI: criação de valor para a Agricultura Digital por meio de APIs

Glauber José Vaz[1], *Daniel Rodrigo de Freitas Apolinário*[1], *Jorge Luiz Correa*[1], *Isaque Vacari*[1], *Luis Eduardo Gonzales*[1], *Débora Pignatari Drucker*[1], *Joice Machado Bariani*[1], *Sílvio Roberto Medeiros Evangelista*[1], *Luciana Alvim Santos Romani*[1]

[1]Embrapa Informática Agropecuária, Campinas, São Paulo, Brasil,

{glauber.vaz, daniel.apolinario, jorge.l.correa, isaque.vacari, luis.gonzales, debora.drucker, joice.bariani, silvio.evangelista, luciana.romani}@embrapa.br

RESUMO

APIs são fundamentais para a transformação digital, que tem provocado profundas mudanças em setores inteiros, inclusive na agricultura, em razão das tecnologias digitais e seus impactos na sociedade. Este trabalho apresenta a plataforma da AgroAPI, que é uma iniciativa da Embrapa para promover a criação de valor na Agricultura com a oferta de dados e serviços via APIs. De forma complementar, neste trabalho, é apresentado o uso de APIs devidamente gerenciadas por meio de uma plataforma tecnológica, o que viabiliza novos negócios para a empresa e para o ecossistema em que está inserida, promovendo inovação institucional e para a Agricultura Digital. Estudos de caso para a validação da AgroAPI envolvem publicações técnico-cientificas, dados experimentais e dados agrometeorológicos. A AgroAPI promove uma série de benefícios: facilita a integração de sistemas de informação, com redução de custo e de tempo, melhora a interface com dispositivos móveis, amplia a capacidade de obtenção e disseminação de dados e informações agropecuárias, possibilita economia de recursos computacionais e compartilhamento de dados e serviços, facilita o estabelecimento de acordo com outras organizações e viabiliza maior alcance dos resultados da empresa e de seus parceiros.

PALAVRAS-CHAVE: Application Programming Interface, Transformação digital, Gerenciamento de APIs, Estratégia de API, Integração de Dados e Sistemas.

ABSTRACT

APIs are the building blocks of digital transformation, which has led to profound changes across whole industries, including agriculture, due to digital technologies and their impacts on society. This work presents the AgroAPI platform, which is an initiative of Embrapa to promote value creation in Agriculture with data and services APIs. In a complementary way, this work presents the use of properly managed APIs through a technological platform. It enables new business for the company and its ecosystem, thus promoting innovation for the institution and for Digital Agriculture. Case studies for the validation of AgroAPI involve technical and scientific publications, experimental and agrometeorological data. AgroAPI promotes several benefits: it facilitates the integration of information systems, with cost and time reduction, improves the interface with mobile devices, increases the capacity of obtaining and disseminating agricultural data and information, enables savings of computational resources and sharing of data and services, facilitates the agreements with other organizations and enables greater reach of the results of the company and its partners.

KEYWORDS: Application Programming Interface, Digital transformation, API management, API strategy, Data and systems integration.

INTRODUÇÃO

Vivemos uma era em que uma significativa vantagem competitiva é proporcionada pela transformação digital. Processos de negócios e setores inteiros, inclusive na agricultura, serão transformados pelas tecnologias digitais e seus impactos na sociedade (DEMIRKAN; SPOHRER; WELSER, 2016). Aqueles que não tirarem vantagem desta nova era podem limitar drasticamente oportunidades para sucesso futuro (BERMAN; BELL, 2011).

A transformação digital está ocorrendo rapidamente nas organizações e revitalizando seus modelos de negócios digitais. Ela fornece os meios para melhorar eficiência, efetividade, sustentabilidade e inovação na oferta de produtos e serviços e está mudando a forma de se organizar, criar e entregar valor nas empresas. As APIs são fundamentais nesta transformação digital e um ecossistema de API foi criado a partir da grande quantidade de APIs disponíveis e a integração entre elas (BASOLE, 2016; DEMIRKAN, SPOHRER, WELSER, 2016).

Vukovic et al. (2016), resumidamente, definem APIs como especificações que governam a interoperabilidade entre aplicações e serviços. Os autores destacam a facilidade de reúso das APIs e o seu uso em contextos que não foram originalmente pensados, de maneira que elas se tornam facilitadoras de negócios. Também defendem que APIs em um

ecossistema são mais valiosas do que isoladas e apontam que membros do ecossistema procuram por provedores que ofereçam APIs de qualidade, confiáveis, de longa duração e que tenham suporte.

Para prover uma plataforma tecnológica para negócios digitais e executar programas de API com sucesso, as organizações precisam fazer o gerenciamento do ciclo de vida completo das APIs, que se refere a planejamento, projeto, implementação, publicação, operação, consumo, manutenção e desativação das APIs (MALINVERNO; O'NEILL, 2016).

Uma plataforma de gerenciamento de APIs fornece uma infraestrutura para o desenvolvimento, a execução e o gerenciamento de APIs, uma vez que certos componentes elementares são comuns a todas as APIs (BIEHL, 2015). Também possibilita a segurança e a governança de forma apropriada das APIs de uma organização. Além de disponibilizar APIs para os ecossistemas de negócios, uma plataforma facilita a internalização de APIs de terceiros, fornecendo assim um mecanismo para estimular o desenvolvimento de novos serviços e gerenciar o uso de APIs (LEHONG et al., 2016).

Porém, a transformação digital não trata apenas do reconhecimento de tecnologias inovadoras, mas também do desafio de aplicá-las aos modelos de negócio existentes (EARLEY, 2014). Por isso, uma boa estratégia de API pode começar por negócios já bem estabelecidos na empresa.

Um exemplo de sucesso na área da agricultura é a plataforma de gerenciamento de APIs construída no âmbito do projeto API-AGRO, conduzido por instituições francesas. Um de seus principais objetivos é promover a distribuição de conjuntos de dados e serviços, assim como facilitar seu reúso em aplicações. A plataforma promove a interoperabilidade, o estabelecimento de parcerias, a emergência de novas oportunidades para inovações tecnológicas, o compartilhamento de melhores práticas, o *crowdsourcing*, o foco no negócio e a incorporação de novos atores no ecossistema, como *start-ups* por exemplo. Portanto, além de estratégica do ponto de vista tecnológico, a API-AGRO é uma ferramenta de inovação aberta, promovendo a criação de valor em uma agricultura cada vez mais digital (SINÉ et al.,2015).

A Embrapa, reconhecida internacionalmente por sua atuação na pesquisa agropecuária, tem a oportunidade de ser um provedor de APIs confiável nesta área. Entre os principais ativos de uma instituição de pesquisa, estão o conhecimento gerado e suas publicações. No caso da Embrapa, a literatura produzida pela área de Pesquisa & Desenvolvimento (P&D) é disponibilizada gratuitamente em seu portal (EMBRAPA, 2017) e em portais especializados,

como a Bases de Dados da Pesquisa Agropecuária (BDPA)[1] e os repositórios institucionais: Repositório Alice[2] e Infoteca-e[3]. Além desses sistemas, com interfaces de usuário próprias, também é possível disponibilizar APIs com dados dessas publicações para que novos serviços e aplicações possam ser construídos.

Os dados experimentais da Embrapa constituem uma das principais fontes de todo o conhecimento gerado pela empresa e por isso possuem um alto valor e grande potencial de reúso. Os experimentos que possuem delineamento estatístico têm seus dados planejados, coletados e armazenados no Sistema de Informação de Experimentos da Embrapa (SIExp), que é disponibilizado para um público restrito e interno à empresa. Tanto os dados brutos dos experimentos quanto os que definem os padrões de variáveis podem ser usados por outros sistemas e aplicações existentes na empresa, justificando assim, a abordagem de criação de APIs primeiramente como uma estratégia de integração de dados e sistemas, mas também vislumbrando a possibilidade de disponibilizá-los de alguma forma ao público no contexto de dados abertos.

A agricultura, de forma geral, é uma atividade econômica fortemente dependente do clima. Variáveis meteorológicas, como temperatura, precipitação pluviométrica, radiação solar e vento, influenciam o processo metabólico das plantas, o que impacta diretamente no desenvolvimento e na produtividade das culturas. Desta forma, diversas instituições que atuam na área agrícola dependem de dados agrometeorológicos para agregar valor aos seus produtos e serviços. Neste contexto, há uma forte demanda por dados agrometeorológicos no mercado, o que justifica o desenvolvimento de APIs como uma forma de tornar disponíveis dados e índices de maneira a possibilitar a geração de novos produtos e tecnologias que possam beneficiar todos os atores no agronegócio.

A AgroAPI é uma iniciativa da Embrapa para promover a criação de valor na Agricultura com a oferta de dados e serviços via APIs. Para mostrar o potencial desta iniciativa para a empresa e para a agricultura no contexto de transformação digital, mostramos três estudos de caso de APIs, para publicações, dados experimentais e dados agrometeorológicos da Embrapa e parceiros. Resultados preliminares mostram que o uso de APIs devidamente gerenciadas por meio de uma ferramenta apropriada traz uma série de benefícios, além de viabilizar novos negócios para a empresa e para o ecossistema em que está inserida, promovendo inovação institucional e para a Agricultura Digital.

[1] Disponível em: <http://www.embrapa.br/bdpa>.

[2] Disponível em: <http://www.embrapa.br/alice>.

[3] Disponível em: <http://www.embrapa.br/infoteca>.

MATERIAL E MÉTODOS

Para o gerenciamento das APIs da AgroAPI, tem sido utilizada a ferramenta WSO2 API Manager. Seu modelo de código aberto é um ponto forte da ferramenta segundo o Gartner Inc (MALINVERNO, P.; O'NEILL, 2016) e a Forrester Research, Inc, que ainda destaca o baixo custo da solução e a qualidade de seus recursos de projeto e criação de APIs (HEFFNER, 2014).

Os principais componentes da ferramenta são: (i) publicador de APIs: interface de usuário para que criadores de API possam desenvolver, documentar e versionar APIs; (ii) loja de APIs ou portal do desenvolvedor: interface colaborativa para que os desenvolvedores hospedem e divulguem APIs e para que consumidores usem-nas de maneira segura, protegida e autenticada. O portal é utilizado para os usuários se registrarem, descobrirem e avaliarem as APIs, bem como para se inscreverem a fim de utilizá-las; (iii) API gateway: protege, gerencia e escalona as chamadas às APIs; e (iv) outros componentes para gerenciamento de chaves, gerenciamento de tráfego e análise de dados (WSO2 INC., 2017).

Uma vez que os provedores de APIs precisam reforçar a segurança, a integridade e a qualidade dos recursos expostos, o gerenciamento de APIs é o primeiro passo para ajudá-los a mitigar riscos por meio da limitação e da proteção do acesso a seus recursos (VUKOVIC et al., 2016). A utilização da ferramenta permite que um único ponto de acesso seja utilizado para o total controle e proteção dos dados disponibilizados. Ela permite um completo gerenciamento de usuários e grupos, com um eficaz sistema de autenticação, a utilização de bases de usuários corporativas, a aplicação de políticas de acesso por grupos, a análise de conteúdo e a geração de grande quantidade de indicadores detalhados para futuras tomadas de decisão.

Para que a AgroAPI seja efetiva, é necessário que os serviços disponibilizados funcionem adequadamente. Em relação a suas publicações técnico-científicas, a Embrapa conta com processos consolidados que envolvem dezenas de bibliotecários e um sistema computacional que gerencia essas publicações e seus processos bibliográficos. Também conta com um processo estabelecido para a gestão de experimentos que envolve planejamento, execução, armazenamento de dados e sua disponibilização. O SIExp dá suporte a esse processo. Por fim, os dados agrometeorológicos gerados e organizados pela Embrapa ao

longo de décadas têm sido usados pela própria empresa e também por outras instituições para aprimorar diversos produtos e serviços, como por exemplo o Agritempo, que é um sistema web para monitoramento agrometeorológico com informações para todo o país. Os resultados obtidos a partir desses ativos - publicações técnico-científicas, dados experimentais e dados agrometeorológicos - podem ser potencializados pelo uso de APIs no oferecimento de serviços digitais.

As organizações podem ser provedoras de APIs em diferentes cenários: (i) uso interno, a fim de aumentar agilidade nos negócios e engajamento dos clientes; (ii) negócios com outras organizações, para otimizar os relacionamentos e os processos da cadeia de valor; e (iii) oferecimento de APIs abertas, de maneira que desenvolvedores possam acessá-las via Web (HEFFNER, 2014).

As APIs envolvendo as publicações técnico-científcias da Embrapa podem ser privadas, nos casos em que seu uso é exclusivo da empresa, ou abertas, nas situações em que disponibilizam informações públicas. Aquelas que disponibilizam dados experimentais visam em princípio ao uso interno. Já o fornecimento de dados agrometeorológicos viabiliza novas soluções para a Agricultura em parceria com outras instituições.

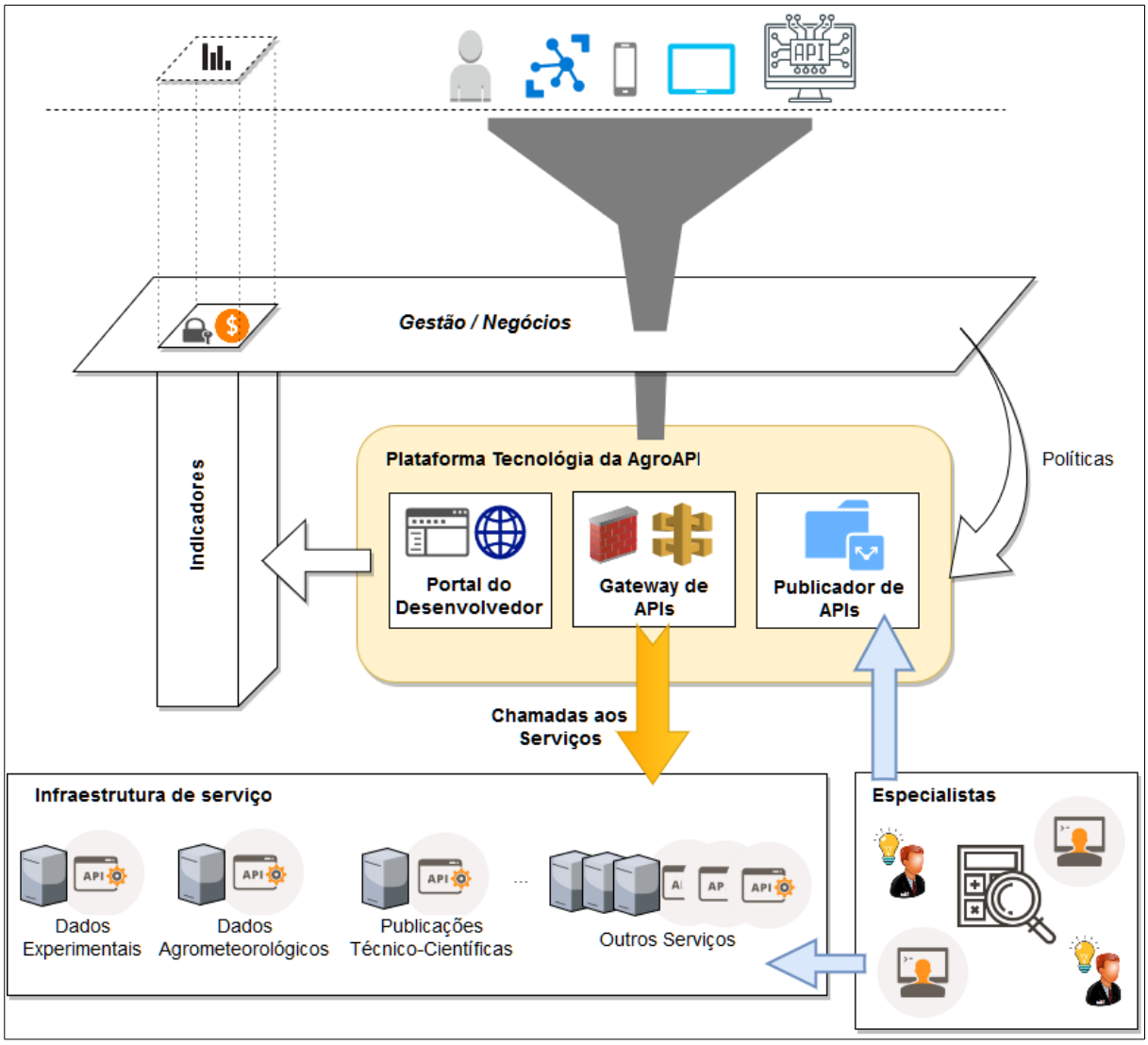
Aliada ao desenvolvimento tecnológico da AgroAPI, modelos de negócios estão sendo propostos para viabilizar o uso desta plataforma no mercado para diferentes segmentos de clientes e usuários de informação. Dada a complexidade e diversidade de produtos pré-tecnológicos que são gerados, bem como diversidade de produtos que podem ser desenvolvidos a partir dessas pré-tecnologias, é necessário oferecer flexibilidade aos licenciados e, simultaneamente, assegurar retorno econômico ou social e segurança jurídica à Embrapa, o que proporciona tanto a manutenção das operações de geração das tecnologias quanto melhorias nas mesmas.

RESULTADOS E DISCUSSÃO

A AgroAPI é representada na Figura 1. Sua plataforma tecnológica apresenta três principais componentes: o portal do desenvolvedor, o gateway e o publicador de APIs. Os serviços que disponibilizam dados experimentais, informações sobre publicações técnico-científicas e dados agrometeorológicos exemplificam o que pode ser oferecido por meio dessa tecnologia. Especialistas da Embrapa e seus parceiros disponibilizam, por meio do publicador, APIs que

fornecem dados e informações em formato digital. A publicação de APIs as torna visíveis no portal do desenvolvedor, usado principalmente por desenvolvedores de software. Também habilita o gateway a receber via Web chamadas às APIs originadas de aplicações diversas: aplicativos de dispositivos móveis, sistemas Web, outras APIs e demais tipos de software, como os embarcados em dispositivos que venham a surgir com o crescimento da Internet das Coisas.

Figura 1- Representação da AgroAPI



A Figura 1 também mostra que as áreas de negócios e de gestão de uma organização provedora de APIs devem estar engajadas na elaboração da estratégia de API. A plataforma tecnológica é utilizada de acordo com as políticas e os modelos de negócio determinados por essas áreas e podem ser ajustados em função dos indicadores coletados por meio da própria

plataforma. Além de servir à gestão da AgroAPI, esses indicadores também podem ser acessados de maneira controlada e segura pelos próprios consumidores de API a fim de tomarem decisões que agreguem valor a seus negócios.

A API que provê acesso aos metadados das publicações da Embrapa possibilita a obtenção de informações como quantidade de registros do acervo, documentos cujos registros sofreram alteração em certo período e documentos cujos identificadores estão dentro de um determinado intervalo de valores.

Esta API é especialmente útil na integração dos sistemas de informação da empresa, pois cria a oportunidade de reduzir custo e tempo de integração devido ao reúso e ao compartilhamento de dados e serviços. Esta API fornece informações, por exemplo, para a construção dos índices que são acessados pela API de busca, tratada a seguir.

A API de busca permite que sejam recuperados os dados de determinada obra da Embrapa por meio de seu identificador ou das publicações que correspondam ao resultado de uma consulta. Para sua validação, construímos uma interface web baseada em componentes visuais (*portlets*) cujo conteúdo é gerado por controladores que fazem chamadas a esta API.

Esta mesma API pode servir a várias outras aplicações, inclusive desenvolvidas por terceiros. Como as publicações são indexadas apenas uma vez, consultas idênticas geram os mesmos resultados, ainda que realizadas por aplicações diferentes. Com isso, também possibilitam economia de recursos, evitando replicações desnecessárias de índices, e um maior controle de acesso, o que possibilita análises mais precisas sobre as consultas realizadas. O uso de APIs para as publicações técnico-científicas da Embrapa, portanto, viabiliza maior alcance de seus resultados de pesquisa e desenvolvimento, economia de recursos e unificação dos métodos de indexação e busca.

O SIExp possui APIs que foram criadas originalmente para viabilizar a integração com uma aplicação para dispositivos móveis cuja principal finalidade é ser uma ferramenta digital para coleta de dados experimentais. Para isso, o SIEXP possui uma API com recursos de consulta e de envio de dados coletados. Dentre seus recursos, destacamos os que fornecem informações de variáveis padronizadas pelos curadores de produtos agropecuários e os que fornecem as informações de experimentos conduzidos no âmbito de projetos de pesquisa, além do que permite o envio de dados de experimento coletados em campo. Como dados experimentais podem ser sensíveis e restritos, a API possui mecanismos para garantir seu acesso exclusivamente por pessoas autorizadas.

Portanto, a API de dados experimentais da Embrapa disponibiliza informações confiáveis e padronizadas de experimentos para serem consumidas por outras aplicações da própria empresa, propiciando maior eficiência para integrar sistemas e dados. O serviço de envio de dados é fundamental para viabilizar o uso de dispositivos móveis para coleta de dados experimentais. Outros dispositivos também podem utilizá-lo, ampliando-se assim o potencial de coleta de dados para o enriquecimento das bases de dados para a pesquisa agropecuária.

Informações meteorológicas e agrometeorológicas, de monitoramento e de previsão, são fundamentais para possibilitar uma avaliação mais completa das condições de tempo permitindo que melhores decisões sejam tomadas, como por exemplo a data certa para o tratamento fitossanitário e o planejamento da colheita, dentre outros. A fim de possibilitar o acesso facilitado a informações agrometeorológicas de todo o país, está sendo desenvolvida uma API para dados agrometeorológicos que busca informações diárias em bases climáticas da empresa permitindo que diferentes aplicações possam ser desenvolvidas para melhorar o processo de tomada de decisão pelo agricultor.

Assim, as APIs de dados agrometeorológicos facilitam a interação com outras organizações no estabelecimento de acordos de cooperação. Como esses dados podem ser úteis em uma grande variedade de aplicações, é possível que as instituições façam uso desses dados e criem interfaces próprias para atender da melhor maneira possível às necessidades dos usuários em nichos específicos de atuação, ampliando-se assim, a capacidade de oferecimento de soluções no ecossistema.

CONCLUSÕES

A AgroAPI é uma iniciativa da Embrapa para promover a criação de valor na Agricultura Digital por meio de APIs. Este trabalho apresenta resultados preliminares que mostram o potencial do uso de APIs devidamente gerenciadas para a inovação na empresa e na Agricultura Digital. Os estudos de caso envolvem APIs abertas, APIs para uso interno e para a realização de negócios com outras organizações. A AgroAPI facilita a integração de sistemas de informação, com redução de tempo e de custo, e a interface com dispositivos móveis e outros equipamentos que possam surgir com o crescimento da Internet das Coisas. Assim, pode-se ampliar a capacidade de obtenção e disseminação de dados e informações agropecuárias. O uso de APIs também possibilita economia de recursos computacionais e

compartilhamento de dados e serviços, além de facilitar o estabelecimento de acordos entre organizações e viabilizar maior alcance dos resultados obtidos pela empresa e seus parceiros.

REFERÊNCIAS

BASOLE, R. C. Accelerating Digital Transformation: Visual Insights from the API Ecosystem. **IT Professional**, v. 18, n. 6, p. 20-25, 2016.

BERMAN, S. J.; BELL, R. Digital transformation: Creating new business models where digital meets physical. **IBM Institute for Business Value**, p. 1-17, 2011.

BIEHL, M. API Architecture: The Big Picture for Building APIs. **API University Series**, v. 2, p. 3, 2015.

DEMIRKAN, H.; SPOHRER, J. C.; WELSER, J. J. Digital Innovation and Strategic Transformation. **IT Professional**, v. 18, n. 6, p. 14-18, 2016.

EARLEY, S. The Digital Transformation: Staying Competitive. **IT Professional**, v. 16, n. 2, p. 58-60, 2014.

EMBRAPA. Portal Embrapa: biblioteca – livros e publicações. Disponível em <https://www.embrapa.br/biblioteca>. Acesso em 2 mai. 2017.

HEFFNER, R. (2014). The Forrester Wave TM: API Management Solutions, Q3 2014. Forrester Research, Inc, 2014.

LEHONG H.; HOWARD C.; GAUGHAN, D.; LOGAN, D. Building a Digital Business Technology Platform. **Gartner Inc**, 2016.

MALINVERNO, P.; O'NEILL, M. Magic Quadrant for Full Life Cycle API Management. **Gartner Inc**, 2016.

SINÉ, M.; HAEZEBROUCKL, T. P.; EMONET, E. API-AGRO: An Open Data and Open API platform to promote interoperability standards for Farm Services and Ag Web Applications. **Agrárinformatika / Journal of Agicultural Informatics**, v. 6, n. 4, p. 56-64, 2015. doi: 10.17700/jai.2015.6.4.209

VUKOVIC, M. et al. Riding and thriving on the api hype cycle. **Communications of the ACM**, v. 59, n. 3, p. 35-37, 2016.

WSO2 INC. WSO2 API Manager documentation: key concepts. Disponível em <https://docs.wso2.com/display/AM210/Key+Concepts>. Acesso em 25 mai. 2017.

Automatic Image-Based Detection and Recognition of Plant Diseases – A Critical View

Jayme Garcia Arnal Barbedo
Embrapa Informática Agropecuária, Campinas, São Paulo, Brasil,
jayme.barbedo@embrapa.br

ABSTRACT

This paper presents a critical analysis of the current state and future perspectives for the use of digital images applied to plant pathology. The differences between the processes of automatic detection and recognition of diseases in plants are presented, with emphasis on the respective current challenges and difficulties. Some of the limitations intrinsic to the use of digital images for detection and recognition of diseases are discussed. Because some of those limitations are mostly inevitable, they may require the use of ancillary data, which may not always be obtained automatically. As a result, depending on the application, the development of completely automatic diagnosis methods may be unfeasible. Thus, the main objective of this paper is to show that one of the main causes for the low relevance attributed to most algorithms proposed so far is the lack of knowledge by the researchers, especially regarding the real difficulties involved in the diagnosis process. The text concludes showing that significant advancements in this area will only be achieved through careful experimental delineation, realistic objectives, and construction of an image database capable of suitably represent all variations expected to occur within the scope of the algorithm to be developed.

KEYWORDS: Disease diagnosis, Image processing, Plant pathology.

RESUMO

Este artigo apresenta uma análise crítica da situação atual e das perspectivas futuras do uso de imagens digitais aplicadas à fitopatologia. São apresentadas as diferenças entre os processos de detecção e reconhecimento automático de doenças em plantas, com destaque para os respectivos desafios e dificuldades enfrentados na atualidade. Em seguida, são discutidas

algumas das limitações intrínsecas ao uso de imagens digitais para detecção e reconhecimento de doenças. Tais limitações, por serem em sua maior parte inevitáveis, podem demandar o uso de dados auxiliares, os quais nem sempre podem ser obtidos automaticamente. Como resultado, dependendo do tipo de aplicação, o desenvolvimento de métodos de diagnóstico totalmente automáticos pode ser inviável. Com isso, chega-se ao principal objetivo deste artigo, que é mostrar que uma das principais causas para a pouca relevância atribuída à maioria dos algoritmos propostos até o presente é a falta de conhecimento por parte dos pesquisadores, especialmente a respeito das reais dificuldades envolvidas no processo de diagnóstico. Conclui-se mostrando que avanços significativos na área só poderão ser alcançados através de cuidadoso delineamento experimental, estabelecimento de objetivos realistas, e construção de bases de imagens capazes de representar adequadamente todas as variações esperadas dentro do escopo do algoritmo a ser desenvolvido.
PALAVRAS-CHAVE: Diagnóstico de doenças, Processamento de imagens, Fitopatologia.

INTRODUCTION

Plant diseases cause significant economic, social and environmental losses. Timely diagnosis is thus very important, so corrective and preventive measures can be quickly implemented. The detection and identification of diseases are mostly performed visually and without a systematic strategy. In this context, the detection and recognition of diseases using digital images can greatly contribute to improve the response to plant pathologies.

At this point, it is important to make a distinction between the detection and recognition (or identification) tasks. The former aims at detecting a specific disease of interest, usually with potential to cause significant damage, while the latter aims at detecting symptoms and determine their origins. The detection task can be seen as a recognition task in which there are only two classes, the disease of interest and everything else (other diseases, healthy tissue, etc.). Thus, although both tasks are challenging, the recognition issue is usually more difficult, especially if the aim is to identify every type of disorder that can affect a given plant.

Among the imaging techniques that can be used for plant disease analysis, the most common are the fluorescence (BAURIEGEL; GIEBEL; HERPPICH, 2010; BELIN et al., 2013), multispectral and hyperspectral (BARBEDO; TIBOLA; FERNANDES, 2015; OBERTI et al., 2014), and conventional photographs (BARBEDO; KOENIGKAN; SANTOS, 2016; POURREZA et al., 2015). The difficulties related to the use of digital images in plant pathology applications were thoroughly explored in Barbedo (2016). The challenges

mentioned in that work will serve as reference for many of the issues discussed throughout this text, especially regarding conventional photographs in the visible spectrum.

The adoption of image-based tools in plant pathology applications has been slow. This is due to the technical limitations that still permeate all of those tools, which, in turn, are due to both technical difficulties and unrealistic or incomplete knowledge about the problem. While Barbedo (2016) focused on the technical aspects, this paper gives more emphasis to the flaws in design, validation and perception that lead to the development of methods of limited usefulness. It also discusses some ways to avoid this kind of problem, offering a perspective on possible solutions to be explored in the future.

CURRENT PROBLEMS, THEIR REASONS AND POSSIBLE SOLUTIONS

Most problems are common to both the detection and recognition issues. However, as these also have some specific issues, this section is divided into three subsections, one general and two dedicated to the specific issues.

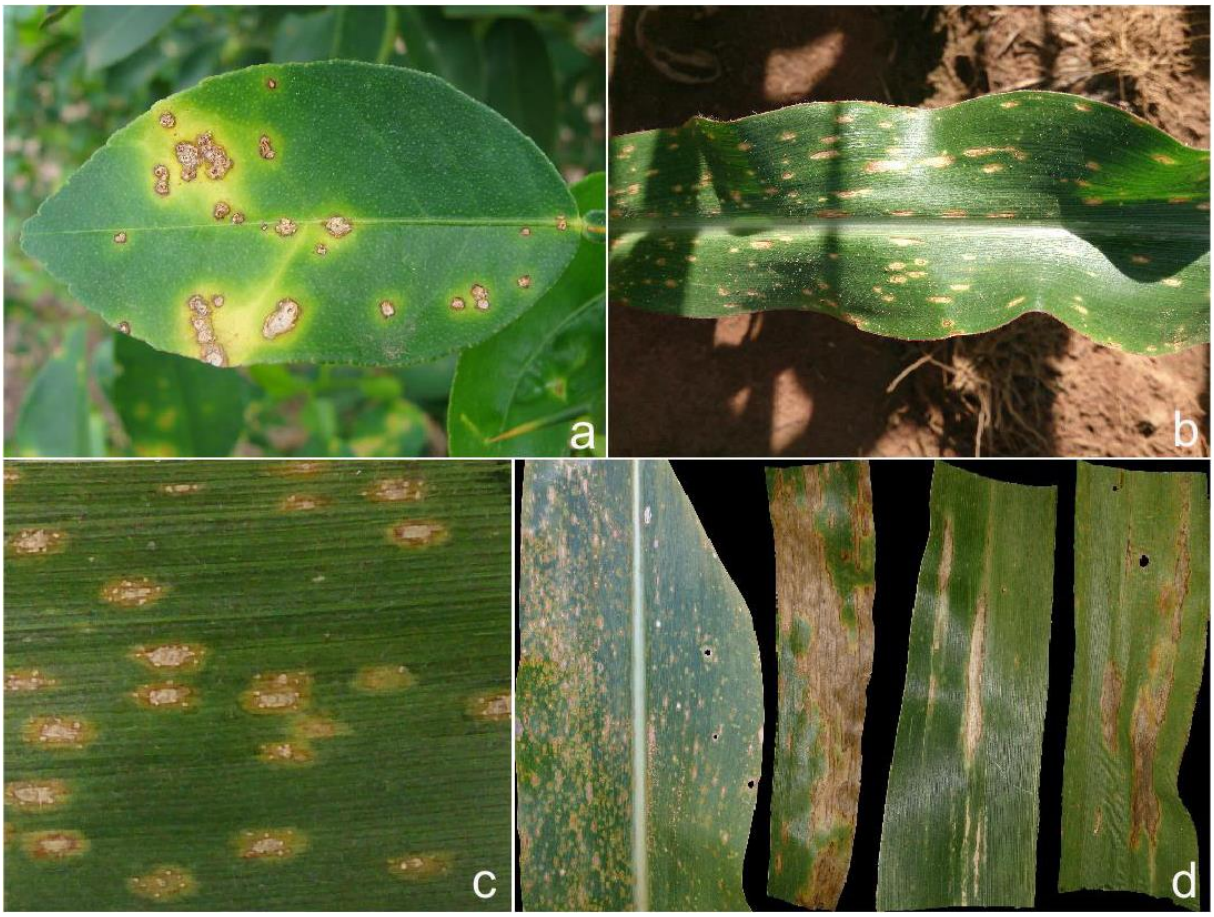
General Issues

Barbedo (2016) identified several challenges that prevent the development of suitable solutions that hold for a wide range of situations (some examples can be found in Figure 1). Those factors should be carefully considered when evaluating any new algorithm, but the majority of studies ignore most or all of them. As a consequence, those algorithms are never validated under realistic conditions. This largely explains why most of those algorithms never find practical applications. A discussion about the consequences of ignoring those factors is presented next.

The background often contains elements that make it very difficult to correctly segment the region of interest where the symptoms are manifesting (BARBEDO, 2016), as seen in Figure 1a. In order to avoid problems associated with this fact, many studies either remove the parts containing the symptoms and image them in a controlled environment with a neutral background (HUANG, 2007), or remove the background manually before applying the algorithm (KRUSE et al., 2014). While the latter prevents the method from being fully automatic, the problem with the former is that it is often not practical to remove parts of the plants, neither to access controlled environments, even if using portable devices (DE CONINCK et al., 2011). A potentially viable solution would be using a panel behind the leaf (MOYA; BARRALESA; APABLAZA, 2005). However, it is not always easy to keep the panel steady and capture the image at the same time. More importantly, the potential client for

the tool will usually detect the symptoms during normal everyday activities, and it would be unlikely that this person would carry the panel at all times. The ideal solution for the problem would be the automatic removal of the background. This has been tried by some authors (ALENYÀ et al., 2013), but this is a very difficult problem by itself that still needs further research.

Figure 1 – a) Example of busy background. b) Example of illumination variation. c) Example of symptoms with unclear edges. d) Variation in symptoms of a corn disease.



Capture conditions in the field can vary greatly, making disease identification very challenging. Again, many authors solve this problem simply by capturing the images under controlled conditions (CLÉMENT et al., 2015). This may be unpractical in real world applications, especially considering that many crop fields are located in remote areas. There are many factor that influence capture conditions: illumination variations (Figure 1b), specular reflections, angle of capture, optical quality of the equipment, image compression, etc. All those factors are relevant in the field, thus if a method fails to take them all into consideration, the probability of failure increases.

Recognition Issues

Most methods involve some kind of segmentation prior to the identification of the disease. Defining the region to be segmented is not a well-defined task, because most symptoms do not have a clear boundary. Instead, they gradually fade into healthy tissue (Figure 1c). As a result, small changes in the parameters of the segmentation procedure (e.g. pixel value threshold) may result in widely diverse segmentations, as more or less of the transition region is considered. In cases like these, there is no ground-truth to be pursued, and the segmentation may be considered correct as long as it includes the whole core of the symptom and does not include healthy tissue. An undesirable consequence of this fact is that the amount of the transition region remaining after the segmentation will vary from image to image, and this may have a great impact on the classification accuracy, especially if it employs features based on color and texture. This is true even when an adaptive approach is used (BARBEDO, 2014). Few methods take this issue into consideration, and even fewer propose a solution for this difficult problem. The most promising approach seems to be avoiding segmentation altogether by formulating the problem directly as an image categorization task, in which context deep convolutional neural networks have played an important role (BARBEDO, 2016; ZHANG et al., 2014).

Intraclass variations are very prominent for most plant diseases (Figure 1d). Those variations may be caused by factors such as plant genotype, variations on healthy tissue color, leaf age, humidity, exposure to sunlight, temperature, wind, etc. (BARBEDO, 2016). However, most techniques proposed in the literature are validated using very small image databases that do not properly cover the variety of symptoms that can be found for each disease. As a result, those techniques tend to fail when used under more realistic conditions. This fact highlights the need for the construction of more comprehensive databases, which is not a trivial task given the difficulties involved (BARBEDO, 2016).

Interclass similarities are also an important factor. Detecting the differences between classes is the primary objective of any disease classification method, so every technique in the literature considers this one way or another. However, since those studies often fail to consider the entire range of symptom variations, two diseases that appear very dissimilar under this limited setup may have relevant similarities when the variety of symptoms is better represented. The consequences and respective solutions for this problem are the same discussed for the intraclass variations.

The vast majority of studies ignore the fact that multiple diseases may manifest simultaneously, with a few exceptions (ZHANG; MENG, 2011). This is a very damaging

limitation, not only because this kind of situation is very common, but also because the interaction between diseases may produce symptoms with diverse characteristics. On the other hand, this is arguably the most difficult limitation to be overcome. As discussed before, it is quite difficult to build image databases encompassing the whole range of symptoms for the diseases of interest, so the task becomes massive if all possible combinations of diseases are to be considered. Thus, a solution for this problem is not expected in the near future (BARBEDO, 2016).

Even if every precaution is taken and state of the art techniques are used, it may not be possible to resolve all ambiguities and reach a reliable diagnosis. Some authors try to minimize this problem by applying fuzzy logic (XU et al., 2011). However, even this kind of approach cannot deal with all uncertainties, which means that some kind of ancillary data, such as geographic distribution of diseases, historical reports, etc., may be necessary. Ultimately, in some cases laboratorial analysis may be the only way to obtain a reliable diagnosis.

Detection Issues

Detection algorithms are usually intended to be used in the field with fixed sensors without supervision, which creates some specific problems that need addressing.

Many diseases, particularly those caused by fungi, have symptoms that manifest on the bottom of the leaves, making them very difficult to be detected by sensors mounted in fixed supports. A possible solution for this issue would be using small robotic rovers capable of going underneath plant canopies. There are some initiatives towards this goal, but much research has yet to be done before it becomes a viable solution.

The lack of sensor mobility may bring other problems. As plants grow, their canopy architecture changes, so the initial position of the sensor may no longer be appropriate. This may be minimized by placing the sensors farther away from the plants, thus covering a larger area. On the other hand, this procedure requires higher resolutions in order to discriminate finer details that may be important for disease detection. Again, mobile sensors seem to be the best solution given their adaptability to different situations.

Monitoring large areas is also challenging. Although there are some areas more likely to be the point of origin for infections, the first symptoms may occur anywhere in the property. In order to assure a timely response, it may be necessary to place a large number of sensors in strategic positions, which may be expensive and, more importantly, may demand a complex communications system for data transmission. Once again, mobile sensors may dramatically

reduce the number of devices needed, as they can theoretically cover large extensions relatively quickly.

Although weather-resistant equipment and devices are common, extreme events may cause considerable damage. Thus, equipment placed in areas prone to such extreme weather events may have to include some extra protection, but even if this is the case, it is important to have replacements ready in case of damage.

DISCUSSION

Most technical issues discussed in the last section depend on the construction of more comprehensive databases to be suitably addressed. Some commercial tools, such as Plantix™ (plantix.net), are starting to break the barrier of database representativeness by developing networks of producers and agronomists to capture and label images of symptoms. However, even these more successful tools have a relatively limited scope regarding plant species and variety of diseases, so a lot more effort has to be spent for their expansion.

As mentioned before, the adoption of image-based tools applied to plant pathology has been slow. The technical difficulties discussed above play an important role, but are not the only reasons for this situation. With the expansion of the smartphone user base, and with most software tools being implemented in the form of applications, it would be reasonable to expect a sharp growth in the use of software tools aimed at agricultural applications. The adoption of this kind of technology by producers is slowed down by two opposite factors. First, many people are naturally skeptical about the ability of technology to replace, with advantages, the processes that have been executed manually for generations. In the opposite side of the spectrum, there are people that have unrealistic high expectations about the performance of those same tools. As a result, they deem the application unsuitable and unreliable, quickly abandoning its use. Thus, it is important to educate potential users about the advantages and limitations of those technologies, making it possible to their potential to be fully realized.

CONCLUSIONS

This paper presented a critical view about some of the main challenges and limitations that still affect the practical adoption of technologies related to the automation of plant pathology activities. It was argued that the main technological barriers can only be overcome, or at least minimized, by the expansion of the existing image databases for plant diseases and by advances in robotics (especially rovers). It was also discussed that unrealistic expectations,

both negative and positive, are a major adverse factor in the adoption of low-cost automation technologies in agriculture. Despite those barriers, image-based technologies applied to agriculture are evolving faster than ever, which means that many of the technical difficulties should soon be overcome.

ACKNOWLEDGEMENTS

The author would like to thank Embrapa (SEG 02.14.01.012.00.00) for funding.

REFERENCES

ALENYÀ, G.; DELLEN, B.; FOIX, S.; TORRAS, C. Robotized plant probing: leaf segmentation utilizing time-of-flight data. IEEE Robotics & Automation Magazine, v. 20, n. 3, 50-59, 2013.

BARBEDO, J. G. A. A review on the main challenges in automatic plant disease identification based on visible range images. Biosystems Engineering, London, v. 144, p. 52-60, Apr. 2016.

BARBEDO, J. G. A.; KOENIGKAN, L. V.; SANTOS, T. T. Identifying multiple plant diseases using digital image processing. Biosystems Engineering, London, v. 147, p. 104-116, July 2016.

BARBEDO, J. G. A.; TIBOLA, C. S.; FERNANDES, J. M. C. Detecting Fusarium head blight in wheat kernels using hyperspectral imaging. Biosystems Engineering, v. 131, p. 65-76, 2015.

BAURIEGEL, E.; GIEBEL, A.; HERPPICH, W. B. Rapid Fusarium head blight detection on winter wheat ears using chlorophyll fluorescence imaging. Journal of Applied Botany and Food Quality, v. 83, n. 2, p. 196-203, 2010.

BELIN, E; ROUSSEAU, D.; BOUREAU, T.; CAFFIER, V. Thermography versus chlorophyll fluorescence imaging for detection and quantification of apple scab. Computers and Electronics in Agriculture, v. 90, p. 159-163, 2013.

CLÉMENT, A.; VERFAILLE, T.; LORMEL, C.; JALOUX, B. A new colour vision system to quantify automatically foliar discolouration caused by insect pests feeding on leaf cells. Biosystems Engineering, v. 133, p. 128-140, 2015.

DE CONINCK, B. M. A.; AMAND, O.; DELAURÉ, S. L.; LUCAS, S.; HIAS, N.; WEYENS, G.; MATHYS, J.; DE BRUYNE, E.; CAMMUE, B. P. A. The use of digital image analysis and real-time PCR fine-tunes bioassays for quantification of Cercospora leaf spot disease in sugar beet breeding. Plant Pathology, v. 61, n. 1, p. 76-84, 2012.

HUANG, K. Y. Application of artificial neural network for detecting Phalaenopsis seedling diseases using color and texture features. Computers and Electronics in Agriculture, v. 57, p. 3-11, 2007.

KRUSE, O. M. O.; PRATS-MONTALBÁN, J. M.; INDAHL, U. G.; KVAAL, K.; FERRER, A.; FUTSAETHER, C. M. Pixel classification methods for identifying and quantifying leaf surface injury from digital images. Computers and Electronics in Agriculture, v. 108, p. 155-165, 2014.

MOYA, E. A.; BARRALESA, L. R.; APABLAZA, G. E. Assessment of the disease severity of squash powdery mildew through visual analysis, digital image analysis and validation of these methodologies. Crop Protection, 24, 785-789, 2005.

OBERTI, R.; MARCHI, M.; TIRELLI, P.; CALCANTE, A.; IRITI, M.; BORGHESE, A. N. Automatic detection of powdery mildew on grapevine leaves by image analysis: optimal view angle range to increase the sensitivity. Computers and Electronics in Agriculture, v. 104, n. 1-8, 2014.

POURREZA, A.; LEE, W. S.; ETXEBERRIA, E.; BANERJEE, A. An evaluation of a vision-based sensor performance in Huanglongbing disease identification. Biosystems Engineering, v. 130, p. 13-22, 2015.

XU, G.; ZHANG, F.; SHAH, S.G.; YE, Y.; MAO, H. Use of leaf color images to identify nitrogen and potassium deficient tomatoes. Pattern Recognition Letters, v. 32, n. 11, p. 1584-1590, 2011.

ZHANG, N.; DONAHUE, J.; GIRSHICK, R.; DARRELL, T. Part based R-CNNs for Fine-Grained category detection. Lecture Notes in Computer Science, n. 8689, p. 834-849, 2014.

Contagem automática de produção frutífera utilizando técnicas de visão computacional e aprendizagem de máquina

Gustavo Ribeiro Teixeira[1], *Victor Mesquita Ferreira*[1], *Juliana Martins de Mesquita Matos*[2], *Alda Mieko Rocha Kimura Vidal*[3], *Flávio de Barros Vidal*[1]

[1]Departamento de Ciência da Computação, Universidade de Brasília - Brasília, DF, Brasil,
[2]CVTAAO - Universidade de Brasília - Brasília, DF, Brasil,
[3]Departamento de Engenharia Florestal, Universidade de Brasília - Brasília, DF, Brasil.

RESUMO

A agricultura de precisão é o ramo da agricultura que utiliza diversos dados para a tomada de decisões em uma colheita. Com a presença das tecnologias para auxílio no campo e a adoção da agricultura de precisão, vários meios de se acompanhar a colheita com um baixo custo vem sendo estudadas e estão se tornando disponíveis para uso na estimativa de produção, em especial a obtenção das sequências de imagens em série de tempo com sistemas de monitoramento. Este trabalho tem por objetivo apresentar um método para realizar a detecção e contagem de frutos automática utilizando imagens capturadas no campo e fazendo uso de técnicas de processamento de imagens e visão computacional. A metodologia proposta segmenta as regiões que contém frutos nas imagens, e analisa a variação dos canais de cor para detectar pontos em que há sobreposição de frutos, separando as regiões segmentadas nestes pontos. Uma contagem do número final de regiões conectadas provê o número estimado de frutos na imagem. Os resultados são avaliados com a métrica *F-measure* a partir das anotações manualmente realizadas das imagens.

PALAVRAS-CHAVE: Contagem de frutos, produção frutífera, visão computacional.

ABSTRACT

Precision Agriculture is a field of agriculture which uses various data to decision–making in a harvest. Using technologies to aid in the field and the adoption of precision agriculture, various ways of overseeing the harvest with low cost have been studied and are being made available to use in production estimative, in particular obtaining time-series image sequences with monitoring systems. This work aims to introduce a method for automated fruit detection and counting using images acquired from the field image processing and computer vision techniques. The proposed method segments the regions that contain fruits in the image, and analyses the variation in the color channels to detect points that are fruits overlaps, then separates them in segmented regions. The estimated number of fruits in the image is obtained by counting the final number of connected regions. The results are evaluated by the F-measure using a created image ground truth labeled dataset.

KEYWORDS: fruit count, fruit harvest, Computer Vision.

Introdução

A agricultura de precisão é o ramo da agricultura que usa dados como clima, qualidade do solo e do ar, maturidade da plantação e até custos de mão-de-obra e de equipamentos, coletados em tempo real para realizar decisões mais inteligentes para a colheita, procurando-se aumentar o retorno econômico e minimizar os efeitos ambientais (INAMASU J.M. NAIME, 2011). O campo de cultivo não é mais visto como uma unidade uniforme como na colheita tradicional e com os dados coletados é possível calcular a quantidade exata de insumos necessários para cada região do campo.

Dentro deste contexto, uma informação importante é a estimativa de produção da colheita. Com essa estimativa, adquire-se informações sobre polinização, estratégias de desbaste, tamanho da produção e planejamento do armazenamento, orçamento financeiro necessário, entre outros. Dentre as diversas formas de se obter/realizar a estimativa de colheita, uma forma simples e eficaz é realizando uma contagem da produção (por exemplo frutos) no campo. No entanto, isso é uma tarefa árdua, normalmente feita visualmente, requerendo que observadores treinados fiquem no campo várias horas por dia, demandando muito tempo, elevada imprecisão (o fator cansaço humano influencia em muito neste processo), além do elevado gasto financeiro para se realizar esta atividade. Com o advento de tecnologias para auxílio na agricultura e a adoção da agricultura de precisão no campo, várias tecnologias estão se tornando disponíveis para auxílio na estimativa de produção. Em particular, a obtenção de sequências de imagens em série de tempo com sistemas de monitoramento de campo de baixo custo como o *Field Server* (HIRAFUJI et al., 2005) facilitam estudos usando técnicas de visão computacional (GUO, 2013). Essas imagens são capturadas automaticamente do campo, de forma não-destrutiva, permitindo assim que sejam feitas análises sobre essas imagens de todo o campo sem interferir na colheita.

Deseja-se a partir neste trabalho apresentar uma metodologia com a utilização de técnicas, a partir de uma única entrada de imagem capturada da produção no campo, auxiliar os agricultores na estimativa de produção de colheita fornecendo uma solução automatizada que realize a contagem de frutos. Essa contagem deve ser realizada de forma não-destrutiva, ou seja, sem interferir na colheita, e portanto deve ser robusta o suficiente para funcionar em condições naturais de iluminação solar e demais condições para o seu uso no campo.

Trabalhos Relacionados

Em Guo (2013) é apresentada uma técnica de segmentação de imagens para a agricultura utilizando um algoritmo de aprendizado de máquina, que permite a classificação voltada para separar solo e folhas a partir de imagens digitais. A robustez da estrutura de árvores de decisão empregada por Guo (2013) garante um grau de invariância à iluminação da fotografia, permitindo que a segmentação seja realizada com fotos adquiridas em condições naturais de iluminação no campo.

Uma outra técnica muito utilizada para segmentação entre solo e planta em imagens de agricultura de precisão é chamada segmentação de *greeness* (WOEBBECKE et al., 1995). No trabalho de Guijarro I. Riomoros (2015) foi proposto um método para melhorar essa técnica utilizando uma transformada *wavelet* discreta. O algoritmo extrai a informação de *greeness* da imagem e aplica a decomposição baseada em *wavelet*, obtendo imagens com coeficientes de aproximação, vertical, ho-

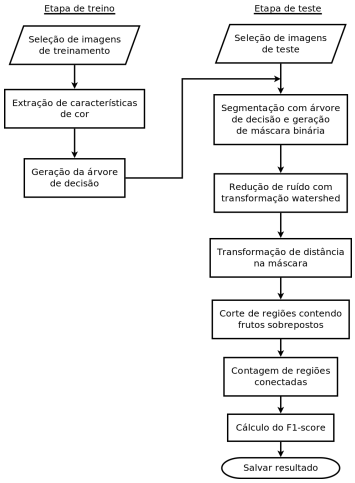
rizontal e diagonal. Descritores de textura são aplicados a essas imagens para obter a variabilidade espacial e então a transformada *wavelet* inversa é realizada, retornando uma imagem com informação de *greeness* melhorada. O método obtém resultados melhores que os métodos de extração de *greeness* convencionais, quantificados em 4,5%. No entanto, sua aplicabilidade a este trabalho é limitada, pois o método destina-se a separar as regiões de planta em geral do solo, enquanto que neste processo de segmentação de frutos estamos interessados em separar os frutos de outros componentes.

No trabalho de Yamamoto W. Guo (2014) é utilizada a técnica de segmentação de imagens de Guo (2013) para segmentar e contar frutos de tomates em imagens, permitindo estimar sua produção. Inicialmente é realizada uma segmentação por cor pixel a pixel seguindo o algoritmo de Guo (2013). Com esta segmentação, são obtidas máscaras indicando regiões que contenham frutos de tomate. No entanto, essas regiões podem conter mais de um fruto sobrepostos, ou em alguns casos, por ruído da segmentação, podem não conter nenhum fruto. Para resolver este problema, é realizada outra classificação com árvores de decisão, desta vez utilizando como entrada características de cor e formato de regiões que podem conter nenhuma fruta, somente uma fruta ou múltiplas frutas. Para isso, identifica-se regiões de superexposição dentro das regiões contendo múltiplas frutas. Essas regiões são causadas pelo *flash* da câmera em cada fruta. Com isso, é possível contar o número de tomates contidos em uma imagem. Mesmo com bons resultados, este método depende de um ambiente controlado de iluminação para a obtenção dos pontos de superexposição e de um método de captura de imagens destrutivo.

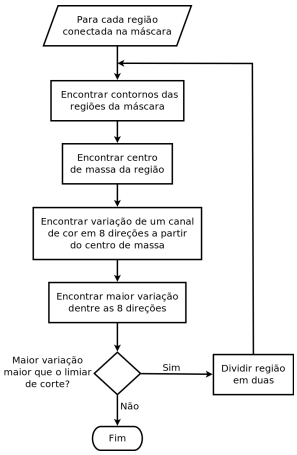
Outro método que faz uso de pontos de superexposição para contagem de frutos é o método proposto por (FONT T. PALLEJÀ; PALACÍN, 2014). Neste trabalho, no entanto, a captura não é feita de forma destrutiva, e sim com o uso de um sistema de captura com *flash* utilizado durante a noite, criando assim os picos de reflexão especular nas uvas que serão contadas. O método proposto realiza uma segmentação das uvas com uma limiarização de *Otsu* sobre o canal *H* do espaço de cor *HSV* e aplicação de filtros morfológicos para redução de ruído. Apesar de a captura de imagens e contagem de frutos ser feita de forma não-destrutiva, as condições de iluminação são altamente controladas ao se capturar imagens somente durante a noite com *flash*. Para Payne K. Walsh (2013) é estimado o número de mangas em um pomar realizando uma segmentação inicial como nos outros métodos, mas optando por uma regressão linear para estimar o número de frutos, compensando por possíveis erros no método. Pixels com predominância de vermelho (canal *R*) sobre o verde (canal *G*) do espaço *RGB*, são selecionados, a imagem é processada com um filtro de variância e limiarizada, removendo pixels de áreas com muitas bordas, como na grama ou folhagem, bem como pixels onde há pouca variação, como no céu.

Metodologia Proposta

O método para a contagem de frutos em imagens proposto neste trabalho utiliza as técnicas de processamento de imagens e visão computacional descritas no fluxograma da Figura 1a descrevendo as etapas do processo, partindo da imagem de entrada até a contagem do fruto, apresentado nas subseções a seguir.



(a) Fluxograma do algoritmo para contagem de frutos em imagens.



(b) Fluxograma do algoritmo de corte de regiões conectadas que contém frutos sobrepostos.

Figura 1: Fluxogramas da Metodologia Proposta.



(a) (b) (c) (d) (e)

Figura 2: Imagens dos estágios de processamento da metodologia proposta.(a) Imagem original. (b) Seleção de regiões de interesse para treinamento do algoritmo. (c) Máscara binária gerada pelo método de segmentação. (d) Máscara binária de segmentação após redução de ruído com *Watershed*. (e) Máscara binária da Figura 2d após realização da transformação de distância.

Segmentação com árvores de decisão

Para que a contagem de frutos possa ser feita na imagem de entrada, inicialmente é realizada uma classificação da imagem, para se realizar o treinamento e separar as regiões que contém frutos das regiões que não contém frutos. Essa classificação é feita pixel a pixel usando características de cor da imagem, resultando em uma classificação binária com uma classe indicando que a região contém frutos e outra indicando que a região não contém frutos. Nesta etapa é empregado uma adaptação da proposta de Guo (2013) utilizando árvores de decisão (BREIMAN J. FRIEDMAN, 1984). As características de cores extraídas da imagem são os valores dos canais de cores de vários espaços de cores, como o *RGB*, *HSV*, *CIE L*a*b*, entre outros. Este algoritmo de aprendizado de máquina é supervisionado e isso significa que deve ser treinado antes de realizar a classificação. O treinamento é feito em um subconjunto de imagens da base de teste usando suas características de cores e ao final é gerada uma árvore binária de classificação. Após a classificação, o método retorna uma máscara binária, representando as duas classes definidas fruta e não fruta[1].

[1]Vale ressaltar que a técnica apresentada neste trabalho pode ser adaptada a qualquer tipo de produção e neste caso optou-se em testar a proposta a para frutos típicos produzidos no Brasil.

Extração de características de cor

As cores de interesse - no caso, as da fruta - serão consideradas como parte de uma classe e cores referentes às folhas, caule, solo, objetos ou demais detalhes em uma foto farão parte de uma segunda classe. O usuário seleciona regiões de pixels nessas imagens de treino, informando a qual das duas classes aquela região de pixels pertence. A Figura 2b mostra uma seleção de regiões realizada em uma imagem da base. Cada pixel selecionado tem seus atributos de cor salvos em uma linha de uma matriz. Esses atributos são os valores de 9 canais de cores - todos os canais dos espaços de cores *RGB*, *CIE L*a*b* e *RGB* após realizada uma equalização de histograma (WOODS, 2010). Uma segunda matriz, contendo o mesmo número de linhas da primeira, guarda o índice da classe a qual pertence o pixel na mesma linha da primeira matriz.

Geração da árvore de decisão

As duas matrizes contendo as informações sobre cada pixel selecionado são usadas para criar a árvore de decisão de acordo com o método do algoritmo *CART* descrito por Breiman J. Friedman (1984). O algoritmo faz uma ordenação dos atributos de cor para descobrir qual deles melhor separa as classes definidas. Esse atributo será usado para a seleção na raiz da árvore. A partir daí, cada novo nó filho faz uma ordenação dos atributos que melhor separam as classes no subconjunto que já foi separado pelo nó pai. Assim, a árvore será construída até que o ganho de informação pela separação não seja significativo. Após isso, é feita uma poda da árvore de decisão para remover nós com seleções que podem confundir o algoritmo de classificação ao levarem em conta ruído nos dados de entrada.

Segmentação e geração de máscara

Com a árvore de decisão criada, é possível segmentar as imagens de teste. Cada pixel da imagem de teste é comparado com os nós da árvore de decisão, a começar pela raiz. Compara-se o valor de uma das características de cor com um subconjunto de valores denotado pelo nó e o algoritmo desce para o nó filho à esquerda ou à direita de acordo com essa comparação. Ao chegar numa folha, o algoritmo obtém a classificação do pixel. Ao classificar cada pixel, uma máscara é preenchida com cor branca em pixels cujo correspondente na imagem original foi classificado como fruto, e com cor preta caso contrário. A Figura 1 mostra um exemplo de uma imagem de acerola sendo segmentada pelo método descrito.

Redução de ruído com Watershed

É possível perceber algum ruído na máscara de segmentação na Figura 2c e para removê-lo faz-se o uso da operação de *Watershed* (WOODS, 2010). Primeiramente, é realizada uma erosão e uma dilatação separadamente na máscara, com cada transformação sendo salva em uma matriz diferente. Na matriz com a dilatação da máscara realiza-se uma limiarização para atribuir a todos os pixels com valor menor que 1 o valor 128, cinza. A transformação *Watershed* é aplicada na soma das duas matrizes. Outra limiarização transforma os pixels na cor cinza em pixels brancos. O resultado é uma

máscara com menos ruído que a original, como demonstra a Figura 2d correspondendo ao resultado para a máscara da Figura 2c.

Transformação de Distância

Com a máscara binária criada, temos informação sobre onde estão os frutos na imagem e é possível contar o número de regiões conectadas para estimar de maneira grosseira o número de frutos. No entanto, essa máscara não fornece informações sobre possíveis sobreposições de frutos na imagem. Em alguns casos, a sobreposição pode ser pequena o suficiente para que uma transformação de distância na máscara possa separar estes frutos. A transformação de distância (TD), reduz o tamanho das regiões brancas na máscara e quando essas regiões possuem um gargalo, que pode indicar uma sobreposição pequena de frutos, esse gargalo fica mais difuso. Com uma limiarização para tornar a máscara binária novamente o gargalo desaparece, deixando duas regiões em seu lugar. A limiarização é feita com um limiar de 0.4 após uma normalização dos valores retornados pela transformação de distância para valores entre 0 e 1. A Figura 2e mostra a mesma máscara da Figura 2d após sofrer a transformação de distância. É possível perceber que uma nova região conectada aparece, correspondendo a um fruto sobreposto a outros na imagem original, a Figura 2a.

Corte de regiões com frutos sobrepostos

Em diversos casos, decorrem problemas em que os frutos não se separam devido à sobreposição no momento da captura. Nem todos eles são separados pela transformação de distância. Para isso, foi criado um método para detectar as regiões conectadas na máscara que contém frutos sobrepostos e separá-las. A Figura 1b mostra o fluxograma do algoritmo para corte de regiões. Para todas as regiões conectadas da máscara, tenta-se encontrar pontos onde possivelmente há uma transição de um fruto para outro. Ao encontrar tal ponto, essa região é separada em duas e o algoritmo tenta encontrar novamente um ponto de corte nas duas regiões novas. Os pontos de transição entre frutos são encontrados a partir de uma análise da variação da luminância ou da cor na imagem. Quando a variação de luminância ou de cor é maior que um limiar estabelecido, aquele ponto é um candidato a um ponto de corte.

Avaliação da contagem de regiões conectadas

Após o método de corte de regiões acima, obtém-se uma máscara com um número de regiões conectadas aproximadamente igual ao número de frutos da imagem. O método de encontrar contornos das regiões pode ser usado para contar o número de regiões encontradas na máscara. Assim, obtém-se o número de frutos reconhecidos em uma imagem. O número de frutos encontrado pode ser comparado com as anotações, o *ground truth* de cada imagem, permitindo assim contar o número de frutos encontrados na imagem (*True Positives*), os frutos que não existem na imagem e foram contados pelo método (*False Positives*) e os frutos presentes na imagem e que não foram contados pelo programa (*False Negatives*). Verdadeiros negativos (*True Negatives*) não serão usados neste problema e os demais valores serão utilizados para o cálculo da *Precision*, do *Recall* e da *F-measure* do algoritmo.

Resultados e Discussões

Base de imagens

Para testar o algoritmo desenvolvido, foram construídas várias bases de imagens, cada uma contendo imagens de um fruto específico. As imagens foram retiradas se sítios na internet, com o único critério de que as fotos fossem capturadas em condições naturais encontradas no campo e que contenham poucos objetos que não sejam as plantas e possivelmente outras características comuns nesta situação (como por exemplo o céu ou solo). Assim, as bases possuem imagens de tamanhos e resoluções variadas, de forma a testar a robustez da solução proposta à esta característica. Esta característica foi definida como um critério que torna a técnica proposta independente do modelo de câmera de captura, incluindo aquelas de mais baixa resolução. Cada base possui entre aproximadamente 50 e 100 imagens, nomeadas com uma ou duas letras referentes ao nome da fruta e um número para a contagem. O número de frutos da imagem foi contado manualmente por um especialista para cada imagem e desenvolvido pelo membros do projeto. Essas anotações são usadas para calcular a métrica de precisão e analisar os resultados. Na Figura 3 são apresentados uma amostra de cada uma das imagens que formam o banco de dados utilizado para os testes da metodologia proposta.



(a) (b) (c) (d) (e) (f) (g) (h)

Figura 3: Exemplos de imagens das bases de imagens de frutas coletadas da internet. (a) Acerola. (b) Ameixa. (c) Caju. (d) Cereja. (e) Laranja. (f) Manga Rosa. (g) Morango. (h) Pêssego.

Realização dos testes

Os testes foram realizados em todas as oito bases de imagens de frutos. Para cada base, foram seguidos dois critérios: um treinamento com 30% das imagens da base e os testes realizados em cima dos outros 70% (caso mais complexo) e também com 70% para treinamento e 30% para testes (caso mais simples). Três parâmetros do algoritmo são alterados para a realização dos testes: a realização ou não da distância de transformação (TD), o canal de análise da variação de cor para corte de região, podendo este ser o canal *L*, *a* ou *b* do espaço de cor *CIE L*a*b*, e o limiar de variação de cor para corte de região, variando entre 10 e 100, com passos de 10 unidades cada. Os resultados de cada imagem são comparados ao *ground truth*, e o número de verdadeiros positivos (TP), falsos positivos (FP) e falsos negativos (FN) é contabilizado para todas as imagens de uma base. Ao final da execução do algoritmo para uma base, esses valores são usados para o cálculo da métrica de precisão *F-measure*(YE et al., 2012). Para cada imagem da base, compara-se o número de frutos obtidos e o valor contido nas anotações. Se esses valores são iguais, adiciona-se esse valor ao número de verdadeiros positivos. Caso o número de frutos obtidos seja maior que o número contido nas anotações, a diferença entre eles é somada ao número de falsos positivos e o *ground truth* é somado aos verdadeiros positivos. Caso o número de frutos obtidos seja menor que as anotações, a diferença entre eles é acrescentada

ao número de falsos negativos e o valor obtido pelo método é somado aos verdadeiros positivos. Como não há imagens sem frutos na base, o número de falsos negativos não é contabilizado.

A Tabela 1 mostra os melhores resultados de *F-measure* para todas as bases de imagens de frutos, utilizando 30% da base para o treinamento e 70% para os testes. A métrica *F-measure* muda de acordo com as variações do limiar de corte entre 10 e 100 e os outros parâmetros, como o canal de corte e a realização ou não da transformação de distância (TD) foram fixados para cada base de acordo com os melhores resultados obtidos e são ilustrados na Tabela 1.

Tabela 1: Principais resultados de *F-measure* para cada base de imagens.

Limiar	*F-measures*							
	Acerola (canal *L*, sem TD)	Laranja (canal *L*, com TD)	Cereja (canal *a*, sem TD)	Caju (canal *a*, com TD)	Ameixa (canal *L*, com TD)	Morango (canal *L*, sem TD)	Pêssego (canal *a*, com TD)	Manga (canal *a*, com TD)
10	0.116741	0.388919	0.231613	0.231969	0.0463065	0.0915758	0.660393	0.241117
20	0.260279	0.613146	0.557225	0.578249	0.136986	0.157907	0.610961	0.361949
30	0.433974	0.646355	0.668596	0.722408	0.369637	0.248445	0.591815	0.482972
40	0.570847	0.658141	0.665639	0.772563	0.582418	0.313268	0.583803	0.557143
50	0.640428	0.670206	0.659341	0.778182	0.71831	0.344688	0.578799	0.524345
60	0.689732	0.677801	0.657233	0.778182	0.746269	0.357548	0.578296	0.524345
70	0.691756	0.686225	0.655118	0.778182	0.763359	0.368397	0.578296	0.524345
80	0.684564	0.674315	0.655118	0.778182	0.752	0.377264	0.578296	0.524345
90	0.679827	0.663385	0.655118	0.778182	0.721311	0.377029	0.578296	0.524345
100	0.669178	0.662064	0.655118	0.778182	0.721311	0.378248	0.578296	0.524345

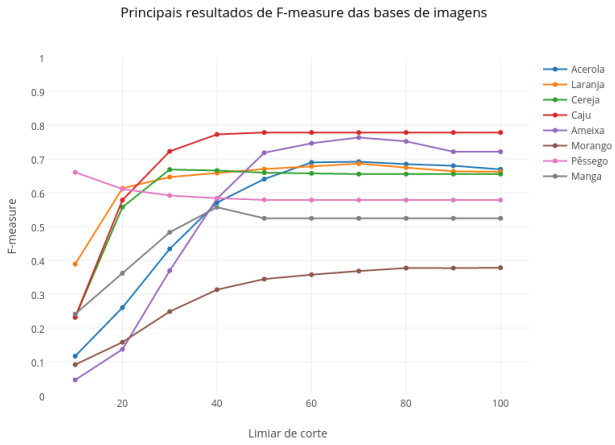
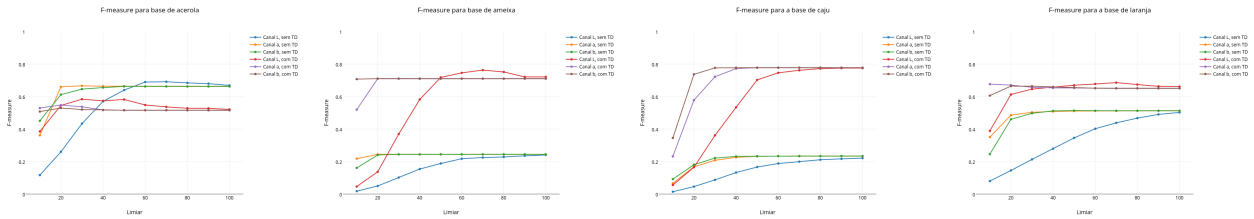


Figura 4: Gráfico dos principais resultados de *F-measure* das bases de imagens.

Os melhores resultados obtidos foram para a base de imagens de caju, com o maior valor de *F-measure* sendo 0.778 para o treinamento em 70% da base. A base com os piores resultados no geral foi a de morango, com o maior valor de *F-measure* chegando somente a 0.38. Somente as bases de caju e ameixa obtiveram resultados acima de 0.7 para o treinamento com 70% da base, embora as bases de acerola e laranja terem chegado bem próximas a esse resultado também. É possível perceber que houve uma melhoria significativa nos resultados quando o treinamento é feito com 70% das bases comparado ao treinamento em 30%. Testes iniciais feitos com um treinamento de poucas

imagens forneceram resultados muito aquém do valor aceitável de 0.7 de *F-measure*, sendo estes não apresentados neste trabalho.



(a) (b) (c) (d)

Figura 5: Gráficos de alguns resultados de *F-measure* ilustrando seu comportamento. Os resultados são das bases de (a) acerola, (b) ameixa, (c) caju e (d) laranja.

Em geral, como é visto nos gráficos das Figuras 5a e 5b, um aumento do limiar leva a um aumento da *F-measure* até um pico claramente definido, a partir do qual os resultados diminuem. Após um certo limiar, os resultados se estabilizam e não se alteram. Isto ocorre pois, a partir deste ponto, a variação do canal de cor selecionado nunca é maior que o limiar, ou seja, a partir desse ponto o algoritmo não realiza mais o corte de regiões. Em alguns casos, o ponto de estabilização pode ser atingido antes de ocorrer um pico claramente definido. Nas Figuras 5c e 5d, é possível ver uma grande disparidade entre os resultados em que foi realizada a transformação de distância e os resultados que não a realizaram. Neste caso específico, bem como em alguns outros, a transformação de distância melhorou bastante os resultados, justificando assim o seu uso, apesar de nem sempre essa melhoria acontecer. Isso pode acontecer pois a transformação de distância em alguns casos remove regiões incorretas na segmentação ou mesmo limita a influência de ruído nos canais de cor no momento do corte de regiões. Em alguns casos, a transformação de distância piora os resultados, como no caso da cereja. Isso pode ocorrer pois o fruto da cereja é relativamente pequeno, e a redução do tamanho das regiões que a transformação de distância faz remove regiões corretas, aumentando o erro da estimativa.

Conclusões e Trabalhos Futuros

Foi desenvolvido um método para detecção e contagem de frutos em imagens voltado para aplicações na área de agricultura de precisão. O método segmenta os frutos da imagem e realiza uma análise de variação de um canal para dividir as regiões que possuem múltiplos frutos. Assim, o método pode contar o número de regiões conectadas na máscara final e obter o número de frutos da imagem. O método obteve resultados satisfatórios, com *F-Measure* acima de 0.7, para as classes de frutos de ameixa e caju. Em outras classes, como acerola, laranja e cereja, muitas vezes os resultados chegaram próximos, passando de 0.6. Isso indica que ainda existe espaço para melhorias do algoritmo, podendo este alcançar resultados mais elevados para a maioria das bases a partir de ajustes finos. O método de segmentação utiliza atributos de cor das imagens para segmentar os frutos e por essa razão pode falhar quando há variação de cor num mesmo fruto, entre os frutos ou quando há cores similares entre classes diferentes. A segmentação é supervisionada, requerendo um treinamento manual realizado antes da segmentação. Os resultados da segmentação dependem altamente desse treinamento.

Uma possível variação futura do método a ser considerado é aplicar uma segmentação com técnicas mais elaboradas, como por exemplo a de *superpixels*. Essa forma de segmentação separaria a imagem em muitas regiões pequenas e o problema se tornaria juntar regiões contíguas ao invés de separar regiões. Assim, esse método poderia substituir tanto a segmentação com árvores de decisão quanto a separação de regiões de frutos sobrepostos.

REFERÊNCIAS

BREIMAN J. FRIEDMAN, C. S. e. R. O. L. *Classification and Regression Trees*. Taylor & Francis, 1984. (The Wadsworth and Brooks-Cole statistics-probability series). ISBN 9780412048418. Disponível em: <https://books.google.com.br/books?id=JwQx-WOmSyQC>.

FONT T. PALLEJÀ, M. T. M. T. D. M. J. M. D.; PALACÍN, J. Counting red grapes in vineyards by detecting specular spherical reflection peaks in rgb images obtained at night with artificial illumination. *Computers and Electronics in Agriculture*, Elsevier, v. 108, p. 105–111, 2014.

GUIJARRO I. RIOMOROS, G. P. e. P. Z. M. Discrete wavelets transform for improving greenness image segmentation in agricultural images. *Computers and Electronics in Agriculture*, Elsevier, v. 118, p. 396–407, 2015.

GUO, U. R. e. S. N. W. Illumination invariant segmentation of vegetation for time series wheat images based on decision tree model. *Computers and Electronics in Agriculture*, v. 96, p. 58–66, August 2013. Disponível em: <http://dx.doi.org/10.1016/j.compag.2013.04.010>.

HIRAFUJI, M. et al. Demo abstract: Field server: Multi-functional wireless sensor network node for earth observation. 2005.

INAMASU J.M. NAIME, A. R. L. B. e. A. B. R. *Agricultura de precisão: um novo olhar*. [S.l.]: São Carlos: Embrapa Instrumentação, 2011.

PAYNE K. WALSH, P. S. e. D. J. A. Estimation of mango crop yield using image analysis–segmentation method. *Computers and electronics in agriculture*, Elsevier, v. 91, p. 57–64, 2013.

WOEBBECKE, D. et al. Shape features for identifying young weeds using image analysis. *Transactions of the ASAE*, American Society of Agricultural Engineers, v. 38, n. 1, p. 271–281, 1995.

WOODS, R. C. G. e R. E. *Digital Image Processing*. 3rd edition. ed. [S.l.]: Tata McGraw Hill Education Private Limited, 2010.

YAMAMOTO W. GUO, Y. Y. e. S. N. K. On plant detection of intact tomato fruits using image analysis and machine learning methods. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 14, n. 7, p. 12191–12206, 2014.

YE, N. et al. Optimizing f-measure: A tale of two approaches. In: *ICML*. icml.cc / Omnipress, 2012. Disponível em: <http://dblp.uni-trier.de/db/conf/icml/icml2012.html>.



Automatic grape bunch detection in vineyards based on affordable 3D phenotyping using a consumer webcam

Thiago Teixeira Santos[1], *Luís Henrique Bassoi*[2], *Henrique Oldoni*[3], *Roberto Luvisutto Martins*[3]

[1]Embrapa Agricultural Informatics
Campinas, São Paulo, Brazil
thiago.santos@embrapa.br

[2]Embrapa Instrumentation
São Carlos, São Paulo, Brazil
luis.bassoi@embrapa.br

[3]Faculty of Agronomic Sciences, Universidade Estadual Paulista – Unesp
Botucatu, São Paulo, Brazil
heriqueoldoni@gmail.com, rlmartins.agro@hotmail.com

RESUMO

O presente trabalho apresenta uma metodologia para a análise fenotípica de vinhas por reconstrução 3-D, a partir de imagens obtidas por uma webcam de alta definição e baixo custo. Um novo software aplicativo integrou componentes de odometria visual e visão estéreo por múltiplas imagens para criar nuvens de pontos tridimensionais densas e precisas para as videiras, em escala milimétrica. Características geométricas e colorimétricas dos pontos foram empregadas em um procedimento de classificação que atingiu 93% de acurácia na detecção de pontos pertencentes às uvas. Os cachos individuais foram automaticamente delimitados e seus volumes estimados. A soma dos volumes estimados por videira apresentou um coeficiente de correlação de $R = 0,99$ ao peso real das uvas observado em cada videira após a colheita.

PALAVRAS-CHAVE: Viticultura, Estimativa de podução, Métodos não-invasivos, Fenotipagem 3-D, Visão estéro múltipla, SLAM.

ABSTRACT

This work presents a methodology for 3-D phenotyping of vineyards based on images captured by a low cost high-definition webcamera. A novel software application integrated visual odometry and multiple-view stereo components to create dense and accurate three-dimensional points clouds for vines, properly transformed to millimeter scale. Geometrical and color features of the points were employed by a classification procedure that reached 93% of accuracy

on detecting points belonging to grapes. Individual bunches were automatically delimited and their volumes estimated. The sum of the estimated volumes per vine presented a coefficient of correlation of $R = 0.99$ to the real grape weight observed in each vine after harvesting.
KEYWORDS: Viticulture, Yield estimation, Non-invasive methods, 3-D phenotyping, Multiple view stereo, SLAM.

INTRODUCTION

Viticulture is an agricultural activity composed by high value perenial crops. Precision viticulture, yield prediction and vine breeding are examples of demanded applications that need some sort of in field phenotyping, but find a bottleneck in laborious measurement by human operators or destructive and sparse sampling. Non-invasive computer vision-based systems have been proposed in research literature, raising as a promising alternative for vineyard phenotyping, and exploiting the ubiquity of digital cameras and their capacity to acquire large amounts of data.

Two different approaches are found in the literature. The 2-D approaches perform grape detection directly on single images, as seen in the works of Nuske et al. (2014). Such methods benefit of their simplified image acquisition step, but occlusions caused by leaves, branches and other bunches are a drawback. An alternative approach are the 3-D based techniques that employ multiple-view stereo vision (HARTLEY; ZISSERMAN, 2004) to create three dimensional models of grapevines in field. Such three-dimensional methods present two clear advantages: they can handle occlusions better than 2-D alternatives, at the cost of a more complex image acquisition step, and they can provide volumetric information, useful on estimating biological attributes as mass, form and fruit quality.

Herrero-Huerta et al. (2015) used a photogrametric software developed by themselves for multiple-view stereo vision, producing sets of 3-D color points, *point clouds*, for vines in field. Convex hulls and solid models were then employed to estimate the bunches volumes and weights. The authors reported a coefficient of determination around 0.77 between their estimated values and the measured biological traits (weight, volume and number of berries). However, their methodology apparently lack an automatic grape detection procedure. Rose et al. (2016) employed a robot carrying a five camera system and a lighting unit to take images of vineyards line at night. The authors used a commercial multiple-view stereo software to produce point clouds. Grape detection was performed using a kernel-based classifier, geometrical features, the *Fast Point Feature Histograms* (FPFH) proposed by Rusu, Blodow e Beetz (2009), and color features in the HSV colorspace. Authors report around 80% of recall and precision for bunches classification in Riesling vineyards cultivated under a vertical trellis system. Another interesting approach is the method proposed by Klodt et al. (2015), an intermediary solution between the 2-D and 3-D approaches: pairs of images are employed to create dense depth maps used in grape classification.

The method proposed in this work is a 3-D based approach. Point clouds were produced

using a novel software, developed by the authors, that combines visual odometry and multiple-view stereo components, resulting in dense, accurate and metric 3-D models for vines in field, from images captured by a simple consumer grade HD webcam. Similar to Rose et al. (2016), FPFH and color features were combined and a support vector classifier was used to detect points belonging to the grapes surfaces. Finally, the volume of grapes was computed for each vine and a regression performed between the estimated volume and the real weight after harvesting.

MATERIALS AND METHODS

Plant material

Five vines of Chardonnay were selected in the vineyards of Guaspari Winery, located at Espírito Santo do Pinhal, São Paulo, Brazil (Lat -22.181018, Lon -46.741618). The winery staff performed dual pruning: one for shaping (after previous year harvest) and one for production, resulting in canopies of lower density. The data gathering was performed few hours before the harvest in May 16, 2017. Images were captured, then bunches were collected and weighted in a scale.

Image data acquisition and 3-D reconstruction

The image data acquisition system was an affordable solution developed by the authors to meet the the idea *affordable and lean phenotyping* for plant science (FIORANI; SCHURR, 2013). The hardware part consisted in a quadri-core computer (an Intel® Core™ i7-based consumer grade notebook) and a HD webcam (a Logitech® c920), able to capture 1920×1080 pixels color frames. The software part consisted in a GNU Linux-based system including:

- a modified version of the ORB-SLAM visual odometry system proposed by Mur-Artal, Montiel e Tardós (2015);
- the patch-based multiple-view stereo system (PMVS) proposed by Furukawa e Ponce (2010), and
- an application implementing the proposed methodology for 3-D reconstruction, *3-Demeter Capture*.

3-Demeter Capture implements an improved version of the methodology previously proposed by Santos e Rodrigues (2016). The camera was set for fixed focus and calibrated using the Zhang (2000) calibration method. This calibration provides the camera *internal parameters* matrix $\mathtt{K}$ (HARTLEY; ZISSERMAN, 2004), immutable along the image acquisition process due to the fixed focus setting. To perform multiple-view stereo, the camera *external parameters*, ie, its rotation and translation parameters for each captured image are also needed. In the robotics and computer vision literature, such estimation of camera external parameters and the needed environment landmarks is called *simultaneous localization and mapping* (SLAM[1])

[1]Actually, the procedure employed in this work is a special case named *monocular visual SLAM* because it employs a single camera and no other sensors.

(SCARAMUZZA; FRAUNDORFER, 2011).

In Santos e Rodrigues (2016), the external parameters were provided by the SVO system proposed by Forster, Pizzoli e Scaramuzza (2014). SVO was able to perform the camera tracking needed for successful 3-D reconstructions, but presented some issues in those experiments: (i) camera tracking showed low robustness to fast camera rotations; (ii) posters containing textures were needed to provide the landmarks employed in camera pose estimation[2], and (iii) SVO implementation did not provide a *re-localization* procedure, able to reset the SLAM system to a sane state after camera tracking is lost, ie, after a pose estimation failure. Instead SVO, 3-Demeter Capture employs a modified version of the ORB-SLAM system (MUR-ARTAL; MONTIEL; TARDóS, 2015). ORB-SLAM has shown superior performance in the authors current phenotyping experiments, being able to find landmarks for SLAM in the surrounding environment, presenting better robustness for rotations and translations and a reliable re-localization module. In the present work, ORB-SLAM code was modified to recover the system *keyframes*. According to Mur-Artal, Montiel e Tardós (2015), ORB-SLAM selects from the video stream a set of frames (keyframes) that lead to robust SLAM, generate a compact and trackable map for the environment and avoid redundancy. This feature is exploited by 3-Demeter Capture for automatic selection of video frames, leading to accurate 3-D reconstructions in the following stereo vision step.

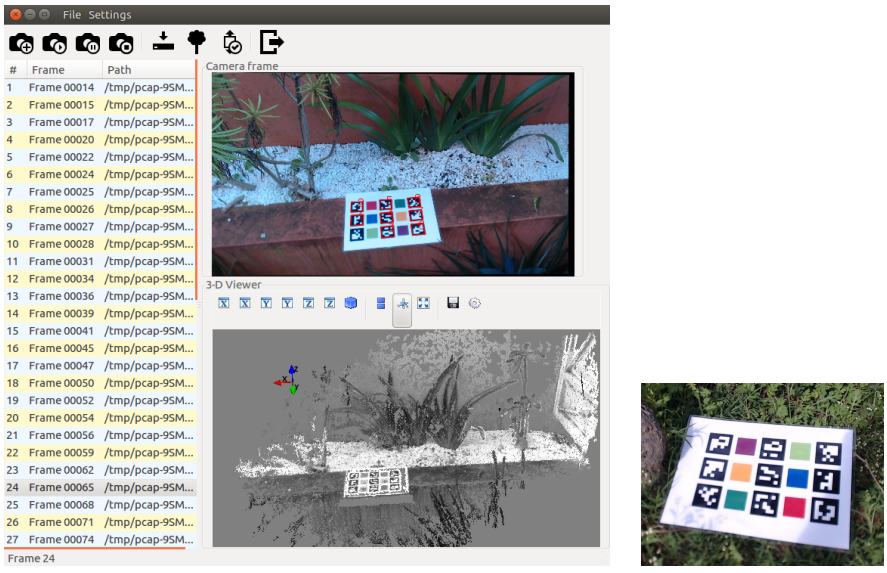
At the end of image acquisition step, 3-Demeter Capture presents a set of data composed by (i) the internal camera parameters matrix $\mathtt{K}$, (ii) a set of N images corresponding to video frames of 1920×1080 pixels and (iii) the camera pose for each image, in the form of rotation matrices $\{\mathtt{R_i}\}_{i=1..N}$ and translation vectors $\{\mathtt{t_i}\}_{i=1..N}$ (HARTLEY; ZISSERMAN, 2004; SCARAMUZZA; FRAUNDORFER, 2011). The application is able to export this data to PMVS and start the multiple-view stereo (FURUKAWA; PONCE, 2010; SANTOS; RODRIGUES, 2016). The point clouds produced by PMVS are actually a set of *surfels*. A surfel is a multidimensional data entry composed by:

- a 3-D point coordinate (x, y, z);
- a vector (n_x, n_y, n_z), corresponding to the normal of the object surface at the point, and
- a color triplet (r, g, b) in the RGB colorspace, corresponding to the observed reflectance of the surface at the point.

However, the *scale* of the produced point cloud is arbitrary. To address this issue, 3-Demeter Capture is able to perform an optional *normalization* step, composed by automatic scaling and rotation to a reference orientation (the Z axis pointing upwards). The application is able to look for a board containing easily recognizable markers presenting known geometry and size (Figure 1). The board geometry is used to transform the point cloud to millimeter scale, a procedure similar to the one employed by Herrero-Huerta et al. (2015). Also, the cloud is transformed to a standard reference frame, where the X and Y axes define the board's plane,

[2]The posters were needed in situations where the environment did not contained very textured objects in the camera field of view.

Figura 1: Screenshot of the 3-Demeter Capture application and the reference board. Application: (a - top) a camera frame where the board was found (the identified markers are highlighted in red); (a - bottom) the normalized 3-D point cloud (note the axis indicator). (b) The reference board with markers.



(a) (b)

and the Z axis is normal to the board. If the board is laid on the ground, the Z axis properly points upwards (see Figure 1).

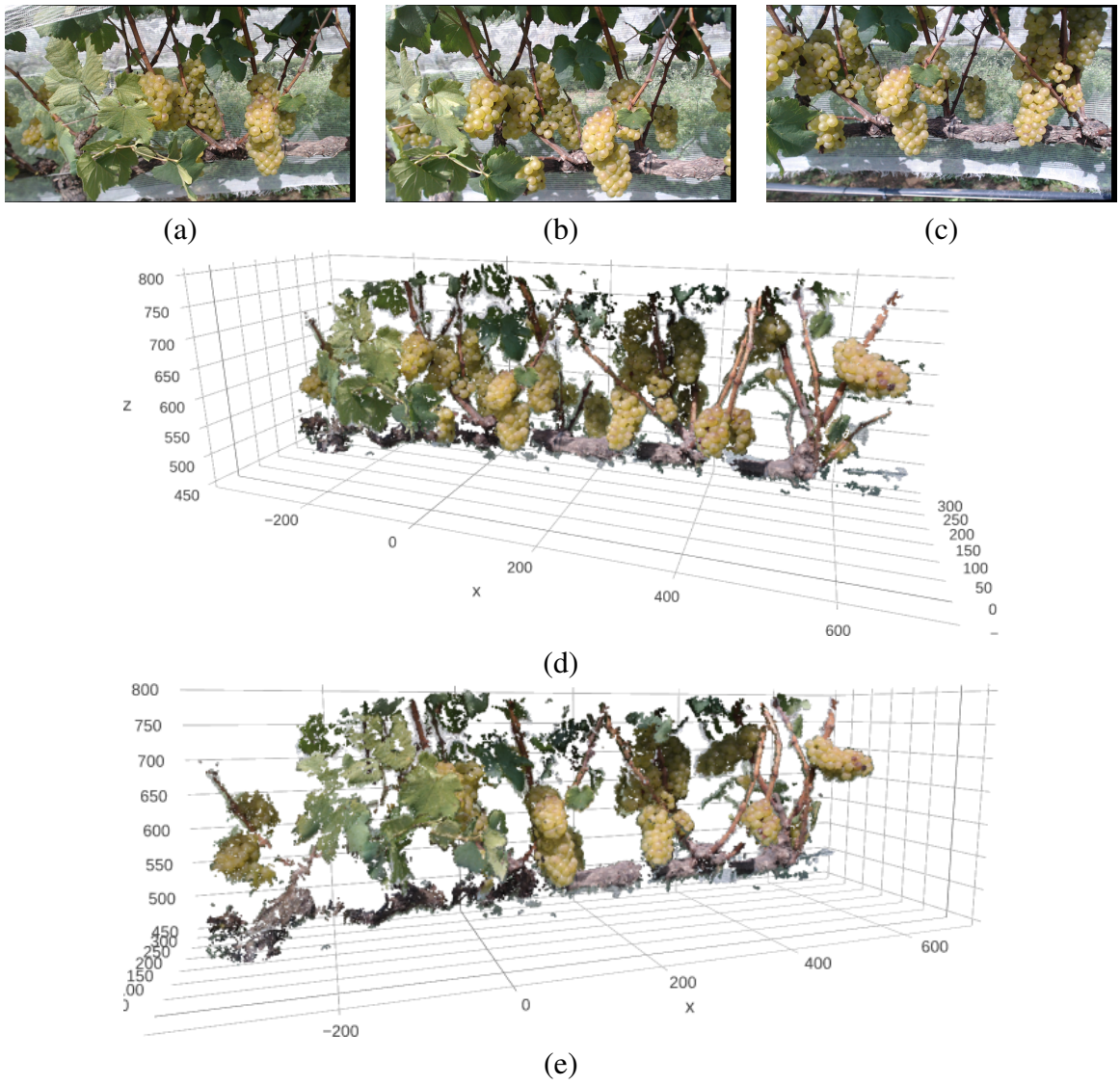
Points classification

The automatic detection of grapes bunches was designed as a supervised classification problem over the set of points in the point cloud. Similar to Rose et al. (2016), we combined geometrical and color information to characterize each point of the cloud as a feature vector. The geometrical features used were the FPFH proposed by Rusu, Blodow e Beetz (2009) and implemented in the Point Cloud Library (PCL) (RUSU; COUSINS, 2011). FPFH represents the geometrical properties of a point's neighborhood by generalizing the mean curvature around the point as a multi-dimensional histogram[3]. This descriptor has been used in several works on 3-D phenotyping for plant parts classification (PAULUS et al., 2013; WAHABZADA et al., 2015; ROSE et al., 2016). A FPFH feature vector is composed by a set of 33 values that are a characterization of the surface around a point: grape bunches, vine leaves, trunks and branches should present different patterns for this descriptor, according to the properties of their surfaces.

Color is also an useful feature for bunches classification. Klodt et al. (2015) used the RGB color space while Rose et al. (2016) employed the three components of the HSV colorspace. In this work, the a^* and b^* components from the CIELAB colorspace were employed. The color information is found in the a^* and b^* components while luminance is expressed on L^*. The L^* channel was discarded in an attempt to make the descriptor more robust to light variations.

[3]See Rusu e Cousins (2011) for a detailed description

Figura 2: Point cloud for a vine. Three of the frames collected by 3-Demeter Capture for this vine, (a), (b) and (c). The point cloud produced by PMVS and normalized by 3-Demeter Capture, seen in two different poses, (d) and (e).



A dataset composed by 121,758 points and their descriptors was employed for classification, under a supervised machine learning framework. The points were manually classified in three classes: (i) grape bunches, (ii) leaves and (iii) trunks and branches, composed by 58,421, 37,258 and 26,079 samples respectively. A support vector machine (SVM) classifier was employed for classification. The classifier penalty parameter C and the used kernel were selected automatically from a set of options by K-Fold cross-validation ($K = 3$) (HASTIE; TIBSHIRANI; FRIEDMAN, 2001). K-Fold cross-validation is employed again to measure the prediction performance of the classifier (using $K = 3$ again), in a *nested cross-validation* approach. The *scikit-learn* (PEDREGOSA et al., 2011) machine learning library was employed for SVM and cross-validation algorithms.

Bunches segmentation, filtering and volume estimation

Points classified as grape surface points were grouped in bunches by finding connected components. Two points were considered connected if the distance between then was up to 3 mm. A connected component is a set of points $\mathcal{B}$ such as for any two points $u_i, u_j \in \mathcal{B}$, there is a path $u_i \rightarrow v_0 \rightarrow v_1 \rightarrow ... \rightarrow v_{P-1} \rightarrow v_P \rightarrow u_j$ linking u_i and u_j, where $v_0, v_1, \ldots, v_P \in \mathcal{B}$ and any two adjacent points v_k, v_{k+1} are connected, ie, are up to 3 mm appart[4]. Each connected components was considered a *bunch*. To filter out noise caused by errors in classification, bunches containing less then 200 points were discarded.

The convex hull for the set of points $u_i \in \mathcal{B}_k$ is computed for each bunch $\mathcal{B}_k$. The convex hull volume is considered here an approximation for the volume of the real grapes bunch in the vine. The convex hull implementation available in SciPy (JONES et al., 2001–) was employed for hull and volume computation. Finally, the individual bunches volumes were summed up to produce a estimation for the vine total volume of grapes.

RESULTS AND DISCUSSION

The data acquisition using 3-Demeter Capture took 1 to 2 min per vine, producing around 90 images per vine (Table 1). PMVS took around 20 minutes per vine to produce a dense point cloud by multiple-view stereo. An example of 3-D reconstruction can be seen in Figure 2. Exploiting the fact the point clouds are properly oriented in a reference frame and in millimeter scale, points belonging to the ground or outside the region of interest were discarded (the grapes are found around 50 cm above the ground).

Tabela 1: Summary of data per vine.

Plant	Num. of images	Num. of 3-D points†	Estimated volume (cm^3)	Weight (g)
Vine 1	89	125,828	1,240.53	1,405.9
Vine 2	96	160,714	566.55	851.4
Vine 3	86	177,218	567.93	750.5
Vine 4	104	172,768	2,311.44	2,420.8
Vine 5	88	121,844	1,585.70	1,771.0

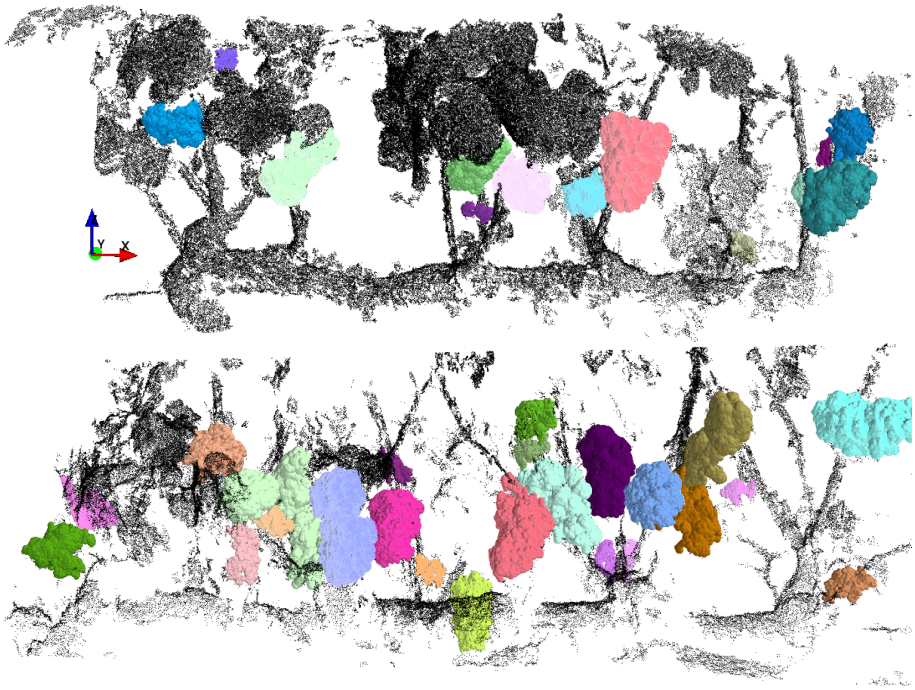
† After the selection of the region of interest.

Classification accuracy was estimated using 3-fold cross validation. The three estimations for accuracy were 93.02%, 92.96% and 93.16% on the annotated dataset, when using the SVM classifier and the 35-features descriptor (33 FPFH features plus the a^* and b^* values). It is interesting to note that using the SVM classifier over just the color descriptors (a^* and b^*) resulted in accuracy around 90% while the 33 FPFH features alone reached less then 70% of accuracy.

[4]This 3 mm threshold produced good results, considering the density of the points clouds.

Figure 3 shows the bunches found for vines 2 and 4. Each bunch is displayed in a different color while points corresponding to leaves and branches are shown in black. Figure 4 shows the linear regression between the estimated total volume of bunches and the true biological yield, the mass of grapes, measured using a digital scale after harvesting (see also Table 1). The correlation coefficient is 0.9984. A larger set of vines should be employed in future experiments to confirm such a good correlation, but this result is not unexpected: a volume estimation made using a 3-D model presenting millimeter precision, combined to a compact grape variety as Chardonnay, should be strongly correlated to weight.

Figura 3: Bunches automatically found in vine 2 (top) and vine 4 (bottom). Each bunch is shown in a different color. Points classified as non belonging to grapes are marked in black.

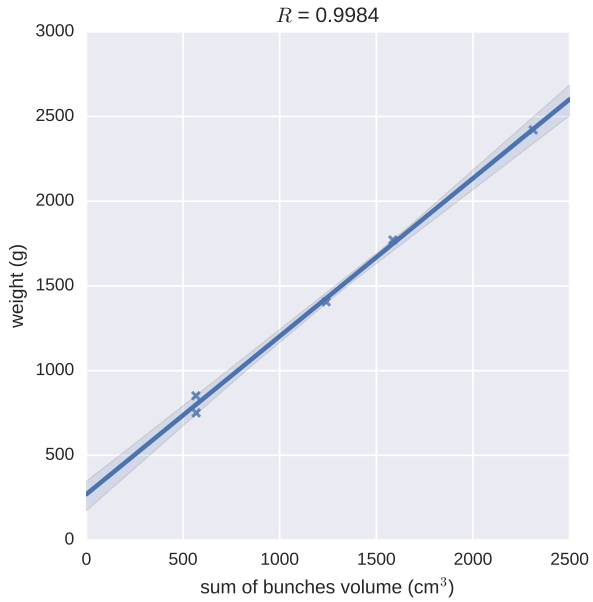


CONCLUSIONS

This work presented an affordable system able to (i) produce three dimensional models of grapevines in field; (ii) recognize bunches, leaves and branches structures in the plants, and (iii) detect bunches of grapes and compute their volume. The computed volumes presented excellent correlation to the vines yields. Applications on plant phenotyping, precision viticulture and agricultural robotics could benefit from these results.

A major advantage of non-invasive methods is the ability to perform non-destructive estimations, enabling the acquisition of developmental data over time. A possible future work would be the monitoring of the entire development cycle until the harvesting.

Figura 4: Regression between the vines estimated volumes and total mass of grapes.



ACKNOWLEDGMENTS

This work was supported by Brazilian Agricultural Research Corporation (Embrapa) under grant 01.14.09.001.05.04. We would like to thank Guaspari Winery for providing access to its wineyards and logistical support.

REFERÊNCIAS

FIORANI, F.; SCHURR, U. Future scenarios for plant phenotyping. *Annual review of plant biology*, v. 64, p. 267–91, jan 2013. ISSN 1545-2123.

FORSTER, C.; PIZZOLI, M.; SCARAMUZZA, D. SVO: Fast Semi-Direct Monocular Visual Odometry. In: *IEEE International Conference on Robotics and Automation (ICRA)*. [S.l.: s.n.], 2014.

FURUKAWA, Y.; PONCE, J. Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 32, n. 8, p. 1362–1376, Aug 2010. ISSN 0162-8828.

HARTLEY, R.; ZISSERMAN, A. Book. *Multiple View Geometry in Computer Vision*. 2. ed. [S.l.]: Cambridge University Press, 2004. ISBN 0521540518.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning*. New York, NY, USA: Springer New York Inc., 2001. (Springer Series in Statistics).

HERRERO-HUERTA, M. et al. Vineyard yield estimation by automatic 3d bunch modelling in field conditions. *Computers and Electronics in Agriculture*, v. 110, p. 17 – 26, 2015. ISSN 0168-1699.

JONES, E. et al. *SciPy: Open source scientific tools for Python*. 2001–. [Online]. Disponível em: <http://www.scipy.org/>.

KLODT, M. et al. Field phenotyping of grapevine growth using dense stereo reconstruction. *BMC Bioinformatics*, v. 16, n. 1, p. 143, 2015. ISSN 1471-2105.

MUR-ARTAL, R.; MONTIEL, J. M. M.; TARDóS, J. D. ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics*, v. 31, n. 5, p. 1147–1163, Oct 2015. ISSN 1552-3098.

NUSKE, S. et al. Automated visual yield estimation in vineyards. *Journal of Field Robotics*, v. 31, n. 5, p. 837–860, sep 2014. ISSN 15564967. Disponível em: <http://doi.wiley.com/10.1002/rob.21541>.

PAULUS, S. et al. Surface feature based classification of plant organs from 3D laserscanned point clouds for plant phenotyping. *BMC Bioinformatics*, v. 14, n. 1, p. 238, 2013. ISSN 1471-2105.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.

ROSE, J. et al. Towards Automated Large-Scale 3D Phenotyping of Vineyards under Field Conditions. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 16, n. 12, p. 2136, dec 2016. ISSN 1424-8220. Disponível em: <http://www.mdpi.com/1424-8220/16/12/2136>.

RUSU, R. B.; BLODOW, N.; BEETZ, M. Fast point feature histograms (FPFH) for 3D registration. In: *2009 IEEE International Conference on Robotics and Automation*. [S.l.: s.n.], 2009. p. 3212–3217. ISSN 1050-4729.

RUSU, R. B.; COUSINS, S. 3D is here: Point Cloud Library (PCL). In: *2011 IEEE International Conference on Robotics and Automation*. [S.l.: s.n.], 2011. p. 1–4. ISSN 1050-4729.

SANTOS, T. T.; RODRIGUES, G. C. Flexible three-dimensional modeling of plants using low- resolution cameras and visual odometry. *Machine Vision and Applications*, v. 27, n. 5, p. 695–707, 2016. ISSN 1432-1769.

SCARAMUZZA, D.; FRAUNDORFER, F. Visual Odometry [Tutorial]. *IEEE Robotics & Automation Magazine*, v. 18, n. 4, p. 80–92, dez. 2011. ISSN 1070-9932.

WAHABZADA, M. et al. Automated interpretation of 3D laserscanned point clouds for plant organ segmentation. *BMC Bioinformatics*, v. 16, n. 1, p. 248, 2015. ISSN 1471-2105.

ZHANG, Z. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 11, p. 1330–1334, Nov 2000. ISSN 0162-8828.



Fusarium Damaged Kernels Detection Using Transfer Learning on Deep Neural Network Architecture

Márcio Nicolau[1,2], *Márcia Barrocas Moreira Pimentel*[1], *Casiane Salete Tibola*[1], *José Mauricio Cunha Fernandes*[1,2], *Willingthon Pavan*[2]

[1]Brazilian Agricultural Research Corporation, National Research Centre for Wheat
Passo Fundo, Rio Grande do Sul State, Brazil
marcio.nicolau@embrapa.br, marcia.pimentel@embrapa.br, casiane.tibola@embrapa.br, mauricio.fernandes@embrapa.br

[2]Graduate Program in Applied Computing (PPGCA), University of Passo Fundo (UPF)
Passo Fundo, Rio Grande do Sul State, Brazil
pavan@upf.br

RESUMO

O presente trabalho mostra a aplicação da transferência de aprendizado em uma rede neural profunda pré-treinada, usando um conjunto pequeno de imagem (≈ 12.000) em uma única estação de trabalho com placa GPU da NVIDIA que levou 1 hora para concluir a tarefa de treinamento e obter uma precisão média global de 94,7%. A rede apresentou classificação incorreta para 20% das imagens de um conjunto de dados externo de teste. A precisão da metodologia proposta é equivalente àquelas que utilizam a metodologia de imagem Hiperespectral (81% - 91%) utilizadas como o mesmo fim, mas com a vantagem de ser mais acessível e de não depender de equipamento especiais para classificar o grão de trigo para sintomas de giberela.

PALAVRAS-CHAVE: Aprendizado Profundo, Grãos com Giberela, ImageNet, TensorFlow, Transferencia de Aprendizado.

ABSTRACT

The present work shows the application of transfer learning for a pre-trained deep neural network, using a small image dataset (≈ 12,000) on a single workstation with enabled NVIDIA GPU card that takes up to 1 hour to complete the training task and archive an overall average accuracy of 94.7%. The DNN presents a 20% score of misclassification for an external test dataset. The accuracy of the proposed methodology is equivalent to ones using HSI methodology (81%-91%) utilized for the same task, but with the affordable advantage of do not depend on special equipment to classify wheat kernel for FHB symptoms.

KEYWORDS: Deep Learning, Fusarium Damaged Kernels, ImageNet, TensorFlow, Transfer Learning.

INTRODUCTION

The wheat is the main source of nutrients to the world population, most of its production is converted into flour for human consumption. In Southern Brazil, where 90% of the domestic wheat is produced, Fusarium head blight (FHB), a fungal disease, is a major concern. Apart from yield loss, the causal agent Fusarium graminearum may cause mycotoxin contamination of wheat products, creating health problems.

Therefore, to avoid potential health risks, Fusarium affected grains must be identified and segregated, before their processing, to avoid its incorporation into food for humans and animal feed.

Usually, the detection of Fusarium head blight (FHB) is carried out manually by human experts using a process that may be both lengthy and tiresome. Moreover, the effectiveness of this kind of detection may drop with factors such as fatigue, external distractions and optical illusions (Barbedo et al., 2015). Thus, improving the detection of Fusarium Head Blight (FHB) in wheat kernels has been a major goal, due to the health risks associated with the ingestion of the mycotoxin, mainly deoxynivalenol (DON).

Methods capable of performing this disease detection in an automatic way are highly demanded in the productive wheat chain, to segregate lots. Most automatic methods proposed to date rely on image processing to perform their tasks (BARBEDO; TIBOLA; FERNANDES, 2015; BARBEDO; TIBOLA; LIMA, 2017; MALONEY et al., 2014).

Barbedo, Tibola and Fernandes (2015) used Hyperspectral imaging (HSI) for detecting Fusarium head blight (FHB) in wheat kernels using an algorithm. The outcome was a Fusarium index (FI), ranging from 0 to 1, that can be interpreted as the likelihood of the kernel to be damaged by FHB. According to the authors, hyperspectral imagery is currently not sensitive enough to estimate DON content directly. However, an indirect estimation from the Fusarium damaged kernels was successfully achieved, with an accuracy of approximately 91% (BARBEDO; TIBOLA; FERNANDES, 2015).

Other study investigated the use of hyperspectral imaging (HSI) for deoxynivalenol (DON) screening in wheat kernels. The developed algorithm achieved accuracies of 72% and 81% for the three- and two-class classification schemes, respectively. The results, although not accurate enough to provide conclusive screening, indicating that the algorithm could be used for initial screening to detect wheat batches that warrant further analysis regarding their DON content (BARBEDO; TIBOLA; LIMA, 2017).

Min and Cho (2015) presented a review about nondestructive detection of fungi and mycotoxins in grains, focusing on spectroscopic techniques and chemometrics. The spectroscopy has advantages over conventional methods including the rapidness and nondestructive nature of

this approach. However, some limitations as expensive and complex setup equipment's and low accuracy due to external interferents exist, which must be overcome before widespread adoption of these techniques.

The application of computer vision on digital images offers a high-throughput and non-invasive alternative to analytical and immunological methods. This paper presents an automated method to detect Fusarium Damaged Kernels, which uses the application of computer vision to digital images.

The main goal of this work is the use of machine learning algorithms and computer vision techniques to detect Fusarium Damaged Kernels in wheat, based on digital images.

MATERIAL AND METHODS

Digital images of Fusarium Head Blight symptomatic and non-symptomatic wheat kernels were available at the National Research Centre for Wheat (Embrapa Wheat), located in Passo Fundo, Rio Grande do Sul State, Brazil.

The images were obtained by recording a video of 06:25 minutes using an Olympus SP-810UZ digital camera with 36x optical zoom, 24mm wide-angle view, and 14-megapixel resolution, 720p HD video and Olympus Lens 36x Wide Optical Zoom ED 4.3-154.8mm 1:2.9-5.

All tasks run on an iMac workstation configured with 32GB of RAM DDR3 1600MHz, a 3.5GHz quad-core Intel Core i7 processor, and a NVIDIA GeForce GTX 780M GPU with 4GB of GDDR5 memory. The TensorFlow 1.0.1 built from source with CUDA Toolkit 8.0 and cuDNN v5.1 to enable GPU support. All scripts were developed using Python 2.7.

Methodology

To classify digital images in predefined classes, we could use one of the several methods developed in recent two decades (SCHMIDHUBER, 2015; GUO et al., 2016). Methodologies to solve this kind of problem was developed both in Artificial Intelligence (AI), a research area in Computer Science, and in Statistics.

The main differences between Statistics and AI approach are the size of the task, for statistics point of view, algorithms and techniques are limited when then input size of image dataset are greater than tens of thousand pictures and for a large number of classification sets.

During the training stage of a system to classify images and objects based only on information content embedded in a single digital file, it is necessary that this system would detect all possible contexts where the object could occur. Small differences in color, luminosity, angle and other could be misinterpreted by the proposed system and results in the wrong or low-level prediction classification.

The key advantage of using AI strategy in the image classification task are related to knowing how to combine layers and manage the relationship between levels of information,

without needing any strong assumption related to the type of dependency or relational structure among input information.

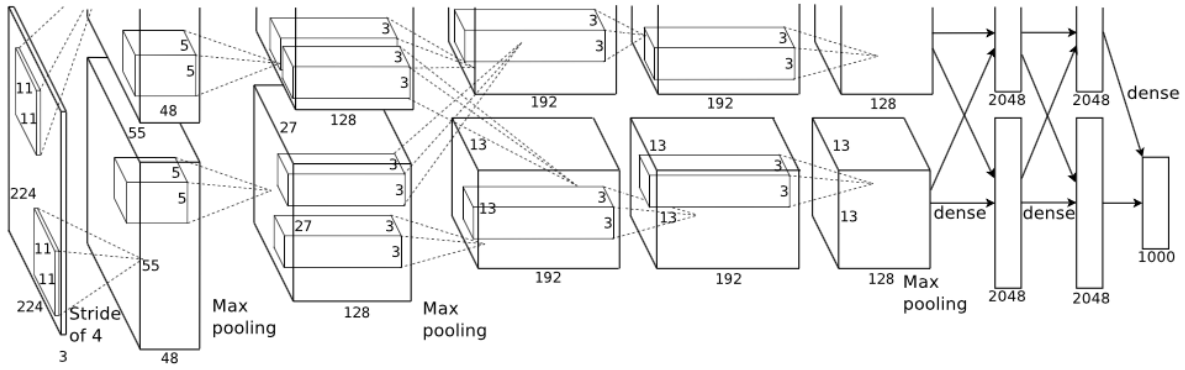
In some cases, the improvements in the accuracy and precision gain are archive using these fine tune setting, but the most of the effort is made only with the input information, in other words, the processes are designed to take the most of the self-learning way.

Deep Convolution Neural Network (DNN)

The most used architecture of the neural network for image classification task is called convolutional neural network (CNN). The convolution operation or, sometimes called convolution layer and is related to the operation to process or respond to "stimuli" in a limited region known as the receptive field.

The receptive field from each neuron contains a partial overlap of information from input layer (raw image) and, in this way, the preprocessing or further operations occur with a minimal amount of effort. Other advantages of this technique are related to the possibility of using distributed algorithms (even GPU versions) to calculate, filters and processing small pieces of information, one each time and aggregated the results when necessary (KRIZHEVSKY; SUTSKEVER; HINTON, 2012).

The deep portion of CNN come from the stacking or combination of several layers where the output of preceding layer is used as an input for the next one. The most common layers employed in DNN are convolution, ReLU (Rectified Linear Units), tanh (Hyperbolic Tangent Function), max pooling, average pooling, fully connected, concat, dropout and softmax. To better understand this relationship see a CNN example in Figure 1.



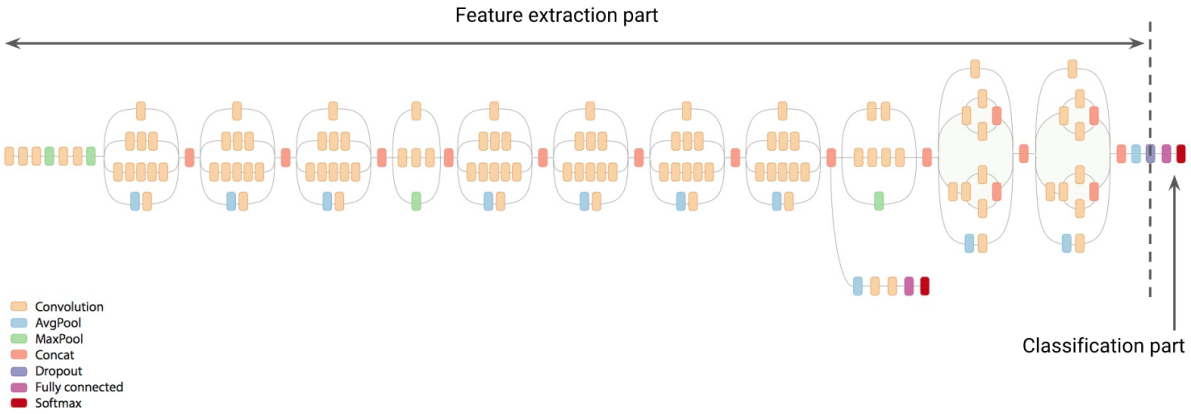
Source: Figure 2 from (KRIZHEVSKY; SUTSKEVER; HINTON, 2012)

Figure 1: An example of relationship between layers in a CNN with convolution, max pooling and dense neuron connections annotated on the illustration.

Transfer Learning

Transfer learning is a machine learning method which utilizes a pre-trained neural network, this technique allows the detachment of the lasts outer layer (classification layer) and uses the

remains structure to retraining and get new weights corresponding the classes of interest – damaged kernel in our case (Figure 2).



Source: Adapted from Figure 5 (DEVIKAR, 2016)

Figure 2: An illustration of a representation of a Deep Convolutional Neural Network, trained on top of ImageNet with detailed information about Feature Extraction part and Classification part. Each small box represents one of the layer types in the Inception Network: Convolution AvgPool, MaxPool, Concat, Dropout, Fully Connected and Softmax.

In this work, we use as a pre-trained neural network the output from (SZEGEDY et al., 2015). Szegedy et al. (2015) developed a 22 layer deep convolution neural network on top of ImageNet for classifying 1000 leaf-node categories, using 1.2 million images for training, 50,000 for validation and 100,000 images for testing.

In brief, transfer learning makes it possible to classify new classes based on a new set of images, reusing the feature extraction part and re-train the classification part of the new picture set. Since feature extraction part of the network it was already trained (which is the most complex part), the new neural network could be trained with less computational resources and time.

Machine Learning Framework

Nowadays we have some options to build and analyze deep neural networks using machine learning algorithms. The final choice would base on the computational infrastructure available to run this task, either the number of classes, the intended purpose and where the final Net will be deployed to handle.

For this work, it is necessary a framework for machine learning that could run on a distributed system with both CPU and GPU, with the possibility to deploy the final network to servers, desktops, mobile applications and embedded systems in an easy way.

Alongside these needs, it is necessary that the chosen framework could easily implement the Transfer Learning techniques, described before. Based on these requirements, the natural choice is TensorFlow (ABADI et al., 2016).

Abadi et al. (2016) presented TensorFlow as an interface for expressing and executing machine learning algorithms that can be performed with little or no change in a broad range of heterogeneous systems, ranging from mobile devices such as phones and tablets up to large-scale distributed systems of hundreds of machines and thousands of computational devices such as GPU cards.

Image Pre-processing

To generate the normal kernel images was used the FFMPEG library[1] to split the 6:25 minutes movie into 11,555 individual files (1280x720). Both normal and damaged kernel images were arranged in the separate folder for later use.

Before using the images for generating the neural network, they as randomly allocated in two distinct image sets: 80% and 20% for training and validation set, respectively. In the process of training the Net, other parts of wheat plant structure like spikes and leaves were used too in the composition of the DNN intending to classify better the wheat damaged kernels.

RESULTS AND RATIONALE

Comparing the effort described in (SZEGEDY et al., 2015) to training a whole DNN from scratch with the number of pictures necessary to get reasonable results. In our case, the number of pictures available at the moment ($\approx$ 12,000) probably will not archive this scores and definitely, the time and computational infrastructure necessary to training and evaluate the resultant DNN must be larger than installed capacity.

A broad range of applications are using transfer learning, Devikar (2016) describe the use in image classification of various dog breeds with an overall accuracy of 96% from 11 dog breeds. Wang et al. (2017) describes the application for remote scene classification and attempt to form a baseline for transferring pre-trained DNN to remote sensing images with various spatial and spectral information.

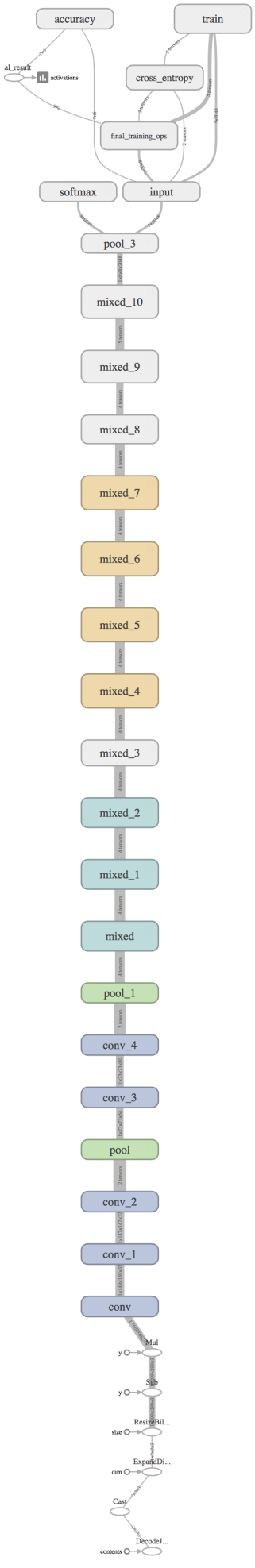
Esteva et al. (2017) describes the use of DNN for dermatologist-level classification of skin cancer, trained end-to-end from images directly, using only pixels and disease labels as inputs for a dataset of 129,450 clinical images.

Tkáč and Verner (2016) presents many business applications, using artificial neural networks, related to financial distress and bankruptcy problems, stock price forecasting, and decision support, with particular attention to classification tasks. For other uses of CNN and DNN, see (SCHMIDHUBER, 2015; GUO et al., 2016).

The training procedure was carried out on the described workstation and took up to 1 hour to finish, and the output retrained neural network archive an overall average accuracy of 94.7%. The main structure of the final neural network is presented in Figure 3.

[1]FFMPEG library, see more information on `http://www.ffmpeg.org`.

Figure 3: An illustration of Deep Convolutional Neural Network, trained using TensorFlow (ABADI et al., 2016) and Transfer Learning techniques from a pre-trained Inception-v3 (SZEGEDY et al., 2015) for an image dataset containing Fursarim Damaged Kernel, normal kernel, spikes and leaves from wheat. The output deep neural network archive an overall average accuracy of 94.7%.



For validation purposes and to check the correctness to classify new images of wheat kernels (with and without Fusarium damage) we choose a new dataset to validate against DNN. Those images were from National Research Centre for Wheat articles published over last decade about this content, alongside other open source images with Fusarium damaged kernel over the Internet.

At this point, we could share our positive experience in using the transfer learning techniques in relation to time to training a DNN with a new set of images and classes and the overall accuracy achieved with this initial image dataset.

The results from validation dataset present a total of 20% of misclassification, and in half (10%) the new images were classified as damaged leaves class. This results could be related to two main reasons: (a) the small number of pictures with Fusarium Damaged Kernels and (b) the prevalence of the normal wheat kernels present in the initial dataset ($\approx$ 80%).

CONCLUSIONS

The associated use of Transfer Learning, TensorFlow, and Inception-v3 cut the time necessary to training and the necessity to have a large image dataset ($\approx$ 120,000) to start the classification procedure with a good accuracy level, compared to training a DNN from scratch.

The misclassification for damaged leaves class could be associated with the characteristics of damage both in kernel and leaves (in the most cases) where the region color of lesions was more blight that the standard wheat kernel.

Unfortunately, the leave damages present in this initial dataset was not separated by diseases what could difficult to claim that some visual characteristics of FHB in leaves could be transferred to kernel evaluations in our context. This supposition needs to be investigated with the addition of new images to actual dataset related to Fusarium Damaged Kernel and wheat leaves with FHB symptoms to test this hypothesis.

Beside this misclassification for a new dataset, the overall average accuracy archived (94.7%) for this Fusarium Damaged Kernels Deep Neural Network (FDK-DNN) exhibits the potential of using this methodology for classifying Fusarium Damaged Kernel without investments on special machinery – only a smartphone camera could to the trick.

The accuracy of the proposed methodology is equivalent to ones using HSI methodology presented by Barbedo, Tibola and Lima (2017).

An interesting future work could be related to using a mixed of RGB pictures, and layers from HSI operational spectra for Fusarium Damaged Kernel proposed by Barbedo, Tibola and Fernandes (2015).

REFERENCES

ABADI, M. et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.

BARBEDO, J. G.; TIBOLA, C. S.; FERNANDES, J. M. Detecting fusarium head blight in wheat kernels using hyperspectral imaging. *Biosystems Engineering*, Elsevier, v. 131, p. 65–76, 2015.

BARBEDO, J. G. A.; TIBOLA, C. S.; LIMA, M. I. P. Deoxynivalenol screening in wheat kernels using hyperspectral imaging. *Biosystems Engineering*, Elsevier, v. 155, p. 24–32, 2017.

DEVIKAR, P. Transfer learning for image classification of various dog breeds. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, v. 5, n. 12, p. 2707–2715, 2016.

ESTEVA, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, Nature Research, v. 542, n. 7639, p. 115–118, 2017.

GUO, Y. et al. Deep learning for visual understanding: A review. *Neurocomputing*, v. 187, p. 27 – 48, 2016. ISSN 0925-2312. Recent Developments on Deep Big Vision.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2012. p. 1097–1105.

MALONEY, P. V. et al. Digital image analysis method for estimation of fusarium-damaged kernels in wheat. *Crop Science*, The Crop Science Society of America, Inc., v. 54, n. 5, p. 2077–2083, 2014.

MIN, H.; CHO, B.-K. Spectroscopic techniques for nondestructive detection of fungi and mycotoxins in agricultural materials: A review. *Journal of Biosystems Engineering*, Korean Society for Agricultural Machinery, v. 40, n. 1, p. 67–77, 2015.

SCHMIDHUBER, J. Deep learning in neural networks: An overview. *Neural Networks*, v. 61, p. 85 – 117, 2015. ISSN 0893-6080.

SZEGEDY, C. et al. Going deeper with convolutions. In: *Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2015.

TKáč, M.; VERNER, R. Artificial neural networks in business: Two decades of research. *Applied Soft Computing*, v. 38, p. 788 – 804, 2016. ISSN 1568-4946.

WANG, J. et al. Transferring pre-trained deep cnns for remote scene classification with general features learned from linear pca network. *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 9, n. 3, p. 225, 2017.

How does atypical weather affect sugar content estimates obtained by machine learning?

Monique Pires Gravina de Oliveira[1], *Luiz Henrique Antunes Rodrigues*[1]

[1] School of Agricultural Engineering, University of Campinas,

monique.oliveira@feagri.unicamp.br, lique@feagri.unicamp.br

ABSTRACT

Current approaches for estimating sugar content in sugarcane culms at harvest do not allow for the inclusion of weather that is expected to occur during the whole cycle. Machine learning techniques have been used to address this concern, as well as to leverage industry data, with good performance. However, the models obtained had not been evaluated regarding the effect of atypical weather. This work investigated estimate errors of models developed with machine learning techniques and data provided by a sugarcane mill. Two evaluation approaches were used: in one of them, the test set was composed from all three years available on the dataset and in the other, the test set isolated one of the years available. In both, the techniques were compared to a simple model based on planting season and ripening pattern. While the advantage of using machine learning was clear when the training set was representative of the test set, this result was not so evident when the test set was composed by an atypical year. This could be due to the approaches used to aggregate the weather time series, meaning better approaches should be sought. Since machine learning models can accommodate more variables, and show more flexibility, they could be further improved, including more detailed genetic information, as well as other environmental and management variables.

KEYWORDS: Weather variability, Crop model, Uncertainty, Sugarcane.

RESUMO

As abordagens atuais para estimativa do teor de açúcar nos colmos da cana-de-açúcar na época da colheita não permitem a inclusão da meteorologia esperada durante o ciclo de crescimento da cana-de-açúcar. Técnicas de aprendizado de máquina têm sido utilizadas para contornar este problema – bem como para extrair informações dos bancos de dados das próprias usinas – e têm apresentado bom desempenho. Entretanto, os modelos já obtidos não foram avaliados quanto ao efeito de eventos meteorológicos atípicos. Este trabalho investigou os erros de modelos desenvolvidos com técnicas de aprendizado de máquina e dados disponibilizados por uma usina de cana-de-açúcar. Duas abordagens para avaliação foram contrastadas: em uma delas, o conjunto de teste continha registros de todas as três safras disponíveis, enquanto na outra, o conjunto de teste era composto por apenas um dos anos-safra. Nos dois casos, as técnicas foram comparadas a modelos simples, baseados na época de plantio e no padrão de maturação. Embora a vantagem no uso de aprendizado de máquina tenha sido evidente quando o conjunto de treinamento era representativo do conjunto de teste, este resultado não foi tão claro quando o conjunto de teste era composto por um ano atípico. Isto pode se dever às abordagens de agregação da série temporal utilizadas, o que significa que abordagens melhores devem ser buscadas. Uma vez que modelos de aprendizado de máquina podem acomodar mais variáveis e mostrar mais flexibilidade, eles ainda podem ser melhorados, incluindo informações genéticas mais detalhadas, bem como outras variáveis ambientais e de manejo.

PALAVRAS-CHAVE: Variabilidade climática, Modelagem de culturas, Incerteza, Cana-de-açúcar.

INTRODUCTION

Sucrose content in sugarcane culms has been estimated by using simple empirical models that include the variety planted and the month of planting or harvesting. This approach was mentioned by Scarpari and Beauclair (2009), and despite the models developed by themselves and by Cardozo *et a*l. (2015), it is still current practice.

One concern regarding these simple models – as well others developed afterward – is that by not allowing for including the weather that is expected to occur during the cycle, they imply an average behavior that was present in the training data, not accounting for variability and thus, not generalizing well. While encouraging the usage of industry data for modeling, Lawes and Lawn (2005) observed how the variables present in the data could be surrogates for other effects.

Recently, Oliveira, Bocca and Rodrigues (2017) proposed data mining and machine learning models derived from production and weather data. Those models addressed this concern because they could accommodate different expected weather data for the whole cycle. While they had shown good performance when they were evaluated under known weather conditions, they have not been evaluated under unknown weather conditions. So, even if the authors have stated that if weather forecasts were precise, the model would present low errors, this has not yet been tested.

Given the influence of weather on sucrose accumulation (CARDOZO; SENTELHAS, 2013; VAN HEERDEN; EGGLESTON; DONALDSON, 2013) and the variability expected due to climate change, it is worth evaluating the impact on model errors of weather conditions that were different from the ones in which the model was developed.

We developed machine learning models following the recommendations of Bocca and Rodrigues (2016). Using a hold-out evaluation approach, we observed the errors when the test set was composed from all three years available on the dataset and when the test set isolated one of the years.

MATERIAL AND METHODS

Software

All analysis were performed using the R environment (R CORE TEAM, 2016). Preprocessing was performed by using packages *dplyr* (WICKHAM; FRANCOIS, 2016), *foreign* (R CORE TEAM, 2015), *lubridate* (GROLEMUND; WICKHAM, 2011), *readxl* (WICKHAM, 2016), *reshape2* (WICKHAM, 2007) and *stringr* (WICKHAM, 2015). Other specific packages will be mentioned in the following topics, along with the respective algorithms.

Data

Data used in this work was provided by the Santa Fe Mill (Usina Santa Fé S. A.), located at Nova Europa (SP). Their annual production area is of 55 thousand hectares, divided in approximately 7,300 fields, which yield 3.5 million tons of fresh mass of sugarcane each year. The spreadsheets that were made available by their team contain:

- Planting and harvesting data from the harvests of 2013/2014, 2014/2015 and 2015/2016.

Each spreadsheet indicates, for each field, how much was harvested and its quality parameters (TRS, pol, purity and Brix). Also, the variety that was planted, planting and harvesting dates and age of the crop.

- Data from their inspections of pests and weeds that happened in 2013, 2014 and 2015.

- Management data regarding the application of fertilizers and agrochemicals, including dates and quantities applied, in the years of 2012, 2013, 2014 and 2015.
- Daily precipitation and maximum, medium and minimum temperatures, taken from the meteorological station of Ibitinga, referring to the years of 2012, 2013, 2014 and 2015.

Pre-processing

Regarding completeness of the data available, from the 1,452 farms present in the dataset, only 674 had information regarding the usage of fertilizers or agrochemicals. Farms that didn't have any information on management were not included in the final dataset. After excluding them, from the original 15,040 records, only 7,465 were left.

Missing values on the weather time series were imputed by being replaced by data from stations located at Nova Europa or Jaboticabal, depending on data availability.

Records with no missing values were then checked for validity of each variable's values. Fields in which the harvest took longer than 30 days were removed as well as those in which the length of the cycle was shorter than 8 months or larger than 20. The removal leaded to a final number of 5,904 records and each record referred to one field in each of the harvesting years.

New variables were created as well, to derive more information for the modeling techniques:

- Weather series were summarized based on sugarcane development stages present in Bocca and Rodrigues (2016) and in Singels, Jones and Van den Berg (2008). The two approaches provided a total of eight periods (four periods each). For maximum, average and minimum temperatures, the median, maximum and minimum values were used for each period, while for precipitation, the sum was used for each period.
- Variables representing the percentage of damaged samples, as determined in the field inspections, referring to damage caused by *Sphenophorus levis* and by *Diatraea saccharalis*.
- Herbicides and insecticides were grouped according to their type of application, and scaled between 0 and 1 by the maximum value applied. Also, herbicides that were applied in the final 100 days were categorized as ripeners.
- Fertilizers' formulas and amounts applied were used to determine how much of each nutrient was applied.

The final dataset had 117 predictor variables, from which 80 refer to the weather, 33 to the management practices and 4 to the crop itself. Tukey's five-number summary for the

target variable – total recoverable sugar (TRS) – in the complete dataset is 80.82, 114.95, 124.61, 134.72 and 189.93 kg Mg^{-1}.

Feature selection

Algorithm RReliefF (KIRA; RENDELL, 1992; KONONENKO; ŠIMEC; ROBNIK-ŠIKONJA, 1997; ROBNIK-ŠIKONJA; KONONENKO, 2003), through the package *CORElearn* (ROBNIK-ŠIKONJA; SAVICKY; ALAO, 2015), was used for feature selection. The number of neighbors was set to 120 and the number of iterations was set to 210, based on a previous evaluation. Also, 10 repetitions of all iterations were used to alleviate the effects of eventual outliers.

Hyperparameter tuning and modeling

Three techniques were used to model sugar content: Support Vector Regression (SVR) with radial kernel, Random Forest (RF) and Regression Tree (RT). We used the packages *e1071* (MEYER *et al*., 2015), *randomForest* (LIAW; WIENER, 2002) and *rpart* (THERNEAU; ATKINSON; RIPLEY, 2015), respectively. Their hyperparameters were defined by random search (BERGSTRA; BENGIO, 2012), within the limits indicated on Table 1. Best parameters were chosen by 10-fold cross validation.

Table 1 – Parameters and inferior and superior limits used on random search for each of the techniques deployed.

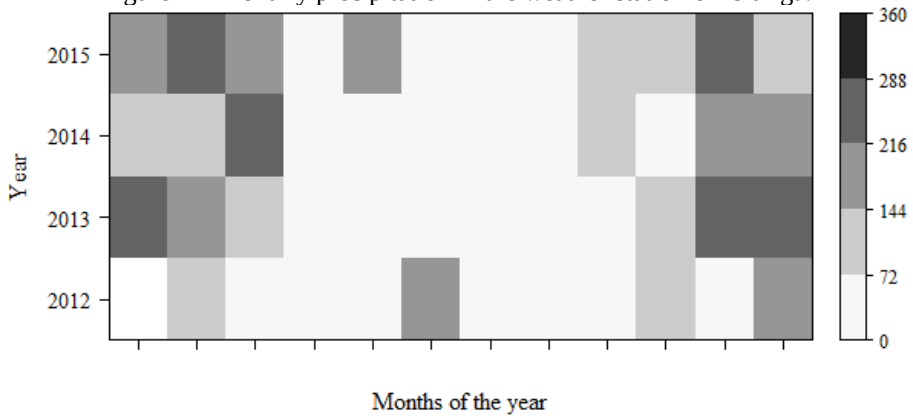
Technique	Number of samples	Parameter	Inferior limit	Superior limit
RT	50	Complexity parameter (*cp*)	0,001	0,1
		Minimum number of observations in a node for attempting a split (*minsplit*)	10	40
RF	50	Number of trees (*ntree*)	100	1000
		Fraction of the number of variables used at each split (*mtry*)	0,13	0,73
		Minimum size of terminal nodes (*node*)	2	25
SVR	50	Cost (*C*)	2^0	2^8
		Gamma (*g*)	2^{-8}	2^1
		Epsilon (*eps*)	0,05	0,2

Evaluation

To assess the impact of weather variability, both a simple hold-out and a leave-one-year-out evaluation were performed. In the former, two-thirds of the dataset was randomly selected for training the models and the remaining one-third, for evaluation, while in the latter, records referring to the harvest in one year were left out from training and used only for evaluation.

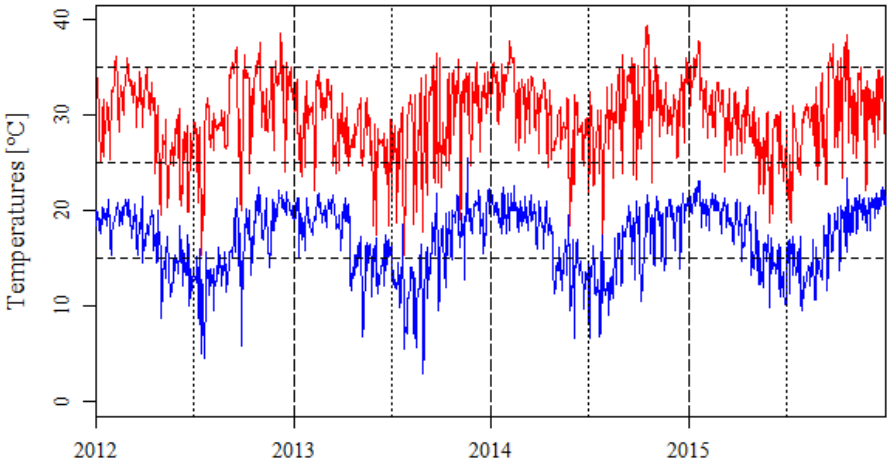
In Figure 1 and Figure 2, one can notice how particularities in weather could lead to very different seasons, e.g. higher temperatures in 2014 and 2015, and the drier period in the summer of 2014.

Figure 1 – Monthly precipitation in the weather station of Ibitinga.



Source: The authors.

Figure 2 – Maximum and minimum daily temperatures in 2012, 2013, 2014 and 2015 at Ibitinga weather station.



Source: The authors.

As a reference of good performance, a null model was defined as the prediction by the mean value on the training set given the ripening pattern (early, medium, late or undefined) and planting season. This model represents the simple models that are currently used.

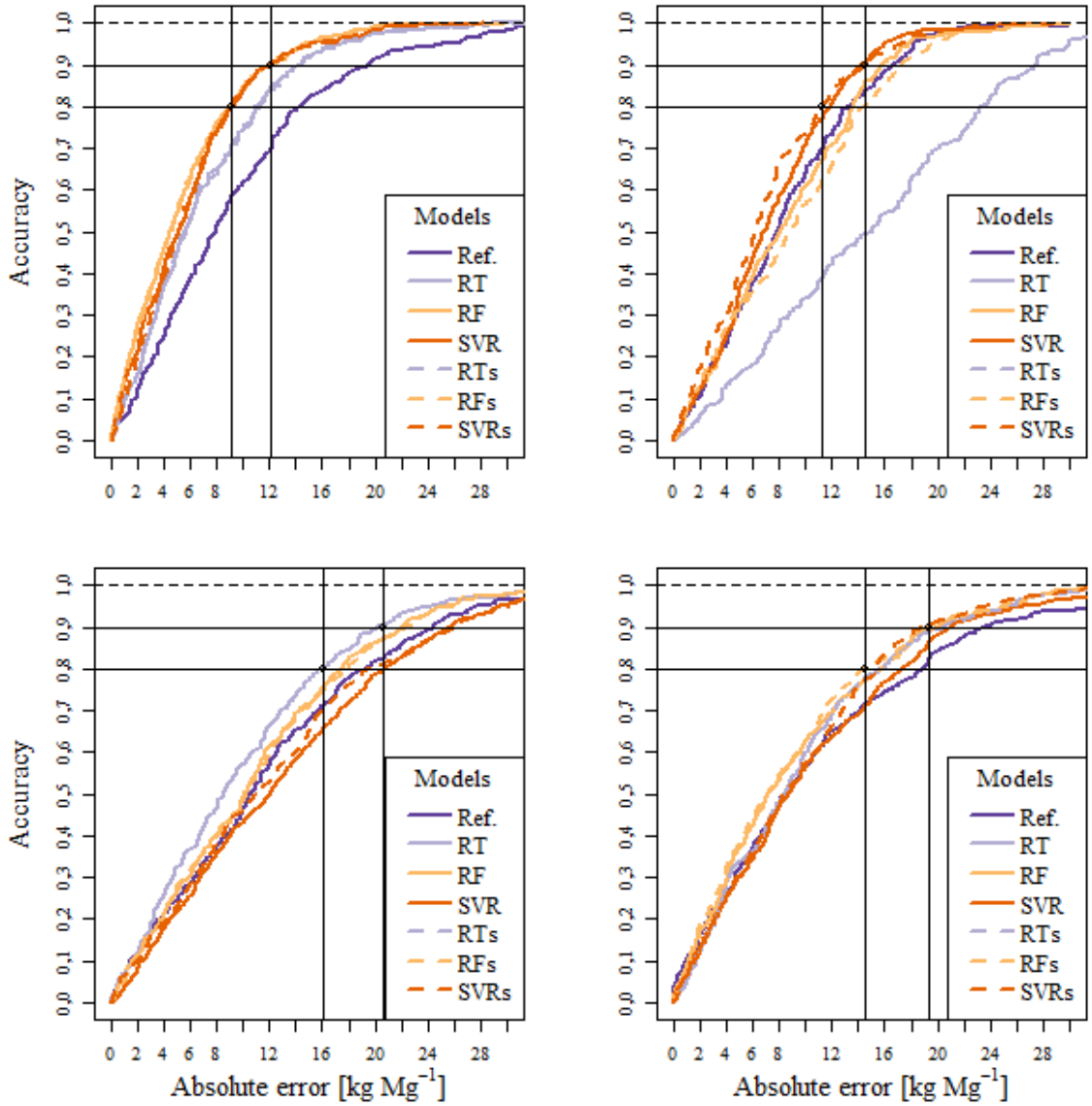
The REC Curve, proposed by Bi and Bennet (2003) for determination of the epsilon value of the SVR, was used to show the cumulative error distribution of all models evaluated.

RESULTS AND DISCUSSION

Similarly to what had been observed by Oliveira, Bocca and Rodrigues (2017), when the training set comprised all years available in the data set, all the models developed with machine learning outperformed the simple model based on planting season (Figure 3a). The

same, however, doesn't happen when the year used for evaluation is not included in the training set. Despite Random Forests with and without feature selection achieving better performances in two out of the three years evaluated, the improvement is not as clear. Additionally, not all techniques presented higher performance (Figure 3b to Figure 3d).

Figure 3 – REC Curves for the Null model (Ref) and the models obtained by Support Vector Regressions (SVR), Random Forests (RF) and the Regression Trees (RT). The letter 's' indicates the variables used were selected by RReliefF. In the first line, left corner (a) are the curves for the simple hold-out evaluation and in the right side (b), for the evaluation on the 2013 season. In the second line, left corner (c) are the curves for the evaluation on the 2014 season and in the right side (d), for the evaluation on the 2015 season.



Source: The authors.

The ranges of errors observed are considered acceptable in production context for all the models, particularly because the results are provided at field scale. It is likely that the models could be further improved by including variables that were not available in the dataset provided by the mill, e.g. those related to soil properties, the occurrence of flowering and closer weather stations.

It could also be the case that the relatively poor performance is due to the weather series aggregation approaches used. The rules proposed by Bocca and Rodrigues (2016) include premises about crop development that do not hold for atypical years, i.e. the length of the developmental stages could vary and the model would not capture this variability when using this aggregation approach. On the other hand, in order to use the stages proposed in Singels, Jones and Van Den Berg (2008), crop specific parameters are needed and few calibrations for sugarcane are available in the literature. This lack of calibrations leaded to using the calibration provided by Marin *et a*l. (2015) for the variety RB86-7515, even though there were other varieties in the dataset. Both approaches had limitations, and the best method has yet to be defined. Everingham *et al*. (2016) used temperature, radiation and precipitation values associated to months of the year, dissociating weather from growth and development. However, given the influence of the weather in sucrose accumulation highlighted by Cardozo and Sentelhas (2013), it is important to include interaction elements between development and weather.

There is not only room for improvement regarding weather series representation, but also, a lot of potential in a new generation of models, that take advantage of machine learning techniques and the recent increase in computational power. Jones *et a*l. (2016) believe the next generation of crop models will be able to optimize the use of resources considering greenhouse gas emissions and expected yield.

Hwang *et a*l. (2016) suggested the inclusion of quantitative trait loci (QTL) in mechanistic crop models, so the models would be less dependent on the calibration of crop specific parameters. For machine learning models, genetic information could replace the surrogate variables variety and ripening pattern. These improvements, however, may not be immediate for sugarcane modeling, due to the crop's genetic complexity (ZHANG et al., 2013).

CONCLUSIONS

While this was a limited experiment, it replicated, with similar and positive results, the approach of modeling sugar content using machine learning and industry data. The advantage of this approach, however, was diminished when the models were confronted with atypical years. Even though the models still performed better than a simple, less robust model, the difference between them was smaller.

There is, however, room for improvement. Machine learning models can accommodate more variables, showing more flexibility, which means they could still include more detailed genetic information, as well as other environmental and management variables.

ACKNOWLEDGEMENTS

We thank Usina Santa Fé S. A. for providing the data and for the support. We also thank the Coordination for the Improvement of Higher Education Personnel (Capes) for providing the scholarship for the first author.

REFERENCES

BERGSTRA, J.; BENGIO, Y. Random Search for Hyper-Parameter Optimization. **Journal of Machine Learning Research**, v. 13, n. feb, p. 281–305, 2012.

BI, J.; BENNET, K. P. Regression error characteristic curves. In: Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), Washington DC. **Anais**... Washington DC: 2003.

BOCCA, F. F.; RODRIGUES, L. H. A. The effect of tuning, feature engineering, and feature selection in data mining applied to rainfed sugarcane yield modelling. **Computers and Electronics in Agriculture**, v. 128, p. 67–76, 2016.

CARDOZO, N. P.; SENTELHAS, P. C. Climatic effects on sugarcane ripening under the influence of cultivars and crop age. **Scientia Agricola**, n. December, p. 449–456, 2013.

CARDOZO, N. P.; SENTELHAS, P. C.; PANOSSO, A. R.; PALHARES, A. L.; IDE, B. Y. Modeling sugarcane ripening as a function of accumulated rainfall in Southern Brazil. **International Journal of Biometeorology**, 2015.

EVERINGHAM, Y.; SEXTON, J.; SKOCAJ, D.; INMAN-BAMBER, G. Accurate prediction of sugarcane yield using a random forest algorithm. **Agronomy for Sustainable Development**, v. 36, n. 2, p. 27, 2016.

GROLEMUND, G.; WICKHAM, H. Dates and Times Made Easy with lubridate. **Journal of Statistical Software**, v. 40, n. 3, p. 1–25, 2011.

HWANG, C.; CORRELL, M. J.; GEZAN, S. A.; ZHANG, L.; BHAKTA, M. S.; VALLEJOS, C. E.; BOOTE, K. J.; CLAVIJO-MICHELANGELI, J. A.; JONES, J. W. Next generation crop models: A modular approach to model early vegetative and reproductive development of the common bean (Phaseolus vulgaris L). **Agricultural Systems**, v. 155, p. 225–239, jul. 2017.

JONES, J. W.; ANTLE, J. M.; BASSO, B.; BOOTE, K. J.; CONANT, R. T.; FOSTER, I.; GODFRAY, H. C. J.; HERRERO, M.; HOWITT, R. E.; JANSSEN, S.; KEATING, B. A.; MUNOZ-CARPENA, R.; PORTER, C. H.; ROSENZWEIG, C.; WHEELER, T. R. Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. **Agricultural Systems**, v. 155, p. 269–288, jul. 2017.

KIRA, K.; RENDELL, L. A. A practical approach to feature selection. In: Proceedings of the ninth international workshop on Machine learning, **Anais**...Morgan Kaufmann Publishers Inc., 1992.

KONONENKO, I.; ŠIMEC, E.; ROBNIK-ŠIKONJA, M. Overcoming the myopia of

inductive learning algorithms with RELIEFF. **Applied Intelligence**, v. 7, n. 1, p. 39–55, 1997.

LAWES, R. A.; LAWN, R. J. J. Applications of industry information in sugarcane production systems. **Field Crops Research**, v. 92, n. 2–3 SPEC. ISS., p. 353–363, jun. 2005.

LIAW, A.; WIENER, M. Classification and Regression by randomForest. **R News**, v. 2, n. 3, p. 18–22, 2002.

MARIN, F. R.; THORBURN, P. J.; NASSIF, D. S. P.; COSTA, L. G. Sugarcane model intercomparison: Structural differences and uncertainties under current and potential future climates. **Environmental Modelling & Software**, v. 72, p. 372–386, 2015.

MEYER, D.; DIMITRIADOU, E.; HORNIK, K.; WEINGESSEL, A.; LEISCH, F. **e1071: Misc Functions of the Department of Statistics (e1071), TU Wien**, 2015.

OLIVEIRA, M. P. G. de; BOCCA, F. F.; RODRIGUES, L. H. A. From spreadsheets to sugar content modeling: A data mining approach. **Computers and Electronics in Agriculture**, v. 132, p. 14–20, 2017.

R CORE TEAM. **foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, ...**, 2015.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**Vienna, Austria, 2016.

ROBNIK-ŠIKONJA, M.; KONONENKO, I. Theoretical and empirical analysis of ReliefF and RReliefF. **Machine learning**, v. 53, n. 1–2, p. 23–69, 2003.

ROBNIK-ŠIKONJA, M.; SAVICKY, P.; ALAO, J. A. **CORElearn: Classification, Regression and Feature Evaluation**, 2015.

SCARPARI, M. S.; BEAUCLAIR, E. G. F. de. Physiological model to estimate the maturity of sugarcane. **Scientia Agricola**, v. 66, n. 5, p. 622–628, 2009.

SINGELS, A.; JONES, M.; VAN DEN BERG, M. DSSAT v4. 5 Canegro sugarcane plant module: scientific documentation. **South African Sugarcane Research Institute**, p. 1–34, 2008.

THERNEAU, T.; ATKINSON, B.; RIPLEY, B. **rpart: Recursive Partitioning and Regression Trees**, 2015.

VAN HEERDEN, P. D. R.; EGGLESTON, G.; DONALDSON, R. A. Ripening and Postharvest Deterioration. In: **Sugarcane: Physiology, Biochemistry, and Functional Biology**. [s.l.] John Wiley & Sons Ltd, 2013. p. 55–84.

WICKHAM, H. Reshaping Data with the reshape Package. **Journal of Statistical Software**, v. 21, n. 12, p. 1–20, 2007.

WICKHAM, H. **stringr: Simple, Consistent Wrappers for Common String Operations**, 2015.

WICKHAM, H. **readxl: Read Excel Files**, 2016.

WICKHAM, H.; FRANCOIS, R. **dplyr: A Grammar of Data Manipulation**, 2016. .

ZHANG, J.; ZHOU, M.; WALSH, J.; ZHU, L.; CHEN, Y.; MING, R. Sugarcane Genetics and Genomics. In: **Sugarcane: Physiology, Biochemistry, and Functional Biology**. [s.l.] John Wiley & Sons Ltd, 2013. p. 623–643.

Active learning e sua aplicação no monitoramento da cana-de-açúcar utilizando o algoritmo *SVM*

João Paulo da Silva[1], *Jurandir Zullo Júnior*[2], *Luciana Alvim Santos Romani*[3]

[1] Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas, Campinas, São Paulo, Brasil, joao.silva@feagri.unicamp.br

[2] Centro de Pesquisas Meteorológicas e Climáticas Aplicadas a Agricultura, Universidade Estadual de Campinas, Campinas, São Paulo, Brasil, jurandir@cpa.unicamp.br

[3] Embrapa Informática Agropecuária, Empresa Brasileira de Pesquisa Agropecuária, Campinas, São Paulo, Brasil, luciana.romani@embrapa.br

RESUMO

A cana-de-açúcar é um dos pilares do agronegócio brasileiro e, por apresentar intensa dinâmica expansionista, demanda metodologias que subsidiem a criação de estratégias políticas e econômicas que promovam a sustentabilidade da produção. Este artigo propõe uma nova abordagem de monitoramento de áreas canavieiras baseada na classificação de séries temporais de imagens de satélite associada à técnica de *Active Learning*. A interação do usuário especialista no aprendizado do algoritmo de classificação através desta técnica utilizando parâmetros sazonais das séries temporais gerou um conjunto de treino otimizado que promoveu a redução do custo operacional de monitoramento da ocupação da cana-de-açúcar. A correlação de cerca de 90% observada entre as análises conduzidas neste trabalho com dados oficiais indica que a metodologia proposta pode ser utilizada no monitoramento agrícola devido à similaridade entre os resultados associada ao baixo custo operacional envolvido.

PALAVRAS-CHAVE: Mineração de dados, Análise sazonal, Classificação, Cana-de-açúcar.

ABSTRACT

Sugarcane is one of the pillars of Brazilian agribusiness and, because of its intense expansionary dynamics, demands methodologies that subsidize the creation of political and economic strategies that promote the sustainability of production. This paper proposes a new approach

for sugarcane areas monitoring based on the classification of satellite images time series associated to the *Active Learning* technique. The interaction of the expert user in the classification algorithm's learning through this technique using time series seasonal parameters generated an optimized training set that promoted operational cost reduction for sugarcane occupation monitoring. The correlation of approximately 90% observed between this study's analysis with official data indicates that the proposed methodology can be used in agricultural monitoring due to the similarity between its results associated with low operating cost involved.
KEYWORDS: Data mining, Seasonal analysis, Classification, Sugarcane.

INTRODUÇÃO

A partir da primeira década de 2000, período de intensificação da produção e comercialização das *commodities* no mundo, a agroindústria canavieira do estado de São Paulo, principal produtor do país, expandiu sua área de ocupação passando de 2,5 milhões para 5 milhões de hectares, segundo a União da Indústria de Cana-de-açúcar (UNICA, 2016). A análise dessa dinâmica expansionista evidencia a necessidade de adotar técnicas de monitoramento da área destinada à cana-de-açúcar para subsidiar a criação de estratégias políticas e técnicas que reduzam seus impactos sobre os recursos naturais existentes (BUCKERIDGE et al., 2012; DAMATTA et al., 2010).

O sensoriamento remoto tem evoluído e se mostrado uma importante ferramenta de monitoramento ambiental. No caso da cana-de-açúcar, destaca-se o programa Canasat[1] (RUDORFF et al., 2010), desenvolvido pelo Instituto Nacional de Pesquisas Espaciais (Inpe) e que, entre 2003 e 2013, monitorou a dinâmica da ocupação canavieira da região Centro-Sul do Brasil. Os levantamentos realizados por esse projeto, entretanto, eram feitos com um grande esforço de interpretação visual, o que poderia limitar sua condução e expansão à disponibilidade de capital humano, especialmente em períodos de desaceleração econômica. Assim, é importante também que as técnicas de monitoramento ambiental que utilizam dados remotos apresentem escalabilidade, para que a intervenção humana e, consequentemente, o seu custo operacional, sejam os menores possíveis.

A utilização de técnicas de mineração de dados em sensoriamento remoto tem demonstrado ser uma alternativa eficiente aos métodos tradicionais para este tipo de aplicação. Diferentes trabalhos envolvendo esta abordagem para monitoramento agrícola vêm sendo propostos para diversas finalidades, dentre as quais está a classificação de áreas agrícolas (XIE

[1] Disponível em: http://www.dsr.inpe.br/laf/canasat/

et al., 2014; ZHANG; FENG; YAO, 2012). Um dos principais entraves neste tipo de tarefa, porém, é a criação de um conjunto de treinamento que expresse as principais características de cada classe envolvida. Para contornar este problema, Crawford, Tuia e Yang (2013) apresentam uma extensa revisão sobre *Active Learning*, uma abordagem para a criação de um conjunto de treinamento baseada na interação do usuário especialista ou em critérios para otimização da criação do conjunto de treino, capaz de reduzir seu tamanho em até 90% sem que haja perda na acurácia dos resultados (TUIA et al., 2009).

Desse modo, esse trabalho teve como objetivo principal apresentar uma metodologia para monitoramento de áreas canavieiras baseada na classificação de séries temporais de imagens de satélite que possa ser adotada como alternativa a métodos tradicionais baseados na interpretação visual de imagens remotas, reduzindo o custo e aumentando a eficiência deste tipo de aplicação.

MATERIAL E MÉTODOS

A área e o período de estudo empregados na condução deste trabalho representam todas as safras agrícolas de cana-de-açúcar ocorridas no estado de São Paulo, principal produtor nacional, entre 2006 e 2012, período de grande expansão da área ocupada com a cultura.

As séries temporais foram construídas a partir de imagens do Índice de Vegetação da Diferença Normalizada (NDVI, em inglês) do sensor MODIS-Terra, com resolução temporal de 16 dias e espacial de 0,25km, sendo cada série representativa do período compreendido entre abril e março do ano seguinte que correspondem, respectivamente, ao início e fim de cada safra.

Preparação e pré-processamento das séries temporais

A preparação e pré-processamento é uma etapa crítica de projetos de mineração de dados que pode influenciar diretamente o sucesso das análises e a robustez dos resultados obtidos. Neste trabalho, esta etapa se deu pela (1) filtragem e extração de parâmetros sazonais das séries temporais; e (2) redução da dimensionalidade do conjunto de dados.

Com o auxílio da ferramenta Timesat (EKLUNDH; JÖNSSON, 2012) foram extraídos parâmetros sazonais das séries temporais relacionados ao tempo cronológico do ciclo vegetativo, como o (1) início; (2) o fim; (3) a duração; e (4) o tempo para máximo vigor vegetativo, e também aos valores assumidos pelo NDVI no ciclo, como (5) o valor máximo; (6) o valor base; (7) a amplitude; (8) a taxa de crescimento; e (9) a taxa de declínio após o máximo vigor vegetativo. Além das séries sazonais (SS) resultantes da extração sazonal, foram mantidas na condução do processamento dos dados séries temporais sem extração da sazonalidade (SES) para avaliação da contribuição deste processo no resultado final.

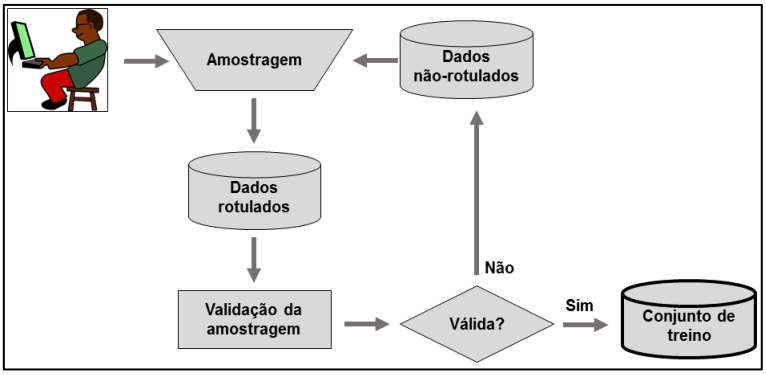
Para a redução de dimensionalidade dos dados, foi aplicada, tanto nas séries SS quanto nas séries temporais SES, a Análise de Componentes Principais (ACP) (RICHARDS, 2012), que por meio da transformação dos atributos originais em autovetores, permite a redução do conjunto de dados para que este contenha apenas os principais autovetores do conjunto inicial. Neste trabalho, tanto para as séries sazonais quanto para as séries temporais SES, foram retidos apenas os autovetores que apresentaram autovalores superiores a 0,7, os quais, segundo Jolliffe (1972) e Jolliffe (1973), detêm cerca de 90% da variância do conjunto de dados original.

Classificação das séries temporais

A classificação das séries sazonais e SES foi feita utilizando-se o algoritmo *Support Vector Machine* (SVM) (CORTES; VAPNIK, 1995), considerado como um dos algoritmos mais sofisticados atualmente para análise de dados remotos (MOUNTRAKIS; IM; OGOLE, 2011) em função do seu desempenho em situações onde o tamanho do conjunto de treino é reduzido (MANTERO; MOSER; SERPICO, 2005) ou quando o número de atributos é elevado, como no caso de séries temporais (WANG et al., 2010). Seu funcionamento baseia-se na escolha de registros essenciais (ou vetores suporte) do conjunto de treino para a determinação de um hiperplano otimizado (e suas margens) de separação dos dados em duas classes distintas.

Para aplicação da abordagem *Active Learning* (Figura 1), foram estabelecidos critérios de seleção por parte do usuário especialista que expressassem, confiavelmente, características espectrais exclusivas de cada uma das classes amostradas. Com o auxílio da ferramenta ENVI, versão 5.2, foram amostradas regiões representativas de seis diferentes classes de cobertura da terra – *área urbana*, *hidrografia*, *área de mata*, *cultura anual*, *cana-de-açúcar* e *pastagem* – a partir de imagens formadas por composições das séries sazonais para serem utilizadas na construção de um conjunto de treino, sendo que o critério adotado para avaliação da distinção entre as amostras foram as medidas de distância Jeffries-Matusita (JM) e a Divergência Transformada (DT) (RICHARDS, 2012). Tomando como referência o estudo de classificação de uso e cobertura da terra realizado por Cai e Zhang (2010), foi definido que a amostragem de uma classe deveria ser refeita até que o grau de separabilidade medido entre esta e as outras classes fosse igual ou superior a 1,9. Para comparação dos resultados da classificação, as regiões amostradas sobre as composições das séries sazonais foram replicadas, sem alterações, na classificação das séries temporais SES.

Figura 1. Fluxograma de aplicação da *Active Learning* para construção do conjunto de treino otimizado.



Fonte: Autor

As estimativas anuais de área plantada geradas a partir da classificação das séries temporais foram cruzadas com dados do projeto Canasat/INPE e de outras metodologias de estimativa de área plantada utilizadas por órgãos oficiais para comparação e validação dos resultados.

RESULTADOS E DISCUSSÃO

Preparação e pré-processamento dos dados

Com a extração dos parâmetros sazonais das séries temporais, além da atenuação primária de ruído, verificou-se também que parâmetros sazonais relacionados ao vigor vegetativo se mostraram mais relevantes que aqueles ligados ao tempo cronológico na distinção de diferentes coberturas vegetais na área estudada. Contudo, os períodos de início, meio e fim do ciclo fenológico da vegetação também foram coincidentes com o clima da região, caracterizado por um período quente e úmido no início e fim do ano, quando o desenvolvimento vegetativo é mais intenso, seguido por outro mais frio e seco no intervalo, quando ocorre o repouso da vegetação, segundo dados do Instituto Nacional de Meteorologia (INMET) (2017).

A partir da aplicação da ACP nas séries sazonais, foram retidos dois dos nove autovetores originais em todas as safras analisadas. Os autovetores retidos corresponderam, em média, a 87% da variância do conjunto original de dados. No caso das séries temporais SES, com exceção das safras 2006-2007 e 2009-2010, para as quais foi retido apenas um, todas as outras safras resultaram também na retenção de dois autovetores, e a variância média contida nestes autovetores resultantes foi de cerca de 93%. Utilizando a abordagem da aplicação da ACP em dados remotos para modelagem ecossistêmica do Pantanal, Almeida et al. (2015) conseguiram descrever 99% da variância contida no conjunto original de dados nos três primeiros autovetores resultantes da aplicação da ACP. Réjichi e Chaabane (2015), por sua vez,

em trabalho de classificação de séries temporais, conseguiram atingir 90% de acerto com o modelo utilizando apenas o primeiro autovetor resultante da aplicação da ACP.

Classificação das séries temporais

A construção do conjunto de treino do modelo de classificação a partir da abordagem *Active Learning* satisfez os critérios estabelecidos em 100% dos casos para as séries sazonais, atingindo, na maioria dos casos, o grau máximo separabilidade (2,0) entre as classes para ambas as medidas de distância, JM e DT. No trabalho realizado por Cai e Zhang (2010), a distância JM obtida pelos autores variou entre 1,86 e 1,98, que na média foram piores que as distâncias obtidas no presente trabalho. Na replicação da medida de distância entre as classes para os mesmos pontos amostrais utilizando as séries SES, contudo, os resultados obtidos foram divergentes em 19% das comparações válidas, principalmente naquelas envolvendo as classes *cultura anual* e *cana-de-açúcar*, para as quais as distâncias JM e DT médias foram, respectivamente, 0,78 e 0,66. Nas safras 2006-2007 e 2009-2010, para as quais a aplicação da ACP resultou na retenção de apenas 1 autovetor, não foi possível realizar o cálculo da distância entre as classes devido a restrições na aplicação das metodologias.

Na validação do mapeamento da área plantada com cana-de-açúcar, foram amostrados pontos aleatórios na área de estudo a partir do mapeamento do projeto Canasat/INPE para comparação com os resultados gerados pela classificação das séries sazonais e séries temporais SES. Os resultados da validação do modelo de classificação para a classe minoritária (*cana-de-açúcar*) e para todas as classes estão descritos, respectivamente, nas Tabelas 1 e 2.

Tabela 1. Valores médios das métricas de avaliação da classificação das séries sazonais e séries temporais sem extração sazonal (SES) referentes à classe minoritária.

Métrica	Séries sazonais	Séries temporais SES
Falso positivo	57%	61%
Falso negativo	60%	59%
Recall	40%	41%
Especificidade	85%	80%
Precisão	43%	39%

Tabela 2. Valores médios das métricas de avaliação da classificação das séries sazonais e séries temporais sem extração sazonal (SES) referentes a classificação geral.

Métrica	Séries sazonais	Séries temporais SES
Kappa	25%	21%
Acurácia Geral	75%	72%
Recall	63%	61%
Especificidade	63%	61%
Precisão	30%	28%

Apesar de ser amplamente utilizado em trabalhos sobre classificação de dados remotos, o uso do SVM engloba, em sua maioria, a classificação de dados com maior resolução espacial

e espectral (MOUNTRAKIS; IM; OGOLE, 2011). Sua aplicação em séries temporais, no entanto, ainda representa um desafio. Em alguns exemplos encontrados sobre o tema (XIE et al., 2014; XUE; DU; FENG, 2014), nota-se a agregação de outros parâmetros, como o uso de imagens de maior resolução espacial e/ou outras bandas espectrais, para subsidiar o algoritmo e melhorar a precisão da classificação final. Neste trabalho, foram utilizadas apenas séries temporais de NDVI e os resultados da sua análise sazonal.

Em trabalho relativo à classificação de séries temporais de NDVI para monitoramento da vegetação, Carrão, Gonçalves e Caetano (2008) concluíram que a variabilidade espectral foi mais determinante que a temporal na classificação da cobertura vegetal, destacando que coberturas que apresentem características semelhantes podem ter sua distinção prejudicada na classificação final. No estado de São Paulo, o comportamento espectral da cana-de-açúcar assemelha-se ao das pastagens (XAVIER et al., 2006), também amplamente cultivada no estado. No trabalho desenvolvido por Cai e Zhang (2010), que apresenta fundamentos similares aos que foram aplicados neste trabalho, os autores obtiveram acurácia geral e índice Kappa acima de 90%, porém, enfatizaram que a sua amostragem para validação da classificação foi feita sobre áreas estrategicamente amostradas, ao passo que neste trabalho a amostragem para validação espacial do modelo foi feita inteiramente ao acaso para evitar o enviesamento dos resultados.

Na comparação da área plantada estimada pelo modelo com as estimativas do Canasat/INPE, Instituto Nacional de Geografia e Estatística (IBGE) (IBGE, 2017) Companhia Nacional do Abastecimento (Conab) (CONAB, 2017) (Tabela 3), foram observadas semelhanças que elucidam o potencial da metodologia apresentada, especialmente em relação à classificação feita utilizando as séries sazonais (SS).

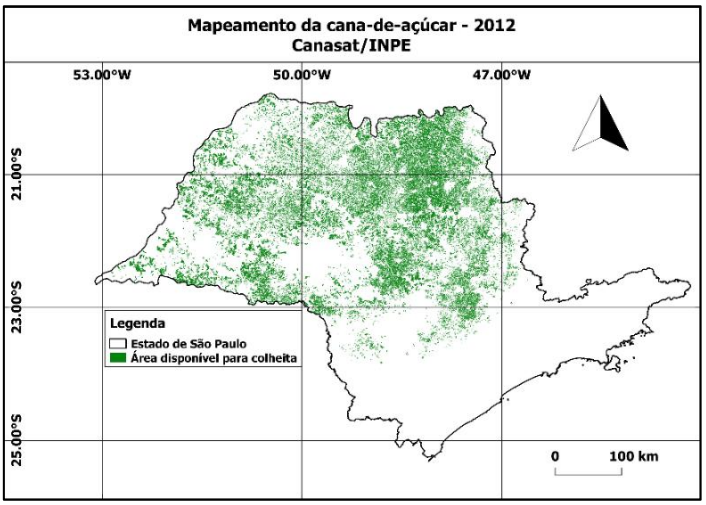
Tabela 3. Estimativa de área plantada com cana-de-açúcar pelo modelo de classificação das séries temporais e por estimativas oficiais.

Safra	Séries sazonais (SS)	Séries temporais (SES)	Canasat/INPE	IBGE	Conab
	Área plantada (hectare)				
2006-2007	4.195.048	2.128.026	3.016.262	3.498.265	3.288.200
2007-2008	2.890.060	4.193.774	3.289.761	3.890.414	3.824.200
2008-2009	4.048.791	2.194.391	3.746.039	4.541.509	3.882.100
2009-2010	2.485.076	6.060.411	4.562.179	4.977.077	4.129.900
2010-2011	4.511.049	2.850.983	4.810.327	5.071.205	4.357.000
2011-2012	4.298.490	6.590.394	4.664.610	5.216.491	4.370.100
2012-2013	4.675.054	8.232.584	4.601.335	5.172.611	4.419.460

Os dados observados na Tabela 3 mostram que as estimativas geradas pela classificação das séries sazonais, com exceção das safras 2006-2007 e 2009-2010, tiveram uma boa correlação com as estimativas oficiais utilizadas para levantamento da área cultivada com cana-de-açúcar. O índice de correlação de Pearson (r), gerado sem considerar as duas safras

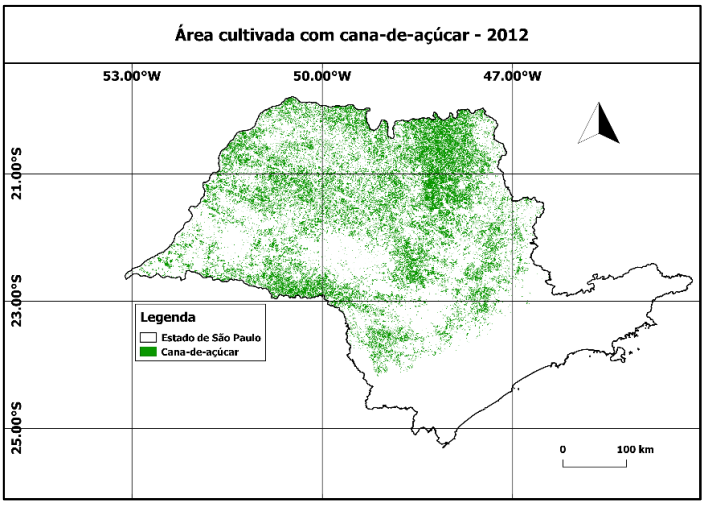
mencionadas acima, indicou correlação de 0,9 entre a estimativa gerada utilizando as séries sazonais com o mapeamento do Canasat/INPE; 0,84 com as estimativas da Conab (2017) e 0,95 com as estimativas do IBGE (2017). No caso das séries temporais SES, este índice foi de, no máximo, 0,61, mesmo após desconsiderar as safras 2006-2007 e 2009-2010. As Figuras 2 e 3 ilustram a distribuição da área plantada com cana-de-açúcar pela metodologia utilizada no projeto Canasat/INPE e pela classificação das séries sazonais.

Figura 2. Mapeamento da área cultivada com cana-de-açúcar pelo projeto Canasat/INPE.



Fonte: Adaptado de RUDORFF et al. (2010)

Figura 3. Mapeamento da área cultivada com cana-de-açúcar pela classificação das séries sazonais.



Fonte: Autor

Conclusão

Modelos de classificação de séries temporais de NDVI pelo algoritmo SVM para identificação de áreas canavieiras tem sua taxa de acerto dependente da cobertura do terreno por outros tipos de vegetação, mesmo quando aplicadas técnicas que promovam melhor diferenciação da cobertura vegetal, como a análise sazonal.

A abordagem *Active Learning* adotada na construção do conjunto de treino do modelo se mostrou eficiente na construção de um conjunto de treino compacto, mas que expressasse as principais características espectrais dos alvos a partir dos critérios de amostragem estabelecidos. Metodologias tradicionais de monitoramento agrícola baseado em imagens de satélite são muito dependentes do esforço humano demandado para a obtenção dos resultados, principalmente aqueles baseados em interpretação visual de imagens.

Excluindo-se as safras ocorridas em 2006-2007 e 2009-2010, a estimativa de área plantada obtida pela classificação das séries sazonais apresentou correlação variando entre 84% e 95% com as estimativas oficiais, porém com custo operacional sensivelmente menor. Com isso, a metodologia proposta pode ser adotada em conjunto com outras fontes de dados e outros mapeamentos afim de aumentar a precisão no mapeamento da cana-de-açúcar no estado de São Paulo.

Referências

ALMEIDA, T. I. R. et al. Principal component analysis applied to a time series of MODIS images: the spatio-temporal variability of the Pantanal wetland, Brazil. Wetlands ecology and management, v. 23, n. 4, p. 737-748, 2015.

BUCKERIDGE, M. S. et al. Ethanol from sugarcane in Brazil: a 'midway'strategy for increasing ethanol production while maximizing environmental benefits. GCB Bioenergy, v. 4, n. 2, p. 119-126, 2012.

CAI, H.; ZHANG, S. Regional land cover classification from MODIS time-series and geographical data using support vector machine. In: Information Computing and Telecommunications, 17610., 2010, Beijing. Anais… Beijing: IEEE, 2010. p. 102-105.

CARRÃO, H.; GONÇALVES, P.; CAETANO, M. Contribution of multispectral and multitemporal information from MODIS images to land cover classification. Remote Sensing of Environment, v. 112, n. 3, p. 986-997, 2008

COMPANHIA NACIONAL DO ABASTECIMENTO. Séries históricas. Disponível em: <http://www.conab.gov.br/ >. Acesso em: 09/05/2017.

CORTES, C.; VAPNIK, V. Support-vector networks. Machine learning: Springer, v. 20, n. 3, p. 273-297, 1995.

CRAWFORD, M. M.; TUIA, D.; YANG, H. L. Active Learning: Any value for classification of remotely sensed data? Proceedings of the IEEE, v. 101, n. 3, p. 593-608, 2013.

DAMATTA, F. M. et al. Impacts of climate changes on crop physiology and food quality. Food Research International, v. 43, n. 7, p. 1814-1823, 2010.

EKLUNDH, L.; JÖNSSON, P. TIMESAT 3.2 with parallel processing-Software Manual. Relatório técnico, 2012. 88p.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Produção agrícola municipal. Disponível em: <http://www.ibge.gov. br/>. Acesso em: 20 mai. 2017.

INSTITUTO NACIONAL DE METEOROLOGIA. Normais climatológicas. Disponível em: < http://www.inmet.gov.br/portal/>. Acesso em: 21 mai. 2017.

JOLLIFFE, I. T. Discarding variables in a principal component analysis. I: Artificial data. Applied statistics, p. 160-173, 1972.

JOLLIFFE, I. T. Discarding variables in a principal component analysis. II: Real data. Applied statistics. p. 21-31, 1973.

MANTERO, P., MOSER, G., SERPICO, S. B. Partially supervised classification of remote sensing images through SVM-based probability density estimation. IEEE Transactions on Geoscience and Remote Sensing, v. 43, n. 3, p. 559–570, 2005.

MOUNTRAKIS, G.; IM, J.; OGOLE, C. Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing, v. 66, n. 3, p. 247-259, 2011.

RÉJICHI, S.; CHAABANE, F. Feature extraction using PCA for VHR satellite image time series spatio-temporal classification. In: Geoscience and Remote Sensing Symposium, 2015, Milan. Anais… Milan: IEEE, 2015., p. 485-488.

RICHARDS, J. A. Remote sensing digital image analysis: An introduction. 3.ed. Berlin: Springer, 2012. 503p.

RUDORFF, B. F. T. et al. Studies on the rapid expansion of sugarcane for ethanol production in São Paulo State (Brazil) using Landsat data. Remote Sensing, v. 2, n. 4, p. 1057-1076, 2010.

TUIA, D. et al. Active Learning methods for remote sensing image classification. Geoscience and Remote Sensing, v. 47, n. 7, p. 2218-2232, 2009.

UNIÃO DAS INDÚSTRIAS DE CANA-DE-AÇÚCAR. Dados da produção canavieira. Disponível em: <http://www.unicadata.com.br/>. Data de acesso: 15/05/2017.

WANG, K. et al. Remote sensing of ecology, biodiversity and conservation: a review from the perspective of remote sensing specialists. Sensors, v. 10, n. 11, p. 9647-9667, 2010.

XAVIER, A. C. et al. Multi-temporal analysis of MODIS data to classify sugarcane crop. International Journal of Remote Sensing, v. 27, p. 755-768, 2006.

XIE, D. et al. Autumn crop identification using high-spatial-temporal resolution time series data generated by MODIS and Landsat remote sensing images. In: Geoscience and Remote Sensing Symposium, 2014, Québec, Anais… Québec: IEEE. 2014., p. 2118-2121.

XUE, Z.; DU, P.; FENG, L. Phenology-driven land cover classification and trend analysis based on long-term remote sensing image series. Journal of Selected Topics in Applied Earth Observations and Remote Sensing, v. 7, n. 4, p. 1142-1156, 2014.

ZHANG, J.; FENG, L.; YAO, F. Improved maize cultivated area estimation over a large scale combining MODIS–EVI time series data and crop phenological information. Journal of Photogrammetry and Remote Sensing, v. 50, n. 9, p. 102-113, 2012.

Agritempo GIS: Apoio para a tomada de decisão no campo

Bruno José Alves[1], *Rachel Scrivani*[1], *Silvio Roberto Medeiros Evangelista*[1], *Luciana Alvim Santos Romani*[1]

[1] Grupo de Sistemas de Informação, Embrapa Informática Agropecuária, Campinas, São Paulo, Brasil, jalvesbruno@gmail.com, rachelscrivani@gmail.com, silvio.evangelista@embrapa.br, luciana.romani@embrapa.br

RESUMO

O uso de aplicativos móveis torna-se cada vez mais importante no meio agrícola, devido à praticidade dos dispositivos que podem ser consultados em qualquer local da propriedade. Tais dispositivos atendem às necessidades do agricultor em termos de espaço e conectividade, facilitando o processo de tomada de decisão. Este artigo apresenta a contribuição do Agritempo GIS para a agricultura, destacando as etapas do seu processo de criação e validação. O aplicativo móvel Agritempo GIS promove o auxílio nos processos de tomada de decisão no campo, ao transportar de maneira concisa as informações georreferenciadas do sistema web Agritempo para os dispositivos móveis. A validação do aplicativo foi realizada com especialistas. Esta validação contribuiu para o desenvolvimento e aperfeiçoamento do aplicativo, por meio da criação da tela de localizações favoritas, do botão de favoritos, do botão GPS, e de métodos para manter a localização selecionada ao trocar de camada, que melhoraram a interação do usuário com o aplicativo.

PALAVRAS-CHAVE: Mapas, Monitoramento Agrometeorológico, Georreferência, Mobile.

ABSTRACT

The use of mobile applications become increasingly important in the agricultural environment, due to the practicality of the devices that can be used anywhere on the farm. Such devices support the farmer's needs in terms of space and connectivity, making the decision-making process easy. This paper presents the contribution of Agritempo GIS to agriculture, highlighting the stages of its creation and validation process. The Agritempo GIS

mobile application promotes aid in decision-making processes in the field by concisely transporting georeferenced information from the Agritempo web system to mobile devices. The validation of the application was carried out by specialists. This validation contributed to the development and refinement of the application creating the favorite locations screen, the Favorites button, the GPS button, and methods to maintain the selected location when switching layers, which improved user interaction with the application.

KEYWORDS: Maps, Agrometeorological Monitoring, Georeference, Mobile.

INTRODUÇÃO

No mundo, aproximadamente 2,5 bilhões de pessoas estão envolvidas em tempo integral ou parcial na agricultura, administrando 500 milhões de pequenas propriedades rurais – das quais 80% estão na mão de pequenos agricultores, e respondem pela produção de mais de quatro quintos de alimentos consumidos nos países em desenvolvimento (SACHS, 2014).

O planejamento rural tem certas peculiaridades que devem ser consideradas no processo administrativo da propriedade rural, pois há vários pontos que interferem nas atividades agrícolas, como o clima, a terra, as pragas e as doenças (METZNER et al., 2014).

O papel da Tecnologia da Informação (TI) na agricultura, segundo ALBERT e ONWUBUYA (2013), se classifica em duas seções gerais. Na primeira, as TIs servem como uma ferramenta para aumentar diretamente a produtividade, enquanto na segunda seriam um instrumento indireto utilizado para a capacitação dos agricultores, a fim de que sejam capazes de tomar decisões para impulsionar seus empreendimentos agrícolas, assim obtendo retornos sobre os seus investimentos.

Corroborando este raciocínio, uma pesquisa realizada com os agricultores de Santa Maria-RS revelou que quase 90% dos produtores rurais possuem telefone móvel. No campo, seu uso pode auxiliar na resolução de problemas da propriedade ao evitar deslocamentos até centro urbanos e no fechamento de negócios, sendo considerado essencial para seu dia a dia (SILVEIRA; SCHWARTZ, 2011).

Além disso, segundo Suarez e Suarez (2013), a implantação de aplicativos móveis resultou em economia para os produtores, pois seu uso permitiu a redução dos custos de produção, diminuição da perda na colheita e diminuição de viagens para descobrir a variação dos preços, o que gerou significativo incremento no lucro dos produtores rurais.

Considerando os benefícios advindos com o uso de dispositivos móveis na agricultura, por meio do uso de aplicativos, é de extrema importância o desenvolvimento dos mesmos. O

aplicativo Agritempo GIS, abordado neste trabalho, foi desenvolvido para ser utilizado na plataforma Android para a versão Icre Cream Sandwich ou superior.

Tal como se encontra na síntese descritiva do aplicativo Agritempo GIS, na Google Play Store, o aplicativo móvel Agritempo GIS, fornece o acesso à funcionalidade WebGIS do sistema Agritempo (www.agritempo.gov.br). O app permite o acesso a dados agrometeorológicos de maneira facilitada para diversos estados e municípios brasileiros. O Agritempo GIS possui as seguintes funcionalidades: monitoramento e previsão agrometeorológicos, índice de seca, e previsão de geadas. O usuário pode selecionar diferentes camadas, incluindo os biomas brasileiros.

Um sistema GIS, tal como o aplicativo abordado neste artigo, segundo Zhang e Francis (2013), para o *Survey* de geologia U.S, é definido como "um sistema computacional capaz de capturar, armazenar, analisar e exibir informações referenciadas geograficamente, ou seja, dados identificados de acordo com a localização".

Desta maneira, o aplicativo Agritempo GIS tem por função, além de auxiliar o agricultor por meio do monitoramento agrometeorológico para uma certa região, a ser escolhida, facilitar a visualização dos dados do sistema Web Agritempo já existente, aprimorando a experiência do agricultor, por meio de uma interface eficiente e intuitiva.

Utilizando este monitoramento, o agricultor pode se orientar melhor com relação à quantia de insumos agrícolas que devem ser utilizados e ao momento correto para utilizá-los, por exemplo. Sendo assim, este trabalho tem por objetivo mostrar, por meio da abordagem das etapas do seu processo de criação, como o aplicativo móvel, Agritempo GIS, pode auxiliar o agricultor nos processos de tomada de decisão no campo.

O restante deste artigo está organizado da seguinte forma: a próxima Seção define os materiais e métodos utilizados, a Seção Resultados e Discussão mostra de forma detalhada os resultados e melhorias obtidos após a fase de Validação e finalmente é apresentada a Conclusão do trabalho.

MATERIAL E MÉTODOS

Inicialmente, para desenvolver os protótipos das interfaces de tela, foram observados aplicativos, tais como o Google Maps e o Waze, que se destacam na Google Play Store, a loja de aplicativos Android. O sucesso desses aplicativos se deve a quantidade de usuários que os qualificaram com cinco estrelas (nota máxima). Segundo dados da Google Play Store, em 4 de maio de 2017, para o Google Maps, tem-se: 5.096.942 avaliações com cinco estrelas, de um total de 7.881.431 avaliações. E, para o Waze, tem-se: 4.399.900 avaliações com cinco

estrelas de um total de 6.224.795 avaliações. Foi observado que estes aplicativos tinham uma tela de fundo, cujo conteúdo é alterado de acordo com a opção escolhida pelo usuário em um menu lateral.

Sendo assim, foi iniciada a construção dos protótipos, sempre objetivando mostrar as informações para o usuário final de uma maneira concisa e prática. Para isso, houve a participação do usuário durante o desenvolvimento do aplicativo. Tal prática, denominada Design Centrado no Usuário (DCU), segundo Magalhães e Romani (2014), faz com que se torne imprescindível a atenção dos desenvolvedores às necessidades e dificuldades específicas do público que será contemplado pela aplicação.

O método utilizado foi o Scrum, pois era preciso gerar um executável usável a cada duas semanas. Além disso, segundo Kenneth (2012), Scrum é um framework leve, que é simples de entender e é capaz de gerenciar produtos complexos auxiliando no desenvolvimento que é estabelecido com algum prazo. Este framework consiste de um time Scrum que tem suas atividades, artefatos e regras. O time Scrum é composto pelo *product owner*, *Scrum master* e pelo time de desenvolvimento. Em Scrum, o produto do software é entregue por meio de uma série de iterações ou incrementos que são predefinidos em uma *time-box* (normalmente de trinta dias), denominada *sprint*.

Junto ao Scrum, foram utilizadas as fases do ciclo de vida para o desenvolvimento de aplicações móveis: *Mobile Application Development Life Cycle* (MADLC) que são: identificação, concepção, desenvolvimento, protótipos, testes, implantação e manutenção. Segundo Kumar e Vithani (2014), esse ciclo de vida também aborda algumas das características distintas de aplicações móveis, como expectativa de vida, funcionalidades complexas, menos interfaces físicas, mais números de telas de interação, uso de bateria e memória, desenvolvimento de plataforma cruzada e manutenção.

Como seria possível aproveitar a estrutura do sistema GIS já disponível na versão web do Agritempo, foi decidido que em vez de converter as camadas existentes para algum serviço de mapas nativo para Android seria usada uma página web com a biblioteca Javascript OpenLayers, que é responsável por aplicar as diferentes camadas agrometeorológicas sobre o mapa.

Ao executar a biblioteca OpenLayers, em um componente de layout do Android chamado WebView, um arquivo HTML é exibido dentro do aplicativo. Como o fluxo de uso planejado para o aplicativo necessitava de troca de informações entre os componentes nativos do Android e a biblioteca OpenLayers, foi realizada uma breve pesquisa e concluiu-se que

seria necessário utilizar a ferramenta JavaScript Interface, que torna métodos Java acessíveis por chamadas JavaScript.

A biblioteca OpenLayers permite o uso de diversos motores de mapa, e a escolha inicial para o projeto foi o Google Maps. Com os primeiros testes do sistema inicial em diversos dispositivos, notou-se uma baixa performance e erros de alinhamento dos mapas com as camadas que eram exibidas. O motor de mapas que se saiu melhor nos testes e, consequentemente, foi escolhido para o projeto foi o OpenStreetMaps. Conforme as funcionalidades eram implementadas, a disposição dos elementos na tela e o layout como um todo (margens, fontes, espaçamento, cores) eram periodicamente revisados e alterados quando necessário, tornando gradual o processo de criação e de testes da interface do aplicativo.

A Figura 1, ilustra a maneira como o aplicativo funciona:

Figura 1 – Modo de funcionamento do aplicativo Agritempo GIS.



Fonte: Alves et al. (2016)

Após 6 meses de desenvolvimento do aplicativo, foi realizada uma validação com 11 participantes, dentre os quais haviam especialistas em Ciência da Computação e Agronomia, pesquisadores, estudantes e profissionais da área de Transferência e Tecnologia. O objetivo desta validação foi a detecção de erros no aplicativo e avaliação da interface, para que menos erros fossem reportados quando o mesmo fosse publicado na Google Play Store. Como realizar esta validação e detectar erros seria algo complexo e o tempo para desenvolver as correções era curto, foi decidido realizar a validação por meio de um questionário baseado em heurísticas, pois, segundo SLOVIC et al. (2000), a heurística é um processo que resulta basicamente da experiência de vida das pessoas que utilizam da memória e dos registros que ela nos permite fazer para "driblar" a complexidade, a escassez de recurso e a pressão do tempo em tomadas de decisão.

Neste questionário baseado em heurísticas, foram propostos quatro cenários com casos de usos para serem realizados em alguns smartphones e tablets Android. No primeiro cenário,

na categoria “Previsão”, acessada pelo menu lateral, deveriam ser selecionados a camada necessidade de irrigação e algum ponto do estado de São Paulo. Em seguida, o *shape* biomas deveria ser selecionado.

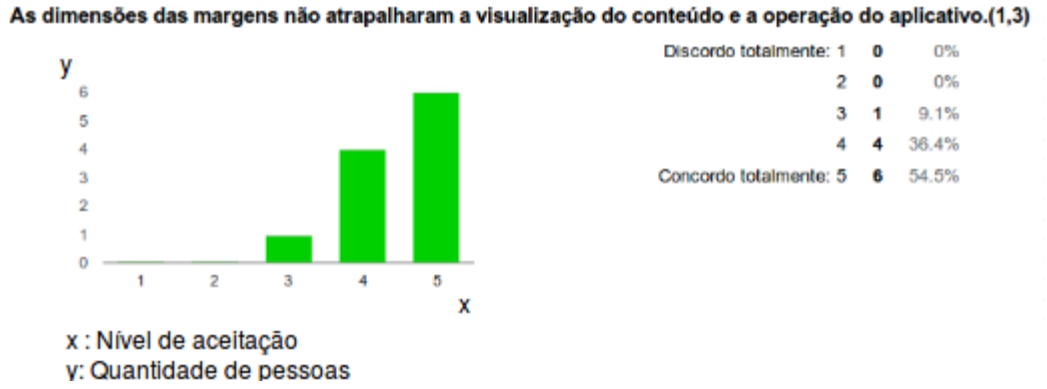
No segundo cenário, foi pedido para realizar a seleção da categoria “Índice de Seca”, pelo menu lateral, e, em seguida, a camada “SPI – Últimos 24 meses” e que a opacidade da camada fosse selecionada. No terceiro cenário, foi solicitado a seleção da categoria “Monitoramento”, por meio do menu lateral, e que, em seguida, fosse selecionada a camada “Temperatura Média” e a funcionalidade GPS fosse ativada.

Após isso, o avaliador deveria clicar em algum ponto da região do mapa e verificar o valor da “Temperatura Média”. No quarto cenário, o avaliador deveria selecionar a categoria “Previsão de geadas”, por meio do menu lateral, e selecionar a camada “A 6°C por 48h”. Após isso, foi pedido para clicar em algum ponto do mapa e verificar se haveria possibilidade de geada para o ponto que foi selecionado. Na próxima Seção são detalhados os resultados do processo de Validação.

RESULTADOS E DISCUSSÃO

Os resultados da validação são apresentados por meio de gráficos (Figura 2 à Figura 4), e pela tabela 1 a seguir. A figura 2 mostra um gráfico sobre a interferência das dimensões das margens de visualização do conteúdo e operação do aplicativo. Segundo 91% dos avaliadores, pertencentes as escalas 4 e 5, as dimensões das margens como definidas no aplicativo não atrapalham a visualização do conteúdo e sua operação.

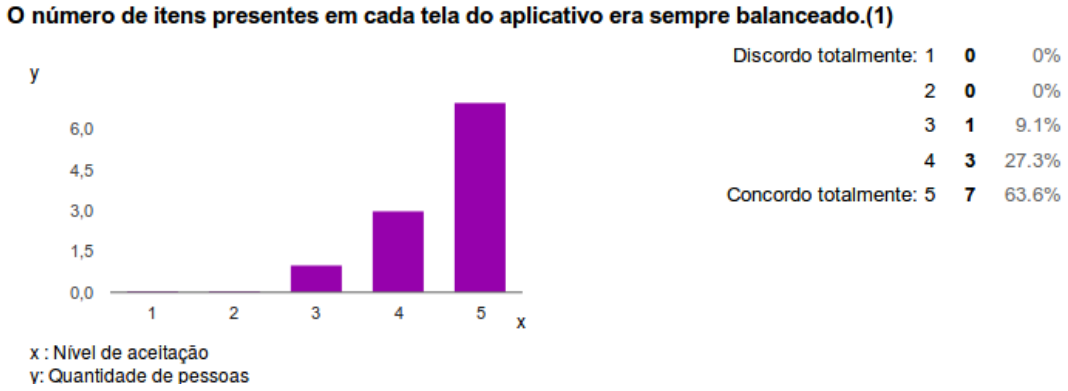
Figura 2 – Avaliação da interferência das dimensões das margens na visualização do conteúdo e operação do aplicativo.



Fonte: Do próprio autor.

De acordo com 91% dos avaliadores, novamente pertencentes as escalas 4 e 5, o número de itens presentes em cada tela do aplicativo está balanceado como mostra o gráfico da Figura 3.

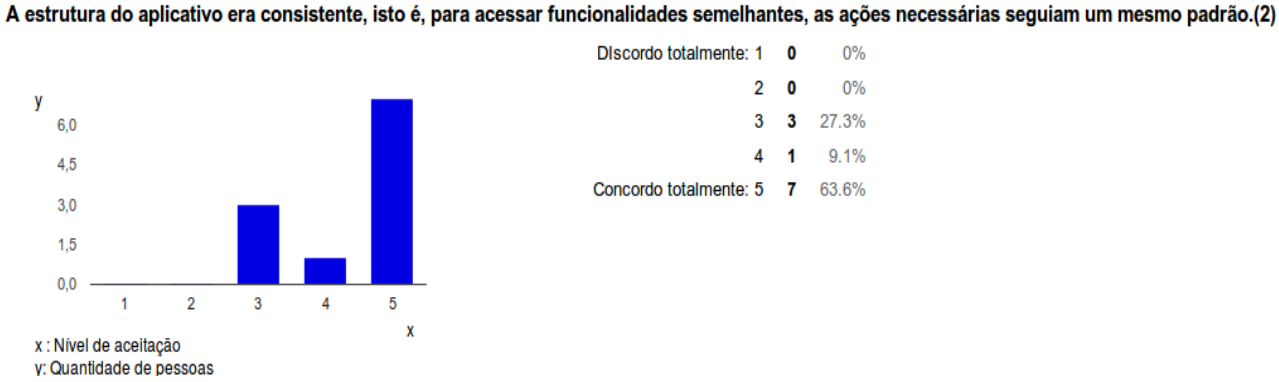
Figura 3 – Avaliação do balanceamento do número de itens por tela do aplicativo.



Fonte: Do próprio autor.

No entanto, quase 30% dos avaliadores, escala 3, consideram que a estrutura do aplicativo no que se refere a sua padronização poderia ser aprimorada (Figura 4). Na Tabela 1 estão listadas cada uma das demais perguntas do questionário baseado em heurística. Cada uma delas foi referenciada por um ID, que foi utilizado para identificar cada questão na Tabela 2.

Figura 4 – Avaliação da consistência do aplicativo para funcionalidades semelhantes.



Fonte: Do próprio autor.

Tabela 1 – ID referente a cada uma das demais perguntas do questionário baseado em heurística.

ID	Pergunta
5	Os ícones utilizados representam adequadamente as respectivas funcionalidades
6	Não houve ambiguidade referente a função de botões e ao que deveria ser preenchido em campos de seleção
7	O espaço disponível na tela do dispositivo foi bem aproveitado pelo aplicativo
8	As funcionalidades do aplicativo estavam igualmente acessíveis

Na Tabela 2, os níveis de aceitação aparecem em uma escala gradual de 1 a 5, em que 1 representa discordar totalmente e 5 representa concordar totalmente. Para cada um dos níveis há a porcentagem de participantes presentes na validação, que aderiram ao referido nível.

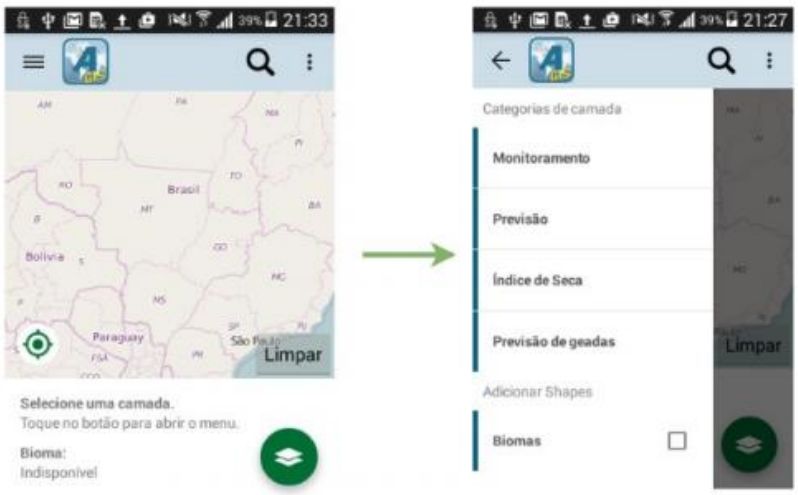
Tabela 2 – Nível de aceitação dos participantes da validação, com relação a cada uma das perguntas.

ID	Nível de aceitação				
	1	2	3	4	5
5	9.1%	0%	18.2%	45.5%	27.3%
6	0%	0%	18.2%	27.3%	54.5%
7	0%	0%	18.2%	54.5%	27.3%
8	0%	0%	0%	36.4%	63.6%

Após a análise dos resultados dos gráficos e da Tabela 2, foram identificados alguns pontos para melhorias, tais como: a falta de um botão de ajuda conceitual para itens, a ausência de um botão para limpar as camadas sobre o mapa, a necessidade de se manter a posição clicada no mapa ao trocar de camada, colocar valor inválido para biomas que estivessem em pontos fora da camada, criação de um botão para a funcionalidade GPS e selecionar o ponto no mapa em que o usuário final se localiza fisicamente ao pressionar este botão, ter um mecanismo de busca para localidades e outro mecanismo que permitisse adicionar as localidades favoritas.

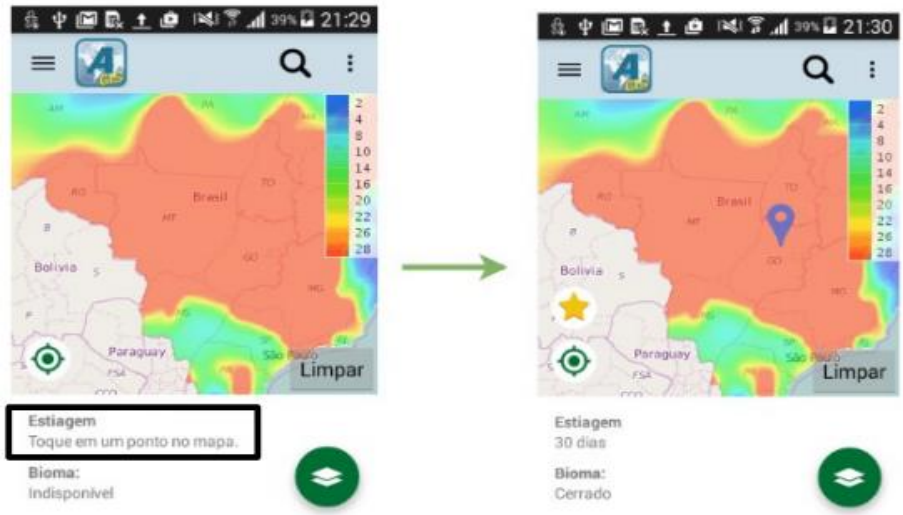
Todas as melhorias foram incorporadas ao aplicativo e uma nova versão foi gerada, como pode ser visualizado nas Figuras 9 e 10, nas quais, respectivamente tem-se como ficou a tela inicial e o menu lateral do aplicativo, e a estiagem, em dias, junto ao bioma, para o estado de Goiás. O Agritempo GIS foi publicado no mês de abril de 2017, na Google Play Store, e já conta com avaliação cinco estrelas (nota máxima), pelos usuários.

Figura 9– Tela inicial do aplicativo e menu lateral.



Fonte: Alves et al. (2016)

Figura 10– Visualização da camada de estiagem junto ao bioma para um ponto do estado de Goiás.



Fonte: Alves et al. (2016)

CONCLUSÕES

Esse trabalho apresentou uma solução móvel baseada em sistema de informação geográfica para apoiar a tomada de decisão no campo. Como smartphones e tablets estão cada vez mais acessíveis, o perfil do usuário torna-se cada vez mais heterogêneo, dificultando a manutenção da usabilidade e da adequação do sistema para usuários distintos. Sendo assim, a inclusão do usuário no processo criativo auxilia na redução da barreira entre a equipe de desenvolvimento e o usuário. Portanto, a etapa de validação colaborativa mostrou a importância de um ambiente favorável à interação entre participantes, resultando em questionamentos essenciais ao projeto e em problemas mapeados que serviram de base para a adequação ao público-alvo e para o *redesign* do aplicativo.

Além disso, como o aplicativo já está publicado na Google Play Store, ele pode ser validado por uma diversidade de usuários muito maior, o que, com certeza, trará grandes melhorias para as futuras versões. Trabalhos futuros incluem as atualizações do aplicativo na Google Play Store, por meio do *feedback* dos usuários nos comentários sobre o aplicativo.

AGRADECIMENTOS

Os agradecimentos são direcionados ao Gabriel Borges, ex-bolsista e colaborador da Embrapa Informática Agropecuária, que fez parte da equipe de desenvolvimento do aplicativo Agritempo GIS, e aos participantes da etapa de validação deste aplicativo.

REFERÊNCIAS

ALBERT, C.; ONWUBUYA, E. ICT APPLICATION IN AGRICULTURAL EXTENSION DELIVERY IN RIVERS STATE: THE PROSPECT. Agricultura, agricultural practice and science journal, v. 87, n. 3-4, 2013.

ALVES, B. J.; ROMANI, L. A. S.; OTAVIAN, A. F. AgritempoGIS: um aplicativo para auxiliar agricultores em processos de tomada de decisão. **In: MOSTRA** DE ESTAGIÁRIOS E BOLSISTAS DA EMBRAPA INFORMÁTICA AGROPECUÁRIA, 12., 2016, Campinas. Resumos expandidos. Brasília, DF: Embrapa, 2016. p. 35-41.

FINUCANE, M. L., ALHAKAMI, A., SLOVIC, P., & JOHHSON, S. M. (2000). The affect heuristic in judgments of risks and benefits. JOURNAL OF BEHAVIORAL DECISION MAKING, 13(1), p. 1-17.

KENNETH S. R. Essential scrum: A practical guide to the most popular agile process, Addison-Wesley, 2012.

KUMAR A., VITHANI T. (2014) Presentation 5. A comprehensive mobile application development and testing lifecycle. **In: PROCEEDINGS** OF IT PROFESSIONAL CONFERENCE (IT PRO), 2014, Gaithersburg, MD, 2014, pp. 1-3.

MAGALHÃES, G. B.; ROMANI, L. A. S. Redesign participativo do aplicativo móvel Agritempo: a importância da interação usuário-desenvolvedor. **In: MOSTRA** DE ESTAGIÁRIOS E BOLSISTAS DA EMBRAPA INFORMÁTICA AGROPECUÁRIA, 10., 2014, Campinas. Resumos. Brasília, DF: Embrapa, 2014. p. 57-60.

METZNER, C. M.; BERTOLINI, G. R. F.; BRAUM, L. M. S.; MARTINI, O. J. Gestão de custos nas propriedades rurais de Toledo com o uso das ferramentas contábeis. **CAP Accounting and Management**, v. 7, n. 7, 2014.

SACHS, I. **Rumo à "agricultura plurifuncional"**. Página 22, n. 76, p. 45, 2014.

SILVEIRA, A. C.; SCHWARTZ, C. TICs e relações afetivo-produtivas na agricultura familiar: enfrentando o isolamento e a exclusão digital. **In:** II CONFERÊNCIA DO DESENVOLVIMENTO. Brasília. 2011.

SUAREZ, S. A.; SUAREZ, A. M. The impact of mobile phone apps. THE AGRICULTURAL PRODUCTION. DAAAM. International Scientific Book, 2013.

ZHANG, Qin; PIERCE, Francis J. (Ed.). **Agricultural automation: fundamentals and practices**. CRC Press, 2013.

Solução móvel de viabilidade econômica para os custos de produção do guaraná e avaliação de usabilidade com a ferramenta *Google TestLab*

Paulo Igor Moraes da Silva[1], *Marcos Filipe Alves Salame*[2]

[1]Bolsista de Iniciação Científica FAPEAM, Embrapa Amazônia Ocidental
Manaus, Amazonas, Brasil, pauloigormoraes@gmail.com
[2]Analista de Tecnologia da Informação, Embrapa Amazônia Ocidental
Manaus, Amazonas, Brasil, marcos.salame@embrapa.br

RESUMO

O guaraná é uma planta originária da Amazônia, no entanto, o seu maior produtor é o Estado da Bahia, tendo margem de área plantada um pouco maior em relação ao Amazonas e apresentando imensa diferença de produtividade. Baseando-se neste cenário, foi desenvolvido um software móvel de cálculos de custos de produção para auxiliar os produtores rurais do Amazonas nas análises, planejamentos e financiamentos. Para atender a um público com pouca experiência no uso de tecnologias da informação, foram utilizadas métricas conhecidas como teoria das ações e o software Google TestLab para validar a usabilidade do aplicativo. Ao final obtivemos o software desenvolvido com várias funcionalidades e a usabilidade aprovada pelo TestLab.

PALAVRAS-CHAVE: Aplicativo, Android, Amazonas

ABSTRACT

Guarana is a plant native to the Amazon, however, its largest producer is the State of Bahia, with a margin of area planted only a little higher in relation to the Amazon, however it presents an immense productivity difference. Based on this scenario, a mobile software of production cost calculations was developed to assist farmers in Amazonas in the analysis, planning and financing. In order to serve a public with little experience in the use of information technologies, metrics known as stock theory and Google TestLab software were used to validate the usability of the application. At the end we obtained the software developed with several functionalities and the usability approved by TestLab.

KEYWORDS: Software, Android, Amazon.

INTRODUÇÃO

O guaraná é uma planta originária da Amazônia e apresenta potencial econômico e social para o Brasil (PINTO; FAJARDO; ATROCH, 2015). O Brasil é um dos únicos produtores de guaraná

do planeta para fins comerciais, juntamente com algumas pequenas regiões da Amazônia venezuelana e peruana. No Brasil, a produção concentrou-se durante muito tempo no Estado do Amazonas e posteriormente houve expansão para os Estados do Pará, Acre, Rondônia, Mato Grosso e Bahia, que também dinamizaram suas plantações comerciais (FURLAN et al., 2003).

O Amazonas já há muito tempo deixou de ser o maior produtor nacional, tendo sido ultrapassado pela Bahia nos quesitos produção e produtividade e pelo Mato Grosso em produtividade (FURLAN et al., 2003). De acordo com os dados do Instituto Brasileiro de Geografia e Estatística (IBGE, 2013), o Amazonas tem uma área destinada à colheita de 6.666 hectares, sendo o segundo Estado com maior área do país, atrás apenas da Bahia, que tem uma área de 6.669 hectares. No entanto, quando se observa a quantidade produzida, obtemos 664 toneladas para o Amazonas e 2.672 toneladas para a Bahia, o que possibilita concluir que no Amazonas há uma baixa produtividade.

A agricultura moderna enfrenta enormes desafios. Atualmente, o setor agrícola cresceu para uma indústria altamente competitiva e globalizada, na qual os agricultores e outros atores devem considerar aspectos climáticos e geográficos locais, bem como fatores ecológicos e políticos, a fim de garantir a sobrevivência econômica e produção sustentável (DENGEL, 2013).

No Amazonas, um fator que dificulta o crescimento do agronegócio, diz respeito ao número de grandes agricultores. Dos 66.784 estabelecimentos rurais do estado, 93% atendem o critério legal de agricultura familiar, o que corresponde a 61.843 propriedades rurais (IBGE, 2006).

O Amazonas apresenta uma vasta extensão territorial e o transporte representa um dos principais problemas para os diversos agricultores do interior do Estado, dificultando deslocamentos e acesso às informações. Somando a esse cenário, a partir de diálogos com pesquisadores da área, foi observado que muitos produtores não mensuram seus lucros líquidos a partir da produção de determinada cultura, culminando em planejamentos e financiamentos mal feitos, prejuízos e endividamentos.

Em muitos ambientes agrícolas no Amazonas, normalmente o dono das lavouras assume papéis como: produtor e administrador, atuando como empreendedor e capitalista e, grande parte desses agricultores não possuem conhecimentos econômicos suficientes para gerir seus negócios, muitas vezes resultando em instabilidades financeiras (GUIDICCI et al., 2015), podendo ocasionar em uma atuação negativa no mercado produtivo.

Diante do exposto, torna-se de grande importância a utilização e desenvolvimento de novas tecnologias para auxiliar de forma estratégica a difusão de conhecimentos e ações para apropriação de novos procedimentos e processos de produção, influenciando diretamente na capacidade produtiva e competitividade dos produtos agrícolas do Estado.

O uso da Computação Aplicada à agricultura pode trazer enormes benefícios para o país e avanços para a ciência e sociedade, possibilitando tomadas de decisão estratégicas, difusão de conhecimentos e auxílios para transferências de tecnologia. Através de um software, por exemplo, é possível ajudar a reduzir planejamentos mal feitos e possíveis endividamentos, auxiliar

produtores rurais e técnicos no processo de tomada de decisão e consequentemente, aumentar a produtividade da cultura em uma região.

Atualmente no mercado, há uma parcela considerável de softwares destinados a controles de custos do produtor. No entanto, foi observado que alguns disponibilizados gratuitamente são difíceis de usar e encontram-se incompletos ou com falhas e os softwares mais completos e robustos apresentam plataforma de uso na Web e/ou *desktop* e são comerciais, sendo inviáveis para uso nos interiores do Amazonas, devido a muitos locais não possuírem conectividade com a Internet e da ausência de computadores nas residências, sem mencionar os custos.

Alguns dos principais são: a ferramenta *e-Agro Web*, que é um sistema destinado a gestão e controle de produção agrícola (EAGRO, 2015). Expõe, de maneira simples e eficaz os dados utilizados pelo agricultor. Por ser um sistema de grande porte, apresenta um cenário mais amplo e com inúmeras funcionalidades, facilidade de uso e flexibilidade para consultar informações descritivas relacionadas aos custos produtivos anuais, mensais e diários, no entanto, esse sistema é comercial e para utilização é necessário um investimento estimado com base no espaço rentável produtivo, em hectares. Outra ferramenta é o aplicativo **Custo Fácil** desenvolvido para análises dos custos de produção e assistência técnica para gestão de granja e funciona em dispositivos móveis (MIELE; BENELLI; SANDI, 2016).

Com base nesse cenário, foi desenvolvido um software para dispositivos móveis que valide os coeficientes técnicos e que ajude produtores rurais e técnicos especialistas nas análises financeiras relacionadas aos custos de produção[1] e lucratividade de acordo com os preços e quantidades de insumos e dos produtos para auxiliar a expansão da cultura do guaraná no Estado do Amazonas.

Além do desenvolvimento, foi utilizado um software do Google destinado a testes para obter informações sobre as interfaces e a usabilidade do aplicativo. No desenvolvimento, buscamos sempre a simplicidade na hora de criar as telas e as funcionalidades pensando de modo integral na experiência do usuário, por se tratar de um público específico sem muitas habilidades no uso de tecnologias.

MATERIAL E MÉTODOS

Foram realizadas entrevistas com especialistas nas áreas de agricultura e economia com intuito de identificar, investigar e filtrar as informações utilizadas no processo de modelagem. Posteriormente, foi executada a fase de levantamento de conhecimentos técnicos computacionais, bibliográficos e documentais.

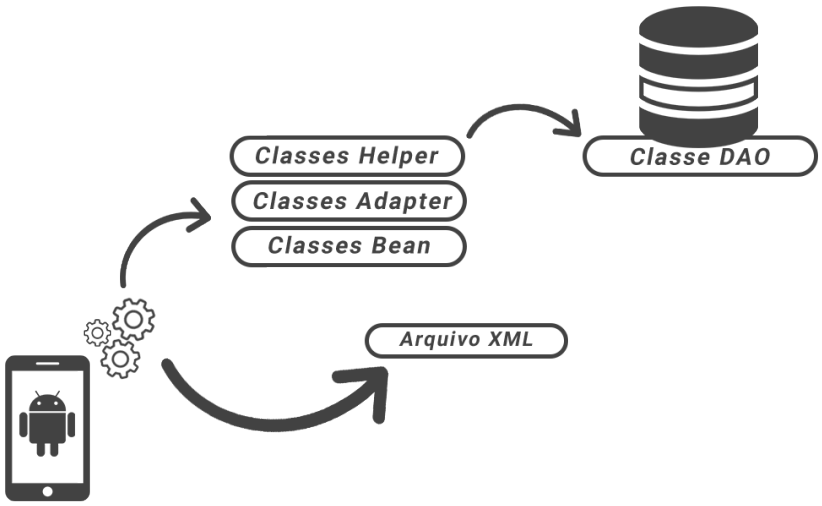
Serviram de base para construção do aplicativo ferramentas como: *Google Android Studio* para escrita, compilação, empacotamento e construção do arquivo executável, a linguagem de programação usada foi o *Java JDK (Kit de Desenvolvimento Java)* e *Android SDK Tools (Kit*

[1]Entende-se por **custo de produção** a superposição de todos os capitais de um processo produtivo, como por exemplo insumos e tarefas realizadas de forma financeira, tendo como objetivo obter quantidade máxima de produto com o mínimo de custo e desperdício, (GUIDICCI et al., 2015).

de Desenvolvimento de Software), além do ambiente *DB Browser* para leitura e interpretação dos dados, o banco de dados foi implementado usando *SQLite*, possibilitando o funcionamento do aplicativo off-line. Algumas bibliotecas como: *PieCharts* e *JXL* foram adicionadas para possibilitar a construção dos gráficos e gerar os arquivos de exportação para o formato XLS *(excel)*, respectivamente.

Durante a etapa de implementação do código foram utilizados conceitos computacionais sobre o paradigma de *programação orientada a objetos* - **(P.O.O)**, facilitando a interpretação e agilidade no processo. Também foram aplicadas algumas definições sobre arquitetura de software a fim de manter a estrutura e organização nos arquivos de classes seguindo padrões da própria IDE *(Ambiente de Desenvolvimento Integrado) Google Android Studio*. A arquitetura e organização do código está ilustrado na Figura 1.

Figura 1: Organização das classes do aplicativo.



A classe DAO é onde estão localizados os métodos do banco de dados bem como as tabelas implementadas. As classes *Helper, Adapter e Bean* são utilizadas para conexão com o banco e também para manipulação e construção dos dados na tela do aplicativo para o usuário. Os arquivos *XML* são onde estão desenhados os modelos visuais, suas telas de listagem, cadastros e telas de análises.

De forma a alcançar o público alvo desejado, foi pensado na usabilidade do aplicativo. Aqui a questão relevante foi aplicar técnicas de interação e usabilidade que impactassem positivamente nos produtores rurais. Para isso, se fez necessário utilizar modelos propostos baseados em engenharia cognitiva[2] para modelagem e implementação do seguimento visual do aplicativo.

O processo de construção da interface foi baseado na **Teoria das Ações**, proposto por Donald Norman em 1986, onde o autor descreve a interação entre dois ciclos de ações, que

[2]*Engenharia Cognitiva* consiste no estudo dos fatores humanos afim de compreender o processo pelo qual o mesmo constrói conhecimento; é a experiência que a mente humana pode ter e induzir sobre o uso de sistemas interativos e de fácil manipulação, (SILVA; BARBOSA, 2010).

é composta por dois extremos, o primeiro sendo o *Golfo da Execução* que está ligado com o esforço mental exercido pelo usuário ao deparar-se com os comandos e funções apresentados pelo sistema; outro, denomina-se *Golfo de Avaliação*, onde o usuário coloca em prática suas ações no sistema (NORMAN; DRAPER, 1986).

Para atravessar os ”golfos”, são necessários alguns passos, com eles precisam ser definidos informações como o objetivo, que é definido como a tarefa pela qual o usuário está interagindo com o sistema (NORMAN; DRAPER, 1986). A partir desse ponto, outras etapas precisam ser desempenhadas, como: a) *intenção*; b) *especificação da ação*, c) *execução*; d) *percepção*; e) *interpretação*; e f) *avaliação*. Conforme, pode ser observado na Figura 2.

Figura 2: Travessia dos golfos. Adaptado de (SILVA; BARBOSA, 2010).



Essas etapas foram aplicadas em todas as funcionalidades do aplicativo com o objetivo de prover uma máxima experiência para o usuário. Também foi seguido o modelo de *design* para dispositivos móveis conhecido como *Material Design*. Esse paradigma é disponibilizado e mantido pelo *Google Design* e respeita todas as etapas descritas na teoria das ações, além de transmitir facilidade, interatividade e conforto para os usuários do sistema (GOOGLE, 2017).

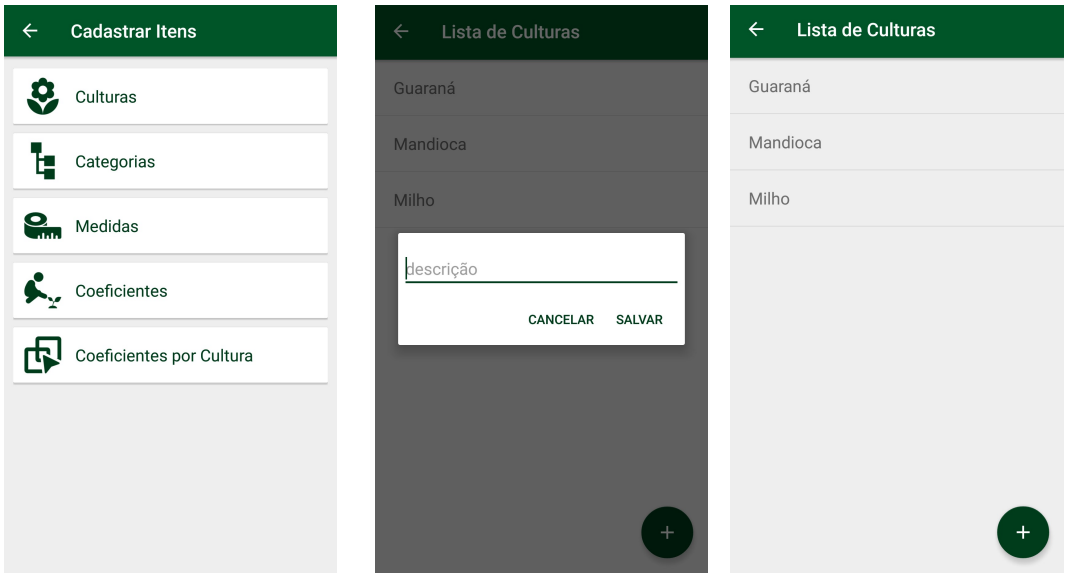
Esse modelo foi construído especificamente para sistemas móveis que usassem o *Google Android S.O*, todavia não se restringiu a esse ambiente, atualmente também é encontrado em sistemas web e sistemas operacionais para computadores, o *Material Design* é embasado em conceitos para identificação de relevos geométricos, também possui compreensão de campos tridimensionais, efeitos de sombreamentos, transições e movimentações das propriedades físicas sem ferir estruturalmente a arquitetura do design (GOOGLE, 2017).

RESULTADOS E DISCUSSÃO

O aplicativo foi desenvolvido para ajudar produtores rurais e técnicos especialistas nas análises financeiras relacionadas aos custos de produção. A Figura 3a ilustra a tela de opções para

cadastro. Essas opções servem para customizar o aplicativo com mais informações, além das cadastradas durante a instalação. A Figura 3b exibe a tela para adicionar uma nova cultura, enquanto que a Figura 3c apresenta a tela com a listagem das culturas cadastradas.

Figura 3: Telas da primeira parte funcional do aplicativo.



(a) Cadastro de item (b) Adicionar cultura (c) Lista de cultura

Caso o usuário deseje usar os coeficientes técnicos padrão de cada cultura, elementos necessários para a produção, que já se encontram no aplicativo, ele pode ir diretamente para o registro dos custos de produção. A Figura 4a apresenta a tela com os campos necessários para criar um registro dos custos. Após isso, surge a tela com os coeficientes técnicos relacionados à cultura escolhida, para serem preenchidos os campos quantidade e valor, conforme apresentados na Figura 4b.

Na Figura 5a, são apresentados todos os registros criados. Ao selecionar qualquer item desta lista, o usuário terá acesso a tela com todos os coeficientes e seus respectivos valores e quantidades, bem como o valor total do custo produtivo da cultura em questão. Os custos podem ser visualizados por categorias, conforme a Figura 5b.

A Figura 5c apresenta um gráfico que ilustra os valores de cada categoria da cultura analisada e, individualmente o valor de cada coeficiente referente à categoria.

Foram realizados testes de usabilidade usando ferramentas oferecidas pela *Google Technology Company*. O processo realizado é denominado *TestLab* e está disponível dentro da plataforma conhecida como *Firebase Console*[3]. O teste é focado em usabilidade em níveis e funciona da seguinte forma: os desenvolvedores geram o arquivo **.apk**, fazem o *upload* para a plataforma, configuram o tempo de execução, que seleciona o nível de navegação, de acordo

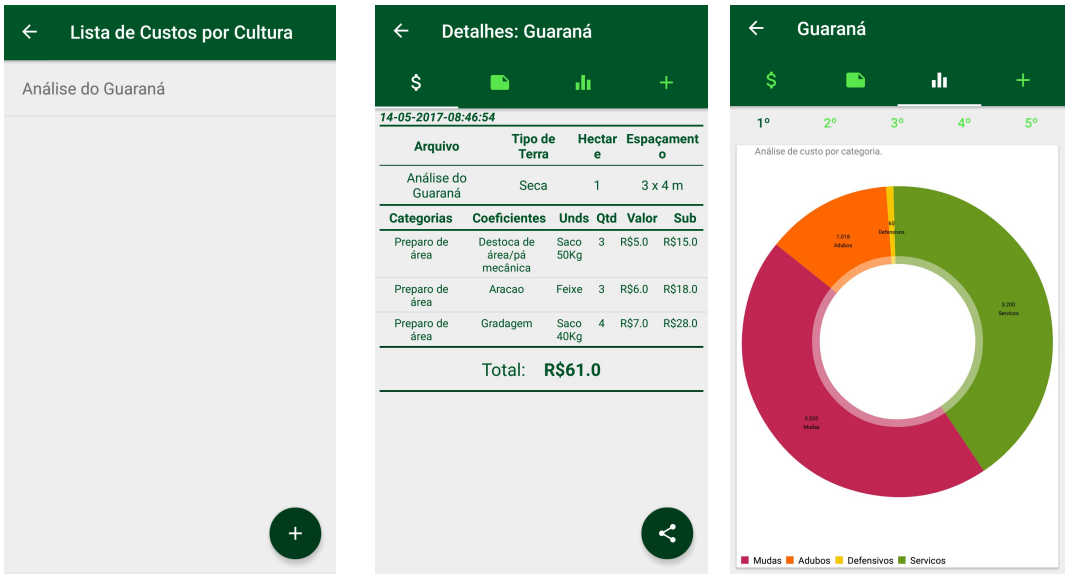
[3]Firebase Console é uma ferramenta do *Google* para desenvolvimento aplicativos, onde os desenvolvedores possam monetizar seus aplicativos de forma fácil, simples, eficaz e sem ferir a experiência de seus usuários, (GOOGLE, 2017).

Figura 4: Telas para cadastro do custo por cultura.



(a) Inserir dados (b) Adicionar dados

Figura 5: Telas da primeira parte funcional do aplicativo.



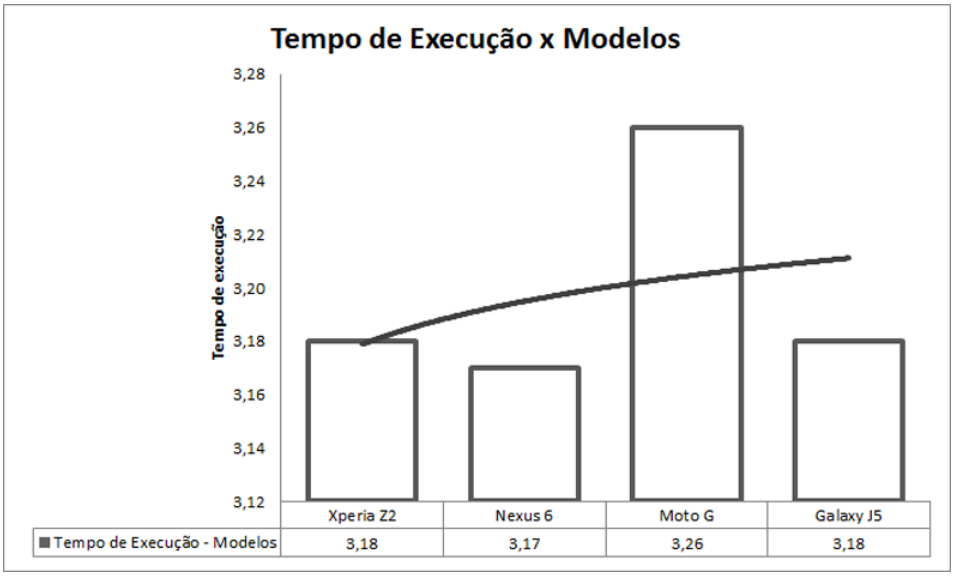
(a) Lista Custo por Cultura (b) Aba de análise (c) Aba de gráfico

com o tempo informado, selecionam os aparelhos físicos e/ou máquinas virtuais com suas respectivas **APIs** - *(Interface de Programação de Aplicativos)* e, por fim, iniciam-se os testes.

O teste é realizado por um robô construído com objetivo de navegar entre as telas analisando contexto de usabilidade. Ele é divido em dois níveis, o primeiro relaciona as configurações dos aparelhos e/ou máquinas virtuais selecionados como: nível de *API*, que verifica se está de acordo com o *API* do aplicativo exposto ao teste, orientação e linguagem do aplicativo e tempo de execução. Este nível é denominado, dimensão do teste. Os dispositivos e configurações eleitos para avaliação foram:

1. Aparelhos físicos (modelos):
    - Xperia Z2, *API level 21*;
    - Nexus 6, *API level 21*;
    - Moto G (3rd Gen), *API level 22*; e
    - Galaxy J5, *API level 23*.
2. Orientação:
    - Retrato; e
3. Linguagem *(localização)*:
    - Brasil, *pt_BR*;
4. Tempo de execução:
    - 4 minutos de teste, com 40 níveis de profundidade de navegação.

Figura 6: Resultado em Tempo de Avaliação.



O segundo destina-se a entrelaçar combinações com a quantidade de telas desenvolvidas, imitando intenções dos usuários, baseando-se no elementos disponíveis nas telas e respeitando o tempo e níveis de navegações selecionada pelo usuário.

Cada modelo teve **4 minutos** e **40 níveis de profundidade** para teste, padrão estipulado pelo *Google*, dado o nível de complexidade e tamanho do aplicativo. Em todos os modelos testados o sistema obteve êxito e ficou dentro do limite de tempo de teste esperado, classificando o aplicativo positivamente em usabilidade.

Na Figura 6, pode ser observado o correlacionamento de cada modelo com seu tempo de execução e a linha média de **4 minutos**, tendo uma variação temporal maior no modelo *Moto G*, mas ainda assim, mantendo uma média similar entre todos os dispositivos, não obtendo nenhuma falha ou grande disparidade e conseguindo realizar todos os testes antes dos 4 minutos.

CONCLUSÕES

Ao longo do trabalho foi apresentado o desenvolvimento de um software com o objetivo de auxiliar os produtores rurais e técnicos especialistas nas análises, planejamentos e financiamentos com relação aos custos de produção do guaraná. Por se tratar de produtores rurais que em sua maioria não apresentam muita experiência com tecnologias da informação, foi dada ênfase nas interfaces, telas e funcionalidades para ter como resultado um aplicativo de fácil utilização.

Foram realizados testes com a ferramenta Google TestLab para verificar a usabilidade do aplicativo, o qual obteve um resultado final positivo, tendo sua estrutura e usabilidade aprovadas.

Avaliadores econômicos foram utilizados no desenvolvimento para compôr os cálculos realizados no aplicativo, os dados manipulados correspondem a estudos realizados por economistas para otimização dos custos produtivos agrícolas (ALMUDI; PINHEIRO, 2013). O aplicativo, por sua vez, mensura possíveis gastos, apresentando ao produtor uma análise de viabilidade econômica, essa análise é gerada a partir de uma relação da quantidade pelo valor de um coeficiente, e está diretamente ligada a uma categoria, que por sua vez está vinculada a uma cultura (FURLAN et al., 2003). Esses itens são adicionados pelo agricultor e o aplicativo se encarrega de realizar a análise e apresentar os resultados.

Como sugestão de trabalho futuro, pode-se fazer uma validação diretamente com o público alvo, utilizando outras metodologias já validadas, para obter mais informações sobre a facilidade, utilidade e eficiência do aplicativo, objetivando aperfeiçoá-lo continuamente.

Espera-se que o aplicativo ajude os produtores rurais e técnicos e que reduza o endividamento dos produtores, que estarão mais conscientes do quanto precisam gastar para poder produzir, e assim, consequentemente ajudar a expansão do guaraná no Amazonas.

AGRADECIMENTOS

O desenvolvimento deste trabalho somente foi possível com o auxílio e colaboração da Fundação de Amparo à Pesquisa do Estado do Amazonas pelo fomento da bolsa e do pesquisador Jose Olenilson Costa Pinheiro, pelo seu empenho, dedicação e transmissão de conhecimentos.

REFERÊNCIAS

ALMUDI, T.; PINHEIRO, J. O. C. *Dados Estatísticos da Produção Agropecuária e Florestal do Estado do Amazonas*. 1ª. ed. [S.l.]: Embrapa Amazônia Ocidental, 2013.

DENGEL, A. Special issue on artificial intelligence in agriculture. 2013.

EAGRO, W. *Sistema para Gestão Agrícola*. 2015. `http://eagroweb.com/`. Acesso em 28 abr. 2017.

FURLAN, L. F. et al. *Projeto potencialidades regionais estudo de viabilidade econômica: Guaraná*. [S.l.]: Superintendência da Zona Franca de Manaus, 2003.

GOOGLE, T. c. *Google Development to Android*. 2017. `https://developer.android.com/index.html`. Acesso em 28 abr. 2017.

GUIDICCI, R. d. C. N. et al. Aspectos metodológicos da análise de viabilidade econômica de sistemas de produção. Empresa Brasileira de Pesquisa Agropecuária, 2015.

IBGE, I. B. d. G. e. E. *Censo Agropecuário*. 2006. 77 p. `http://biblioteca.ibge.gov.br/visualizacao/periodicos/51/agro_2006.pdf`. Acesso em: 28 abr. 2017.

IBGE, I. B. d. G. e. E. *Produção Agrícola Municipal: Culturas temporárias e permanentes*. 2013. 77 p. `http://biblioteca.ibge.gov.br/visualizacao/periodicos/66/pam_2013_v40_br.pdf`. Acesso em: 28 abr. 2017.

MIELE, M.; BENELLI, D. A.; SANDI, A. J. *Aplicativo Custo Fácil: Suínos e Frangos de Corte*. 1ª. ed. [S.l.]: Embrapa Amazônia Ocidental, 2016.

NORMAN, D. A.; DRAPER, S. W. *Usercentered System Design: New Perspectives on Human-Computer Interaction*. 1ª. ed. [S.l.: s.n.], 1986.

PINTO, C. E. D. L.; FAJARDO, J. D. V.; ATROCH, A. L. Avaliação da adaptabilidade e estabilidade de clones de guaraná no estado do amazonas. 2015.

SILVA, B. S. d.; BARBOSA, S. D. J. Interação humano-computador: Projetando a experiência perfeita. 2010.

Seleção de atributos agressiva para estimativa de teor de nutrientes em soja utilizando espectroscopia NIR

Luiz Rodolfo Machado[1], *Antônio David Viniski*[2], *Pablo Henrique Ferreira*[2], *Alaine Margarete Guimarães*[3], *Eduardo Fávero Caires*[4]

[1] Curso de Engenharia de Computação – Bolsista do Programa de Iniciação Científica, Universidade Estadual de Ponta Grossa, Ponta Grossa, Paraná, Brasil, luizrodolfo.m@gmail.com

[2] Programa de Pós-Graduação em Computação Aplicada, Universidade Estadual de Ponta Grossa, Ponta Grossa, Paraná, Brasil, davidviniski@gmail.com, pablo.henrikef@gmail.com

[3] Departamento de Informática, Universidade Estadual de Ponta Grossa, Ponta Grossa, Paraná, Brasil, alainemg@uepg.br

[4]Departamento de Ciência do Solo e Engenharia Agrícola, Universidade Estadual de Ponta Grossa, Ponta Grossa, Paraná, Brasil, efcaires@uepg.br

RESUMO

A análise de dados de espectroscopia vem sendo utilizada como alternativa na estimativa do teor de nutrientes para a agricultura, pois oferece vantagens como a minimização da quantidade de amostras, tempo e resíduo gerado. Um problema da utilização de espectroscopia é a alta dimensionalidade da base de dados, o qual pode ser solucionado por meio da seleção de atributos. O objetivo do trabalho foi avaliar o desempenho da seleção agressiva de atributos na estimativa do teor de nutrientes em grãos de soja, a partir de dados de espectroscopia. A base de dados foi elaborada a partir de 285 amostras de grãos de soja com os respectivos teores de nutrientes. As amostras foram analisadas em um espectrofotômetro na região do infravermelho próximo (NIR). A seleção de atributos foi feita pelo método convencional disponível no software SAS Enterprise Miner e também adotando um critério de seleção que resultou na redução agressiva da dimensionalidade. A mineração de dados foi feita por meio da rede neural artificial "AutoNeural". A comparação dos modelos foi feita através do erro RMSE normalizado em relação à amplitude. Os modelos gerados obtiveram RMSE menor do que 30% da amplitude para todos os nutrientes da base de dados.

Os elementos N, P, Ca, Mg, Fe, B e a Produtividade se beneficiaram da redução agressiva da dimensionalidade. Os resultados obtidos indicam que a seleção de atributos agressiva pode trazer benefícios para a estimativa do teor de nutrientes por meio da espectroscopia NIR.
PALAVRAS-CHAVE: Redução da dimensionalidade, Refletância espectral, Redes Neurais Artificiais.

ABSTRACT

Spectroscopy data analysis has been used in agriculture as an alternative in the estimation of nutrient content, minimizing the amount of samples, time and residue generated. A problem related to the use of spectroscopy is the high dimensionality of the database, which can be solved through feature selection. The goal of this paper was to evaluate the performance of the aggressive feature selection in the estimation of nutrient content in soybean grains, based on spectroscopy data. The data base was obtained from 285 samples of soybean grains with the respective nutrient contents. The samples were analyzed in a near infrared spectrophotometer (NIR). The feature selection was done by the conventional method available in the software SAS Enterprise Miner and adopting a selection criterion that resulted in dimensionality aggressive reduction. Data mining was done through the "AutoNeural" artificial neural network. The comparison of the models was done through the normalized RMSE error related to the amplitude. The generated models presented RMSE less than 30% of the amplitude for all the nutrients of the database. The N, P, Ca, Mg, Fe, B nutrients and the yield were benefited by the use of dimensionality aggressive reduction. The results indicate that aggressive feature selection can bring benefits to the estimation of nutrient concentration through NIR spectroscopy.
KEYWORDS: Dimensionality reduction, Spectral reflectance, Artificial Neural Network.

INTRODUÇÃO

A análise de dados de espectroscopia vem sendo utilizada como alternativa na estimativa do teor de nutrientes em diversas áreas, inclusive na agricultura, sendo alvo de pesquisas em folhas, grãos e até mesmo no solo. Essa técnica oferece vantagens em relação às metodologias correntes de análise, sendo a principal delas, a possibilidade de analisar amostras sólidas e líquidas sem nenhum tratamento prévio (BALABIN; SMIRNOV, 2011). Além disso, necessita de pequeno volume de amostra e minimiza drasticamente a quantidade de resíduos gerados, preservando as amostras utilizadas (BEN-DOR; BANIN, 1995).

A espectroscopia também permite a análise de um número maior de amostras com custo e tempo reduzidos. Um fator negativo na utilização dessa técnica é a alta dimensionalidade, pois o número de comprimentos de onda dentro do espectro analisado fica geralmente entre 200 e 2000, sendo que cada um deles se torna um atributo na base de dados gerada.

A mineração de dados consiste na descoberta automática de informação útil em grandes depósitos de dados, sendo uma parte integral do processo de Descoberta de Conhecimento em Bases de Dados (KDD – do inglês, *Knowledge Discovery in Databases*). Uma das etapas desse processo é o pré-processamento, o qual inclui a redução de dimensionalidade, realizado para transformar os dados em um formato apropriado para a mineração (TAN; STEINBACH; KUMAR, 2006).

A redução da dimensionalidade em bases de dados, por meio da seleção de atributos, tem o objetivo de selecionar apenas os atributos mais relevantes para a descrição do objeto de estudo, eliminando aqueles que poderiam interferir negativamente no resultado (TAN; WANG; TSANG, 2009). Essa técnica antecede a mineração de dados, melhorando o desempenho e diminuindo a complexidade dos modelos (PIRAMUTHU, 2004). Com isso, a seleção de atributos torna-se uma etapa crítica na análise de dados de espectroscopia (BALABIN; SMIRNOV, 2011), pois a redução do número de atributos sem perda de capacidade de predição pode simplificar o processo de estimativa dos teores de nutrientes.

De acordo com Liu e Motoda (2007), uma estratégia de redução da dimensionalidade é a seleção de atributos agressiva, na qual a predição é realizada por pouquíssimos atributos (mais de 90% dos atributos são descartados) com o mínimo de perda de performance e o máximo da redução da complexidade. Se for possível realizar a redução agressiva da dimensionalidade em dados espectrais mantendo resultados similares, a estimativa de teor de nutrientes se torna mais simples utilizando entre um e cinco comprimentos de onda ao invés de espectros inteiros ou seleções mais abrangentes.

O objetivo deste trabalho foi avaliar o desempenho da seleção agressiva de atributos na estimativa do teor de nutrientes em grãos de soja, utilizando mineração de dados de espectroscopia na região do infravermelho próximo (NIR).

MATERIAL E MÉTODOS

Essa seção está dividida em três etapas, que representam a coleta dos dados de nutrientes em grãos de soja, a leitura dos valores de refletância espectral e geração da base de dados, e por fim o pré-processamento e mineração de dados.

Coleta de Dados

Foram utilizadas neste estudo, 285 amostras de grãos de soja, coletadas de seis fazendas localizadas na região dos Campos Gerais, no estado do Paraná, durante o ano agrícola de 2013-2014. As amostras foram fornecidas pelo Laboratório de Fertilidade de Solo da UEPG (Universidade Estadual de Ponta Grossa) em conjunto com os resultados de análises laboratoriais de determinação dos teores dos nutrientes N, P, K, Ca, Mg, S, Cu, Mn, Fe, Zn e B. Como informação complementar, o valor da produtividade também foi observado.

Poltronieri (2015) utilizou essa mesma base de dados em seu trabalho, onde cita o método de obtenção de cada nutriente:

> O N foi extraído por meio de digestão sulfúrica e determinado pelo método semi-micro-Kjeldahl. As análises de P, K, Ca, Mg, S, Cu, Fe, Mn e Zn foram realizadas por meio de digestão nítrico-perclórica e determinação por espectrofotometria de absorção molecular para P, fotometria de chama para K, turbidimetria do sulfato de bário para S e espectrofotometria de absorção atômica para Ca, Mg, Cu, Fe, Mn e Zn. O B foi extraído por incineração e determinado por espectrofotometria de absorção molecular, pelo método da azometina H.

Espectroscopia NIR

A obtenção dos dados de espectroscopia foi feita através do espectrofotômetro NIRS DA1650 da marca FOSS (Figura 1), no qual foram feitas as leituras de absorbância espectral de todas as amostras de grãos de soja. O volume amostral utilizado foi de 40ml.

Figura 1 – Espectrofotômetro FOSS NIRS DA1650 utilizado para a obtenção dos dados espectrais.



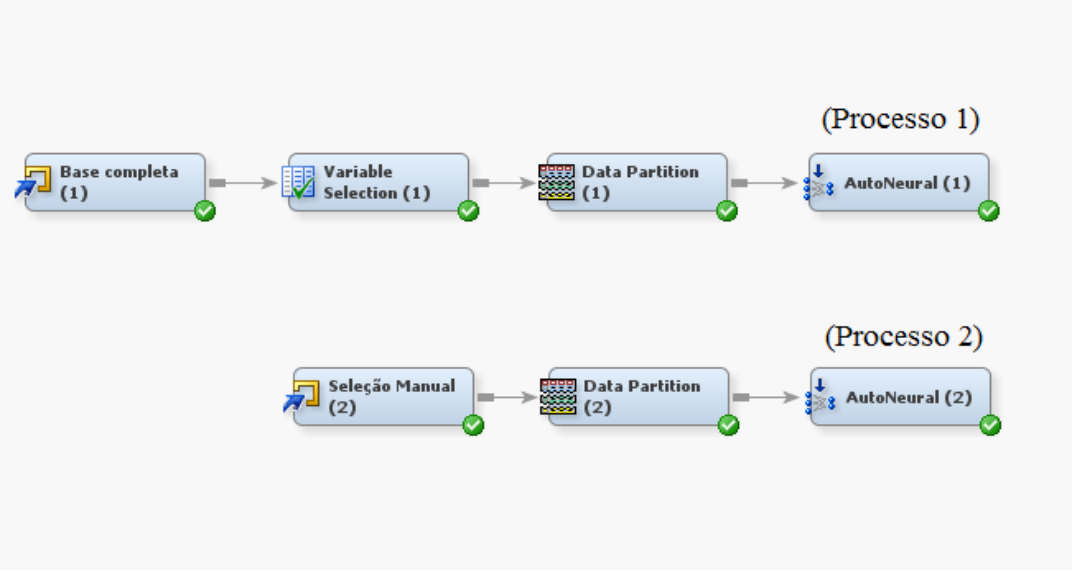
Fonte: (O autor)

As leituras de comprimentos de onda das amostras foram realizadas na região do infravermelho próximo, no intervalo de 1100nm a 1648nm, de dois em dois nanômetros, totalizando 274 atributos. Os valores de absorbância foram convertidos para refletância espectral para cada um dos atributos, e a base de dados gerada foi composta por esses valores em conjunto com os teores de nutrientes de cada amostra.

Pré-processamento e mineração de dados

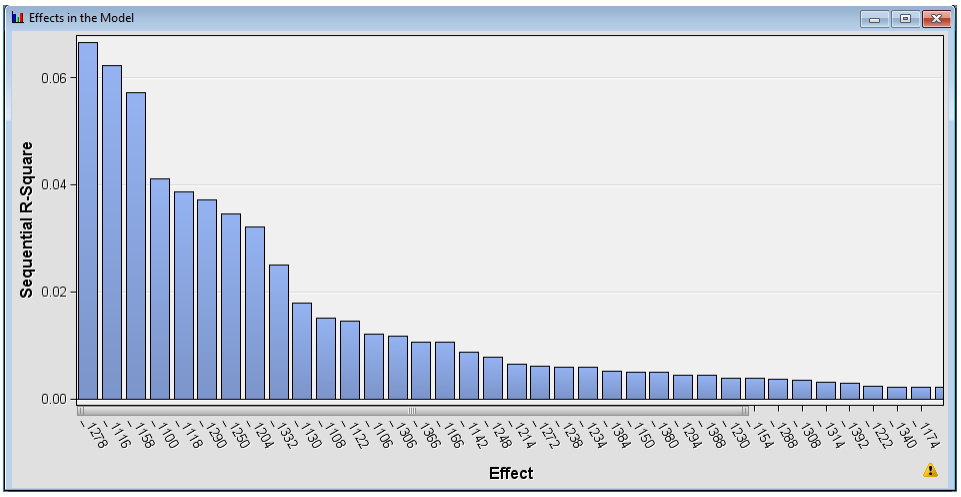
As etapas de pré-processamento e mineração de dados foram realizadas utilizando o software SAS On Demand For Academics: Enterprise Miner, que é uma versão acadêmica gratuita do módulo de mineração do SAS (Statistical Analysis System). Esse ambiente utiliza Diagramas de Fluxo de Processos, onde as etapas de importação da base de dados, seleção de atributos, particionamento e geração do modelo de predição representam os nós do diagrama. A Figura 2 apresenta o Diagrama de Fluxo de Processos definido para esse trabalho.

Figura 2 – Diagrama de Fluxo de Processos da mineração de dados.



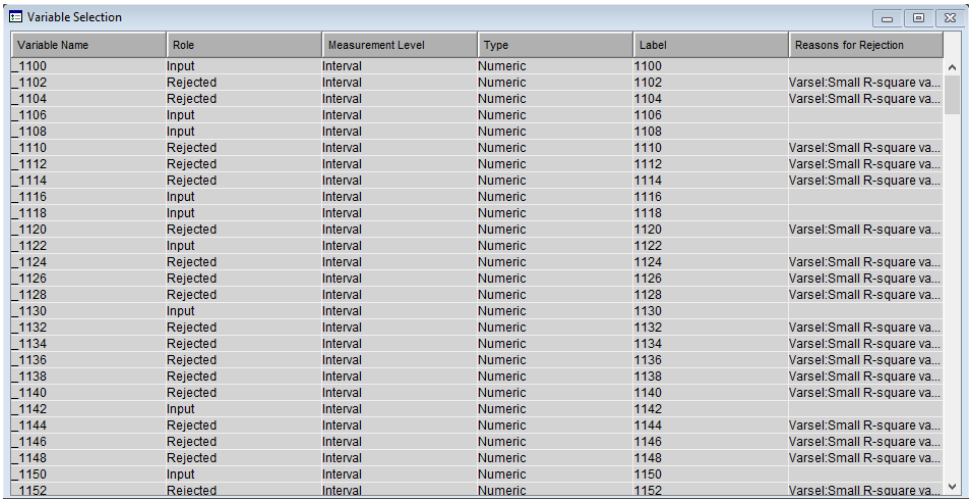
Os processos representados no diagrama diferem entre si quanto à seleção de atributos utilizada. No primeiro caso (Processo 1) foi realizada a seleção de atributos padrão do software SAS, através do nó “Variable Selection”. A saída desse nó apresenta diversas janelas com resultados, sendo que os mais relevantes para esse trabalho são o gráfico de efeitos no modelo e a tabela de atributos selecionados. A Figura 3 representa o gráfico de efeitos de cada um dos atributos no modelo, em ordem decrescente do R^2 sequencial.

Figura 3 – Gráfico de efeitos no modelo gerado no software SAS Enterprise Miner.



A Figura 4 representa uma parte da a tabela de atributos selecionados para serem usados no modelo.

Figura 4 – Tabela de atributos selecionados no gerado no software SAS Enterprise Miner.



Variable Selection

Variable Name	Role	Measurement Level	Type	Label	Reasons for Rejection
_1100	Input	Interval	Numeric	1100	
_1102	Rejected	Interval	Numeric	1102	Varsel:Small R-square va...
_1104	Rejected	Interval	Numeric	1104	Varsel:Small R-square va...
_1106	Input	Interval	Numeric	1106	
_1108	Input	Interval	Numeric	1108	
_1110	Rejected	Interval	Numeric	1110	Varsel:Small R-square va...
_1112	Rejected	Interval	Numeric	1112	Varsel:Small R-square va...
_1114	Rejected	Interval	Numeric	1114	Varsel:Small R-square va...
_1116	Input	Interval	Numeric	1116	
_1118	Input	Interval	Numeric	1118	
_1120	Rejected	Interval	Numeric	1120	Varsel:Small R-square va...
_1122	Input	Interval	Numeric	1122	
_1124	Rejected	Interval	Numeric	1124	Varsel:Small R-square va...
_1126	Rejected	Interval	Numeric	1126	Varsel:Small R-square va...
_1128	Rejected	Interval	Numeric	1128	Varsel:Small R-square va...
_1130	Input	Interval	Numeric	1130	
_1132	Rejected	Interval	Numeric	1132	Varsel:Small R-square va...
_1134	Rejected	Interval	Numeric	1134	Varsel:Small R-square va...
_1136	Rejected	Interval	Numeric	1136	Varsel:Small R-square va...
_1138	Rejected	Interval	Numeric	1138	Varsel:Small R-square va...
_1140	Rejected	Interval	Numeric	1140	Varsel:Small R-square va...
_1142	Input	Interval	Numeric	1142	
_1144	Rejected	Interval	Numeric	1144	Varsel:Small R-square va...
_1146	Rejected	Interval	Numeric	1146	Varsel:Small R-square va...
_1148	Rejected	Interval	Numeric	1148	Varsel:Small R-square va...
_1150	Input	Interval	Numeric	1150	
_1152	Rejected	Interval	Numeric	1152	Varsel:Small R-square va...

Fonte: (SAS)

No segundo caso (Processo 2) foi realizada a seleção dos principais atributos apontados pelo nó “Variable Selection” (aqueles com maiores efeitos) para serem utilizados na geração do modelo de predição. O critério de seleção adotado foi utilizar todos os atributos cujo valor de efeito atingisse no mínimo 50% do maior valor, o que resultou na redução agressiva do número de atributos.

O particionamento da base de dados foi feito através do nó “Data Partition”, sendo selecionados 66% para a base de treinamento (geração do modelo de predição) e 34% para a validação. A seleção foi feita de forma aleatória, sendo utilizada a mesma semente nos processos 1 e 2, para que as bases fossem idênticas.

O método utilizado na mineração de dados foi o nó "Auto Neural", que consiste em uma rede neural artificial que procura encontrar os parâmetros ótimos para a execução do algoritmo de forma automática, gerando assim o modelo.

Após a execução, foram analisados os resultados através dos valores de RMSE (raiz do erro quadrático médio) apresentados na saída do nó "Auto Neural". O RMSE é comumente utilizado em comparações no software SAS. Tso e Yau (2007) utilizaram o RMSE para comparar diferentes métodos do software SAS Enterprise Miner na predição do consumo de energia elétrica em Hong Kong. Kim (2008) também utilizou o SAS para comparar os métodos de Redes Neurais, Árvores de Decisão e Regressão Linear, utilizando o RMSE.

Foi realizada a comparação entre os processos 1 e 2, analisando os resultados obtidos em relação aos diferentes métodos de seleção de atributos. Também foi utilizado o valor do NRMSE, que elimina a dependência da unidade de medida original através da normalização do RMSE, dividindo o erro pelo valor da amplitude do teor de determinado nutriente nas amostras. Isso facilita a posterior comparação dos resultados de diferentes nutrientes.

RESULTADOS E DISCUSSÃO

Os valores dos erros RMSE e NRMSE resultantes da mineração de dados para estimativa do teor de cada nutriente e da produtividade, referente aos processos 1 e 2, bem como o resultado da seleção de atributos em cada um, além do percentual de redução da quantidade de atributos do Processo 2 em relação ao Processo 1 são apresentados na Tabela 1.

Tabela 1 – Resultados da mineração de dados dos processos 1 e 2.

Atributo Meta	Processo 1			Processo 2			
	Quantidade de atributos	RMSE	NRMSE	Atributos	RMSE	NRMSE	Redução percentual
N	79	5,61	9,60	1460	6,35	10,86	98,73%
P	25	9,22	21,57	1122; 1156	9,62	22,53	92,00%
K	45	2,14	6,68	1510; 1512; 1558; 1648	4,13	12,85	91,11%
Ca	78	0,76	21,18	1450	0,49	13,63	98,71%
Mg	83	0,80	17,51	1238; 1332	0,86	18,70	97,59%
S	28	5,30	13,29	1122; 1156	10,41	26,09	92,85%
Cu	49	2,65	16,99	1194; 1414; 1436; 1532	3,12	20,06	91,83%
Mn	47	69,68	8,34	1100; 1116; 1118; 1158; 1250; 1278; 1290	94,72	11,34	85,10%
Fe	68	52,11	7,74	1100; 1102; 1482; 1542; 1576	54,88	8,16	92,64%
Zn	84	6,84	13,90	1132; 1134; 1180	9,21	18,70	96,42%
B	81	4,49	14,86	1468	3,96	13,08	98,76%
Produtividade	76	705,84	12,81	1118; 1410; 1502; 1522; 1582	570,41	10,35	93,42%

‡ NRMSE = RMSE/Amplitude do atributo meta na base de dados.

Valores de erro NRMSE abaixo de 10% em relação à amplitude foram obtidos nos nutrientes N, K, Mn e Fe, sendo que nenhum dos demais nutrientes ultrapassou a faixa de 30% de erro. Isso indica que é possível obter bons resultados de predição do teor de nutrientes por meio de espectroscopia NIR.

Dentre os elementos, apenas um não apresentou a redução agressiva na quantidade de atributos do Processo 2 em relação ao Processo 1, sendo esse o Manganês, que obteve aproximadamente 85% de redução da dimensionalidade.

Alguns elementos se beneficiaram mais da redução agressiva do número de atributos, sendo eles: N, P, Ca, Mg, Fe, B e a Produtividade, os quais apresentaram erros próximos ou até menores para o Processo 2 em relação ao Processo 1. Considerou-se benefício tanto a diminuição do erro quanto a obtenção de um valor próximo, pois nesse caso a vantagem seria a simplificação do modelo sem perda considerável de qualidade. No caso do Ca, B e Produtividade, além da simplificação do modelo, obteve-se resultados melhores para o Processo 2. Isso acontece porque a seleção de atributos agressiva pode remover a interferência de comprimentos de onda não relevantes e manter somente o necessário para a predição.

Os demais nutrientes apresentaram valores de erro menores para o Processo 1 em relação ao Processo 2. Isso ocorre porque em alguns casos a redução da dimensionalidade não produz resultados melhores, devido à perda excessiva de informação para o modelo de predição.

A técnica de seleção agressiva de atributos é amplamente utilizada na categorização de texto, que assim como em espectroscopia, possui alta dimensionalidade. Gabrilovich e Markovitch (2004) obtiveram acurácia de 0,769 com 100% dos atributos, e de 0,853 com 0,5% dos atributos, na categorização de texto utilizando máquina de vetor de suporte (SVM).

No ramo da espectroscopia, existem trabalhos que comparam métodos de seleção de atributos, e é possível observar melhores resultados à medida que se reduz a dimensionalidade da base de dados. Balabin e Smirnov (2011) fizeram a comparação de métodos de seleção de atributos para a predição de parâmetros de qualidade em biodiesel por meio de espectroscopia NIR, obtendo resultados melhores conforme o número de atributos diminuiu: para o teor de metanol, os resultados extremos foram de RMSE igual a 306 ppm, utilizando 316 atributos e 107 ppm utilizando 122 atributos. Friedel, Patz e Dietrich (2013) utilizaram o espectro de infravermelho por transformada de Fourier (FT-IR) para a análise de componentes do vinho e realizaram a seleção de atributos agressiva por meio do método de Algoritmo Genético e Mínimos Quadrados Parciais (GA-PLS – do inglês, *Genetic Algorithm and Parcial Least*

Squares), reduzindo em cerca de 92,5% o número de atributos e obtendo resultados mais baixos de RMSE em relação ao espectro inteiro.

Quanto ao número de amostras utilizadas, Ferreira, Poppi e Pallone (2015) utilizaram 80 amostras, obtendo coeficiente de determinação R^2 de 0,80 para a estimativa de fibras dietéticas em grãos de soja; Ferreira et. al. (2014) utilizaram 40 amostras para determinação de parâmetros de qualidade em grãos de soja, obtendo valores de R^2 superiores a 0,80 na maioria dos casos; Santos, Santos e Korndörfer (2012) utilizaram 226 amostras, obtendo coeficiente de correlação R de 0,79 e 0,85, na estimativa de nitrogênio foliar em milho e soja, respectivamente; todos eles utilizando espectroscopia. Sendo assim, a quantidade de amostras utilizadas nesse trabalho foi ligeiramente superior à maioria dos trabalhos correlatos encontrados na literatura. Mesmo assim, essas quantidades podem ser consideradas pequenas, mas devido às análises laboratoriais que demandam custos de tempo e recursos para serem efetuadas, quantidades muito maiores são impraticáveis.

Os resultados obtidos nesse trabalho são satisfatórios, pois na maioria dos casos o modelo de predição não perdeu em qualidade, dada a diminuição massiva do número de atributos e a simplificação que isso traz ao modelo. Sendo assim, a técnica de seleção de atributos agressiva unida à mineração de dados de espectroscopia pode ser uma alternativa para a simplificação da estimativa do teor de nutrientes em grãos de soja.

CONCLUSÃO

Os resultados da mineração dos dados das diferentes bases para a estimativa do teor de nutrientes de grãos de soja indicam que a seleção de atributos agressiva é capaz de manter ou até melhorar a qualidade da predição, além de simplificar o modelo e melhorar o desempenho do processamento, quando aplicada em dados de espectroscopia NIR.

AGRADECIMENTOS

Ao CNPq pelo apoio financeiro.

REFERÊNCIAS

BALABIN, R. M.; SMIRNOV, S. V. Variable selection in near-infrared spectroscopy: Benchmarking of features selection methods on biodiesel data. Analytica Chimica Acta, v. 692, p. 63-72, 2011.

BEN-DOR, E.; BANIN, A. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. Soil Science Society of America Journal, Madison, v. 59, n. 2, p. 364-372, Mar./Apr 1995.

FERREIRA, D.S. et al. Comparison and application of near-infrared (NIR) and mid-infrared (MIR) spectroscopy for determination of quality parameters in soybean samples. Food Control, [S.l], v.35, p. 227-232, jan. 2014.

FERREIRA, D.S.; POPPI, R.J.; PALLONE, J.A.L. Evaluation of dietary fiber of Brazilian soybean (Glycine max) using near-infrared spectroscopy and chemometrics. Journal of Cereal Science, [S.l], v.64, p. 43-47, mai. 2015.

FRIEDEL, M.; PATZ, C. D.; DIETRICH, H. Comparison of different measurement techniques and variable selection methods for FT-MIR in wine analysis. Food Chemistry, v. 141, p. 4200-4207, 2013.

GABRILOVICH, E.; MARKOVITCH, S. Text Categorization with Many Redundant Features: Using Aggressive Feature Selection to Make SVMs Competitive with C4.5. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 21., 2004, Banff. Anais... 2004.

KIM, Y. S. Comparison of the decision tree, artificial neural network, and linear regression methods based on the number and types of independent variables and sample size. Expert Systems with Applications, v. 34, p. 1227-1234, 2008.

LIU, H.; MOTODA, H. Computational Methods of Feature Selection. Boca Raton: Chapman & Hall/crc, 2007. 419 p.

PIRAMUTHU, S. Evaluating Feature Selection Methods for Learning in Data Mining Applications. European Journal of Operational Research, v. 156, n. 2, p. 483-494, 2004.

POLTRONIERI R. Modelos Matemáticos e computacionais para avaliação do estado nutricional da soja por meio de diagnose em folhas e grãos. 2015. 86 f. Dissertação (Mestrado em Computação Aplicada) – Setor de Ciências Agrárias e de Tecnologia, Universidade Estadual de Ponta Grossa, Ponta Grossa, 2015.

SANTOS, G. A.; SANTOS, A. P.; KORNDÖRFER, G. H. Sistema por Infravermelho próximo (NIR) para análise de nitrogênio foliar. Biosci. J., Uberlândia, v. 28, n. 1, p. 83-90, mar. 2012.

TAN, P.; STEINBACH, M.; KUMAR, V. Introduction to data mining. Pearson Education, 2006.

TAN, M.; WANG, L.; TSANG, I. W. Learning Sparse SVM for Feature Selection on Very High Dimensional Datasets. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 27., 2010, Haifa. Anais... 2010.

TSO, G. K. F.; YAU, K. K. W. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. Energy, v. 32, n. 1, p. 1761-1768, 2007.

Determinação da severidade de *Septoria lycopersici* em folhas de tomateiro por meio de imagens digitais

Amanda do Prado Mattos[1], *Adriana Terumi Itako*[2], *João Batista Tolentino Júnior*[2]

[1] Estudante do curso de Agronomia, Campus de Curitibanos, Universidade Federal de Santa Catarina, Curitibanos, Santa Catarina, Brasil, amanda.mattos@grad.ufsc.br

[2] Departamento de Agricultura, Biodiversidade e Florestas, Centro de Ciências Rurais, Campus de Curitibanos, Universidade Federal de Santa Catarina, Curitibanos, Santa Catarina, Brasil, adriana.itako@ufsc.br, joao.tolentino@ufsc.br

RESUMO

A determinação da severidade de doenças em plantas pode ser facilitada com a utilização de softwares que quantifiquem a área foliar da planta afetada a partir de análises de imagens digitais. Sendo assim, foi realizada a coleta e obtenção de imagens digitais de folhas sadias e doentes de tomateiro em diferentes graus de severidade da doença causada pelo fungo *Septoria lycopersici*. Inicialmente, as imagens foram submetidas à edição no software GIMP para delimitação das áreas lesionadas. A partir de um script escrito em R com o pacote EBImage, a imagem foi decomposta em três níveis de cor (RGB) e, por meio do processo de limiarização, o fundo da imagem foi desconsiderado, e posteriormente, a área lesionada da folha foi separada da área sadia da folha. A metodologia descrita foi capaz de identificar e quantificar as áreas lesionadas das folhas com sintomas de doenças.

PALAVRAS-CHAVE: EBImage, software R, visão computacional, *Solanum lycopersicum*.

ABSTRACT

The determination of the severity of plant diseases can be facilitated using software that quantifies the affected leaf area from the analysis of digital images. Digital imagen of leaves with and without symptoms were collected of tomato in different degrees of severity of the disease caused by the fungus *Septoria lycopersici*. Initially, the images were submitted to editing in the GIMP software to delineate the injured areas. From a script written in R with the

EBImage package, the image was decomposed into three color levels (RGB) and, through the thresholding process, the background of the image was disregarded, and subsequently, the injured area of the leaves was separated from the healthy leaf area. The described methodology could identify and quantify the injured areas of leaves with disease symptoms.

INTRODUÇÃO

A epidemiologia agrícola é um ramo da fitopatologia que trata das populações de plantas e dos patógenos, assim como as interações entre si, com o ambiente e com o ser humano, tendo como objetivo criar medidas e procedimentos viáveis, práticos e eficientes para os agricultores, no manejo de doenças em plantas. Para tanto, a quantificação de doenças nas plantas é um dos principais pilares para a estudo e análise das epidemias, através da obtenção de dados quantitativos sobre a ocorrência e desenvolvimento das doenças. A intensidade da doença na planta pode ser determinada pela incidência, que é o número de unidades doentes visíveis na planta, e/ou pela severidade, sendo a área do tecido da planta doente (VALE et al., 2004).

A severidade é a que melhor demonstra a quantidade de tecido lesionado na planta, e os principais métodos utilizados para a sua determinação são a escala descritiva, que determina a intensidade da doença por meio de graus ou notas e a escala diagramática, que é uma ilustração da planta ou parte dela com os sintomas típicos da doença em diferentes níveis de severidade (BERGAMIN FILHO & LOPES, 1995).

A determinação da severidade de uma doença pela análise visual, diferente da incidência, é demorada e sujeita a erros, visto que tal método de quantificação é subjetivo, necessitando de grande conhecimento sobre a doença estudada. O olho humano tem capacidade limitada em interpretar pequenas diferenças na severidade, sendo muito difícil a determinação visual da severidade, e o que normalmente ocorre é a subjetividade da percepção da quantidade da doença pelo avaliador, ocorrendo variações entre diferentes avaliadores, levando a uma superestimação ou subestimação da doença (VALE et al., 2004).

Sendo assim, tarefas como a determinação da severidade de doenças em plantas pode ser facilitado e ter uma maior precisão, com a utilização de softwares, que quantifiquem a área foliar da planta afetada a partir de análises de imagens digitais. O sistema de reconhecimento de imagens digitais é capaz de extrair características dos objetos das imagens digitais e relacionar com a grandeza de interesse, facilitando a determinação correta da severidade.

A septoriose é uma doença que ataca principalmente o tomateiro sendo severa em áreas de umidade alta e temperatura de 20 a 25°C. Aparecem inicialmente nas folhas mais velhas em forma de manchas circulares pequenas que se unem causando queima e desfolha das plantas. As manchas possuem coloração marrom escura nas bordas e cinza no centro, podendo haver ou não um halo clorótico, região ao redor da necrose de coloração amarelada. No caule, pecíolo e sépala as manchas são mais escuras e menores, e o fruto é raramente afetado. A desfolha da planta expõe os frutos a queima pela luz solar, diminuindo o vigor da planta (REIS et al., 2006).

A quantificação deve apresentar resultados precisos e com exatidão para a correta interpretação das variações na intensidade da doença e precisão do seu efeito na produção. Assim, a utilização de softwares para facilitar este processo torna-se uma ótima opção. O que se espera, portanto, é a obtenção de dados precisos e exatos, com o aumento da eficiência e facilitação do processo de quantificação, em relação à avaliação apenas visual, e com isso facilitar processos que demandam de uma correta quantificação de severidade. E ainda, utilizando de imagens digitais obtidas de câmeras fotográficas comuns analisadas em um software oferecido gratuitamente.

A partir disso, o objetivo do trabalho é determinar a severidade de doenças foliares em plantas por meio da análise computacional de imagens digitais de folhas afetadas.

MATERIAL E MÉTODOS

O experimento foi realizado no Laboratório de Fitopatologia da Universidade Federal de Santa Catarina, Campus de Curitibanos, e a metodologia foi composta pelas etapas de coleta das folhas doentes, obtenção das imagens digitais e quantificação da severidade tradicional.

A coleta das folhas foi realizada em lavouras comerciais da região e em experimentos realizados para este fim. Foram coletadas 37 folhas de plantas de tomateiro (*Solanum lycopersicum*) com a ausência e a presença de lesões da doença septoriose causada pelo fungo *Septoria lycopersici*, sendo estas selecionadas de acordo com o grau de desenvolvimento da doença, variando desde a ausência da doença até elevada severidade na folha, evitando-se folhas totalmente mortas. As doenças foram identificadas a partir da análise de sua sintomatologia conforme Kimati *et al*. (2005).

A coleta das folhas foi realizada com o auxílio de uma tesoura de poda, com corte na base da folha. Para evitar desidratação, e consequente influência nas estimativas, logo após sua coleta, as folhas foram colocadas dentro de sacos de polietileno fechados, e transportadas dentro de caixa de isopor para o laboratório (DUARTE; LOBO JR.; CORTES, 2013).

Cada folha foi disposta sobre fundo branco em um ambiente com boa iluminação. A folha foi fotografada, com câmara fotográfica digital, na face abaxial e adaxial, e a imagem obtida transferida para um microcomputador. Foi utilizado um notebook com processador Intel ® Core ™ i-3-4025U, 1.9 GHz, com 4,00 GB de memória RAM, e sistema operacional de 64 bits, com configuração de tela de 1366 x 768 pixels.

Posteriormente à obtenção das imagens digitais, foi realizado a identificação visual da área afetada de cada folha. As imagens foram submetidas a tratamento no software GIMP para a sua edição e a delimitação das áreas lesionadas, e posterior mensuração das áreas pelo software R.

A partir da utilização do software R e seu pacote EBImage, as imagens digitais editadas foram analisadas. Dentre os comandos utilizados, aqueles que se destacam pela essencialidade no processo de quantificação, foram, *channel*, correção gama (^), *otsu* e *computeFeatures.shape*.

O comando *channel* (canal), foi utilizado primeiramente para alterar a imagem para o modo cinza, convertendo a imagem colorida para apenas uma camada de imagem (*red, green* ou *blue*) (PAU et al., 2010). Tal escolha depende da coloração da lesão e da folha, sendo no presente trabalho utilizado a camada *green*, visto a coloração da folha, em sua predominância, verde, e das lesões de coloração escura. A correção gama foi utilizada com o intuito de alterar o contraste da imagem. Em um primeiro momento este comando foi empregado para tornar a folha mais escura em relação ao fundo (expoente 2,1), e em um segundo momento tornando as lesões mais evidentes com o clareamento da folha (expoente 0,3).

Os procedimentos descritos anteriormente possibilitam o processo de limiarização (*threshold*), sendo este utilizado com o objetivo de separar o fundo, da folha, e posteriormente as lesões, do fundo. Tal processo ocorre a partir da determinação de um limiar, que define o limite entre o fundo e o objeto de interesse (MOREIRA, 1999) para a transformação da imagem cinza em uma imagem binária (preto e branco). O cálculo do limiar foi realizado com a função *otsu*, que separa o histograma em duas classes (objeto e fundo) pela minimização da variância intraclasse.

A partir da separação foi utilizado o comando *ComputeFeatures.shape*, que analisa as regiões em destaque da imagem e realiza cálculos das suas dimensões, dentre elas, a área dos objetos da imagem (pixels^2), essencial para o cálculo da proporção de área lesionada em relação a folha.

RESULTADOS E DISCUSSÃO

As fotografias das folhas foram editas no software GIMP para delimitação das áreas lesionadas. Neste procedimento, as manchas nas folhas foram pintadas na cor preta para garantir o máximo contraste com a cor da folha. Optou-se pela delimitação apenas das áreas necrosadas, não considerando as áreas cloróticas (Figura 1).

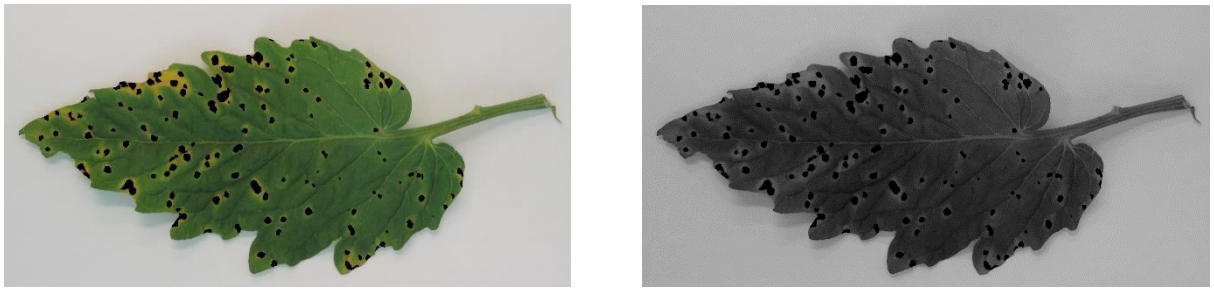
Figura 1. Fotografia editada para delimitação da área foliar lesionada pela doença causada pelo fungo *Septoria lycopersici*.



Fonte: Autores

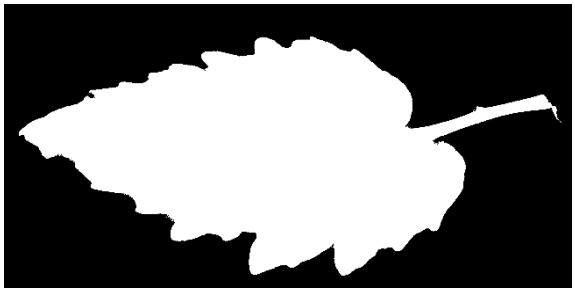
Os comandos utilizados para separação do canal de cor e correção gama resultaram em uma imagem com elevado contraste entre a folha e o fundo, onde avaliou-se que quanto maior o valor utilizado para a correção gama, mais escura a folha em relação ao fundo (Figura 2). Esse processo evidenciou o fundo, facilitando a correta limiarização, com consequente separação folha/restante da imagem (Figura 3).

Figura 2. Resultado do processo de correção gama (expoente 2,1)



Fonte: Autores

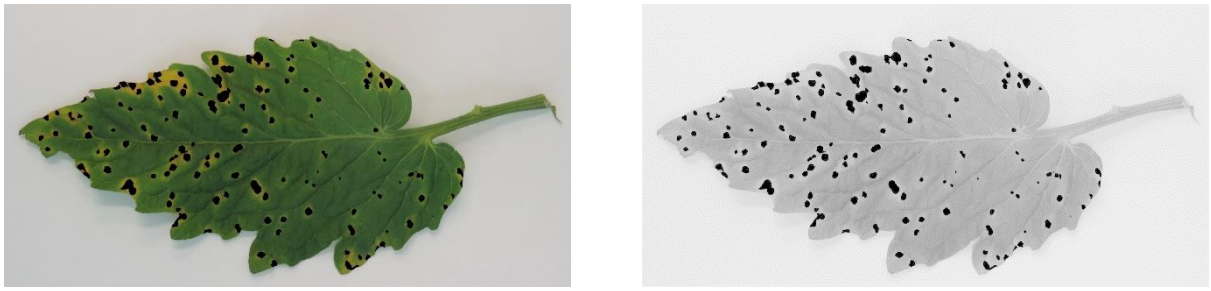
Figura 3. Separação folha/restante da imagem a partir da limiarização



Fonte: Autores

Com o expoente de correção gama menor, a folha apresentou menor brilho, ficando mais clara e evidenciando as lesões (Figura 4). Seguindo o mesmo processo de limiarização, apresentou uma ótima separação de lesão/restante da imagem (Figura 5).

Figura 4. Resultado do processo de correção gama (expoente 0,3)



Fonte: Autores

Figura 5. Separação lesão/restante da imagem a partir da limiarização



Fonte: Autores

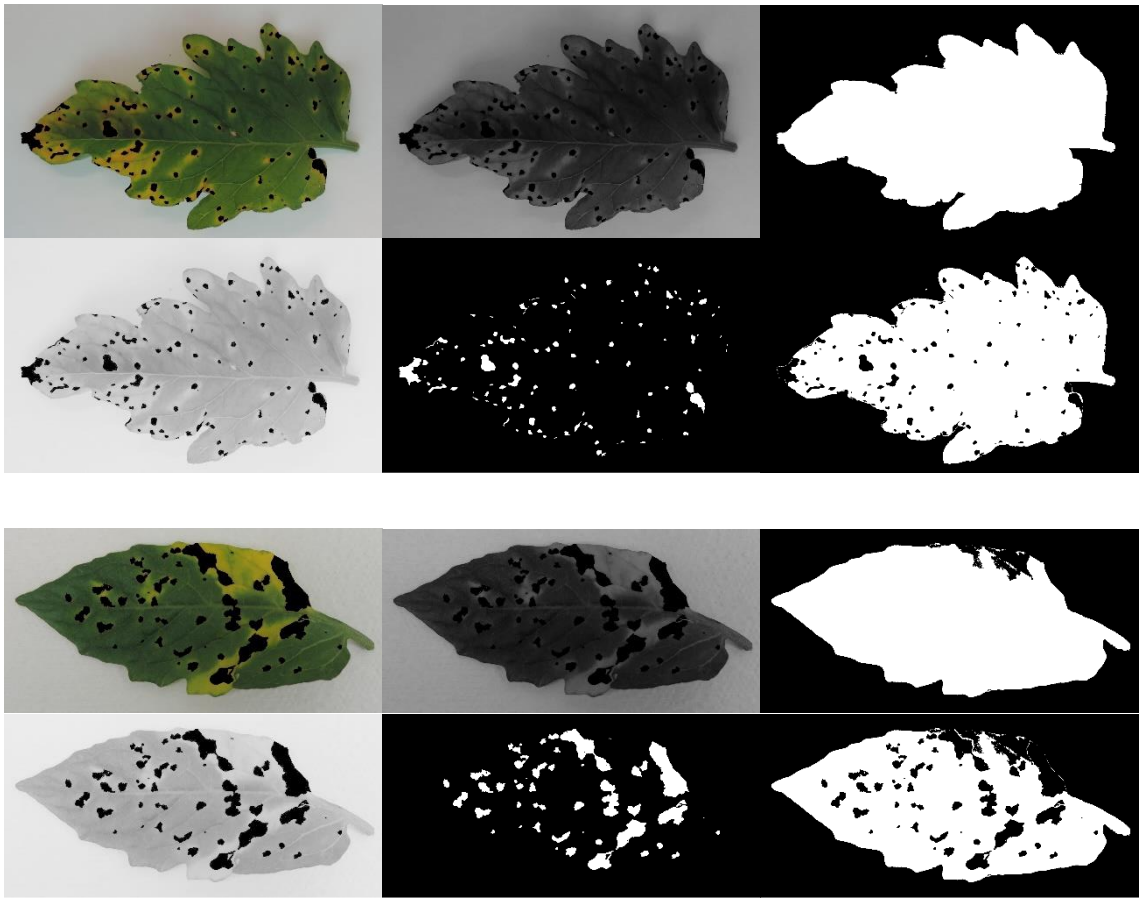
Com as áreas de interesse devidamente delimitadas, e com a utilização do comando *ComputeFeatures.shape*, foi possível obter os resultados de área foliar e área de lesões (Tabela 1) e calcular a proporção de área foliar lesionada.

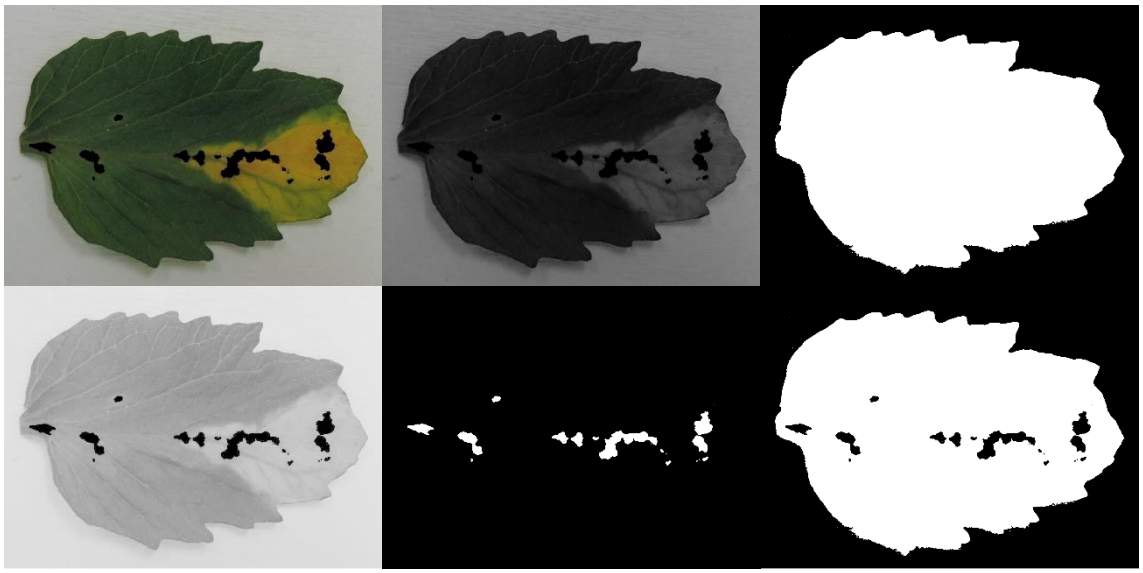
Tabela 1. Resultados de área foliar, de lesões e a relação entre estas no software R, da folha de tomateiro exemplificada acima.

Folha					
s.area	s.perimeter	s.radius.mean	s.radius.sd	s.radius.min	s.radius.max
641676	4762	510,04	151,61	280,58	847,30
Lesões					
s.area	s.perimeter	s.radius.mean	s.radius.sd	s.radius.min	s.radius.max
30724	24	3,39	1,78	0,68	6,06
% Área foliar afetada					
4,79					

Fonte: Autores

Figura 6 – Exemplos do processo de limiarização da área lesionada da área foliar total para outras três imagens diferentes.





Fonte: Autores

Na figura 6, é possível observar o resultado do processo de limiarização para outras três imagens de folhas com diferentes níveis de severidade de doença. De modo geral, a limiarização é bastante satisfatória, sendo possível a correta determinação da proporção de área foliar lesionada, de 5,22%, 13,89% e 2,94% respectivamente.

Em trabalhos futuros, esta metodologia pode ser adaptada para outras doenças de importância agrícola que causam manchas foliares. Mesmo que as diferentes doenças causem sintomas bastante diferentes, em relação à cor, tamanho, formato, espera-se que a metodologia proposta neste trabalho possa identificar as lesões e quantificar de forma adequada.

CONCLUSÕES

Todas as 37 imagens analisadas apresentaram resultados esperados, havendo algumas diferenciações quanto a delimitação de folha/restante da imagem, porém nada que tenha interferido de forma significativa nos resultados finais, visto que os pixels em discordância eram poucos. A metodologia descrita foi capaz de identificar e quantificar as áreas lesionadas das folhas com sintomas de doenças.

REFERÊNCIAS

AGRIOS, G. N. **Plant Pathology.** 5. ed. Burlington: Elsevier Academic Press, 2005. 922 p.

AMORIM, L.; BERGAMIM FILHO, A. Fenologia, Patometria e Quantificação de danos. In: AMORIM, L.; REZENDE, J. A. M.; BERGAMIM FILHO, A. **Manual de fitopatologia**: Princípios e Conceitos. 4. ed. v. 1. Piracicaba: Agronomica Ceres, 2011. 704 p.

AMORIM, L.; REZENDE, J. A. M.; BERGAMIM FILHO, A. **Manual de fitopatologia**: Princípios e Conceitos. 4. ed. v. 1. Piracicaba: Agronomica Ceres, 2011. 704 p.

BERGAMIN FILHO, A.; LOPES, D. B. **Avaliação dos danos causados por doenças de plantas.** Revisão Anual de Patologia de Plantas. São Paulo: Editora Agronômica Ceres 1995. v.3, p. 133-184.

CONCI, A.; AZEVEDO, E.; LETA, E. F. R. **Computação Gráfica:** Processamento e Análise de Imagens Digitais, v. 2 Campus/Elsevier. 2008

DUARTE, L. T.; LOBO JR., M.; CORTES, M. V. C. B. **Recomendações para Coleta e Envio de Material ao Laboratório de Fitopatologia para Diagnóstico de Doenças em Plantas.** Documentos, Embrapa Arroz e Feijão, 2013. 20 p. Disponível em: <http://www.cnpaf.embrapa.br/transferencia/informacoestecnicas/publicacoesonline/seriedocumentos_271.pdf> Acesso em: 27 abr. 2016.

FERREIRA, T.; RASBAND, W. **ImageJ:** User Guide. IJ 1.46r. 2012. Disponível em: <https://imagej.nih.gov/ij/docs/guide/user-guide.pdf> Acesso em: 27 abr. 2016.

KIMATI, H.; AMORIM, L.; REZENDE, J. A. M.; BERGAMIM FILHO, A.; CAMARGO, L. E. A. **Manual de Fitopatologia:** Doenças de Plantas Cultivadas. 4. ed. v. 2. São Paulo: Agronomica Ceres, 2005. 663 p.

MARQUES FILHO, O.; VIEIRA NETO, H. **Processamento Digital de Imagens**. Rio de Janeiro: Brasport, 1999. 307 p.

MORAES, S.A. **Quantificação de doenças de plantas**. Infobibos, 2007. Disponível em: <http://www.infobibos.com/Artigos/2007_1/doencas/index.htm>. Acesso em: 21 abr. 2016

PAU, G.; FUCHS F.; SKLYAR O.; BOUTROS M.; HUBER W. EBImage—an R package for image processing with applications to cellular phenotypes. **Bioinformatics,** v. 26, n. 7, p. 979-981, 2010.

PEDRINI, H.; SCHWARTZ, W. R. **Análise de imagens digitais:** Princípios, Algoritmos e Aplicações, Cengage Learning, 2007

REIS, A.; BOITEUX, L. S.; LOPES, C. A. **Mancha-de-septória: doença limitante do tomateiro no período de chuvas.** Comunicado Técnico, 37 – Embrapa, 2006. Disponível em: <http://ainfo.cnptia.embrapa.br/digital/bitstream/item/103062/1/cot-37.pdf> Acesso em: 27 abr. 2016

VALE, F. X. R.; JESUS JR., W. C.; ZAMBOLIM, L. **Epidemiologia aplicada ao manejo de doenças de plantas**. Belo Horizonte: Editora Perffil, 2004. 531p

Influência da resolução espacial em imagens obtidas por RPA na estimativa de doenças na soja

David Patrick Zampier[1], Antônio David Viniski[2], Gislaine Gabardo[3], Alessandro Stachak[2], Alaine Margarete Guimarães [4], Maristella Dalla Pria[5]

[1] Curso de Engenharia de Computação – Bolsista do Programa de Iniciação Científica, Universidade Estadual de Ponta Grossa, Ponta Grossa, Paraná, Brasil, davidzampier@gmail.com

[2] Programa de Pós-Graduação em Computação Aplicada, Universidade Estadual de Ponta Grossa, Ponta Grossa, Paraná, Brasil, davidviniski@gmail.com

[3] Programa de Pós-Graduação em Agronomia, Universidade Estadual de Ponta Grossa, Ponta Grossa, Paraná, Brasil, gislainegabardo2007@yahoo.com.br

[4] Departamento de Informática, Universidade Estadual de Ponta Grossa, Ponta Grossa, Paraná, Brasil, alainemg@uepg.br

[5] Departamento de Fitotecnia e Fitossanidade, Universidade Estadual de Ponta Grossa, Ponta Grossa, Paraná, Brasil, mdallapria@uol.com.br

RESUMO

A utilização de imagens obtidas por meio de aeronave remotamente pilotada (RPA) pode facilitar a estimativa de doenças na soja. Uma dificuldade está na determinação da resolução espacial das imagens mais adequada para tal estimativa, visto que trabalhos recentes mostraram que nem sempre um maior nível de detalhamento dos dados significa melhores resultados. Este trabalho teve como objetivo analisar a influência da resolução espacial de imagens obtidas por RPA, com resoluções de 3,4 e 15 cm/pixel, na estimativa da ferrugem asiática e do oídio na soja. Foram efetuados voos nas duas resoluções utilizando câmeras RGB e NIR numa área experimental com duas cultivares diferentes da soja, situada em Ponta Grossa - PR. Foram gerados mapas de refletância a partir das imagens coletadas, os quais foram submetidos a uma correção geométrica. Os dados extraídos destes mapas foram minerados utilizando algoritmo de Máquina de Vetores de Suporte para Regressão (SVR), disponível na ferramenta R. Como resultado, os dados extraídos dos mapas com resolução de

15 cm/pixel mostraram correlação maior ou igual em relação aos de 3,4 cm/pixel. Em adição, os dados extraídos dos mapas da câmera RGB proporcionaram uma maior correlação, comparados com da câmera NIR. Pode-se concluir que as imagens com maior resolução espacial não são necessariamente mais adequadas para a estimativa de doenças na soja, podendo-se, nesse caso, utilizar satisfatoriamente resolução mais baixa o que facilita tanto o processamento das imagens como a mineração dos dados.

PALAVRAS-CHAVE: Sensoriamento remoto, Severidade de doenças, Aeronave Remotamente Pilotada.

ABSTRACT

The use of images obtained by remotely piloted aircraft (RPA) can facilitate the severity estimation of soybean diseases. Since recent studies have shown that not always a higher level of detail of the images data means better results, a challenge is to determine the best images spatial resolution for the severity estimation. In this paper, we present the influence of spatial resolution of RPA images, using 3.4 and 15 cm/pixel resolutions, on the estimation of Asian rust and powdery mildew on soybean. The two resolutions were analyzed using RGB and NIR cameras in an experimental area with two different soybean cultivars located in Ponta Grossa - PR. Reflectance maps were generated from the collected images, which were submitted to a geometric correction. Data extracted from these maps were mined using Regression Support Vector Machine (SVR) algorithm, available in R tool. As result, the data extracted from maps with 15 cm/pixel resolution showed equal or higher correlation than those with 3.4 cm/pixel, with both RGB and NIR cameras. Data extracted from images generated with RGB camera provided a higher correlation compared to the NIR camera. It can be concluded that images with higher spatial resolution are not necessarily more adequate for the severity estimation of soybean diseases. In this case, lower resolution can be satisfactorily used, making easier both the image processing and the data mining.

KEYWORDS: Remote Sensing, Disease severity, Remotely Piloted Aircraft.

INTRODUÇÃO

A quantificação de danos é um ponto chave na definição de qualquer estratégia de controle de doenças (Hikishima et al, 2010). Tradicionalmente, estimativas visuais identificam uma doença com base em seus sintomas característicos na planta (lesões, ferrugem, perda de vigor, murchas, apodrecimento, etc.) ou sinais visíveis do patógeno. Essas estimativas visuais são

realizadas por especialistas treinados e têm sido objeto de intensas pesquisas e investigações (Mahlein, 2016).

As técnicas de sensoriamento remoto, que consistem da obtenção de informações de um objeto sem que haja contato físico entre o instrumento de medida e o objeto, vêm sendo empregadas com sucesso em estudos de manejo de doenças. As medidas de refletância espectral constituem o principal exemplo da aplicação destas técnicas na fitopatologia, uma vez que existem diferenças nas respostas espectrais de uma folha sadia e uma folha doente (Silva et al., 2000).

Atualmente, as Aeronaves Remotamente Pilotadas (RPAs – *Remotely Piloted Aircraft*), também conhecidas como DRONE, VANT ou ainda UAV (do inglês *Unmanned Aerial Vehicle*), vêm se tornando alternativas às técnicas convencionais de sensoriamento remoto, devido ao seu baixo custo, altas resoluções espaciais e temporais das imagens. Segundo Mahlein et al. (2012), a resolução espacial tem uma forte influência na detecção de doenças em plantas ou em interações planta-patógeno. Nas altas resoluções espaciais, cada pixel da imagem representa uma área menor do alvo que se deseja detectar, conforme se diminui a resolução espacial, cada pixel passa a representar uma área maior do alvo. Assim, a resolução espacial de 5 cm/pixel (custo computacional e nível de detalhe maiores) é maior que a resolução espacial de 10 cm/pixel (custo computacional e nível de detalhe menores).

Di Gennaro et al. (2016) realizaram um estudo comparando o uso de imagens multiespectrais de alta resolução (5 cm/pixel) adquiridas através de RPA com a análise da planta em campo na identificação de sintomas foliares da doença da folha da videira (GLSD – do inglês, *Grapevine Leaf Stripe Disease*). Após testes estatísticos, os autores encontraram que as imagens obtidas pela RPA podem discriminar diferentes níveis de estresse na videira, em relação aos sintomas visuais da GLSD.

Garcia-Ruiz et al. (2013) obtiveram resultados melhores em seu estudo com a resolução de 5,45 cm/pixel, comparada com 50 cm/pixel, na detecção da *Greening* nas plantas cítricas. Também, Lacerda (2017) obteve bons resultados com resolução alta (3,4 cm/pixel) em seu estudo, que visava estimar a severidade de ferrugem asiática na soja.

Associado a resolução espacial, está o tipo de câmera utilizado para a captura das imagens, que podem ser câmeras do espectro visível (RGB - *red, green, blue*) ou infravermelho-próximo (NIR - *near infrared*), até câmeras multiespectrais e hiperespectrais. O tipo de câmera adequado dependerá da finalidade na qual ela será empregada. Gerke (2017) realizou um estudo comparando diferentes resoluções espaciais (3,4 cm/pixel e 20 cm/pixel) e diferentes câmeras (RGB e NIR) em imagens obtidas por RPA, visando a estimativa da

produtividade do trigo. Os melhores resultados foram obtidos com a resolução mais baixa (20 cm/pixel), para a câmera RGB.

No entanto, ainda são poucos os estudos comparando o comportamento de diferentes resoluções espaciais em estimativas agrícolas, sendo que estas interferem diretamente no nível de detalhamento dos dados. Ainda é um desafio determinar qual resolução espacial de imagem é mais apropriada para a estimativa satisfatória de severidade de doenças em plantas. Com isso, este trabalho teve como objetivo avaliar a influência de diferentes resoluções espaciais de imagens obtidas por RPA com diferentes câmeras, na estimativa da severidade da ferrugem asiática e do oídio na soja (*Glycine max*).

MATERIAL E MÉTODOS

Neste estudo, realizou-se o acompanhamento da cultura da soja, sendo efetuados voos semanais na área de estudo, utilizando RPA. Em paralelo, foram realizadas avaliações semanais da cultura em campo, documentando a severidade das doenças presentes e outras características fisiológicas das plantas.

Área de estudo

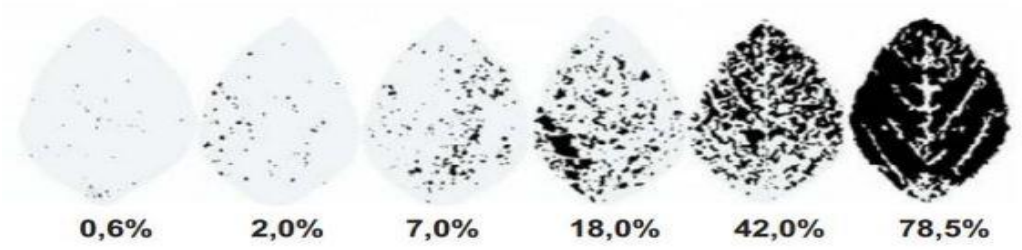
A área de estudo situa-se na Fazenda Escola da Universidade Estadual de Ponta Grossa, localizada na cidade de Ponta Grossa – PR e o delineamento experimental compreende quatro parcelas (denominadas A3, A4, B3 e B4). Cada parcela foi subdividida em quatro blocos, com 11 tratamentos por bloco, onde cada tratamento corresponde a um insumo específico aplicado. A soja foi semeada no dia 24/10/2016 nas parcelas A3 e A4, e no dia 13/12/2016 nas parcelas B3 e B4.

As parcelas A3 e B4 continham a cultivar 5969 Nidera, com população de 14 ou 15 plantas por metro. Já as parcelas A4 e B3 correspondiam a cultivar TMG 7262, com a população de 10 plantas por metro.

Medida da Severidade em campo

Para obtenção da severidade das doenças em campo foram avaliadas, no centro de cada parcela, sete plantas escolhidas ao acaso, as quais foram divididas e analisadas em três partes: superior, média e inferior, sendo o valor final da severidade calculado a partir da média dos valores das três partes e das sete plantas. Para quantificar a severidade do oídio foi utilizada a escala de Mattiazzi (2003) e para a ferrugem a escala de Godoy, Koga e Canteri (2006). A Figura 1 apresenta um exemplo de escala diagramática da ferrugem na soja.

Figura 1 – Escala diagramática para quantificação da severidade da ferrugem na soja.



Fonte: Godoy, Koga e Canteri (2006).

Obtenção das imagens por meio da RPA

As imagens foram obtidas por meio da RPA modelo eBee (SenseFly, Suíça), a qual é equipada com sensores de altitude, vento, acelerômetro, giroscópio, bússola, GPS, entre outros, tendo autonomia de voo de aproximadamente 40 min (Figura 2). As câmeras acopladas a RPA, utilizadas neste estudo, foram Canon S110 NIR (12 megapixels) e Sony *Cybershot* WX RGB (18.2 megapixels).

Figura 2 - RPA modelo eBee utilizada no estudo.

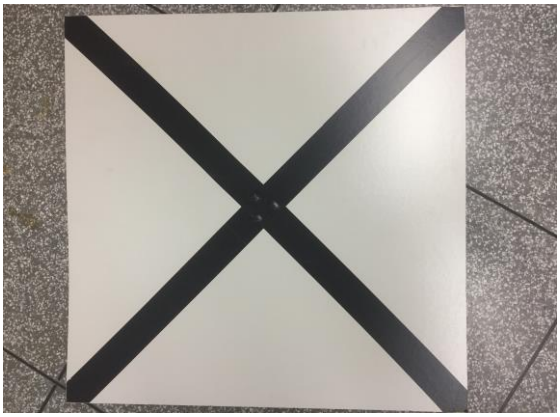


Fonte: http://www.angeleyesuav.com.

A geração e acompanhamento da execução do plano de voo foi feita com o software eMotion 2 (SenseFly, Suíça). O plano de voo consiste em demarcar a área a ser sobrevoada, assim como definir a resolução espacial, área de pouso, entre outras configurações. As resoluções espaciais utilizadas foram 3,4 cm/pixel (em média 97 e 120 metros de altura para a câmera NIR e RGB, respectivamente) e 15 cm/pixel (em média 420 e 530 metros de altura para a câmera NIR e RGB, respectivamente). Foram efetuados quatro voos semanais, entre as 11h e 14h, sendo escolhido o dia de semana que apresentasse condições climáticas favoráveis.

Antes de cada voo, foram posicionados cinco pontos de controle no terreno (GCP – do inglês, *Ground Control Point*), medindo 60x60cm (Figura 3), dos quais foram obtidas as coordenadas geográficas do ponto central, para correção do georreferenciamento dos mapas de refletância, utilizando o receptor GNSS (Sistema de navegação global por satélite) Trimble Juno SC (Trimble, USA).

Figura 3 – GCP utilizado para corrigir as coordenadas geográficas dos mapas.



Fonte: Os Autores (2017).

Processamento das imagens e extração dos dados de refletância espectral

Para análise neste trabalho, foram processadas as imagens dos voos realizados no dia 21/02/2017, data em que a soja mais antiga apresenta maior índice de severidade que as demais coletas. Essa data representa 120 dias após o plantio das áreas A3 e A4 e 70 dias após o plantio das áreas B3 e B4. Vale ressaltar que na data da coleta das imagens, o dia estava ensolarado, sem a presença de nuvens, ou seja, em condições ideais de voo.

Após a coleta das imagens, realizou-se o pré-processamento das mesmas no software eMotion 2. Esta etapa consiste em atribuir as informações do voo nos metadados de cada imagem. Dentre essas informações pode-se destacar: latitude, longitude, altitude, resolução da imagem, modelo da câmera e modelo da aeronave.

Em seguida, foram gerados os mapas de refletância de cada voo utilizando o software Pix4D Mapper (Pix4D SA), que também realiza automaticamente as correções geométrica e radiométrica, necessárias às imagens. Devido às distorções provocadas na aquisição das imagens, o software QGIS (Open Source) foi utilizado para correção do georreferenciamento de cada mapa de refletância, com base nas coordenadas dos GCPs.

Utilizou-se também o software QGIS para extração dos dados dos mapas de refletância gerados, como mostrado na Figura 4. Os pontos em vermelho representam os seis GCPs, eos pontos e linhas em amarelo representam as divisões dos blocos dentro de cada parcela. Os retângulos verdes delimitam a área para extração dos dados da imagem em cada tratamento, na qual foram definidos 200 pontos aleatórios (representando aproximadamente 10% de sua área total), sendo que esses mesmos pontos foram usados para extrair os valores referentes as bandas espectrais das imagens NIR (infravermelho próximo, verde, vermelho) bem como das imagens RGB (vermelho, verde e azul). Associado aos dados de refletância

espectral, foram atribuídos para esses pontos o respectivo índice de severidade das doenças ferrugem e oídio.

Figura 4 - Mapa de Refletância da área de estudo do dia 21/02/2017 gerado a partir das imagens da câmera RGB na resolução de 3,4 cm/pixel.



Fonte: Os Autores (2017).

Mineração de Dados

As bases foram submetidas a testes estatísticos na ferramenta R (R Core Team, 2014), versão 3.3.3. Em cada base, realizou-se a análise de regressão a fim de estimar os índices de severidade da ferrugem e oídio. O algoritmo de regressão utilizado foi a Máquina de Vetores de Suporte para Regressão (SVR – do inglês, *Support Vector Regression*) (VAPNIK et al., 1997), opção eps-regression (tipo de SVM para problemas de regressão usando o parâmetro *epsilon*) com o kernel *Gaussian RBF* (Função de Base Radial Gaussina) (SMOLA; SCHÖLKOPF, 2004), presente no pacote e.1071 (DIMITRIADOU et. Al. 2011), método de Máquina de Vetor de Suporte (SVM – do inglês, *Support Vector Machine*) (VAPNIK, 1963). Os parâmetros *cost, gamma e épsilon,* utilizados na SVR, foram definidos com os valores 1, 0,33 e 0,1, respectivamente. Para validação dos modelos de regressão utilizou-se o método de validação cruzada, no qual o conjunto de testes foi dividido em dez partes iguais (*folds*) e a predição foi aplicada a cada parte separadamente.

RESULTADOS E DISCUSSÃO

Os índices de correlação obtidos com o algoritmo SVR para a estimativa de severidade das doenças, considerando as diferentes câmeras e resoluções espaciais, são mostrados na Tabela 1. Para ambas as doenças, as melhores correlações foram obtidas com os mapas de refletância gerados a partir das imagens obtidas com a câmera RGB. Para estimativa da severidade da ferrugem asiática, os melhores resultados foram obtidos com essa câmera e a resolução

espacial de 15 cm/pixel (r = 0,82). Já para o oídio, o melhor resultado de correlação foi igual (r = 0,73) para ambas as resoluções, também com a câmera RGB. Isso mostra que imagens com menos detalhes, ou seja, com resolução espacial mais baixa (15 cm/pixel) apresentam resultado de correlação igual ou melhor do que a resolução maior (3,4 cm/pixel) na estimativa de severidade do oídio e da ferrugem na soja.

Tabela 1 – Resultados obtidos com o algoritmo SVR para as diferentes doenças, considerando as diferentes câmeras e resoluções espaciais utilizadas.

Atributo Meta	Índice de Correlação/RMSE							
	RGB 3.4 cm/pixel	RMSE	RGB 15 cm/pixel	RMSE	NIR 3.4 cm/pixel	RMSE	NIR 15 cm/pixel	RMSE
Oídio	0.73	12.02	0.73	12.58	0.54	10	0.62	10.85
Ferrugem	0.76	8.91	0.82	8.76	0.63	7.86	0.70	8.28

RMSE – Raiz Quadrada do Erro Quadrático Médio

Garcia-Ruiz et al. (2013) realizaram um estudo comparando o impacto da resolução espacial das imagens na detecção da doença Greening (*Huanglongbing*/HLB), que ataca plantas cítricas. Foram comparadas as resoluções de 50 cm/pixel e 5,45 cm/pixel. As imagens de 50 cm/pixel foram adquiridas utilizando um avião equipado com uma câmera hiperespectral. Já as imagens de 5,45 cm/pixel, foram obtidas através de uma RPA equipada com uma câmera multiespectral. Os autores concluíram que os conjuntos de dados baseados na RPA produziram uma melhor precisão na classificação (67-85%) e menores falsos negativos (7-32%) do que os conjuntos de dados correspondentes baseados no avião (61-74% e 28-45%, respectivamente). O fato de em tal estudo, a resolução mais alta (5,45 cm/pixel) ter apresentado melhores resultados, pode ser justificado por ter sido comparado com uma resolução praticamente dez vezes menor (50 cm/pixel). Salienta-se também, que a diferença entre as resoluções espaciais do trabalho de Garcia-Ruiz et al. (2013) é significativamente maior (5,45 e 50 cm/pixel) que as deste trabalho (3,4 e 15 cm/pixel) e que as imagens não foram capturadas com o mesmo tipo de equipamento.

O resultado satisfatório obtido neste trabalho está de acordo com Lacerda (2017), que obteve coeficiente de correlação de até 0,96 na estimativa da ferrugem na soja utilizando imagens obtidas por RPA com resolução espacial de 3,4 cm/pixel. Já neste trabalho, a maior correlação obtida esta mesma resolução foi de 0,76 na estimativa da ferrugem. Esta diferença pode ter sido ocasionada por diversos fatores, como o clima, época de semeadura e estádio fenológico da planta, pois correspondem a safras diferentes e tais variáveis não foram analisadas nos modelos de regressão. Vale ressaltar também, que Lacerda (2017) realizou a

predição utilizando o algoritmo SMOREG, do software WEKA, enquanto neste trabalho utilizou-se SVR no software R.

Vale ressaltar que diversos fatores influenciam na precisão dos dados. A luminosidade do dia na hora em que as imagens é um desses fatores, pois sombras causam manchas escuras nas imagens alterando o valor do pixel. O vento também é um fator a ser analisado, pois pode aumentar o erro das coordenadas geográficas da imagem, se este estiver muito intenso.

CONCLUSÕES

Este trabalho mostrou que um maior nível de detalhamento dos dados de imagens não significa necessariamente que serão obtidos melhores resultados na estimativa de doenças. Os mapas gerados a partir das imagens com resolução de 15 cm/pixel proporcionaram a estimativa da ferrugem e do oídio na soja com correlações melhores ou iguais as obtidas com mapas de resolução de 3,4 cm/pixel.

A câmera RGB se mostrou ser mais eficiente para a estimativa de severidade tanto do oídio com da ferrugem asiática. Das doenças estudadas a ferrugem asiática respondeu melhor à análise por imagens, tendo apresentado a melhor correlação de estimativa de severidade ($r = 0,82$).

Como trabalhos futuros, outras resoluções serão comparadas em novos experimentos, a fim de estabelecer um padrão de resolução espacial que possa ser adotado para a estimativa de doenças na soja.

AGRADECIMENTOS

A Capes e a Fundação Araucária, pelo apoio financeiro.

REFERÊNCIAS

CALDERÓN, R. et al. High-resolution airborne hyperspectral and thermal imagery for early detection of Verticillium wilt of olive using fluorescence, temperature and narrow-band spectral indices. Remote Sensing of Environment, v. 139, p. 231-245, 2013.

DI GENNARO, S. F. et al. Unmanned Aerial Vehicle (UAV)-based remote sensing to monitor grapevine leaf stripe disease within a vineyard affected by esca complex. Phytopathologia Mediterranea, v. 55, n. 2, p. 262-275, 2016.

DIMITRIADOU, E. et al. e1071: Misc Functions of the Department of Statistics (e1071), TU Wien. R package version 1.5-27. 2011.

GARCIA-RUIZ, F. et al. Comparison of two aerial imaging platforms for identification of Huanglongbing-infected citrus trees. Computers and Electronics in Agriculture, v. 91, p. 106-115, 2013.

GERKE, T. Mineração de dados de imagens obtidas com aeronave remotamente pilotada para estimativa de produtividade do trigo. Dissertação (Mestrado em Computação Aplicada) – Setor de Ciências Agrárias e de Tecnologia, Universidade Estadual de Ponta Grossa, Ponta Grossa, 2017.

GODOY, C.V.; KOGA, L. J.; CANTERI, M.G. Diagrammatic scale for assessment of soybean rust severity. Fitopatologia Brasileira, Brasília, DF, v. 31, n. 1, p. 63-68, 2006.

HIKISHIMA, M. et al. Quantificação de danos e relações entre severidade, medidas de refletância e produtividade no patossistema ferrugem asiática da soja. Tropical Plant Pathology, v. 35, n. 2, p. 96-103, 2010.

LACERDA, V. S. Estimativa do índice de severidade de ferrugem asiática na cultura da soja por meio de imagens obtidas com aeronave remotamente pilotada. Dissertação (Mestrado em Computação Aplicada) – Setor de Ciências Agrárias e de Tecnologia, Universidade Estadual de Ponta Grossa, Ponta Grossa, 2017.

MAHLEIN, A. et al. Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases. Plant methods, v. 8, n. 1, p. 3, 2012.

MAHLEIN, A. Plant disease detection by imaging sensors–parallels and specific demands for precision agriculture and plant phenotyping. Plant Disease, v. 100, n. 2, p. 241-251, 2016.

MATTIAZZI, P. Escala diagramática do oídio da soja. Efeito do oídio na produção e duração da área foliar sadia da soja, 2003. 63 f. Dissertação (Mestrado em Agronomia), Escola Superior de Agricultura “Luiz de Queiroz” Piracicaba, 2003.

R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria, 2014. Disponível em: <http://www.r-project.org>.

SILVA, E. A. et al. Perspectivas no manejo e controle localizado de doenças das culturas: uma abordagem no manejo do nematóide do cisto da soja. Viçosa, Editora UFV. 453p, 2000.

SMOLA, A. J.; SCHOLKOPF, B. A tutorial on support vector regression. Statistics and computing, Springer, v. 14, n. 3, p. 199–222, 2004.

VAPNIK, V. Pattern recognition using generalized portrait method. Automation and remote control, v. 24, p. 774–780, 1963.

VAPNIK, V. et al. Support vector method for function approximation, regression estimation, and signal processing. Advances in neural information processing systems, Morgan Kaufmann Publishers, p. 281–287, 1997.

Efeito da autocorrelação espacial na avaliação de modelos empíricos de produtividade na cana-de-açúcar

Matheus Agostini Ferraciolli[1], *Felipe Ferreira Bocca*[1],*Luiz Henrique Antunes Rodrigues*[1]

[1]Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas,
Campinas, São Paulo, Brasil
matheus.ferraciolli@feagri.unicamp.br, felipe.bocca@feagri.unicamp.br,
lique@feagri.unicamp.br.

RESUMO

Técnicas de aprendizado de máquina podem ser utilizadas para modelagem de fenômenos na agricultura. Entretanto, nestes dados é característica a presença da autocorrelação espacial, que é uma medida da dependência entre observações em um espaço geográfico. Isso viola a premissa de independência dos dados para o desenvolvimento de modelos. Uma alternativa é agrupar espacialmente os registros em blocos e utilizar estes blocos na separação de conjuntos de treino e teste, reduzindo as chances de que registros similares sejam usados em ambos os conjuntos. Neste trabalho, utilizamos técnicas de aprendizado de máquina para modelar a produtividade de cana-de-açúcar nas áreas de três usinas. A abordagem clássica, em que cada registro é considerado independente e a validação cruzada espacializada k-fold, em que os registros são divididos em blocos foram utilizadas nas etapas de treino, ajuste de hiperparâmetros e teste dos modelos. O desempenho de ambas as abordagens foi comparado calculando o erro médio absoluto de previsão em um conjunto de validação de cada usina. Apesar da separação diferenciada dos dados, a maioria dos modelos apresentou os mesmos parâmetros finais. Além disso, embora a abordagem clássica tenha resultado em menor erro médio do que espacializada, os desempenhos de ambas não foram satisfatórios para estimar o erro no conjunto de validação.

PALAVRAS-CHAVE: Modelagem de culturas, Análise de dados, Aprendizado de máquinas.

ABSTRACT

Machine learning techniques can be used to model phenomena in agriculture. However, agricultural data usually presents spatial autocorrelation, which is a measure of the dependence between observations in a geographic space, and violates the premise of data independence for model development. An alternative is to spatially group the records into blocks and use these blocks in separating training and test sets, reducing the chances that similar records will be used

in both sets. In this work, we used machine learning techniques to model the yield of sugarcane in three mill areas. The classical approach, in which each record is considered independent and the k-fold spatial cross-validation, in which the records are divided into blocks, were used in the steps of model training, hyperparameter adjustment, and model testing. The performance of both approaches was compared by calculating the mean absolute error of prediction in a validation set for each mill area. Despite the differentiated data separation, most models presented the same final parameters. In addition, although the classical approach resulted in a lower mean absolute error than the spatial approach, the both performances were not satisfactory for estimating the error in the validation set.
KEYWORDS: Crop Modeling, Data analysis, Machine Learning.

INTRODUÇÃO

O aumento da disponibilidade de dados e de recursos computacionais relacionados à produção agrícola tem permitido o uso de métodos mais robustos para modelagem de fenômenos na área. Para isso, são usadas técnicas de regressão capazes de lidar com respostas não-lineares por meio da definição de thresholds e da consideração da interação e da autocorrelação entre atributos. Entre as técnicas utilizadas estão Random Forests (RF), Support Vector Machine (SVM) e Boosted Regression Trees (BRT), com aplicações em diferentes áreas: na medicina, para diagnóstico ou previsão de ocorrência de doenças (BRENNING; LAUSEN, 2008); na ecologia, para modelagem da distribuição de espécies (DORMANN et al., 2007) e para quantificação de biomassa por imagens de satélites (ZANDLER; BRENNING; SAMIMI, 2015) e na geologia, para determinação de locais susceptíveis a desmoronamentos (BRENNING; LAUSEN, 2008).

Na agricultura, modelos empíricos de previsão podem auxiliar em diversas tomadas de decisão, com benefícios para toda a cadeia de produção. Devido às relações entre o resultado agrícola e o meio ambiente, diversas restrições sobre os procedimentos de validação – e.g. as relacionadas à autocorrelação – também se aplicam para validação de modelos na agricultura (ROBERTS et al., 2017).

Autocorrelação significa que uma variável está relacionada com ela mesma. A autocorrelação espacial é a correlação entre valores de uma única variável estritamente atribuível à proximidade destes valores em um espaço geográfico (GRIFFITH, 2013), ou seja, localizações adjacentes apresentam características parecidas, sendo mais comumente observada em mapas na forma de manchas ou gradientes (HOEF; CRESSIE; GLENN-LEWIN, 1993).

A presença de autocorrelação viola a premissa de independência dos registros. Neste caso, a estimativa de parâmetros é enviesada (DORMANN et al., 2007). Em geral, dados similares acabam nos conjuntos de treino e teste do modelo, retornando uma estimativa demasiada otimista do erro de predição do modelo em questão (RUSS; BRENNING, 2010), resultando em overfitting do modelo aos dados de treino.

De forma geral, a estrutura dos dados e as características do objeto de modelagem pautam

as modificações que devem ser feitas nos procedimentos de validação (ROBERTS et al., 2017), considera-se que o efeito da autocorrelação espacial é o predominante. Para este caso, o procedimento recomendado é a estratificação por blocos. No estudo de previsão de colheita, Ruß e Brenning (2010) dividiram o conjunto de dados de dois talhões em clusters com aproximadamente o mesmo número de dados. Estes clusters foram utilizados então no método validação cruzada k-fold espacializada (Spatial Cross Validation, SCV), com conjuntos de treino e teste pertencentes a diferentes clusters, garantindo a diminuição da autocorrelação espacial entre os dados. Os autores observaram que também poderia ser implementado um método de amostragem que levasse em conta a localização espacial dos clusters, para evitar a presença de pontos adjacentes às bordas.

Na produção de cana-de-açúcar, a disponibilidade de dados de colheita de talhões torna possível a aplicação de técnicas de mineração de dados para estudar os efeitos da autocorrelação espacial em diferentes etapas do processo de modelagem da produtividade. O objetivo deste trabalho é, então, a identificação de um método para lidar adequadamente com estas características dos dados. Foram utilizados dados de diferentes usinas, com áreas e configurações de talhões variadas. O impacto da autocorrelação espacial (AE) foi analisado nas etapas de separação dos dados em folds durante o ajuste de parâmetros dos modelos e de separação dos conjuntos de treino e teste na avaliação da performance do modelo obtido, com a proposta de obter-se um processo válido para lidar com dados onde exista AE.

Devido à semelhança entre as análises, o método utilizado por Ruß e Brenning (2010) para avaliar a estimativa de produtividade de trigo nos talhões foi adaptado para este contexto, com os dados da produção de usinas de cana-de-açúcar.

MATERIAL E MÉTODOS

A avaliação do modelo em um cenário com AE dos dados envolve etapas similares à da modelagem tradicional. Os dados de colheita foram separados em conjuntos de validação, treino e teste para desenvolvimento do modelo com dois métodos diferentes. Um dos métodos considera que os dados são independentes e o outro leva em conta a AE entre dados. Posteriormente, os modelos gerados pelos diferentes métodos foram comparados utilizando conjunto de validação. Foram utilizados dados de produção de três usinas diferentes.

Os métodos utilizados em cada usina foram padronizados, com aproximadamente a mesma proporção da área total dividida em conjuntos de treino, teste e validação. Em cada usina, foi realizada a identificação, por inspeção no SIG, de blocos de talhões com distância mínima de 3 km para as outras áreas, para minimizar o efeito da autocorrelação espacial entre blocos. Estes blocos compuseram o conjunto de validação. Os demais talhões foram divididos novamente e três-quartos foram destinados para treino e um-quarto, para teste dos modelos.

A divisão entre os conjuntos de treino e de teste, porém, não é completamente aleatória. O método que lida com AE é o mesmo utilizado por Ruß e Brenning (2010). O algoritmo k-

means foi utilizado para dividir os dados em k clusters ou grupos, formando um mosaico que considera, dentre todos os atributos dos pontos amostrados, apenas a relação espacial. O número de clusters foi escolhido de forma que os blocos tivessem, aproximadamente, 5 observações. Após essa divisão, os k grupos formados foram utilizados no processo de SCV, com a vantagem de os conjuntos de treino dos dados possuírem pouca AE com o conjunto de teste.

As técnicas de regressão utilizadas foram Support Vector Machine (SVM) e Random Forest (RF). Frequentemente, algoritmos de aprendizagem envolvem um processo de otimização de uma função baseada em parâmetros obtidos com a análise dos dados, porém, o próprio algoritmo possui parâmetros ajustáveis, chamados hiperparâmetros. Para determinação dos hiperparâmetros dos diferentes algoritmos, foi utilizada a busca aleatória, a fim de evitar a explosão combinatória no ajuste de parâmetros, além de obter resultados superiores aos da busca de parâmetros em grade (BERGSTRA; BENGIO, 2012). Os intervalos de busca de cada técnica constam na Tabela 1.

Tabela 1: Técnicas e parâmetros ajustados por busca aleatória. Para cada parâmetro são apresentados os limites inferior e superior para amostragem sob distribuição uniforme.

Technique	**Parameter**	**Inferior Limit**	**Superior Limit**
Support Vector Machine	Cost	1	1024
(MEYER; BISCHL; WEIHS, 2014)	Gamma	2^{-5}	2^{-1}
	Epsilon	0.05	0.3
Random Forest	Number of trees	300	1000
(LIAW; WIENER et al., 2002)	% of attributes for division	20	100
	Node size	20	100

Após o ajuste de hiperparâmetros e desenvolvimento dos modelos, as duas técnicas tiveram seus desempenhos comparados em duas etapas. A primeira validou se a divisão treino e teste foi efetiva na estimativa do erro de generalização, comparando o erro médio absoluto no conjunto de teste e de validação. A segunda comparou o erro dos modelos no conjunto de validação, com o objetivo de testar o impacto das diferentes formas de divisão no modelo obtido. Os modelos foram comparados utilizando o erro médio absoluto.

Os dados correspondem aos dados de produção das usinas da Odebrecht Agroindustrial: UCR – Usina Costa Rica (Costa Rica – MS), URC – Usina Rio Claro (Caçu - GO) e USL – Usina Santa Luzia (Nova Alvorada do Sul – MS) e se referem às colheitas nos anos de 2012, 2013, 2014 e 2015. Foram utilizados dados relativos a aplicações de insumos, classificação e propriedades do solo e condições meteorológicas durante o ciclo de produção, totalizando 37 atributos. De forma geral, o procedimento de criação de atributos foi realizado da mesma forma que em trabalhos anteriores de modelagem de cana (BOCCA; RODRIGUES, 2016).

RESULTADOS E DISCUSSÃO

Para o desenvolvimento dos modelos de SVM nas diferentes usinas, apenas na URC foi obtido um modelo diferente em função do uso de SCV (Tabela 2). Para o desenvolvimento dos modelos de RF, nenhum modelo diferente foi obtido em função do uso de SCV (Tabela 3) .

Tabela 2: Técnicas e parâmetros ajustados por busca aleatória. Para cada parâmetro são apresentados os limites inferior e superior para amostragem sob distribuição uniforme.

Usina	**Tratamento**	**Custo**	**Gamma**	**Epsilon**	**MAE [t ha-1]**		
					CV Treino	**Teste**	**Validação**
UCR	SCV	10.13	0.06	0.06	9.37	8.94	18.75
UCR	Conv.	10.13	0.06	0.06	7.5	6.82	18.75
URC	SCV	4.74	0.08	0.1	8.48	7.32	13.88
URC	Conv.	31.24	0.16	0.06	6.86	6.68	14.55
USL	SCV	5.15	0.09	0.07	6.22	6.14	17.42
USL	Conv.	5.15	0.09	0.07	5	4.23	17.42

Tabela 3: Técnicas e parâmetros ajustados por busca aleatória. Para cada parâmetro são apresentados os limites inferior e superior para amostragem sob distribuição uniforme.

Usina	**Tratamento**	**Nº de árvores**	**Atributos**	**Tamanho do nó**	**MAE [t ha-1]**		
			para split	**para split**	**CV Treino**	**Teste**	**Validação**
UCR	SCV	10.13	0.06	0.06	9.37	8.94	18.75
UCR	Conv.	10.13	0.06	0.06	7.5	6.82	18.75
URC	SCV	4.74	0.08	0.1	8.48	7.32	13.88
URC	Conv.	31.24	0.16	0.06	6.86	6.68	14.55
USL	SCV	5.15	0.09	0.07	6.22	6.14	17.42
USL	Conv.	5.15	0.09	0.07	5	4.23	17.42

No geral, os erros estimados pelo método convencional foram menores que os erros estimados pelo SCV. Este resultado também foi observado por Ruß e Brenning (2010) em um estudo de previsão de colheita de trigo realizado na Alemanha.

Os erros de validação, porém, se mostraram maiores que o esperado, ou seja, discrepantes, quando comparados aos da avaliação no teste. Uma possível explicação para isso é de que em divisões de conjuntos de forma espacial, é possível que nem todos os intervalos dos valores das variáveis preditoras estejam contemplados no conjunto de treino (ROBERTS et al., 2017). Isso faz com que os modelos tenham de realizar extrapolações na predição, o que possivelmente acarretou maiores erros. Outra possibilidade é de que o agrupamento com um número pequeno de registros não tenha sido efetivo em atenuar o efeito da AE entre talhões vizinhos.

Ruß e Brenning (2010) indicaram que o número de agrupamentos pauta um trade-off entre a precisão da predição e a validade estatística do modelo. Nesse contexto, o valor pequeno pode ter comprometido a validade dos modelos criados. Cabe destacar que a discrepância observada entre as estratégias de validação não foi tão grande quanto a discrepância observada em relação aos dados de validação. Ruß e Brenning (2010) também afirmaram que a determinação no número de grupos deve ser feita por uma heurística. Nenhum dos trabalhos consultados indicou de forma específica como obter a estrutura de blocos para validação. Em termos do volume

de dados utilizados para treino/teste/validação, além da diferença de amplitudes entre preditores no treino/teste e validação, o volume alocado para validação pode ter sido muito grande. Dessa forma, o modelo pode não ter sido capaz de aprender com condições presentes apenas no conjunto de validação.

CONCLUSÕES

Embora exista larga evidência de que a autocorrelação espacial afeta as estratégias de validação, a estratégia de agrupamento de registros utilizando o algoritmo k-means empregada neste trabalho não se mostrou efetiva. Os modelos desenvolvidos com a incorporação da informação espacial na estratégia de validação não diferiram dos modelos obtidos com a estratégia convencional de validação. Isso pode ser atribuído à criação de agrupamentos com poucos registros. Foi observado também uma grande diferença de performance entre os erros estimados no conjunto de teste e no conjunto de validação. Isso pode ser atribuído a diferenças nos intervalos dos atributos preditores no conjunto de validação. Nessa condição, os modelos têm que extrapolar para realizar predições, incorrendo em maiores erros. Uma redução na fração de registros alocados para validação poderia diminuir esse efeito. Trabalhos futuros devem considerar a compatibilidade de intervalos entre os blocos utilizados nas avaliações. Esforços futuros serão direcionados para avaliar o efeito do tamanho dos agrupamentos utilizados na capacidade da estratificação por blocos atenuar o efeito da autocorrelação espacial na validação de modelos.

AGRADECIMENTOS

Agradecemos à Odebrecht Agroindustrial pela cessão dos dados, ao PIBIC/CNPq pela concessão da bolsa do primeiro autor e à CAPES e ao CNPq pela concessão da bolsa do segundo autor.

REFERÊNCIAS

BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, v. 13, n. Feb, p. 281–305, 2012.

BOCCA, F. F.; RODRIGUES, L. H. A. The effect of tuning, feature engineering, and feature selection in data mining applied to rainfed sugarcane yield modelling. *Computers and Electronics in Agriculture*, Elsevier, v. 128, p. 67–76, 2016.

BRENNING, A.; LAUSEN, B. Estimating error rates in the classification of paired organs. *Statistics in medicine*, Wiley Online Library, v. 27, n. 22, p. 4515–4531, 2008.

DORMANN, C. F. et al. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, Wiley Online Library, v. 30, n. 5, p. 609–628, 2007.

GRIFFITH, D. A. *Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization*. [S.l.]: Springer Science & Business Media, 2013.

HOEF, J. M. V.; CRESSIE, N. A.; GLENN-LEWIN, D. C. Spatial models for spatial statistics: some unification. *Journal of Vegetation Science*, Wiley Online Library, v. 4, n. 4, p. 441–452, 1993.

LIAW, A.; WIENER, M. et al. Classification and regression by randomforest. *R news*, v. 2, n. 3, p. 18–22, 2002.

MEYER, O.; BISCHL, B.; WEIHS, C. Support vector machines on large data sets: Simple parallel approaches. In: *Data Analysis, Machine Learning and Knowledge Discovery*. [S.l.]: Springer, 2014. p. 87–95.

ROBERTS, D. R. et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, Wiley Online Library, 2017.

RUSS, G.; BRENNING, A. Data mining in precision agriculture: management of spatial information. *Computational intelligence for knowledge-based systems design*, Springer, p. 350–359, 2010.

ZANDLER, H.; BRENNING, A.; SAMIMI, C. Quantifying dwarf shrub biomass in an arid environment: Comparing empirical methods in a high dimensional setting. *Remote Sensing of Environment*, Elsevier, v. 158, p. 140–155, 2015.

Melhorias no processamento dos dados de um espectrômetro de baixo custo para uso agrícola

Rafaella Pironato Amaro[1], *Felipe Ferreira Bocca*[1], *Diego Della Justina*[1], *Luiz Henrique Antunes Rodrigues*[1]

[1]Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas, Campinas, São Paulo, Brasil, rafaella.amaro@feagri.unicamp.br, felipe.bocca@feagri.unicamp.br, diego.justina@feagri.unicamp.br, lique@feagri.unicamp.br.

RESUMO

Dado o potencial de sensores de baixo custo para a agricultura, foi avaliado um espectrômetro desenvolvido a partir de componentes de baixo custo quanto a suas características estáticas – exatidão, precisão e reprodutibilidade – em folhas de alface crespa e couve-manteiga. Os resultados obtidos com o sensor de baixo custo foram comparados com os obtidos utilizando um FieldSpec 4 após quatro tipos de processamento. Em termos de precisão, os sensores de baixo custo apresentam desempenho duas ordens de magnitude pior que o sensor de referência, independentemente do tipo de processamento. Nas análises de reprodutibilidade e exatidão o erro foi duas vezes maior que o obtido na análise de precisão. O comportamento geral das métricas estudadas em relação aos diferentes tratamentos foi similar, não havendo tendência clara de melhoria. Ao se analisar o erro ao longo dos diferentes comprimentos de onda, as regiões de menor erro correspondem aos picos de sensibilidade dos canais. Estas regiões de menor erro podem orientar a investigação dos comprimentos de onda com maior sensibilidade e, portanto, gerar potenciais áreas de aplicação para o sensor.

PALAVRAS-CHAVE: Public Lab, Sensores open hardware, Validação de sensores.

ABSTRACT

Given the potential of inexpensive sensors in agriculture, a spectrometer developed from low cost components was evaluated regarding its static characteristics - accuracy, precision and reproducibility - in leaves of crisp lettuce and kale. The results obtained with the low-cost sensor were compared with those obtained using a FieldSpec 4 after being processed in four different ways. In terms of accuracy, low cost sensors performances were two orders of

magnitude worse than the reference sensor, regardless the processing method used. In the analyses of reproducibility and accuracy the error was two-fold greater than that obtained in the precision analysis. The general behavior of the metrics studied regarding the different treatments was similar, with no clear trend of improvement. By analyzing the error along the different wavelengths, the regions of smaller error correspond to the sensitivity peaks of the channels. These smaller errors regions allow for the investigation of the wavelengths with greater sensitivity, leading to potential areas of applications for the sensor.
KEYWORDS: Public Lab, Open hardware sensors, Sensor validation.

INTRODUÇÃO

No monitoramento de estresse nos tecidos vegetais são utilizadas tecnologias multiespectrais e hiperespectrais, em que comprimentos de onda na região do visível e do infravermelho próximo (VIS/NIR) são comumente exploradas, devido a fatores como a água, nutrientes e temperatura (COZZOLINO, 2014). As regiões VIS/NIR também são ricas em informações sobre a composição química e qualidade dos alimentos e produtos agrícolas (PARK; LU, 2015).

A quantidade de dados fornecida por sensores espectrais é um dos fatores que leva à sua ampla utilização e visando aumentar a qualidade das medidas e facilitar as medições, cada vez mais novos sensores surgem. Já é sabido, porém, que quando os produtores agrícolas não possuem recursos para investir em tecnologias para maior produção de alimentos, a baixa produtividade se mantém (GODFRAY et al., 2010). Além disso, uma vez que na produção agrícola, o preço dos produtos finais não é estabelecido pelo produtor, a coleta de informações não pode ser cara, pois inviabilizaria o resultado econômico do produto final. Então, se por um lado, o uso de sensores pode aumentar o detalhamento na coleta de dados, por outro, o aumento do custo tornaria seu uso inviável. Considerando que a obtenção de informações mais detalhadas pelos métodos convencionais pode ter custos elevados, já que as áreas agrícolas possuem grandes extensões, a caracterização da variabilidade da área agrícola demandaria um sensor de qualidade e, simultaneamente, de baixo custo (ZHANG; KOVACS, 2012).

O desenvolvimento de espectrômetros de baixo custo está intimamente ligado ao barateamento das tecnologias de aquisição de imagens digitais. Dessa forma, o fácil acesso a celulares e *webcams* é um fator facilitador para desenvolvimento de tecnologias de espectrometria de baixo custo. Por outro lado, características desejáveis das câmeras digitais para uso convencional se mostram como obstáculos para seu uso em espectrômetros de baixo

custo (SCHEELINE, 2016). De forma geral, um espectrômetro de baixo custo é composto por uma estrutura que reúne um colimador, um meio de difração e uma câmera digital. Elementos que podem ser adicionados são: fontes de luz, lentes para foco e suportes para amostras. Um espectrômetro de baixo custo popular foi concebido e é distribuído com licença de open-hardware (CERN) pela Public Lab[1]. O espectrômetro utiliza um colimador de acetato, uma câmara escura feita com papel cartonado, um fragmento de DVD como meio de difração e uma *webcam* como detector. Um kit para montagem do equipamento é vendido por U$ 47.

Enquanto diversas melhorias em um espectrômetro de baixo custo podem ser feitas com melhorias nos componentes utilizados (SCHEELINE, 2016), outro aspecto que pode ser abordado são os procedimentos de tratamento dos dados. Considerando os procedimentos utilizados na obtenção dos dados a partir da plataforma disponibilizada pela Public Lab, o objetivo deste trabalho foi avaliar o desempenho relativo dos espectrômetros da Public Lab em comparação a um espectrômetro de referência, em suas diferentes formas de processamento.

Para isso, diferentes propostas de tratamento dos dados obtidos pelo espectrômetro foram avaliadas quanto à precisão, à exatidão e à reprodutibilidade. Os espectrômetros de baixo custo foram utilizados na medição da resposta espectral de duas hortaliças: alface (*Lactuca sativa* var. *crispa)* e couve (*Brassica oleracea*, grupo Acephala). As medições foram comparadas com um sensor de referência (FieldSpec 4).

MATERIAL E MÉTODOS

Procedimentos gerais

A refletância de 5 folhas de couve e 5 folhas de alface foram medidas utilizando dois espectrômetros: Public Lab (v3 – faixa espectral: 400-900 nm; resolução espectral: 1-5 nm) e um FieldSpec® 4 Standard-Res (faixa espectral: 350-2500 nm; resolução espectral: 3 nm @ 700 nm e 10 nm @ 1400/2100 nm).

Para cada folha (amostra), foram coletadas 5 replicatas de espectro, sendo as replicatas obtidas por meio da rotação da amostra. Foram analisadas a precisão e a reprodutibilidade de dois sensores da Public Lab. Os dois sensores serão denominados P e L.

Na análise de precisão, foram comparados os espectros obtidos entre as replicatas de uma mesma amostra. Na análise de reprodutibilidade, os espectros de cada replicata obtido

[1] https://publiclab.org/notes/warren/09-01-2016/calibrating-your-spectrometer-in-spectral-workbench

por um sensor para uma superfície foram comparados a cada uma das replicatas obtidas pelo outro sensor na mesma superfície.

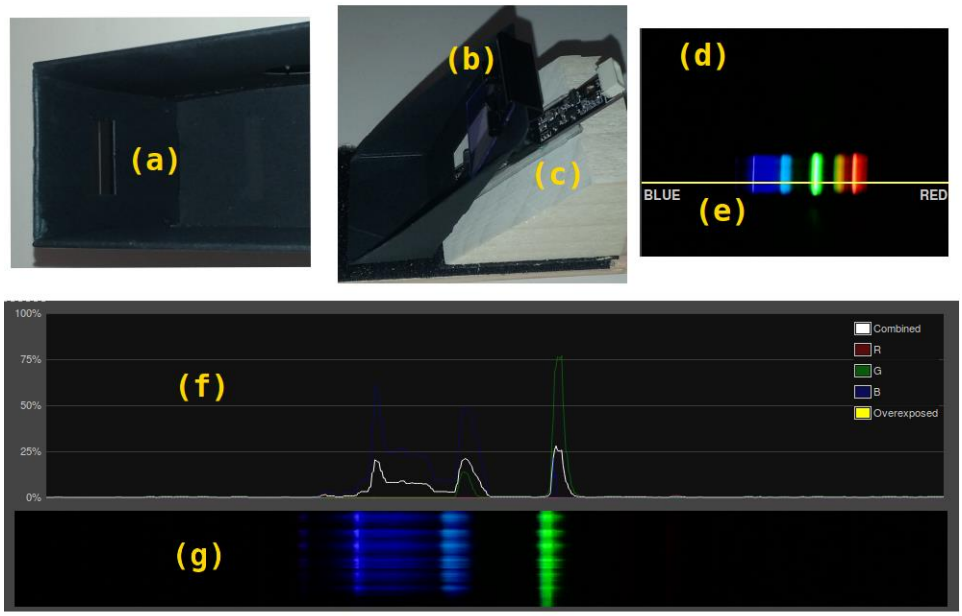
A precisão do FieldSpec (FS) foi utilizada como parâmetro de comparação para a precisão e reprodutibilidade dos espectrômetros da Public Lab. A análise de exatidão foi realizada assumindo o espectro medido pelo FS como valor de referência. Todas as análises de dados utilizaram a linguagem R (R CORE TEAM, 2016).

Espectrômetro Public Lab

Funcionamento geral

Um feixe de luz passa pela entrada identificada na Figura 1a e incide no fragmento de DVD (Figura 1b), sofrendo difração. A *webcam* (Figura 1c) captura uma imagem (exemplificada na Figura 1d), da qual uma linha é selecionada (Figura 1e). Nesta linha, é medida a intensidade luminosa nos diferentes comprimentos de onda através da intensidade de cor medida nos diferentes canais da câmera (escala de 0 a 255) para cada coluna. Por meio da calibração com uma lâmpada fluorescente compacta, é possível estabelecer correspondência entre comprimentos de onda específicos e as colunas da imagem. Para isso, são identificadas em quais colunas ocorrem os picos de emissão do mercúrio, seguindo procedimento estabelecido pela Public Lab. Dessa forma, é possível relacionar as intensidades de radiação a comprimentos de onda específicos. Cabe destacar que a intensidade luminosa não tem significado físico direto. Um exemplo de espectro resultante pode ser visto na Figura 1f e Figura 1g.

Figura 1 – Detalhes do sensor Public Lab.



Fonte: (PUBLIC LAB, 2017).

Aquisição dos dados

A plataforma Spectral Workbench (PUBLIC LAB, 2017) foi utilizada para captura de imagens no modo *offline*. Neste formato, uma imagem no formato PNG é gerada com uma estrutura RGBA, contendo as intensidades luminosas nos três canais da câmera, com 150 pixels de altura e 640 colunas. Isso corresponde a 150 medições de intensidade luminosa para os 640 comprimentos de ondas estabelecidos, que vão de 450 a 630 nm. Utilizando a plataforma *online*, a primeira linha desta imagem é utilizada para compor o espectro medido. Para isso, é utilizada a média entre os três canais (RGB).

Processamento dos dados

- ***Convencional***

Análises preliminares indicaram que utilizar entre 20 e 30 linhas da imagem reduz a variância dos valores medidos, e que não há vantagem em utilizar mais do que 30 linhas. Os dados das imagens PNG foram processados para extrair as 30 primeiras linhas e obter as médias das 30 leituras para cada canal. Esse procedimento é análogo ao processamento realizado na plataforma Spectral Workbench, porém utiliza um número maior de linhas. Dados que foram obtidos por meio deste procedimento serão denominados “originais”.

- ***Obtenção da refletância***

Foram realizadas diversas tentativas de obter a resposta espectral de uma superfície branca de referência. Quando foi utilizada como referência uma placa Spectralon (LABSPHERE, 2017), foi necessário afastar a fonte luminosa para que não houvesse saturação dos canais da câmera. Na distância necessária para não haver saturação da câmera, não era possível captar respostas da superfície folhas utilizadas. Para o estabelecimento de uma intensidade luminosa de referência, foi utilizada uma folha de papel branco, sendo determinada a intensidade de referência para cada sensor da Public Lab. Os espectros das superfícies das plantas obtidos pelo método apresentado no tópico “Aquisição dos dados” foram então divididos pela intensidade de referência obtida com a folha de papel. Dessa forma, o valor passa a representar a fração refletida em relação à superfície de referência. Dados que foram obtidos por meio deste procedimento serão denominados “normalizados”.

- ***Obtenção da refletância com segmentações de canais***

Dado que ao se utilizar a média dos diferentes canais, são misturados dados de regiões de baixa sensibilidade de um canal com dados de alta sensibilidade de outro canal, é possível que

ruído esteja sendo introduzido nas medições. Sendo assim, foi analisada a intensidade luminosa da referência branca para três medições. Quando havia medições de dois canais para um mesmo comprimento de onda, foi identificado o canal de onda com maior resposta. O espectro resultante foi obtido com a composição dos canais de maior resposta. Os dados que receberam este tratamento foram denominados "segmentados". Cabe destacar que este procedimento também foi aplicado para as amostras utilizadas para determinar a normalização. Dados que foram obtidos por meio deste procedimento foram denominados "segmentados-m".

- ***Pós-processamento***

Após obtenção de intensidade luminosa, visando a comparação com os dados obtidos pelo FS, os espectros obtidos com os espectrômetros da Public Lab foram interpolados para coincidir com os comprimentos de ondas da saída do FS. Em função da baixa sensibilidade (menor que 50% da resposta máxima) nos comprimentos de onda inferiores a 450 nm e superiores a 630 nm, regiões nesses comprimentos de onda foram excluídos da análise.

FieldSpec

Antes de se obter os dados com o FS, foi realizado o procedimento de otimização do equipamento e estabelecida a refletância da superfície de referência utilizando uma placa de Spectralon.

Avaliação do desempenho

As comparações foram feitas para cada comprimento de onda, e então agregadas de acordo com as análises realizadas. Para análise de precisão, os valores médios foram obtidos para cada amostra, e então agregados para os pares de Superfície vs Sensor. Para análise de reprodutibilidade, foram comparados os espectros obtidos pelos espectrômetros P e L, sendo os valores médios obtidos para cada permutação, e agregados para os pares de Amostra vs Superfície. Para análise de exatidão, os espectros obtidos pelos espectrômetros P e L foram comparados individualmente para cada replicata com os valores obtidos pelo FS. Os valores foram agregados para os pares de Amostra vs Superfície para cada sensor (P e L).

Avaliações de correlação entre o espectro medido pelos espectrômetros da Public Lab e o espectro medido pelo FS foram realizados para cada par de medidas de comprimento de onda para as diferentes combinações de Amostra vs Replicata. Foi utilizada a correlação de Pearson para esta avaliação.

Nas avaliações de reprodutibilidade e precisão foi utilizada a métrica Diferença Média de Porcentagem Absoluta (MAPD, sigla em inglês *Mean Absolute Percentage Deviation*), calculada pela Equação 1. Na equação, M1 corresponde à medida 1, M2 à medida 2 e n ao número de observações para cada comprimento de onda.

$$MAPD = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{M1 - M2}{\left(\frac{M1 + M2}{2}\right)}\right| \qquad \text{Equação 1}$$

RESULTADOS E DISCUSSÃO

Utilizando a métrica MAPD foram geradas as métricas de avaliação da precisão, exatidão e reprodutibilidade. De forma geral, o MAPD obtido para o FS nas condições de experimento foi de 0,10% para as folhas de alface e 0,14% para as folhas de couve. Para os sensores P e L foi no geral de 60% para as folhas de alface. Nas folhas de couve, o sensor L obteve precisão de 43 a 48%, sendo os menores valores para os tratamentos original e normalizado. O sensor P obteve menores valores de MAPD, próximos a 36%. Isso significa que no geral, uma diferença de aproximadamente 40% no valor medido é esperada entre diferentes medições dos sensores Public Lab em uma mesma superfície.

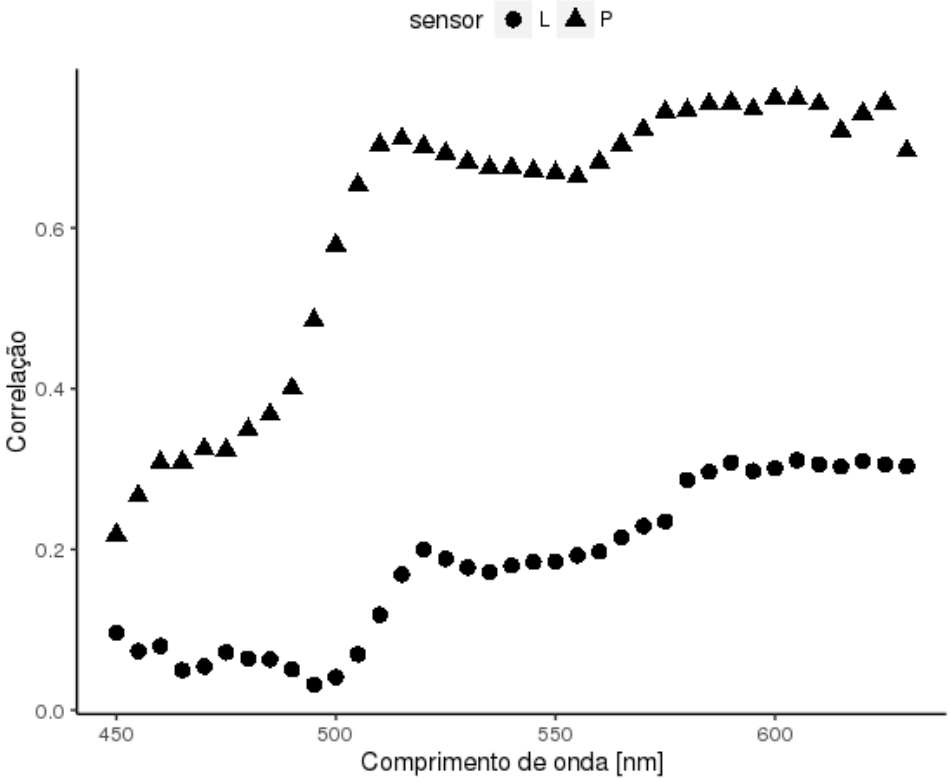
Na análise da reprodutibilidade, foram obtidos valores de aproximadamente 80% de MAPD. Em linhas gerais, isso significa que a diferença entre medir uma superfície com diferentes sensores pode ter o dobro da variação de uma medida com o mesmo sensor. Diferenças entre as medições dos sensores podem ser explicadas por falhas em se manter a mesma montagem dos dois equipamentos e diferenças no fragmento de DVD utilizado.

Na análise de exatidão, foram obtidos valores de 67% (sensor P, análise original para folha de alface) a 92% (sensor L, análise normalizada para folha de couve). No geral, os valores de MAPD foram de aproximadamente 75%, sendo o desempenho do sensor P melhor do que o sensor L (MAPD 5 a 10% menor).

Sendo assim, considerando as análises de precisão, reprodutibilidade e exatidão, não houve tendência clara de melhoria no uso dos diferentes tratamentos de dados.

A partir dos dados obtidos foram gerados gráficos para a correlação entre os sensores P e L do Public Lab e FS (Figura 2) e para as diferentes formas de processamento de acordo com as métricas analisadas (Figura 3).

Figura 2 - Correlação dos sensores P e L com o FS, por comprimento de onda utilizando o processamento original.

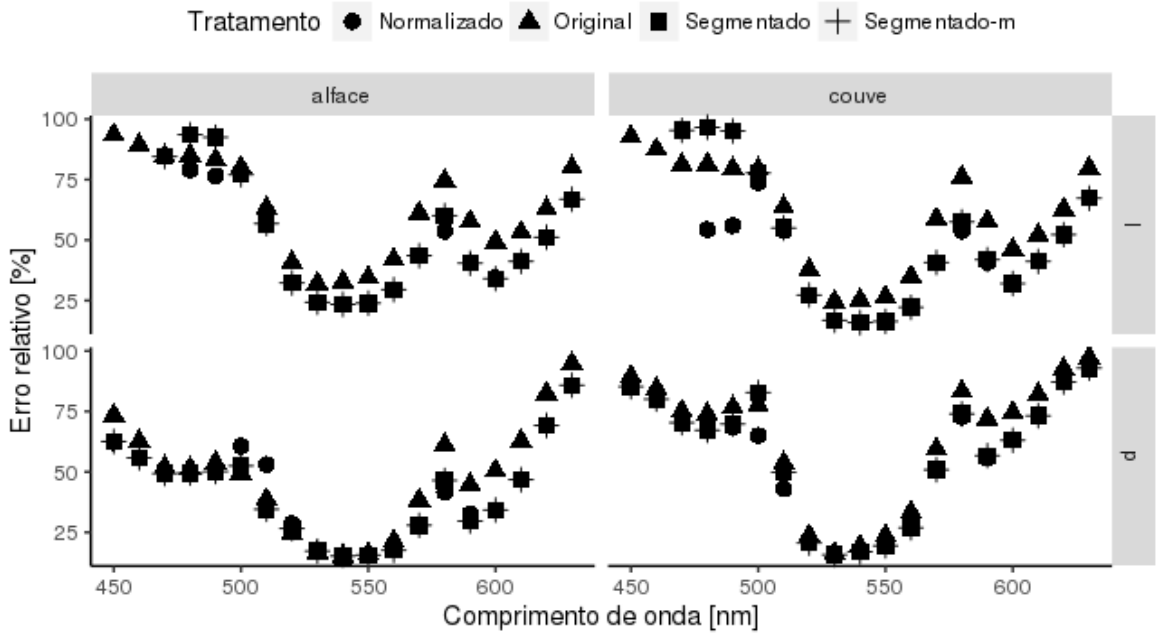


Fonte: Produção dos autores.

Quanto à correlação entre os valores medidos pelos sensores P e L em relação ao FS, o comportamento dos dois sensores é similar. A correlação é maior nos comprimentos de onda maiores do que 525 nm, sendo os maiores valores obtidos para comprimentos de onda maiores que 590 nm. As correlações obtidas pelo sensor L são baixas (entre 0,05 e 0,35), enquanto os valores obtidos para o sensor P são consideradas altas (acima de 0,70 para comprimentos de onda superiores a 510 nm). Picos de correlação foram obtidos para regiões que correspondem aos comprimentos de onda de maior sensibilidade dos canais da câmera. Em uma análise dos erros médios em função do comprimento de onda, pode-se observar um comportamento análogo, onde as regiões de menor erro correspondem aos picos de sensibilidade dos canais. Analisando os gráficos de erro, nota-se que em algumas regiões, erros relativos menores que 25% são obtidos para os processamentos com segmentação. Esse resultado indica que é possível isolar algumas regiões do espectro para uso dos sensores com melhor desempenho. Regiões de menores erros corresponderam aos comprimentos de onda relativos a cor verde. O sensor utilizado no canal verde apresentou maior sensibilidade em todas as análises realizadas, e esse pode ser o motivo dos menores erros. É possível que isso se dê também pelos mecanismos de compensação de cores da *webcam* utilizada.

Investigações similares para erros em função de comprimentos de onda de imagens predominantemente vermelhas ou azuis podem esclarecer melhor esse resultado.

Figura 3: Erro relativo das diferentes formas de processamento em relação ao comprimento de onda para os sensores P e L nas diferentes superfícies.



Fonte: Produção dos autores.

A forma de processar não influenciou os resultados gerais, porém em alguns comprimentos os processamentos com segmentação produziram menores erros, em especial nas regiões do verde e do vermelho. Os erros observados na região do azul (400-500 nm) são resultado da baixa intensidade registrada pelo sensor. Isso ocorre devido ao alto nível de interação entre esse comprimento de onda e o alvo.

CONCLUSÕES

Em todos os processamentos avaliados, as análises de reprodutibilidade e exatidão indicaram que ao serem realizadas as medidas com sensores diferentes o erro esperado é o dobro do obtido nas análises de precisão. Na precisão o erro na referência foi similar a diferença entre P e L. Em situações onde o valor da refletância é relevante, os valores de erro encontrado se mostram insatisfatórios. Em outras condições onde a variação do espectro é a variável de interesse, é possível que os sensores da Public Lab sejam adequados, tendo em vista os valores de correlação obtidos.

Apesar da aplicação de diferentes tratamentos de dados, não houve tendência clara de melhoria para a precisão, reprodutibilidade e exatidão.

Em uma análise dos erros relativos em função do comprimento de onda, percebeu-se que o desempenho das regiões associados ao comprimento de onda do verde e do vermelho foram melhores que o do azul. O comportamento geral das curvas em relação aos diferentes tipos de tratamento foi aproximado, sendo as regiões de menor erro correspondentes aos picos de sensibilidade dos canais. A partir das regiões de menor erro, a investigação dos comprimentos de onda com maior sensibilidade pode gerar potenciais áreas de aplicação para o sensor.

AGRADECIMENTOS

Agradecemos ao PIBIC/CNPq pela concessão da bolsa da primeira autora, à CAPES pela concessão da bolsa do segundo autor e ao CNPq pela concessão da bolsa do segundo e do terceiro autores.

REFERÊNCIAS

COZZOLINO, D. Use of Infrared Spectroscopy for In-Field Measurement and Phenotyping of Plant Properties: Instrumentation, Data Analysis, and Examples. **Applied Spectroscopy Reviews**, v. 49, n. 7, p. 564–584, 2014.

GODFRAY, H. C. J.; BEDDINGTON, J. R.; CRUTE, I. R.; HADDAD, L.; LAWRENCE, D.; MUIR, J. F.; PRETTY, J.; ROBINSON, S.; THOMAS, S. M.; TOULMIN, C. Food security: the challenge of feeding 9 billion people. **Science**, v. 327, n. 5967, p. 812–818, 2010.

LABSPHERE. **Spectralon Diffuse Reflectance Material**. Disponível em: <https://www.labsphere.com/labsphere-products-solutions/materials-coatings-2/coatings-materials/spectralon/>. Acesso em 15 mar 2017.

PARK, B.; LU, R. **Hyperspectral Imaging Technology in Food and Agriculture**. [s.l: s.n.]. New York: Springer, p. 403, 2015.

PUBLICLAB. **SpectralWorkbench**. Disponível em: <https://www.labsphere.com/labsphere-products-solutions/materials-coatings-2/coatings-materials/spectralon/>. Acesso em 10 mai 2017.

R CORE TEAM. **R: A Language and Environment for Statistical Computing** Vienna, Austria, 2016. . Disponível em: <http://www.r-project.org>.

SCHEELINE, A. Cell phone spectrometry: Science in your pocket? **TrAC Trends in Analytical Chemistry**, v. 85, p. 20–25, dez. 2016.

ZHANG, C.; KOVACS, J. M. The application of small unmanned aerial systems for precision agriculture: A review. **Precision Agriculture**, v. 13, n. 6, p. 693–712, 2012.

Modelagem da eficiência, fator importante na rentabilidade da ovinocaprinocultura, dos produtores do Semiárido

Urbano G. P. Abreu[1], Lisiane D. de Lima[2], Fernando H. M. A. R. de Albuquerque[2], Juan D. F. Souza[2]

[1]Embrapa Pantanal, Corumbá, Mato Grosso do Sul, Brasil, urbano.abreu@embrapa.br
[2]Embrapa Caprinos e Ovinos, Sobral, Ceará, Brasil, lisiane.lima@embrapa.br, fernando.albuquerque@embrapa.br, juan.souza@embrapa.br

RESUMO

Neste artigo propõe-se utilizar abordagem DEA para avaliar o desempenho de sistemas de ovinocaprinocultura no semiárido do Brasil. Por meio de questionário com 127 variáveis, foram coletadas informações de 254 produtores do município de Tauá - CE, as varáveis de *input* número de ovinos e número de bovinos, e a variável *output* receita com produção animal. Foi realizada a modelagem DEA de orientação produto com retornos constantes a escala (RCE), e com retornos variáveis a escala (RVE). Os modelos RCE e RVE cujas medidas de eficiência técnica médias foram 9,72% e 23,65%, e que apenas 2 e 8 produtores obtiveram eficiência igual a um, respectivamente. Observamos que a grande maioria dos produtores operam na faixa de retorno crescente, e são ineficientes. Ou seja, a maioria dos produtores estão trabalhando abaixo da escala ótima.

PALAVRAS-CHAVE: Sistema de Produção, Retorno Constante, Retorno Variável.

ABSTRACT

This paper proposes to use the DEA approach to evaluate the performance of ovinocaprinoculture systems in the semi-arid region of Brazil. Using questionnaire with 127 variables, and information from 254 producers in the municipality of Tauá - CE, with the variables of 'input' number of sheep and number of cattle, and the variable 'output' revenues with animal production. DEA model with constant returns to scale (RCE), and with variable returns to scale (RVE) and were used. The RCE and RVE models, whose average technical efficiency measures were 9.72% and 23.65%, and that only 2 and 8 producers obtained efficiency equal to one in each model, respectively. The majority of producers operate in the

increasing return range, and are inefficient. That is, the vast majority of producers are working below the optimum scale, which leads to the exclusion of producers.
KEYWORDS: Production System, Constant Returns, Variable Returns.

INTRODUÇÃO

A conceituação e a busca por medidas rigorosas de eficiência envolvem o estabelecimento de critérios para determinar a melhor situação possível que pode ser obtida pelos agentes econômicos na repartição dos bens produzidos. A busca da eficiência técnica (a menor utilização de insumos), de escala (nível de produção mais adequado) e alocativa (menor custo e maior receita aos preços de mercado) é componente importante nas estratégias de rentabilidade (FERREIRA; GOMES, 2009).

A modelagem de sistemas agropecuários por meio de análise envoltória de dados (DEA) tem evoluído de maneira consistente e rápida, em função da flexibilidade do método, e do esforço teórico no processo de adaptação da metodologia aos mais diferentes sistemas produtivos (GOMES, 2008; EMROUZNEJAD; YANG, 2017).

Vários fatores dão à ovinocaprinocultura importância socioeconômica na definição de opções de atividades de produção. Assim sendo, podem-se destacar a elevada tolerância dos ovinos e caprinos às altas temperaturas e a combinação adequada entre os seus hábitos alimentares e a flora existente. Convém evidenciar, também, as vantagens dos reduzidos investimentos necessários ao criatório, e o papel destacado que desempenha a atividade no suprimento alimentar e na geração de renda, com origem na comercialização da carne, da pele e de subprodutos. Por outro lado, há necessidade de superar os problemas da baixa produtividade, e da reduzida rentabilidade da agropecuária no semiárido (CAMPOS; CAMPOS, 2013).

Neste artigo propõe-se utilizar abordagem DEA para avaliar o desempenho de sistemas de ovinocaprinocultura no semiárido. Pressupõem-se que o conhecimento da eficiência dos sistemas de produção auxiliará para aumento da rentabilidade da atividade, e para melhoria da gestão e dos processos tecnológicos, que refletirá em melhores índices produtivos da cadeia produtiva como um todo.

MATERIAL E MÉTODOS

Por meio de questionário estruturado com 127 variáveis (com informações de ordem econômica, social e zootécnica) foram entrevistados 324 produtores do município de Tauá - CE, sendo que após a retirada de dados perdidos, informações de 254 produtores permaneceram na amostra analisada. Os dados foram inseridos em planilha eletrônica, e

estatísticas exploratórias foram estimadas, com objetivo de verificar a consistência dos dados, conforme descrito por Souza et al. (2014).

Com objetivo de estabelecer tipologias para os sistemas de produção Abreu et al. (2016), por meio da análise multivariada de fatores verificaram que os dois primeiros fatores explicaram 80% da (co)variação total das informações. Por meio da análise das estimativas das cargas fatoriais foram mantidas as variáveis: quantidade de ovinos - QuantOv, quantidade de bovinos - QuanBov, receita total anual com produtos de origem animal - Recprodoani, receita total anual - RecT, e receita total anual agropecuária- RecTagro.

A análise envoltória de dados (DEA) foi originalmente proposta por Charnes, et al., (1978) como metodologia para avaliação da eficiência relativa de unidades homogêneas que possuem autonomia na tomada de decisão (*Decision Making Units – DMU*). Os modelos DEA são baseados em amostra de dados (insumos e produtos) observados para diferentes DMU's. O objetivo é construir um conjunto de referência convexo, e então classificar as DMU's em eficientes ou ineficientes, tendo como referencial a superfície formada.

O modelo desenvolvido por Charnes et al. (1978), é chamado de modelo com retornos constantes à escala (RCE), sendo reformulado por Banker et al. (1984) que desenvolveram outro modelo com objetivo de possibilitar a análise no caso de retornos variáveis à escala (RVE). Na análise formula-se o modelo a ser utilizado, e a sua resolução calcula a fronteira de produção como um envoltório nos dados, determinando para cada DMU, se a mesma, está ou não na fronteira de eficiência . A utilização da metodologia não requer especificação de uma forma funcional que relacione os dados, por ser uma técnica não paramétrica. A metodologia é baseada em programação linear, que visa mensuração do desempenho relativo de unidades organizacionais, onde ocorre presença de múltiplos insumos e produtos relativamente homogêneos, o que causa dificuldade nas comparações. No caso simples onde o processo de produção possui apenas um insumo e apenas um produto, eficiência pode ser definida como:

Eficiência = Produto / Insumo

De forma geral as organizações possuem um número grande de produtos e de insumos, e esta complexidade pode ser incorporada na mensuração da eficiência pela definição de eficiência como:

Eficiência =Soma ponderada dos Produtos / Soma ponderada dos Insumos

O que requer série de pesos que são difíceis de serem estimados, particularmente se uma série de pesos necessitam ser aplicados para um número grande de unidades

organizacionais com diferentes características. Este problema pode ser resolvido pela definição de que cada unidade individual pode possuir seu próprio sistema de valores e, consequentemente definir seu próprio sistema de pesos para ponderação (COOPER, et al., 2006). A pressuposição fundamental na técnica DEA é que, se dada DMU é capaz de produzir Y(A) unidades de produto, utilizando-se X(A) unidades de insumos, outras DMU's poderiam fazer o mesmo, caso elas estejam operando eficientemente. Da mesma forma, se outra DMU é capaz de produzir Y(B) unidades de produto, utilizando X(B) insumos, então outras firmas podem fazer o mesmo esquema de produção. Caso as duas DMU's sejam eficientes, poderiam ser combinadas para formação de DMU composta que utiliza combinação de insumos para produzir combinação de produtos. Se esta DMU composta não existir ela passa a ser denominada DMU virtual. A metodologia DEA, consiste em encontrar o melhor DMU virtual para cada DMU da amostra. No caso da DMU virtual ser melhor que a DMU original, ou por produzir mais com a mesma quantidade de insumos, ou por produzir a mesma quantidade com menos insumos, a DMU original será ineficiente.

Charnes et al. (1978) propuseram que a eficiência da unidade j_0 pode ser avaliada pela solução do seguinte modelo:

$$Max\ h_0 = \frac{\sum_{r=1}^{t} u_r y_{rj0}}{\sum_{i=1}^{m} v_i x_{ij0}} \qquad (M1)$$

sujeito a :

$$\frac{\sum_{r=1}^{t} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1, \quad j = 1, \ldots, n,$$

$u_r, v_i \geq \varepsilon, \quad \forall \quad r$ e i,

onde:

y_{rj} = total de saídas r, oriunda da unidade produtiva j;
x_{ij} = total de saídas i, oriunda da unidade produtiva j;
u_r = peso dado aos insumos r;
v_i = peso dado aos produtos i;
n = número de unidades produtivas;
t = número de produtos;
m = número de insumos; e
ε = número pequeno e positivo.

Na solução deste modelo a eficiência da unidade j_0 , é maximizada sujeita à restrição de que as medidas de eficiência de todas as outras unidades produtivas sejam menores ou iguais a um. A característica chave do modelo acima é que os pesos u_r e v_i são tratadas como

incógnitas, sendo então escolhidas de forma que a eficiência de j_0 seja maximizada. Se a eficiência de j_0 for igual a 1, será considerada eficiente em relação a outras unidades produtivas do contrário, é considerada ineficiente. No caso de uma unidade ineficiente a solução identifica unidades eficientes correspondentes que formam um grupo de referência (*benchmark*) para unidades ineficientes.

O modelo DEA acima descrito é um modelo linear fracionário, mas pode ser convertido na forma linear, de maneira que os métodos de programação linear possam ser aplicados. O modelo na forma de programação linear é mostrado abaixo:

$$Max\ h_0 = \sum_{r=1}^{t} u_r y_{rj0} \qquad (M2)$$

sujeito a:

$$\sum_{i=1}^{m} v_i x_{ij0} = 1,$$

$$\sum_{r=1}^{t} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0,$$

$j = 1, \ldots, n,$

$-u_r \leq -\varepsilon,\ r = 1, \ldots, t,$

$-v_i \leq -\varepsilon,\ i = 1, \ldots, m.$

A função objetivo sendo linearizada permite reconhecer, que durante a maximização a razão (magnitudes relativas do numerador e denominador) é o fator importante na análise, e não seus valores iniciais. Desta forma, no modelo (M2) o denominador é igual a uma constante (arbitrariamente 1 na notação) sendo o numerador maximizado. O modelo é aplicado a cada unidade com objetivo de obtenção uma a uma das medidas de eficiência. A solução é computacionalmente facilitada, pois a maioria das restrições é a mesma para diferentes unidades produtivas.

Para calcular a eficiência relativa de uma DMU, resolve-se o problema *dual*, modelo que pode ser descrito para no caso de retornos constantes à escala (CCR), de acordo com a notação de Coelli (1996), da seguinte maneira:

$$Min_{\theta,\lambda}\ \theta \qquad (M3)$$

sujeito a:

$$-y + Y\lambda \geq 0$$

$$\theta x_i - X\lambda \geq 0$$

$$\lambda \geq 0$$

onde:

y = produto do modelo sob análise;

x = insumo do modelo sob análise;

X = matriz de insumos K x N;

Y = matriz de produtos M x N;
θ = é uma escalar que multiplica os vetores dos insumos;
λ = é um vetor de constantes N x 1 que multiplica a matriz de insumos e produtos; e
N = número de DMU's.

O modelo *dual* permite a análise da eficiência relativa fornecida nos casos em que existam folgas (*slacks*) ou reduções não radiais nos insumos. Para que uma DMU seja considerada tecnicamente eficiente, θ será igual a um, e as folgas iguais a zero. O problema linear deve ser resolvido N vezes obtendo-se para cada iteração a eficiência relativa de cada DMU.

Banker et al. (1984), desenvolveram importante extensão do modelo DEA com retornos constantes à escala (RCE). Os autores modificaram o programa linear de maneira a incorporar uma restrição a convexidade ($N1'\lambda=1$), permitindo a análise de rendimentos variáveis à escala (RVE). O modelo é descrito da seguinte maneira:

$$Min_{\theta,\lambda}\ \theta \qquad \text{(M4)}$$

sujeito a:

$$-y + Y\lambda \geq 0$$

$$\theta x_i - X\lambda \geq 0$$

$$N1'\lambda = 1$$

$$\lambda \geq 0$$

onde:
$N1$ = vetor unitário.

Esta modificação permitiu a decomposição da eficiência técnica em duas: a eficiência técnica pura e a eficiência de escala. Para tal, devemos resolver o problema linear através dos dois modelos (M3 e M4) descritos, e se houver uma diferença entre as duas soluções em uma DMU em particular, significa que a DMU, possui ineficiência de escala, e o valor da ineficiência é a razão entre os valores encontrados nos modelos RCE e RVE.

A eficiência técnica pura coincide com a solução para o modelo RVE. A ineficiência de escala origina-se da produção em nível deficiente de escala. O nível ótimo de escala é considerado o que é obtido nas comparações com as firmas eficientes (RCE = 1). A eficiência técnica global é o produto das duas eficiências (técnica pura e escala), e sua medição coincide com o modelo RCE.

O modelo DEA pode ter duas orientações. Realizando a otimização na combinação de insumos (modelo insumo orientado) para a produção de produtos, ou realizando a otimização para produção de produtos (modelo produto orientado), diferenças e detalhes da metodologia podem ser vistos em Cooper et al. (2006).

A interpretação das ineficiências depende do modelo ser insumo ou produto orientado. O primeiro, nos permite conhecer qual a proporção de insumos que foram transformados em produtos nas diferentes DMU's. O segundo visão, permite inferir o quanto foi incremento de produto em todos os DMU's eficientes. Um indicador interessante que visa diferenciar a ineficiência de escala de uma DMU, pode ser calculado visando verificar se a DMU esta operando com retornos decrescentes à escala (RNC) ou numa área de retornos crescentes em escala (RND). Para tal devemos substituir a restrição $N1'\lambda = 1$ por $N1'\lambda \leq 1$. Desta forma, incorpora-se nas restrições, a impossibilidade de rendimentos crescentes a escala, se o novo valor obtido ao executar a análise for igual ao obtido no modelo RVE, significa que a DMU esta operando no setor das curvas de rendimentos decrescentes a escala. Se o contrário ocorrer, significa que a DMU, esta operando no setor de rendimentos crescentes a escala. Logicamente as DMU's nas quais os resultados dos modelos RVE e RCE são iguais possuem escala ótima e não são consideradas nesta classificação.

Os dados de cada produtor levantados *in loco*, foram considerados uma DMU, sendo as variáveis de input, quantidade de ovinos - QuantOv, quantidade de bovinos - QuanBov,. E a variável output, receita total anual com produtos de origem animal - Recprodoani. A modelagem DEA utilizada foi de orientação produto com retornos constantes a escala (RCE) e com retornos variáveis a escala (RVE), de acordo com o descrito por Cooper et al. (2006).

RESULTADOS E DISCUSSÃO

Na Tabela 1, os totais e percentuais dos produtores, em relação ao retorno constante (RCE), e variáveis (RVE), e a eficiência de escala. Sob a pressuposição de retornos constantes de escala, verificou-se que, da amostra total de 254 produtores, apenas 2 obtiveram máxima eficiência técnica. O nível médio de ineficiência técnica foi de 90,28%, com direcionamento de produto (output). O que significa que os produtores poderão, em média, aumentar até 90,28% sua renda com a melhor combinação na produção de ovinos/bovinos nos sistemas de produção dos pequenos produtores. O modelo RVE cuja medida de eficiência técnica média foi de 23,65%, e que apenas 8 produtores obtiveram medida de eficiência igual a um. Ou seja são modelos de eficiência para o conjunto de produtores analisados. Isto significa que, dos 8 produtores com eficiência técnica igual a um, no modelo com retornos variáveis, apenas 2 deles são igualmente eficientes, no modelo com retornos constantes.

A medida de Eficiência de Escala é obtida pela razão entre as medidas de eficiência técnica dos modelos com retornos constantes e com retornos variáveis. Se essa razão for igual

a um, o produtor estará operando na escala ótima. Se for menor que um o produtor será tecnicamente ineficiente, pois não estará operando na escala ótima.

Tabela 1 – Número e percentuais de produtores classificados de acordo com relação ao retorno constante (RCE), e retorno variáveis (RVE), e a eficiência de escala.

Eficiência	Números e Percentuais de Produtores		
	RCE	RVE	Eficiência de Escala
E < 0,1	182 (71,65%)	85 (33,46%)	8 (3,15%)
0,1 ≤ E < 0,2	48 (18,90%)	66 (25,98%)	29 (11,42%)
0,2 ≤ E < 0,3	10 (3,94%)	30 (11,81%)	48 (18,90%)
0,3 ≤ E < 0,4	5 (1,97%)	33 (12,99%)	37 (14,57%)
0,4 ≤ E < 0,5	3 (1,18%)	12 (4,72%)	34 (13,39%)
0,5 ≤ E < 0,6	2 (0,79%)	5 (1,97%)	23 (9,06%)
0,6 ≤ E < 0,7	1 (0,39%)	7 (2,76%)	18 (7,09%)
0,7 ≤ E < 0,8	1 (0,39%)	2 (0,79%)	19 (7,48%)
0,8 ≤ E < 0,9	0	5 (1,96%)	16 (6,30%)
0,9 ≤ E < 1,0	0	1(0,39%)	18 (7,09%)
E = 1,0	2 (0,79%)	8 (3,14%)	4 (1,57%)
Total	254	254	254
Medidas de Eficiência			
Média	9,72	23,65	46,96
Desvio-padrão	12,92	23,10	26,37
Coeficiente de variação	1,33	0,98	0,56

Para detectar se a natureza das ineficiências de escala são devidas ao fato do produtor operar na faixa de retornos crescentes ou na faixa de retornos decrescentes foi imposto a restrição de retornos não crescentes de escala. Se o valor da medida de eficiência encontrado nesse modelo for igual ao valor detectado no modelo com retorno variáveis, então o produtor encontra-se na faixa de retorno decrescente a escala, ou seja, está operando acima da escala ótima. Do contrário, o produtor situa-se na faixa de retornos crescentes. Na Tabela 2, mostra-se o resultado do número de produtores segundo o tipo do retorno.

Tabela 2 - Número e distribuição dos produtores de Tauá - CE, segundo o tipo de retorno.

Tipo de Retorno	Eficientes	Ineficientes	Total
Crescente	0	230	230
Constante	2	0	2
Decrescente	2	20	22
Total	4	250	254

Claramente observamos que a grande maioria dos produtores operam na faixa de retorno crescente, e são ineficientes. Ou seja, 90% dos produtores estão trabalhando abaixo da escala ótima com utilização de muito 'input' e alcançando pouco 'output', o que direciona a

exclusão dos produtores como consequência da falta de uso de tecnologia, das imperfeições de mercado, e das políticas públicas que discriminam a pequena produção. Para aumentar a eficiência técnica, e a chegar a escala ótima é necessário aumentar a produção, porém esse aumento deve ocorrer com balanço entre as quantidades utilizadas de 'input' e o volume de 'output'.

De acordo com Ferreira e Gomes (2009) o progresso tecnológico é uma das poucas soluções viáveis para impedir a ineficiência na produção resultante da tentativa de aumentar a produção. A tecnologia é conhecimento criado pela pesquisa e desenvolvido pelos produtores em sistema de produção. De acordo com Alves et al. (2012), poucos estabelecimentos foram capazes de fazer essa organização, seja porque contaram com assistência técnica especializada, seja porque seus administradores eram competentes nessa especialização. Porém não ocorreu o mesmo com milhões de estabelecimentos. À pesquisa, no caso da agricultura familiar, em conjunto com a assistência técnica e extensão rural, cabe organizar os sistemas de produção, tendo como critério de organização a compreensão dos agricultores e a lucratividade.

Por outro lado, Silva e Costa (2014) enfatizam que qualquer proposta de reestruturação dos estabelecimentos rurais de menor porte econômico do Semiárido a partir da adoção de pacotes tecnológicos, que não levarem em consideração, por exemplo, a salinização dos solos, a desertificação de áreas, o desaparecimento de espécies nativas, escolaridade dos produtores, etc., aspectos típicos do Bioma, que se não forem considerados, levarão à inadequação tecnológica desses pacotes.

Apesar das condições favoráveis ao criatório, o sistema de produção de ovino e caprino, na região é desenvolvida de maneira extensiva, caracterizando-se por alimentação deficiente, manejo e profilaxia inadequados, o que resulta em baixa produtividade, baixo desfrute e, em consequência, insatisfatórios resultados econômico-financeiros. O que direciona ao subdesenvolvimento da cadeia produtiva, sendo a atividade caracterizada pela prevalência de arranjos produtivos pouco organizados, incapazes de gerar competitividade para o sistema agroindustrial da ovinocultura (SOUZA et al., 2014).

CONCLUSÕES

O sistema de produção dos ovinocaprinocultores de Tauá-CE apresentam em sua maioria ineficiência de escala, e operam na faixa de retornos crescentes. Há amplo espaço para o desenvolvimento tecnológico dos sistemas de produção. Em paralelo, políticas públicas, ajustadas às peculiaridades locais, devem ser desenvolvidas para organização da produção e desenvolvimento tecnológico da região.



Uma Aplicação Desktop Interativa para Desenvolvedores e Usuários de Modelos de Sistemas Dinâmicos em R

Bruno Henrique Pereira[1], *Cristina Freitas Bazzano*[1],
Luís Gustavo Barionni[1], *Adauto Luiz Mancini*[1], *Márcio Nicolau*[2]

[1] Embrapa, Av. André Tosello, nº 209 Campus da Unicamp,
Barão Geraldo, Campinas, São Paulo, Brasil
hbrunopereira@gmail.com, crisfbazz@gmail.com,
luis.barioni@embrapa.br, adauto.mancini@embrapa.br

[2]Embrapa, BR 285 Km 294, Passo Fundo, Rio Grande do Sul, Brasil
marcio.nicolau@embrapa.br

RESUMO

A aplicação apresentada neste artigo propõe facilitar o desenvolvimento e uso de simuladores de modelos de sistemas dinâmicos contínuos. Seu ambiente, intuitivo e didático, permite a definição e simulação dos modelos de forma rápida e organizada. Os modelos desenvolvidos podem ser salvos em arquivos, possibilitando a criação de repositórios e o compartilhamento de simuladores. Assim, a aplicação permite que usuários de diversas áreas do conhecimento, tal como a agropecuária, colaborem no processo de desenvolvimento e simulação de modelos de sistemas dinâmicos.

PALAVRAS-CHAVE: Modelagem e simulação, Compartilhamento de Código, RShiny.

ABSTRACT

The application presented herein proposes to facilitate the development and use of simulators of continuous dynamic system models. Its environment, intuitive and didactic, allows the definition and simulation of the models in a fast and organized way. The developed models can be saved in files, making it possible to create repositories and share simulators. Thus, the application allows users of different areas of knowledge, such as agriculture and livestock, to collaborate in the process of development and simulation of dynamic system models.

KEYWORDS: Modeling and simulation, Code-sharing, RShiny.

INTRODUÇÃO

R é uma linguagem para computação estatística e gráficos (R Core Team, 2017). O ambiente colaborativo do R permite extensão de sua funcionalidade por meio de pacotes desenvolvidos por uma ampla comunidade de pesquisadores e desenvolvedores, incluindo novos algoritmos e técnicas gráficas. Os pacotes deSolve (SOETAERT; PETZOLDT; SETZER, 2010), Simecol (PETZOLDT; RINKE, 2007), Rodeo (KNEIS, 2017) e sdsim dão suporte à implementação de modelos de sistemas dinâmicos (MSDs) contínuos, baseados em sistemas de equações diferenciais ordinárias (EDOs) de primeira ordem.

O desenvolvimento e uso de simuladores de MSDs requer equipes de profissionais com diferentes perfis. Desde profissionais ligados à computação até especialistas de domínio, que utilizam modelos apenas para compreensão do comportamento de um sistema, desinteressados pelos detalhes de sua implementação (FREUA et al., 2014).

Na agropecuária a modelagem de sistemas tornou-se essencial ao longo das últimas décadas. A partir de experimentos baseados em simulações de MSDs é possível entender e predizer o comportamento e performance do sistema real, facilitando a tomada de decisões e a adequação a diversos cenários. Essa abordagem pode ser usada, por exemplo, para prever o crescimento bovino a partir de dados sobre sua dieta. (JONES et al., 2016).

O pacote R sdsim, desenvolvido pelos autores deste artigo, utiliza uma moderna abordagem de implementação orientada a objetos para sistemas dinâmicos (SD) com rigorosa separação do código e algoritmos de solução do modelo e é o único, atualmente, que dá suporte explícito ao desenvolvimento de modelos acoplados. O pacote sdsim também suporta o armazenamento e posterior e recuperação de modelos, o que viabiliza e facilita a distribuição de simuladores, possibilitando que pesquisadores usem o pacote como um meio de comunicação na implementação de MSDs.

A especificação e o uso de modelos criados com o pacote sdsim em script R é intuitivo para usuários com conhecimento intermediário na linguagem. Porém a prototipação de modelos com auxílio de uma aplicação com interface gráfica de usuário pode reduzir drasticamente a curva de aprendizado e a ocorrência de erros de especificação.

O objetivo deste trabalho foi desenvolver uma aplicação, por meio do framework Shiny e do pacote sdsim, para a prototipagem, execução e compartilhamento de simuladores de MSDs atômicos de forma simples e acessível para usuários de diferentes áreas.

MATERIAL E MÉTODOS

A aplicação aborda todo o processo de definição de MSDs atômicos, da especificação até a execução de simulações. Esses modelos são especificados por variáveis de entrada (e.g. parâmetros, constantes, *inputs* e *switches*), séries temporais, variáveis auxiliares e variáveis de estado, EDOs de primeira ordem, funções opcionais de inicialização e pós processamento, e funções associadas ao disparo e execução de eventos.

Shiny

A aplicação foi desenvolvida por meio do framework Shiny (CHANG et al., 2017), um *framework* para criação de interfaces web interativas associadas a códigos desenvolvidos em R. O Pacote Shiny utiliza programação funcional reativa em seu desenvolvimento, idéia esta obtida de outro *framework* web chamada Meteor (Meteor Development Group, 2016). Utilizou-se pacotes adicionais, específicos para o Shiny, como o shinydashboard (CHANG; RIBEIRO, 2017) para o layout da aplicação, o shinyAce (ALLEN, 2016) para gerar campos de edição de código e o rhandsontable (OWEN, 2016) para gerar planilhas editáveis.

sdsim

O pacote sdsim provê a funcionalidade necessária para execução de simulações. Esse pacote possui suporte a eventos, diversos métodos de integração numérica, processamento dos dados de entrada (e.g. séries temporais, variáveis auxiliares, função de inicialização de parâmetros), e exibição de resultados em tabela e gráficos.

As funções e variáveis dos modelos informadas na aplicação são encapsuladas em objetos da classe SDModel do pacote sdsim. A aplicação usa os métodos *Run* e *Plot* dos objetos instanciados. O método *Run* usa o método de integração numérica, o tempo e o passo da simulação informados na aplicação para a execução da simulação do modelo, e retorna os dados resultantes. O método *Plot*, gera os gráficos do resultado da simulação a partir das configurações de plotagem definidas na aplicação.

Especificação de Modelos

A representação de um modelo atômico (M) utilizada neste trabalho é definida por meio da tupla 1. Os elementos obrigatórios da tupla foram representados por meio do símbolo * (asterisco).

$$M = \{S*, C, P, W, I, A, T, \delta d*, \delta i, \delta R, \delta E, \delta G\} \quad (1)$$

Onde,

- S (*state*) é o vetor de variáveis de estado, contendo o estado inicial do modelo;
- C (*const*) é o vetor de constantes do modelo (imutáveis entre simulações e inacessíveis para calibração estatística);
- P (*parms*) é o vetor de parâmetros do modelo (constantes somente ao longo de uma simulação);
- W (*switches*) é o vetor de variáveis discretas que podem ser utilizadas como seletores no modelo;
- I (*inputs*) é o vetor contendo variáveis de entrada;
- T (*timeSeries*) é o vetor contendo as variáveis de séries temporais, cujos valores dependem do tempo da simulação;

- A (*aux*) é o vetor de variáveis auxiliares, cujos valores são equações dependentes de outras variáveis do sistema avaliadas a cada passo de simulação;
- δd (*Differential Equations*) é uma função que define o vetor das diferenciais das variáveis de estado do sistema;
- δi (*Parameter Initialization*) é a uma função, executada antes do início da simulação, que pode ser usada para estabelecer valores iniciais para os parâmetros e para as variáveis de estado;
- δR (*Root*) é uma função que verifica as condições para a ocorrência de um evento;
- δE (*Events*) é uma função que define a resposta a um determinado evento;
- δG (*Global Functions*) é um vetor de funções adicionais que podem ser executadas dentro do escopo de quaisquer funções definidas no modelo.

Variáveis

Os MSDs representados contêm um conjunto de variáveis de entrada (parâmetros, constantes, *inputs*, *switches*), de séries temporais, de saída, de estado e auxiliares relacionadas entre si por meio das equações diferenciais de primeira ordem. Essas variáveis são definidas por meio de planilhas editáveis.

Essas variáveis são definidas por meio de planilhas editáveis, que possuem as seguintes colunas: (a) *Variable*, armazena o nome das variáveis; (b) *Value*, armazena o valor das variáveis. (c) *Unit* (opcional)), armazena a unidade de medida das variáveis e; (d) *Description* (opcional), armazena breves descrições das variáveis.

A coluna *Value* é preenchida de maneira diferente em algumas planilhas. Nas planilhas *State*, *Const*, *Parms*, *Inputs* e *Switches* cada variável possui um valor estático. Na planilha das variáveis auxiliares a coluna *Value* deve ser preenchida com equações escritas na linguagem R. Essas fórmulas serão computadas em cada passo da simulação. A planilha que define as séries temporais se diferencia por conter duas colunas adicionais: (a) *Times*, armazena um vetor com os tempos das séries temporais; (b) *Interpolation*, armazena o método de interpolação (funções degrau, linear ou spline). Além disso, para essas variáveis a coluna *Value* deve ser preenchida com um vetor de valores. Esses valores correspondem a cada tempo do vetor definido na coluna *Times*, formando uma matriz de valor por tempo que descreve a série temporal. Se o campo *Times* não for preenchido, *Value* deve conter apenas um valor e será considerado constante para todos os tempos de simulação.

Funções

As funções necessárias para a especificação do simulador são escritas em linguagem R, e recebem como argumento as variáveis definidas nas planilhas. Esses argumentos podem ser: (a) *t*, que contém o tempo atual da simulação; (b) *state*, que contém as variáveis de estado; (c) *vars*, que contém as listas *const*, *parms*, *inputs* e *switches*; (d) *timeSeries*, que contém as variáveis de

séries temporais, cujo valor é automaticamente calculado para o tempo *t* da simulação. (e) *aux*, que contém as variáveis auxiliares, cujo valor é automaticamente calculado para o tempo *t* da simulação.

Armazenamento de modelos

O compartilhamento de um modelo especificado com a aplicação é suportado por meio do formato nativo RData, que permite encapsular todos os dados de um modelo em um único arquivo. O arquivo RData armazenado contém um objeto de simulação da classe SDModel, e por isso pode ser carregado e executado via comandos em ambiente R utilizando o pacote sdsim.

A aplicação também permite armazenar o modelo em um projeto R. Isso é útil por fornecer a flexibilidade da linguagem R e toda a funcionalidade do pacote sdsim. O projeto R armazenado possui: (a) arquivos com os parâmetros de entrada; (b) um arquivo de configurações para gráficos; (c) um arquivo com configurações de diretórios e formatos de dados e; (d) um script que descreve as funções do modelo e instancia o objeto de simulação usando o pacote sdsim. O projeto gerado pode então ser executado em ambiente R a partir da recuperação de um objeto de simulação.

Detecção de erros

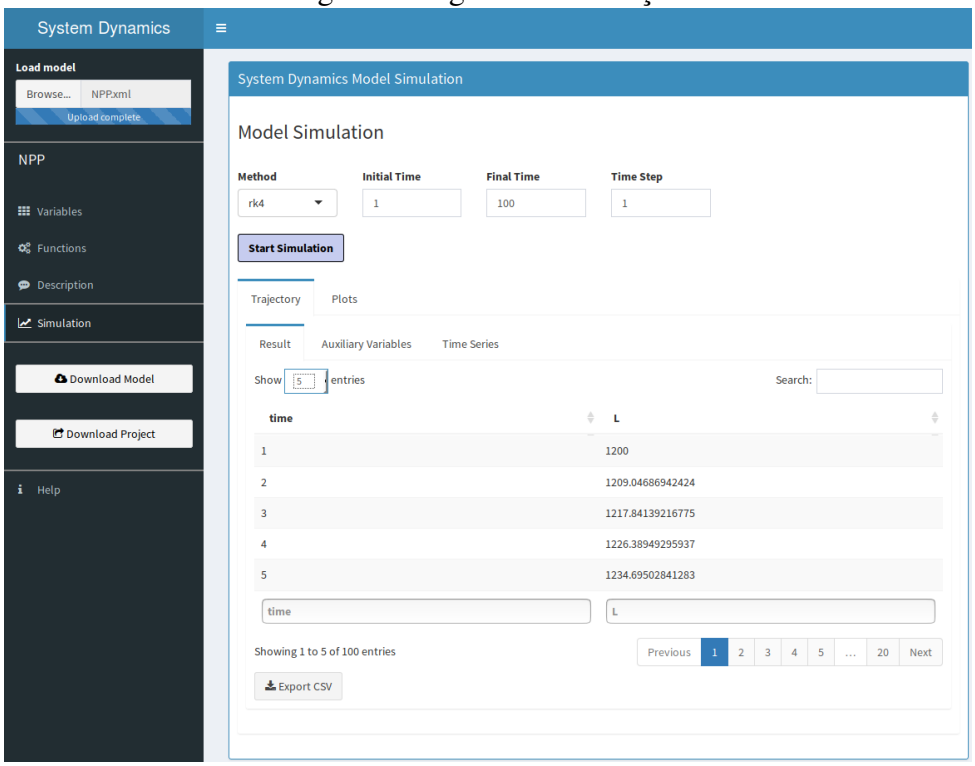
A aplicação é beneficiada pela funcionalidade de deteção de erros inclusa no pacote sdsim. Esta deteção de erros consiste em: (a) checagem de tipo de parâmetros, que notifica o usuário caso algum dos parâmetros tenha sido fornecido de forma incorreta; (b) detecção de valores nulos, ausentes ou infinitos, que podem indicar a presença de erros nos cálculos ou parâmetros do modelo; (c) visualização de dados intermediários para verificar o funcionamento do simulador.

RESULTADOS E DISCUSSÃO

A aplicação (Figura 1) apresenta toda a estrutura de especificação intuitivamente organizada, sendo necessário apenas preencher os campos correspondentes. Para facilitar o uso, instruções de preenchimento foram organizadas em breves descrições que acompanham cada campo de edição. Além disso, um menu de ajuda oferece um tutorial sobre como utilizar a aplicação.

O exemplo utilizado nas figuras é uma simplificação do modelo de crescimento de massa foliar de uma pastagem. A taxa de crescimento (dL) da massa foliar (L) do modelo é influenciada positivamente pela radiação solar diária, pela partição de fotoassimilados e pela quantidade de água disponível no solo (W). Usando o pacote sdsim um modelo acoplado pode ser criado a partir do modelo de crescimento foliar e um modelo de água disponível no solo, que calcula a variação de W ao decorrer da simulação. Neste exemplo o valor de W é constante de 50mm durante a simulação, pois a aplicação só suporta modelos atômicos, e L possui valor inicial de 1200 kg/ha.

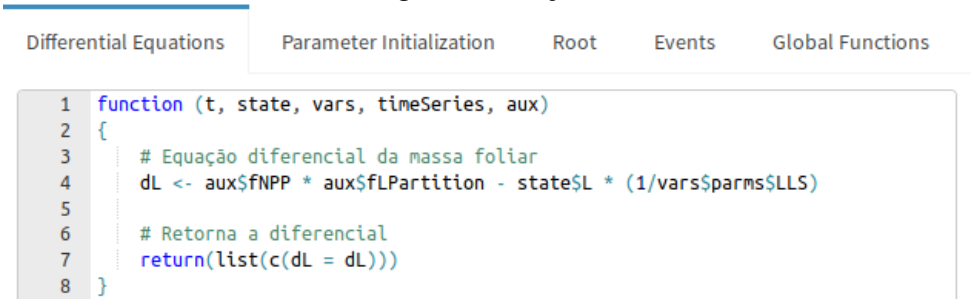
Figura 1: Página de Simulação



Especificação de funções

As funções são preenchidas em campos para edição (Figura 2). Esses campos possuem numeração de linhas, coloração do texto seguindo a formatação de código R, predição de palavras, entre outras funcionalidades que facilitam a definição das funções em linguagem R.

Figura 2: Funções



Especificação de variáveis

As variáveis de um modelo, disponíveis no menu Variables, são definidas usando as planilhas (Figuras 3, 4, 5 e 6) editáveis, onde cada linha representa uma variável.

Figura 3: Variável de Estado

	Variable	Value	Unit	Description
1	L	1200	kg/ha	Massa foliar

Figura 4: Variáveis de Entrada

	Variable	Value	Unit	Description
1	W	50	mm	Água em estoque
2	Q	18	MJ/m^2/day	Radiação global

Figura 5: Parâmetros

	Variable	Value	Unit	Description
1	SLA	0,002857143	m2/kg	Área foliar específica
2	k	0,5		Coeficiente de extinção da radiação
3	SWconst	0,7		As constantes são: 0,7 para areia, 0,6 para silte arenoso, 0,5 para silte argiloso, 0,4 para argila
4	epsilon	8	gDM/MJ	Eficiência de uso da radiação fotossinteticamente ativa
5	y	0,47		Coeficiente de respiração
6	Lmax	3000	kg/ha	Massa foliar máxima
7	LLS	50	dias	Tempo de vida da folha
8	SWpower	9		A expoente na equação para modificadores SW: 9 para areia, 7 para silte arenoso, 5 para silte argiloso e 3 para argila
9	Wmax	80	mm	Capacidade máxima de armazenamento de água do solo

Figura 6: Variáveis Auxiliares

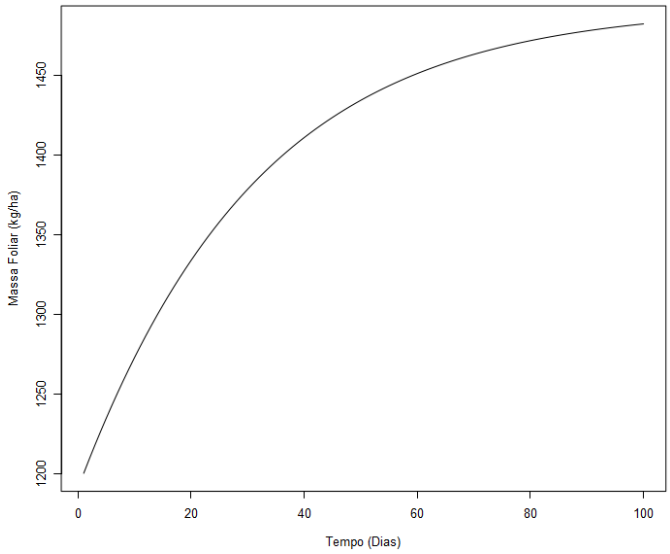
	Variable	Value	Unit	Description
1	fLAI	stateSL * varsSparmsSSLA		Índice da Área da Folha
2	fLIF	1 - exp(-varsSparmsSk * auxSfLAI)		Índice de Intercepção de Luz
3	fRADint	varsSinputsSQ * auxSfLIF		Radiação Interceptada
4	fWRatio	min((varsSinputsSW/varsSparmsSWmax), 1)		Razão entre estoque e capacidade máxima de água no solo
5	fSW	1/(1 + ((1 – auxSfWRatio)/varsSparmsSSWconst)^varsSparmsSSWpower)		Absorção de água pelo solo
6	fGPP	varsSparmsSepsilon * auxSfSW * auxSfRADint		Produção Primária Bruta
7	fNPP	auxSfGPP * varsSparmsSy		Produção Primária líquida
8	fLPartition	(varsSparmsSLmax – stateSL)/varsSparmsSLmax		Partição de Nutrientes

Execução de simulações

A execução do simulador é feita por meio do menu '*Simulation*'. O usuário escolhe o método de integração numérica e informa o tempo e o passo da simulação. Os métodos de integração disponíveis são os mesmos disponibilizados pelo pacote deSolve (e.g. euler, radau, runge-kutta, adams). Quando o botão '*Start Simulation*' é pressionado a simulação do modelo é executada e a trajetória das variáveis de estado, das variáveis auxiliares e das séries temporais são exibidas em uma tabela (Figura 1). Todas as trajetórias podem ser exportadas para arquivos em formato *csv* (valores separados por vírgula), que podem facilmente ser visualizados em planilhas de cálculo como o Calc e o Excel, por exemplo.

Os resultados podem ser exibidos graficamente (Figura 7) correlacionando as variáveis contidas nas trajetórias ou em função do tempo de simulação. Esse comportamento pode ser definido através da planilha de configuração de gráficos. Caso nenhuma configuração gráfica seja definida, por padrão, apenas as variáveis de estado serão projetadas em relação ao tempo.

Figura 7: Trajetória da Massa Foliar



Documentação de modelos

O menu *Description* contém um campo de texto para documentação de detalhes sobre o modelo e seu funcionamento. Esse mecanismo facilita o entendimento de modelos compartilhados por terceiros e permite seu armazenamento em repositórios.

Salvar e carregar modelos

Um modelo criado pela aplicação pode ser armazenado em formato de arquivo RData usando o botão '*Download Model*'. Este botão abre uma janela em que o usuário pode informar o nome do arquivo e baixar o modelo. Após armazenado, o modelo pode ser carregado de volta à aplicação por meio do menu '*Load Model*'.

Exportação de projetos

Um modelo pode ser exportado para formato de um projeto R, que pode ser armazenado por meio de um arquivo compactado (*zip*) usando o botão '*Download Project*'. Este botão abre uma janela em que o usuário pode informar o nome do projeto, configurar formatos de arquivo para os parâmetros do modelo (planilhas em formato excel ou formato *csv*), e baixar o projeto.

CONCLUSÕES

A aplicação desenvolvida apresenta vários benefícios para a modelagem e simulação de sistemas agropecuários, pois ela proporciona um ambiente amigável, organizado e intuitivo para o desenvolvimento de simuladores de modelos de sistemas dinâmicos contínuos. Seu uso facilita o aprendizado de modelagem e não requer conhecimento avançado de programação; além de

promover documentação, a expressividade e a manutenção de modelos de maior porte sem se apoiar na construção de diagramas complexos e extensos.

Além disso, o uso da aplicação acelera o processo de desenvolvimento de simuladores. Para criar um simulador usando a linguagem R é necessária uma estrutura por meio da qual as variáveis serão declaradas, as funções especificadas e o simulador inicializado e executado. A aplicação cria essa estrutura automaticamente, sendo necessário apenas o preenchimento de cada campo que compõe o modelo, simplificando o desenvolvimento e favorecendo a padronização, a checagem e a auditoria do modelo.

Os modelos criados pela aplicação possuem uma estrutura bem definida e padronizada. Esses modelos podem ser documentados, encapsulados em objetos e armazenados em arquivos no formato RData. Esse método facilita a portabilidade, o entendimento, o reuso e a manutenção de modelos.

A partir dos modelos armazenados em RData e projetos R também é possível criar repositórios de compartilhamento de modelos. Esses modelos podem então ser executados, calibrados e modificados de acordo com a necessidade de cada usuário. Isso também possibilita que esses modelos sejam carregados e manipulados em scripts R, fornecendo aos modelos maior funcionalidade do sdsim, como o acoplamento de modelos, assim como todos os outros recursos disponíveis para a linguagem R.

Trabalhos futuros incluem adição de funcionalidades na aplicação, e.g. definir séries temporais por meio de arquivos e permitir calibração de modelos. Pretende-se hospedar a aplicação em um servidor web, permitindo seu uso em ambiente online, sem necessitar da instalação local do R e do pacote sdsim, e adicionar suporte ao acoplamento de modelos.

AGRADECIMENTOS

Agradecimentos aos membros envolvidos no desenvolvimento deste trabalho e à Embrapa Informática Agropecuária pela oportunidade concedida.

REFERÊNCIAS

ALLEN, J. *Ace Editor Bindings for Shiny*. [S.l.], 2016. Disponível em: <https://CRAN.R-project.org/package=shinyAce>.

CHANG, W. et al. *Web Application Framework for R*. [S.l.], 2017. Disponível em: <https://CRAN.R-project.org/package=shiny>.

CHANG, W.; RIBEIRO, B. B. *Create Dashboards with 'Shiny'*. [S.l.], 2017. Disponível em: <https://CRAN.R-project.org/package=shinydashboard>.

FREUA, M. C. et al. A comparison between three different approaches to omplement a system dynamic model: an assessment by a multidisciplinary team. *Embrapa Informática*

Agropecuária. Boletim de pesquisa e desenvolvimento, 36, 2014. ISSN 1677-9266. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/117682/1/Livro-BolPesq36.pdf>.

JONES, J. W. et al. Brief history of agricultural systems modeling. 2016. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0308521X16301585>.

KNEIS, D. *A Code Generator for ODE-Based Models*. [S.l.], 2017. Disponível em: <https://CRAN.R-project.org/package=rodeo>.

Meteor Development Group. *Meteor: Build Apps with JavaScript*. [S.l.], 2016. Disponível em: <https://www.meteor.com>.

OWEN, J. *Interface to the 'Handsontable.js' Library*. [S.l.], 2016. Disponível em: <https://CRAN.R-project.org/package=rhandsontable>.

PETZOLDT, T.; RINKE, K. simecol: An object-oriented framework for ecological modeling in r. *Journal of Statistical Software*, v. 22, n. 9, p. 1–31, 2007. ISSN 1548-7660. Disponível em: <http://www.jstatsoft.org/v22/i09>.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2017. Disponível em: <https://www.R-project.org>.

SOETAERT, K.; PETZOLDT, T.; SETZER, R. Solving differential equations in r: Package desolve. *Journal of Statistical Software, Articles*, v. 33, n. 9, p. 1–25, 2010. ISSN 1548-7660. Disponível em: <https://www.jstatsoft.org/v033/i09>.



sdsim: Um Pacote para Modelagem e Simulação de Sistemas Dinâmicos Contínuos em R

Cristina Freitas Bazzano[1], *Bruno Henrique Pereira*[1],
Luís Gustavo Barionni[1], *Adauto Luiz Mancini*[1], *Márcio Nicolau*[2]

[1] Embrapa, Av. André Tosello, nº 209 Campus da Unicamp,
Barão Geraldo, Campinas, São Paulo, Brasil
hbrunopereira@gmail.com, crisfbazz@gmail.com,
luis.barioni@embrapa.br, adauto.mancini@embrapa.br

[2]Embrapa, BR 285 Km 294, Passo Fundo, Rio Grande do Sul, Brasil
marcio.nicolau@embrapa.br

RESUMO

Este artigo apresenta o pacote sdsim, um framework livre para definição, prototipação e simulação de modelos de sistemas dinâmicos contínuos em R, e sua aplicação. Por meio de uma abordagem orientada a objetos, o pacote define uma arquitetura genérica e estruturada para facilitar a implementação de simuladores e o compartilhamento de modelos. A integração com interface gráfica simplifica e torna amigável o uso do pacote por usuários com diferentes níveis. Seu suporte inovador ao acoplamento de modelos permite a definição de sistemas complexos que interagem entre si. Como resultado, o sdsim conta com um repositório de modelos que podem auxiliar no desenvolvimento de novos modelos e fomentar o uso do pacote na comunidade de modelagem e simulação em diversas áreas do conhecimento.

PALAVRAS-CHAVE: Equação Diferencial, Programação orientada a objetos, Compartilhamento de código.

ABSTRACT

This paper presents the sdsim package, a free framework for defining, prototyping and simulating continuous system dynamics models in R, and it's application. Through an object-oriented approach, the package defines a generic and structured architecture to facilitate the implementation of simulators and the sharing of models. The integration with a graphical interface simplifies and makes a user friendly environment for different users levels of the package. It's innovative support for coupling models allows the definition of complex systems that interact with each other. As a result, sdsim has a repository of models that can assist in the development

of new models and foster the use of the package in the modeling and simulation community.
KEYWORDS: Differential Equation, Object-oriented programming, Code-sharing.

INTRODUÇÃO

Dinâmica de sistemas ou "System Dynamics" (SD) é um método formal de análise e modelagem de sistemas introduzido por Forrester (1961). Esse método alia rigor e generalidade para representação e simulação de sistemas contínuos complexos (STERMAN, 2000) de forma a permitir melhor avaliação de opções e a engenharia de sistemas mais eficientes. O paradigma de SD assume que a dinâmica é determinada por seu estado presente, i.e. o estado atual reflete toda a informação necessária sobre a trajetória pregressa (passado) do sistema que possa influenciar sua trajetória futura.

Em geral, aplicam-se regras iterativas que definem as mudanças do estado do sistema (JOST, 2005). Em modelos de sistemas dinâmicos (MSDs) contínuos, essa regra matemática é definida por um sistema de equações diferenciais ordinárias (EDOs) de primeira ordem, onde cada equação define a taxa de variação de uma variável de estado. Modelos atômicos são implementados como componentes elementares de um software de simulação. Em contrapartida, modelos acoplados são especificados como estruturas contendo um conjunto de sub-modelos (atômicos ou acoplados) com conexões entre si e com entradas e saídas do modelo acoplado (ZEIGLER; KIM; PRAEHOFER, 2000).

Esse método abriu possibilidades para o surgimento de novos modelos que possuem funções que descrevem as mudanças nos estados do sistema, em resposta a ações externas e outros componentes do sistema. Nos sistemas agropecuários iniciou-se uma nova era de nutrição dinâmica, levando a melhores tipos de modelos, incluindo de culturas, pecuária e agricultura (JONES et al., 2016). A simulação desses modelos predizem, por exemplo, o crescimento diário de uma planta baseado em dados meteorológicos diários e informações do solo, da gestão e da genética da planta (GRAEFF et al., 2012). Algumas dessas informações podem ser obtidas via acoplamento, por exemplo, de modelos de balanço hídrico e de nutrientes no solo. Na pecuária, a simulação também pode predizer o crescimento diário do animal baseado, por exemplo, em uma dieta contínua e na interação com um modelo de crescimento do pasto que pode ser estático, contínuo ou acoplado.

Existe uma miríade de aplicações voltadas à implementação de MSDs com foco em diferentes públicos. Entretanto, diversos critérios definem a ferramenta mais apropriada para o tipo de desenvolvimento e equipe (FREUA et al., 2014). O R, uma linguagem e ambiente de computação científica, destaca-se nesse contexto, por: (a) ser de domínio público; (b) ser compatível com as principais ferramentas do mercado; (c) ser suportado nos principais sistemas operacionais; (d) possuir suporte à entrada, manipulação, análise e visualização de dados; (e) ser uma linguagem de programação funcional com suporte a programação orientada a objetos (OO); (f) possuir elegante suporte à diversos algoritmos para solução de EDOs, por meio do

pacote deSolve (SOETAERT; PETZOLDT; SETZER, 2010), e; (g) ser open-software de natureza, o que facilita o entendimento e a troca de informação entre membros da comunidade (MATLOFF, 2011).

Embora alguns pacotes já se proponham a facilitar a implementação de MSDs em R, tais como o Simecol (PETZOLDT; RINKE, 2007) e o Rodeo (KNEIS, 2017), a implementação de modelos mais complexos, continua sendo dificultado pela rigidez das representações utilizadas atualmente. O pacote sdsim, objetivando suprir as dificuldades encontradas na prototipação e simulação de MSD, utiliza uma abordagem diferenciada, particularmente por ampliar a generalidade, definição e armazenamento dos modelos, suportar acoplamento de modelos e possibilitar a integração com interface gráfica intuitiva construída automaticamente para prototipação e execução de modelos atômicos.

O uso crescente do pacote sdsim pode fomentar a criação de repositórios de modelos na padronização proposta, o que contribuirá para o estudo e aplicação de MSDs às mais diferentes áreas das ciências e da engenharia, inclusive aos sistemas agropecuários.

MATERIAL E MÉTODOS

Com o objetivo de oferecer uma arquitetura genérica para implementação de MSDs contínuos, dentro do paradigma de EDOs de primeira ordem, um design orientado a objetos (OO) foi aplicado. A estrutura dessa arquitetura ajuda na abstração dos simuladores ao encapsular em uma única estrutura de dados toda a especificação de um MSDs (suas equações, variáveis, parâmetros, etc) e possuir métodos que possibilitam a simulação do modelo instanciado.

Utilizou-se a estrutura de classe do pacote R6 (CHANG, 2017) do R, que é, dentre as possibilidades de escrever código OO no R, aquela que melhor atende às necessidades da arquitetura proposta. Suas capacidades comuns a outras linguagens OO, tais como, atributos e métodos públicos e privados (possibilitando encapsulamento de funcionalidades) e herança (superclasses que possibilitam suporte futuro a mais regras matemáticas), apresentam um ótimo desempenho e justificam sua escolha.

As classe criadas com o R6, e descritas abaixo, padronizam a implementação de simuladores ao encapsular toda a informação necessária para simular um MSD em um único objeto, e dão suporte ao acoplamento de modelos. O compartilhamento desses objetos promove a distribuição e reuso de simuladores de forma simples e compacta, permitindo que modeladores usem o sdsim como um meio de comunicação na implementação de MSDs.

Para oferecer usabilidade na prototipação, simulação e reuso de modelos de sistemas dinâmicos, uma interface gráfica de usuário (GUI) foi construída para o pacote. Nessa aplicação todas as funcionalidades disponíveis para modelos atômicos são apresentadas de maneira visual e clara, o que facilita a interação com todos os tipos de usuários, acelerando a curva de aprendizagem e contribuindo para a quebra da barreira de acesso às novas tecnologias sobre modelagem e simulação.

Além da estrutura para armazenamento dos MSDs também são necessárias rotinas matemáticas para integração temporal das EDOs que modulam o sistema e possibilitam sua simulação. Tais rotinas são implementadas por diversos algoritmos do pacote R deSolve, usado neste projeto para resolver os problemas de valor inicial de um sistemas de ODEs de primeira ordem. As funções do deSolve também oferecem suporte a eventos (mudanças temporais nas variáveis do sistema quando desejado) e permitem o uso de código compilado para um melhor desempenho na rotina de integração.

Classe SDModel

A classe *SDModel* pode armazenar em seus membros toda a especificação de um MSD atômico, tais como as variáveis de entrada (e.g. parâmetros, constantes, etc.), as variáveis de estado, as equações diferenciais e auxiliares, funções opcionais de inicialização e pós processamento, e funções opcionais associadas ao disparo e execução de eventos. A função de inicialização pode ser utilizada, por exemplo, para uma inicialização customizada de variáveis de estado e parâmetros do sistema. As funções associadas a eventos podem ser utilizadas para determinar em que condições um evento ocorre e qual a resposta a esse evento, como, por exemplo, alterações abruptas no valor das variáveis de estado do sistema. A rotina de pós-processamento pode ser utilizada para processar estatisticamente os dados de trajetória das variáveis simuladas ou calcular novas variáveis e indicadores.

Um objeto *SDModel* também contém métodos para simulação de um MSD e para visualização do resultado. Na integração do sistema a regra matemática utilizada é a função que calcula as equações diferenciais e o estado atual do sistema é informado nas variáveis de estado. As equações auxiliares são pré avaliadas a cada passo da simulação para serem utilizadas, juntamente com as variáveis de entrada e o estado atual, no cálculo das EDOs do sistema.

No resultado da simulação também são armazenadas, caso existam, as trajetórias das variáveis auxiliares e das séries temporais do modelo, possibilitando a visualização e checagem de resultados intermediários da simulação.

Criar um objeto SDModel

No pacote sdsim, foi implementada a seguinte função construtora para instanciar um objeto *SDModel*: *CreateSDModel(Nome do Modelo Atômico, Função que Calcula as EDOs do Modelo, Função de Inicialização, Função de Pós Processamento, Função que Dispara Eventos, Função que Executa Eventos)*. Esta função cria um objeto *SDModel* com identificação *Nome do Modelo Atômico* que deve especificar pelo menos a *Função que Calcula as EDOs do Modelo*. A especificação das equações do sistema é feita via funções R, o que requer um conhecimento básico da linguagem de script. No script, as variáveis do modelo são inicializadas pelo método *InitiModel* do objeto *SDModel* criado e são informados via dados em memória ou em arquivos. As variáveis podem estar em arquivos texto dentro de um diretório ou em um único arquivo

EXCEL, onde cada arquivo ou aba representa um tipo de variável do sistema.

Caso existam variáveis de entrada que sejam definidas por séries temporais, essas podem ser convertidas automaticamente em funções unárias do tempo. A interpolação das variáveis temporais foi implementada utilizando algoritmos do pacote *stats*, os quais recebem como argumentos o método de interpolação (degrau, linear ou spline), o tempo e os valores da variável para cada instante de tempo, e retornam a função unária correspondente.

A criação e simulação de um objeto *SDModel* pode ser feita, de maneira mais interativa, por meio da GUI implementada utilizando o R Shiny (CHANG et al., 2017) e exportada com o pacote sdsim. Essa interface gráfica, mostra de forma visual os campos necessários para especificar, inicializar e simular um MSD. Existem campos para editar os parâmetros do modelo e campos para inserir o script das funções R necessárias na definição do modelo.

Funcionalidades de um objeto SDModel

Com um objeto *SDModel* inicializado é possível executar a simulação. Para tanto, foi implementado o seguinte método: *Run(Método de Integração Temporal das EDOs, Sequência de Tempo, Variáveis de Estado, Variáveis de Entrada, Booleanas para Detecção de Eventos, Inicialização e Checagem de Erros)*. Este método irá utilizar as variáveis padrão inicializadas no objeto, mescladas com as variáveis passadas como argumento que se queira adicionar ou alterar na lista padrão, para integrar a função que calcula as EDOs do sistema ao longo da *Sequência de Tempo*. O *Método de Integração Temporal das EDOs* define o algoritmo do pacote deSolve que será utilizado na simulação. Os demais parâmetros desse método auxiliam no controle e execução de eventos, inicialização de variáveis e checagem de erros adicional.

Caso o objeto *SDModel* possua a função para detecção de eventos, quando a condição do evento é satisfeita, a função de disparo de eventos é executada. Caso seja especificada a função para a detecção mas não a função de disparo, a simulação é interrompida no primeiro evento encontrado.

Para plotar o resultado da simulação de um MSD foi implementado o método *Plot* na classe do objeto *SDModel*. Esse método é capaz de plotar a trajetória de todas as variáveis de estado e, caso existam, as variáveis auxiliares e as séries temporais. Os gráficos gerados podem correlacionar as variáveis ou ser em função do tempo da simulação, seguindo as configurações de plot armazenadas no objeto. Se nenhuma configuração é definida, apenas as variáveis de estado são plotadas em função do tempo.

Um objeto *SDModel* também possui métodos para checar a descrição e unidade das variáveis, para recuperar informações sobre a última simulação (e.g. método de integração utilizado, tempo de simulação, etc.), para salvar em arquivos de texto o resultado da última simulação executada com seus respectivos dados de entrada e para limpar o objeto de todos os dados referentes a última simulação executada (e.g. as trajetórias das variáveis).

Classe CoupledSDAssembler

A classe *CoupledSDAssembler* utilizada para estruturar a criação de simuladores de MSDs acoplados (MSDAs) também foi especificada usando a estrutura de classe do pacote R6. Essa classe representa um MSDA composto por objetos *SDModel* inicializados e uma lista de conexões que definem a comunicação entre os modelos, i.e. a atribuição de valores de variáveis de estado de um modelo para outro(s) modelo(s). As conexões possibilitam especificar laços de feedback de informação do sistema modelado. Dessa forma, modelos atômicos existentes podem ser reutilizados na definição de outros sistemas mais complexos.

A solução numérica do modelo acoplado é feita executando-se todas as conexões e em seguida calculando-se as diferenciais de todos os modelos componentes antes de cada transição de estado. Isso evita o efeito colateral de assincronia entre componentes do sistema, o que poderia gerar erros na simulação.

Criar um objeto CoupledSDAssembler

Foi implementada a seguinte função construtora para classe *CoupledSDAssembler*: *CreateCoupledSDModel(Nome do Modelo Acoplado, Lista de Objetos SDModels, Lista de Conexões)*. Esta função cria um objeto *CoupledSDAssembler* com identificação *Nome do Modelo Acoplado* que possui como componentes a *Lista de Objetos SDModels*, inicializados e identificados pelo *Nome do Modelo Atômico* de cada objeto, e a *Lista de Conexões* que definem a comunicação entre os modelos.

Os métodos *AddModels*, *AddConnections*, *RemoveModels* e *RemoveConnections* de um objeto da classe CoupledSDAssembler podem auxiliar na reestruturação de modelos acoplados, possibilitando o acoplamento ou a remoção de modelos do sistema complexo, sempre mantendo a lista de conexões consistente com os modelos componentes.

Funcionalidades de um objeto CoupledSDAssembler

Na classe *CoupledSDAssembler* o método *Run(Método de Integração Temporal das EDOs, Sequência de Tempo, Booleanas para Detecção e Configuração de Eventos)* executa a simulação de um MSDA. Este método irá utilizar as variáveis padrão inicializadas nos objetos *SDModel* acoplados para integrar a função que calcula as EDOs de todos os modelos componentes simultaneamente ao longo da *Sequência de Tempo*. O *Método de Integração Temporal das EDOs* define o algoritmo do pacote deSolve que será utilizado na simulação. Os demais parâmetros desse método auxiliam no controle e execução de eventos.

A classe *CoupledSDAssembler* também possui suporte a eventos e visualização gráfica dos resultados semelhante à classe *SDModel*.

Tratamento de erros no pacote sdsim

A função *tryCatch* do R (MATLOFF, 2011) é utilizada ao longo do pacote sdsim para possibilitar a detecção de avisos e erros, customização das mensagens exibidas e controle do valor de retorno.

Ademais, as classes *SDModel* e *CoupledSDAssembler* possuem uma validação das EDOs que definem o sistema. Essa validação checa por valores inválidos (e.g. resultados nulos, ausentes ou infinitos) gerados na execução do primeiro passo da simulação e que podem decorrer de erros na especificação do modelo. O usuário recebe um aviso para cada variável com valor inválido encontrada na checagem, o que pode indicar que o modelo deva ser refatorado.

RESULTADOS E DISCUSSÃO

Uma versão preliminar do pacote sdsim pode ser obtida no seguinte endereço eletrônico: `https://goo.gl/SQeUTu`. Essa versão já conta algumas das melhorias previstas para o pacote e não necessariamente corresponde a versão que foi utilizada no exemplo da Figura 1.

A Figura 1 exemplifica o uso deste pacote em script para implementação de um MSDA que representa a interação do modelo simplificado de crescimento foliar de pastagens (CFP) com um modelo simplificado de quantidade de água disponível no solo (AS).

```
library(sdsim)

## MODELO DE CRESCIMENTO FOLIAR DE PASTAGENS (CFP)
# Q Radiação Solar
# L Massa de folhas
# SLA Área especifica da folha
# LLS Vida útil da folha (dias)
# fSW Absorção de água pelo solo
# fWRatio Razão entre estoque e capacidade máxima de água no solo
# fLPartition Partição de Nutrientes
auxCpf <- list(
  fLAI =  "state[['L']] * vars$par$SLA",        ## Índice da Área da Folha
  fLIF =  "1 - exp(-vars$par$k * auxSfLAI)",    ## Índice de Intercepção de Luz
  fRADint =  "timeSeries$Q * auxSfLIF",         ## Radiação Interceptada
  fWRatio = "min( (vars$inputs$W / vars$par$Wmax), 1)",
  fSW =  "1 / (1 + ((1 - auxSfWRatio) / vars$par$SWconst) ^ vars$par$SWpower)",
  fGPP =  "vars$par$epsilon * auxSfSW * auxSfRADint", ## Produção Primária Bruta
  fNPP =  "auxSfGPP * vars$par$y",                 ## Produção Primária liquida
  fLPartition =  "(vars$par$Lmax - state[['L']]) / vars$par$Lmax")

# CFP Model Definition
dCpf <- function(t, state, vars, timeSeries, aux)
{
  dL <- aux$fNPP * aux$fLPartition - state[['L']] * (1/vars$par$LLS)
  return(list(c(dL = dL)))
}

# Variáveis do modelo em planilha excel
cpfInputConfig <- list(defaultInputPath = "CFP.xlsx", timeSeriesDirectory = ".")
# Cria e inicializa o modelo CFP
cfp <- CreateSDModel("CFP", ModelDefinition = dCpf)
cfp$InitModel(inputConfig = cpfInputConfig, aux = auxCpf,
              timeS = seq(from = 1, to = 100, by = 1), method = "rk4")
```

```
## MODELO DE QUANTIDADE DE ÁGUA DISPONÍVEL NO SOLO (AS)
# ETo Evaporação do Tanque Classe A
# SSHC Condutividade hidráulica saturada do solo
# fPP Quantidade de Água Drenada do Solo Saturado
# fCanCond Condução Canopy
# fETR Evapotranspiração Foliar
auxAs <- list(
  fCanCond = "vars$par$MaxCond * min(1, vars$inputs$L / vars$par$Lmax)",
  fETR = "vars$inputs$ETo * auxSfCanCond * state[['W']] / vars$par$Wmax",
  fPP = "max(timeSeries$Rain + state[['W']] - vars$par$Wmax, 0)*vars$par$SSHC")

# AS Model Definition
dAs <- function(t, state, vars, timeSeries, aux)
{
  dW <- timeSeries$Rain + vars$inputs$Irrig - auxSfETR - auxSfPP
  return(list(c(dW)))
}

# Variáveis do modelo em planilha excel
asInputConfig <- list(defaultInputPath = "AS.xlsx", timeSeriesDirectory = ".")
# Cria e inicializa o modelo CFP
as <- CreateSDModel("AS", ModelDefinition = dAs)
as$InitModel(inputConfig = asInputConfig, aux = auxAs,
             timeS = seq(from = 0, to = 100, by = 1), method = "rk4")

## MODELO ACOPLADO DO CFP COM O AS
connectionSimpleGrowth <- list(c("CFP", "W", "AS", "W"),
                               c("AS", "L", "CFP", "L"))
sg <- CreateCoupledSDModel("SimpleGrowth", models = list(cfp, as),
                           connections = connectionSimpleGrowth)
sg$Run()
sg$Plot()
```

Figura 1: Implementação do Modelo Acoplado de Crescimento Foliar de Pastagens Simplificado

Os modelos componentes foram criados com a classe *SDModel*, inicializados com dados em planilhas e acoplados utilizando a classe *CoupledSDAssembler*. A taxa de crescimento da massa foliar (L) do modelo CFP é influenciada positivamente pela radiação solar diária, pela partição de nutrientes disponíveis e pela quantidade de água disponível no solo (W) dada pelo acoplamento, e é afetado negativamente pelo decaimento das folhas. A taxa de crescimento da água disponível no solo (W) do modelo AS é influenciada positivamente pela chuva diária

e pela irrigação aplicada, e é afetada negativamente pela evapotranspiração foliar da pastagem (em função de L) acoplada e pela drenagem do solo saturado.

O resultado da simulação do modelo acoplado presente na Figura 2 mostra que uma maior quantidade de massa foliar implica em uma maior evapotranspiração e consequente queda da quantidade de água disponível no solo. Essa queda da água implica em uma queda no crescimento foliar e assim segue a interação entre os componentes.

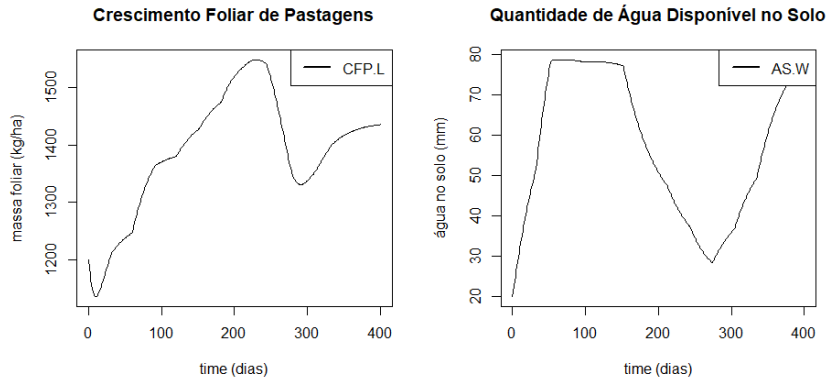


Figura 2: Resultado da Simulação do Modelo Acoplado da Figura 1

O pacote sdsim conta com um repositório de modelos implementados e prontos para serem executados. Estes modelos podem ser carregados com a função *data* e permitem que os usuários do R possam se familiarizar com o uso do pacote.

CONCLUSÕES

A arquitetura proposta pelo pacote sdsim destina-se a solucionar a estruturação da definição de MSD atômicos no R. Essa arquitetura com estrutura OO incorpora dados e funções em um único objeto *SDModel* capaz de modelar a representação de sistemas dinâmicos baseados em EDOs de primeira ordem, permitindo a padronização generalizada tão desejada. Este padrão acelera o processo de desenvolvimento de simuladores por apresentar a estrutura que o modelista deve seguir para definir e armazenar seus modelos.

Os simuladores criados usando o sdsim podem ser salvos em formato *Rdata* com a função *save* do R, a função padrão para salvar objetos R, para serem armazenados, por exemplo, em repositórios compartilhados. O entendimento e uso da estrutura proposta auxilia na compreensão de simuladores de MSD de terceiros, facilitando a distribuição e reuso dos mesmos. Estes modelos podem então ser executados e modificados de acordo com a necessidade de cada usuário.

A integração do pacote com uma GUI construída automaticamente para prototipação e execução de modelos atômicos contribui para a simplicidade e facilidade de uso do pacote por programadores pouco experientes. Essa interface poderá auxiliar na difusão de conhecimentos sobre modelagem e simulação a todos os níveis de usuários.

Ao nosso conhecimento o sdsim é o primeiro pacote R a suportar formalmente o acoplamento de modelos por meio de uma abordagem orientada a objetos. Modelos acoplados são objetos *CoupleSDAssembler* contendo um conjunto de modelos atômicos componentes (objetos *SDModel's*) com conexões entre si. Dessa forma o entendimento do padrão adotado para modelos atômicos é suficiente para criação de modelos acoplados e, portanto, estes também usufruem da facilidade de definição e compartilhamento que a estrutura OO confere.

O acoplamento de modelos permite a definição de modelos mais complexos que interagem entre si. Dessa forma se torna viável, por exemplo, a construção de um modelo de crescimento de cultura acoplado com modelos de doenças que atacam essa plantação. A simulação deste modelo reproduz a interação entre seus componentes e auxilia na obtenção de informações para estimar os impactos no crescimento e no rendimento da cultura.

Na nova fase de desenvolvimento do sdsim os conceitos de simulador, modelo e cenário (valor das variáveis do sistema para um ambiente específico) serão desagregados para dar lugar a uma nova estrutura de classes. Essa estrutura possuirá novas classes para definir um modelo atômico, um modelo acoplado e um cenário, e possuirá uma função para executar a simulação de um modelo em um dado cenário. O resultado da simulação será armazenado em um novo objeto com funcionalidades para analise e visualização das trajetórias obtidas.

Futuramente, pretende-se estender a funcionalidade do pacote de forma a incluir métodos para calibração, intercomparação, otimização, filtragem e outras análises para MSDs atômicos e acoplados, nas classes *SDModel* e *CoupledSDAssembler*, respectivamente. Além disso, uma GUI com suporte a modelos acoplados e a todas as futuras funcionalidades pode diferenciar a usabilidade do pacote.

O fato da classe R6 possuir herança, possibilita que outros paradigmas sejam utilizados na regra matemática que integra o sistema. Dessa forma, em versões futuras do pacote, além de EDOs, pode-se incluir outras famílias de modelos (e.g. estatísticas, eventos discretos, baseado em indivíduos, espacialmente explícitos) aplicadas individualmente ou acopladas.

Seguindo a filosofia open-software do R, o pacote sdsim pretende oferecer um framework livre para definição e simulação de MSDs atômicos e acoplados. O código completo do pacote será publicado sob uma licença que garantirá uso irrestrito nas diversas áreas do conhecimento, e possivelmente a inclusão de modelos feitos por terceiros.

AGRADECIMENTOS

Agradecemos aos membros da equipe da Embrapa Informática Agropecuária e a todos os outros colaboradores que empenharam esforços no desenvolvimento do pacote sdsim, tornando este artigo possível.

REFERÊNCIAS

CHANG, W. R6: Classes with reference semantics. 2017. Disponível em: <https://cran.r-

project.org/package=R6>.

CHANG, W. et al. Web application framework for r. 2017. Disponível em: <https://CRAN.R-project.org/package=shiny>.

FORRESTER, J. W. *Industrial Dynamics*. 1st. ed. [S.l.]: Pegasus Communications, 1961. ISBN 1614275335.

FREUA, M. C. et al. A comparison between three different approaches to omplement a system dynamic model: an assessment by a multidisciplinary team. *Embrapa Informática Agropecuária. Boletim de pesquisa e desenvolvimento, 36*, 2014. ISSN 1677-9266. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/117682/1/Livro-BolPesq36.pdf>.

GRAEFF, S. et al. *Crop Models as Decision Support Systems in Crop Production*. 1st. ed. Crop Production Technologies, 2012. ISBN 978953307781. Disponível em: <http://www.intechopen.com/books/crop-production-technologies/crop-modelsas-decision-support-systems-in-crop-production>.

JONES, J. W. et al. Brief history of agricultural systems modeling. 2016. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0308521X16301585>.

JOST, J. *Dynamical systems*: Examples of complex behaviour. 1st. ed. [S.l.]: Springer, 2005. ISBN 9783540229087.

KNEIS, D. A code generator for ode-based models. 2017. Disponível em: <https://CRAN.R-project.org/package=rodeo>.

MATLOFF, N. *The art of R programming*: A tour of statistical software design. 1st. ed. [S.l.]: No Starch Press, 2011. ISBN 9781593273842.

PETZOLDT, T.; RINKE, K. simecol: An object-oriented framework for ecological modeling in r. *Journal of Statistical Software*, v. 22, n. 9, p. 1–31, 2007. ISSN 1548-7660. Disponível em: <http://www.jstatsoft.org/v22/i09>.

SOETAERT, K.; PETZOLDT, T.; SETZER, R. Solving differential equations in r: Package desolve. *Journal of Statistical Software, Articles*, v. 33, n. 9, p. 1–25, 2010. ISSN 1548-7660. Disponível em: <https://www.jstatsoft.org/v033/i09>.

STERMAN, J. D. *Business dynamics*: systems thinking and modeling for a complex world. 1st. ed. [S.l.]: McGraw-Hill Higher Education, 2000. ISBN 0072311355.

ZEIGLER, B.; KIM, T.; PRAEHOFER, H. *Theory of Modeling and Simulation*. 2nd. ed. [S.l.]: Academic Press, 2000. ISBN 9780127784557.



Análise de protocolos de roteamento em Redes de Sensores sem Fio para Monitoramento Bovino

João Felipe Resende Nacer[1], *Irineu Sotoma*[2]

[1]Instituto Federal de Mato Grosso do Sul
Campo Grande, Mato Grosso do Sul, Brasil
joao.nacer@ifms.edu.br

[2]Faculdade de Computação, Universidade Federal de Mato Grosso do Sul
Campo Grande, Mato Grosso do Sul, Brasil
isotoma@facom.ufms.br

RESUMO

Este trabalho apresenta uma RSSF projetada para monitorar posições de bovinos em tempo real. Nesta RSSF, cada bovino possui um colar com sensor GPS e um chip XBee, que utiliza o mesmo protocolo de roteamento para transmitir as informações até que estas alcancem um nó sorvedouro. Em uma estação base, as posições recentes dos bois, armazenadas no nó sorvedouro, podem ser visualisadas pelo pecuarista. Após um experimento na RSSF e simulações, concluiu-se que entre vários protocolos de roteamento para RSSF, aqueles disponibilizados pelos chips XBee têm bom desempenho nas métricas de taxa de entrega, consumo de energia e latência.

PALAVRAS-CHAVE: pecuária de precisão; redes de sensores sem fio; protocolos de roteamento; zigbee, xbee.

ABSTRACT

This paper presents a wireless sensor network (WSN) developed to monitor bovine localizations in real time. In this WSN, each bovine has a collar with GPS sensor and XBee chip, which utilizes the same routing protocol to send information until information reach a sink node. At a base station, the recent bovine localizations stored in sink node can be viewed by a farmer. After an experiment in the WSN and simulations, the conclusion is, among several routing protocols for WSNs, those available in XBee chips provide good performance in the metrics delivery rate, energy consumption and latency.

KEYWORDS: precision livestock farming; wireless sensor networks; routing protocols; zigbee, xbee.

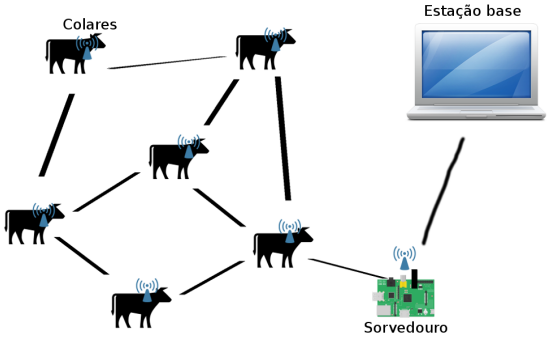
INTRODUÇÃO

A pecuária de precisão, no contexto dos ecossistemas pastoris, é a forma moderna de gerenciar os sistemas de produção animal a pasto. Consiste na medição de diferentes parâmetros dos animais, a modelagem desses dados para selecionar a informação que se quer, e o uso desses modelos em tempo real visando o monitoramento e o controle de animais e rebanhos (BERCKMANS, 2004).

Recentemente, Jesus (2014) desenvolveu um sistema de monitoramento bovino baseado nas posições dos animais no pasto. Nesse sistema, as posições dos bois são coletadas em um colar desenvolvido para cada animal composto por um sensor de GPS (*Global Positioning System* - Sistema de Posição Global) e um cartão de memória, para o armazenamento das informações. Ao final de cada experimento as informações são coletadas manualmente para análises do comportamento dos animais. O sistema pode detectar 4 comportamentos dos animais: comendo, andando, em pé e deitado.

Lomba (2015) complementou o trabalho de Jesus (2014) ao incluir novos sensores de luminosidade e acelerômetro no colar para identificar o comportamento bovino por meio da aceleração do animal e monitoramento da luminosidade do ambiente.

Figura 1: Imagem ilustrativa da organização da RSSF



Fonte: Elaborada pelo autor

Este artigo apresenta a implementação de uma Rede de Sensores Sem Fio (RSSF) (AKYILDIZ et al., 2002) de modo a melhorar o tempo de entrega entre a coleta das informações dos colares até sua entrega ao pecuarista, tornando esse processo automático, ao invés de manual. A Figura 1 apresenta a visão geral da RSSF, em que cada bovino possui um nó sensor, que é o colar desenvolvido por Jesus (2014) estendido com um chip XBee, para transmissão das informações entre os colares até alcançar o nó sorvedouro desenvolvido neste projeto. As informações do nó sorvedouro podem então ser acessadas pelo pecuarista a partir de uma estação base.

Em uma RSSF, para que os dados monitorados pelos nós sensores sejam enviados até o nó sorvedouro, é necessário que sejam estabelecidas rotas entre a origem e o destino. Os protocolos que têm como objetivo determinar uma rota entre a origem e o destino são chamados protocolos de roteamento.

Para que o pecuarista possa ter uma RSSF com longo tempo de vida e que forneça as informações dos bois rapidamente, a escolha do protocolo de roteamento utilizado é fundamental. Os autores não encontraram na literatura uma comparação de vários protocolos de roteamento para redes de sensores sem fio no contexto de monitoramento bovino. Portanto, este trabalho comparou vários protocolos de roteamento e concluiu que os protocolos disponibilizados pelos chips XBee utilizados fornecem resultados satisfatórios equivalentes a outros protocolos de roteamento existentes. Outras questões sobre a melhoria do roteamento também são apresentadas.

MATERIAL E MÉTODOS

Metodologia

A infraestrutura da RSSF desenvolvida depende dos colares, do nó sorvedouro e da estação base. A escolha dos protocolos de roteamento depende de estudos de artigos da literatura e das especificações do chip XBee utilizado nos colares.

Os estudos dos protocolos de roteamento encontrados na literatura e nas especificações do chip Xbee foram realizados paralelamente ao desenvolvimento dos colares e do nó sorvedouro. A seção *Protocolos de roteamento* detalha os protocolos de roteamento avaliados.

A extensão do colar desenvolvido por Jesus (2014) para permitir a comunicação entre os colares é apresentada na seção *Materiais utilizados nos colares*. O desenvolvimento do nó sorvedouro está detalhado na seção *Materiais utilizados no Nó sorvedouro*. A estação base utilizada está descrita na seção *Estação base*.

Uma vez que os colares, o nó sorvedouro e a estação base ficaram prontos, realizou-se um experimento com 8 bois para coletar as posições reais obtidas por meio de GPS. A seção *Experimento* descreve como o experimento foi realizado.

Por fim, para avaliar os protocolos de roteamento utilizou-se um simulador e as posições reais coletadas pelo experimento. A seção *Simulações* detalha o simulador utilizado nas simulações.

Protocolos de roteamento

No contexto de monitoramento animal, para determinar o melhor algoritmo de roteamento para este projeto, vários protocolos existentes foram estudados, e alguns foram selecionados para o trabalho considerando algumas particularidades, tais como: uso de GPS para construção de rotas (método geográfico), formação de rotas de forma reativa ou proativa, opção de mobilidade dos nós, e outros. Os protocolos utilizados nas simulações são listados a seguir com uma descrição breve e justificativa da escolha:

- *Flooding* (AKYILDIZ et al., 2002): protocolo em que os nós sensores realizam o *broadcast* das informações coletadas para todos os nós dentro do alcance de transmissão. Nesse protocolo não há a preocupação com algoritmos de roteamento e manutenção de rotas.

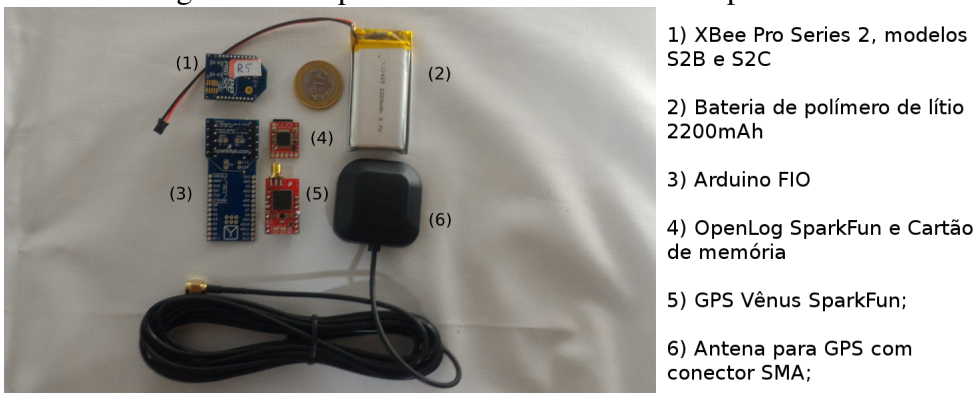
No *flooding*, cada nó que recebe um pacote de dados repete o envio por *broadcast*, até que o número máximo de saltos do pacote seja alcançado ou que o destino de um pacote seja alcançado. O *Flooding* foi escolhido para simular a propagação de pacotes na RSSF por meio de *broadcast*.

- *Gossiping* (AKKAYA; YOUNIS, 2005): o protocolo *Gossiping* é uma versão levemente melhorada do *Flooding*. Nesse protocolo ao invés de um sensor enviar dados para todos os vizinhos até chegar ao nó de destino, um nó escolhe aleatoriamente um dos vizinhos para enviar o pacote e esse processo se repete até que o protocolo alcance o nó sorvedouro. Uma desvantagem dessa abordagem é que como a escolha é aleatória, o nó pode enviar para um vizinho que está mais distante do destino. O *Gossiping* foi escolhido para simular envio de dados de forma aleatória na rede.
- AODV (*Ad-hoc On-Demand Distance Vector Routing*): o AODV (PERKINS; BELDING-ROYER; DAS, 2003) é um protocolo reativo, ou seja, cada nó constrói uma rota até o nó sorvedouro apenas quando precisa enviar informações. Sempre que esse nó precisa enviar uma informação a algum destino no qual não possui na sua tabela de destinos, inicia-se um novo processo de descobrimento de rota. Este protocolo foi escolhido para estudo por ser base da implementação do protocolo ZigBee, que é o padrão utilizado no rádio utilizado no trabalho.
- HGR (*Hybrid Geographic Routing for Flexible Energy*) (CHEN et al., 2009): o HGR é um protocolo geográfico híbrido. Neste protocolo, os sensores escolhem o nó transmissor baseados em sua posição e distância do nó sorvedouro. O HGR foi escolhido por ser um protocolo geográfico e une duas características utilizadas em algoritmos que utilizam a posição dos nós para a escolha do próximo salto, a distância e a direção.
- IRP (*Implicit Routing Protocol*) (KWONG et al., 2012): é um protocolo que foi desenvolvido com o propósito de monitorar o comportamento bovino. Esse protocolo possui um algoritmo de simples implementação baseado no número de saltos do nó sorvedouro até os nós sensores. O protocolo é dividido em duas fases. A primeira fase é o envio de uma mensagem em *broadcast* pelo nó sorvedouro com a ID do sorvedouro e uma contagem de saltos iniciando em 0. A cada salto a contagem é incrementada. Dessa forma, quanto mais saltos de distância do sorvedouro, maior será essa contagem de saltos, e o número de saltos é gravado em cada nó sensor. A segunda fase é a fase de envio de dados do nó sensor ao sorvedouro. Quando um nó sensor precisa enviar um pacote de dados, este envia o pacote de dados juntamente com o número de saltos. Todos os nós que receberem o pacote e possuirem o número de saltos maior ou igual irão descartar o pacote. Os nós, que receberem o pacote e tiverem número de saltos menor que o do sensor que está enviando, reencaminharão a mensagem. Esse processo é repetido até a informação alcançar o nó sorvedouro.
- *Many-to-one*: semelhante ao IRP, o protocolo *Many-to-one* ocorre em duas fases. Na operação de descoberta de rota *Many-to-one*, a rota para o sorvedouro é estabelecida em

todos os dispositivos: O sorvedouro transmite uma mensagem de criação de rota *Many-to-one* tendo ele mesmo como o endereço de destino. Os módulos XBee que recebem esta mensagem armazenam uma entrada de tabela de roteamento *Many-to-one* reversa para criar um caminho de volta para o sorvedouro. O pacote de configuração da rede *Many-to-one* deve ser enviado periodicamente para atualizar as rotas reversas na rede. Este é outro protocolo escolhido por ser possível de uso no XBee utilizado no projeto.

- *History-based protocol* (JUANG et al., 2002): o protocolo utiliza um contador que é incrementado cada vez que algum sensor faz uma transmissão direta ao nó sorvedouro. Desse modo, quanto maior o número do contador, maior a probabilidade do sensor estar mais próximo ao sorvedouro. Cada vez que um nó procura por nós vizinhos ele solicita o contador de cada um dos vizinhos e depois envia os pacotes ao vizinho com o contador de maior número e isso é repetido até a informação encontrar o nó sorvedouro. Este protocolo foi desenvolvido inicialmente para ser utilizado no projeto ZebraNet, que é um projeto de monitoramento de Zebras com RSSF. Seu objetivo era compreender alguns padrões de comportamento desses animais. Os sensores podiam detectar a posição pelo GPS e outros dados como frequência cardíaca e temperatura corporal. O protocolo *History-based* e o IRP foram escolhidos para o estudo pois foram utilizados previamente em pesquisas com monitoramento de animais.

Figura 2: Componentes do colar com moeda para escala



Fonte: Elaborada pelo autor

Materiais utilizados nos colares

Ao colar desenvolvido por Jesus (2014) foram adicionados os rádios XBee modelo PRO Series 2. Os nós sensores utilizados no trabalho são compostos pelos itens indicados na Figura 2.

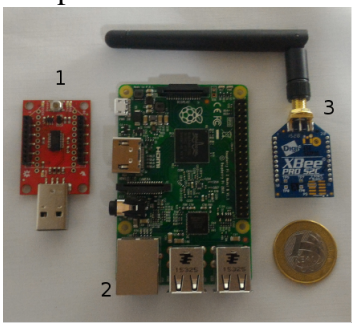
Materiais utilizados no Nó sorvedouro

A topologia da rede escolhida foi a topologia em malha, devido a mobilidade dos bois. Todos os nós foram configurados para atuar como roteadores da rede e enviando as mensagens de posição para o nó sorvedouro. O nó sorvedouro é composto pelo itens indicados na Figura 3.

O nó sorvedouro é o nó coordenador da rede, responsável por receber e tratar as informações enviadas pelos sensores. Este nó foi desenvolvido utilizando um XBee RPSMA, que permite adicionar uma antena externa, acoplado a um Raspberry PI 2 (Modelo B). Por ser o coordenador da rede, este nó deve estar ligado todo o período para receber os dados enviados pelos colares. O Raspberry foi ligado diretamente à rede elétrica por meio de uma fonte de energia e ligado à rede de dados da Embrapa Gado de Corte por meio de cabo Ethernet durante os experimentos realizados.

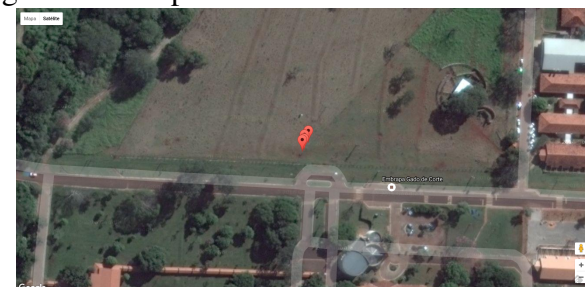
Neste projeto, o Raspberry foi programado para receber todas as mensagens do XBee em uma API desenvolvida em Java e disponibilizada pela fabricante dos módulos XBee (DIGI, 2016). Um banco de dados em PostgreSQL foi criado no nó sorvedouro para o registro dessas posições. A cada posição recebida pelo sorvedouro, cria-se um novo registro de posição com a identificação do colar. Essas posições gravadas podem ser acessadas por um serviço web.

Figura 3: Componentes do nó sorvedouro com moeda para escala



Fonte: Elaborada pelo autor

Figura 4: Mapa com posição mais atual de alguns bois no pasto



Fonte: Elaborada pelo autor

Figura 5: Área do experimento



Fonte: Elaborada pelo autor

Estação base

Um notebook funciona como estação base e via um navegador web, acessa-se o serviço web do nó sorvedouro pela rede da Embrapa, para visualizar o mapa do local das coletas, e a cada recebimento de um pacote pelo sorvedouro o mapa era atualizado com as posições dos bois (Figura 4). Cada bovino é representado por um ponto no mapa do Google Maps na posição atual e é possível acompanhar pelo navegador as posições atuais dos bois.

Experimento

O experimento com 8 colares foi realizado na Embrapa Gado de Corte para a obtenção das posições de GPS. Cada coleta de posição dos bois era realizada a cada 5 minutos, a escolha do intervalo foi baseada no trabalho de Mezzalira et al. (2011) que considera esta frequência importante para estudos de comportamento do animal. Cada registro composto com os seguintes dados: id do colar, latitude, longitude, altitude, data, hora, número de satélites, timestamp. O experimento de coleta dos dados iniciou no dia 17/10/2016 no espaço chamado Mangueiro Digital (Figura 5), na Embrapa Gado de Corte, e a remoção dos os colares foi concluída no dia 24/10/2016.

Simulações

Os protocolos de roteamento descritos na seção *Protocolos de Roteamento* foram implementados no simulador Castalia (BOULIS et al., 2011). O simulador Castalia é uma extensão do simulador OMNeT++ e foi escolhido entre outros devido ao suporte de modelos mais realísticos de rádio e de canal de comunicação sem fio. As posições de GPS coletadas no experimento na Embrapa foram pré-processadas e inseridas no simulador para a comparação dos protocolos de roteamento com as posições reais dos bois.

O protocolo MAC utilizado nas simulações foi o que mais se aproximava ao CSMA-CA, utilizado pelo XBee. Nas simulações, a camada MAC não realizou tentativas de retransmissão de pacotes, de modo que a camada MAC tentava enviar apenas uma vez um pacote de um nó sensor a outro nó sensor ou de um nó sensor ao nó sorvedouro.

As simulações em Castalia avaliaram os protocolos de roteamento em relação às métricas de consumo de energia, taxa de recebimento dos pacotes de dados e latência de entrega. Essas métricas possibilitam escolher protocolos de roteamento que atendam aos requisitos que um pecuarista gostaria em um sistema de monitoramento bovino, como durabilidade da rede de sensores sem fio (consumo de energia), recebimento das informações dos bois (taxa de entrega de pacotes), e rapidez na visualização das informações dos bois (latência de entrega de pacotes).

RESULTADOS E DISCUSSÃO

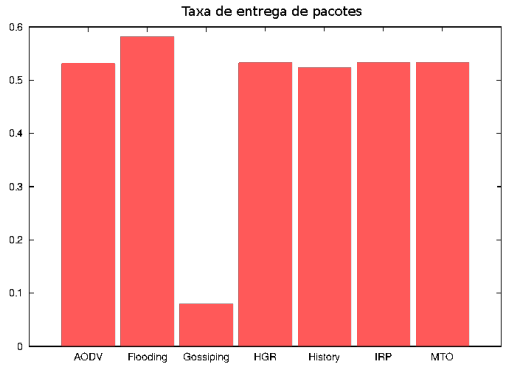
Esta seção apresenta e discute os resultados da simulações. O gráfico da Figura 6 apresenta a taxa de pacotes entregues ao sorvedouro por protocolo. O quantitativo apresentado no gráfico é dado pela razão dos pacotes de dados não duplicados recebidos na camada de aplicação do nó sorvedouro pela quantidade de pacotes criados pelos nós sensores.

Devido a característica de mobilidade dos bovinos, de seguirem um líder e andar em grupos, em alguns momentos do experimento os bovinos se afastaram do nó sorvedouro o que ocasionou desconexões entre os nós sensores e o nó sorvedouro. Por isso, o gráfico da Figura 6 mostra que nenhum protocolo simulado alcançou 100% de taxa de entrega. O protocolo *Gossiping* foi o protocolo com menor taxa de entrega devido à característica de protocolo de escolha

aleatória do próximo salto. O protocolo *Flooding* foi o protocolo com a maior taxa de entrega devido às suas várias retransmissões que compensam perdas de pacotes. Os demais protocolos tiveram taxas de entrega similares.

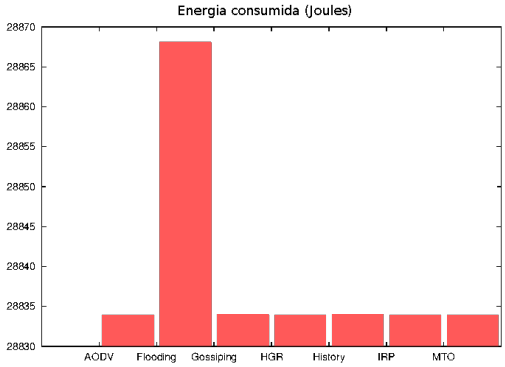
O gráfico da Figura 7 mostra a estimativa da energia consumida pelos colares. Este gráfico apresenta o consumo total dos colares, somando o consumo do circuito (GPS+Antena, OpenLog e Arduino) e o consumo dos XBees. Exceto pelo Flooding, a diferença do consumo de energia entre os protocolos é pequena devido ao consumo base do colar, que é exacerbado pelo elevado consumo do circuito de GPS. Houve experimentos adicionais para desligar e ligar novamente o GPS mas, devido ao tempo para a sincronização e ao intervalo de envios dos dados na rede que é de 5 minutos, não houve melhora. Por isso, a faixa de apresentação do consumo de energia no gráfico da Figura 7 ficou nivelada.

Figura 6: Taxa de entrega de Pacotes



Fonte: Elaborada pelo autor

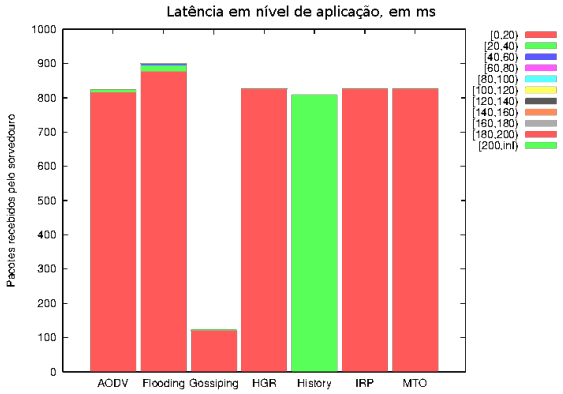
Figura 7: Energia consumida pelo colar



Fonte: Elaborada pelo autor

O rádio transmissor utilizado, XBee Pro Series 2, não permite que uma nova camada de roteamento seja implementada em seu *firmware* devido à implementação fechada. Mesmo assim, no contexto estudado, as simulações mostraram que os protocolos existentes no XBee, AODV e *Many-to-one*, possuem desempenho equivalente aos protocolos HGR e IRP, e melhores que Flooding, Gossiping e History.

Figura 8: Latência na entrega de Pacotes



Fonte: Elaborada pelo autor

O gráfico da Figura 8 apresenta a latência (em ms) de entrega de pacotes. A latência é o tempo que demora para um pacote originário de um nó sensor alcançar o nó sorvedouro. O protocolo de pior latência foi o *History-based*, com pacotes com latência acima de 200ms, devido à busca por um nó vizinho que esteve mais vezes próximo ao nó sorvedouro, tendo que esperar a resposta de todos os vizinhos, foi estipulado um tempo de espera baseado no *timeout* de envio de pacote *unicast* do XBee, a busca por vizinho sempre era feita antes de enviar um pacote de dados. AODV e Flooding tiveram pacotes com latência nas faixas até 60 ms, o AODV devido às idas e voltas de pacotes de estabelecimento de rotas entre o nó sensor e o nó sorvedouro, e o Flooding devido a uns poucos pacotes que utilizaram mais saltos a partir de um nó sensor até alcançar o nó sorvedouro. Os demais protocolos tiveram latência similar, na faixa até 20 ms. Mesmo com essas diferenças de latência, qualquer um dos protocolos avaliados poderia ser utilizado porque o intervalo de 5 minutos de coleta das posições é muito maior do que a latência máxima observada na maioria dos protocolos.

CONCLUSÕES E TRABALHOS FUTUROS

A rede de sensores sem fio proposta permitiu melhorar os trabalhos de Lomba (2015) e Jesus (2014) ao realizar a comunicação dos colares e recebimento automático dos dados dos sensores dos colares durante os experimentos, sem precisar realizar a coleta de forma manual e sem esperar o término dos experimentos. Os colares permitem a adição de novos sensores, e estes poderiam utilizar a mesma arquitetura de rede de sensores sem fio proposta. Pôde-se observar nos resultados das simulações para o cenário estudado, que os protocolos de roteamento AODV e *Many-to-one* dos chips XBee utilizados tiveram um consumo de energia e taxa de entrega próximos, com uma pequena vantagem do *Many-to-one* em relação à latência.

Entre vários trabalhos futuros, um deles seria a utilização de pontos auxiliares fixos para ajudar no roteamento das mensagens para enfrentar as desconexões ocorridas na RSSF devido à mobilidade em grupo dos bovinos e para permitir que alguns nós sensores possam ser configurados como *end devices* para utilizar o modo *sleep* e diminuir o consumo de energia. Outro trabalho futuro seria o desenvovimento de alguns colares sem o GPS, que calculariam a posição aproximada do animal por cálculos matemáticos, utilizando a potência de recebimento dos pacotes dos outros colares como uma métrica para estimar a distância de outros sensores. A utilização de protocolos das classes MULE (Mobile Ubiquitous LAN Extension) (SHAH et al., 2003) e SDN (Software Defined Networking) (XIA et al., 2015) também é uma opção a ser estudada pois os nós não precisam comunicar-se entre si na ausência do sorvedouro.

AGRADECIMENTOS

Agradecimentos à FUNDECT pelo apoio financeiro via Projeto Universal com termo de outorga Nº 102/2014, e à Embrapa Gado de Corte pela disponibilização da área e dos bovinos para os experimentos e pelo apoio do Dr. Pedro Paulo Pires e de Quintino Izidio dos Santos Neto.

REFERÊNCIAS

AKKAYA, K.; YOUNIS, M. A survey on routing protocols for wireless sensor networks. *Ad hoc networks*, Elsevier, v. 3, n. 3, p. 325–349, 2005.

AKYILDIZ, I. F. et al. A survey on sensor networks. *IEEE communications magazine*, IEEE, v. 40, n. 8, p. 102–114, 2002.

BERCKMANS, D. Automatic on-line monitoring of animals by precision livestock farming. *Proceedings of the ISAH Conference on Animal Production in Europe: The Way Forward in a Changing World - Saint-Malo, France*, p. 27–31, 2004.

BOULIS, A. et al. Castalia: A simulator for wireless sensor networks and body area networks. *NICTA: National ICT Australia*, 2011.

CHEN, M. et al. Hybrid geographic routing for flexible energy—delay tradeoff. *IEEE Transactions on Vehicular Technology*, IEEE, v. 58, n. 9, p. 4976–4988, 2009.

DIGI. *XBee Java Library*. 2016. Disponível em: <https://docs.digi.com/display/XBJLIB/XBee+Java+Library>.

JESUS, L. d. Identificação do comportamento bovino por meio do monitoramento. *Dissertação de Mestrado, Universidade Federal de Mato Grosso do Sul*, 2014.

JUANG, P. et al. Energy-efficient computing for wildlife tracking: Design tradeoffs and early experiences with zebranet. In: ACM. *ACM Sigplan Notices*. [S.l.], 2002. v. 37, n. 10, p. 96–107.

KWONG, K. H. et al. Practical considerations for wireless sensor networks in cattle monitoring applications. *Computers and Electronics in Agriculture*, Elsevier, v. 81, p. 33–44, 2012.

LOMBA, L. F. D. Identificação do comportamento bovino a partir dos dados de aceleração do animal e monitoramento da luminosidade do ambiente. 2015.

MEZZALIRA, J. C. et al. Aspectos metodológicos do comportamento ingestivo de bovinos em pastejo. *Revista Brasileira de Zootecnia*, v. 40, n. 5, p. 1114 – 1120, 2011.

PERKINS, C.; BELDING-ROYER, E.; DAS, S. *Ad hoc On-Demand Distance Vector (AODV) Routing*. [S.l.], July 2003. (Accessed on 11/15/2016). Disponível em: <https://www.ietf.org/rfc/rfc3561.txt>.

SHAH, R. C. et al. Data mules: modeling a three-tier architecture for sparse sensor networks. In: *Proceedings of the First IEEE International Workshop on Sensor Network Protocols and Applications, 2003.* [S.l.: s.n.], 2003. p. 30–41.

XIA, W. et al. A survey on software-defined networking. *IEEE Communications Surveys Tutorials*, v. 17, n. 1, p. 27–51, Firstquarter 2015. ISSN 1553-877X.

Integração de Redes de Sensores Sem Fio eKo a Sistemas de Informação para Rastreabilidade do Vinho

Igo Romero Costa de Souza[1], Leonardo Barreto Campos[1,2], Carlos Eduardo Cugnasca[2]

1 Grupo de Pesquisa Sistemas de Hardware e Software, Instituto Federal de Educação, Ciência e Tecnologia da Bahia, Vitória da Conquista, Bahia, Brasil, igoromero@ifba.edu.br, leonardobcampos@gmail.com

2 Departamento de Engenharia da Computação e Sistemas Digitais, Escola Politécnica, Universidade de São Paulo, São Paulo, São Paulo, Brasil, carlos.cugnasca@poli.usp.br

RESUMO

A rastreabilidade da cadeia produtiva do vinho vem sendo viabilizada por meio de Redes de Sensores Sem Fio (RSSF's). Entre as RSSF's mais usadas no agronegócio está a linha *eKo* da *MEMSIC* que permite monitorar constantemente diversas variáveis ambientais em qualquer etapa na cadeia. Entretanto, essa importante rede de sensores não permite a interoperabilidade com outros sistemas, tornando os dados subutilizados. Dessa forma, o presente trabalho descreve o desenvolvimento de uma solução capaz de compartilhar as informações coletadas pelas RSSF's eKo na cadeia de suprimentos do vinho para demais sistemas.

PALAVRAS-CHAVE: RSSF eKo, Rastreabilidade, Vinho, Sistemas de Informação.

ABSTRACT

The traceability of the wine production chain has been made possible through Wireless Sensor Networks (WSNs). Among the most commonly used WSNs in agribusiness is MEMSIC's eKo line, which allows constant monitoring of various environmental variables at any stage in the chain. However, this important network of sensors does not allow interoperability with other systems, making the data underutilized. Thus, the present work describes the development of a solution capable of sharing the information collected by WSNs eKo in the wine supply chain for other systems.

KEYWORDS: eKo Wireless Sensor Network, Traceability, Wine, Information System.

INTRODUÇÃO

O conhecimento das ações e condições que transformaram uma matéria-prima em produto final está entre as exigências que mais crescem entre consumidores e receptores de mercadorias, especialmente para produtos exportados e serviços de segurança alimentar (MACHAD E NANTES, 2007; YAN et al., 2011). Os consumidores finais estão se interessando cada vez mais em conhecer por exemplo, a origem e características dos produtos adquiridos, dados sobre a região produtora, insumos utilizados, fatores que foram expostos durante a produção e transporte. Esse anseio tem encontrado repouso nas RSSF's. Com a RSSF é possível efetuar coleta de dados de forma distribuída. Uma RSSF pode ser formada por dezenas ou até milhares de sensores posicionados dentro do fenômeno a ser observado ou próximo a ele (MACEDO; ARAÚJO;SOUSA, 2013).

Um exemplo de RSSF é a solução *eKo* fornecida pela *Memsic Corporation*. Esta solução conta com um conjunto de nós equipados com sensores capazes de monitorar variáveis ambientais. Mesmo sendo uma tecnologia avançada e amplamente utilizado no meio bem como outras RSSFs, a solução *eKo*, esbarra no problema do compartilhamento das informações para outros sistemas. A falta de integração entre as diferentes tecnologias, a heterogeneidade na comunicação de controle dos dispositivos e a dificuldade de padronização do mercado resultam em problemas relacionados à interoperabilidade entre as diferentes soluções existentes (WACKS, 2002).

A ausência de interoperabilidade faz com que não seja possível a utilização dos dados da RSSF por outros sistemas de informação, tornado os subutilizados. Dessa forma, este trabalho apresenta o desenvolvimento de uma solução escalável e interoperável capaz de recuperar as informações de uma RSSF *eKo*, e disponibilizar essas informações para um sistema de informação voltado a rastreabilidade do vinho. Para melhor entendimento, o restante desse artigo está estruturado da seguinte forma. A seção seguinte apresenta o Material e Métodos utilizados. A terceira seção apresenta a solução implementada para a integração das Redes de Sensores Sem Fio da linha *eKo* à Sistemas de Informação. A quarta seção mostra uma prova de conceito. Por fim, a ultima seção apresenta as conclusões do trabalho.

MATERIAL E MÉTODOS

O presente trabalho foi desenvolvido em uma parceria do Instituto Federal de Educação Ciência e Tecnologia Campus Vitoria da Conquista com a colaboração do Laboratório de Automação Agrícola (LAA) da Universidade de São Paulo (USP) e a Vinícola Santa Maria localizada na

cidade de Lagoa Real – PE (Vale do Rio São Francisco) adotando o Material e Métodos descritos na subseção a seguir.

Material

Para realização da pesquisa, foi utilizado os equipamentos descrito na Tabela 1.

Tabela -1 Material Utilizado

Variáveis	Modelo	Sensores	Imagem
Temperatura e Humidade	ES-1201	Sensor Humidade e Temperatura	
Potencial Hídrico do Sol	ES-1110	Sensor de medição de potencial hídrico do solo	
Radiação Solar	ES-1401	Sensor de radiação foto-sinteticamente ativa	
Humidade do Solo	ES-1101	Sensor de umidade do solo	
Umidade Foliar	ES-1301	Senso de umidade foliar	

A escolhas destes sensores apoiam-se nas pesquisas realizadas durante o levantamento bibliográfico, e analise dos trabalhos como (ANASTASI et al., 2009) e (GAY-FERNÁNDEZ; CUIÑAS, 2011) que utilizam as RSSF para monitorar humidade, temperatura, velocidade e direção do vento além do índice de irradiação solar na produção de vinho e (BURRELL; BROOKE; BECKWITH, 2004) que utilizam técnica de etnografia para estudar e documentar as necessidades e prioridades das pessoas que trabalham na vinha a fim de obter uma melhor compreensão do potencial para redes de sensores na agricultura. Outros trabalhos que influenciaram essa escolha foram os de (CAMPOS; CUGNASCA, 2013) onde os autores ressaltam as vantagens e a importância das RSSF para a rastreabilidade nos componentes da cadeia do vinho.

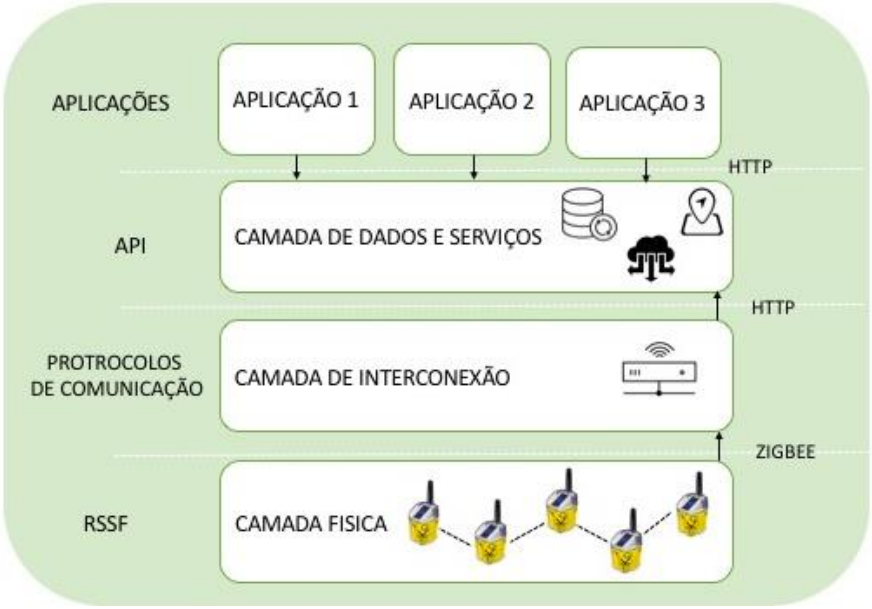
Métodos

Inicialmente realizou-se uma revisão bibliográfica para compreender o funcionamento e identificar os principais problemas entre as RSSF e o compartilhamento das informações neste domínio. Com base nas informações obtidas, modelou-se então uma arquitetura candidata para o desenvolvimento da solução. Essa arquitetura foi projetada seguindo a abordagem SOA (*Service-Oriented Architecture*), sendo organizada em três camadas principais: coleta, comunicação e apresentação de dados. A partir da arquitetura proposta foi desenvolvido um protótipo em escala reduzida utilizando a RSSF *eKo*. Uma vez comprovado a eficiência do protótipo na coleta e comunicação com os demais componentes da solução implantou-se a solução em um ambiente de produção como uma prova de conceito.

RESULTADOS E DISCUSSÃO

A partir dos materiais e métodos apresentados na seção anterior, tornou-se possível a interoperabilidade e o compartilhamento dos dados em campo para demais aplicações. Para entender melhor sobre a solução e como estes se comunicam a Figura 1 apresenta os componentes, tecnologias e protocolos em forma de uma arquitetura separada por camadas.

Figura 1 – Arquitetura do Sistema



Fonte: Autor

Camada Física

A Camada Física é formada pelos nós e sensores da RSSF eKo. Essa camada é responsável pela coleta e envio dos dados dos nós em campo para o *Gateway* presente na Camada de

Interconexão. A comunicação entre o *Gateway* e os Nós é realizada por meio da tecnologia Zigbee, além disso a rede conta com mecanismos para correção do enlace de forma autônoma caso ocorre-se algum problema. Para cobrir área monitorada foram utilizados: um *Gateway*, cinco nós e um total de quatorze sensores espalhados na área cultivada, conforme mostra a Figura 2.

Figura 2 – Disposição dos Nós instalados



Fonte: Autor

Uma vez instalado os nós em campo, realizou-se a instalação e configuração do *Gateway*. O *Gateway* (base), foi montado em uma sala de apoio da própria vinícola, uma vez que a mesma possui conexão de energia e internet que são essenciais para o seu funcionamento. Com a RSSF configurada e funcionando, deu-se início a coleta de dados que se estendeu todo do mês de setembro a dezembro de 2016.

Camada de Interconexão

A Camada de Interconexão visa proporcionar comunicação entre os nós da RSSF, *Gateway* e o Serviço. Assim que os dados coletados pelos Nós chegam ao *Gateway*, os mesmos são organizados e armazenados em uma base de dados interna. O *Gateway* utiliza como Sistema Operacional (SO) uma distribuição baseada no *Linux*. Dessa forma, para sincronizar os dados armazenados e a camada de Dados e Serviços foi desenvolvido um *script* com a linguagem *Python.* Quando executado, o script recupera os últimos registros armazenados e os organiza em um documento no formato *JSON* (JavaScript Object Notation), enviando para a API (*Application Programming Interface*) por meio de uma requisição HTTP. A Figura 3 apresenta

um exemplo de documento JSON contendo os dados armazenados no *Gateway*. Para automatizar essa tarefa, utilizou-se as ferramentas *contrab* e *cron* presentes no SO. O *crontab* é um programa do *Unix* que edita o arquivo onde são especificados as configurações e os comandos a serem executados pelo *cron*. No *cron* são definidos o horário, usuário e frequência em que as ações serão executadas. No projeto o *cron* foi configurado para executar os scripts em um intervalo de quinze minutos.

Figura 3 – Exemplo de um Documento JSON Contendo os Dados do *Gateway*


```
{
  "registros": [
    {
      "_id": "57efdd9a29f19b73c597832a",
      "data": "2016-10-01T12:45:04.000Z",
      "noid": 2,
      "bv": 4.37,
      "sv": 4.6,
      "tempInterna": 41.76
    },
    {
      "_id": "57efdd9b29f19b73c597832d",
      "data": "2016-10-01T12:45:31.000Z",
      "noid": 9,
      "bv": 4.22,
      "sv": 4.51,
      "tempInterna": 40.73
    },
```


Fonte: Autor

Camada de Dados e Serviços

Por fim a Camada de Dados e Serviços é responsável por armazenar e disponibilizar os dados para outras aplicações através da *Web* por meio de uma API. A API foi implementada seguindo o modelo arquitetural REST (*Representational State Transfer*). Neste modelo todos os recursos e informações são disponibilizados por meio de URLs. Toda a troca de informações entre as aplicações com esta camada é realizada por meio requisições HTTP implementadas pelos métodos (GET, POST, PUT, DELETE). Para gerenciar as requisições da API foi escolhido como tecnologia *server-side* o *Node.js* e o *framework Web Express*.

As aplicações *Node.js* são projetadas visando maximizar o *troughput* e a eficiência, usando operações de entrada e saída não bloqueantes, e eventos assíncronos. Essas aplicações são escritas com a linguagem *JavaScript* e executadas com motor de interpretação V8 *JavaScript Engine*. Essa tecnologia possui, uma comunidade ativa e conta com uma grande quantidade de módulos já desenvolvidos que podem ser utilizados para os mais diversos fins. Na Figura 4 é apresentado o arquivo de configuração "*package.json*" contendo os módulos e configurações utilizados para desenvolvimento do projeto.

Figura 4 – Dependências Utilizadas no Projeto



```
{
  "name": "api-nodejs-mongodb",
  "version": "1.0.1",
  "description": " API com NodeJS e MongoDB",
  "main": "app.js",
  "dependencies": {
    "body-parser": "^1.15.2",
    "cors": "^2.7.1",
    "express": "^4.14.0",
    "getmodule": "^1.0.0",
    "mongoose": "^4.5.4"
  },
  "devDependencies": {},
  "scripts": {
    "start": "node ./bin/www",
    "test": "echo \"Error: no test specified\" && exit 1"
  },
  "author": "Igo Romero",
  "license": "MIT"
}
```

Fonte: Autor

Um dos módulos combinados com o *Node.js* foi o *Express*. O *Express* estende as capacidades do servidor padrão do *Node.js* adicionando *middlewares* e outras capacidades como *views* e rotas (ALMEIDA, 2015). Os recursos desse *framework web*, foram importantes para a criação e definição de rotas personalizas do serviço. Por meio do *Express* mapeou-se cada sensor da RSSF para uma rota especifica, fazendo com que toda a interação entre os sistemas e a solução desenvolvida ocorressem por meio dos métodos HTTP. A Tabela 2 apresenta as rotas da API REST, com os respectivos verbos HTTP e a descrição das ações que cada que podem ser executadas.

Tabela 2 – Rotas da API

URL	Verbo HTTP	Sensor	Variáveis Monitoradas
http://ipservidor:2500/v1/registros/es1201	GET/POST	ES1201	Umidade, Temperatura, Ponto de Orvalho
http://ipservidor:2500/v1/registros/es1110	GET/POST	ES1110	Umidade Externa do Solo
http://ipservidor:2500/v1/registros/es1301	GET/POST	ES1301	Molhamento Foliar
http://ipservidor:2500/v1/registros/es1401	GET/POST	ES1401	Radiação Solar
http://ipservidor:2500/v1/registros/es1100	GET/POST	ES1100	Umidade Interna do Solo

Fonte: Autor

A cada requisição a uma dessa URLs, combinado com o verbo HTTP GET tem como resposta um documento JSON contendo todos os registros armazenados de um determinado sensor. Dessa forma caso uma aplicação necessite de apenas uma variável, a API poderá disponibilizar de forma separada, evitando assim que a mesma não precise de tratar dados não necessários. Além dos sensores externos, os dispositivos internos para o acompanhamento da capacidade energética solar e carga da bateria do Nós também estão disponíveis na API. Para persistência dos dados foi utilizado o banco de dados *MongoDB*. O *MongoDB* é considerado um dos bancos de dados *NoSQL* mais populares no mercado.

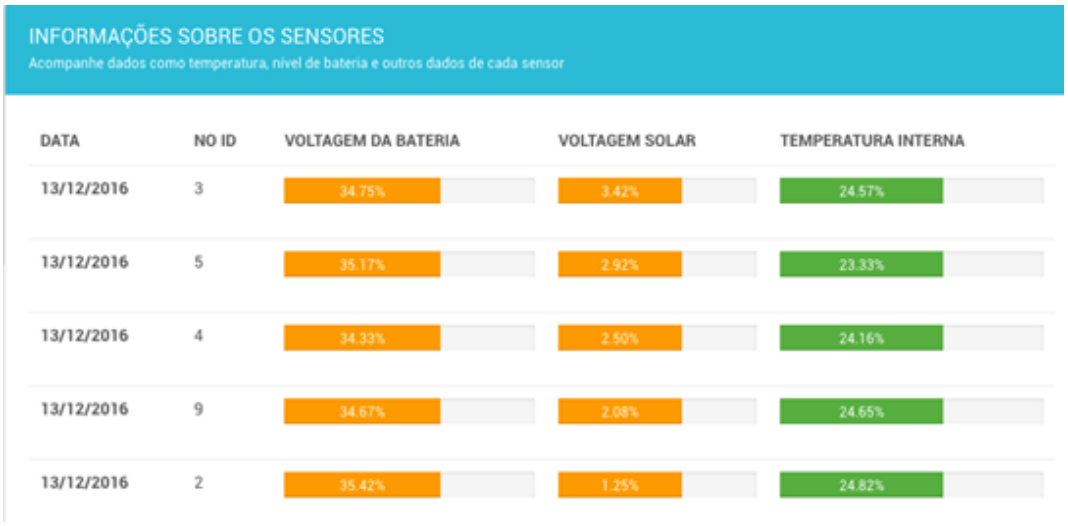
Em vez de armazenar dados em estruturas rígidas, como tabelas, eles os armazenam em documentos vagamente definidos conhecido como documentos. Para organizar as informações, foram realizados mapeamentos das tabelas do banco de dados *SQLite* para coleções do *MongoDB* garantindo a assim a integridade dos dados. As coleções foram criadas utilizando o modulo *Mongoose*. O *Mongoose* é uma biblioteca do *Node.js* para modelagem de objetos *MongoDB* escrita em *JavaScript*, funcionando como ODM (*Object Data Mapping*). No *Monogoose* são definidos os esquemas de dados contendo os campos e tipos de informações que serão armazenadas nas coleções.

PROVA DE CONCEITO

Como uma prova de conceito foi implementado uma aplicação web para exibição dos dados da API. A aplicação foi desenvolvida em forma de uma SPA (*SINGLE PAGE APLICATION*) utilizando as bibliotecas da *Google*: *Angular, Google MAPS*. A aplicação tem como objetivo, permitir que o produtor ou responsável técnico possa o recuperar dados referentes as características ambientais, físicas e climáticas capturadas pela RSSF, em um determinado espaço temporal em que uma safra especifica ficou exposta durante a produção, permitindo assim a aplicação da rastreabilidade ainda na primeira etapa da cadeia de suprimento do vinho.

Para isso a aplicação utiliza os dados disponíveis da API desenvolvida, para alimentar gráficos e relatórios personalizados das principais variáveis envolvidas na produção e onde estão localizadas. Uma vez cadastrados os Nós, a aplicação permite acompanhar informações como: nível de bateria, radiação solar e temperatura interna além dos sensores em campo. Estes dados podem ser acompanhados por meio uma página Web desenvolvida em um formato de tabela interativa, e que pode ser vista na Figura 5. Através desta tela o usuário pode acompanhar as condições que o Nó se encontra e o momento da última leitura.

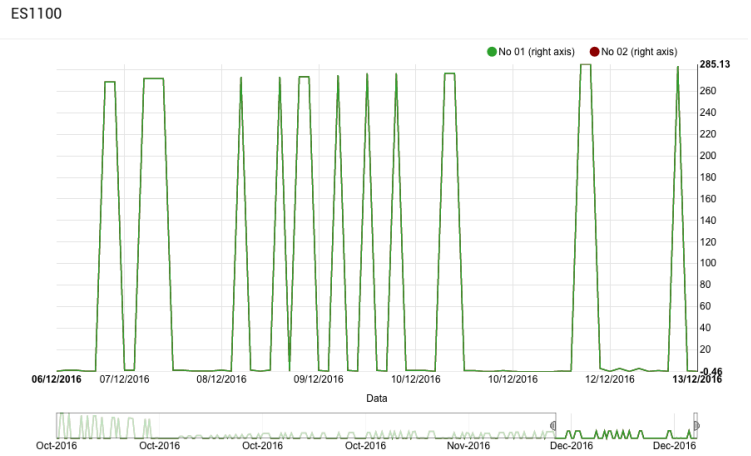
Figura 5 – Dashboard para apresentação dos Dados dos Nós



Fonte: Autor

Para apresentação dos dados dos sensores, a aplicação web reorganiza as informações dos sensores em formato de gráficos interativos. A Figura 6 apresenta um exemplo de uma página contendo as informações do sensor ES1100 responsável por monitorar umidade do solo. Os dados dos sensores foram organizados em forma temporal, podendo ser alterado de forma dinâmica quando necessário.

Figura 6 – Gráfico com os Dados do Sensor de Umidade ES1100



Fonte: Autor

CONCLUSÕES

O presente trabalho possibilitou o estudo e desenvolvimento de um serviço (mecanismo) para prover interoperabilidade e compartilhamento de dados entre a RSSF eKo e os sistemas voltados a rastreabilidade na cadeia de suprimentos do vinho. O desenvolvimento da pesquisa culminou em um modelo de comunicação baseado em camadas, capaz de recuperar as informações coletadas nos sensores em campo e disponibiliza-las na web através de uma API

REST. A solução implantada na vinícola, permitiu com que os dados coletados no campo pela RSSF eKo, se tornassem disponíveis de forma organizada e modular para serem utilizados por qualquer outro sistema. A disponibilização destes dados, permite por exemplo, a criação de aplicações móveis para monitorar e recuperar informações sobre as características físicas de uma determinada área que produziu ótimas safras, ou então a integração de um sistema de irrigação inteligente capaz de irrigar apenas as áreas necessárias com base nas informações de umidade do solo geradas pela RSSF. Como trabalhos futuros pretende-se implementar a RSSF nas demais fases da cadeia de suprimentos como fermentação, amadurecimento e armazenamento fazendo com que os dados de todas as etapas possam sejam integrados e possam estar disponíveis por meio da API desenvolvida neste trabalho. Ou então estender a capacidade da solução para outras tecnologias e outras culturas.

REFERÊNCIAS

ALMEIDA, Flavio Henrique de Souza (Ed.). **MEAN: Full Stack JavaScript para aplicações web com MongoDB**, Express, Angular e Node. São Paulo: Casa do Código, 2015. 377 p.

ANASTASI, Giuseppe et al. Monitoring High-Quality Wine Production using Wireless Sensor Networks. **2009 42nd Hawaii International Conference On System Sciences,** [s.l.], p.298-305, 5 Não é um mês valido! 2009. Institute of Electrical & Electronics Engineers (IEEE). http://dx.doi.org/10.1109/hicss.2009.313.

BURRELL, J.; BROOKE, T.; BECKWITH, R. Sensor and actuator networks - Vineyard computing: sensor networks in agricultural production. **Ieee Pervasive Comput**., [s.l.], v. 3, n. 1, p.38-45, jan. 2004. Institute of Electrical & Electronics Engineers (IEEE). http://dx.doi.org/10.1109/mprv.2004.1269130

CAMPOS, Leonardo Barreto; CUGNASCA, Carlos Eduardo. **Um Modelo de Rastreabilidade da Cadeia do Vinho Baseado em Internet das Coisas Utilizando Redes de Sensores Sem Fio e RFID**. Workshop de Pós-Graduação da Área de Concentração Engenharia de Computação do Ppgee-epusp, São Paulo, p.73-74, 10 out. 2013.

GAY-FERNÁNDEZ, José A.; CUIÑAS, Iñigo. DEPLOYMENT OF A WIRELESS SENSOR NETWORK IN A VINEYARD. Wireless Information Networks And Systems (winsys)**, 2011 Proceedings Of The International Conference On,** Seville, p.35-40, 18 jul. 2011.

MACED, Diego Carvalho P.; ARAÚJO, Harilton da S.; SOUSA, Aldir S. **UMA PROPOSTA PARA CONTROLE DE DENSIDADE EM REDES DE SENSORES SEM FIO UTILIZANDO INTELIGÊNCIA COMPUTACIONAL**. Simpósio Brasileiro de Automação Inteligente, Fortaleza, p.1-1, 13 out. 2013.

MACHADO, J.; NANTES, J. F. D. a. **A Visão Institucional do Processo de Rastreabilidade da Carne Bovina**: Capturado em 2007.

WACKS, K. **Home systems standards: achievements and challenges**. IEEE Communications Magazine, v. 40, n. 4, p. 152–159, 2002.

Rastreabilidade da cadeia do vinho brasileiro: um modelo baseado em tecnologia de radiofrequência e rede de sensores sem fio

Bruno Rogora Kawanor[1], Renata Maria Marér[2], Roberto Fray da Silva[3], Carlos Eduardo Cugnasca[4]

[1] Departamento de Engenharia da Computação e Sistemas Digitais, Escola Politécnica, Universidade de São Paulo, São Paulo, SP, Brasil, bruno.kawano@usp.br, renatamare@gmail.com, roberto.fray.silva@gmail.com, carlos.cugnasca@usp.br

RESUMO

O vinho produzido no Brasil e comercializado no mercado internacional ainda não está amparado por um sistema eficiente de rastreabilidade. Além disso, as leis que regulam as indústrias de alimentos e boa parte das cadeias de suprimentos de produtos agroindustriais ao redor do mundo estabelecem que a rastreabilidade em toda a produção, fases de transformação e distribuição devem ser garantidas. Nesse sentido, Redes de Sensores sem Fio que utilizam Identificação por Radiofrequência, tem se mostrado uma solução promissora para fornecer o histórico de dados, em tempo real, associados a um dado produto, desde a sua fabricação até o descarte. Dessa forma, o presente projeto propõe a elaboração de um modelo de rastreabilidade que forneça informações úteis aos membros da cadeia de suprimentos. Para tanto será concebido e avaliado um modelo de rastreabilidade utilizando as tecnologias de Identificação por Radiofrequência e Redes de Sensores Sem Fio para a cadeia do vinho.

PALAVRAS-CHAVE: Rastreabilidade, Internet das Coisas, Redes de Sensores sem Fio, RFID, Cadeia do Vinho

ABSTRACT

Wine produced in Brazil and marketed in the international market is not yet supported by an efficient traceability system. In addition, laws regulating food industries and most of the agro-industrial supply chains around the world state that traceability across production, transformation and distribution phases must be guaranteed. In this sense, Wireless Sensor Systems using Radio Frequency Identification, has shown a promising solution to provide the

real-time data history associated with a given product since its manufacture Until disposal. Thus, the present project proposes the elaboration of a traceability model that provides useful information to the members of the supply chain. For this, a traceability model will be designed and evaluated using Radio Frequency Identification and Wireless Sensor Networks technologies for the wine chain.

KEYWORDS: Traceability, Internet of Things, Wireless Sensor Networks, RFID, Wine Supply Chain.

INTRODUÇÃO

Acontecimentos ocorridos nos últimos 15 anos envolvendo contaminação de produtos alimentícios, entre eles o caso da Encefalopatia Espongiforme Bovina ocorrido na União Europeia em 1992 (LATOUCHE; RAINELLI; VERMERSCH, 1998) e a contaminação de alimentos para animais na Bélgica em 1999 (BERNARD et al., 2002), contribuíram para que o Parlamento Europeu e o Conselho da União Europeia regulamentassem a produção e importação de quaisquer produtos alimentícios ou substâncias que venham ou possam vir a ser incorporadas em seus produtos alimentícios (COMMISSION, 2002).

A capacidade de identificar automaticamente e rastrear individualmente objetos tornaram-se fatores essenciais em diferentes processos de negócios e domínios, tais como, fabricação, logística, varejo, estoque, transporte, identificação humana, saúde, etc (ILIE-ZUDOR et al., 2011). Entre essas aplicações, o setor agroindustrial tem atraído esforços consideráveis de pesquisa para elevar a precisão na obtenção de dados, segurança, confiabilidade e a agilidade nos processos de produção, distribuição e comercialização de seus produtos (YAN et al., 2011).

O conhecimento das ações e condições que transformaram uma matéria-prima em produto final estão entre as exigências que mais crescem entre consumidores e receptores de mercadorias, especialmente para produtos exportados (MACHADO; NANTES, 2007).

Desta forma, pretende-se apresentar neste trabalho um modelo de rastreabilidade aplicado à cadeia de suprimentos do vinho, de forma que seria possível a gestão de dados obtidos por sensores e de chips de RFID (Identificação por Radiofrequência) a fim de otimizar a cadeia, reduzindo custos e aumentando a confiabilidade do produto e garantia da rastreabilidade para todos os agentes da cadeia.

CADEIA LOGÍSTICA DO VINHO

A logística é a área de estudos multidisciplinar que foca no transporte e na armazenagem de produtos, além dos fluxos de informações e de dinheiro entre as empresas envolvidas no atendimento da demanda de um segmento consumidor. O conceito de CS adotado atualmente contempla o aumento da eficiência de cadeias de produtos, levando às interações entre áreas das empresas e à análise das interações entre as diferentes empresas ao longo da cadeia estudada (BALLOU, 2006; CHOPRA e MEINDL, 2011).

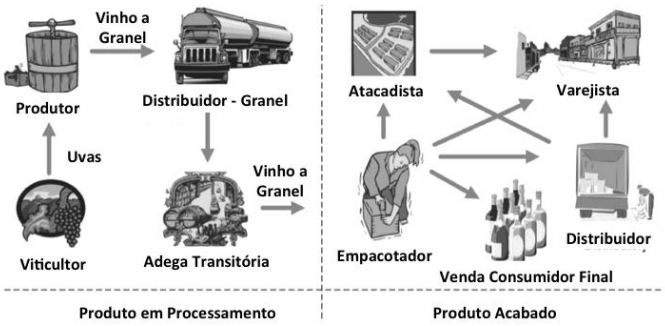
Além disso, com a promulgação do Regulamento (CE) 178/2002, tornou-se obrigatória a manutenção de dados de rastreabilidade por todos os envolvidos na comercialização de produtos alimentícios com a União Europeia, a partir de 2005 (GOGLIANO SOBRINHO et al., 2010).

Uma das cadeias sujeitas a essa adequação é a cadeia de suprimentos do vinho, Figura 1, objeto deste estudo. A Figura 1 demonstra a cadeia do vinho como ocorre atualmente, sem a adoção de um sistema unificado de rastreabilidade e acompanhamento em tempo real do produto durante seu transporte. Neste trabalho, vai ser modelada a cadeia do vinho com a adoção de um sistema de rastreabilidade baseada nas tecnologias de RFID e de RSSF, conforme já foi citado e pode ser verificado no item de resultados e discussão.

Na legislação europeia, considera-se o vinho como um produto alimentício e, portanto, é mandatória aos produtores brasileiros, voltados à exportação para os países locais, a adoção de práticas de rastreabilidade (GOGLIANO SOBRINHO et al., 2010).

Segundo padrões internacionais (ISO, 1994), rastreabilidade é o processo de descrever a história, aplicação, eventos e localização de um produto a uma determinada organização.

Figura 1 - Cadeia de Suprimentos do Vinho



Fonte: Adaptado de CIMINO; MARCELLONI, 2012

RASTREABILIDADE E TECNOLOGIAS PARA A MODELAGEM

Por meio da rastreabilidade, é possível saber, entre outras informações referentes a um produto, quais foram os lotes de cada matéria-prima utilizada na sua composição, sua data de validade, os seus laudos de análises, o fornecedor, pesagem de cada ingrediente que compõe esse produto, tempo de processamento, condições em que foram feitos.

Em diversos cenários de aplicação, umas das principais tecnologias é conhecida como Identificação por Radiofrequência (*Radio Frequency Identification* – RFID). Esta tecnologia pode ser definida como "uma pequena etiqueta (*tag*) contendo uma pastilha de circuito integrado e uma antena capaz de responder, via ondas de rádio transmitidas ao leitor RFID, informações do processo e armazenamento" (WU et al., 2006).

No geral, sistemas que usam RFID possuem, essencialmente, etiquetas, leitores e sistemas *backend*, ou simplesmente, "um típico sistema RFID consiste em etiquetas (*tags*), leitores, aplicações de software, hardwares e *middlewares*" (CHEUNG; KAILING; SCHONAUER, 2007).

A tecnologia RFID permite o registro automático e digital de dados, minimizando erros em processos de rastreabilidade. No entanto, esta tecnologia não permite por si só o monitoramento das variáveis ambientais às quais o produto está exposto, e esta é a principal razão pela qual a tecnologia de Redes de Sensores Sem Fio (RSSF) pode auxiliar, de forma significativa, no desenvolvimento de melhores modelos de rastreabilidade (EXPÓSITO; GAY-FERNÁNDEZ; CUIÑAS, 2013).

Uma Rede de Sensores Sem Fio pode ser definida como um grande número de sensores em atuação em uma área delimitada, sendo que a principal utilidade destes é captar informações do ambiente em que estão inseridos e comunicar-se entre si (WILSON, 2004; GLISIC; LORENZO, 2009). Os sensores são acoplados a nós, que podem estar inseridos ou não no fenômeno a ser medido, que podem ou não estar conectados a estações centrais, chamadas de estação base (AGUILAR, 2008). As principais vantagens citadas no uso da tecnologia de RSSF é que se evitam os problemas de infraestrutura advindos do uso de fios e cabos, como no caso das redes de sensores comuns, além de uma maior velocidade na transmissão de informações, permitindo a transferência de informações a tempo real via internet (WILSON, 2004).

Uma RSSF é possui três componentes principais: infraestrutura (características dos sensores, seu número e disposição na área de abrangência do fenômeno a ser medido), pilha

de protocolos (software utilizado na implementação das camadas de protocolos em cada nó sensor da rede) e aplicação (componente que representa os interesses e consultas dos usuários) (AGUILAR, 2008).

Dessa forma, a aplicação destas tecnologias em cadeias de suprimentos, contribui para assegurar a qualidade dos insumos e dos produtos finais, evita a mescla de produtos, facilita chamadas de retorno (*recall*) e torna menos onerosa a localização de falhas e correção das mesmas (JURAN; GODFREY, 1999; NUKALA et al., 2016).

MATERIAL E MÉTODOS

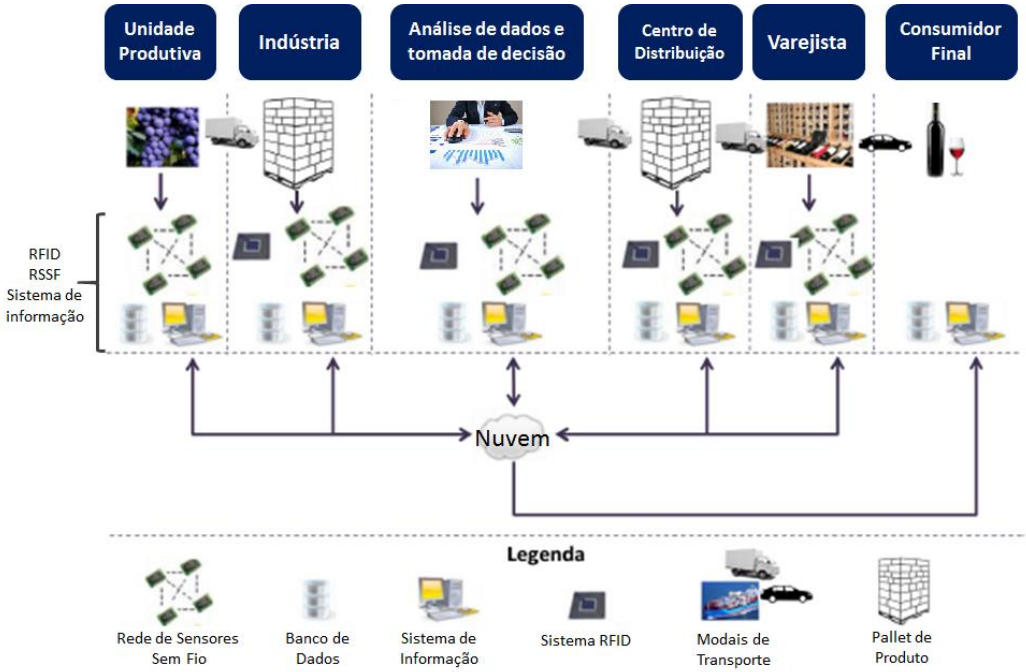
Este artigo faz parte de um projeto de Edital Universal do CNPq justamente sobre a temática da rastreabilidade da cadeia do vinho e neste trabalho, buscou-se realizar um breve levantamento bibliográfico sobre a logística da cadeia do vinho e também sobre a relevância da rastreabilidade e as tecnologias envolvidas para o modelo desenvolvido e apresentado neste artigo.

Em seguida, é apresentado um modelo de rastreabilidade com base nas características de cada tecnologia como o RFID e das RSSF. Levantou-se também, em cada etapa da cadeia do vinho, onde cada uma dessas tecnologias pode ser inserida e de que forma elas tornam o sistema mais eficiente, otimizado e com potencial de redução de custos e aumento dos benefícios também para o consumidor.

RESULTADOS E DISCUSSÃO

De acordo com a pesquisa e o levantamento das informações da cadeia logística do vinho e também sobre as tecnologias de RFID e de RSSF, concebeu-se o modelo de rastreabilidade para a cadeia do vinho, conforme pode ser verificado na Figura 2.

Figura 2 - Modelo de rastreabilidade da cadeia do vinho



Fonte: Os autores

Primeiramente, sensores de temperatura do ar, umidade relativa do ar e do solo, bem como sensores pluviométricos seriam instalados no campo, na Unidade Produtiva da uva para a produção do vinho. Estes sensores seriam dispostos em RSSF de forma que estes dados seriam utilizados para gerenciar todo o processo produtivo de manejo da cultura da uva, de forma a otimizar utilização de insumos agrícolas como fertilizantes e defensivos.

Neste modelo, estes dados seriam sincronizados com a nuvem, através do *middleware* desenvolvido para esta CS. Esta etapa termina com o transporte da matéria-prima, a uva, até a Indústria, ou seja, até a unidade onde a uva ser processada e transformada em vinho. Em muitos casos, na própria Unidade Produtiva, se produz o vinho nas chamadas vinícolas. Os caminhões para o transporte teriam um chip de RFID de forma a identificá-lo, por meio de portais, quando este chegar na Indústria de forma que tanto o produtor quanto o comprador da carga saibam o caminho da carga.

Na Indústria, o vinho já processado, vai ser disposto em garrafas, que em seguida vão ser dispostas em caixas e por fim em paletes. Nestas etapas tanto as garrafas, como as caixas quanto os paletes podem receber as etiquetas de RFID, sendo que a é disposta junto à garrafa é uma etiqueta especial que evita a interferência do líquido, sendo este modelo presente no mercado.

Entretanto, dependendo da escala da Indústria, talvez fique mais viável economicamente a inserção das etiquetas de RFID somente nas caixas e nos paletes. Estas etiquetas teriam uma dupla função, tanto na identificação do produto no caminho que este faz internamente na Indústria, por meio de portais, quanto no momento em que estes chegam no Centro de Distribuição (CD), no mercado Varejista.

Como o vinho é um produto sensível às variações de temperatura, quando ele sai da Indústria para o Centro de Distribuição, no compartimento de carga de transporte dos caminhões seriam instalados sensores de temperatura para que haja o monitoramento em tempo desta importante variável para a qualidade deste produto.

Este monitoramento das condições ambientais às quais o produto se encontrará exposto nos diferentes elos da CS, RSSF serão montadas em pontos de processamento (indústria), durante o transporte (dentro dos caminhões), e em pontos estacionários (CDs, atacadistas e varejistas), sendo os dados coletados continuamente por um software baseado em web (Sistema Monitorar® disponível no endereço eletrônico http://www.monitorar.net.br/), responsável pela geração e armazenamento de histórico de dados bem como, pela geração de alarmes de valores críticos eventualmente observados nessas etapas (KAWANO et al, 2015).

O cenário genérico de aplicação do Sistema Monitorar é um ambiente (edificação, contêiner ou meio de transporte) provido de uma rede de sensores adequados às variáveis de interesse. Por meio de um gateway conectado à Internet (Ethernet, GPRS), os dados são transmitidos continuamente a um servidor web, onde são armazenados e processados. Uma interface customizada e dedicada é disponibilizada aos usuários autorizados, que podem consultar os dados por meio de dispositivos conectados à Internet, acessando gráficos diversos, dashboards de valores instantâneos, planilhas, ou mesmo recebendo alarmes de valores críticos (SMS, e-mail). Esta arquitetura segue parâmetros da arquitetura de referência IoT-A segundo trabalho de Joachin (2012). Todos estes dados serão enviados à nuvem e ficariam disponíveis para todos os agentes da cadeia envolvidos.

A Análise de Dados e Tomada de Decisão, é a etapa anterior ao CD, sendo que o gestor da vinícola, de posse dos dados de produção e movimentação da carga dentro da indústria, bem como a gestão de estoques, que seria otimizada com a utilização das etiquetas nas caixas e paletes, este poderia conferir uma inteligência na gestão dessa carga. Assim, ele poderia programar de uma forma mais eficiente a quantidades e localidades para onde sua carga irá ser despachada, bem como gerenciar e controlar o estoque.

Chegando no CD, a carga seria novamente identificada por chips de RFID nos caminhões e, da mesma forma que na indústria, as movimentações internas delas seriam mais

facilmente acompanhadas por meio de portais de leitura e também leitores portáteis manuais que identificam os chips. Os gestores dos CD, munidos dos dados contidos na nuvem também teria sua gestão facilitada de forma que estes poderiam programar mais facilmente o destino de suas cargas para os Varejistas.

Os Varejistas, também estariam conectados ao sistema de controle e rastreabilidade da carga ser recebida de forma que todos os sistemas dos agentes das cadeias seriam compatíveis e interoperáveis, característica esta fundamental para um sistema que pretende ligar todos os componentes da cadeia de forma eficiente e com redução de custos, baseado na escala.

No mercado varejista, os agentes desta etapa podem utilizar tanto os portais para identificar as cargas recebidas dos CD´s quanto os leitores de RFID portáteis e manuais, de forma que estes leitores manuais facilitariam também o processo de inventário de produtos em seus estoques e também em seus pontos específicos de venda. Caso as etiquetas sejam inseridas nas garrafas de vinho, o Consumidor Final, poderia ler as informações do histórico deste vinho por meio de um leitor de RFID disponível nos pontos de venda, de forma que esta informação completaria o fluxo da rastreabilidade e informação.

CONCLUSÕES

Por meio deste trabalho, verificou-se que a rastreabilidade é um ponto fundamental para o acompanhamento de informações relevantes aos agentes da cadeia. Da mesma forma, as tecnologias levantadas possuem potencial de contribuir para a cadeia de suprimentos do vinho a fim de reduzir os custos, otimizar o acompanhamento e gestão logística da movimentação deste produto, bem como ser fonte de informações que alimentam a nuvem.

A interoperabilidade dos sistemas no modelo proposto é um fator preponderante para que haja o controle até mesmo sanitário por instituições governamentais de controle de qualidade, caso haja algum problema relacionado. Assim, espera-se que este modelo possa incentivar mais estudos de adoção das tecnologias citadas, bem como os modelos logísticos e de sistemas de informação relacionados ao modelo, de forma que a cadeia do vinho possa se tornar mais competitiva e eficiente, bem como ser base para futuros trabalhos na área.

AGRADECIMENTOS

Agradecemos ao apoio financeiro do Edital Universal do CNPq aprovado no edital MCT/CNPq Nº 014/2012, às bolsas de pesquisa do CNPq (161109/2013-6) e também à Fapesp.

REFERÊNCIAS

AGUILAR, J. C. P. Abordagem Semântica Aplicada ao Gerenciamento de Dados em Redes de Sensores sem Fio. Escola Politécnica, Universidade de São Paulo (USP), 2008.

BALLOU, R. H. Gerenciamento da Cadeia de Suprimentos. Editora Bookman, Porto Alegre – RS, 2006.

BIBI, F.; GUILAUMME, C. GONTARD, N. SORLI, B. A review: RFID technology having sensing aptitudes for food industry and their contribution to tracking and monitoring of food products. Trends in Food Science & Technology, v. 62, 2017.

CAIXETA-FILHO, J. V.; GAMEIRO, A. H. Transporte e logística em sistemas agroindustriais. São Paulo: Atlas, 2001.

CATARINUCCI, L.; CUINAS, I.; COLALLA, R.; FERNANDEZ, J. RFID and WSNs for Traceability of Agricultural Goods from Farm to Fork: Electromagnetic and Deployment Aspects on Wine Test-Cases. Software, Telecommunications and Computer Networks (SoftCOM), 2011 19th International Conference, 2011.

COMMISSION, E. Regulation (EC) No 178/2002 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety. Official Journal of the European Communities L, v. 31, n. 1, p. 1-24, 2002.

CHEUNG, A.; KAILING, K.; SCHONAUER, S. Theseos: a query engine for traceability across sovereign, distributed RFID databases. Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on, 2007, IEEE. p.1495-1496.

CHOPRA, S.; MEINDL, P. Supply Chain Management: Strategy, Planning and Operation. Fourth Edition, Pearson Education, New Jersey, USA, 2010.

CIMINO, M. C. A.; MARCELLONI, F. Enabling Traceability in theWine Supply Chain. Methodologies and Technologies for Networked Enterprises, v. 7200, pp 397-412, 2012.

EXPOSITO, I.; GAY-FERNANDEZ, J. A.; CUINAS, I. A Complete Traceability System for a Wine Supply Chain Using Radio-Frequency Identification and Wireless Sensor Networks. IEEE Antennas and Propagation Magazine, Vol. 55, No. 2, 2013.

ILIE-ZUDOR, E. et al. A survey of applications and requirements of unique identification systems and RFID techniques. Computers in Industry, v. 62, n. 3, p. 227-252, 4// 2011. ISSN 0166-3615. Disponível em: < http://www.sciencedirect.com/science/article/pii/S0166361510001521 >.

JOACHIN, W.; WALEWSKI, S. Internet of Things Architecture IoT-A. Deliverable D1.4. Converged archictural reference model for the IoT, v.2.0, 2012.

KAWANO, B. R. et al. Smart agriculture in the ubiquitous computing context: state of the art

and a proposal for an applied model. In: 12th International Conference of Information Systems and Technology Management (CONTECSI), 2015, São Paulo. Proceedings … São Paulo: FEA-USP, 2015.

LATOUCHE, K.; RAINELLI, P.; VERMERSCH, D. Food safety issues and the BSE scare: some lessons from the French case. Food policy, v. 23, n. 5, p. 347-356, 1998. ISSN 0306-9192.

LUIGI, A.; IERA, A.; MARABITO, G. The Internet of Things: A survey. Computer Networks, v. 54, 2010.
MACHADO, J.; NANTES, J. F. D. a. A Visão Institucional do Processo de Rastreabilidade da Carne Bovina: Capturado em 2007.

NUKALA, R.; PANDURU, K.; SHIELDS, A.; RIORDAN, D.; DOODY, P.; WALSH, J. Internet of Things: A review from 'Farm to Fork'. Signals and Systems Conference (ISSC), 2016.

RAIMUNDO, L.; TAISCH, M.; TERZI, S. State of the art of technology in the Food sector value chain towards the IoT. 2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI), 2016.

SOBRINHO, O. G. Modelagem de um Sistema de Informação para Rastreabilidade na Indústria Vinícola Baseado em uma Arquitetura Orientada a Serviços. 2008. 132 (Mestrado). Departamento de Engenharia de Computação e Sistemas Digitais, Universidade de São Paulo.

SOBRINHO, O. G.; CUGNASCA, C. E. RASTREABILIDADE NA CADEIAS DE SUPRIMENTOS DO VINHO BRASILEIRO. Revista Brasileira de Agroinformática, v. 7, n. 1, p. 44-57, 2004.

WILSON, J. S. Sensor technology handbook. Newnes, 2004.

WU, N. C. et al. Challenges to global RFID adoption. Technovation, v. 26, n. 12, p. 1317-1323, 12// 2006. ISSN 0166-4972. Disponível em: < http://www.sciencedirect.com/science/article/pii/S016649720500146X >.

YAN, M. et al. The Application of the Internet of Things in Agriculture. 2011 International Conference on Future Computers in Education (Icfce 2011), Vol I, p. 232-235, 2011 2011.



Análise de agrupamento de dados climatológicos e de plantio de alface baseada no algoritmo *K-Means*.

José Leonardo Dos Santos Costa[1], *Diego Saqui*[2], *Tiago Gonçalves Botelho*[1], *Paulo César dos Santos*[1]

[1] Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais, Muzambinho, Minas Gerais, Brasil, joseleonardodacosta@gmail.com, tiago.botelho@muz.ifsuldeminas.edu.br, paulo.santos@muz.ifsuldeminas.edu.br

[2] Instituto Federal de Educação, Ciência e Tecnologia de Mato Grosso do Sul, Corumbá, Mato Grosso do Sul, Brasil, diego.saqui@ifms.edu.br

RESUMO

Atividades ligadas à agricultura são responsáveis por boa parte do PIB do brasil. Uma das grandes preocupações do ramo é a qualidade e quantidade produzida, pois variáveis incontroláveis podem afetar a produção. Por exemplo, algumas dessas variáveis são o clima e as pragas. Técnicas para auxiliar o produtor em tomadas de decisão são essenciais para o controle desses fatores. O objetivo deste trabalho é a construção de uma base de dados de produção de alface da horta Costa na cidade de Nova Resende/MG e a sua correlação com os dados climatológicos para posteriormente aplicar técnicas de aprendizado de máquina e auxiliar o produtor na tomada de decisão na plantação. Por meio da aplicação dos algoritmos *K-means*, criou-se grupos semelhantes na base de dados e junto a um especialista foi possível tirar conclusões da produção de alface. Um dos resultados foi a percepção de que a alta quantidade de chuva e média elevada da temperatura, influencia no nível de praga na horta e consequentemente em alfaces do tipo crespa, que são categorizadas como ruins. Outra

observação importante é que a alface do tipo lisa com muita chuva se sai melhor do que a crespa. Todos os resultados servem como base a decisões futuras na horta Costa.
PALAVRAS-CHAVE: *K-means*, Verdura, Base de dados, Agrupamento de dados, Mineração de dados.

ABSTRACT

Agriculture-related activities account for much of Brazil's PIB. One of the major concerns of the industry is the quality and quantity produced, since uncontrollable variables can affect production. For example, some of these variables are climate and plagues. Techniques to assist the producer in making decisions are essential to control these factors. The objective of this work is the construction of a database of the lettuce production of Costa's vegetable garden in the city of Nova Resende / MG and its correlation with the climatological data, to later apply machine learning techniques and to assist the producer in planting decisions. By means of the K-means algorithm, similar groups were created in the database and together with a specialist it was possible to draw conclusions about lettuce production. One of the results was the perception that the high amount of rain and high average temperature influences the amount of pest in the garden and consequently in crisp type lettuces that are categorized as bad. Another important observation is that lettuce of the smooth type with much rain does better than the crisp. All results serve as a basis for future decisions in Costa's vegetable garden.
KEYWORDS: K-means, Greenery, Database, Data grouping, Data mining.

INTRODUÇÃO

Segundo o Centro de Estudos Avançados em Economia Aplicada (CEPEA), de janeiro a dezembro de 2016, o produto interno bruto do agronegócio brasileiro apresentou acúmulo de crescimento de 4,48%, o ramo agrícola seguiu em alta com um aumento de 0,18%.

Uma das grandes preocupações atualmente na agricultura é qualidade e quantidade na produção. Pragas e fungos influenciam diretamente nos números e na qualidade da produção final. Um exemplo dessa relação é a lagarta *Helicoverpa armigera*, que tem sido um grande problema em muitas lavouras do país (RODRIGUES, 2013). A praga ataca com gravidade culturas como feijão, soja, algodão, milho e hortaliças. Desde sua entrada nas terras brasileiras a lagarta tem dado um prejuízo estimado aos agricultores que ultrapassa R$ 10

bilhões. Atualmente a lagarta foi localizada em todo o brasil e se espalha rapidamente devido a facilidade de migração das mariposas adultas. Técnicas de cultivos inconvenientes, como grandes extensões de monocultura e o uso sem controle de agrotóxicos, contribuem para a maximização da praga.

A plantação de alface no sul de Minas Gerais tem a agricultura familiar como um dos principais produtores da verdura. Esses produtores não dispõe de informações que possam ajudar a fazer uma correta avaliação de suas produções. Como há muitas variáveis que influenciam na produção da alface, eles enfrentam muitos problemas e não sabem o motivo. Métodos computacionais podem auxiliar esses produtores na investigação dessas dificuldades ocorridas no campo.

Segundo FACELI et al. (2011), técnicas de Aprendizado de Máquina podem auxiliar em problemas da agricultura como: manejo racional e do solo, combinação de diferentes culturas; barateamento e expansão da produção agrícola, sem prejuízos ao meio ambiente; previsão e controle de pragas e epidemias, agricultura de alta precisão. Conforme descrito por Mitchell (1998, p.2) "[...] um computador é dito aprender a partir de uma experiência E com respeito a alguma tarefa T e alguma medida de desempenho P, se seu desempenho na tarefa T, que é mensurada por P, melhora conforme a experiência E [...]". Essa definição caracteriza Aprendizado de Máquina que entendemos como uma área que estuda algoritmos que com o aumento de tarefas realizadas e maior experiência, tem uma melhora em seu desempenho.

Trabalhos de classificação, regressão, agrupamentos de dados e mineração de padrões frequentes estão sendo utilizados em vários dos problemas mencionados. Principalmente quando aplicado a agricultura de precisão, onde se busca reduzir custos e contaminação do solo com o emprego de tecnologias computacionais.

Com a mineração de dados e a utilização de algoritmos de agrupamentos espera-se reconhecer padrões que possam ser utilizados para serem aplicados na melhoria das plantações de alface.

MATERIAL E MÉTODOS

Levantamento de dados na horta de verduras

A horta de verduras onde foram levantados os dados está localizada no sul do Estado de Minas Gerais, na cidade de Nova Resende. Ela encontra na zona rural no bairro Penha, no Km 6 da rodovia LMG-846 que liga Nova Resende a cidade de Bom Jesus Da Penha.

Figura 1 – Imagem aérea de Horta de verdura Costa.



Fonte: GOOGLE MAPS. [2017]. Acesso em 19/02/2017

Na Figura 1 é apresentado a visão aérea da horta de verduras Costa. Pela medição feita através da ferramentas do GOOGLE MAPS, a horta tem área total de 10.685m^2.

Para o registro das informações foi usada a seguinte estratégia na plantação:

- No dia que as verduras foram plantadas foi criado um lote, com a data de plantação e o tipo da alface plantada, no canteiro foi colocado um placa de identificação com os dados relacionados.

- Durante o desenvolvimento do lote o especialista classificou o nível de praga do mesmo. O nível de praga é um valor inteiro variando de 0 a 10, sendo que, quanto maior o número de pragas, maior é o nível. Foram examinadas também outras variáveis de desenvolvimento da produção, para futuramente serem analisados juntos aos dados da base.

- Foi registrado também o dia em que o lote será finalizado para comercialização, ou

seja, será possível saber quantos dias o lote ficou ativo na horta. Sabendo essa informação, poderão ser finalizadas as variáveis climatológicas que também serão levantadas para relacionar o clima com a produção, assunto que será tratado no próximo tópico.

Aquisição dos dados climatológicos para a base de dados

Além de informações relacionadas a produção da alface, foi inserido na base de dados, conceitos relacionados ao clima durante o tempo que o lote ficou ativo na horta. A estação meteorológica a qual foi coletado os dados, fica também na cidade de Nova Resende/MG. Está localizado a uma latitude de -21,130678, longitude de -46,433896 e altitude de 1.200,0 m. A estação pertence a Cooperativa Regional de Cafeicultores de Guaxupé (COOXUPÉ), na unidade de Nova Resende. As informações da estação meteorológica são disponibilizados no site da COOXUPÉ.

Na Figura 2, é apresentado a imagem aérea entre a horta de verdura e a estação meteorológica. A distância entre elas é de 5,60 Km, essa medida foi obtida através da ferramentas do GOOGLE MAPS.

Figura 2 –Imagem da Distância entre a Horta e a Estação Metereológica..



Fonte: GOOGLE MAPS. [2017]. Acesso em 28/04/2017

As informações da estação são divulgadas em forma de relatório de período de tempo. Dentro desse relatório contém o dia e as informações climatológica dele, como:

- Temperatura média em ºc;
- Temperatura mínima em ºc;
- Temperatura máxima em ºc;
- Quantidade de chuva em mm;
- Porcentagem da umidade do ar;
- Porcentagem mínima umidade do ar;
- Porcentagem máxima umidade do ar;
- Velocidade média do vento em Km/h.

Análise da base de dados

Depois da fase de aquisição das informações, ou seja, levantamento dos dados da plantação de alface e aquisição das informações climatológicas de cada lote, foi iniciado a fase de pré-processamento de dados.

Nessa fase foi aplicado técnicas como eliminação manual de atributos, portanto, variáveis que não forem relevantes serão desconsiderados para uma futura análise. Foi executado nesta etapa a integração dos dados entre informações da plantação com dados climatológicos. Continuando a aplicação das técnicas, foi elaborado amostragem de dados e o tratamento dos dados desbalanceados. Por fim, executou a limpeza de dados onde foi tratado dados com ruídos, inconsistentes, redundantes e incompletos. Com a aplicação dessas técnica, espera-se melhorar a qualidade dos dados, minimizar ou eliminar problemas existentes na base (FACELI et al, 2011).

Aplicação do algoritmo de agrupamento

Foi aplicado na base de dados o algoritmo de agrupamento *K-means*, através do software de mineração de dados WEKA. O algoritmo *K-means* também conhecido como k-médias é utilizado procedimentos particionais, Segundo Filho (2015) "métodos de particionamento procuram obter uma única partição dos dados de entrada em um número fixo de grupos, geralmente por meio da otimização de uma função objetivo".

BARBARA (2000) apresentam os passos para encontrar os *clustering*:

1. Selecione K pontos como os centróides iniciais.
2. Atribuir todos os pontos ao centróide mais próximo.
3. Recalcule o centróide de cada *cluster*.

4. Repita os passos 2 e 3 até que os centróides não mudem.

Uns dos pontos essenciais do algoritmo é a escolha dos centróides iniciais, pois é feito de maneira aleatória. Isso pode levar ao problema dos mínimos locais, que é uma solução, mas não é a melhor. O algoritmo tem como objetivos criar *clustering*, que é a criação de grupos de dados com semelhanças (ALSABTI; RANKA; SINGH, 1997).

Dentro do WEKA pode escolher o número de *clustering* que o algoritmo vai agrupar. Na base de dados foi executado o algoritmo com a quantidade de grupos variando de 4 a 6. O resultado de cada execução foi apresentado a um especialista da área para a interpretação dos grupos gerados.

RESULTADOS E DISCUSSÃO

Com a execução do algoritmo *K-means* variando o número dos *cluster*, o especialista pode notar quais atributos da base influenciam na produção de alface. Empiricamente foi observado que com 4 e 6 *clusters* foi o ideal para fazer a análise.

Como pode-se observar na Tabela 1, para a alface do tipo crespa ser considerada ótima, o nível de chuva é o mais baixo em relação a outros *cluster* formado para mesma. Fica claro também, que o nível de praga do *cluster* 3 e a temperatura média são menores do que os grupos 1 e 4 da alface do tipo crespa que são considerados médios, enquanto o grupo 3 é ótimo.

Tabela 1 – Resultado da execução do algoritmo K-means com 4 agrupamento

Atributos	*Cluster*			
	1	2	3	4
Tipo da alface	crespa	lisa	crespa	crespa
Nível de Praga	7	8	6	8
Temperatura média	21.3996	21.3395	21.141	21.731
Quantidade de chuva	248.4875	298.1407	218.6611	240.4
Classificação	média	média	ótimo	média

Fica claro também que com o aumento do número dos *clusters* as informações relacionadas aos demais tipos de alface começam a aparecer. Na Tabela 1 só tinha informações relacionadas ao tipo crespa e lisa, já na Tabela 2 com seis *clusters*, acrescentou-se o tipo americana.

Com a análise dos dados da Tabela 2, no *cluster* número três verifica que o grande volume da chuva e a média da temperatura elevada influência no nível de praga e

consequentemente na classificação da alface do tipo crespa. Uma observação muito importante para próximas plantações, foi observada através do *cluster* número seis, pois observa-se um grande volume de chuva e classificação ótima para alface do tipo lisa.

Tabela 2 – Resultado da execução do algoritmo *K-means* com 6 agrupamento

Atributos	*Cluster*					
	1	2	3	4	5	6
Tipo da alface	lisa	crespa	crespa	americana	crespa	lisa
Nível de Praga	8	7	8	6	6	6
Temperatura média	21.4187	20.9594	21.617	21.8967	21.3442	21.3862
Quantidade de chuva	229.375	239.8609	309.5444	257.4143	173.4353	318
Classificação	média	boa	ruim	média	ótimo	ótimo

CONCLUSÕES

Com os resultados obtidos através da aplicação do algoritmo *k-means*, conclui-se que com a aplicação das etapas de processamento de dados, em uma base de dados da produção de alface correlacionados com clima, pode-se extrair decisões que auxiliam os produtores de verdura. Através dos resultados verificou-se que o clima influencia diretamente no nível de praga e na classificação final da alface. Assim, com a aplicação de técnicas de agrupamento o especialista foi favorecido na análise dos resultados e com isso, pôde planejar-se para tomada de decisão das próximas plantações de alface com clima semelhante.

REFERÊNCIAS

ALSABTI, A.; RANKA, S.; SINGH, V. An efficient k-means clustering algorithm. 1997.

BARBARA, D. An Introduction to Cluster Analysis for Data Mining. 2000.

CEPEA; PIB do Agronegócio BRASIL**.** Dez. 2016.

FACELI, K. et al. Inteligência Artificial : Uma Abordagem de Aprendizagem de Máquina. 1. ed. Rio de Janeiro, RJ, 2011.

FILHO , V. M. ALGORITMOS PARTICIONAIS SEMISSUPERVISIONADOS COM PONDERAÇÃO AUTOMÁTICA DE VARIÁVEIS. Universidade Federal de Pernambuco, Pernambuco: Recife, 2015.

MITCHELL, T, M. Machine Learning, 1997.

RODRIGUES, P. O Ataque da Lagarta Práticas de cultivo inadequadas facilitam a disseminação da praga Helicoverpa Armigera. Hortaliças em Revistas, n. 9, p. 6-9, 2013.

Identificação de padrões tecnológicos do sistema de pecuária de corte desenvolvido no Cerrado

Stanley R. de M. Oliveira[1], Urbano G P de Abreu[2], Maria do Carmo Ramos Fasiaben[1], Luis Gustavo Barioni[1], Helano Póvoas de Lima[1], Maxwell Merçon Tezolin Barros Almeida[3], Octávio Costa de Oliveira[3]

[1] Embrapa Informática Agropecuária - CNPTIA, Campinas, São Paulo, Brasil
{stanley.oliveira, maria.fasiaben, luis.barioni, helano.lima}@embrapa.br
[2] Embrapa Pantanal - CPAP, Corumbá, Mato Grosso do Sul, Brasil
urbano.abreu@embrapa.br
[3] Instituto Brasileiro de Geografia e Estatística - IBGE, Rio de Janeiro, Brasil
{maxwell.almeida, octavio.oliveira}@ibge.gov.br

RESUMO

O objetivo deste trabalho é inferir sobre os padrões de desenvolvimento tecnológico dos sistemas de produção de pecuária de corte no Cerrado. Para tanto, foram empregadas técnicas de aprendizagem de máquina e análise multivariada sobre dados do Censo Agropecuário 2006, em escala municipal. Os resultados foram considerados aderentes à realidade, permitindo distinguir diferentes tipos de exploração pecuária entre os municípios do Cerrado e seus padrões de distribuição geográfica.

PALAVRAS-CHAVE: Tipologias, Clusterização, Sistemas de produção.

ABSTRACT

The objective of this work is to infer about the technological development patterns of the production systems of beef cattle in the Cerrado. To do so, machine learning techniques and multivariate analysis were applied to the data from the 2006 Census of Agriculture, at a municipal scale. The results were considered adherent to the reality, allowing to distinguish different types of types of livestock farming among the municipalities of the Cerrado and their patterns of geographical distribution.

KEYWORDS: Typologies, Clustering, Production systems.

INTRODUÇÃO

O Censo Agropecuário brasileiro de 2006, realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), contabilizou 62,7 milhões de bovinos no Cerrado (35,6% do rebanho nacional), o mais numeroso efetivo bovino dos biomas brasileiros. Dos 799.534 estabelecimentos agropecuários contabilizados no Cerrado, 104.288 informaram possuir mais de 50 cabeças de bovinos e ter o corte como a principal finalidade do rebanho.

O Cerrado Brasileiro, localizado na porção central do país, ocupa área contínua de 204,7 milhões de hectares, isto é, aproximadamente 24% da superfície do Brasil. Suas distintas condições ecológicas, seu diferenciado padrão espaço-temporal de ocupação e de investimentos em infraestrutura nas diferentes localidades do Bioma, bem como a conjuntura macroeconômica e os estímulos de mercado ao longo das últimas quatro décadas, determinaram um mosaico de características econômicas e sociais explicando os impactos sobre os sistemas de produção de pecuária de corte (MUELLER; MARTHA, Jr. 2008).

O número de bovinos no Cerrado aumentou em 91,7% entre 1975 e 1996, passando de 24,9 milhões para 47,8 milhões de cabeças. No período de 1990-2005, o rebanho bovino cresceu 34,5% (de 72,17 milhões para 97,07 milhões de cabeças) correspondendo a uma taxa média anual de crescimento de quase 2%. Esse crescimento, no entanto, foi mais rápido no período de 1996-2005 (2,5% ao ano), em comparação ao período de 1990 a 1996 (1,2% ao ano). Para comportar esse rápido crescimento do rebanho bovino no Cerrado, nas décadas citadas, houve aumentos significativos na área de pastagem cultivada, de maior capacidade de suporte do que as pastagens nativas (MARTHA Jr. et al., 2007).

Uma alternativa para identificação de padrões de desenvolvimento tecnológico no Cerrado é o uso de técnicas de aprendizado não-supervisionado, notadamente clusterização, também conhecida como análise de clusters ou análise de agrupamentos (LINDEN, 2009). Um dos métodos de clusterização mais utilizados é o de particionamento, no qual forma um conjunto inicial de k partições, sendo o parâmetro k o número de partições a serem consideradas para o conjunto de dados (RAJESH, 2011). O algoritmo de particionamento mais conhecido na literatura é o K-Means (McQUEEN, 1967).

A motivação pelo uso do algoritmo K-Means para identificação de padrões de desenvolvimento tecnológico se dá pelos seguintes aspectos: a) o algoritmo é simples e escalável; b) pode ser encontrado em diversos pacotes de software (WEKA, R, SPSS, SAS, MATLAB, entro outros); c) é sensível a ruídos, isto é, agrupa todos os objetos (no caso em estudo, os objetos representam municípios) disponíveis no conjunto de dados sob análise.

O objetivo deste artigo é o de, por meio de técnicas de aprendizagem de máquina e análise multivariada, inferir sobre os padrões de desenvolvimento tecnológico dos sistemas de produção de pecuária de corte no Cerrado em escala municipal, com a utilização de dados do Censo Agropecuário 2006.

MATERIAL E MÉTODOS

A região de estudo compreende o Cerrado brasileiro, incluindo os estados da região Centro-Oeste (Goiás, Mato Grosso, Mato Grosso do Sul e Distrito Federal), além do sul do Pará e Maranhão, interior do Tocantins, oeste da Bahia e Minas Gerais, bem como o norte de São Paulo. É constituído por várias características de vegetação, é classificado em subsistemas: de campo, de cerrado, de cerradão, de matas, de matas ciliares e de veredas e ambientes alagadiços. A vegetação do Cerrado encontra-se em uma região onde o clima que predomina é o tropical, apresenta duas estações bem definidas, sendo uma chuvosa, entre outubro e abril, e outra seca, entre maio e setembro.

Para atender ao presente trabalho, foi utilizada tabulação especial do Censo Agropecuário 2006, fornecida pelo IBGE (IBGE, 2011). A tabulação consistiu de dados de agregados por município de estabelecimentos agropecuários do Bioma Cerrado com mais de 50 cabeças de bovinos e que declararam ter o corte como principal finalidade do rebanho. Por questão de sigilo estatístico foram descartados os municípios com menos de três estabelecimentos.

Após o tratamento dos dados, foi realizado o processo essencial de extração dos padrões de desenvolvimento tecnológico dados para o bioma Cerrado. A técnica escolhida foi a análise de clusters (clusterização), por meio do método de particionamento. A similaridade entre objetos (municípios) foi medida pela distância Euclidiana entre dois objetos (HAN et al., 2011). Quanto menor for a distância entre dois objetos, maior será a similaridade entre eles. O algoritmo utilizado nas análises foi o K-Means (McQUEEN, 1967), que é um método de particionamento, em que o centro de cada cluster é representado pelo valor médio dos objetos no cluster. Possui como entrada o número de clusters (k) e o conjunto de dados (D) contendo n objetos, e como saída um conjunto de k cluster.

Para determinar o número de clusters, foram testados vários valores de k (k = 2, ..., 16). Em seguida, para minimizar a subjetividade na escolha do valor de k, foi utilizado o método do cotovelo (HAN et al., 2011). Este método é simples e permite uma boa aproximação de k por meio do menor erro quadrático médio, definido pelo ponto de máxima curvatura para a trajetória da dissimilaridade interna e externa. A determinação do número

apropriado de clusters foi então definido no ponto onde o gráfico apresenta uma curva acentuada, à semelhança de um cotovelo. O algoritmo K-Means e o método do cotovelo foram implementados no ambiente RStudio (RStudio, 2017), que é um software livre de ambiente de desenvolvimento integrado para R, uma linguagem de programação para gráficos e cálculos estatísticos.

A discriminante canônica é uma técnica para auxiliar na classificação ou na pré-classificação de um elemento em determinado grupo, além de testar a hipótese de que os centróides do grupo de um conjunto de variáveis independentes para os outros grupos são iguais. O teste para a significância da função discriminante é uma medida generalizada da distância entre os centróides dos grupos (HAIR et al., 2009). O método foi inicialmente desenvolvido por Fisher (1936), com objetivo de maximizar a distância entre grupos, não levando em consideração o problema de minimização das probabilidades de erro no processo discriminatório. A função discriminante linear de Fisher (FDLF) consiste em separar objetos, sendo separados ou classificados com base nas medidas de p variáveis aleatórias.

O objetivo foi a criação dessa regra de reconhecimento de padrões e classificação. Assim as observações multivariadas X's foram transformadas nas observações univariadas Y's, tal que os Y's dos agrupamentos π_1 e π_2 fossem separados em relação às médias dos dois grupos tanto quanto possível. A ideia de estimar combinações lineares de X para criar os Y's, dado que essas combinações são funções de X. Seja $\mu_1 Y$ a média dos Y's obtidos dos X's pertencentes a π_1, e μ_2 Y a média dos Y's obtidos dos X's pertencentes a π_2 e, considerando a matriz de covariância, Σ, como sendo a mesma para ambos grupos, selecionou-se a combinação linear que maximiza a razão entre o "quadrado da distância entre as médias" e a "variância de Y", ou seja: Essa razão é maximizada por: $C = \Sigma^{-1} (\mu_1 - \mu_2)$, assim: $Y = C'X = (\mu_1 - \mu_2)' \Sigma^{-1} X$, que é conhecida como FDLF (MARQUES; LIMA, 2002).

Os agrupamentos dos municípios formados pelo método de Wrapper foram classificados, e em seguida submetidos à análise de discriminante canônica com objetivo de verificar se a classificação dos municípios foi significativamente diferente. Seis variáveis foram utilizadas:

1) número de estabelecimento de cria/número de estabelecimento de gado de corte (NECNGC); 2) número de estabelecimento de cria e recria /número de estabelecimento de gado de corte (NECRNGC); 3) número de estabelecimentos que fizeram suplementação com ração mais estabelecimentos que utilizaram ração e sal mineral / número de estabelecimento de gado de corte (NERNGC); 4) número de estabelecimento que receberam orientação técnica/ número de estabelecimento de gado de corte (NEONGC); 5) somatório das áreas de

lavoura permanente mais áreas com lavoura temporárias / somatório das áreas de lavoura permanente; áreas com lavoura temporárias; áreas com forrageiras para corte; área com viveiro de mudas; áreas de pastagens; áreas de silvicultura e áreas com sistemas agroflorestais (SATSALP); e 6) valor da produção dos bovinos / valor total da produção (VPBVT), em reais.

RESULTADOS E DISCUSSÃO

Para determinar o número de clusters, foram testados vários valores de k (k = 2, ..., 16). A escolha do número apropriado de clusters foi definido pelo método do cotovelo, indicando k = 6 para o banco de dados de desenvolvimento tecnológico da pecuária de corte para o Cerrado brasileiro (Figura 1).

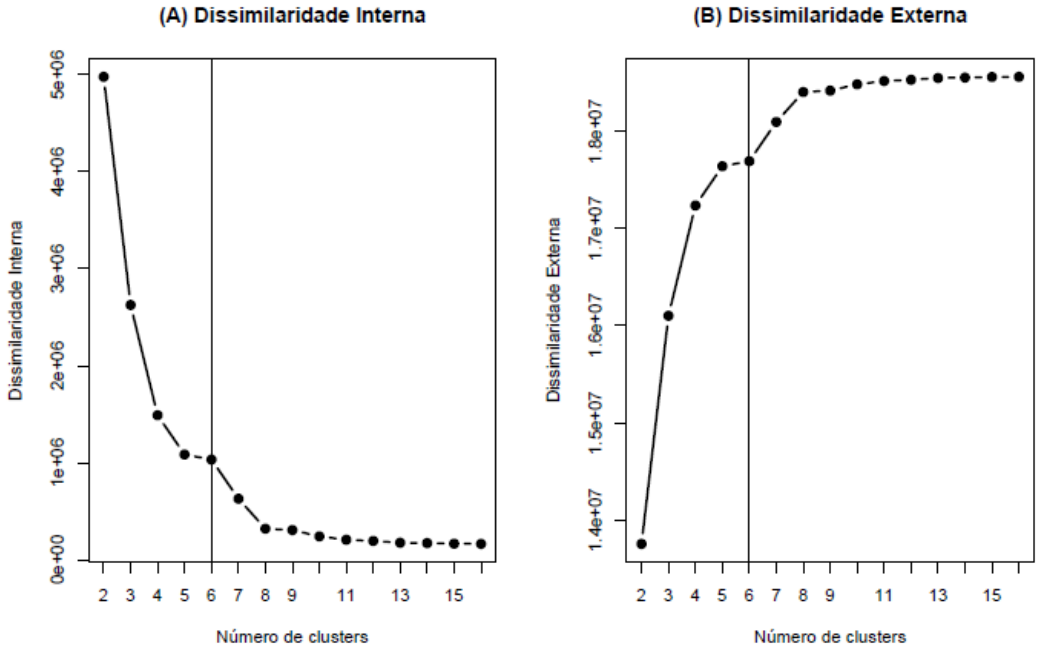


Figura 1 – Dissimilaridade interna e externa em função do número de clusters para o banco de dados de desenvolvimento tecnológico da pecuária de corte para o Cerrado brasileiro.

Na Tabela 1 são mostradas as médias dos percentuais das variáveis de cada cluster, em escala municipal. Observamos que os sistemas com predominância de sistemas de cria (NECNGC) são predominantes nos Clusters 0 e 1, sendo interessante notar que a diferença entre eles é o maior percentual de sistemas de cria e recria (NECRNGC) no Cluster 1. Por outro lado, os Clusters 1 e 2 apresentaram maior valor da produção de bovinos em relação ao valor de produção total (VPBVT), e em paralelo maior presença de área com pastagens (SAPSAT). O Cluster 3 também apresentou o mesmo perfil, entretanto com percentual menor nas duas variáveis. A utilização de suplementos (NERNGC) e orientação técnica (NEONGC),

percentualmente foram mais presentes nos grupos 4 e 5, provavelmente em função da necessidade da orientação técnica para fornecimento adequado de suplementos.

Após a classificação dos municípios os grupos foram submetidos à técnica multivariada de análise discriminante canônica (FDLF). Na Figura 2, são apresentadas as duas primeiras variáveis canônicas dos seis grupos. Todos os grupos apresentaram a distância de Mahalanobis entre suas médias canônicas significativamente diferentes ($p < 0,01$), o que nos permite inferir que os grupos formados possuem diferentes padrões tecnológicos no desenvolvimento da atividade de pecuária de corte.

Tabela 1 – Média e erro padrão da média, das variáveis selecionadas para a 'clusterização' dos municípios do Cerrado, e número de municípios em cada cluster observado.

	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
	143 municípios	241 municípios	251 municípios	139 municípios	88 municípios	194 municípios
NECNGC (%)	0,2857± 0,0140	0,3442±0,0113	0,1334 ± 0,0053	0,1917±0,0102	0,1288±0,0131	0,1296± 0,0066
NECRNGC (%)	0,1604± 0,0105	0,2601± 0,0100	0,1720 ± 0,0063	0,0646±0,0049	0,1259±0,01091	0,1478±0,0079
NERNGC (%)	0,2600± 0,0138	0,2017±0,0097	0,4141± 0,0110	0,1973±0,013	0,5347±0,0261	0,4359±0,0115
NEONGC (%)	0,2603±0,0116	0,2386±0,0095	0,5119± 0,0106	0,2232±0,0119	0,7491±0,0209	0,6482± 0,0110
SAPSAT (%)	0,7529±0,0138	0,9007±0.0069	0,9071±0.0065	0,7991±0,0148	0,3406±0,0161	0,7547±0,0084
VPBVT (%)	0.2728±0.0134	0,8009±0.0091	0,8136±0.0085	0,7839±0,0129	0,1731±0,0191	0,2859±0.0117

Fonte: Dados do Censo Agropecuário 2006 trabalhados pelos autores.

Nota: NECNGC- número de estabelecimento de cria/número de estabelecimento de gado de corte, NECRNGC- número de estabelecimento de cria e recria /número de estabelecimento de gado de corte, NERNGC-número de estabelecimentos que fizeram suplementação com ração mais estabelecimentos que utilizaram ração e sal mineral / número de estabelecimento de gado de corte, NEONGC-número de estabelecimento que receberam orientação técnica/ número de estabelecimento de gado de corte, SAPSAT- somatório das áreas de pastagem natural mais áreas com pastagens degradadas e não degradadas / somatório das áreas com lavoura permanente e não permanente, áreas com forrageiras para corte, cultivo de flores, casa de vegetação e viveiro de mudas, áreas com silvicultura, sistemas florestais, pastagem natural mais áreas com pastagens degradadas e não degradadas, VPBVT-valor da produção dos bovinos / valor total da produção em reais.

Após a classificação dos municípios os grupos foram submetidos a técnica multivariada de análise discriminante canônica (FDLF). Na Figura 2, são apresentadas as duas primeiras variáveis canônicas dos seis grupos. Todos grupos apresentaram a distância de Mahalanobis entre suas médias canônicas significativamente diferentes ($p < 0,01$), o que nos permite inferir que os grupos formados possuem diferentes padrões tecnológicos no desenvolvimento da atividade de pecuária de corte.

Apresenta-se a seguir a interpretação e descrição dos diferentes clusters e a Figura 3 mostra a sua localização geográfica.

Cluster 0 - "Municípios com predomínio da fase de cria, produção extensiva"

Abrange 14% dos municípios (143 municípios) e 13% dos estabelecimentos com bovinos de corte no Bioma Cerrado, respondendo por 10% do total de bovinos e 12% da área total de pastagens. Nestes municípios, o rebanho médio é de 332 cabeças, a área média de pastagens é de 381 ha e a área média dos estabelecimentos que têm bovinos é de 1537 ha. Da área explorada pelos estabelecimentos desse cluster, 75% são destinados a pastagens e 11%, a lavouras. No Cluster 0, em 26% dos estabelecimentos se declarou usar ração e em 7% adubar as pastagens. O percentual de animais confinados não atingiu 2%. Somente 26% dos

estabelecimentos produtores de bovinos de corte dos municípios desse cluster declararam receber orientação técnica. Embora a maior parte da área explorada seja destinada a pastagens, nesse cluster apenas 27% do valor total da produção provêm da bovinocultura. A taxa de lotação é de 0,69 UA/hectare, a menor entre todos os clusters. Os municípios do Cluster 0 se encontram especialmente numa faixa que abrange o Sul do Maranhão, Sudoeste do Piauí, Oeste da Bahia e Centro-Norte de Minas Gerais.

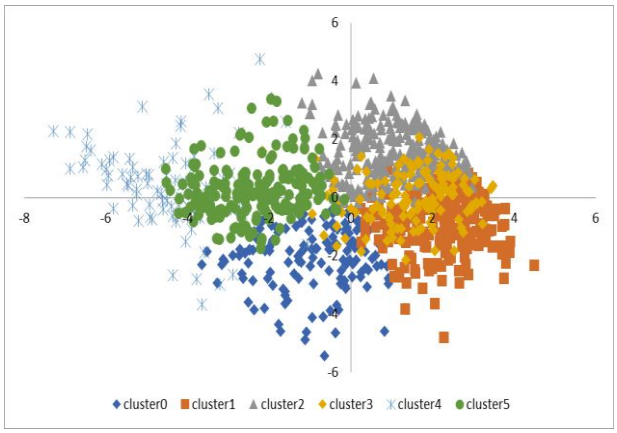


Figura 2 – Variáveis canônicas discriminadas nos seis grupos de municípios do Cerrado, conforme perfil tecnológico do sistema de produção de gado de corte.

Cluster 1 - "Municípios com importância das fases de cria e recria"

Representa 23% dos municípios (241 municípios), 29% dos estabelecimentos, 22% do total de bovinos e 25% do total de pastagens dos municípios do Bioma Cerrado (produtores de bovinos de corte). Nestes municípios, o tamanho médio do rebanho é de 338 cabeças, a área média de pastagens é de 374 ha e a área total média dos estabelecimentos, 934 ha. Aqui, 90% da área explorada pelos estabelecimentos com bovinos de corte são destinadas a pastagens. No Cluster 1, 20% dos estabelecimentos declararam usar ração e 6% adubar as pastagens. O percentual de animais confinados foi de 1,6% (o menor entre todos os clusters). Nos municípios que compõem o Cluster 1, 24% dos estabelecimentos declararam receber orientação técnica. A taxa de lotação é de 0,75 UA/ hectare. Nesse cluster, 80% do valor da produção correspondem à bovinocultura de corte. Os municípios que fazem parte do Cluster 1 estão localizados predominantemente nos estados de Goiás, Tocantins e Maranhão e leste do Mato Grosso.

Cluster 2 - "Municípios com importância da bovinocultura de corte de ciclo completo"

Neste cluster enquadram-se 24% dos municípios do Bioma Cerrado produtores de bovinos de corte (251 municípios), 32% dos estabelecimentos, 49% do total de bovinos e 45% do total de pastagens. A área média dos estabelecimentos com bovinos dos municípios que

compõem o Cluster 2 é de 1093 ha, com uma média de 600 ha de pastagens. O número médio de bovinos por estabelecimento é de 668 cabeças. Da área explorada com agropecuária pelos estabelecimentos, 91% são destinados a pastagens. Constata-se, neste cluster, uma maior importância da produção de bovinos de corte a ciclo completo (cria, recria e engorda). Nos municípios que compõem esse cluster, 41% dos estabelecimentos declararam usar ração e perto de 15% fazer adubação de pastagens. A porcentagem de animais confinados chegou a 7%. Pouco mais da metade dos estabelecimentos produtores de bovinos de corte dos municípios desse cluster declararam receber assistência técnica (51%). A maior parte do valor da produção deste grupo provém da produção de bovinos (81%). O cluster apresenta uma taxa de lotação média de 0,98 UA/hectare. Os municípios que compõem o Cluster 2 distribuem-se especialmente nos estados do Mato Grosso do Sul, Goiás e interior de Minas Gerais.

Cluster 3 - "Municípios com importância da produção a ciclo completo e da fase de engorda a pasto"

Representa 13% dos municípios (139 municípios), 8% dos estabelecimentos, 5% do total de bovinos e 5% do total de pastagens. A área média dos estabelecimentos que têm bovinos é de 1028 ha, a área média de pastagens/estabelecimento é de 253 ha e o número médio de bovinos por estabelecimento é de 264 cabeças. Cerca de 80% da área explorada pelos estabelecimentos são destinados a pastagens. Referindo-se à tecnologia empregada na bovinocultura, 20% dos estabelecimentos dos municípios que fazem parte do Cluster 3 declararam usar ração, e 4% informaram adubar as pastagens. O percentual de animais confinados foi de pouco mais de 3%. Apenas 22% dos estabelecimentos produtores de bovinos de corte dos municípios declararam receber orientação técnica. Neste cluster, 78% do valor da produção provêm da bovinocultura de corte. A taxa de lotação é de 0,86 cabeças/hectare. Os municípios que pertencem ao Cluster 3 concentram-se na porção mais ao norte do Cerrado brasileiro, em especial no estado do Maranhão.

Cluster 4 - "Municípios com diversificação lavoura/pecuária e maior presença de confinamento"

O Cluster 4 é composto por 88 municípios (8%), engloba 3% dos estabelecimentos e abriga 2% do total de bovinos e 2% do total de pastagens dos seis clusters de bovinos de corte do Cerrado. Nesse cluster, a área média dos estabelecimentos é de 2789 ha, a área média de pastagens/estabelecimento é de 258 ha e o número médio de cabeças por estabelecimento é de 351. Nesse cluster, 61% da área explorada se destinam a lavouras, enquanto 34% se referem a pastagens. Aqui, 53% dos estabelecimentos declararam usar ração e 26% fazer adubação de pastagens. O percentual de animais confinados foi de 18%, e em 75% dos estabelecimentos se

declarou a presença de orientação técnica. A maior parte do valor da produção deve provir da produção vegetal, já que os bovinos representam um percentual de apenas 17% desse valor total. A taxa de lotação é a mais alta entre todos os clusters: 1,73 UA/hectare. Os municípios que integram o Cluster 4 se encontram em maior número no Sul de Minas, Norte de São Paulo e região Central de Mato Grosso.

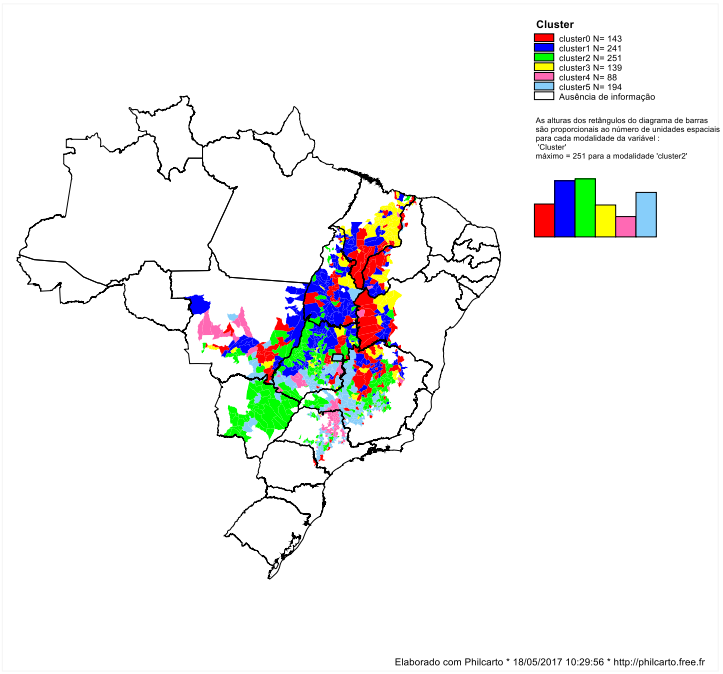


Figura 3- Localização geográfica dos clusters de municípios produtores de bovinos de corte no Cerrado em 2006.

Cluster 5 - "Municípios com importância da bovinocultura de corte de ciclo completo e de lavouras"

Abrange 18% dos municípios (194 municípios) e detém 16% dos estabelecimentos, que respondem por 13% do total de bovinos e 11% do total da área de pastagens. Nos municípios desse cluster, a área média dos estabelecimentos é de 1093 ha e a área média de pastagem/estabelecimento é de 311 ha. O tamanho médio do rebanho/estabelecimento é se 368 cabeças. Da área explorada pelos estabelecimentos desse cluster, 75% são destinados a pastagens e 19%, a lavouras. No Cluster 5, em 44% dos estabelecimentos se fez uso de ração e em 19% se adubou as pastagens. O percentual de animais confinados foi de 7%. Em 65% dos estabelecimentos nos municípios desse cluster se declarou receber orientação técnica. Nesse cluster, 29% do valor total da produção provêm da bovinocultura. A taxa de lotação é de 1,06 UA/hectare, atrás apenas do Cluster 4. Os municípios do Cluster 5 aparecem especialmente no Centro-Sul de Goiás, DF, Centro-Leste de Minas Gerais e região Centro-Sul de São Paulo.

CONCLUSÕES

Os resultados demonstram que os dados obtidos pelo Censo Agropecuário 2006 são suficientes para distinguir tipos de exploração pecuária entre os municípios do Cerrado. Os agrupamentos possuem padrão geográfico definido e dentro do esperado. Na região mais ao sul do Cerrado (Sul de Minas, Mato Grosso do Sul, Sul de Minas e Sul de Goiás) predominam os clusters 2, 4 e 5, enquanto que os clusters 0, 1 e 3 predominam nos estados mais ao norte do bioma. Os biomas mais ao sul estão associados a sistemas mais tecnificados, envolvendo integração com lavouras, ciclo completo e confinamento. Os municípios mais ao norte estão associados à produção mais extensiva e às fases de cria e recria, embora haja a presença de ciclo completo e engorda, particularmente no Maranhão. O Mato Grosso foi o estado menos padronizado em termos de tipos de sistema produtivo, apresentando todos os tipos identificados.

REFERÊNCIAS

FISHER R. A. The use of multiple measurements in taxonomic problems. **Annals of eugenics**, v. 7, p. 179-188, 1936.

HAIR Jr; BLACK, W.C.; BABIN, B. J.; ANDERSON, R. E.; TATHAM, R. L. **Análise multivariada de dados.** 6. ed. Porto Alegre: Artmed, 2009. 688p.

HAN, J.; KAMBER, M.; PEI, J. **Data mining:** concepts and techniques. 3rd edition. San Francisco: Morgan Kaufmann Publishers, p.770, 2011.

IBGE. **Tabulações especiais do Censo Agropecuário 2006**. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística, 2011.

LINDEN, R. Técnicas de Agrupamento. Revista **de Sistema de Informação da FSMA**, n. 4, 2009.

MARQUES, J. M.; LIMA, J. D. de. A estatística multivariada na análise econômico-financeira de empresas. **Revista da FAE**, v. 5, p. 51-59, 2002.

MARTHA JR., G. B., VILELA, L., BARCELLOS, A. O., SOUZA, D. M. G., BARIONI, L. G. Pecuária de corte no Cerrado: aspectos históricos e conjunturais. In: CERRADO: USO EFICIENTE DE CORRETIVOS E FERTILIZANTES EM PASTAGENS, 2007, Brasília, **Anais...** Brasília:Embrapa Cerrados, 2007. p. 17-42.

McQUEEN, J. B. Some Methods for Classification and Analysis of Multivariate Observations. In: BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTIC AND PROBABILITY, 5., 1967, Berkeley. Proceedings, p. 281-297.

MUELLER, C. C.; MARTHA Jr, G. B. A agropecuária e o desenvolvimento sócio-econômico recente do Cerrado. In: SIMPÓSIO NACIONAL CERRADO, 9.; SIMPÓSIO INTERNACIONAL SAVANAS TROPICAIS, 2., 2008, Brasília. **Anais...** Brasília: Embrapa Cerrados, 2008. p. 1-41.

RAJESH, D. Application of Spatial Data mining for Agriculture. **International Journal of Computer Applications**, v. 15, n. 2, 2011.

RStudio – Open source and enterprise-ready professional software for R. Disponível em: https://www.rstudio.com. Consultado em 3 de maio de 2017.

Metodologia baseada em Algoritmos de Aprendizado de Máquina para geração de séries espaço-temporais de precipitação

Henrique Lima Cará de Oliveira[1], *Stanley Robson de Medeiros Oliveira*[2], *José Eduardo Boffino de Almeida Monteiro*[2]

[1] Bolsista Iniciação Científica, Embrapa Informática Agropecuária, Campinas, São Paulo, Brasil, henriquelimacoliveira@gmail.com
[2] Embrapa Informática Agropecuária, Campinas, São Paulo, Brasil,

{stanley.oliveira, eduardo.monteiro}@embrapa.br

RESUMO

O objetivo deste trabalho foi desenvolver uma metodologia baseada em algoritmos de aprendizado de Máquina (*Random Forest* e *OversamplingR Intervals*) para gerar séries espaço-temporais de precipitação. Foi definida uma região de estudo de formato retangular, incluindo a metade norte do Estado de São Paulo e parte do sul de Goiás, a metade sul de Minas Gerais e o Estado do Espírito Santo. A região foi escolhida por ser climaticamente bastante heterogênea e por conter muitas estações meteorológicas de diversas instituições, principalmente ANA e INMET. Foram utilizadas as séries temporais de precipitação e de temperatura máxima e mínima disponíveis na região, compreendendo o período de 01/01/1999 a 31/12/2013. Também foram utilizadas as bases externas AgMERRA, AgCFSR, radiação solar GL 1.2 CPTEC, TRMM e Nasa Power, cujos dados estão espacialmente dispostos em grids que cobrem a região de estudo. A região de estudo foi subdividida em formato de grade regular com resolução de 0,5º (latitude e longitude), resultando em 280 quadrículas, sendo 28 na horizontal e 10 na vertical. Para cada quadrícula foram ajustados vários modelos preditivos para precipitação diária e acumulada de 10 dias. Os resultados revelaram um bom ajuste dos modelos preditivos em relação aos valores previstos e observados para as quadrículas avaliadas, indicando um grande potencial da metodologia proposta tanto para a imputação de dados ausentes quanto para a geração de séries espaço-temporais em regiões sem a presença de dados medidos.

PALAVRAS-CHAVE: Séries temporais, Otimização, Modelos preditivos, Agrometeorologia, Imputação de dados.

ABSTRACT

This work aimed to develop a methodology, based on the Random Forest algorithm, to generate space-time series of temperature and precipitation. A rectangular study region was defined, including the northern half of the State of São Paulo and part of southern Goiás, the southern half of Minas Gerais and the State of Espírito Santo. This region was chosen because it is climatically very heterogeneous and also because it contains many meteorological stations of several institutions, mainly ANA and INMET. We used time series of precipitation

and maximum and minimum temperature available in the region, covering the period from 01/01/1999 to 12/31/2013. We also used the external bases AgMERRA, AgCFSR, solar radiation GL 1.2 CPTEC, TRMM and Nasa Power, whose data are spatially arranged in grids covering the region of study. The study area was divided into a regular grid with a resolution of 0.5º (latitude and longitude), resulting in 280 squares, 28 horizontally and 10 vertically. For each square, several predictive models were fit for daily and accumulated 10-day precipitation. The results showed a good fit of the predictive models in relation to the predicted and observed values for the evaluated squares, indicating a great potential of the proposed methodology both for the imputation of absent data and for the generation of space-time series in regions without the presence of measured data.

KEYWORDS: Time series, Optimization, Predictive modeling, Agrometeorology, Data imputation.

INTRODUÇÃO

A variabilidade da distribuição espaço-temporal das chuvas é uma característica muito importante para os diversos setores econômicos, principalmente para a agricultura de sequeiro, que depende diretamente do regime pluviométrico para seu crescimento e produção. O Zoneamento Agrícola de Risco Climático, instrumento de política agrícola e fundamental para modernização tecnológica da agricultura, também depende de informações sobre precipitação. O conhecimento dessa característica pode orientar a tomada de decisão sobre o planejamento agrícola e as medidas necessárias para minimizar os danos decorrentes da irregularidade dessa variável.

Seguramente, um dos grandes desafios relacionados ao planejamento agrícola refere-se à baixa disponibilidade e qualidade de informações meteorológicas no Brasil. Por exemplo, interrupções e erros em séries de dados gerados por estações meteorológicas são relativamente comuns devido a várias razões (DUMEDAH; COULIBALY, 2011; MWALE et al., 2012). Em estações automáticas, essas razões incluem desde o simples esgotamento de bateria até falha permanente de sensores ou da estação, que podem permanecer inativos até seu eventual reparo ou substituição, por dias ou meses. Em estações convencionais, as falhas estão mais associadas à ausência do observador meteorológico ou à quebra ou falha dos instrumentos de medição. Por outro lado, mesmo quando registrados e armazenados em um banco de dados, os valores medidos estão sujeitos a imprecisões e erros de medida, de procedimentos de cálculo, de anotação pelo observador, de cópia ou transmissão de dados, de armazenamento ou transposição para tabelas. Em estudos agrometeorológicos, erros desse tipo implicam, quase sempre, na perda dos resultados e da análise de todo um ciclo de cultura, safra ou ano, no local da ocorrência.

Outro aspecto importante é a distribuição da informação no espaço. Estudos de grande abrangência espacial, na escala estadual, regional ou nacional, envolvendo muitas séries temporais, aumentam enormemente o volume e as necessidades em processamento de dados. A análise espacializada do território nacional é, muitas vezes, limitada pela baixa densidade dos pontos de observação, ou seja, de estações meteorológicas em operação (XAVIER et al., 2016).

Usuários de dados muitas vezes se utilizam de métodos para estimar dados em locais sem medida ou completar as falhas nas séries temporais, simplesmente porque não querem ou não podem arcar com a perda daquele local e ano em sua análise ou estudo. Geralmente, o preenchimento, ou imputação de dados, é adotado em falhas curtas, de poucos dias. No entanto, quando as falhas são recorrentes ou por períodos mais longos, esse processo de imputação pode se tornar tanto mais trabalhoso quanto deletério devido ao aumento da incerteza (MWALE et al., 2012).

Dessa forma, a formação de uma base de dados meteorológicos requer, muitas vezes, um processo de reconstrução de séries espaço-temporais (FENG; QIAN, 2004; VICENTE-SERRANO et al., 2010). Tanto para a geração de estações virtuais, em locais sem medidas de estação meteorológica, quanto para o preenchimento de falhas, diversos métodos independentes ou complementares podem ser empregados (STEINACKER et al., 2011; CAMPOZANO et al., 2014).

Técnicas de mineração de dados são uma alternativa promissora para estimar dados em séries temporais de precipitação, já que alguns algoritmos utilizados no processo de descoberta de conhecimento são capazes de capturar a relação não-linear desses dados, uma característica presente em dados de precipitação.

Em particular, o algoritmo *Random Forest* (BREIMAN, 2001) é eficiente para estimar séries temporais em várias áreas do conhecimento (VILLANUEVA, 2006; KANE et al., 2014). Suas principais vantagens são: (a) tem natureza não-paramétrica; (b) é um dos algoritmos de aprendizado de máquina com alta taxa de acurácia; (c) pode lidar com milhares de variáveis de entrada sem a necessidade prévia de redução de dimensionalidade; (d) identifica quais as variáveis são mais importantes na classificação; (e) é eficaz para estimar os dados faltantes, mesmo quando boa parte dos dados não está disponível; (f) é robusto na presença de ruído e de variáveis sem importância; (g) pode ser altamente flexível para realizar vários tipos de análise de dados, incluindo regressão, classificação e aprendizagem não supervisionada.

Neste trabalho, o algoritmo *Random Forest* foi utilizado para geração de séries temporais de precipitação. Além disso, como em séries temporais de precipitação existem eventos raros (precipitação diária acima de 50 mm), será utilizado um novo algoritmo (*OversamplingR Interval*s), que é uma proposta de adaptação do algoritmo *SmoteR* (TORGO et al., 2013), que consiste em aplicar um método de amostragem *Over-sampling* com geração de casos sintéticos no conjunto de treinamento. A amostragem visa equilibrar a distribuição dos casos menos representados com as observações mais frequentes. Esse procedimento é feito somente no conjunto de treinamento; o conjunto de teste não pode ser alterado para não comprometer a qualidade do modelo gerado.

Nesse contexto, o objetivo deste estudo foi desenvolver uma metodologia baseada em algoritmos de Aprendizado e Máquina (*Random Forest* e *OversamplingR Interval*s) para gerar séries espaço-temporais de precipitação.

MATERIAL E MÉTODOS

Foi definida uma região de estudo de formato retangular, entre as latitudes de -18º e -22º e as longitudes de -52º e -39º, incluindo a metade norte do Estado de São Paulo e parte do sul de Goiás, a metade sul de Minas Gerais e o Estado do Espírito Santo. Essa região foi escolhida por conter um grande número de estações meteorológicas de diversas instituições, como ANA, INMET, IAC, CEMIG, Embrapa, EPAMIG, Universidades e outras. Além disso, foi escolhida por ser climaticamente bastante heterogênea, tanto no tempo (estacionalidade) como no espaço, com oito classificações diferentes pelo sistema de Köppen (Af, Am, Aw, Cfb, Cwa, Cwb, Cfa), devido à atuação de diversos condicionantes climáticos como áreas com relevo e altitudes bastante distintos, áreas montanhosas e áreas de planícies, influência litorânea, maior ou menor continentalidade vs. maritimidade nos extremos, com influências estacionais das massas de ar tropical atlântica, tropical continental, equatorial continental e polar atlântica.

A metodologia deste trabalho foi dividida em três principais etapas, como segue:

Etapa 1 - Aquisição e pré-processamento de dados: Todos os dados climáticos analisados consistem de séries temporais compreendidas no período de 01/01/1999 a 31/12/2013. Foram

extraídas séries temporais diárias incompletas de precipitação e de temperatura máxima e mínima das estações meteorológicas. As de precipitação estão presentes em todas as estações, já as de temperaturas máxima e mínima estão presentes em apenas 20,87% delas. Além disso, foram utilizadas as bases externas AgMERRA, AgCFSR, Modelo de radiação solar GL 1.2 CPTEC, TRMM e Nasa Power, cujos dados estão espacialmente dispostos em grids que cobrem a região de estudo, das quais foram extraídas séries temporais completas de precipitação (TRMM, Nasa Power, AgMERRA, AgCFSR), radiação solar (Nasa Power, GL 1.2), evapotranspiração potencial (Nasa Power) e temperaturas máxima, mínima e média (Nasa Power, AgMERRA, AgCFSR), para o mesmo período.

Com o objetivo de gerar os modelos preditivos de precipitação diária e precipitação acumulada de 10 dias e, a partir deles, séries espaço-temporais, a região de estudo foi dividida num grid com resolução de 0,5º (latitude e longitude), resultando em 280 quadrículas, sendo 28 na horizontal e 10 na vertical, as quais podiam conter ou não estações meteorológicas (Figura 1). Em seguida, foram definidas oito subáreas distintas, cada uma contendo 6 quadrículas, totalizando 48 quadrículas. Quadrículas que contêm uma ou mais estações meteorológicas são consideradas factíveis, isto é, a partir delas os modelos preditivos serão gerados.

Figura 1 – Número de estações com séries de precipitação por quadrícula num grid com resolução de 0,5º (latitude e longitude) subdividido em 8 áreas de interesse.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	X21	X22	X23	X24	X25	X26	X27	X28
Y1	1	3	1	4	4	1	4	7	2	1	3	1	3	1	2	4	2	0	2	5	4	4	3	1	4	3	1	0
Y2	0	1	1	0	3	3	4	2	0	7	8	3	5	3	9	2	3	9	1	2	1	2	4	7	5	3	0	0
Y3	0	1	2	3	2	8	3	5	4	3	2	6	5	4	0	2	5	2	1	3	3	7	4	8	6	6	0	0
Y4	0	4	2	2	1	4	3	2	0	3	3	8	3	6	8	5	5	2	6	9	4	3	5	6	4	0	0	0
Y5	2	1	2	3	5	3	3	2	5	7	6	5	3	1	2	6	14	15	7	2	7	2	7	9	6	1	0	0
Y6	1	0	7	14	8	12	6	12	11	15	10	1	6	6	4	10	14	12	4	4	4	4	5	12	7	0	0	0
Y7	2	5	10	6	7	9	12	17	13	13	13	6	4	2	2	4	6	6	4	7	5	8	9	10	1	0	0	0
Y8	1	6	10	13	14	13	8	15	17	16	15	11	3	8	6	13	9	9	6	7	2	4	5	1	0	0	0	0
Y9	3	6	12	16	9	11	8	13	18	16	17	13	5	5	11	6	11	4	8	6	6	4	4	0	0	0	0	0
Y10	5	6	6	10	9	13	9	7	10	19	11	16	3	9	6	9	17	11	11	14	11	2	2	0	0	0	0	0

Dada uma quadrícula factível, *quad*, foi adaptado o algoritmo heurístico de resolução do problema da mochila, modificado com pesos variáveis para o problema em questão (PISINGER, 1995). O algoritmo é responsável por escolher uma estação *e* em *quad* e, também, um conjunto de estações vizinhas *S*, cujos elementos são estações que não necessariamente precisam estar contidas em *quad*.

A motivação para a utilização desse algoritmo heurístico é devido ao alto nível de incompletude das séries espaço-temporais, sendo que há 28,84% de dados faltantes no total das séries de precipitação. Este percentual influi negativamente na qualidade dos resultados previstos, logo, em razão disso, nessa solução, para selecionar as estações vizinhas.

O problema da mochila (em inglês, *Knapsack problem*) é um problema de optimização combinatória. O objetivo é que se preencha uma mochila com o maior valor possível, não ultrapassando o peso máximo. No caso em questão, os itens correspondem às séries espaço-temporais das estações vizinhas; dado um item *i*, o peso de *i* é igual a porcentagem de dados perdidos nas intersecções de *i* com os outros itens (séries espaço-temporais) já escolhidos, o valor de *i* é igual ao inverso da distância entre as posições da estação à qual o item *i* pertence a estação *e*, e, por fim, o peso máximo é igual a porcentagem máxima de dados faltantes.

O número de estações vizinhas escolhidas varia entre n_{min} e n_{max}, e as estações escolhidas dessa forma satisfazem a exigência de que a porcentagem de dados faltantes na intersecção das séries espaço-temporais, não supere um valor *p*. Quanto ao modo de operação,

o algoritmo tenta, a cada iteração, adicionar as estações vizinhas mais próximas, sujeitas à condição anterior, e de tal forma que o raio médio da distância entre todas as estações que já foram adicionadas r_{avg} não supere um valor limite *threshold*, mas, eventualmente, este valor pode ser incrementado sucessivamente por um valor *incr* até que os critérios anteriores sejam satisfeitos.

Etapa 2 - Modelagem de dados e amostragem de eventos raros: Nesta etapa, seja *quad* uma quadrícula factível, para a qual pretende-se gerar um modelo preditivo. Aplica-se o algoritmo do problema da mochila em *quad*, em seguida, são extraídas as séries espaço-temporais advindas de *e* (I) e de *S* (II). Feito isso, são extraídas as séries espaço-temporais completas de cada uma das variáveis meteorológicas das bases externas nos pontos que recaem sobre *quad* (III) – Vale notar que essa extração é sempre possível porque as bases externas recobrem todo o grid. Todas essas séries são, então, reunidas compondo um único conjunto de dados, de forma que (I) é a variável alvo de predição e as séries espaço-temporais (II) e (III) são as demais variáveis dos modelos preditivos. Em seguida, desse conjunto de dados, são eliminadas todas as observações que contêm dados faltantes.

Considerando que prever precipitação diária é uma tarefa árdua, pois a maioria dos valores são iguais a zero, e isso tende a polarizar o processo de aprendizado, o que dificulta inferir eventos raros (precipitação diária acima de 50 mm); propôs-se uma alteração na etapa de modelagem dos dados, que consiste aplicar um método de amostragem *Over-sampling* com geração de casos sintéticos no conjunto de treinamento. O objetivo é amostrar os eventos raros para que o algoritmo possa aprendê-los. A alteração tem por objetivo equilibrar a distribuição dos casos menos representados (mas igualmente importantes) com as observações mais frequentes (TORGO et al., 2013). Esse procedimento é feito somente no conjunto de treinamento; o conjunto de teste não pode ser alterado.

A realização do *Over-sampling* no conjunto de treinamento, é feita por meio de uma adaptação no algoritmo *SmoteR* (TORGO et al., 2013). As modificações consistem em: não realizar Under-sampling, mas apenas Over-sampling; e o fato de, para cada intervalo de dados da variável alvo, gerar uma porcentagem bem definida de exemplos sintéticos; assim, o algoritmo modificado é bem flexível, pois permite ao analista de dados gerar diferentes porcentagens de valores raros nos intervalos.

Etapa 3 - Geração e validação dos modelos preditivos: Para cada uma das quadrículas factíveis nas quais quis-se gerar os modelos preditivos, obteve-se um conjunto de dados pelos procedimentos descritos nas etapas anteriores. Em seguida, esse conjunto foi dividido aleatoriamente em dois subconjuntos (com o mesmo número de variáveis), sendo que 2/3 dos dados deram origem ao conjunto de treinamento D_{train} e os outros 1/3 deram origem ao conjunto de validação D_{test}. Ao conjunto D_{train} foram aplicados os procedimentos de amostragem, e, por fim, o conjunto obtido foi utilizado na geração dos modelos preditivos baseados no algoritmo *Random Forest* (BREIMAN, 2001), para precipitação diária e precipitação acumulada de 10 dias.

O *Random Forest* é uma técnica de classificação e regressão desenvolvida por BREIMAN (2001), que consiste num conjunto de árvores de decisão combinadas para solucionar problemas de classificação. Cada árvore de decisão é construída utilizando uma amostra aleatória inicial dos dados e, a cada divisão desses dados, um subconjunto aleatório de *m* atributos é utilizado para a escolha dos atributos mais informativos. No final, *Random Forest* gera uma lista dos atributos mais importantes no desenvolvimento da floresta, que são determinados pela importância acumulada do atributo nas divisões dos nós de cada árvore da floresta.

Por fim, os modelos preditivos foram validados com o uso do conjunto D_{test} (método *Hold Out*).

As implementações dos algoritmos do problema da mochila, *Random Forest*, *SmoteR* adaptado (que a partir desse trabalho recebe o nome *OversamplingR Intervals*), bem como os procedimentos da etapa de pré-processamento foram realizados no ambiente RStudio (RStudio, 2017), que é um software livre de ambiente de desenvolvimento integrado para R, uma linguagem de programação para gráficos e cálculos estatísticos.

RESULTADOS E DISCUSSÃO

Em todas as execuções do algoritmo heurístico de resolução do problema da mochila foram utilizados os seguintes valores dos parâmetros: $n_{min} = 6$, $n_{max} = 12$, $p = 0{,}15$, *threshold* = 1 e *incr* = 1.

Foram gerados modelos preditivos de precipitação diária e precipitação acumulada de 10 dias em todas as quadrículas factíveis; mas, por uma questão de espaço, dado o alto número de modelos, estão expostas aqui as avaliações de desempenho dos modelos obtidos em apenas duas dessas quadrículas: uma na qual os modelos tiveram bons desempenhos, e a outra em que os desempenhos foram inferiores.

Nas execuções em que houve amostragem *Over-sampling*, as Tabelas 1, 2, 3 e 4 apresentam algumas das configurações testadas de parâmetros do algoritmo *OversamplingR Intervals*, os valores de correlação e os erros dos modelos preditivos. Em particular, para o intervalo [0, 20] não foi aplicou o procedimento de *Over-sampling*, pois os eventos nesse intervalo são frequentes.

Tabela 1 – Valores dos parâmetros do algoritmo *OversamplingR Interval*s para divisão em 3 intervalos distintos: [0, 20],]20, 40] e]40, +∞[; valores de correlação e erros dos modelos preditivos de precipitação diária.

Quad	%obs]20, 40]	%obs]40, +∞[	Correlação	RMSE	R^2
(2, 1)	100	100	0,58	8,571	0,3357
(2, 1)	200	500	0,54	9,564	0,2970
(12, 9)	100	100	0,85	6,298	0,7285
(12, 9)	200	500	0,84	6,851	0,7068

Tabela 2 – Valores dos parâmetros do algoritmo *OversamplingR Intervals* para a divisão em 5 intervalos distintos: [0, 20],]20, 40],]40, 60],]60, 80] e]80, +∞[; valores de correlação e erros dos modelos preditivos de precipitação diária.

Quad	%obs]20, 40]	%obs]40, 60]	%obs]60, 80]	%obs]80, +∞[	Correlação	RMSE	R^2
(2, 1)	100	100	200	500	0,57	8,881	0,3229
(2, 1)	200	500	500	500	0,56	9,516	0,3139
(12, 9)	100	100	200	500	0,84	6,655	0,7137
(12, 9)	200	500	500	500	0,84	6,884	0,7060

Tabela 3 – Valores dos parâmetros do algoritmo *OversamplingR Intervals* para divisão em 3 intervalos distintos: [0, 20],]20, 40] e]40, +∞[; valores de correlação e erros dos modelos preditivos de precipitação acumulada de 10 dias.

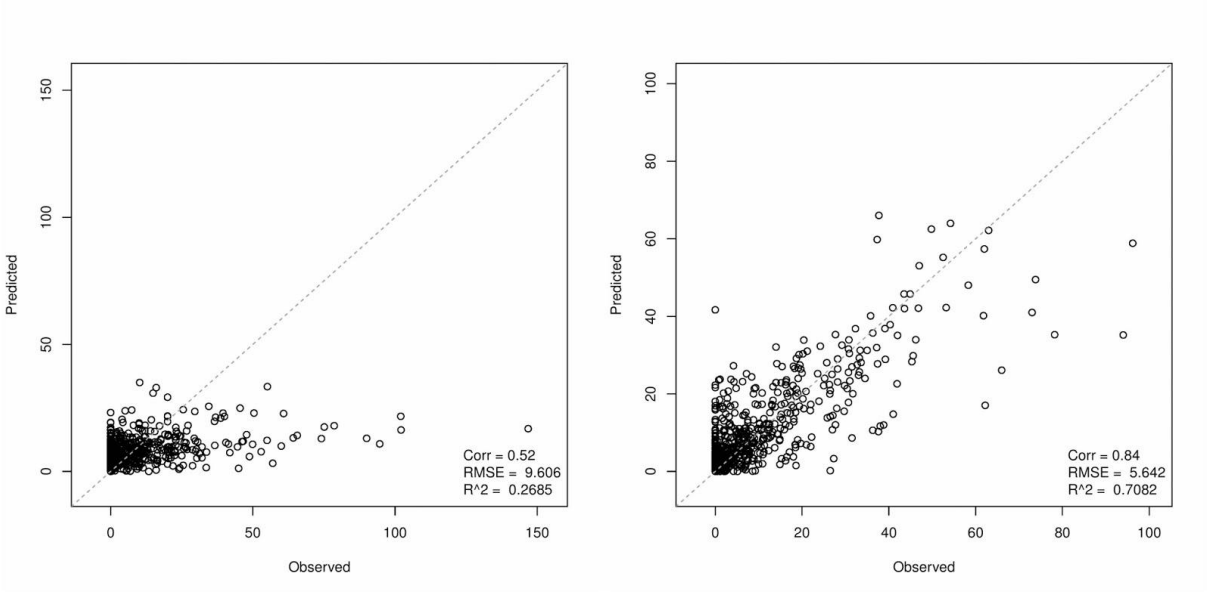
Quad	%obs]20, 40]	%obs]40, +∞[	Correlação	RMSE	R^2
(2, 1)	100	100	0,78	32,08	0,6071
(2, 1)	200	500	0,79	31,55	0,6212
(12, 9)	100	100	0,96	15,58	0,9123
(12, 9)	200	500	0,95	16,91	0,9005

Tabela 4 – Valores dos parâmetros do algoritmo *OversamplingR Intervals* para a divisão em 5 intervalos distintos: [0, 20],]20, 40],]40, 60],]60, 80] e]80, +∞[; valores de correlação e erros dos modelos preditivos de precipitação acumulada de 10 dias.

Quad	%obs]20, 40]	%obs]40, 60]	%obs]60, 80]	%obs]80, +∞[	Correlação	RMSE	R^2
(2, 1)	100	100	200	500	0,78	32,52	0,6056
(2, 1)	200	500	500	500	0,80	31,11	0,6330
(12, 9)	100	100	200	500	0,96	15,49	0,9133
(12, 9)	200	500	500	500	0,95	16,50	0,9024

A Figura 2 exibe a avaliação de desempenho dos modelos preditivos de precipitação diária sem amostragem *Over-sampling*, utilizando o algoritmo *Random Forest* e a Figura 3 refere-se ao gráfico de avaliação de desempenho dos modelos preditivos de precipitação diária com amostragem *Over-sampling* na melhor configuração, ou seja, com divisão em 3 intervalos, %obs]20, 40] = 100 e %obs]40, +∞[= 100 para ambas as quadrículas.

Figura 2 – Gráficos de avaliação de desempenho dos modelos preditivos baseados no algoritmo *Random Forest* para precipitação diária, sem amostragem *Over-sampling*, obtidos nas quadrículas: (2, 1) (esquerda), e (12, 9) (direita).



Em relação aos modelos preditivos de precipitação diária obtidos na quadrícula (2, 1), o uso da técnica de amostragem (Figura 3, esquerda) resultou num aumento de aproximadamente 10% no valor da correlação, quando comparado ao valor obtido sem o uso de amostragem (Figura 2, esquerda); já para os modelos preditivos de precipitação diária obtidos na quadrícula (12, 9), o uso da técnica de amostragem (Figura 3, direita) resultou num aumento de aproximadamente 1% no valor da correlação, quando comparado ao valor obtido sem o uso de amostragem (Figura 2, direita). Essas diferenças ocorrem porque a amostragem faz um esforço no sentido de equilibrar a distribuição dos casos menos representados, sendo esse um problema muito presente na quadrícula (2, 1), e menos presente na quadrícula (12, 9).

A Figura 4 ilustra o gráfico de avaliação de desempenho dos modelos preditivos de precipitação acumulada de 10 dias sem amostragem *Over-sampling*, enquanto a Figura 5 apresenta os resultados com amostragem *Over-sampling* na melhor configuração, ou seja, com divisão em 5 intervalos, %obs]20, 40] = 200, %obs]40, 60] = 500, %obs]60, 80] = 500 e %obs]80, +∞[= 500 para a quadrícula (2, 1); e %obs]20, 40] = 100, %obs]40, 60] = 100, %obs]60, 80] = 200 e %obs]80, +∞[= 500 para a quadrícula (12, 9).

Figura 3 – Gráficos de avaliação de desempenho dos modelos preditivos baseados no algoritmo *Random Forest* para precipitação diária, com amostragem *Over-sampling* na melhor configuração, obtidos nas quadrículas: (2, 1) (esquerda), e (12, 9) (direita).

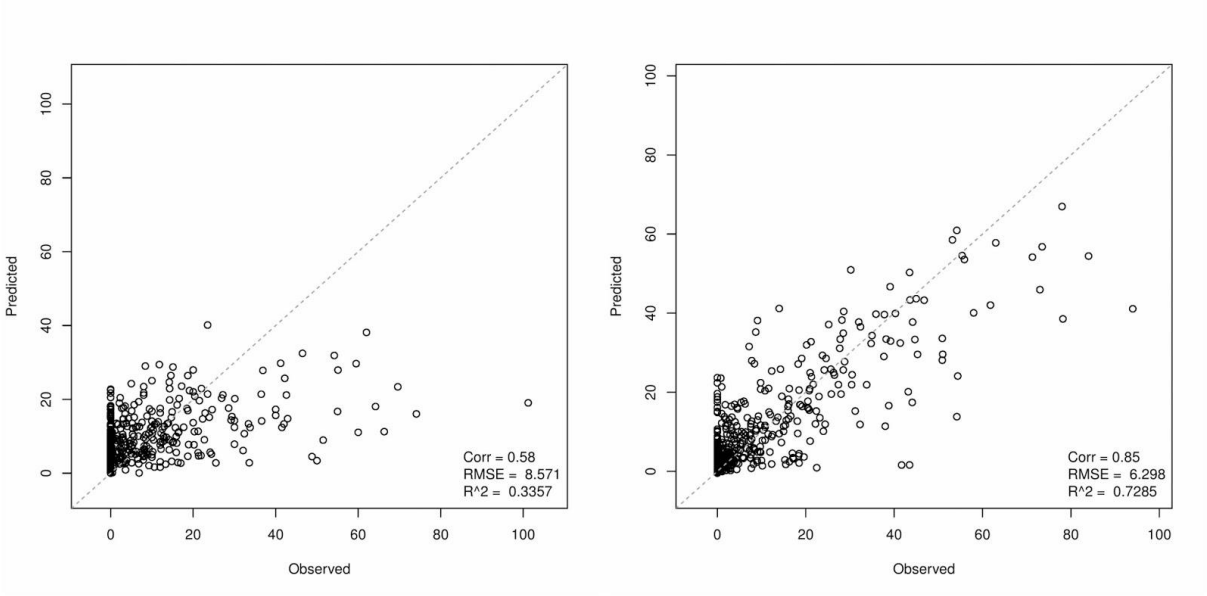


Figura 4 – Gráficos de avaliação de desempenho dos modelos preditivos baseados no algoritmo *Random Forest* para precipitação acumulada de 10 dias, sem amostragem *Over-sampling* na melhor configuração, obtidos nas quadrículas: (2, 1) (esquerda), e (12, 9) (direita).

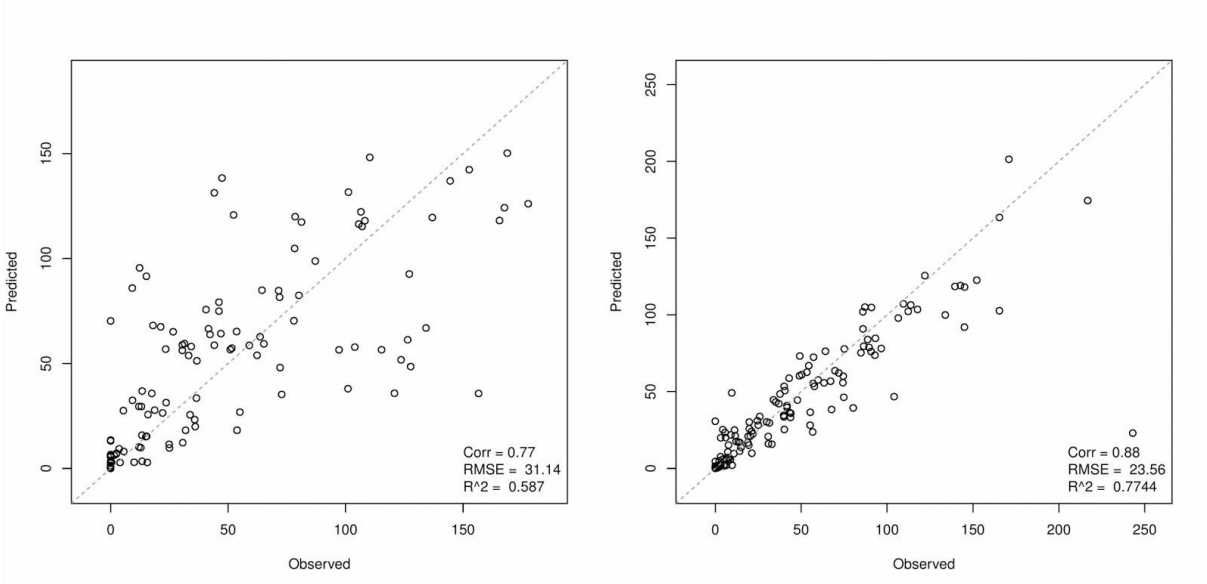
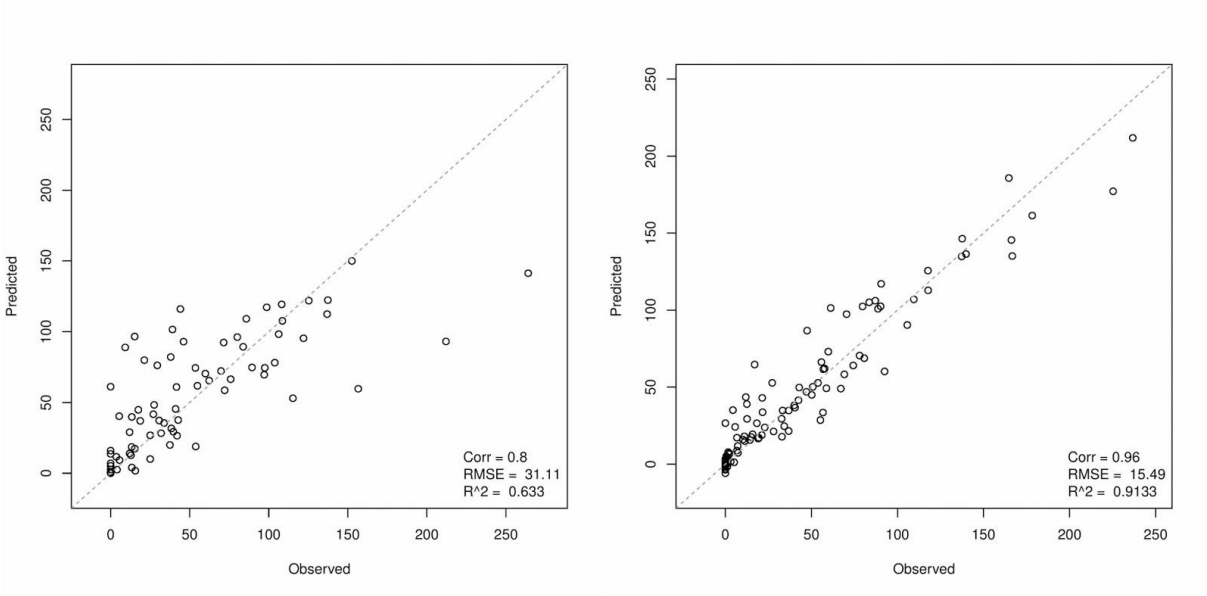


Figura 5 – Gráficos de avaliação de desempenho dos modelos preditivos baseados no algoritmo *Random Forest* para precipitação acumulada de 10 dias, com amostragem *Over-sampling*, obtidos nas quadrículas: (2, 1) (esquerda), e (12, 9) (direita).



As dificuldades para quantificar, estimar ou espacializar o valor da chuva são conhecidas e muito tratada na literatura com diferentes abordagens (NOBRE; MOLION, 1988, PLOUFFE et al., 2015). Isso decorre da alta variabilidade espacial e temporal da precipitação, às quais somam-se limitações para a manutenção de uma rede densa de postos de medida, além de falhas e erros no próprio processo de medição.

Esse conjunto de limitações vem motivando o desenvolvimento de outras técnicas, com destaque para as metodologias baseadas em Sensoriamento Remoto Orbital (SRO), como os dados provenientes do TRMM (Tropical Rainfall Measuring Mission) (SOARES et al., 2016). A grande vantagem do SRO é a abrangência ou cobertura espacial mas, uma limitação é a baixa resolução de 25km, muitas vezes insuficiente para representar ocorrências pontuais, típicas da chuva. Nesse sentido, a metodologia proposta se apresenta como uma alternativa muito eficiente para identificar e modelar correlações entre variáveis correlacionadas, combinando de forma complementar dados de diferentes fontes e naturezas.

CONCLUSÕES

A metodologia proposta, baseada no algoritmo de aprendizado de máquina *Random Forest* e no algoritmo *OversamplingR Interval*s permitiu a geração de séries espaço-temporais de precipitação diária e acumulada de 10 dias com bom ajuste em relação aos valores observados, mostrando grande potencial de uso para imputação de dados ausentes ou na geração de séries de precipitação em quadrículas ou regiões com falhas de dados ou sem a presença de estações de medição. Essa iniciativa é fundamental para apoiar o planejamento de atividades agrícolas que são dependentes de variáveis meteorológicas, notadamente em áreas onde não é possível otimizar modelos locais.

AGRADECIMENTOS

Ao programa CNPq/PIBIC pela concessão da bolsa de Iniciação Científica, processo Nº 145155/2016-1.

REFERÊNCIAS

BREIMAN, L. Random Forests. In Machine Learning, v.45, p.5–32, 2001.

CAMPOZANO, L.; SANCHEZ, E.; AVILES, A.; SAMANIEGO, E. Evaluation of infilling methods for time series of daily precipitation and temperature: The case of the Ecuadorian Andes. MASKANA, v. 5, n. 1, p. 99-115, 2014.

DUMEDAH, G.; COULIBALY, P. Evaluation of statistical methods for infilling missing values in high-resolution soil moisture data. Journal of Hydrology, v. 400, p. 95-102, 2001. doi:10.1016/j.jhydrol.2011.01.028.

FENG, S.; QIAN, W. Quality control of daily meteorological data in China, 1951-2000: a new dataset. International Journal of Climatology, Chichester, v. 24, p. 853-870, 2004.

KANE, M. J.; PRICE, N.; SCOTCH, M.; RABINOWITZ, P. Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. BMC Bioinformatics, v.15, n.1, 2014.

MWALE, F.D.; ADELOYE, A.J.; RUSTUM, R. Infilling of missing rainfall and streamflow data in the Shire River basin, Malawi – A self organizing map approach. Physics and Chemistry of the Earth, 50–52, p. 34–43, 2012.

NOBRE, C. A.; MOLION, L. C. B. The climatology of drought prediction. Norwell: Kluwer, 1988.

PISINGER, D. Algorithms for Knapsack Problems. PhD thesis, Department of Computer Science, University of Copenhagen, February 1995.

PLOUFFE, C. C. F.; ROBERTSON, C.; CHANDRAPALA, L. Comparing interpolation techniques for monthly rainfall mapping using multiple evaluation criteria and auxiliary data sources: a case study of Sri Lanka. Environ. Model. Software, v. 67, p 57-71, May 2015.

RStudio – Open source and enterprise-ready professional software for R. Disponível em https://www.rstudio.com. Consultado em 3 de Maio de 2017.

STEINACKER, R.; MAYER, D.; STEINER, A. Data Quality Control Based on Self-Consistency. Monthly Weather Review, v. 139, p. 3974–3991, 2011. doi: http://dx.doi.org/10.1175/MWR-D-10-05024.1

TORGO, L.; RIBEIRO, R. P.; PFAHRINGER, B.; BRANCO, P. Smote for regression. In: PORTUGUESE CONFERENCE ON ARTIFICIAL INTELLIGENCE, 16., 2013, Angra do Heroismo. **Progress in artificial intelligence**: proceedings. New York: Springer, 2013. p. 378-389. (Lecture notes in artificial intelligence, 8154). EPIA 2013.

SOARES, A.S.D.; PAZ, A.R.; PICCILLI, D.G.A. Avaliação das estimativas de chuva do satélite TRMM no Estado da Paraíba. Revista Brasileira de Recursos Hídricos, v. 21, n. 2, p. 288 – 299, 2016.

VICENTE-SERRANO, S.M., SANTIAGO, B., LÓPEZ-MORENO, J.I., GARCÍA-VERA, M.A.; STEPANÉK, P. A complete daily precipitation database for northeast Spain: reconstruction, quality control, and homogeneity. International Journal of Climatology, v. 8, p. 1146-1163, 2010.

VILLANUEVA, W. J. P. **Comitê de Máquinas de Predição de Séries Temporais**. 2006. p.150. (Dissertação) - Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas.

XAVIER, A. C.; KING, C. W.; SCANLON, B. R. Daily gridded meteorological variables in Brazil (1980-2013). **International Journal of Climatology**, v. 36, n. 6, p 2644-2659, 2016.



Uma abordagem paralela para a execução do modelo de simulação de cultura CSM-Cropsim: Wheat

Angela Mazzonetto[1], *Carlos Amaral Hölbig*[1], *Marcelo Trindade Rebonatto*[1], *Willingthon Pavan*[1]

[1]Programa de Pós Graduação em Computação Aplicada
Universidade de Passo Fundo, Brasil
144981@upf.br, holbig@upf.br, rebonatto@upf.br, pavan@upf.br

RESUMO

Este trabalho apresenta uma abordagem computacional paralela que visa a ampliação da área de cobertura e/ou da série temporal a ser simulada pelo modelo de crescimento da cultura do trigo chamado de CSM-Cropsim: Wheat. A abordagem de paralelização da execução do modelo escolhida foi a mestre-escravo, juntamente com a utilização da biblioteca de comunicação MPI, que mostrou ser adequada, visto que cada uma das execuções (rodadas do modelo) são independentes dos resultados das outras no modelo avaliado. Os resultados obtidos com os testes realizados demonstraram-se satisfatórios, demonstrando a validade da aplicação desta abordagem na execução em larga escala do CSM-Cropsim: Wheat.

PALAVRAS-CHAVE: CSM-Cropsim: Wheat, execução paralela, modelo de cultura.

ABSTRACT

This paper shows a parallel computational approach that aims to expand the coverage area and/or the time series to be simulated by the wheat growth model called CSM-Cropsim: Wheat. The parallelization approach of the execution of the chosen model was the master-slave, along with the use of the MPI communication library, which proved to be adequate, since each of the executions (model rounds) are independent of the results of the others in the evaluated model . The results obtained with the tests performed proved to be satisfactory, demonstrating the validity of the application of this approach in the large scale execution of CSM-Cropsim: Wheat.

KEYWORDS: CSM-Cropsim: Wheat, parallel execution, crop model.

INTRODUÇÃO

No atual cenário científico e tecnológico muitos fatores contribuem para o aumento do volume de dados que são utilizados para a solução de aplicações reais nas mais variadas áreas. Com a

crescente evolução tecnológica, a presença de novos coletores de dados meteorológicos torna-se comum. Com esses coletores, a massa de dados aumenta consideravelmente, o que ajuda na qualidade da cobertura de dados observados, utilizados como dados de entrada por centros e grupos de pesquisa e demais interessados em utilizá-las em seus ramos de atividade Chan et al. (2007).

Uma das áreas de maior aplicação destes dados é a área agrícola, onde modelos de simulação de culturas e doenças os utilizam. Os resultados gerados por estes modelos são posteriormente analisados e tem por objetivo auxiliar no processo de tomada de decisão. Neste caso em específico, entretanto, esta crescente disponibilidade de dados pode gerar dificuldades de execução quando pretende-se utilizar a sua totalidade, fazendo com que o número de execuções exceda o poder de processamento das máquinas dos pesquisadores devido à sua área de abrangência ou ao período a ser simulado. Além dos dados observados, dados gerados por outros modelos como, por exemplo, de previsão de tempo e clima, também são utilizados.

A necessidade de abranger uma maior região geográfica ou maior período de simulação torna a execução dos modelos repetitiva, pois o mesmo modelo tende a ser executado muitas vezes para abranger estas diferentes regiões, áreas ou períodos. Porém, com este aumento da granularidade, a quantidade de execuções aumenta exponencialmente nas mesmas proporções, e a quantidade de dados necessários para serem processados muitas vezes pode resultar em um despendimento maior de tempo, tornando inviável a utilização dos resultados gerados pelos modelos. Por estes motivos torna-se indispensável a utilização de ferramentas ou técnicas computacionais de alto desempenho que visem a otimização e a melhora do desempenho computacional da execução destes modelos. A alternativa escolhida por esta pesquisa baseia-se no uso do processamento paralelo, possibilitando a execução dos modelos de simulação por esta abordagem.

O processamento paralelo tem por objetivo a execução de um programa ou tarefas simultaneamente em vários processadores. Contudo, é importante saber quais técnicas serão utilizadas para que a paralelização seja efetivamente eficiente na solução do problema proposto, sendo, para isto, necessário conhecer detalhes sobre o programa que terá sua execução em paralelo. Entre as maneiras de paralelizar um programa, estão o uso de ferramentas como o MPI, o OpenMP, o framework Hadoop, a linguagem R e seus pacotes, ou bibliotecas desenvolvidas para este propósito. Juntamente com a abordagem paralela, é importante saber o ambiente computacional no qual o programa será executado, podendo ser em *grid* de computadores, em computadores com processadores *multicore*, em computadores com placa gráfica ou em clusters de computadores. Com o objetivo de apresentar alternativas viáveis de paralelização da execução de modelos de simulação ligados à área agrícola, este trabalho selecionou um modelo como estudo de caso para validar estas alternativas mediante alguns testes. Este estudo de caso é composto por um modelo de simulação de cultura do trigo, chamado de CSM-Cropsim: Wheat (HUNT; PARARAJASINGHAM, 1995), utilizado pela Embrapa Trigo e pela UPF. Portanto, o problema de pesquisa deste trabalho foi o de definir uma estratégia de paralelização da execução

deste modelo.

MATERIAL E MÉTODOS

Entre as diversas espécies de diferentes culturas, o trigo destaca-se como um importante alimento para a humanidade. É cultivado em diversos lugares do mundo e por isso possui grande relevância econômica e social. Porém, sua produção pode ser afetada por diversos fatores como as condições do solo, do clima, do manejo, entre outras. Por isso são necessários meios de evitar que estes fatores hajam de forma negativa na produtividade da espécie. Os modelos de simulação de culturas combinam dados meteorológicos, dados do solo, fisiologia vegetal, etc e geram informações sobre possíveis impactos com a variação destes fatores no crescimento da cultura (Lazzaretti (2013)). Entre estes modelos está o modelo CSM-Cropsim: Wheat, estudado neste trabalho e apresentado a seguir.

Descrição do modelo e suas características

O modelo de simulação de crescimento do trigo CSM-Cropsim: Wheat, foi desenvolvido por Leslie A. Hunt Hunt e Pararajasingham (1995) na linguagem de programação Fortran e, atualmente, faz parte do sistema Decision Support System for Agrotechnology Transfer (DSSAT http://dssat.net/), composto por diversos outros modelos de simulação de culturas. O CSM-Cropsim: Wheat simula o crescimento e desenvolvimento da cultura do trigo. Devido aos fatores já destacados neste texto, o CSM-Cropsim: Wheat necessita que sua execução deve ser organizada a fim de se obter uma melhora em seu rendimento computacional em situações onde há uma grande área a ser coberta pelo modelo e/ou onde a série temporal a ser simulada é muito grande. Este modelo é utilizado no grupo de pesquisa Mosaico da Universidade de Passo Fundo e na Embrapa Trigo.

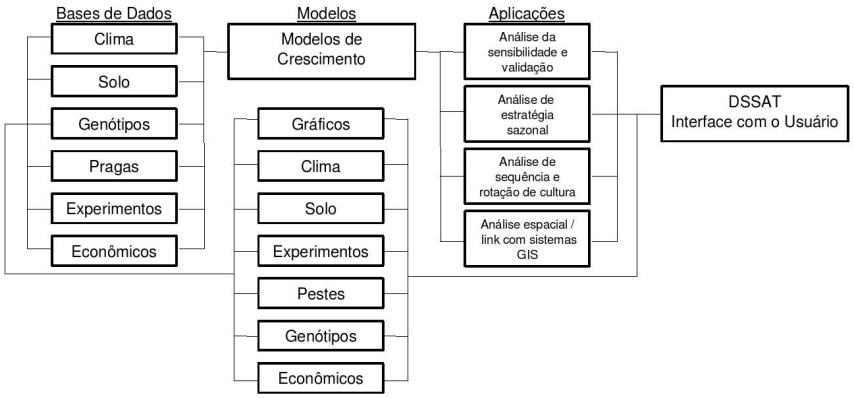
No Brasil, o CSM-Cropsim: Wheat tem sido testado, calibrado e validado por diversos pesquisadores, sendo utilizado para simular o processo do desenvolvimento de cultivares de trigo, além de contribuir para a tomada de decisão de pesquisadores e produtores da área agrícola Pavan (2007).

Segundo Pavan (2007), o CSM-Cropsim: Wheat é composto por um módulo planta de trigo que se conecta com o módulo de clima e solo, pertencentes à suite do DSSAT, os quais calculam a energia e a água disponíveis para o crescimento da planta de trigo, ao passo que o módulo planta de trigo simula os eventos fenológicos, expansão foliar, acúmulo de carboidratos e a partição entre a parte aérea e raízes.

A funcionalidade deste modelo integra uma série de dados de entrada que compreendem dados meteorológicos, de solo, de fisiologia vegetal e de genótipo da planta, conforme representado na Figura 1. Eles estão parametrizados em arquivos de formato texto, sendo os resultados do modelo, com as informações de pós processamento, também armazenados em arquivos texto.

Figura 1: Estrutura da suíte do Sistema de Apoio à Decisão para Transferência de Agrotecnologia (DSSAT)

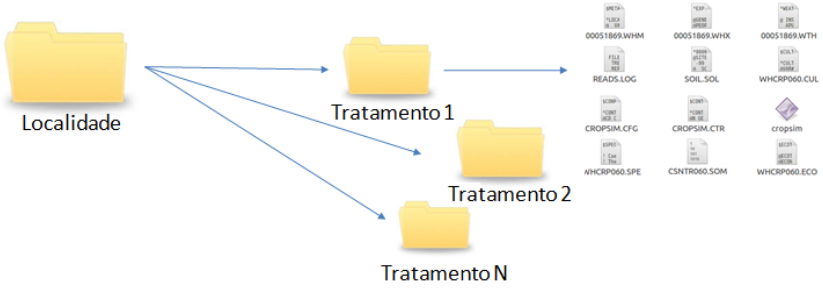
Fonte (PAVAN, 2007)



Para um melhor gerenciamento destes arquivos, a aplicação realizada no trabalho de Lazzaretti (2013) integrou o modelo CropSim com um banco de dados. Nele, os dados são organizados de forma que cada diretório contém arquivos textos com os dados do tratamento, num determinado período e, também, o executável do modelo. Após o modelo ser executado, os arquivos de saída são armazenados no mesmo diretório correspondente. Ainda conforme Lazzaretti (2013), os arquivos gerados pela sua aplicação são os seguintes:

- `.WTH`: arquivo com os dados meteorológicos. Os principais dados necessários nessa etapa são: temperaturas do ar (máxima e mínima), precipitação, radiação solar e concentração de CO2 na atmosfera;
- `.SOIL`: arquivo com os dados de solo por camadas;
- `.WHX`: arquivo que gerencia a execução, no qual são descritos os tratamentos, a identificação dos arquivos de solo e clima, detalhes da semeadura, condições iniciais e manejo;
- `.CUL`: arquivo que contém os coeficientes da cultivar como, por exemplo: sensibilidade ao fotoperíodo, características fenológicas, taxa de fotossíntese à luz saturada, área foliar específica, peso máximo de sementes;
- `.ECO`: arquivo que contém os coeficientes de resposta da planta em relação ao ambiente (ecotipo). Possui alguns atributos genéticos que permitem diferenciar cultivares de hábito de crescimento determinado e indeterminado;
- `.CFG`: arquivo com as configurações para a execução do CSM-Cropsim: Wheat, tais como diretórios, módulos, arquivos e programas;
- `.SOM`: abreviatura de *Soil Organic Matter*. Arquivo que contém os parâmetros necessários para a simulação da matéria orgânica no solo.
- `.SPE`: arquivo que caracteriza a espécie. Possui coeficientes que caracterizam a composição básica dos tecidos e alguns processos da planta como: fotossíntese, respiração, assimilação de nitrogênio, partição de fotoassimilados, senescência, fenologia e crescimento.
- `.OUT, .LOG e .WHM.` arquivos de saída gerados após a simulação.

Figura 2: Estrutura dos diretórios montada para a execução do modelo CSM-Cropsim: Wheat.



Conforme a Figura 2, uma localidade está relacionada a um experimento. Dentro do diretório estão vários tratamentos, e cada tratamento contém conjuntos de dados de solo, épocas de semeadura e um conjunto de dados meteorológicos. Estes dados, obtidos do banco de dados AgroDB Lazzaretti (2013), são dados que abrangem dez épocas de semeadura e trinta anos de dados meteorológicos observados do estado do Paraná (Brasil) entre os anos de 1980 e 2009.

Por exemplo, nos dados utilizados neste trabalho, são processados 30 anos, com cinco épocas de semeadura e 16 *ensembles* de dados meteorológicos, totalizando 2400 tratamentos que equivalem à 2400 execuções do modelo.

Execução do CropSim

O modelo de simulação CropSim é executado via linha de comando no Linux `./cropsim 00051962.WHX 1 1`, onde `./cropsim` é o comando para a execução do arquivo binário do modelo de simulação, `00051962.WHX` é o arquivo de configuração responsável por comandar o processo de execução do modelo, ou seja, utilizar os dados de todos os outros arquivos de entrada, e os parâmetros ”1 1”são valores que controlam qual tratamento do modelo deve ser executado.

Com esta linha de comando, que é correspondente à apenas uma execução de um tratamento, o tempo de execução sequencial é de aproximadamente $0,070$ segundos. Porém, para cada localidade, que contém vários tratamentos, são necessárias inúmeras execuções deste modelo, o que acaba acarretando em mais tempo de simulação para que este possa abranger todas os experimentos da mesma localidade. Geralmente, um mesmo experimento contém inúmeros tratamentos. Portanto, para executá-lo, são necessários vários arquivos de entrada e, para sua execução, várias chamadas desta linha de comando.

Estratégia de Paralelização

Com o objetivo de executar o modelo de simulação CSM-Cropsim: Wheat de forma eficiente e rápida, abrangendo mais localidades e utilizando uma série histórica maior, buscou-se definir uma maneira viável de paralelizar sua execução visando alcançar esta finalidade.

A paralelização deste modelo baseou-se no algoritmo paralelo mestre-escravo. Esta abordagem foi a que melhor se adaptou à estrutura de funcionamento do modelo, visto que cada uma

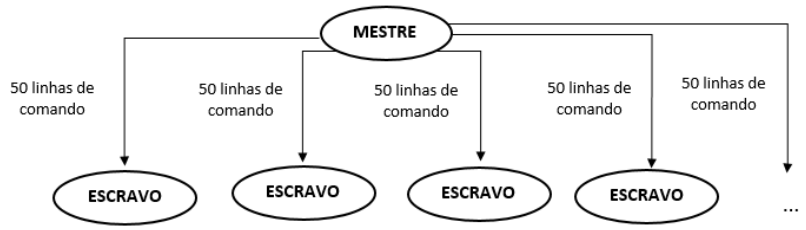
de suas execuções (rodada do modelo) é independente das demais. Este método foi escolhido pois permite realizar a programação paralela da execução do modelo sem alterar seu código fonte, ou seja, conforme os processos vão terminando suas tarefas de processamento dos dados, o mestre pode enviar outras até que não haja mais nenhum dado para ser processado. Optou-se por não paralelizar o modelo porque o cálculo para uma execução de um tratamento é de apenas 0,055 segundos. O maior tempo gasto com a execução do modelo deve-se ao processo de leitura/escrita de arquivos. Anteriormente a este trabalho foram realizadas tentativas pelo grupo de pesquisa da Universidade de Passo Fundo para realizar a paralelização implícita do modelo, porém, após a análise do desempenho obtido e devido à estrutura de funcionamento do mesmo, chegou-se à conclusão de que esta paralelização não traria benefícios significativos.

Como exemplo de uma paralelização de modelo baseada no algoritmo mestre-escravo, onde houve melhora no desempenho, pode ser citado o trabalho de Hadka e Reed (2015) onde foi desenvolvida a implementação paralela do algoritmo Borg Multiobjective Evolutionary Algorithm (MOEA) voltado para resolver problemas em fontes de águas. Por meio de testes com até 16.384 processadores obteve-se resultados satisfatórios. Outro trabalho com este tipo de abordagem é o realizado por (TIEJIAN WANG GUANGQIAN, 2011), onde é apresentado o desenvolvimento de um algoritmo paralelo dinâmico para a realização de simulações de modelos hidrológicos, que utilizam a metodologia mestre-escravo e a comunicação com o padrão MPI. Os resultados desta aplicação revelam que o algoritmo é eficiente no envio das tarefas de simulação entre os processos e que há um aumento da velocidade e eficiência em função da largura da bacia.

A estratégia de paralelização da execução do modelo CSM-Cropsim: Wheat funciona de forma que um conjunto de comandos de execução de simulações sejam enviadas para cada processador. Cada linha de comando contém a execução de um tratamento, e um tratamento corresponde a um conjunto de dados referente a uma localidade. Por exemplo, se são necessárias 9600 execuções do modelo (isso corresponde a 4 localidades com 2400 tratamentos cada uma) serão enviadas 50 linhas de comandos de execução para cada processo até completar o total das 9600 execuções. Este envio é feito pelo processo mestre. Logo após, o processo escravo começa as execuções. Quando este chega ao final das 50 linhas de comando recebidas, envia um aviso ao mestre informando que finalizou suas execuções. Se ainda houver linhas de comando a serem executadas, o mestre as envia aos seus escravos até que não as tenha mais (Figura 3). Quando todo o trabalho tiver chegado ao final, o mestre envia um sinal de fim de trabalho aos escravos.

Esta troca de mensagens é feita pelo padrão MPI, que é uma interface de programação para comunicação de dados entre processos paralelos. Esta ferramenta oferece infra-estrutura para modalidades de computação paralela quando existe a necessidade de passar informações entre os vários processadores ou nodos de um *cluster* com memória distribuída (Sena e Costa (2008)). Também, conforme descrito em Sena e Costa (2008), a interface do MPI foi estendida com funções para gerenciamento dinâmico de processos, entrada/saída paralela e acesso

Figura 3: Modelo paralelo de execução do modelo CSM-Cropsim: Wheat.



a memória remota, além de ser adaptada a arquiteturas com memória híbrida (compartilhada entre os núcleos e distribuída entre os nós).

Outra questão que afeta o desempenho da execução paralela é o aumento do número de processadores que, em um dado momento, pode tender a um aumento de tempo de execução em paralelo. Um exemplo desta situação é o trabalho descrito por Zhang et al. (2013), onde os autores comentam sobre a crescente disponibilidade de dados e o uso cada vez maior de modelos de bacias hidrográficas distribuídos por grandes áreas com alta resolução espacial e temporal. O autor descreve que nesse trabalho, foi desenvolvido um software paralelo para melhorar a eficiência de calibração dos modelos. Resultados de testes realizados em um *cluster* de computadores com sistema operacional Linux mostraram que o software desenvolvido pode alcançar um melhor desempenho dependendo da complexidade do modelo e que, aumentando a quantidade de processadores além de certo limite, não necessariamente melhorou a eficiência, pois a concorrência de intensificação do recurso resultou em um gargalo de I/O. Portanto, existe a possibilidade de melhora no desempenho até certa quantidade de processos. Isto vai depender das características de cada modelo e da tecnologia usada na implementação.

RESULTADOS E DISCUSSÃO

A análise do desempenho da implementação paralela do modelo CSM-Cropsim: Wheat foi realizada por meio da execução com diferentes números de processadores de um sistema constituído por várias estações de trabalho, buscando obter mais desempenho computacional na divisão de trabalho entre os núcleos de processamento da máquina *Network of Workstations* (NOW) (UFRGS, 2006).

Utilizou-se um aglomerado homogêneo composto por 16 máquinas formado pelas máquinas do Laboratório Central de Informática (LCI) da Universidade de Passo Fundo , cada qual com a seguinte configuração: Processador Intel Core i7 2600, 3.4 GHz, memória RAM 8Gb, 4 núcleos físicos e sistema operacional Linux (Ubuntu 12.10) de 64-bits.

Os dados utilizados para a realização dos testes foram obtidos do banco de dados desenvolvido pelo trabalho descrito em Lazzaretti (2013) e correspondem a quatro localidades do sul do Brasil. O tempo para a execução sequencial do modelo para estes dados foi de aproximadamente 43 minutos.

Após a realização dos testes foi construída a Tabela 1 demonstrando que há melhora no desempenho após a paralelização do modelo.

Tabela 1: CSM-Cropsim: Wheat: 50 linhas de execução e dados distribuídos.

Processos	Tempo médio (seg)	*Speedup*	Eficiência (%)	Custo
2	4326,14	0,60	29,96%	8652,28
4	2938,07	0,88	22,06%	11752,28
8	1632,72	1,59	19,84%	13061,76
16	500,56	5,18	32,36%	8008,96
32	251,25	10,32	32,24%	8040,00
64	187,13	13,85	21,64%	11976,32

Levando em consideração o melhor resultado para o número de linhas de execução, analisando a Tabela 1, obteve-se melhor desempenho quando foram utilizados 64 processos. Obteve-se um *speedup*(razão entre o tempo para executar o algoritmo sequencialmente e o tempo executando o algoritmo paralelo, Ferlin (2012)) de $13,85$ indicando em quantas vezes foi acelerada a versão sequencial que foi de, aproximadamente, 2592 segundos.

Com base nas observações anteriores, pode se verificar a influência de certos fatores referentes ao tamanho das mensagens e à quantidade de arquivos lidos para a execução do modelo e escritos após o processamento dos dados. Observando os valores obtidos de *speedup* e eficiência o melhor desempenho foi obtido com 64 processos. O desempenho computacional melhorou aproximadamente 90% comprovando que a paralelização do modelo de simulação CSM-Cropsim: Wheat mostrou ser uma opção viável para agilizar o processamento dos dados e aumentar a área de abrangência geográfica simulada por este modelo. Trabalhos futuros poderão investigar a possibilidade de melhora no desempenho que ainda pode ser feito com outras abordagens computacionais e também com o aumento do número de processadores. Porém sabe-se que em alguns casos uma quantidade muito maior de processadores poderia acarretar e aumento do tempo de processamento por motivos como tempo de troca de mensagens ou leitura e escrita.

CONCLUSÕES

O modelo de simulação utilizado pela Universidade de Passo Fundo e Embrapa Trigo, simula o crescimento do trigo, CSM-Cropsim: Wheat, o qual auxilia profissionais responsáveis na tomada de decisão com o intuito de melhorar a produtividade da cultura em questão. Observou-se que havia a necessidade da utilização de maiores quantidade de dados e, com isso, aumentar a área de abrangência geográfica e/ou também de dados históricos processados pelo modelo. Devido a este fato iniciou-se o estudo do funcionamento e da execução sequencial do modelo. Posteriormente, avaliou-se possíveis técnicas computacionais que proporcionassem o processamento de maiores quantidades de dados de forma efetiva e em tempo computacional hábil que favorecesse o uso destes dados dentro de um prazo útil. Entre estas técnicas estão algoritmos,

linguagens e bibliotecas para a computação paralela. Conforme as características dos modelos e das aplicações a que eles simulam, optou-se pela utilização da biblioteca MPI, com a metodologia do algoritmo mestre/escravo. Para validar esta escolha foram realizados testes utilizando a arquitetura computacionais composta de um sistema constituído de 16 máquinas do Laboratório Central de Informática da UPF.

Aplicando esta abordagem obteve-se resultados satisfatórios na paralelização da execução do modelo, pois, com 64 processos, obteve-se um tempo de $187,13$ segundos com *speedup* de $13,85$, enquanto que o tempo sequencial para processar a mesma quantidade de dados foi de 2592 segundos. Com isso, conclui-se que o desempenho computacional da execução paralela da execução de modelos de simulação depende de alguns fatores principais, entre eles está se o modelo realiza mais ou menos cálculos e/ou mais leitura e escrita de arquivos. Outro fator importante é a arquitetura computacional a ser utilizada para a execução dos modelos. A partir dos resultados deste trabalho, levando em consideração os recursos computacionais disponíveis na UPF, obteve-se melhora no desempenho da execução do modelo.

Por fim, com este trabalho, buscou-se otimizar o desempenho computacional da execução dos modelos de simulação objetivando uma maior abrangência na utilização da quantidade de dados processados por estes modelos e, assim, auxiliar com a prática do desenvolvimento de programação paralela para os profissionais da computação e meteorologia envolvidos nos modelos, para os produtores rurais na tomada de decisão em relação às culturas e para pesquisadores interessados nos benefícios da utilização dos modelos de simulação de culturas e de previsão de tempo em aplicações na área agrícola.

AGRADECIMENTOS

Agradeço à colaboração da Universidade de Passo que contribuiu com a estrutura computacional para a realização dos testes deste trabalho e também aos professores envolvidos na elaboração, desenvolvimento e demais contribuições na elaboração do mesmo.

REFERÊNCIAS

CHAN, C. S. et al. Refinamento estatístico das previsões horárias de temperatura a 2 m do modelo eta em estações do nordeste do brasil. *Revista Brasileira de Meteorologia*, Scielo, v. 22, p. 287 – 296, 12 2007. ISSN 0102-7786. Disponível em: <http://www.scielo.br/pdf/rbmet/v22n3/01.pdf>. Acesso em: 20 Jul. 2015.

FERLIN, E. P. *As Métricas de Desempenho dos Computadores (Parte III): Speedup e Eficiência*. 2012. Acesso em: 13 Jul. 2015.

HADKA, D.; REED, P. Large-scale parallelization of the Borg multiobjective evolutionary algorithm to enhance the management of complex environmental systems. *Environmental*

Modelling & Software, Elsevier Ltd, v. 69, p. 353–369, 2015. ISSN 13648152. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S1364815214003041>. Acesso em: 28 Jan. 2015.

HUNT, L. A.; PARARAJASINGHAM, S. Cropsim-wheat: A model describing the growth and development of wheat. *Canadianjournal Plant Science*, Department of Crop Science, University of Guelph, Guelph, Ontario, Canada, 1995. Disponível em: <http://pubs.aic.ca/doi/pdf/10.4141/cjps95-107>. Acesso em: 11 Jul. 2014.

LAZZARETTI, A. T. *Interação de banco de dados e modelos de simulação de culturas para estimar o impacto de mudanças do clima no rendimento de grãos e na severidade da giberela no trigo.* Tese (Doutorado) — UPF - Faculdade de Agronomia e Medicina Veterinária, Programa de Pós Graduação em agronomia., 2013. Disponível em: <www.ppgagro.upf.br/images/stories/alexandre-tagliari-lazzareti.pdf>. Acesso em: 20 Ago. 2014.

PAVAN, W. *Técnicas de Engenharia de Software Aplicadas à Modelagem e Simulação de Doenças de Plantas.* Tese (Doutorado) — UPF - Faculdade de Agronomia e Medicina Veterinária, Programa de Pós Graduação em agronomia., 2007. Disponível em: <www.ppgagro.upf.br/download/Willingthon_Pavan.pdf>. Acesso em: 15 Ago. 2014.

SENA; COSTA. *Tutorial OpenMP C/C++*. [S.l.], 2008. Disponível em: <http://www.lncc.br>. Acesso em: 13 Out. 2014.

TIEJIAN WANG GUANGQIAN, C. J. W. H. L. Dynamic parallelization of hydrological model simulations. *Environmental Modelling and Software*, Elsevier Ltd, v. 26, n. 12, p. 1736–1746, 2011. ISSN 13648152. Disponível em: <http://dx.doi.org/10.1016/j.envsoft.2011.07.015>. Acesso em: 8 Mar. 2015.

UFRGS. *Caderno dos Cursos Permanentes das Escolas Regionais de Alto Desempenho*. [S.l.: s.n.], 2006.

ZHANG, X. et al. Efficient multi-objective calibration of a computationally intensive hydrologic model with parallel computing software in Python. *Environmental Modelling and Software*, Elsevier Ltd, v. 46, p. 208–218, 2013. ISSN 13648152. Disponível em: <http://dx.doi.org/10.1016/j.envsoft.2013.03.013>. Acesso em: 20 Jan. 2015.



Inferência da estimativa de produtividade do solo a partir de dados de produção pecuária

Marina Silva Gomes[1], *Ana Paula Lüdtke Ferreira*[1], *Vinícius do Nascimento Lampert*[2]

[1]Programa de Pós-Graduação em Computação Aplicada, Universidade Federal do Pampa
Bagé, Rio Grande do Sul, Brasil
gomes.marina93@gmail.com, ana.ferreira@unipampa.edu.br

[2]Embrapa Pecuária Sul
Bagé, Rio Grande do Sul, Brasil
vinicius.lampert@embrapa.br

RESUMO

A *pecuária de precisão* visa aumentar a eficiência da produção pecuária levando a melhores resultados econômicos. Para que possa ser implantada com efetividade, são necessárias a coleta e a análise de uma grande quantidade de dados. A automatização dessa tarefa, por meio de técnicas matematicamente robustas, é um objetivo a ser atingido. Neste trabalho são apresentados dois modelos de inferência da capacidade produtiva do solo em função dos dados de produção, baseados em árvores de decisão e redes Bayesianas – dois modelos que implementam raciocínio probabilístico. Adicionalmente, usa-se os modelos construídos para descobrir quais são as variáveis mais influentes nos resultados, com vistas à melhoria dos processos de tomada de decisão.

PALAVRAS-CHAVE: Produtividade do Solo, Pecuária de Precisão, Classificação.

ABSTRACT

Livestock precision aims to increase production efficiency leading to better financial results. Effective implementation of livestock precision requires the gathering and analysis of massive amounts of data. Automation of such task, by means of mathematical sound techniques is still a goal to be achieved. This paper presents two distinct models to perform inference on soil productivity based on production data: decision trees and Bayesian networks – two models designed to perform probabilistic reasoning. Addionally, those models allow to find the most influential variables to be used in decision making processes.

KEYWORDS: Soil Productivity, Livestock Precision, Classification.

INTRODUÇÃO

A pecuária de corte possui importante representatividade na economia do país (BRASIL, 2015): segundo dados da Associação Brasileira das Indústrias Exportadoras de Carnes (ABIEC), o Brasil exportou em 2016 cerca de 1,4 milhão de toneladas de carne bovina, com faturamento de aproximadamente US$ 5,5 bilhões. Dados da Associação Brasileira de Proteína Animal (ABPA) informam que as exportações de carne suína do Brasil cresceram 32% em 2016, com um volume de 732,9 mil toneladas entre produtos *in natura* e processados. A receita do ano foi de US$ 1,483 bilhão, 16% a mais que em 2015. No entanto, segundo (LUPINACCI; ZEFERINO, 2000a), os índices de produtividade nessa atividade a descrevem como ineficiente quanto aos seus objetivos, não sendo capaz de competir de forma lucrativa com outras modalidade de exploração econômica da terra.

A chamada *pecuária de precisão* visa diminuir a quantidade de recursos utilizados para a produção e, ao mesmo tempo, maximizar a produção da propriedade – levando a melhores resultados econômicos diretamente. A pecuária de precisão lida com sistemas automatizados de coleta de informações referentes à produção, com vistas à geração de dados que, após tratamento, geram informação referente aos processos de produção que podem ser utilizados para melhorias e otimizações significativas desses processos. Lupinacci e Zeferino (2000b) afimam que a adoção de medidas de planejamento associadas com ações de manejo de pastagens e dos animais podem estabelecer sistemas de produção eficientes e produtivos. Porém, muitos pecuaristas resistem à ideia de realizar registros sobre a sua produção, e se fazem não dão a devida importância ao dados. No entanto o conhecimento sobre estes dados pode melhorar os resultados econômicos da atividade.

As fases de *recria* e *terminação*, por exemplo, tem grande sensibilidade a intervenções no processo de produção, que podem acarretar mudanças na qualidade dos resultados da produção e consequentemente na eficiência econômica da atividade (LAMPERT et al., 2002). O conhecimento sobre a capacidade de suporte das áreas de pastagem em função dos resultados de produção também pode ser usado para ajustar as práticas de manejo, levando a melhores resultados. Este trabalho descreve o uso de métodos automatizados de classificação de dados para construir modelos para predição da produtividade do solo (produção por hectare) na pecuária. Tem-se como objetivo também verificar quais variáveis, dentro do conjunto de dados disponíveis, influenciam na produtividade do solo.

O processo de classificação consiste em encontrar um conjunto de modelos e funções que descrevem e distinguem classes ou conceitos, com o propósito de predizer a classe de pertinência de objetos ainda não avaliados (AMO, 2004). O processo possui duas etapas (HAN; PEI; KAMBER, 2011): (i) construção de um modelo, a partir de dados conhecidos, que descreva o mapeamento entre os dados (atributos de entrada) e as classes (atributos de saída) e (ii) classificar as novas instâncias a partir do modelo construído.

O intuito deste trabalho é investigar a pertinência de duas técnicas clássicas de descoberta

de conhecimento – nomeadamente, *árvores de decisão* e *redes Bayesianas* – para a predição da produtividade do solo a partir de dados de produção. Como efeito colateral, procura-se verificar quais são as variáveis mais relevantes para essa conclusão. Uma *árvore de decisão* é uma estrutura binária onde cada nodo contém um teste. A avaliação do teste determina se a decisão seguirá o ramo esquerdo ou o ramo direito. As folhas da árvore são os nodos rotulados com as categorias de classificação. O algoritmo de indução de árvores de decisão é baseado na Teoria da Informação de Shannon (SHANNON; WEAVER, 1949), que busca particionar o conjunto de dados de acordo com a máxima entropia de informação associada a cada variável em cada ponto de decisão. Uma *rede Bayesiana* (RULFF; MADSEN, 2008) implementa raciocínio probabilístico, tanto indutivo como abdutivo, e consiste em um grafo dirigido acíclico (DAG) (CORMEN et al., 2001) em que os vértices representam as variáveis do domínio do problema e os arcos representam dependências causais entre elas. A rede representa ainda a distribuição de probabilidade conjunta, fatorizada por um conjunto de probabilidades condicionais determinadas pela estrutura do grafo.

Este trabalho está organizado como se segue: nesta seção foi introduzido o problema a ser tratado e as abordagens selecionadas para a solução do problema; na seção seguinte serão explicitados os materiais e métodos selecionados; na próxima seção são apresentadas discussões sobre os resultados obtidos e por último, são apresentadas as conclusões do trabalho.

MATERIAIS E MÉTODOS

Este trabalho segue a metodologia de pesquisa experimental aplicada, visto ter por objetivo investigar a pertinência de técnicas clássicas de raciocínio probabilístico do domínio da pecuária de precisão. A base de dados utilizada neste trabalho consiste em medições sobre informações zootécnicas da pecuária de corte coletadas dentro de projeto de um dos autores. A base de dados contém as seguintes informações, para cada animal do rebanho: (i) produtividade do solo (Kg PV/ha), (ii) produtividade do animal (%), (iii) carga animal (UA/ha), (iv) taxa de desmame (%), (v) idade de acasalamento (anos), (vi) idade de venda (anos) (vii) peso do boi gordo em quilogramas (kg), (viii) peso da vaca gorda em quilogramas (kg) e (ix) taxa de mortalidade (%). O conjunto de dados é formado por 10.000 instâncias e 9 atributos. Destas, 70% foram utilizadas para treinamento e 30% para teste. Note-se que existe a necessidade da separação dos dados, para assegurar que as técnicas adotadas são capazes de extrapolar o conhecimento adquirido e não meramente memorizar instâncias de dados.

Os valores atribuídos originalmente para cada atributo/variável no conjunto de dados foram numéricos. Desta forma, a produtividade do solo, atributo que se quer classificar e verificar a importância das outras variáveis sobre a sua classificação, foi transformado para um atributo discreto, tendo suas classes como *54a124*, *124a 176*, *176a229* e *229a410*. A nomenclatura dessas classes significa intervalos da variável produtividade do solo, por exemplo: produtividade do solo de 54 a 124 (kg PV/ha). Para realizar a divisão das classes para o atributo produtividade

do solo, foram calculados seus valores máximo, mínimo, mediana, 1º quartil e 3º quartil. Após a discretização do conjunto de dados foi realizada a geração dos modelos e a classificação da produtividade do solo sobre os dados de treinamento e de teste.

Tabela 1: Variáveis do modelo de classificação.

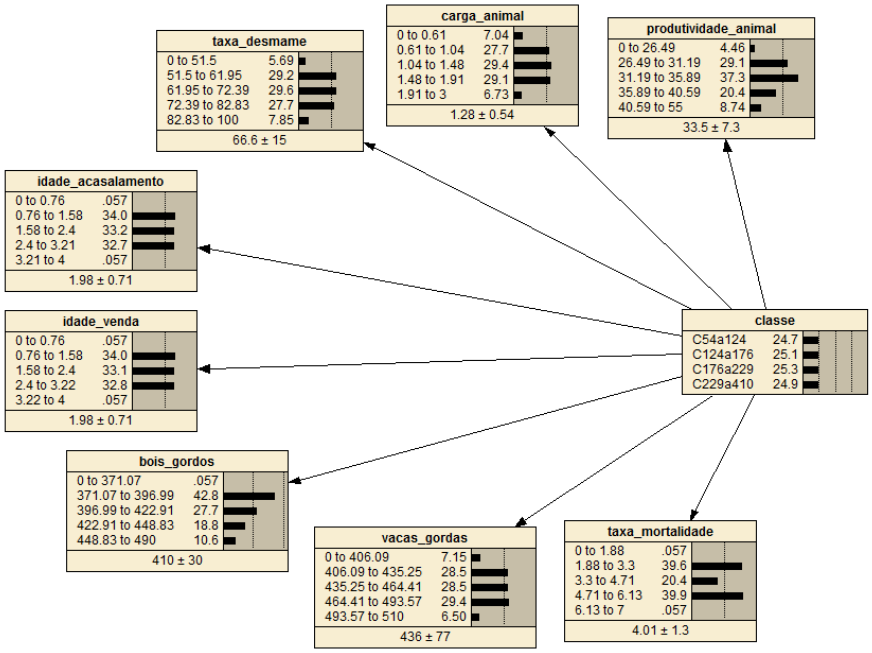
Descrição	Valores
(ii) Produtividade do animal (%)	<=26.49
	>26.49 e <=31.19
	>31.19 e <=35.89
	>35.89 e <=40.59
	>40.59
(iii) Carga animal (UA/ha)	<=0.61
	>0.61 e <=1.04
	>1.04 e <=1.48
	>1.48 e <=1.91
	>1.91
(iv) Taxa de desmame (%)	<=51.50
	>51.50 e <=61.95
	>61.95 e <=72.39
	>72.39 e <=82.83
	>82.83
(v) Idade de acasalamento e (vi) Idade de venda (anos)	<=0.76
	>0.76 e <=1.58
	>1.58 e <=2.40
	>2.40 e <=3.21
	>3.21
(vii)Bois Gordos (kg/animal)	<=371.07
	>371.07 e <=396.99
	>396.99 e <=422.91
	>422.91e <=448.83
	>448.83
(viii)Vacas Gordas (kg/animal)	<=406.09
	>406.09 e <=435.25
	>435.25 e <=464.41
	>461.41 e <=493.57
	>493.57
(ix) Taxa de mortalidade (%)	<=1.88
	>1.88 e <=3.30
	>3.30 e <=4.71
	>4.71 e <=6.13
	>6.13

Classificação com Redes Bayesianas

A modelagem de uma rede Bayesiana é dividida em três etapas: seleção de variáveis, construção da estrutura da rede e estimação de probabilidades (FARIA, 2014). A seleção de variáveis é feita com base no conhecimento sobre o domínio do problema e nas informações de interesse. A

construção da estrutura pode ser realizada manualmente, levando em conta relações de causalidade entre as variáveis selecionadas, ou por meio do uso de algoritmos. A última etapa, de estimação de probabilidades, pode ser realizada também com base em conhecimentos subjetivos (caso não haja quantidade suficiente de dados) ou a partir do uso de algoritmos. As variáveis usadas na construção da rede foram as já mencionadas informações coletadas por animal; o domínio das variáveis foi definido segundo a média mais dois desvios padrão e a média menos dois desvios padrão; e a estrutura da rede foi definida de acordo com o conhecimento sobre o domínio do problema. Essa informação encontra-se sumarizada na Tabela 1.

Figura 1: Rede Bayesiana para classificação da produtividade do solo



Fonte (Autor, 2017)

A rede construída foi descrita com apoio do software *Netica*[1], conectando as variáveis selecionadas de forma a se obter um grafo acíclico direcionado que representasse o modelo construído. A construção da estrutura da rede foi realizada manualmente. Por se tratar de um modelo para classificação, o nodo que se quer predizer foi ligado a todos os outros nodos. A Figura 1 apresenta a rede construída para este problema de classificação.

A etapa da estimação das probabilidades condicionais para cada nodo da rede foi implementada pelo software Netica, a partir da carga de dados de treinamento. Após a definição das probabilidades condicionais de cada nodo, foram utilizados dados de teste como evidência para verificar se o modelo desenvolvido classifica corretamente o atributo produtividade do solo. O programa gerou a matriz de confusão do teste, que indica quais foram as amostras corretamente e incorretamente classificadas. O modelo desenvolvido possui uma taxa de erro de 30.45%. A Tabela 2 apresenta a matriz de confusão para a rede, onde é possível verificar que as classes C176a229 e C124a176 são confundidas no processo de classificação.

[1]Netica Application: https:// www.norsys.com/netica.html

Tabela 2: Matriz de confusão para o modelo com redes bayesianas.

Classe	C54a124	C124a176	C176a229	C229a410
C54a124	690	64	3	0
C124a176	186	432	118	0
C176a229	0	306	311	101
C229a410	0	16	107	625

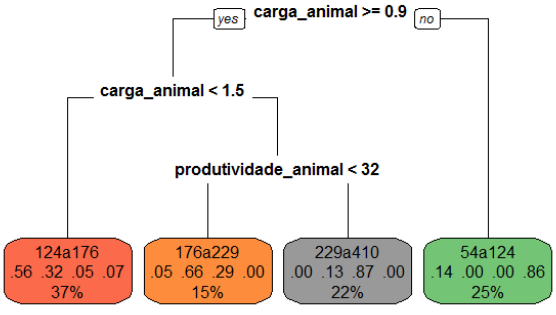
Classificação com Árvores de Decisão

O método de classificação por árvore de decisão funciona como um fluxograma em forma de árvore, onde cada nodo interno (não folha) indica um teste feito sobre um valor. As ligações entre os nós representam os valores possíveis do teste do nó superior, e as folhas indicam a classe (categoria) a qual o registro pertence. Após a árvore de decisão montada, para realizar a classificação de um novo registro, basta seguir o fluxo na árvore (mediante os testes nos nós não-folhas) começado no nó raiz até chegar a uma folha, que determina a classe de pertencimento do elemento.

Para a construção de uma árvore de decisão para o conjunto de dados trabalhado foi utilizado o pacote *rpart* (THERNEAU; ATKINSON; RIPLEY, 2014) do pacote probabilístico R. Os dados que foram anteriormente transformados e divididos em dados de treinamento e teste foram utilizados para construção (dados de treino) e validação (dados de teste) da árvore.

Após a utilização da função *rpart* para construção da árvore de decisão, foi utilizado o comando plot() para mostrar a árvore gerada. Além disso a árvore também foi podada, retirando ramos que tornavam a classificação redundante. A Figura 2 apresenta a árvore de decisão gerada.

Figura 2: Árvore de decisão



Fonte (Autor, 2017)

A partir da árvore de decisão gerada é possível verificar as principais variáveis utilizadas para classificar a produtividade do solo, que são a carga animal e produtividade do animal. A Tabela 3 apresenta a matriz de confusão para a árvore de decisão gerada, de onde é possível derivar a taxa de erro para o modelo – que é de 29.84%. Novamente é possível verificar a confusão na classificação da classe C176a229.

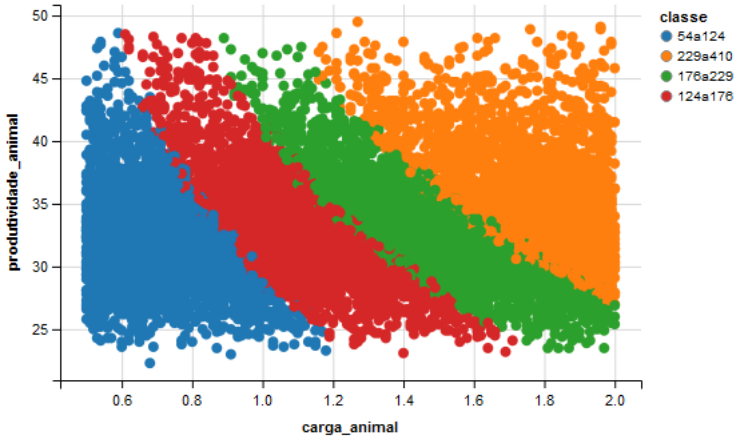
Tabela 3: Matriz de confusão para o modelo com árvores de decisão

Classe	C54a124	C124a176	C176a229	C229a410
C54a124	686	108	0	0
C124a176	106	605	30	0
C176a229	0	352	282	97
C229a410	0	59	141	526

RESULTADOS E DISCUSSÃO

A partir da geração da matriz de confusão dos modelos, pode-se perceber que os modelos obtiveram uma precisão muito próxima, com uma taxa de erros de 29.84% para árvores de decisão e 30.45% para redes Bayesianas. É possível verificar também que há uma confusão na classificação da classe C176a229, classificando um número considerável de amostras como sendo da classe C124a176 tanto no modelo com rede Bayesiana quanto no modelo de árvore de decisão. Isso pode ter ocorrido devido a intersecção/sobreposição de dados de ambas as classes ou os atributos disponíveis para classificação não foram os adequados, dificultando a distinção das classes, ou ainda pelo particionamento não adequado dos valores possíveis para as variáveis escolhidas. A Figura 3 apresenta a distribuição do conjunto de dados, que mostra que existe sobreposição de valores para as diferentes classes.

Figura 3: Árvore de decisão



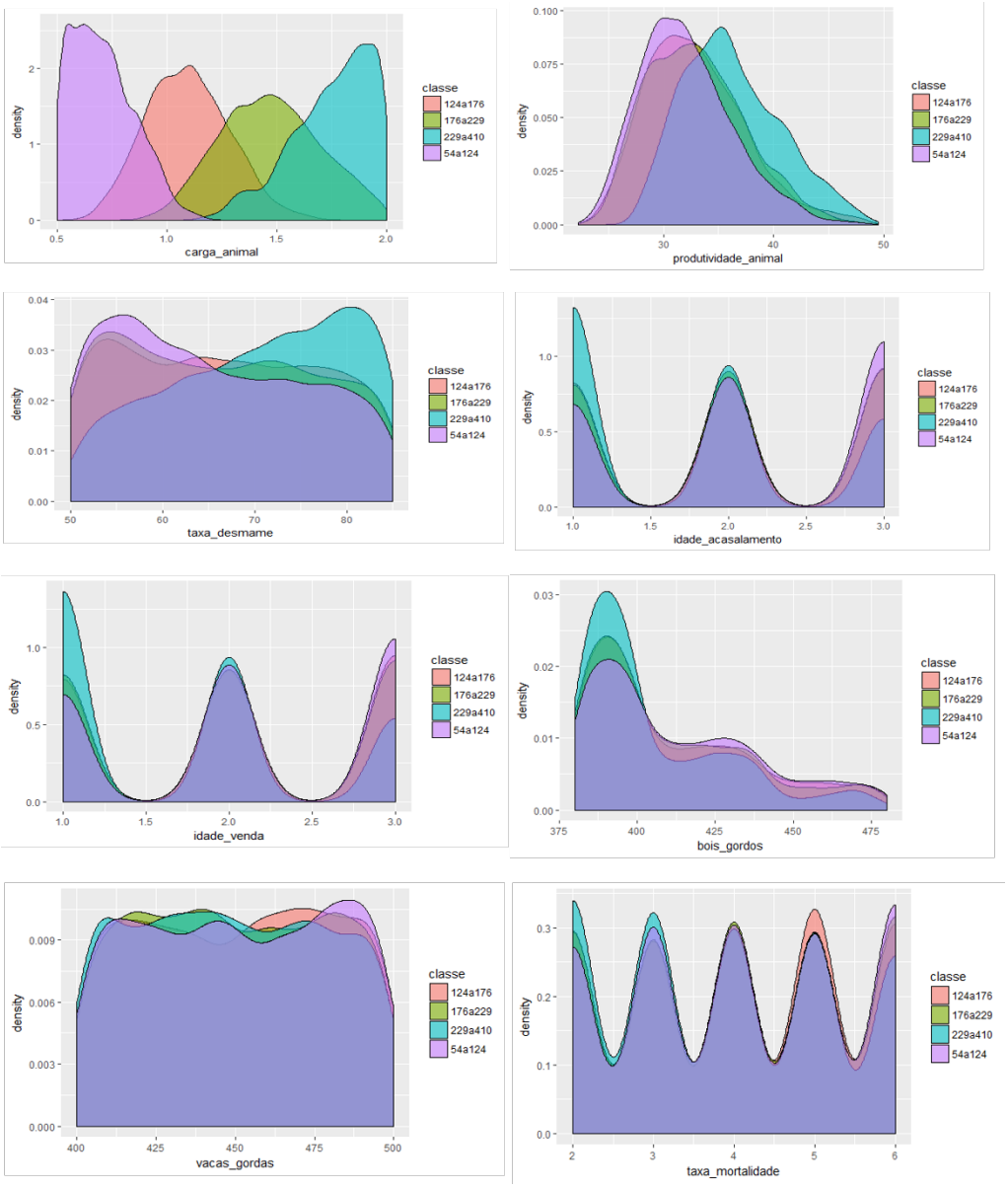
Fonte: (Autor, 2017)

A investigação deste trabalho de buscar dentre as variáveis do conjunto de dados quais delas influenciam na produtividade do solo foi possível por meio do resultado produzido pela árvore de decisão e de testes sobre a rede Bayesiana, onde verificou-se que a carga animal combinada com a produtividade do animal influenciam todas as classificações da produtividade do solo. Isso determina que não há necessidade de uso de mais variáveis quando se deseja inferir a produtividade do solo a partir dos dados de produção recolhidos.

A causa da escolha das variáveis para a árvore de decisão pode ser vista pela Figura 4, que

apresenta os gráficos de densidade das amostras do conjunto de dados. É possível identificar que a variável carga animal é a que melhor consegue diferenciar entre as classes de produtividade do solo. As redes Bayesianas não utilizam apenas uma ou duas variáveis para definir a classificação de um atributo, mas sim a combinação das probabilidades, porém neste caso a partir de testes também foi possível verificar que a carga animal discrimina bem entre uma classe e outra.

Figura 4: Gráficos de densidade do conjunto de dados



Fonte: (Autor, 2017)

CONCLUSÕES E TRABALHOS FUTUTOS

Este trabalho apresentou um modelo baseado em redes Bayesianas e outro em árvores de decisão para inferência da produtividade do solo a partir de dados de produção pecuária. Os resultados iniciais mostraram que ambos os métodos podem ser usados com uma taxa de erro

ainda considerada alta. Essa quantidade de erros precisa ser investigada com mais detalhes, para verificar se é um problema na qualidade dos dados, no particionamento (discretização) dos valores possíveis para cada variável do modelo ou se é um erro inerente aos métodos usados.

O uso dos modelos usados para verificação das dependências causais entre as variáveis do modelo é mais promissor, permitindo que dados de produção coletados possam ser sistematizados em termos de suas dependências, permitindo a construção de modelos corretos mas mais enxutos, com menos variáveis e computacionalmente mais eficientes. Esses modelos devem permitir a construção de sistemas de apoio à decisão matematicamente bem fundamentados.

REFERÊNCIAS

ABIEC. *Associação Brasileira das Indústrias Exportadoras de Carnes*. 2016. Disponível em: <http:// http://www.abiec.com.br/>.

AMO, S. de. Técnicas de mineração de dados. *Jornada de Atualizaçao em Informatica*, 2004.

BRASIL. *Estatisticas. Ministério da Agricultura, Pecuária e Abastecimento - MAPA*. 2015. Disponível em: <www.agricultura.gov.br/acessoainformacao/estatistica>.

CORMEN, T. H. et al. *Introduction to algorithms*. [S.l.]: MIT press Cambridge, 2001.

FARIA, R. C. *Redes probabilísticas: aprendendo estruturas e atualizando probabilidades*. Tese (Doutorado) — Universidade de São Paulo, 2014.

HAN, J.; PEI, J.; KAMBER, M. *Data mining: concepts and techniques*. [S.l.]: Elsevier, 2011.

LAMPERT, V. d. N. L. et al. Sistema de apoio às decisões econômicas no manejo de novilhas de corte. Universidade Federal de Viçosa, 2002.

LUPINACCI, A. V.; ZEFERINO, C. V. *Índices de produtividade da pecuária de corte no Brasil. Parte 1/3*. 2000. Disponível em: <www.beefpoint.com.br/cadeia-produtiva/conjuntura-de-mercado/indices-de-produtividade-da-pecuaria-de-corte-no-brasil-parte-13-3877/>.

LUPINACCI, A. V.; ZEFERINO, C. V. *Índices de produtividade da pecuária de corte no Brasil. Parte 3/3*. 2000. Disponível em: <www.beefpoint.com.br/cadeia-produtiva/conjuntura-de-mercado/indices-de-produtividade-da-pecuaria-de-corte-no-brasil-parte-33-3879/>.

RULFF, U. B. K.; MADSEN, A. L. *Bayesian Networks and Influence Diagrams – A Guide to Construction and Analysis*. [S.l.]: Springer, 2008. (Information Science and Statistics). ISBN 978-0-387-74100-0 (Print) 978-0-387-74101-7 (Online).

SHANNON, C. E.; WEAVER, W. The mathematical theory of information. University of Illinois Press, 1949.

THERNEAU, T.; ATKINSON, B.; RIPLEY, B. rpart: Recursive partitioning and regression trees. r package version 4.1-8. *Available at CRAN. R-project. org/package= rpart. Accessed May*, 2014.

Sistema para avaliação de risco de resistência de insetos a toxinas Bt expressas em culturas transgênicas

Aline de Holanda Nunes Maia[1], *Marcelo Gonçalves Narciso*[2], *Durval Dourado Neto*[3]

[1] Embrapa Meio Ambiente, Jaguariúna, São Paulo, Brasil, aline.maia@embrapa.br

[2] Embrapa Arroz e Feijão, Santo Antônio de Goiás, Goiás, Brasil, marcelo.narciso@embrapa.br

[3] Departamento de Produção Vegetal, Escola Superior de Agronomia Luiz de Queiroz, Universidade de São Paulo, Piracicaba, São Paulo, Brasil, ddourado@usp.br

RESUMO

O sistema RRiskBtWeb foi desenvolvido para avaliação de risco de evolução de resistência de insetos a toxinas Bt expressas em culturas transgênicas. A evolução de resistência de insetos a qualquer agente de controle é um processo natural e contribui para a redução da vida útil da tecnologia. RRiskBtWeb é uma versão WEB do software RRiskBt, desenvolvido em Visual Basic, em 2003. Naquela época, não era ainda permitido o cultivo de plantas transgênicas no Brasil e o software foi construído para auxiliar no estabelecimento de estratégias preventivas de manejo de resistência, anteriores à liberação da tecnologia de plantas transgênicas inseticidas. Recentemente, após 2013, começaram a surgir falhas de controle de importantes lepidópteros-praga nas lavouras de milho transgênico, acarretando sérios danos econômicos nos sistemas de produção. Esse evento motivou o desenvolvimento da versão mais acessível aos usuários em geral, o RRiskBtWeb, que pode ser executado em qualquer computador com um navegador de Internet. Esse sistema simula o processo de evolução da resistência em diferentes cenários, gerando informações cientificamente embasadas para auxílio à tomada de decisão sobre estratégias de manejo da resistência, especialmente o tamanho de refúgio.

PALAVRAS-CHAVE: RRiskBtweb, *Bacillus thuringiensis*, culturas inseticidas.

ABSTRACT

The RRiskBtWeb System was developed for assessing the risk of insect resistance evolution to Bt toxin expressed in transgenic insecticidal crops. Insect resistance evolution in response to any control agent is a natural process, which contributes to reduction of the technology

useful life. RRiskBtWeb is a Web version of the RRiskBt software, developed in Visual Basic language in 2003. At that time, cultivation of transgenic crops was not still allowed in Brazil; the software was constructed as a tool for helping definition of preventive resistance management strategies, before liberation of the insecticidal plant technology. Recently, after 2013, control failures of important Lepidopteran pests started to appear in transgenic crop fields, leading to serious economic damage in the corn production systems. Such event motivated the development of a more accessible version, the RRriskBtWeb, which can be run in any computer with Internet browser. This system simulates resistance evolution process in different scenarios thus generating scientifically sound information for helping decision making on resistance management strategies, specially the refuge size.
KEYWORDS: RRiskBtweb, *Bacillus thuringiensis*, insecticidal crops.

INTRODUÇÃO

Culturas transgênicas que expressam genes que codificam para a produção de toxinas da bactéria *Bacillus thuringiensis* (Bt), são cultivadas em vários países desde 1996. As toxinas Bt são ativas contra algumas espécies de pragas das ordens Lepidoptera, Coleoptera e Diptera. Por esse motivo, as culturas nas quais esses genes foram incorporados são chamadas plantas transgênicas inseticidas. O principal risco ambiental associado à adoção das culturas inseticidas é a evolução de resistência das pragas-alvo a essas toxinas, processo que pode ser retardado pela adoção de estratégias de manejo. A principal estratégia adotada é o uso de materiais genéticos que expressem a toxina Bt em alta concentração nos tecidos da planta, combinada com adoção de refúgios estruturados. Os refúgios são áreas ocupadas por plantas do mesmo híbrido ou variedade da cultura transgênica, mas que não expressem a toxina (ILSI-HESI, 1998; USEPA, 2001; SCIENTIFIC ADVISORY PANEL, 2001). A modelagem matemática é a ferramenta utilizada para definição do tamanho e localização dos refúgios, uma vez que experimentos em campo para monitorar evolução de resistência em diferentes cenários são impraticáveis.

O sistema RRiskBtWeb é uma versão Web do Software RRiskBt (não registrado), desenvolvido em Visual Basic por Maia et al. (2003). Naquela época, não era ainda permitido o cultivo de transgênicos no Brasil e o uso de modelos seria útil para o estabelecimento de estratégias preventivas anteriores à adoção da tecnologia de plantas inseticidas. O software RriskBT foi utilizado no âmbito do projeto internacional intitulado *International Project on GMO Environmental Risk Assessment Methodologies*, (http://www.gmoera.umn.edu/) num

estudo de caso sobre avaliação de risco de resistência da lagarta-do-cartucho (*Spodoptera frugiperda*) a toxinas Bt expressas na cultura do algodão.

O modelo matemático que serviu de base para o desenvolvimento do software RRiskBt, codificado em linguagem Java (Caprio,1998a), estava disponível, à época, sem restrição de acesso; seus fundamentos teóricos estão descritos em Caprio (1998b). Trata-se de um modelo determinístico para simular a evolução de resistência via projeções pontuais da frequência do alelo de resistência (FreqR) à toxina Bt incorporada numa cultura transgênica, na população de uma praga-alvo, ao final de cada geração da praga, em diferentes cenários. Os cenários são definidos pelo tamanho da área de refúgio (área ocupada pela cultura convencional, não transgênica) e pelo manejo de pragas nessa área (USEPA, 2001). No software RRiskBt, foram incorporados módulos para análise de sensibilidade e incertezas não contempladas no Modelo de Caprio (Caprio,1998a). Recentemente, após 2013, começaram a surgir falhas de controle de diversas pragas do milho em materiais Bt, o que ocasionou o resgate do interesse pelos modelos de evolução de resistência de insetos a culturas inseticidas. Esse evento, motivou a construção de uma versão Web do RRriskBt, para atendimento à demanda dos setores envolvidos.

O sistema RRiskBtWeb foi construído para ser executado a partir de um navegador e pode ser acessada em qualquer ambiente (IBM AIX, Linux, FreeBSD, Windows, MacOS, etc.) enquanto a versão em Visual Basic (RRiskBt) roda apenas em ambiente Windows. A disponibilização desse sistema é uma demanda de diferentes setores envolvidos na pesquisa sobre manejo da evolução da resistência de insetos em culturas transgênicas inseticidas.

O objetivo deste trabalho é apresentar o sistema RRiskBtWeb e demonstrar sua utilidade na geração de informações técnico-científicas para auxílio à definição de políticas públicas relacionadas à regulamentação do manejo preventivo da resistência de insetos, fundamentado na adoção de áreas de refúgio.

MATERIAL E MÉTODOS

O Sistema RRiskBtWeb foi desenvolvido utilizando as linguagens HTML, CSS, e JavaScript, para simular o processo de evolução da resistência à toxina Bt, em populações de insetos-praga de culturas transgênicas inseticidas. É uma versão do software RRiskBt (não registrado), desenvolvido em 2003 (MAIA, 2003).

O software RRIskBt

O software RRiskBt, construído em linguagem Visual Basic, foi baseado no modelo determinístico de Caprio (CAPRIO, 1998a e CAPRIO, 1998b), para gerar projeções da frequência de insetos resistentes à toxina Bt, em populações de pragas-alvo das culturas de milho e algodão Bt, a partir de informações sobre: i) características da cultura transgênica; ii) genética de resistência, sobrevivência e dispersão da praga-alvo e iii) fatores operacionais do manejo de pragas área da cultura não transgênica (refúgio) (MAIA, 2003; MAIA e DOURADO-NETO, 2004).

No modelo de Caprio e versões dele derivadas (RRIskBt e RRiskBtWeb), considera-se que a resistência é governada por um par de genes com dominância incompleta: os indivíduos homozigotos recessivos (RR) são resistentes a qualquer concentração da toxina Bt expressa nos tecidos da planta, os heterozigotos (RS) podem ser resistentes ou susceptíveis, dependendo da concentração da toxina e os indivíduos SS são susceptíveis.

A evolução da resistência é quantificada pelas frequências do alelo R (FreqR) ou do genótipo RR (FreqRR), ao longo das gerações da praga, até ser atingida uma frequência crítica (FreqCritica) definida pelo usuário. Os parâmetros de entrada do modelo (*input data*) são descritos na Tabela 1.

Tabela 1 - Dados de entrada para simular a evolução da resistência à toxina Bt em pragas-alvo, utilizando o modelo determinístico de Caprio.

Parâmetro	Descrição
AreaDeRefugio	Proporção da área total ocupada pelo refúgio
FreqInicial	Frequência inicial do alelo de resistência (R) na área de refúgio
FreqCritica	Frequência crítica do alelo R na área da cultura transgênica
SFImatRef	Sobrevivência da lagarta-do-cartucho durante a fase imatura, na área de refúgio.
CoefEndogamia	Coeficiente de endogamia da população da lagarta-do-cartucho
SFImatSStran	Sobrevivência da subpopulação SS durante a fase imatura na área da cultura transgênica
SFImatRRtran	Sobrevivência da subpopulação RR durante a fase imatura na área da cultura transgênica
SFImatRStran	Sobrevivência da subpopulação RS durante a fase imatura na área da cultura transgênica
DispPRErand	Proporção de indivíduos de cada *habitat* que se dispersa de forma aleatória na área total antes do acasalamento
DispPOSrand	Proporção de indivíduos com dispersão pré-acasalamento restrita aos respectivos *habitats* de origem que se dispersa de forma aleatória na área total após o acasalamento

No RRiskBt, também foram incorporados ao modelo de Caprio, módulos de análise de sensibilidade e de análises de incertezas. A análise de sensibilidade permite avaliar a influência de variações nos parâmetros de entrada do modelo sobre as variáveis prognósticas

(FreqR ao longo das gerações da praga e número de gerações até que a FreqCritica seja atingida, NGer). A análise de incertezas considera incertezas associadas a alguns parâmetros de entrada do modelo, representadas via distribuições de probabilidade. Permite obter projeções probabilísticas das métricas da evolução de resistência e dessa forma quantificar o de risco de resistência na população de insetos ao longo das gerações, em diferentes cenários (MAIA e DOURADO-NETO, 2004).

O Sistema RRiskBtWeb

O sistema RRiskBtWeb foi construído para ser executado a partir de um navegador, e assim pode ser acessado por intermédio de um servidor ou ainda rodar em qualquer computador que tenha navegador Firefox, Safari ou Google Chrome. A versão em Visual Basic (RRiskBt) roda apenas em ambiente Windows, enquanto a versão Web pode ser acessada em qualquer ambiente (IBM AIX, Linux, FreeBSD, Windows, MacOS, etc.) desde que haja um navegador disponível, conforme descrito anteriormente. A linguagem usada para os cálculos foi JavaScript, que é executada no computador do usuário, não sendo necessário o uso de um servidor, embora o servidor possa ser usado para disponibilizar as páginas do sistema.

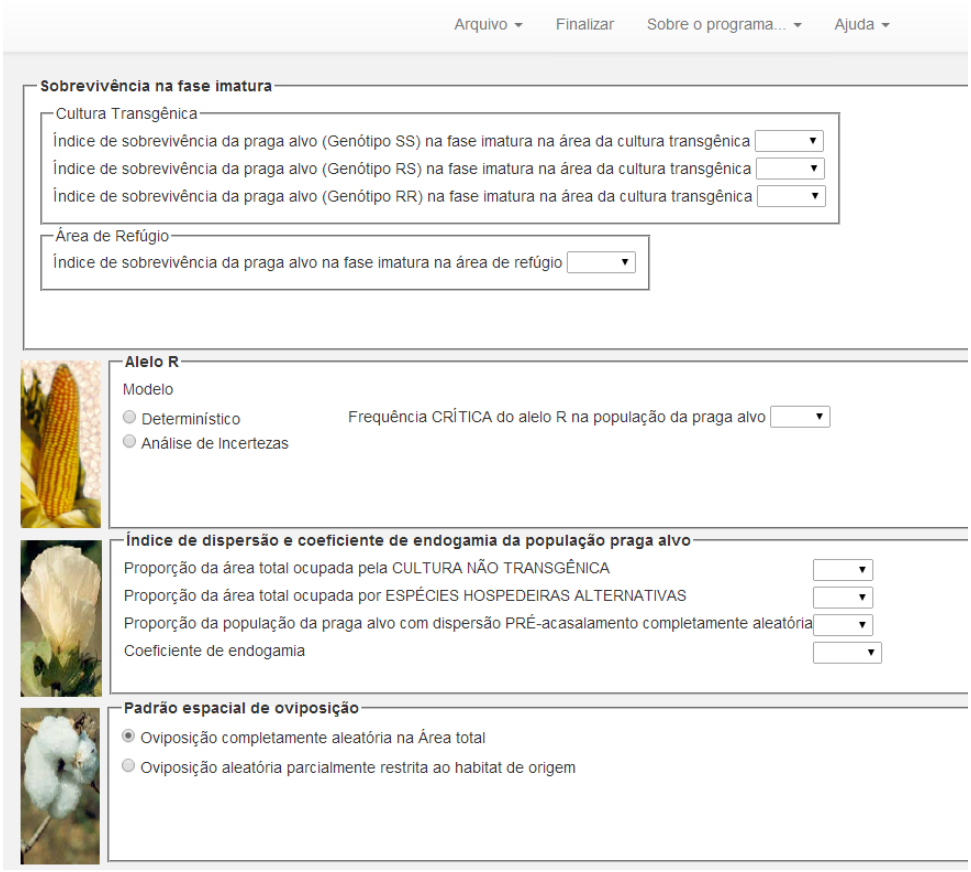


Figura 1 – Tela inicial do RRiskBtWeb com campos para inserção das seguintes informações: a) valores dos parâmetros de entrada do modelo, descritos na Tabela 1; b) escolha do tipo de análise: determinística ou incorporando análise de incertezas, e c) padrão espacial de oviposição.

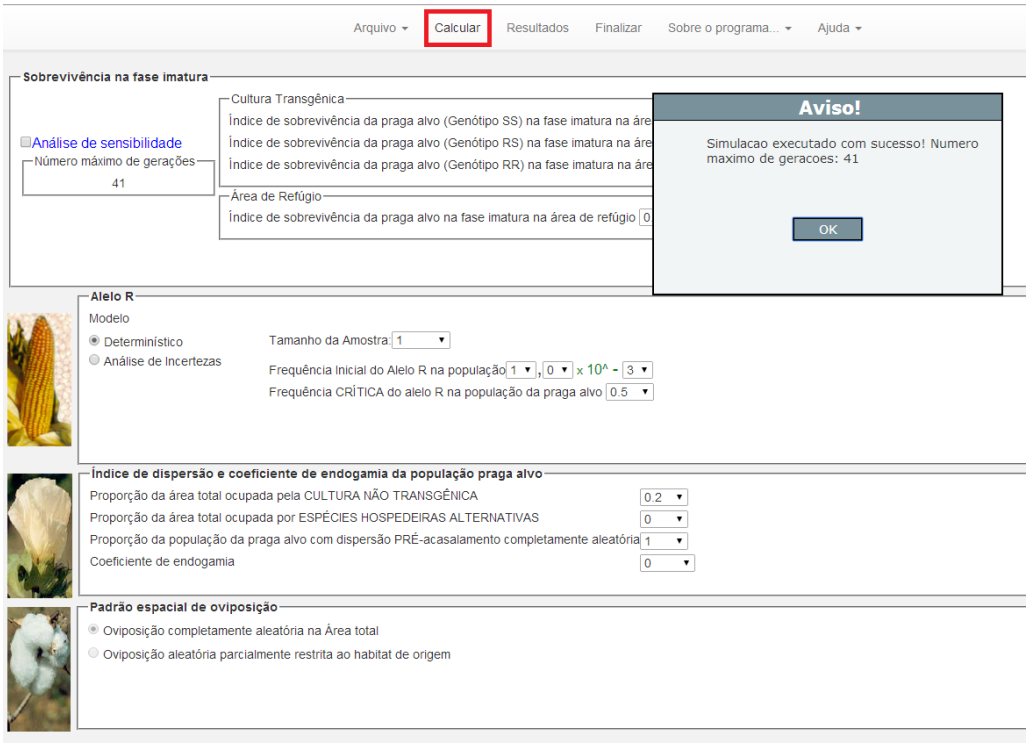


Figura 2 – Tela do sistema RRiskBtWeb obtida após execução de uma simulação do modelo determinístico, após informação sobre os valores dos parâmetros de entrada do modelo, descritos na Tabela 1.

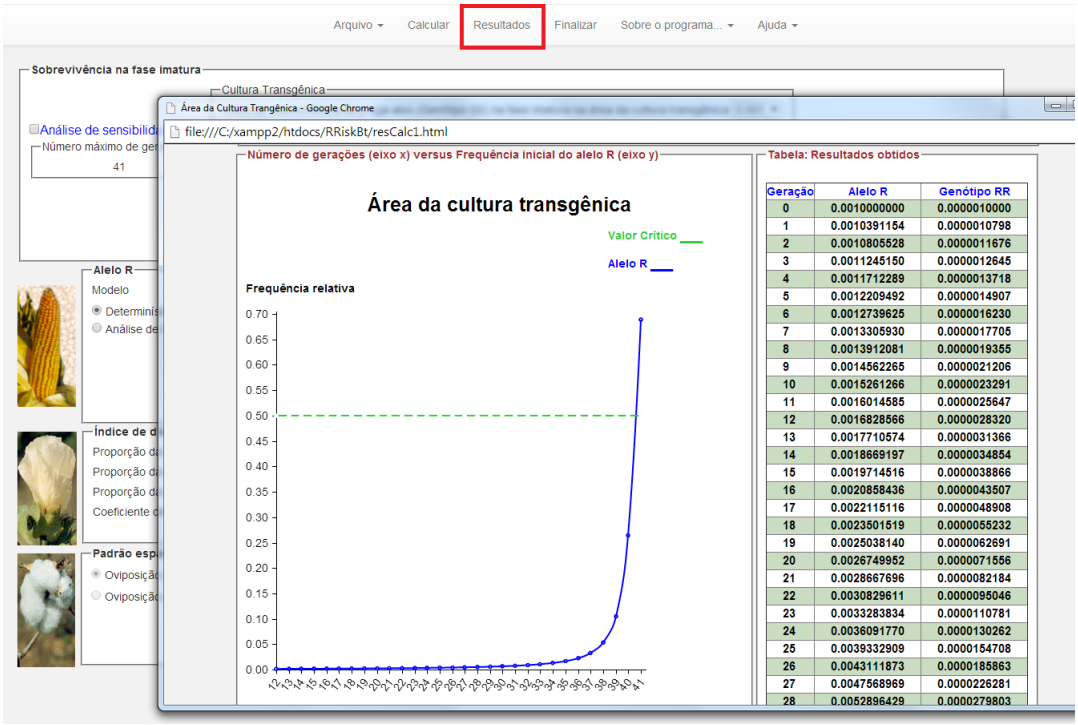


Figura 3 – Tela do sistema RRiskBtWeb obtida após execução de uma simulação do modelo determinístico, ao escolher a opção “Resultados”. O gráfico representa a evolução de resistência, expressa como frequência relativa do alelo R, ao longo das gerações da praga-alvo da toxina Bt. A tabela mostra os valores da frequência do alelo R e do genótipo RR, ao longo das gerações da praga, até que a frequência crítica de 0,5 (50%) seja atingida.

O programa permite que as informações sobre os parâmetros sejam digitadas na própria tela inicial (Figura 1) ou carregadas pela leitura de um arquivo de dados. Após a leitura e carregamento desse arquivo, o sistema fornece opções para executar o modelo e mostrar os resultados. Na Figura 2, é ilustrada a tela gerada após a execução de uma análise determinística, obtida ao clicar na opção “Calcular”;na Figura 3, tabelas e gráficos relativos ao resultado da simulação, obtidos ao clicar na opção “Resultados".

RESULTADOS E DISCUSSÃO

Neste sistema, a evolução da resistência pode ser simulada de duas formas distintas, a saber:

a) Determinística: são obtidos valores pontuais de FreqR ou FreqRR, ao longo das gerações, até que a FreqCritica seja atingida. Os dados de entrada são valores pontuais dos parâmetros do modelo.

b) Probabilística: são obtidos valores pontuais de FreqR ou FreqRR, ao longo das gerações, até que a FreqCritica seja atingida, para cada valor ou combinação de valores dos parâmetros do modelo considerados na análise de incertezas. Assim, são obtidas distribuições de probabilidade de FreqR, ao longo do tempo e calculado o risco de resistência (prob[FreqR>FreqCritica]) ao longo das gerações da praga (ANDERSON e HATTIS, 1999).

No primeiro caso, é gerada uma tabela com os valores de FreqR e FreqRR a cada geração, até que FreqR exceda a FreqCritica, e um gráfico, mostrando a evolução de FreqR ao longo das gerações da praga alvo. No modelo que incorpora análise de incertezas, são gerados os seguintes resultados: (i) um arquivo com os n valores de FreqInicial amostrados da distribuição de probabilidade escolhida para representação da incerteza desse parâmetro (Normal truncada, Triangular ou Uniforme) e o histograma correspondente; (ii) um arquivo com os valores de FreqR para cada geração da praga, correspondentes a cada valor amostrado de FreqInicial; (iii) um gráfico do risco de resistência em função do número de ordem da geração da praga-alvo; (iv) uma tabela com a informação utilizada para gerar o gráfico e (v) um arquivo com a seguinte informação: distribuição de probabilidade escolhida para

caracterizar a incerteza de FreqInicial, tamanho da área de refúgio utilizada na simulação, número de ordem da geração da praga e risco de resistência.

Apresentamos um estudo de caso onde é simulado o processo de evolução de resistência na população de uma praga-alvo, numa cultura transgênica com apenas uma toxina Bt inserida, expressa nos tecidos da planta numa concentração considerada 'alta dose', usando a abordagem probabilística. Os valores dos parâmetros de entrada do modelo foram os seguintes: SFImatRRTran=1; SFImatRSTran=0,002; SFImatSSTran=0,001; FreqCritica=0,5; DispPRErand=1; DispPOSrand=1 e CoefEndogamia=0. A evolução da resistência foi simulada para dois cenários (Tabela 2): (I) sem área de refúgio e (II) com área de refúgio de 20% da área total, com aplicação de um inseticida com eficiência de 80%, o que corresponde a uma mortalidade de 20% da fase imatura da praga na área de refúgio (SFImatRef=20% ou 0,2). Foi usado o valor médio de FreqInicial de 1×10^{-3}, com amplitude de 0 a 2×10^{-3} com base em informações de Andow e Alstad (1998). As distribuições usadas para caracterizar a incerteza desse parâmetro foram a Uniforme (U), Triangular (T) e Normal Truncada Simétrica (NTS).

Tabela 2 - Cenários utilizados para a geração da distribuição empírica de FreqR para diferentes combinações de tamanho de refúgio e sobrevivência da fase imatura da praga-alvo na área de refúgio.

Cenário	AreaDeRefugio	SFImatRef[(1)]	Distribution
I	0	-	Uniforme
	0	-	Triangular Simétrica
	0	-	Normal Truncada Simétrica
II	20	20[(2)]	Uniforme
	20	20	Triangular Simétrica
	20	20	Normal Truncada Simétrica

[1] Sobrevivência da fase imatura da praga-alvo na área de refúgio; [2]Aplicação de inseticida com 80% de eficiência para a praga-alvo na área de refúgio.

Observa-se uma rápida evolução do risco de resistência no cenário sem refúgio (I), com o risco atingindo 100% já na oitava geração da praga, independente da distribuição utilizada na análise de incertezas. Isso corresponde a quatro anos, numa região com duas gerações da praga-alvo por ano. Já no cenário com área de refúgio de 20% (II), o risco atinge 100% em torno da geração 200 – o que corresponde a 100 anos, considerando a mesma região do cenário I (Figura 4).

De posse dessas informações, é possível comparar os riscos associados a cada cenário e avaliar qual tamanho de refúgio seria mais adequado, considerando os custos envolvidos e o tempo de duração pretendido para a tecnologia Bt. Quanto maior o refúgio, maior a duração

da tecnologia, mas isso implica em perda de produtividade pelo ataque da praga no refúgio (MAIA e DOURADO-NETO, 2004).

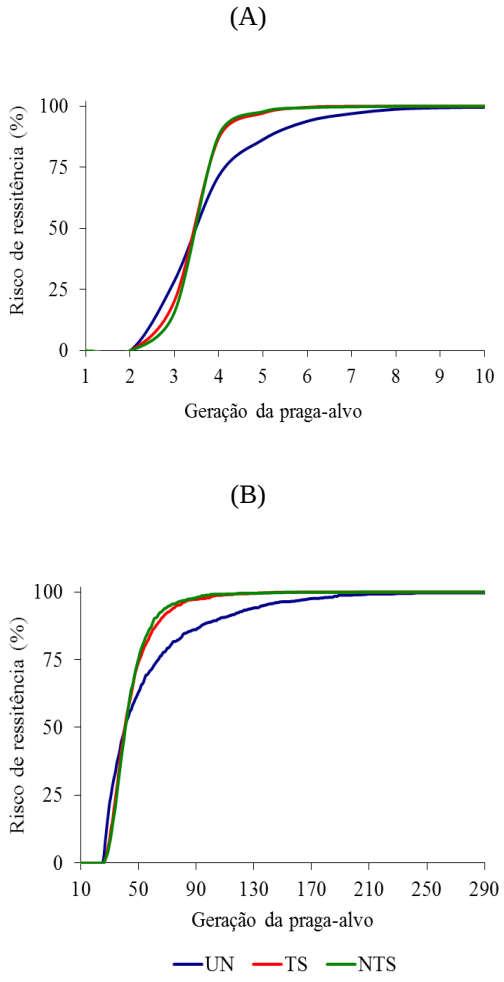


Figura 4 – Risco de resistência à toxina Bt expressa numa cultura transgênica, na população de uma praga alvo, em função do número de gerações da praga, para 0 (A) e 20% (B) de área de refúgio, considerando as diferentes distribuições para caracterizar a incerteza da frequência inicial do alelo de resistência (FreqInicial): Uniforme (UN), Triangular simétrica (TS) e Normal truncada simétrica (NTS)

CONCLUSÕES

O Software RRiskBtWeb apresenta a vantagem, em relação ao RRiskBt, de ser mais acessível aos usuários, podendo ser executado em qualquer computador com acesso à Internet. A abordagem probabilística para quantificar o risco de evolução de resistência é

fundamental para fornecer informações cientificamente embasadas sobre o tamanho de refúgio a ser recomendado, num estágio em que o conhecimento sobre os parâmetros dos modelos, especialmente para FreqInicial e DFRes, é ainda extremamente incerto. A análise de incertezas permite que amplitudes de variação dos parâmetros de entrada do modelo (*input parameters*) observadas na literatura, ou mesmo o conhecimento subjetivo de especialistas, seja incorporado à simulação via distribuições de probabilidade. O modelo implementado (modelo de Caprio) é relativamente simples, quando comparado aos modelos espaciais estocásticos, e portanto, mais adequado quando não há informação detalhada sobre os padrões de dispersão da praga nos períodos de pré e pos-oviposição. O software RRiskBtWeb permite a geração de informações cientificamente embasadas para auxílio à tomada de decisão sobre estratégias de manejo preventivo da resistência de insetos a toxinas expressas em plantas transgênicas, fundamentado na adoção de áreas de refúgio. O sistema também é útil como ferramenta didática para entendimento dos fatores que influenciam o processo de evolução da resistência e da importância das ferramentas probabilísticas para auxílio à tomada de decisão em ambientes de incerteza.

REFERÊNCIAS

ANDERSON, E. L.; HATTIS, D. Foundations: uncertainty and variability. **Risk analysis**, v.19, n.1, p.47-49, 1999.

ANDOW, D. A.; ALSTAD, D.N. F2 Screen for Rare Resistance Alleles. **Journal of Economic Entomology**, v.91, n.3, p.572-578, 1998.

CAPRIO, M.A. Non random mating model, 1998a. Disponível em: <www.msstate.edu/Entomology/PgJava/ILSI model.html> Acesso em: 12 out. 2003.

CAPRIO, M. A. Evaluating resistance management for multiple toxins in presence of external refuges. **Journal of Economic Entomology**, v.91, n.5, p.1021-1031, 1998b.

ILSI/HESI. An evaluation of insect resistance management in Bt field corn: a science based framework for risk assessment and risk management. 1999. Disponível em:<https://research.cip.cgiar.org/confluence/download/attachments/3443/Q41.pdf> Acesso em: 21 set. 2003.

MAIA, A. de H. N. Modelagem da evolução da resistência de pragas a toxinas Bt expresas em culturas transgênicas: quantificação de risco utilizando análise de incertezas. Piracicaba, 2003.



Sensor de umidade do solo por pulsos de calor em ponteira simples utilizando transistor bipolar

José Cláudio da Silva Junior[1], *Maria Bernadete de Morais França*[1]

[1]Departamento de Engenharia Elétrica, Universidade Estadual de Londrina
Londrina, Paraná, Brasil
joseclaudio.silvajr@gmail.com, mbmorais@uel.br

RESUMO

O controle da irrigação é de fundamental importância à produtividade agrícola. Como os recursos hídricos têm se tornado cada vez mais limitados, e, por consequência, onerosos, a implementação de técnicas que auxiliem na medição do nível de umidade do solo se faz necessária. Com o intuito de desenvolver um sensor de umidade do solo confiável, com custo reduzido e de baixo consumo, foi proposto o uso de dispositivos sensores por pulso de calor em ponteira simples utilizando transistores bipolares em blocos de gesso e cimento. Para tanto, foi projetado um circuito de condicionamento juntamente com uma interface microcontrolada capaz de interpretar e transmitir as informações coletadas. Para avaliação do sistema, testes foram realizados com umidades gravimétricas variando de 0 a 45 %, executando ensaios a cada hora. Os resultados obtidos foram analisados por meio de redes neurais *feed-forward* com *backpropagation*, obtendo-se coeficientes de determinação superiores a 0,99 para blocos de cimento e 0,95 para o gesso em comparação aos valores gravimétricos.

PALAVRAS-CHAVE: Conteúdo de água no solo, Produtividade agrícola, Irrigação.

ABSTRACT

The irrigation control is primordial to rural productivity. As hydric resources have been more limited and, therefore expensive, the implementation of techniques that help in the measuring of soil moisture level is required. In order to develop a reliable soil moisture sensor with reduced cost and low power consumption, it was proposed the use of single probe heat pulse sensors using bipolar transistors in blocks of gypsum and cement. Therefore, it was designed a conditioning circuit with a microcontrolled interface capable of interpreting and transmitting the collected data. For the evaluation of the system, tests were performed with gravimetric moisture varying from 0 to 45 %, running tests every hour. The results were analyzed through feed-forward neural networks with backpropagation, obtaining determination coefficients above

0.99 for cement blocks and 0.95 for gypsum compared to gravimetric values.
KEYWORDS: Soil water content, Rural productivity, Irrigation.

INTRODUÇÃO

Para o setor agrícola, a determinação da umidade do solo apresenta um papel fundamental. Como aponta Kirkham (2014), o crescimento de hortaliças é dependente de quatro fatores físicos principais: impedância mecânica, aeração, temperatura e água, sendo a última o agente de maior importância. A concentração de água no solo impacta diretamente na germinação de sementes e na decomposição de matéria orgânica, bem como na transformação de nutrientes nas raízes (BITTELLI, 2011).

Além disso, após a crise hídrica enfrentada entre 2014 e 2015, discussões sobre a gestão consciente deste recurso se intensificaram no país. Entidades governamentais iniciaram debates para adoção de novos instrumentos econômicos, como impostos, subsídios e taxas, com intuito de promover o uso racional da água, como exposto pelo Conselho Nacional de Recursos Hídricos (CNRH) (2016). Tais intervenções incidirão essencialmente no setor agrícola, principal consumidor do insumo.

A Agência Nacional de Águas (2016) estima que, dos 2.275,07 m^3/s de água retirada no Brasil em 2015, 55 $\%$ foram aplicados no setor de irrigação. Quando se é considerado o percentual consumido, ou seja, sem retorno direto ao ambiente, este índice salta para 75 $\%$. Neste contexto, o conhecimento do teor de umidade no solo também é relevante à gestão eficiente de sistemas para irrigação, estando inserido na temática da agricultura de precisão, já que visa a otimização do uso de recursos.

Atualmente, existem diversas técnicas para determinação de umidade do solo, as quais podem ser classificadas em métodos diretos, que determinam a massa de água em dada porção de terra, ou indiretos, os quais estimam a umidade em função de sua correlação com outros fenômenos físicos. O principal método direto é o gravimétrico, nele amostras de solo são retiradas para análise em laboratório, sendo a determinação da porção de água dada pela perda de massa com a secagem em estufa (ISO 17892-1:2014, 2014). Além de custoso, tal processo demanda muito tempo, sendo impraticável em campo.

Dentre as técnicas indiretas, podem ser citados os sensores por pulsos de nêutron, reflectometria, eletromagnetismo, capacitância, resistência e transferência de calor (LEKSHMI; SINGH; BAGHINI, 2014). Como aponta Dias (2012), técnicas envolvendo eletromagnetismo, pulsos de nêutron ou reflectometria tendem a ser mais caras e de pouca aplicação em cultivos, já os sensores resistivos e capacitivos sofrem com resultados errôneos pela variação de temperatura ou condutividade do solo, apresentando pouca confiabilidade e repetibilidade (ARAVIND et al., 2015). Métodos por transferência de calor tendem a ser mais baratos que as técnicas por reflectometria e mais confiáveis que os dispositivos capacitivos, sendo opções mais convenientes para aplicações em campo.

Sensores de umidade por pulsos de calor podem ser organizados basicamente em três categorias: ponteira dupla, *dual probe heat pulse* (DPHP) (KAMAI; KLUITENBERG; HOPMANS, 2015), botão, *button probe heat pulse* (BPHP) (FRANÇA, 2015), e simples, *single probe heat pulse* (SPHP) (LI et al., 2016). Nos dispositivos DPHP, o sensor de temperatura e o elemento aquecedor são dispostos em hastes distintas, tornando-os suscetíveis a desvios pelo desalinhamento das mesmas. OS BPHP visam reduzir este efeito, fixando os dois elementos à uma distância fixa em um mesmo material com boa condutividade térmica. Por fim, nos SPHP, sensor e aquecedor são unidos em um mesmo elemento. Neste método, os pulsos de calor apresentam duração mais longa se comparado aos DPHP, em contrapartida, necessitam de menores potências, viabilizando sua operação em sistemas de baixo consumo.

Sensores SPHP com transistores bipolares já foram empregados por Dias et al. (2013) em blocos de gesso, exibindo sensibilidade superior as técnicas convencionais com termistores. Contudo, apesar do presente trabalho também utilizar o mesmo elemento sensor, o mesmo apresenta a proposta de um circuito de condicionamento distinto. Visando viabilizar a operação de um conjunto com até quatro sensores bipolares por meio de baterias comuns, esta pesquisa tem, como objetivo principal, desenvolver um sistema confiável, com custo reduzido e de baixo consumo para quantificação da umidade presente no solo. Em função da elevada sensibilidade, buscou-se também investigar a resposta destes em blocos cimentícios, opção mais resistente à degradação se comparada ao gesso. Os resultados obtidos são discutidos ao longo deste trabalho, comparando os mesmos ao método gravimétrico.

MATERIAL E MÉTODOS

Elemento sensor

Para o método SPHP, o valor da condutividade térmica (λ) é comumente estimada com base no modelo definido por Blackwell (1954), no qual o sensor é considerado uma fonte de calor cilíndrica infinita colocada num meio homogêneo em equilíbrio térmico. Uma aproximação obtida por esta modelagem à correlação entre a variação de temperatura e a condutividade térmica em pulsos de longa duração é exposta na Equação 1.

$$\Delta T(t) \; = \; \frac{Q}{4\pi\lambda} ln(t) + C \tag{1}$$

Nela, $\Delta T(t)$ indica a variação de temperatura, Q a potência dissipada ($W \cdot m^{-1}$) e t a duração do pulso (s). C é uma constante para ajuste da curva.

Para o transistor bipolar, elemento sensor empregado neste trabalho, uma forma clássica de se mensurar a temperatura de operação do mesmo é forçar correntes conhecidas e medir a variação de tensão entre base e emissor (V_{BE}), como exposto na Equação 2.

$$\Delta V_{BE} \; = \; \frac{nkT}{q} ln \left(\frac{I_{C1}}{I_{C2}} \right) \tag{2}$$

Onde n é o fator de idealidade do transistor, k a constante de Boltzmann, q a carga do elétron e T a temperatura (K). I_{C1} e I_{C2} são as correntes aplicadas durante a medição. Caso haja uma diferença de potência muito elevada durante os pulsos ao longo de um certo período, o sensor sofrerá com o auto-aquecimento, caracterizando-o como um SPHP. Para uma potência constante, a variação de temperatura em função do auto-aquecimento é dada pela Equação 3.

$$V_{BE}(T) \; = \; V_{G0} \cdot \left(1 - \frac{T}{T_0}\right) \; + \; V_{BE0} \cdot \left(\frac{T}{T_0}\right) \; + \; \left(\frac{\gamma k T}{q}\right) ln\left(\frac{T_0}{T}\right) \tag{3}$$

Sendo V_{G0} a tensão de *bandgap* (1,2 V) e γ um parâmetro de fabricação (3,2). T_0 e V_{BE0} são valores de referência para temperatura e respectivo V_{BE} em 25 ^oC.

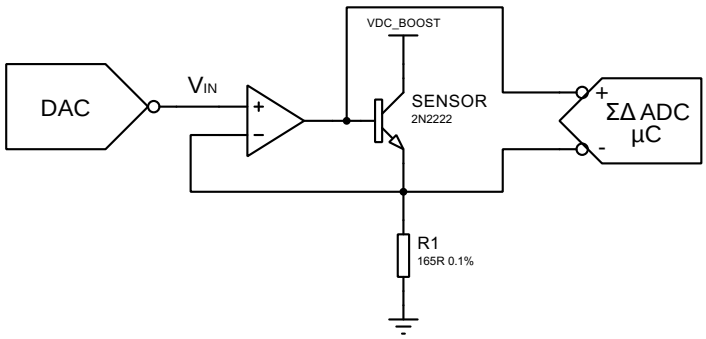
Apesar da condutividade térmica ser obtida com certa facilidade por meio deste tipo de sensor, a correlação com a umidade do solo não é trivial, sendo muito dependente das características físicas de cada solo. Para reduzir a influência de tais efeitos, os sensores por pulso de calor normalmente são postos em blocos de gesso ou epóxi.

Neste trabalho, foram produzidos dois pares de blocos em gesso e cimento com uma polegada de diâmetro e cinco centímetros de comprimento, nos quais foram inseridos transistores bipolares NPN modelo 2N2222 com encapsulamento metálico TO-18. Serão apresentados, no entanto, os resultados obtidos com apenas um par de sensores.

Condicionamento do sinal

Sensores de umidade por pulso de calor necessitam que uma fonte constante de energia seja aplicada em um período de tempo conhecido. Tal característica pode ser obtida assegurando dissipação fixa de potência por meio de um circuito capaz de controlar o acionamento do pulso, mantendo tensão e corrente constantes. Desta forma, foi proposto neste trabalho o condicionamento exposto na Figura 1.

Figura 1: Diagrama do circuito de condicionamento para um canal



Fonte (Autor, 2017)

O circuito conta com o conversor *boost* TPS61041, da *Texas Instruments*, o qual é capaz de gerar, com baixo *ripple*, tensões de até 28 V advindas de baterias do tipo lítio ou níquel cádmio (NiCd). O mesmo foi configurado para aplicar um pulso de 25 V nos sensores, os quais

podem ser empregados em um circuito como fonte de corrente linear. Nesta configuração, a tensão sobre o resistor R_1 controla a corrente sobre o transistor, como exposto na Equação 4:

$$I_{sensor} = \frac{V_{IN}}{R_1} \tag{4}$$

Assim, a potência dissipada no sensor pode ser aproximada pela Equação 5:

$$P_{sensor} \approx I_{sensor} \cdot (V_{DC\_Boost} - V_{IN}) \tag{5}$$

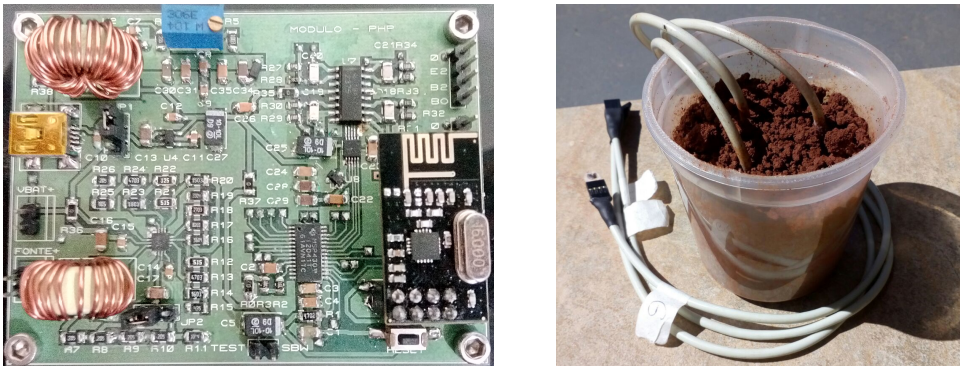
Para assegurar um melhor rendimento do sistema, uma resistência de baixo valor ôhmico, 165 Ω, com 0,1 $\%$ de tolerância é utilizada. O conversor digital-analógico (DAC) DAC084S085, de quatro canais e 8 bits, é empregado no controle dos pulsos de calor pelo fornecimento da tensão de referência V_{IN}. Assim, foram definidas duas potências de trabalho, 1 mW para *stand-by* e 80 mW em operação.

Como há o interesse no uso de transistores como elementos sensores e aquecedores, há a necessidade de se mensurar V_{BE} durante o processo. Assim, uma medição diferencial de tensão se fez necessária. Optou-se por utilizar o microcontrolador MSP430I2041, da *Texas Instruments*, o qual conta com quatro conversores analógico-digitais (ADC) diferenciais sigma-delta de 24 bits com ganho programável. Esta unidade lógica também é responsável por comandar o acionamento dos pulsos de calor e transmitir os dados adquiridos para o computador.

Para testes, uma placa de circuito impresso foi confeccionada, Figura 2a. Ao todo, o sistema apresentou um custo aproximado de U$ 20. Para cada sensor, foi definida a aplicação de um pulso de 20 s, sendo adquiridas 10 amostras por segundo. Além disso, foram coletados os dois segundos anteriores e posteriores à aplicação do pulso.

Os sensores foram posicionados em uma mesma amostra de solo, Figura 2b, a qual teve sua massa de água modificada em intervalos de 60 minutos, tempo utilizado para cada medição. Tal período foi definido para viabilizar uma distribuição homogênea de água na amostra. Os resultados obtidos foram então comparados ao método gravimétrico.

Figura 2: Sistema desenvolvido para testes



(a) Placa projetada (b) Três sensores em amostra de solo

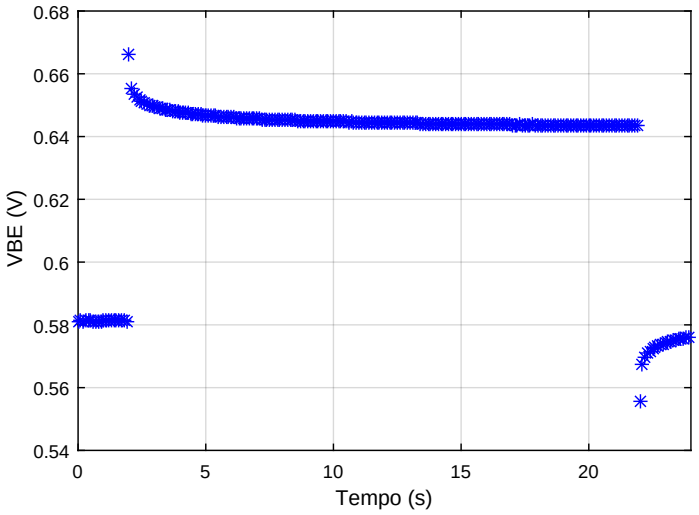
Fonte (Autor, 2017)

RESULTADOS E DISCUSSÃO

Foram realizadas cinco baterias de testes em dias distintos. Cada bateria foi caracterizada por um acréscimo de 5 % de massa d'água por hora na amostra de solo, variando a umidade entre 0 % e 45 %, realizando sua posterior secagem ao final de cada bateria.

Como há a necessidade de se caracterizar os sensores desenvolvidos, o registro completo de V_{BE} é realizado, como ilustra a Figura 3.

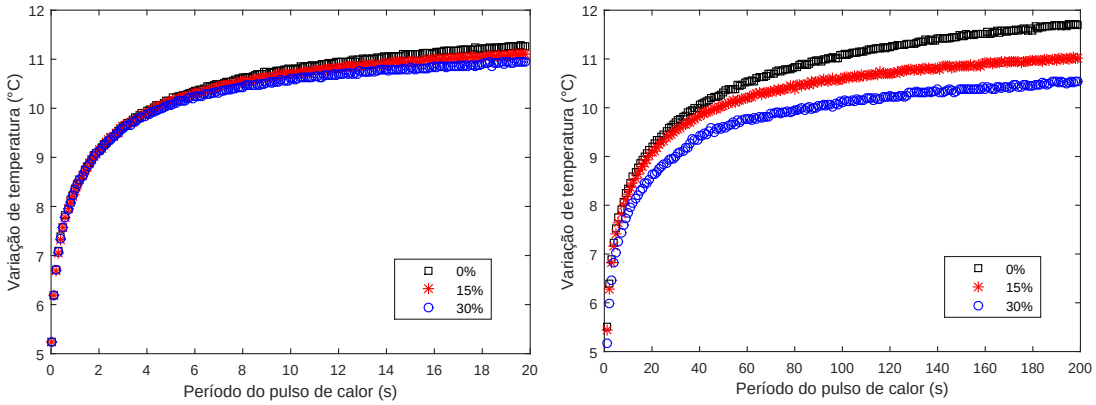
Figura 3: Variação de V_{BE} em função da duração do pulso de calor



Fonte (Autor, 2017)

Com base na Equação 3, as curvas com o registro de V_{BE} foram convertidas em variação de temperatura. A Figura 4 exibe a resposta obtida em um ensaio com os blocos de cimento e gesso para concentrações gravimétricas de 0 %, 15 % e 30 %. Deste modo, é possível notar que o sistema com gesso apresenta maior sensibilidade à variação de umidade.

Figura 4: Variação de temperatura em função da duração do pulso de calor



(a) Sensor no bloco de cimento

(b) Sensor no bloco de gesso

Fonte (Autor, 2017)

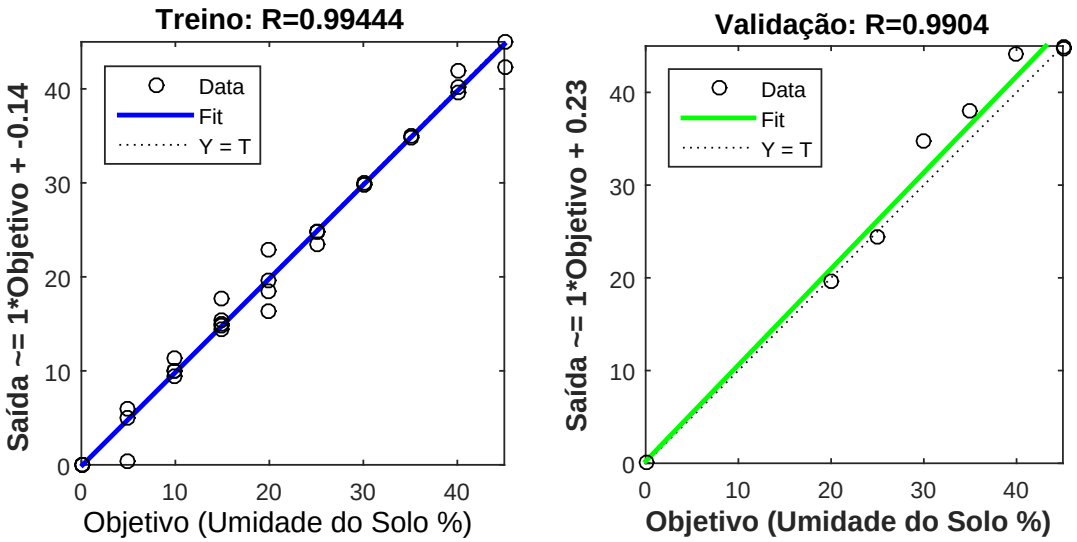
Para o bloco de gesso e uma dissipação de 1,5 J, foi registrada uma variação de temperatura de 1,8 ^{o}C entre o solo seco e com 30 % de massa d'água adicionada. Este *range* é inferior aos 3 ^{o}C obtidos por Dias et al. (2013) com pulsos de 150 mW em 10 s. No entanto, esta sensibilidade é superior aos trabalhos de Kamai, Kluitenberg e Hopmans (2015), 0,8 ^{o}C, e Campbell et al. (1994), 1,2 ^{o}C, para uma mesma energia.

Pelos ensaios realizados, foram registrados elevados desvios entre medições de variação de temperatura para um mesmo registro de umidade gravimétrica. Como os testes foram desenvolvidos ao longo de dias distintos, sem um devido controle da temperatura ambiente, testou-se a hipótese da mesma influenciar, de forma significativa, a resposta do sistema desenvolvido.

Para avaliar esta possibilidade, redes neurais de arquitetura *feed-forward* com *backpropagation* foram treinadas tendo como entradas a temperatura inicial medida pelo transistor e a sua respectiva variação em função do pulso de calor. Como saída, as redes buscam estimar a concentração gravimétrica de água no solo, função objetivo.

A Figura 5 exibe o resultado do treino da rede neural para o bloco cimentício. As previsões da rede para o treino e validação mostram um coeficiente de correlação (r) superior a 0,99, resultando em um respectivo coeficiente de determinação (r^2) de 0,9889 e 0,9809, o que significa que mais de 98% da variação na saída da rede pode ser definida por sua relação com a concentração real pelas regressões lineares apresentadas. Em termos absolutos, a predição exibiu até 15 % de desvio em relação à umidade gravimétrica.

Figura 5: Rede *feed-forward* treinada para o bloco de cimento



(a) Dados para treino da rede

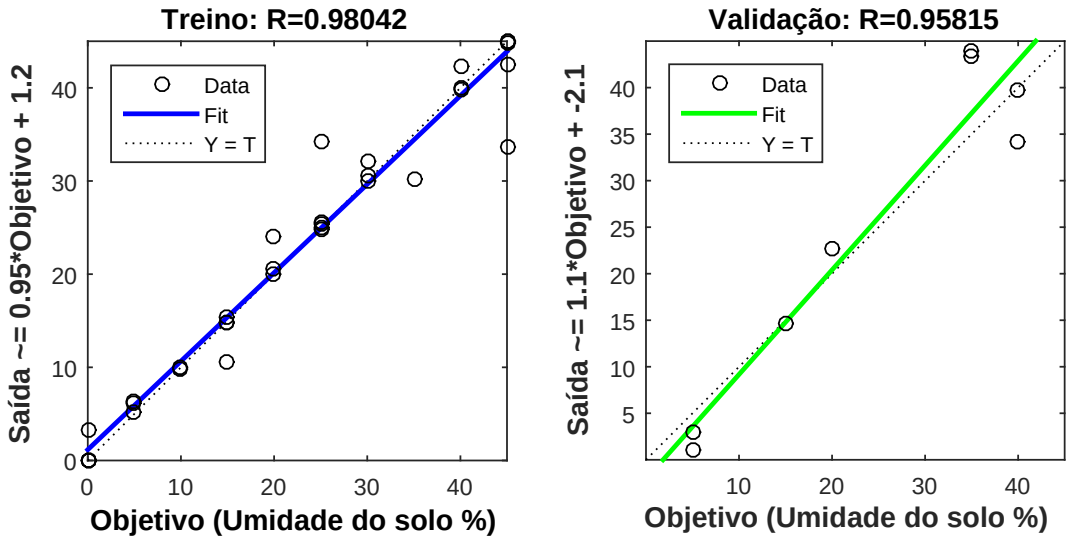
(b) Dados para validação da rede

Fonte (Autor, 2017)

A Figura 6 expõe o resultado do treino da rede neural para o bloco de gesso. As previsões da rede exibem r superior a 0,958, resultando em um respectivo coeficiente de determinação (r^2) de 0,9612 e 0,9180, ou seja, as equações lineares explicam mais de 91% da variância dos dados na saída da rede. É notório que os maiores desvios ocorreram para predição de umidades

elevadas, superiores a 30 %, nestas concentrações os sensores tenderam à saturação.

Figura 6: Rede *feed-forward* treinada para o bloco de gesso



(a) Dados para treino da rede (b) Dados para validação da rede

Fonte (Autor, 2017)

CONCLUSÕES

O sensor proposto utiliza a aplicação de pulsos de calor em ponteira simples com transistores bipolares para determinação da umidade contida no solo. Com base nos resultados obtidos, verificou-se que ambas as estruturas implementadas para o sensor bipolar, de cimento e gesso, apresentam elevada correlação com a umidade do solo observadas pelo método gravimétrico. No entanto, tais registros só foram obtidos pelo treinamento de redes neurais *feed-forward*, as quais ponderaram a influência da temperatura ambiente sobre as medições realizadas. Tal dependência já foi descrita para sensores por pulsos de calor com termistores (CAMPBELL et al., 1994), no entanto, esta característica deve ser melhor investigada para o sistema proposto ao longo do processo de pesquisa.

Assim, a continuidade dos ensaios se faz necessária para a determinação de uma modelagem matemática adequada à resposta do sistema em desenvolvimento. Assim, os sensores deverão ser aplicados em diferentes tipos de solos e testados em condições diversas de temperatura e pressão. Testes em campo também estão previstos para avaliação do dispositivo em condições normais de uso, neste ponto, o uso de baterias comuns será facilitado pelo emprego do circuito de condicionamento discutido no trabalho.

AGRADECIMENTOS

Os autores gostariam de agradecer o fomento concedido pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Ministério da Educação (Capes - MEC) ao desenvolvimento da pesquisa.

REFERÊNCIAS

Agência Nacional de Águas. *Conjuntura dos recursos hídricos: Informe 2016*. [S.l.: s.n.], 2016.

ARAVIND, P. et al. A wireless multi-sensor system for soil moisture measurement. IEEE Sensors, 2015.

BITTELLI, M. *Measuring Soil Water Content: A Review*. [S.l.]: HortTechnology, 2011.

BLACKWELL, J. *A Transient-Flow Method for Determination of Thermal Constants of Insulating Materials in Bulk Part I—Theory*. [S.l.]: Journal of Applied Physics, 1954.

CAMPBELL, G. S. et al. Predicting the effect of temperature on soil thermal conductivity. *Soil Science*, Lippincott Williams and Wilkins, v. 158, 11 1994.

Conselho Nacional de Recursos Hídricos (CNRH). Resolução 181 - prioridades, ações e metas do plano nacional de recursos hídricos para 2016-2020. p. 15, 2016.

DIAS, P. C. Um novo sensor de umidade de solo de pulso de calor de alta sensibilidade, baseado em um Único transistor bipolar de junção npn. 2012.

DIAS, P. C. et al. *A high sensitivity single-probe heat pulse soil moisture sensor based on a single npn junction transistor*. [S.l.]: Computers and Electronics in Agriculture - Elsevier, 2013. 139-147 p.

FRANÇA, M. B. d. M. *Contribuição ao estudo e fabricação de sensores de umidade do solo usando a técnica da transferência de calor*. [S.l.]: Universidade Estadual de Campinas, 2015.

ISO 17892-1:2014. *Geotechnical investigation and testing – Laboratory testing of soil – Part 1: Determination of water content*. [S.l.]: International Organization for Standardization, 2014.

KAMAI, T.; KLUITENBERG, G. J.; HOPMANS, J. W. *A Dual-Probe Heat-Pulse Sensor with Rigid Probes for Improved Soil Water Content Measurement*. [S.l.]: Soil Physics and Hydrology - Soil Science Society of America Journal, 2015. 14 p.

KIRKHAM, M. B. *Principles of Soil and Plant Water Relations*. 2. ed. [S.l.]: Elsevier, 2014.

LEKSHMI, S. S.; SINGH, D.; BAGHINI, M. S. *A critical review of soil moisture measurement*. [S.l.]: Measurement, vol. 54, no. 0, pp. 92– 105, Elsevier, 2014.

LI, M. et al. Single-probe heat pulse method for soil water content determination: Comparison of methods. Vadose Zone Journal, p. 13, 7 2016.

Desenvolvimento de um perfilômetro laser para a geração de modelos 3D aplicados à Engenharia Agrícola

Antônio Carlos Loureiro Lino[1], *Marcos Valério Gebra da Silva*[2], *Edson d'Avila*[3], *Denival da Costa Silva*[1]

[1] Centro de Engenharia e automação, Instituto Agronômico de Campinas, Jundiaí, São Paulo, Brasil, lino@iac.sp.gov.br, costasilvadenival@gmail.com

[2] Faculdade de Jaguariúna, Jaguariúna, São Paulo, Brasil, marcosgebra@terra.com.br

[3]Instituto Federal de São Paulo, São Paulo, Brasil, davila@ifsp.edu.br,

RESUMO

O objetivo deste trabalho foi desenvolver uma metodologia e construir um protótipo de um scanner tridimensional baseado em triangulação laser. A perfilometria a laser, baseada na triangulação, é um método para a medição de profundidade de pontos no espaço em que se utilizam a um feixe de laser para sondar o objeto. A imagem de uma linha laser projetada sobre uma superfície de um objeto é capturada e tratada em software de análise de imagens, que calcula o seu perfil. A união de uma sequência de perfis permite construir os modelos digitais 3D de objetos. Foi construído um protótipo constituído de uma estrutura que possui um eixo sem-fim, acionado por um motor elétrico de velocidade variável. Este eixo ao girar desloca um carro com um diodo que produz uma linha laser na superfície do objeto e uma câmera que filma o objeto, com um angulo conhecido que contém as informações para gerar o modelo 3d do objeto. Foram feitos testes em objetos regulares calotas, pirâmides e paralelepípedos e os resultados mostraram que os modelos 3D produzidos tinham as mesmas dimensões. Também foram feitos testes em objetos irregulares como superfície do solo e frutas. O protótipo mostrou-se robusto para gerar modelos 3D de objetos.

PALAVRAS-CHAVE: Processamento de imagem, triangulação laser, modelos tridimensionais.

ABSTRACT

The laser profilometry based on triangulation is a method for measuring depth of points in space that use a laser beam to scan the object. The image of a laser line projected onto a surface of an object is captured and treated in image analysis software, which calculates your profile. The union of a sequence of profiles allows the construction of the digital 3D models of objects. one consisting of a prototype was built structure which has a worm shaft driven by an electric motor of variable speed. This axis spin move a car with a diode that produces a laser line on the surface of the object and a camera filming the object, with an angle of 30 °, as the car moves. It developed a routine that breaks the film into frames, transforms each in binary images being black regarding the background and the white line laser, then creates a vector corresponding to the coordinates of this line. The following corrects the pixel dimensions to millimeter and deformations due to the angle of observation. The union of these vectors forms a matrix that contains the information to generate the 3D model of the object. Tests were made at regular caps objects, pyramids and cobblestones and the results showed that the produced 3D models had the same dimensions. Tests were also made on uneven objects such as the ground surface and fruits. The prototype was robust to generate 3D models of objects.

KEYWORDS: Image processing, laser triangulation, three-dimensional models.

INTRODUÇÃO

Técnicas perfilométricas ópticas podem ser utilizadas para o estudo em diversas área da engenharia agrícola tais como: comportamento mecânico, analise de forma e volume de materiais biológicos, rugosidade do solo entre outros.

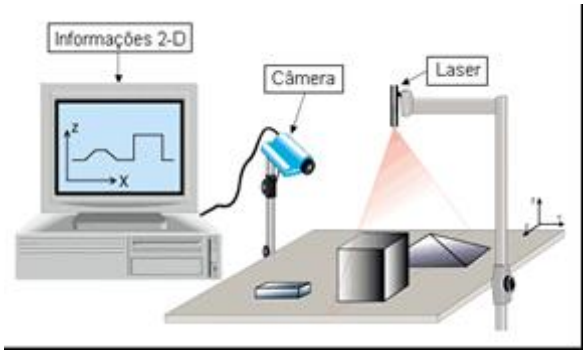
Dentre as varias técnicas a triangulação é provavelmente o método mais antigo para a medição de profundidade de pontos no espaço, sendo provavelmente uma das técnicas mais comuns (FERNANDES, 2007). Normalmente se utiliza um feixe de laser para sondar o objeto. Ela ilumina pontos na superfície do objeto e através de uma câmera colocada a um determinado ângulo localiza este ponto e posteriormente através de processada a informação é possível determinar a distância dos pontos iluminados. Dependendo da distância que a superfície do objeto se encontra, o ponto aparece em diferentes lugares na imagem. Essa técnica é denominada triangulação porque o emissor de laser, a câmera e o ponto gerado pelo laser na superfície do objeto formam um triangulo.

STORBECKA e DAANB (2001) utilizaram a técnica de triangulação para reconhecer a espécie de peixes através do volume. O sistema conseguiu classificar de maneira satisfatória

aproximadamente de 98% dos peixes. Já NOORDAM (2007) cita que a empresa GREEN VISION desenvolveu um sistema de alta velocidade para seleção de pimentão baseado na análise de forma tridimensional também por triangulação laser.

Segundo CLARK (2000), o princípio de funcionamento do sensor óptico tipo “Folha de Luz” consiste na projeção de um plano de luz (laser) na superfície a ser medida. A interseção desse plano com a superfície define uma linha que é então capturada por um sistema de visualização posicionado, conforme a técnica de triangulação, a uma distância e inclinação conhecidas em relação ao plano de luz. Analisando a imagem obtida, é possível determinar as coordenadas de pontos pertencentes ao objeto estudado em um sistema de coordenadas conhecido. Um desenho esquemático é apresentado na Figura 1.

Figura 1 – Figura 1. Esquema de Funcionamento do Sensor Óptico tipo “Folha de Luz”.



Fonte: CLARK (2000)

O objetivo deste trabalho foi desenvolver uma metodologia e construir um prototipo de um scanner tridimensional baseado em triangulação laser.

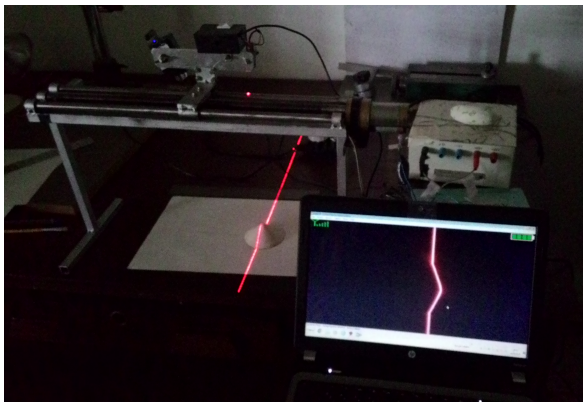
MATERIAL E MÉTODOS

O trabalho foi realizado no Laboratório de imagem do Centro de Engenharia e Automação do Instituto Agronômico Campinas (IAC), localizado em Jundiaí, SP. O arranjo experimental será constituído por: Microcomputador com processador modelo Core i5-2410M, com 3GB de memória RAM, Câmera wireless camera 2.4G RF, acoplada ao computador através de uma porta USB e diodo laser em linha com 5mW e 650ηm.

Foi construído um protótipo (FIGURA 2) constituído de um pórtico que possui um eixo sem-fim, acionado por um motor elétrico de velocidade variável. Este eixo ao girar desloca um carro com o diodo que produz uma linha laser na superfície do objeto e uma câmera que filma o objeto, com um angulo de 18o, à medida de o carro se desloca.

Para a captura e tratamento das imagens serão utilizados o software ImageJ. Desenvolveu-se uma rotina que quebra o filme em quadros, transforma cada um em tons de cinza, onde a região mais clara corresponde a linha laser, que contém as informações sobre a topografia do objeto em estudo. A seguir corrige as deformações devido ao angulo de observação e cria uma nova imagem que contém as informações topográficas do objeto.

Figura 2 – Setup do scanner laser.



Para testar o escâner 3d foram utilizados um cone de com 74,15 mm de diâmetro por 26,35mm de altura, uma modelo de de um fruto de banana confeccionado em cera e um recipiente com a superfície coberta de terra na qual foi escrito a palavra “IAC” em relevo.

RESULTADOS E DISCUSSÃO

A FIGURA 3 mostra a palavra “IAC” escrita em baixo relevo na superfície da de terra, enquanto que, a FIGURA 4 mostra a mesma superfície digitalizada, com o relevo representado em tons de cinza, onde as regiões mais escuras possuem cotas mais baixas e as mais claras as de cotas maiores. Já FIGURA 5 mostra a mesma superfície em falsa cor para ressaltar o relevo e em *wireframe*.

Figura 3 – Palavra IAC escrita na terra em baixo relevo.



Figura 4 – Superficie digitalizada em tons de cinza.

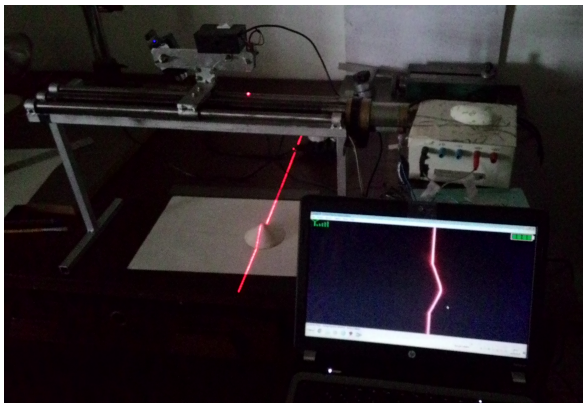
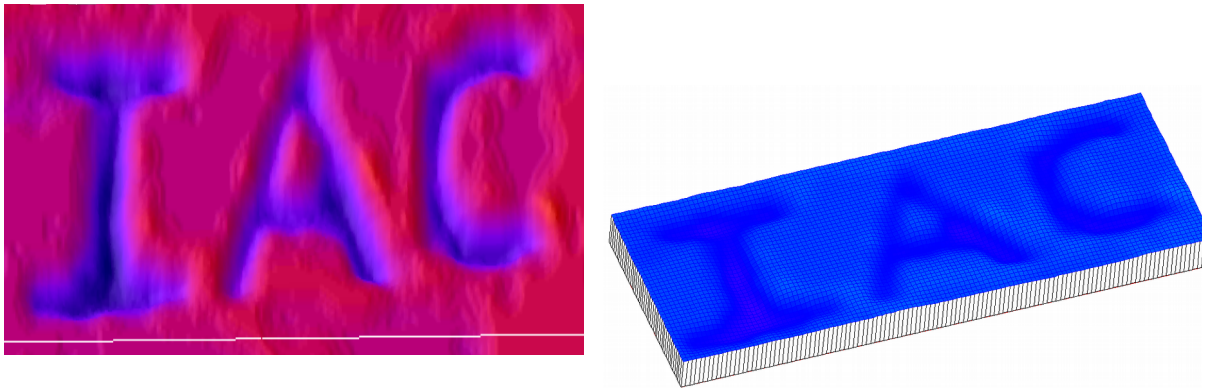
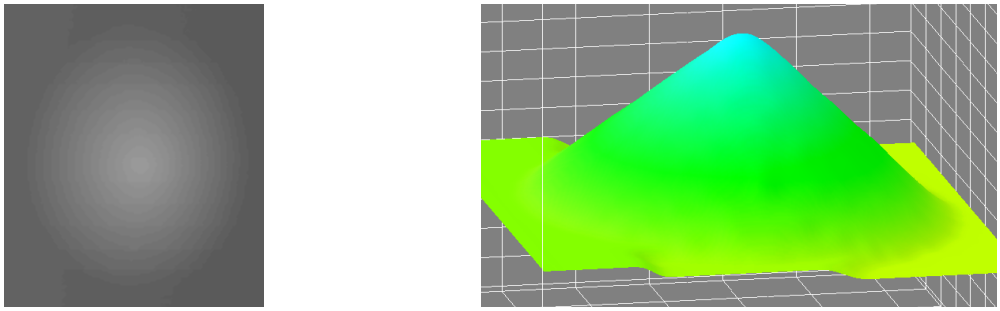


Figura 5 – Representação da superfície da terra representada em cor falsa e em *wireframe.*



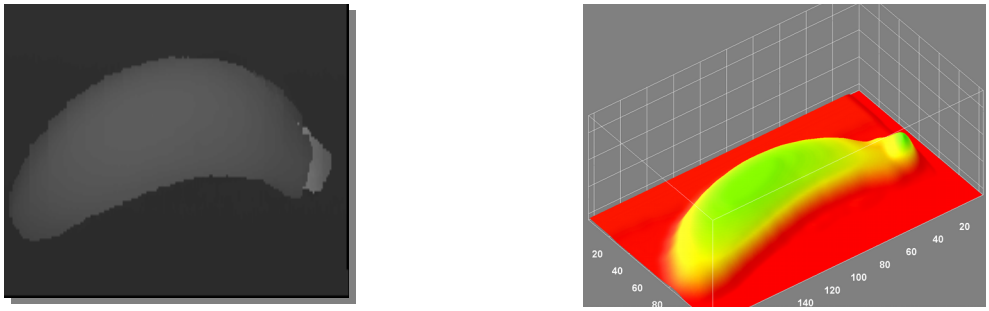
A FIGURA 6 mostra a digitalização tridimensional do cone, com o relevo representado em tons de cinza, onde as regiões mais escuras possuem cotas mais baixas e as mais claras as de cotas maiores (FIGURA 6A) e o mesmo objeto em falsa cor, evidenciando o seu relevo (FIGURA 6B). e a FIGURA 7 mostra os contornos topográficos do modelo da banana em tons de cinza, (FIGURA 7A) e o mesmo objeto em falsa cor, evidenciando o seu relevo (FIGURA 7B).

Figura 6 – Representação da topografia do Cone, em tons de cinza (A) e em cor falsa (B).



(A) (B)

Figura 7 – Representação da topografia do modelo da banana, em tons de cinza (A) e em cor falsa (B).



(A) (B)

CONCLUSÕES

O escâner 3d baseado em triangulação laser, foi capaz de digitalizar os contornos topográficos de vários tipos de objetos, entre eles superfície do solo, objetos com formas regulares e objetos com formas irregulares.

REFERÊNCIAS

FERNANDES; L. A. F.. Estudo de Métodos para Extração de Formas e Realização de Medidas a Partir de Imagens.
<http://www.inf.ufrgs.br/~laffernandes/fernandes_TI_1159.pdf>. 10 de abril de 2011.

CLARK, J. - Implementing Non-Contact Digitisation Techniques Within the Mechanical Design Process, Sensor Review, v. 20, No. 3, p. 195-201, 2000.

NOORDAM, J.C. Innovative Applications in the Agro and Food industry. http://greenvision.wur.nl/. 20 de maio de 2007.

STORBECKA, Frank; DAANB, Berent Fish species recognition using computer vision and a neural network. Fisheries Research, v. 51, p. 11-15, 2001.

Desenvolvimento de um instrumento medidor de área foliar

Edson d'Avila[1], Antônio Carlos Loureiro Lino[2], Inácio Dal Fabbro[3], Kirlian Kitzinger d'Avila[1]

[1] Instituto Federal de São Paulo, São Paulo, Brasil, davila@ifsp.edu.br, kirliandavila@hotmail.com

[2] Centro de Engenharia e automação - CEA do Instituto Agronômico de Campinas - IAC, Jundiaí, São Paulo, Brasil, lino@iac.sp.gov.br

[3] UNICAMP, Campinas, SP, inacio@feagri.unicamp.br

RESUMO

Este trabalho apresenta o atual estágio do desenvolvimento de um instrumento para medir área foliar. O protótipo é composto basicamente de duas esteiras de transporte que acondicionam, transportam e pressionam a superfície da folha; um sistema de captura de imagens; um negatoscópio para iluminação de contorno e dois pedais para controle de avanço da folha e disparo da câmera fotográfica. O protótipo demonstrou-se de fácil calibragem e utilização. O sistema fotográfico proporcionou precisão para os cálculos do *ImageJ* em toda a extensão da janela de captura de imagens. O disparo da câmera pelo pedal se demonstrou útil porque permitiu ao operador do instrumento manter as duas mãos livres. O avanço e parada do transportador pelo segundo pedal também traz vantagens porque sincronizou a velocidade de operação do instrumento com a dinâmica do operador. O controle de direção da esteira facilitou a operação para pessoas que usam preferencialmente a mão esquerda. Na próxima etapa do desenvolvimento do protótipo a instrumento fornecerá, ao usuário, informações de comprimento, largura, entre outras.

PALAVRAS-CHAVE: Processamento digital de imagem, Macro para *ImageJ*, Área foliar processo fotográfico, *ImageJ*.

ABSTRACT

The present work showcases the state of development of an instrument to measure leaf area. The prototype consists, basically, of two transport belts which house, transport and pressurize the leaf's surface; an image capture system; outline illumination and two pedals which control the leaf's advance and photographical camera activation. The prototype has shown to be easy to calibrate and operate. The photographical system allowed for precise calculations in ImageJ including the whole extension of the pictures taken. The camera activation by the pedal has shown to be useful because it allowed the instrument operator to have both hands free. The advances and stops of the belts controlled by the second pedal also displays advantages because it allowed sinchronized operation of the instrument with the operator's dynamic. The direction control of the belts made for an easier operation for left-handed staff. In the next stage of development of the prototype, the instrument will relay to the user information of length, width among others.

KEYWORDS: Digital image processing, Macro to ImageJ, Leaf area photographic process, ImageJ.

INTRODUÇÃO

Medir a área de folhas permite a obtenção de indicadores para a compreensão das respostas da planta a fatores ambientais, como por exemplo, crescimento, a adaptação ecológica, a competição com outras espécies e a produtividade de um ecossistema vegetal. O crescimento de uma planta pode ser observado pelo aumento do volume e peso. Para medir o crescimento podemos, por exemplo, medir a altura ou diâmetro do caule ou a área de uma folha (SALISBURY e ROSS, 2012).

Considerando a folha como principal agente da planta a realizar a fotossíntese e transpiração, avaliar e acompanhar suas dimensões é de relevância para produtores e estudiosos da área. As folhas apresentam grande polimorfismo e o cálculo de área de formas irregulares, com precisão e rapidez, não é uma tarefa das mais fáceis (MARACAJÁ et. al., 2008).

Entre os métodos utilizados para a medição da área foliar estão os medidores eletrônicos e as equações matemáticas para estimativa da área foliar a partir do comprimento

ao longo da nervura principal (Campostrini e Yamanishi, 2001), ou da largura máxima (Queiroga et al. 2003) ou ainda a relação entre largura e comprimento.

Este estudo visa desenvolver um instrumento eletrônico capaz de medir com precisão folhas com qualquer forma e tamanho pela retirada da folha da planta e submetida ao instrumento.

MATERIAL E MÉTODOS

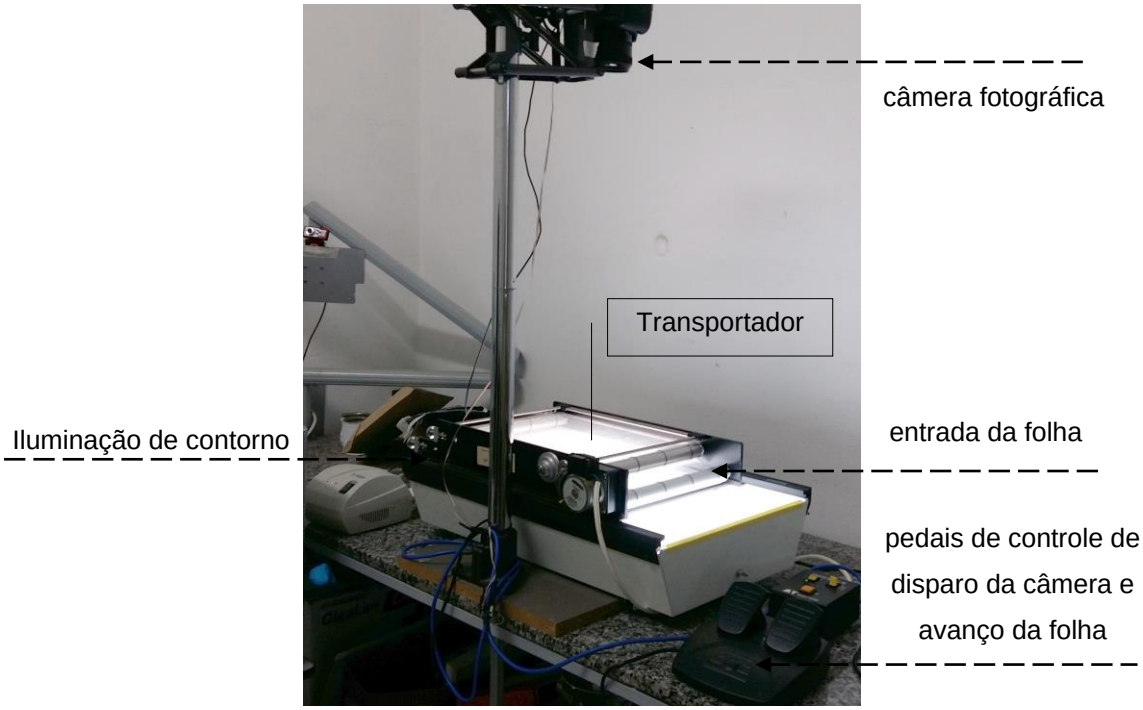
O protótipo de um medidor de área foliar foi desenvolvido no Laboratório de Imagem do Centro de Engenharia e Automação - CEA do Instituto Agronômico Campinas - IAC, localizado em Jundiaí, SP.

O arranjo experimental esquematizado na Figura 1 apresenta a disposição dos elementos básicos de transporte e acomodação da folha; de captura de imagens, de iluminação de contorno com LED 6500K e dois pedais para controle de avanço da folha e disparo da câmera fotográfica.

O sistema de captura de imagens é composto por uma câmera Canon T5 com sensor de 18,0 Megapixels, CMOS (APS-C - fator de crop: 1.6), processador de imagem DIGIC 4 e lente Canon EF 50 mm f/1.2L USM.

O sistema de transporte é composto por duas esteiras que acondicionam, transportam e pressionam a superfície da folha para se obter a máxima precisão na medida da área foliar. Um motor de corrente contínua é responsável pelo acionamento de quatro roletes e pela direção do transporte, da esquerda para a direita ou da direita para a esquerda.

Figura 1 – Instrumento para medir área foliar



A câmera é acoplada ao pedal de disparo e a um computador pessoal pela porta USB. O EOS Utility é um programa de operação remota da Canon que permite ajustar todos os parâmetros da câmera para uma captura de imagem digital com nitidez, contraste, saturação e matiz que irá minimizar a pós-produção da imagem. Trata-se de um software para estabelecer uma conexão, com fio, entre a câmera e um computador. A interface amigável na tela do PC, oferece, basicamente, os mesmos recursos que estão disponíveis no visor da câmera, figura 2.

Figura 2 – Interface de controle da câmera fotográfica



A macro "calcarf" desenvolvida para o programa *ImageJ* calcula da área da folha a partir das imagens capturadas pelo programa EOS Utility. Através de ferramentas nativas do *ImageJ* as imagens são analisadas. Ferramentas de segmentação destacam a folha do fundo, "*Set Scale*" converte medidas em pixel para milímetros e "*Mesure*" calcula a área, Tabela 1.

Tabela1 – Trecho da macro "calcarf" para cálculo da área foliar

selectWindow("IMG_0028.JPG");
run("Split Channels");
selectWindow("IMG_0028.JPG (blue)");
setAutoThreshold("Default");
run("Set Scale...", "distance=4976.023 known=290 unit=mm global");
run("Measure");

RESULTADOS E DISCUSSÃO

Para verificar a precisão do protótipo, um papel milimetrado foi fotografado e dez áreas foram delimitadas e, em seguida, calculadas pela macro "calcarf". A tabela 2 apresenta a comparação entre o resultado da "calcarf" e os valores medidos sobre o papel milimetrado.

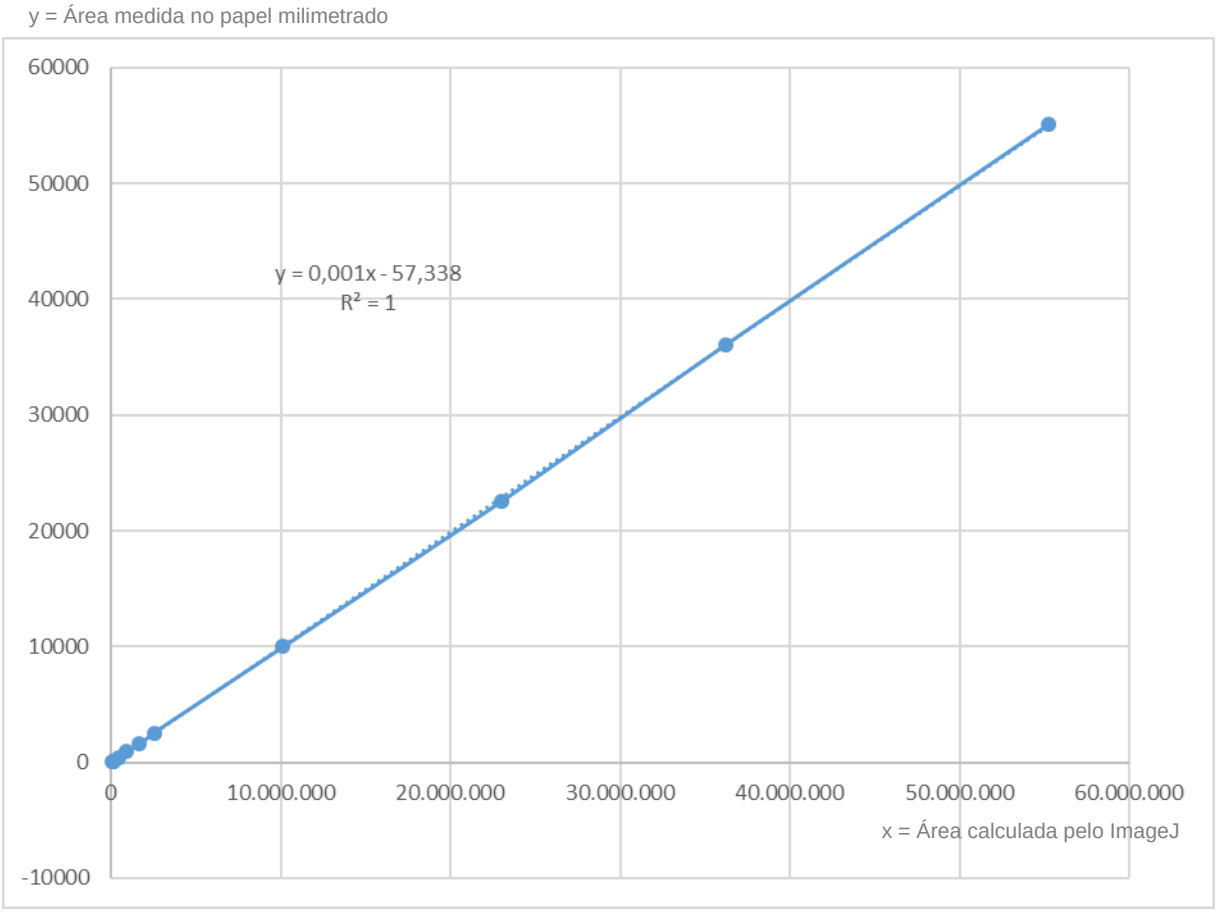
Tabela 2 – Resultados comparativos entre a área calculada e a medida

Experimento	Área calculada pelo *ImageJ* mm2	Medida no papel milimetrado mm2	dimensões da figura	Desvio %
1	25,413	25	5 x 5	1,65
2	102,829	100	10x10	2,83
3	414,884	400	20x20	3,72
4	923,681	900	30x30	2,63
5	1649,970	1600	40x40	3,12
6	2547,110	2500	50x50	1,88
7	10118,118	10000	100x100	1,18
8	22995,396	22500	150x150	2,20
9	36229,063	36100	190x190	0,36
10	55207,582	55100	190x290	0,20

O gráfico 1 apresenta a correlação entre a área calculada pela macro do *ImageJ* e a área medida sobre o papel milimetrado. O valor de R^2, varia entre 0 e 1, neste caso igual a 1, logo a linha de tendência linear é confiável e indica que o instrumento consegue medir as

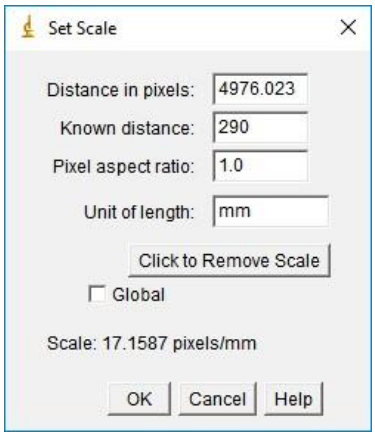
áreas propostas com precisão. Quanto maior o R^2, melhor expressa a capacidade do instrumento de medir a área da amostra.

Gráfico 1 – Linha de tendência e R^2



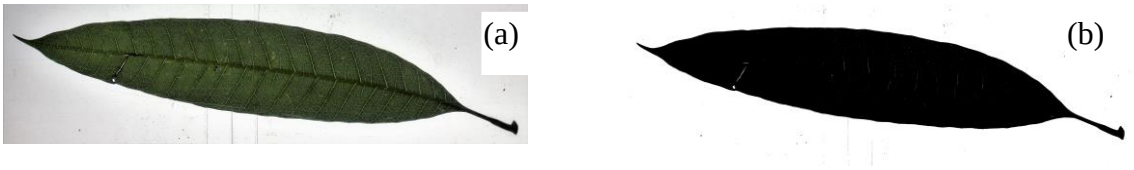
Na sequência a macro "calcarf" calculou a área de uma folha de manga em 7529,253 mm^2, com a calibração do "*SetScale*" padrão obtido com o papel milimetrado, figura 3.

Figura 3 – Caixa de diálogo de configuração do SetScale do ImageJ



A figura 4 (a) e (b) apresenta a imagem digital capturada e após o "*threshold*" que transformou a imagem em branco e preto, ou seja, em binário 0 e 1.

Figura 4 – Imagem digital (a) e após tratamento representada em binário (b)



CONCLUSÕES

O protótipo desenvolvido para medir área foliar demonstrou-se de fácil calibragem e utilização. O sistema fotográfico composto pela câmera T5 e a lente 50 mm F/1.2L da Canon proporcionou precisão para os cálculos da macro do *ImageJ* em toda a extensão da janela de captura de imagens. O disparo da câmera pelo pedal também se demonstrou útil porque permitiu ao operador do instrumento medidor de área foliar manter as duas mãos livres. O pedal de avanço e de parada do transportador demonstrou-se, igualmente, útil porque sincronizou a velocidade de operação do instrumento com a dinâmica do operador. O controle de direção da esteira facilitou a operação para pessoas que usam preferencialmente a mão esquerda.

Na próxima etapa do desenvolvimento do protótipo, usuários deste tipo de equipamento serão entrevistados e serão convidados a operarem o protótipo para que melhorias sejam acrescentadas. O *ImageJ* possui ainda outras ferramentas que poderão ser disponibilizadas aos usuários com pequenas alterações na macro. O comprimento máximo e a largura máxima da folha, por exemplo, podem ser apresentadas rapidamente por ferramenta nativas do ImageJ.

REFERÊNCIAS

CAMPOSTRINI, E.; YAMANISHI, O. K. *Estimation of papaya leaf area using the central vein length*. Scientia Agricola, 2001, vol. 58 p. 39-42.

MARACAJÁ, P. B et ali. Estimativa de área foliar de juazeiro por dimensões lineares do limbo foliar. **Revista Verde** (Mossoró – RN – Brasil) v.3, n.4, p0-05 outubro/dezembro de 2008.

QUEIROGA, J. L. et ali. Estimativa da área foliar de feijão-vagem (Phaseolus vulgaris L.) por meio da largura máxima do folíolo central. Horticultura Brasileira, 2003, vol. 21(1): 64-68.

SALISBURY, Frank B.; ROSS, Cleon W. *Fisiologia das plantas*. São Paulo: Cengage Learning, 2012.

Produção de *Agaricus blazei* em função de diferentes umidades acionadas automaticamente por sensores.

Gustavo Carolino Girardi[1], Pedro Luis de Paula Filho[1], Itamar Iliuk[2] ,Daniel Rodrigues Blanco[2], Vitor Alex Alves de Marchi[2], Jonathan Galdino da Silva[2], Danilo Luiz Cavazim Neto[2]

[1] Departamento de informática, Universidade Tecnológica Federal do Paraná, Medianeira, Paraná, Brasil, gustavogirardi23@gmail.com, plpf2004@gmail.com.

[2] Departamento de informática, Universidade Tecnológica Federal do Paraná, Santa Helena, Paraná, Brasil, danielrblanco.utfpr@gmail.com, Itamar.iliuk@gmail.com, vmarchi@utfpr.edu.br, johnzballad@gmail.com, dcavazim@gmail.com.

RESUMO

O Cogumelo *Agaricus blazei*, popularmente conhecido como Cogumelo do Sol, é um fungo que se desenvolve em climas temperados. É considerado um suplemento alimentar e consumido em decorrência de seu valor medicinal, ação protetora e fortalecedora do sistema imunológico, além de possuir ação anticancerígena, contribuindo para o aumento da imunidade das pessoas. Neste trabalho avaliou-se o efeito de 6 diferentes umidades de canteiros de *A. blazei*, oscilando em um percentual de 90% até 65% de umidade, com variação de 5% de umidade para cada canteiro, a quantidade de substrato utilizada foi de 88,300 kg e 0,883 gramas de sementes em cada metro quadrado de experimento. O delineamento experimental foi planejado e analisado com seis níveis. A indução automática de água tendeu a apresentar produção em todos os canteiros. Obteve como resultado que a média de colheita depende da umidade. O acionamento foi totalmente automatizado com sensores de umidade aferidos, que passava as informações para a plataforma Arduino que acionava, quando necessário, as válvulas solenoides aspergindo água.

PALAVRAS-CHAVE: Temperatura, Umidade, *Agaricus blazei*.

ABSTRACT

The Mushroom Agaricus blazei, popularly known as Mushroom of the Sun, is a fungus that develops in temperate climates. It is considered a food supplement and consumed as a result of its medicinal value, protective action and strengthening of the immune system, besides having anticancer action, contributing to increase the immunity of the people. In this work the effect of 6 different humidity of A. blazei beds was evaluated, ranging from 90% to 65% of humidity, with a variation of 5% of humidity for each plot, the amount of substrate used was 88,300 Kg and 0.883 grams of seeds in each square meter of experiment. The experimental design was planned and analyzed with six levels. The automatic induction of water tended to present production in all the beds. As a result, the average harvest depends on the humidity. The drive was fully automated with calibrated humidity sensors, which passed the information to the Arduino platform that triggered, when necessary, the solenoid valves sprinkling water.

KEY WORDS: Temperature, Humidity, Agaricus blazei.

INTRODUÇÃO

O agronegócio familiar brasileiro caracteriza-se pela sua rápida expansão, dados do instituto brasileiro de geografia e estatística IBGE confirmam que 85% das propriedades rurais constituem-se de núcleos familiares, sendo seu desenvolvimento crescente (INCRA/FAO, 2000).

A maioria dessas propriedades efetua plantio de soja e milho, culturas características da região brasileira, contudo, o cultivo de fungos é uma possibilidade de complemento de renda para essas pequenas propriedades, entre eles, o cogumelo *A. blazei*, que possui características alimentares favoráveis ao fortalecimento do sistema imunológico (ZHANXI; ZHANHUA, 2001; SHAH; AHARAF; SHTIAQ, 2004 apud VARGAS A., 2011). É uma cultura de baixo custo e seu manejo é totalmente artesanal, o que significa a necessidade de desenvolvimento de tecnologias que possam inclui-lo na agricultura digital, umas das possibilidades de automação da produção deste fungo é a utilização de sensores de umidade para monitoramento dessas variáveis bióticas, constituindo um sistema de irrigação automatizado, que pode influenciar diretamente no desenvolvimento da safra.

Por ser um fungo que apresenta melhor produtividade em ambientes úmidos (OEI, 2003; GRIENSVEN, 1988) a utilização de tecnologias que possibilitem o monitoramento das condições climáticas ambientais possibilita ao produtor melhorar sua qualidade e produtividade, bem como dispêndio de tempo para controle da umidade, uma vez que a

programação de sensores pode acionar automaticamente uma válvula que fará a irrigação do canteiro.

A irrigação efetuada por meio de aspersor assemelha-se à chuva, uma vez acionada conduz a água por uma tubulação até o canteiro a ser irrigado, sendo um sistema de fácil instrumentalização (LUCIETTI, 2014).

A programação do sistema de irrigação pode ser efetuada através da plataforma Arduino, de computação de código aberto fundamentado em uma placa micro controladora ATmega2560 (ARDUINO.CC, 2015), que dispõe de um espaço de desenvolvimento para escrever o código e introduzi-lo no micro controlador com comunicação serial. A plataforma Arduino tem como propósito principal descomplicar o desenvolvimento de sistemas embarcados capazes de interagir com o mundo exterior por meio de sensores e atuadores. O plano traçado na plataforma Arduino pode ser de forma autônoma ou capaz de se comunicar com um *software* executado em um computador. Os circuitos podem ser montados à mão ou adquiridos pré-montados. A obtenção do *software* de programação é código-livre e gratuito. A linguagem de programação da plataforma Arduino é a linguagem C e aplica bibliotecas similares as do projeto *Wiring* (GIRARDI, 2017). O ambiente de programação é similar ao ambiente da linguagem de programação multimídia *Processing* (ARDUINO.CC, 2015).

MATERIAL E MÉTODOS

O projeto foi desenvolvido através da necessidade que o *A. blazei* possui de ser umidificado constantemente. Cultivado na cidade de Santa Helena, estado do Paraná, na zona rural, o experimento foi realizado em um barracão de 5 m x 4 m de comprimento, construído com madeira rústica, cobertura de telha de barro, pé direito de 3 m de altura e fechamento das laterais com lonas. A ventilação, em seu interior acontece de forma natural através da passagem de ar durante o período em que as lonas forem retiradas. A metodologia seguiu os padrões adotados por Braga (2006).

Foram construídos 6 canteiros para experimentos de 1 m² cada nomeados com as letras A, B, C, D, E, F. Para cada experimento utilizamos um conjunto de sensores e atuadores, contendo nesse conjunto uma válvula solenoide 220V com um bico de micro aspersão de água na extremidade ligada diretamente a um rele 12V que recebe o sinal digital da plataforma Arduino para o acionamento e realização das medidas de umidade do solo, executada pelo sensor de umidade YL69 conectado por meio de cabo de rede de ethernet CAT5 até a plataforma Arduino MEGA2560.

Para que a plataforma Arduino atuasse como termo higrômetro e exibisse o sinal analógico em números, um resistor no valor de 1KΩ foi utilizado em um circuito com funcionamento de *PULLDOWN*, alimentação de 5V e saída analógica entre os resistores, a ser lido e de *PULLDOWN*, após esse funcionamento foi feito a programação na plataforma Arduino.

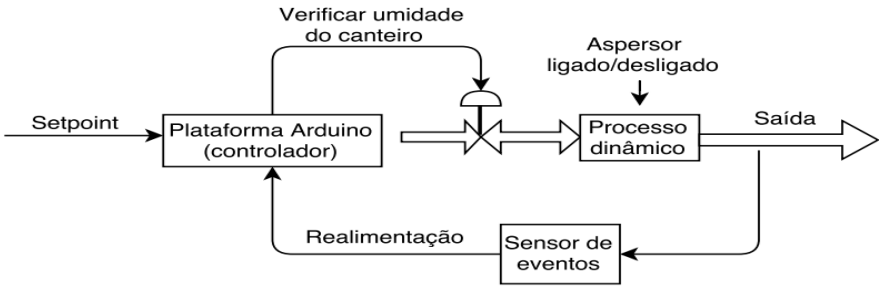
O software foi desenvolvido através da ferramenta plataforma Arduino que é de código e uso livre, assim como todas as bibliotecas usadas no desenvolvimento do trabalho. O microcontrolador foi calibrado para gerar diversas medidas via software com o tempo de 15 minutos entre elas. Como resultado, o microcontrolador apresenta no notebook a umidade de cada canteiro e grava a data e o horário dos dados coletados em uma planilha on line no Google Docs.

Método de controle da umidade do canteiro

O controle de umidade é uma dinâmica de evento discreto, as variáveis internas e de saída alteram-se sempre como consequência de variações quase instantâneas das entradas e de certos sensores internos, obedecendo a rígidas regras de causa e efeito, assumindo somente valores pertencentes a um conjunto finito, como desligado, ligado (CASTRUCCI, 2011). Um sistema de controle de malha fechada tem como ponto de partida um *Setpoint* e a realimentação desse sinal para compará-lo (BISHOP, 2013). De acordo com a ação que atua nos processos os controladores automáticos podem ser classificados do tipo Liga-Desliga (OGATA, 2011).

O *Setpoint* do canteiro A foi programado na plataforma Arduino com umidade de 90%. No canteiro B foi programado umidade de 85%. No canteiro C o Setpoint foi de 80%. No canteiro D 75%. No canteiro E, 70% e no canteiro F, na plataforma Arduino, uma umidade de 65%. Processo de controle demonstrado na figura 1.

Figura 1 - Demonstração do controle

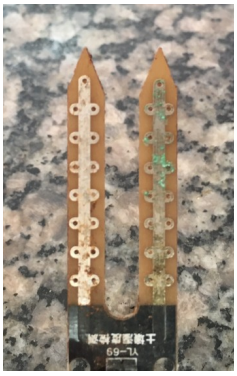


Fonte: Próprio autor.

Umidade do canteiro

Para aferir a umidade do canteiro com o sensor de umidade YL69 (Figura 2), foi coletado amostras do substrato de cada canteiro com uma pinça em aço inox, nesse mesmo período a plataforma Arduino registrava no computador os valores que cada sensor estava indicando (Quadro 1), para garantir a verificação da umidade de cada canteiro.

Figura 2 - Foto do sensor de umidade de solo



Fonte: Próprio autor.

Fonte: Próprio autor.

RESULTADOS E DISCUSSÃO

Saber a umidade do substrato é de fundamental importância, pois indica em que condições hídricas encontra-se o mesmo. Vários métodos são usados para definir a umidade, no Brasil os métodos oficiais foram publicados em uma Instrução Normativa do Ministério da agricultura pecuária e abastecimento (MAPA) n.° 17 (IN n.° 17), de maio de 2007 (Brasil, 2007), alterada pela Instrução Normativa n.° 31 (IN n.° 31), do MAPA de 23 de outubro de 2008 (Brasil, 2008).

Quadro 1 - Registro dos valores de cada sensor.

Canteiro [A] = 79.00 %
Canteiro [B] = 77.00 %
Canteiro [C] = 76.00 %
Canteiro [D] = 75.00 %
Canteiro [E] = 78.00 %
Canteiro [F] = 73.00 %

Essa alteração se deu com relação ao método de avaliação de densidade da água, visto que no país não consideram esse tipo de atributo. Por não ser uma alteração significativa, a IN n.° 17, ainda vigente, traz em sua redação os métodos analíticos oficiais para Análise de Substratos, onde também demonstra os passos que foram seguidos por este trabalho para as amostras serem adequadas.

Em um primeiro momento foram retiradas 3 três amostras de aproximadamente 100 g de substrato de cada canteiro. As mesmas foram passadas em peneira de malha 19 x 19 mm (IN n.° 17), a quantidade foi escolhida para uma melhor amostragem (Tabela 1), após essa etapa foram todas vedadas e levadas para o Laboratório de microbiologia da Universidade Tecnológica Federal do Paraná – Campus Santa Helena.

Foram pesadas cada amostra em balança de precisão (BIOSCALE), sendo colocadas em um Becker com suas respectivas identificações. Nos 6 canteiros coletamos 3 amostras em cada um, totalizando 18 amostras inseridas na estufa (LUCADEMA), com temperatura de (65ºC ± 5,0ºC), durante 2 dias (IN n.° 17) para a determinação da umidade atual. Após 48 horas foi retirado e novamente pesado, resultando através da formula % = pu - ps / pu *100, em uma média geral de umidade de 76,84%.

Tabela 1 – Tabela de demonstração das médias da diferença entre o peso seco e úmido, aplicado a fórmula.

AMOSTRAS	Peso das amostras Úmidas	Peso das amostras Secas	Porcentagem de umidade (Formula)
Média amostra (A)	122,1063	26,81073333	78,04312035
Média amostra (B)	121,0811333	27,48823333	77,29767423
Média amostra (C)	122,0292667	27,4915	77,47138801
Média amostra (D)	121,8313667	28,6229	76,50613238
Média amostra (E)	120,5544667	27,74043333	76,98929447
Média amostra (F)	121,2105	30,59083333	74,76222494

Fonte: Próprio autor**.**

A plataforma Arduino, usada na experimentação com as devidas calibrações de hardware e programações de software, gravam, ordenam e apresentam os dados, a cada 15 minutos, em uma planilha Google Docs (Figura 3).

Figura 3 - Apresentação dos dados armazenados na planilha Google Docs.

	A	B	C	D	E	F	G
1	Dia & Hora	Umidade do Solo (A)	Umidade do Solo (B)	Umidade do Solo (C)	Umidade do Solo (D)	Umidade do Solo (E)	Umidade do Solo (F)
2	08/04/2017	87.00	82.00	82.00	82.00	80.00	81.00

Fonte: Próprio autor**.**

A produção foi determinada através da quantidade colhida de cogumelos in-natura em quatro colheitas (Gráfico 1), no período compreendido entre fevereiro a maio do ano de 2017. Verifica-se, neste experimento, que no canteiro A à produção total foi de 10,489 kg, do

canteiro B 7,569 kg, e no canteiro C a produção foi de 7,656 kg, especificamente esse canteiro teve um problema de contaminação externa, oriunda do próprio fornecedor do substrato, o que justifica uma menor produção, a produção do canteiro D foi de 12,022 kg, do canteiro E 9,381 kg e a produção do canteiro F de 8,439 kg.

Como motivo de investigação do comportamento do *A. blazei* nas condições de Santa Helena – PR, seguimos a ideia de que possivelmente, o processo de umedecimento do composto e boa homogeneização com as sementes, tenham contribuído para o crescimento.

No continente europeu, umedecer o composto e homogeneizar o substrato são atividades consideradas decisivas para obtenção de sucesso na colheita, além de facilitar a elevação do padrão de qualidade (KOPYTOWSKI, 2006).

Os dados foram submetidos a análise de variância (Quadro 2), conhecida como ANOVA, que pode ser usada para comparar médias quando existem mais de dois níveis de um único fator (MONTGOMERY, 2013). Esses níveis são as quantidades de colheitas efetuadas durante todo o período de safra e o fator é a umidade de cada canteiro.

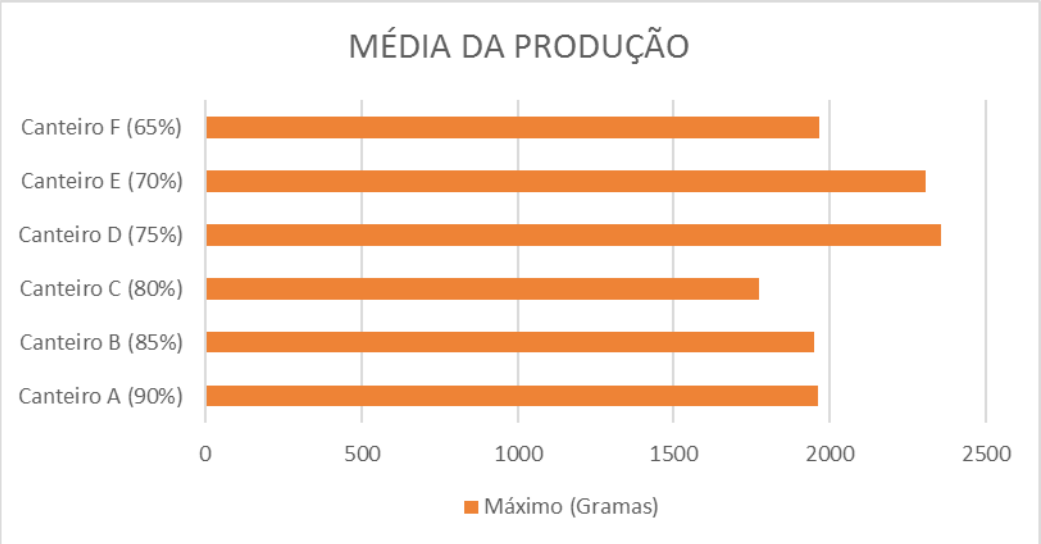
Quadro 2 - ANOVA da Produção de Agaricus blazei.

Causa Variação	SQ	GL	QM	F
Tratamentos	3.021.800,27	5	604.360,05	1,50
Resíduos	26.908.396,60	67	401.617,86	
Total	29.930.196,88	72		

Fonte: Próprio autor**.**

O gráfico 1 a seguir apresenta o intervalo de confiança de 95% da produção em cada canteiro, o que corrobora com o resultado obtido no Quadro 2, demonstrando que durante a produção pode não ocorrer colheita em um determinado canteiro e em outro pode chegar a quase 2500 gramas de cogumelos dependendo da umidade.

Gráfico 1 - Demonstração do intervalo da média de produção com 95% de confiança.



Fonte: Próprio autor.

CONCLUSÕES

A partir dos resultados apresentados pode-se concluir que na cidade de Santa Helena – PR, a umidade controlada pelo conjunto de sensores e atuadores obteve como melhor desempenho o canteiro D (75%) durante os quatro meses da safra de *A. blazei*.

No presente trabalho a produção de *A. blazei* em função de diferentes umidades acionadas automaticamente por sensores, trouxe um resultado onde a programação na plataforma Arduino interferiu na produção, a umidade usada de 75% foi bem aceita na colheita do cogumelo in-natura. Como os resultados encontrados neste artigo não são conflitantes com o encontrado na literatura, mostra que controlar a umidade por sensores e atuadores pode ser uma alternativa para o agricultor.

O processo carece de mais investigações com outros fatores que podem ser influentes no processo de crescimento e cultivo da produção.

REFERÊNCIAS

ARDUINO.CC, 2015. Arduino. [Online] Available at: https://www.arduino.cc/> Acesso em: 27 fev 2017.

BISHOP RH, DORF RC. Sistemas de controle modernos. Gen LTC, Rio de Janeiro, Brasil, 2013; 12° Edição. ISBN 978-85-216-1995-6: p.2.

BRAGA GC, MONTINI RMC, SALIBE AB. Parâmetros da produção de Agaricus blazei sob diferentes condições ambientais de cultivo. Scientia Agraria Paranaensis, Marechal Candido Rondon, Brasil. 2006; v.5 n°1: 47-56.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento – MAPA. Secretaria de Defesa Agropecuária. Instrução Normativa Nº 17, de 21 de maio de 2007. Aprova os Métodos Analíticos Oficiais para Análise de Substratos e Condicionadores de Solos, na forma do Anexo à presente Instrução Normativa. Diário Oficial da República Federativa do Brasil, Brasília, DF, 24 maio 2007. Seção 1, p.8.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento – MAPA. Secretaria de Defesa Agropecuária. Instrução Normativa N.º 31, de 23 de outubro de 2008. Altera os subitens 3.1.2, 4.1 e 4.1.2, do Anexo à Instrução Normativa SDA n.º 17, de 21 de maio de 2007. Diário Oficial da República Federativa do Brasil, Brasília, DF, 24 out. 2008. Seção 1, p.20.

CASTRUCCI PL, BITTAR A, SALES RM. Controle automático. Gen LTC, Rio de Janeiro, Brasil, 2011; ISBN 978-85-216-1786-0: p.1.

GRIENSVEN, L. J. L. D.(Ed.) The cultivation of mushrooms. Horst: Mushroom Experimental Station, 1988. p.515.

GIRARDI GC, ILIUK I, BLANCO DR. Recursos de tecnologias computacionais para a facilitação no manejo do cultivo do cogumelo do sol. CONFERENCIA CIENTIFICA INTERNACIONAL. UNIVERSIDADE DE HOLGUIN. Cuba. 2017; 8° Edição. ISBN 978-959-16-3272-2: p.34.

INCRA/FAO. Novo retrato da agricultura familiar: o Brasil redescoberto. Brasília: INCRA/FAO, 2000.

KOPYTOWSKI FILHO, J. Produtividade e eficiência biológica de agaricus blazei (murrill) heinemann, em diferentes condições de cultivo. Tese apresentada à Faculdade de Ciências Agronômicas da UNESP. Bocutacu, 2006.

LUCIETTI, D. Irrigação das hortaliças: cultivo orgânico. Santa Catarina, 2014.Disponível em: < http://cultivehortaorganica.blogspot.com.br /2014/01/irrigacao-das-hortalicas.html>. Acesso em: 25 ago. 2016.

MONTGOMERY D.C, RUNGER G.C. Estatística aplicada e probabilidade para engenheiros. Gen LTC, Rio de Janeiro, Brasil. 2013; Edição 5; ISBN 978-85-216-1902-4: p.338.

OEI, P. Mushroom cultivation: appropriate technology for mushroom growers. Backhuys Publishers: Leiden, 2003. 426 p.

OGATA, K. Engenharia de Controle Moderno. 2011. 5º ed. Editora Pearson.

PARDO A, JUAN A, PARDO E. Characterization of diferente substrates for possible use as casing in mushroom cultivation. *Sci. Technol. 1*: 2003. 107-114.

VARGAS, A. M. Análise Diagnóstica da Cadeia Produtiva de Cogumelos do Distrito Federal. Dissertação (mestrado) – Universidade de Brasília (UNB). Programa de Pesquisa e Pós-graduação em Agronegócios– Brasília, 2011.

SISTEMA TELEMÉTRICO VIA RFID EM UHF PARA MEDIÇÃO DA UMIDADE DO SOLO

Adriano Ricardo de Abreu Gamba [1], *Sérgio Francisco Pichorim* [1 e 2]

[1] DAELN - Departamento Acadêmico de Eletrônica, UTFPR - Universidade Tecnológica Federal do Paraná, Curitiba, PR, Brasil, adriano-gamba@hotmail.com

[2] CPGEI - Curso de Pós-Graduação em Eng. Elétrica e Informática Indust., UTFPR - Universidade Tecnológica Federal do Paraná, Curitiba, PR, Brasil, pichorim@utfpr.edu.br

RESUMO

O uso de sistemas de sensores remotos pode ajudar a monitorar o grau de umidade no solo em áreas cultivadas. A tecnologia de RFID em UHF, que apresenta os recursos de operar sem fio (telemetria *wireless*) e sem baterias (passivamente), tem grande potencial de aplicação na agricultura. Os parâmetros elétricos do solo, em especial a condutância/resistência e a capacitância, variam com a presença da umidade. Desta forma, pequenos sensores que detectam esses parâmetros elétricos podem ser usados para medir a umidade do solo. Neste trabalho é apresentado um sistema de telemetria passiva que integra a tecnologia de RFID em UHF com sensores de umidade em solo. Para tal, foram avaliados dois formatos de sensores capacitivos e um de sensor resistivo. Também foi feita uma análise da integração destes sensores em um circuito integrado SL900A para o desenvolvimento do sistema telemétrico completo de medição de umidade via leitor de RFID em UHF. Os erros médios foram de 11,5% para o sistema telemétrico com sensor capacitivo e 7,7% para o resistivo.

PALAVRAS-CHAVE: Umidade em solo, Sensor capacitivo, Sensor resistivo, RFID.

ABSTRACT

The use of remote sensing systems can help to monitor the degree of soil moisture in cultivated areas. The UHF RFID technology, which is capable of wireless operation (telemetry) and without batteries (passively), has great potential for application in agriculture. The electrical parameters of the soil, especially the conductance/resistance and the

capacitance, vary with the presence of humidity. In this way, small sensors, that detect these electrical parameters, can be used to measure soil moisture. This work presents a passive telemetry system that integrates the UHF RFID technology with soil moisture sensors. For that, two designs of capacitive sensors and one of resistive sensor were evaluated. An analysis of the integration of these sensors into a SL900A integrated circuit was also carried out for the development of the complete telemetric system for humidity measurement via an UHF RFID reader.

KEYWORDS: Soil moisture, Capacitive sensor, Resistive sensor, RFID.

INTRODUÇÃO

A umidade do solo é um aspecto que varia com o ambiente e cuja análise evidencia várias características do meio onde se encontra. Assim sendo, a medição da umidade do solo vem a ser importante em várias áreas, dentre elas, na agricultura, na engenharia civil e na biologia, pois permite, por exemplo, estudos relacionados à proliferação de vida e ao transporte de nutrientes no interior da terra, e ainda na supervisão da quantidade de água em encostas como controle de riscos de acidentes (PEREIRA et al, 2015; PICHORIM et al, 2014; AEMC, 2002).

Redes de sensores sem fio para medição de umidade em solo têm sido utilizadas como uma solução para a agricultura, em especial onde se tem a necessidade de irrigação e controle da demanda da quantidade de água (PASIEKA et al, 2015), resultando em economia e maior eficiência no processo produtivo.

Nas últimas duas décadas, sistemas de identificação via radiofrequência (RFID em inglês) têm sido utilizados para rastreabilidade na indústria, na cadeia de produção agrícola e na pecuária (ERADUS e JANSEN, 1999; DA SILVA et al, 2015). Contudo, o uso da tecnologia de RFID tem sido pouco explorado como uma rede de sensores para sensoriamento remoto.

Diante disso, foi desenvolvido um sistema de telemetria via RFID, onde sensores de umidade foram incorporados ao dispositivo passivo (*tag*). Os sensores detectam a capacitância ou a condutância do meio onde estão inseridos (solo). E, uma vez que estas características elétricas estão diretamente relacionadas com a umidade do solo, pode-se usar este tipo de sensor para se medir a umidade (BAI et al, 2013). Assim, neste trabalho, foram avaliados três tipos de sensores: capacitivo espiral quadrado, capacitivo paralelo retangular e resistivo cilíndrico, os quais foram usados para detectar a umidade do solo na faixa de 0% até 30%. Também foram realizados testes de configuração e leitura destes sensores pelo circuito

integrado do *tag*, bem como testes de comunicação do sistema de telemetria (*tag* e leitor) via enlace RFID em UHF (868 MHz).

MATERIAL E MÉTODOS

O sistema de telemetria para medição da umidade no solo é constituído de três partes: 1) um sensor que vai enterrado no solo, 2) um módulo passivo (*tag*) de leitura do sensor e transmissão da informação via UHF (ver figura 1A), e 3) um leitor de RFID que recebe os dados lidos dos *tags* remotos. Neste trabalho foram avaliados três tipos de sensores para umidade: dois capacitivos e um resistivo. Os sensores capacitivos (figura 1B), que foram desenhados em dois formatos, de trilhas em espiral e de trilhas em paralelo, podem medir a capacitância ou a condutância (*Cs* ou *Gs*, respectivamente) do solo. O sensor resistivo (figura 1C) mede a resistência do solo (*Rs*). Estas três grandezas elétricas (*Cs*, *Rs* e *Gs*) têm uma alta correlação com o grau de umidade presente no solo. Para a leitura e transmissão passiva (sem fio e sem bateria) destas grandezas foi utilizado um módulo de RFID (identificação por radiofrequência) em UHF (868 ou 915 MHz) com entradas externas para sensores do tipo capacitivo, resistivo e condutivo. Este módulo passivo (*tag*) é construído com o circuito integrado SL900A da empresa austríaca AMS. Para o leitor remoto dos *tags* foi utilizado o leitor Newton desenvolvido pela mesma empresa.

Descrição e caracterização dos sensores

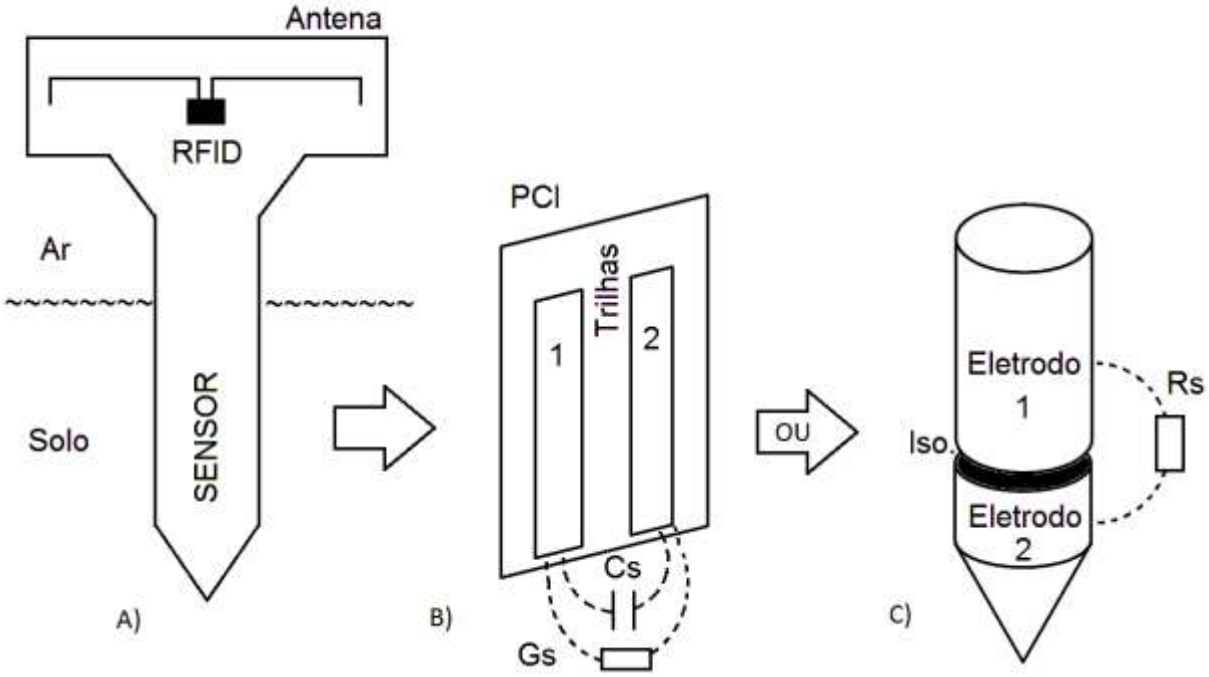
Ao todo, três possibilidades de sensor que medem a umidade do solo através da medição da capacitância, da resistência ou da condutância foram desenvolvidos e avaliados, conforme as figuras 1B, 1C e 2A. Os sensores utilizados, bem como suas características construtivas, são descritos a seguir:

a) Sensor capacitivo paralelo: é constituído de 12 trilhas de cobre dispostas em paralelo sobre uma placa de circuito impresso (PCI), com largura das trilhas de 2 mm e 0,5 mm de separação. Esta PCI é revestida com uma tinta isolante para proteção. Toda a área útil deste sensor é de 30 × 70 mm²;

b) Sensor capacitivo espiral: é constituído de um par de trilhas de cobre dispostas em um desenho espiral quadrado sobre uma PCI, com trilhas de largura de 0,5 mm, com 0,5 mm de separação. Esta PCI também é revestida com uma tinta isolante e a sua área útil é de 38 × 38 mm²; e

c) Sensor resistivo, que é constituído de uma haste metálica com dois eletrodos de aço inox em formato cilíndrico, com diâmetro de 10 mm, separados por um isolante de 5 mm e com área útil de aproximadamente 20 cm².

Figura 1: Sistema de telemetria com *tag* de RFID em UHF, os dois tipos de sensores utilizados e as grandezas medidas (*Cs, Gs* e *Rs*). A) Dispositivo RFID com a antena sobre o solo e extensão (sensor) inserida no solo. B) Sensor capacitivo construído em PCI para medir a capacitância *Cs* e a condutância *Gs* entre as trilhas. C) Sensor de haste metálica para medir a resistência *Rs* entre os eletrodos 1 e 2.



Fonte: Autores

Sistema telemétrico

Com o propósito de tornar possível a realização das medidas de capacitância, condutância e resistência à distância, foi usado um módulo de comunicação RFID passivo (sem bateria) que, além de medir a capacitância e a resistência dos sensores a ele conectados, pode ser programado e lido remotamente via sinais de UHF por uma unidade leitora.

O módulo remoto (ou *tag*) é composto do circuito integrado (CI) SL900A, uma antena dipolo e o sensor, conforme a figura 1A. O *tag* trabalha de forma passiva, ou seja, sem bateria, de maneira que a sua alimentação vem do próprio campo eletromagnético de frequência ultra-alta (UHF de 868 ou 915 MHz) gerado pela unidade leitora afastada de poucos metros. O alcance da comunicação depende apenas da unidade leitora, da sua potência

e do ganho da sua antena. O *tag* vai operar parcialmente enterrado no solo, de forma que o CI e a antena ficam fora do solo, ligeiramente acima dele (em torno de 10 a 15 cm).

O CI possui entradas e modos de operação específicos para sensores resistivos/condutivos e capacitivos, os quais foram testados a fim de se obter a configuração que permitisse a melhor leitura possível da faixa de valores que são medidos pelos sensores.

Para a unidade leitora, foi utilizado o Newton *demo reader* da empresa AMS, o qual é um leitor pré-programado para se trabalhar com o CI SL900A. O leitor possui uma interface que permite a leitura e a configuração do *tag* a partir de um *software* dedicado no computador. Isso possibilita a mudança das configurações da interface sensora do *tag*, de seu modo de funcionamento, e a leitura da capacitância e/ou da resistência lidas nas entradas analógicas do CI.

Metodologia dos experimentos

Os experimentos realizados para se avaliar a umidade no solo com os sensores, bem como a caracterização mais específica dos sensores, do *tag* e da unidade leitora de RFID estão descritos a seguir.

Uma amostra de solo foi coletada e peneirada com o intuito de evitar a presença de pedras, folhas, galhos, etc. Após isso, o solo foi seco em um forno. Em seguida, foi escolhido um copo de béquer no qual 641 g de solo e os sensores capacitivos espiral e paralelo (figura 2A) seriam colocados para se realizar o experimento. As massas do béquer (197 g), dos sensores (23 g, ambos os sensores) e do solo foram medidas com uma balança de precisão. Por fim, o solo foi colocado no béquer e os sensores foram enterrados de modo que suas áreas úteis ficassem completamente submersas dentro do solo (figura 2B).

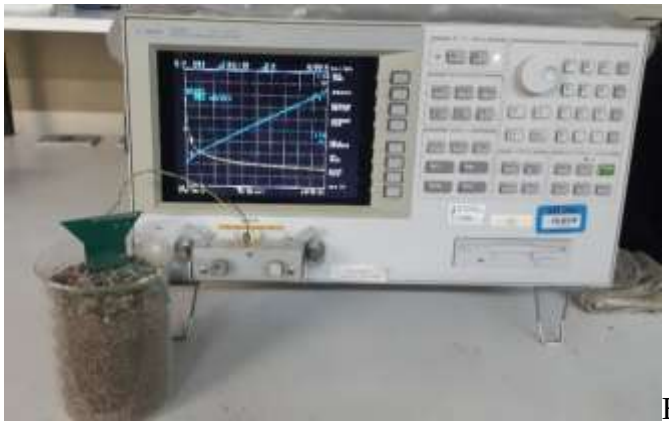
Para se calcular a quantidade de água presente no solo (PICHORIM et al, 2014), foram utilizadas as massas da amostra seca e úmida. Desta forma, a umidade gravimétrica (*H*) do solo é dada por

$$H = \frac{m_w - m_d}{m_d}.100\% = \frac{m_a}{m_d}.100\% \quad (1)$$

onde m_w é a massa de terra com água, m_d é a massa de solo seco utilizada na amostra, e m_a é a massa de água da amostra.

Após registrados os dados iniciais da amostra, incluindo as medidas de capacitância (pF) e condutância (μS) do solo seco, um volume de água filtrada foi colocado no béquer com o solo. Assim, para a inserção de 141 g de água, a umidade do solo ficou em aproximadamente 20%.

Figura 2: Fotografia do experimento de caracterização dos sensores capacitivos. A) Sensor paralelo à esquerda e sensor espiral à direita. B) Amostra de terra com os sensores e o analisador de impedância utilizado.


A)

B)

Fonte: Autores.

As medidas de capacitância (*Cs*) e condutância (*Gs*) detectadas pelos sensores foram lidas pelo analisador de impedâncias Agilent 4294A (figura 2B), sempre na frequência de 2 kHz. Esta é a mesma frequência que o CI SL900A utiliza para a leitura da capacitância em sua entrada.

Ao longo do processo de secagem, o solo foi medido periodicamente durante aproximadamente 2 meses, de maneira a se poder verificar a variação de sua umidade de uma maneira mais natural, onde o ambiente foi responsável por fornecer a energia térmica suficiente para secá-lo.

Análise dos dados

Sobre os dados coletados de *Cs* e *Gs* foi feita a plotagem de gráficos para cada sensor. Uma análise estatística foi aplicada, principalmente para o ajuste destes valores através de uma função sigmoide, dada por

$$f(x) = \frac{A_1 - A_2}{1 + e^{\frac{x - x_0}{dx}}} + A_2 \qquad (2)$$

onde os parâmetros indicam: A_1, o limite inferior de $f(x)$; A_2, o limite superior da função; x_0, o ponto de inflexão da curva; e dx é a largura (VASILIEF et al, 2001). Os ajustes e os gráficos foram gerados utilizando-se o *software* SciDAVis.

A partir dessa análise, foram obtidas as curvas e as equações que relacionam a capacitância e a condutância medidas pelos sensores com a umidade do solo.

Análise do sensor resistivo

O sensor resistivo teve sua curva de característica levantada na prática através de um procedimento de comparação com um instrumento de referência. Para isso, foi utilizado o medidor de umidade MO750 da empresa Extech Instruments Corporation. Com ele, foram medidos valores de umidade que o sensor apresentava para vários valores de resistência, o que resultou em um gráfico que relaciona a resistência medida com a umidade do solo.

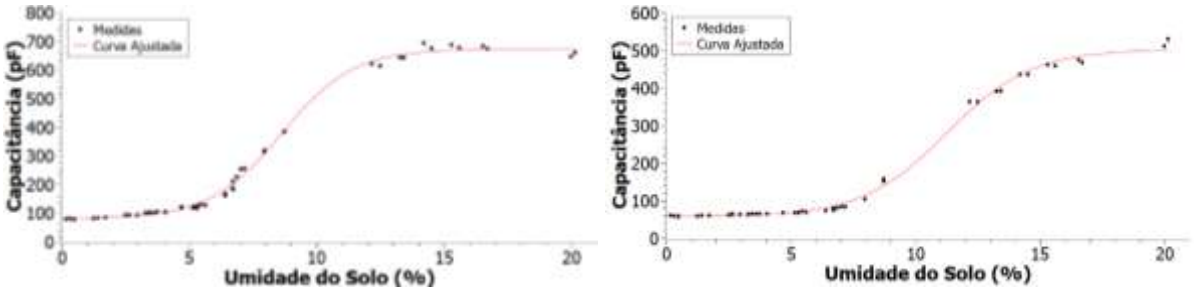
RESULTADOS E DISCUSSÃO

A partir das medições realizadas no experimento com o solo, foram feitos gráficos a fim de se representar as medidas realizadas e a aproximação destas com as respectivas curvas ajustadas. Foram obtidos também, através do programa SciDAVis, os valores de correlação R^2, que dizem respeito à concordância entre os pontos e a curva ajustada, e os coeficientes das equações das curvas sigmoides.

Assim sendo, a figura 3 apresenta a variação da capacitância (*Cs*) em função da variação da umidade do solo (*H*) com valores de R^2 de 0,9973 e 0,9971 para os sensores capacitivos de trilhas paralelas e em espiral, respectivamente. Por outro lado, a figura 4 apresenta a variação da condutância (*Gs*) em função da variação da umidade do solo (*H*) com valores de R^2 de 0,9875 e 0,9950 para os mesmos sensores paralelo e espiral, respectivamente.

Os resultados mostram que a capacitância e a condutância aumentam conforme a umidade do solo aumenta, e que estas se aproximam muito de uma função sigmoide (correlação sempre maior que 0,98). A variação dos parâmetros elétricos *Cs* e *Gs* foi mais intensa da faixa de umidade entre 5% até 15%. Discrepâncias nos valores medidos devem-se principalmente à distribuição não homogênea da água na terra, à disposição dos sensores na amostra e à geometria dos sensores.

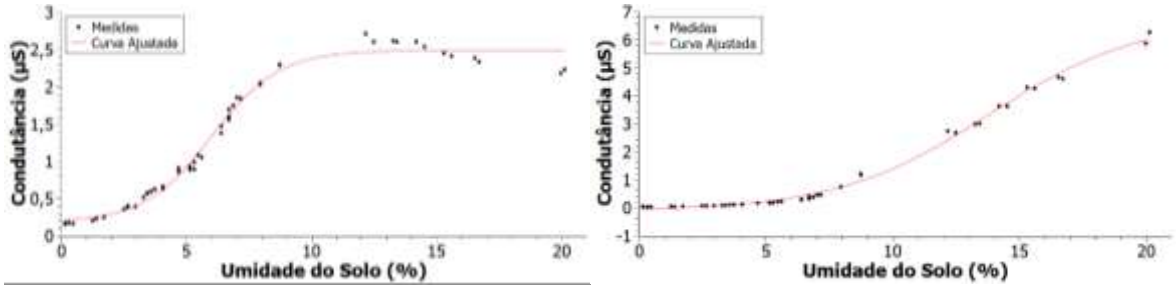
Figura 3: Curvas da relação da capacitância (*Cs*) dos sensores em função da porcentagem de umidade da amostra. A) Sensor com trilhas em paralelo. B) Sensor com trilhas em espiral.



A) B)

Fonte: Autores.

Figura 4: Curvas da relação da condutância (*Gs*) dos sensores em função da porcentagem de umidade da amostra. A) Sensor com trilhas paralelas. B) Sensor com trilhas em espiral.



A) B)

Fonte: Autores.

Os coeficientes obtidos para as curvas sigmoides geradas, baseando-se na equação 2, estão apresentados na tabela 1. A partir deles, é possível ter equações de calibração do sistema de telemetria. Ou seja, partindo-se das leituras remotas (via RFID) dos valores de capacitância ou da condutância dos sensores, pode-se determinar os valores da umidade do solo. Desta forma, na equação 2, a função $f(x)$ passa a representar H (a umidade do solo) e a variável x é a leitura da capacitância ou da condutância medida remotamente no leitor.

Tabela 1: Coeficientes da equação 2 determinados pela análise feita no *software* SciDAVis.

Sensor	Grandeza	Coeficientes			
		A_1	A_2	x_0	dx
Paralelo	Capacitância *Cs*	80,699	673,391	8,622	1,406
	Condutância *Gs*	0,181	2,467	6,012	1,347
Espiral	Capacitância *Cs*	57,167	503,002	11,254	1,703
	Condutância *Gs*	-0,122	6,711	13,760	2,973

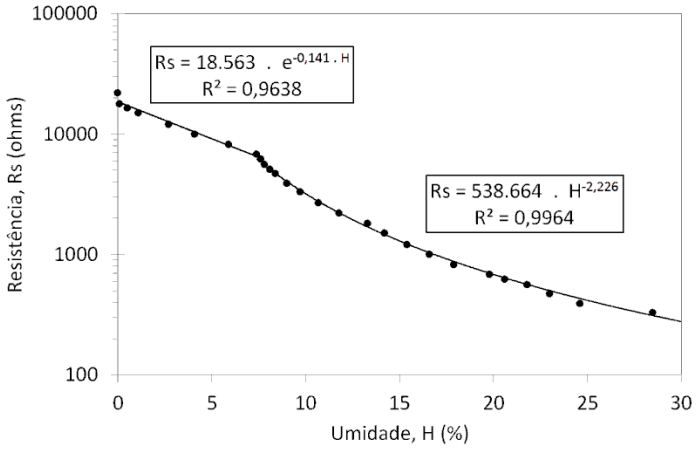
A tabela 2 mostra um comparativo entre os valores de capacitância (*Cs*) e condutância (*Gs*) para ambos os sensores com o solo seco antes de começar o experimento e com o solo seco ao final do experimento. Nota-se que as capacitâncias e as condutâncias medidas ficaram próximas, indicando que os parâmetros elétricos do solo não mudaram significativamente com a realização do experimento.

Tabela 2: Comparativo da terra seca antes e depois do experimento.

Situação	Sensor Paralelo		Sensor Espiral		Umidade
	Cs (pF)	*Gs* (µS)	*Cs* (pF)	*Gs* (µS)	
Antes	75,82	0,182	65,59	0,068	0%
Depois	79,61	0,152	60,49	0,034	0,2%

Para o sensor resistivo constituído dos dois eletrodos cilíndricos da figura 1C, obteve-se um comportamento diferenciado quando do solo mais seco em relação ao mais úmido. A figura 5 apresenta o gráfico obtido, utilizando-se o sensor resistivo, a partir da análise da variação da resistência *Rs* em função das umidades medidas pelo instrumento de referência (MO750). Observa-se que para solo seco (umidade menor que 7% ou *Rs* maior que 7 kΩ) a curva de tendência é uma função exponencial com R^2 de 0,9638. Já para solo mais úmido (*H*>7% ou *Rs* menor que 7 kΩ) a curva é uma função potência com R^2 de 0,9964. Assim, como nos sensores capacitivos, os coeficientes das equações de ajuste (*boxes* na figura 5) podem ser utilizados para a calibração do sistema telemétrico via UHF RFID.

Figura 5: Variação da resistência *Rs* do sensor resistivo com a umidade do solo.



Fonte: Autores

Testes de leitura telemétrica de capacitância e resistência via RFID foram realizados. Estes testes consistiram na comparação dos valores de *Cs* e *Rs* presentes nas entradas analógicas do CI SL900A com os valores lidos pelo leitor de RFID a uma distância de 30 cm. Nas medições da capacitância e resistência, os erros médios foram de 11,5% (faixa entre 33 e 680 pF) e 7,7% (faixa de 470 Ω a 47 kΩ), respectivamente. Os maiores erros de capacitância (19,8%) e de resistência (54,6%) foram encontrados quando os valores estavam nos limites extremos da faixa de operação.

CONCLUSÕES

Os parâmetros elétricos dos sensores desenvolvidos têm boa correlação com a umidade do solo e, através das equações de calibração, pode-se utilizá-los para medição desta grandeza. O uso da tecnologia de RFID com o CI SL900A permite que sensores possam ser incorporados

ao *tag* passivo, desta forma, as informações podem ser transmitidas em UHF para um leitor e para computador. Assim, o sistema via RFID em UHF, aqui apresentado, se mostra capaz de medir telemetricamente a umidade do solo. Em virtude disso, esta solução pode facilitar as atividades em várias áreas da agricultura, além de outros setores, e ainda, abre espaço para que outras aplicações possam ser estudadas e implantadas no futuro.

AGRADECIMENTOS

O presente trabalho foi realizado com o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, processo nº 307509/2015-0) e da Universidade Tecnológica Federal do Paraná (UTFPR).

REFERÊNCIAS

AEMC Instruments. Why Measure Soil Resistivity?, 2002, Disponível em <http://lgstatic.transcat.com/media/pdf/App-Ground-SoilResistivity.pdf>. Acessado em 14 maio, 2017.

BAI, W.; KONG, L.; GUO, A. Effects of physical properties on electrical conductivity of compacted lateritic soil. Journal of Rock Mechanics and Geotechnical Engineering. Wuhan, 3 jun. 2013. p.406.

DA SILVA, R. F., KAWANO, B. R., MARÈ, R. M., CUGNASCA, C. E. Rastreabilidade de açúcar refinado no mercado interno: elaboração de um modelo utilizando as tecnologias de RFID e redes de sensores sem fio. Anais do X Cong. Bras. de Agroinformática, Ponta Grossa, Pr, outubro 2015.

ERADUS, W. J., M. B. JANSEN. Animal identification and monitoring. Computer and Electronics in Agriculture, v. 24, n. 1-2, p. 91-98, 1999.

PASIEKA, T. J., SHUBEITA, F. M., CARAFFA, M. Monitoramento on-line de umidade do solo com auxílio de Rede de Sensores Sem Fio. Anais do X Cong. Bras. de Agroinformática, Ponta Grossa, Pr, outubro 2015.

PEREIRA, R. S., SCHOVEIGERT, I. D., MATOS, J., FALATE, R., AYUB, R. A. Projeto de um sistema de monitoramento para casas de vegetação. Anais do X Cong. Bras. de Agroinformática, Ponta Grossa, Pr, outubro 2015.

PICHORIM, S. F.; MARCIS, V. A.; LASKOSKI, G. T. Humidity in sandy soil measured by passive, wireless, and resonant sensor with bifilar coil, in proceedings of First Latin-American Conference on Bioimpedance, Joinville, Santa Catarina, Brazil, 2014.

VASILIEF, I; GADIOU, R; FRANKE, K. The SciDAVis Handbook, Chapter 6. Analysis of data and curves. Disponível em <http://scidavis.sourceforge.net/manual/x4267.html>. Acessado em 14 maio 2017.



Sistema para Monitoramento da Movimentação Bovina e Aferição dos Comportamentos

Luiz Fernando Delboni Lomba[1], *Leandro de Jesus*[1], *Hana Karina Salles Rubinsztejn*[2], *Luciano Gonda*[2], *Pedro Paulo Pires*[3]

[1]Instituto Federal de Mato Grosso do Sul (IFMS)

[2]Faculdade de Computação (FACOM) - Univ. Fed. de Mato Grosso do Sul (UFMS)

[3]Embrapa Gado de Corte

{luiz.lomba,leandro.jesus}@ifms.edu.br,{hana,gonda}@facom.ufms.br
pedropaulo.pires@embrapa.br

RESUMO

A produção pecuária é uma das principais fontes econômicas do Brasil e a utilização de tecnologias na produção e na gestão dos negócios pecuários apresenta-se como uma poderosa ferramenta para estimular a competividade e qualidade dos produtos. Este artigo apresenta o desenvolvimento de um sistema para automatizar a coleta de dados de bovinos, a fim de aferir automaticamente os seus principais comportamentos. Além de detalhes da solução proposta, são apresentados os resultados de uma experimentação realizada em campo.

PALAVRAS-CHAVE: Pecuária de Precisão, Comportamento Bovino, Reconhecimento de Padrões.

ABSTRACT

Livestock production is one of the main economic sources in Brazil and the use of technologies in the production and management of livestock businesses presents as a powerful tool to stimulate competitiveness and product quality. This paper presents the development of a system to automate the collection of livestock data to to automatically classify main behaviors. In addition of the proposed solution, the results of an experiment carried out in the field are presented.

KEYWORDS: Precision Livestock Farming, Animal Behavior, Pattern Recognition.

1 Introdução

No Brasil, a pecuária representa cerca de 30% do valor gerado pelo agronegócio (BRASIL, 2014), movimentando cerca de 400 bilhões de reais por ano. A competitividade mundial exige o au-

mento da produção de carne de qualidade, de maneira eficaz e competitiva. Essa produção tem evoluído em função do uso de ferramentas de precisão, utilizadas para prever eventos, realizar diagnósticos e auxiliar na tomada de decisões (RUIZ-GARCIA; LUNADEI, 2011).

Para auxiliar o pecuarista no monitoramento do rebanho, a fim de avaliar o bem-estar animal e a produção, é utilizada a Pecuária de Precisão - definida como a gestão de bovinos a partir do uso da Tecnologia da Informação e Comunicação (CÁCERES et al., 2011). Um dos métodos comumente utilizado para obter informações sobre o comportamento dos animais, a fim de prover o manejo eficiente que tome proveito das características e potencialidades do grupo, subgrupo ou indivíduo em especial, é a utilização de colares equipados com sensores que capturam a movimentação do animal.

O *Electronic Collars to Track Cattle* é um exemplo de solução comercial que realiza o monitoramento de bovinos utilizando dados de posicionamento GPS, criado pela *Operational Management and Geodecisional Prototype to Track and Trace Agricultural Production*, para a rastreabilidade bovina (IICA, 2009). A solução apresentada pela empresa é constituída por um colar eletrônico, a estação base e um protocolo de comunicação. Entretanto, ela apresenta os problemas típicos das soluções comercias: alto valor para aquisição dos colares (cada unidade custa alguns milhares de dólares) e o uso de protocolo de comunicação totalmente fechado (JESUS, 2014).

O grande desafio é desenvolver ferramentas de baixo custo, que possam ser incorporadas a outros sistemas, mas que principalmente apresentem uma acurácia aceitável. Baseado neste contexto, o principal objetivo deste trabalho é desenvolver uma solução para o monitoramento de bovinos, que obtenha dados sobre a movimentação dos animais utilizando colares, e faça a predição dos comportamentos utilizando algoritmos de aprendizado de máquina.

Para tal, este artigo está estruturado em mais três seções, além desta introdução. A Seção 2 apresenta os trabalhos relacionados, apresentando a abordagem utilizada por cada um. A Seção 3 apresenta o experimento em campo, o equipamento utilizado e a metodologia utilizada para realizar a predição dos comportamentos. A Seção 4 discute os resultados obtidos comparado aos trabalhos relacionados. E por fim, a Seção 5 apresenta as conclusões e as possibilidades de trabalhos futuros.

2 TRABALHOS RELACIONADOS

As pesquisas visando automatizar o monitoramento do comportamento animal têm se tornado cada vez mais atraentes, tanto para fins científicos quanto para fins produtivos. Entre os trabalhos analisados observam-se diferentes características quanto ao processo de coleta e processamento dos dados. A Tabela 1 faz um comparativo entre o tipo de animal monitorado, a posição em que o equipamento foi instalado no animal e os sensores utilizados por cada um.

A instalação do equipamento no animal, em geral, é realizada por meio de colares, salvo exceções, como nos trabalhos de Moreau et al. (2009), Scheibe e Gromann (2006) e Robert et

Tabela 1: Comparativo das características dos trabalhos relacionados.

			Sensores	
Trabalho	**Posição do Equipamento**	**Espécie Animal**	**Acelerômetro**	**GPS**
Diosdado et al. (2015)	Pescoço	Bovinos	sim	não
Dutta et al. (2015)	Perna	Bovinos	sim	sim
Guo et al. (2009)	Pescoço	Bovinos	sim	sim
González et al. (2015)	Pescoço	Bovinos	sim	sim
Jesus (2014)	Pescoço	Bovinos	não	sim
Martiskainen et al. (2009)	Pescoço	Bovinos	sim	não
Moreau et al. (2009)	Pescoço, Dorso e Costa	Caprinos	sim	sim
Oliveira (2013)	Pescoço	Bovinos	não	sim
Robert et al. (2009)	Perna	Bovinos	sim	não
Scheibe e Gromann (2006)	Pescoço e Perna	Bovinos e Equinos	sim	não
Watanabe et al. (2005)	Pescoço	Felinos	sim	não

al. (2009): os dois primeiros buscaram comparar os resultados da aferição dos comportamentos a partir dos dados coletados de diferentes partes do corpo do animal; e o último tinha interesse em capturar a movimentação de pernas dos animais.

Os principais sensores utilizados são os de movimentação, com destaque para o acelerômetro tridimensional, e o GPS. Dos 11 trabalhos avaliados apenas 4 fizeram a combinação dos dois: Moreau et al. (2009), González et al. (2015), Dutta et al. (2015) e Guo et al. (2009).

Com exceção dos trabalhos de Moreau et al. (2009) e Watanabe et al. (2005), todos monitoraram o comportamento bovino. Os comportamentos comumente classificados são os naturais do animal: Pastando, Parado (Em Pé ou Deitado) e Andando. Dutta et al. (2015), González et al. (2015) e Martiskainen et al. (2009) também monitoraram o comportamento Ruminando.

3 MATERIAL E MÉTODOS

A metodologia está separada em três sub-seções, para detalhar cada uma das partes desenvolvidas no trabalho: os detalhes dos colares utilizados; as informações do trabalho em campo; e a metodologia utilizada para aferir os comportamentos.

3.1 Colares

O colar é uma solução de hardware, composta por um módulo GPS, responsável por capturar os dados de posicionamento do animal baseado nos dados do satélite GPS e uma placa de medição inercial triaxial, composto pelo chip LSM303D - responsável pela captura do movimento de aceleração e de medições magnéticas - e o chip L3GD20H - responsável pela captura dos movimentos de rotação. Para armazenar os dados foi utilizado o *OpenLog*, um *data logger* que permite a gravação de até 16 GB de dados.

Para controlar os sensores e processar as coletas foi utilizado o Arduino, uma plataforma *open-source* de prototipagem eletrônica de hardware livre baseada em uma placa microcontroladora. O modelo utilizado foi o Arduino Fio, que além de atender aos requisitos do trabalho, suporta o protocolo ZigBee, desenvolvido pela *ZigBee Aliance* a partir do padrão IEEE 802.15.4. O protocolo foi projetado para ser utilizado em aplicações de sistemas embarcados,

possibilitando a formação de uma rede de sensores sem fio e a transmissão *on-line* dos dados.

Para integrar todos os itens que formam o colar, foi construída uma placa com o circuito. Foi utilizada uma placa de fenolite, com soquetes para encaixar os componentes e encapsulada em uma caixa plástica. A caixa foi fixada em um cinto de couro (com espessura suficiente para evitar que arrebentasse caso o animal enroscasse em um galho ou na cerca) utilizando uma fita constituída de filme de polietileno e trama de tecido com adesivo à base de resina e borracha, servindo para protegê-la contra eventuais impactos e sujeiras.

Quatro unidades do colar foram confeccionadas e cada uma colorida em uma cor. A coloração possibilita distingui-los visualmente, quando utilizados simultaneamente em campo. Após a construção dos colares, foi realizada a etapa de campo, onde foram coletados os dados dos animais.

3.2 Relato da Experiência em Campo

A área em que o experimento foi realizado - cedido pela Embrapa Gado de Corte - possui uma área de 4,09 hectares e estava com oito novilhas soltas. Na área há dois mangueiros para o manejo dos animais, que foram mantidos apascentados em pastagem cultivada e com suplementação mineral adequada, água à vontade e tratamento sanitário, pelos profissionais da Embrapa.

Para avaliar o funcionamento dos colares, foram realizados dois períodos de testes em campo, cada um com dois animais selecionados aleatoriamente no rebanho. Os bovinos foram equipados com o colar, como mostra a Figura 1.



Figura 1: Instalação do colar realizada com auxílio de especialistas.

O primeiro período foi utilizado para validar o funcionamento dos colares. Foram mais de 200 horas de coleta de dados em um período de 10 dias. Os resultados serviram para ajustar detalhes no encapsulamento do circuito, do posicionamento do colar no animal e avaliar o tempo de vida útil da bateria.

O segundo período soma mais de 500 horas de dados coletados, em um intervalo de aproximadamente 25 dias. Neste experimento, foram realizadas observações *in loco* (aproximadamente 70 horas), para rotular os dados com os comportamentos.

Nesta fase do experimento, o interesse era formar um *dataset* com os dados dos colares (registros dos sensores) e os comportamentos dos animais (observados e anotados pelo especialista), para criar o modelo de classificação dos comportamentos. A coleta e gravação dos dados eram realizadas em intervalos de 1 segundo.

Simultaneamente, foi realizada a rotulação dos dados, indicando quais os comportamentos dos animais. Os principais comportamentos bovinos apontados na literatura são os comportamentos Pastando/Procurando, Andando, Em Pé e Deitado (KILGOUR, 2012). O modelo de classificação aqui apresentado considerou estes comportamentos, que possuem as seguintes características:

- Pastando/Procurando: caracterizado pelo animal sobre as quatros patas, com a cabeça baixa procurando ou mastigando o capim. O animal pode ou não estar em movimento, já que ele pode estar se deslocando à procura de capim;
- Andando: o animal também está sobre as quatros patas, porém com o pescoço reto (apontando o focinho para frente) e se deslocando pela área de pasto;
- Em Pé: o animal está sobre as quatro patas, com a cabeça erguida e não há deslocamento;
- Deitado: o animal está com as patas abaixadas e com a barriga tocando o solo.

Para construir um modelo é necessário que um conjunto de dados, com o atributo alvo rotulado e denominado conjunto de treinamento, seja processado com um algoritmo de classificação. O modelo então é validado com um conjunto de teste, com registros com rótulos de classes desconhecidos.

Nos trabalhos de análise do comportamento animal a rotulação dos dados é geralmente feita por meio de anotações realizadas em campo ou por análise de vídeos (gravados durante o experimento), como fizeram Scheibe e Gromann (2006), Martiskainen et al. (2009) e Nadimi et al. (2012). Os dados registrados em manuscritos são transcritos para um arquivo digital, para então serem processados pelos algoritmos que fazem o reconhecimento dos padrões.

A utilização de softwares para rotular os *datasets*, além de reduzir o tempo da atividade, também elimina a possibilidade de erros na transcrição das anotações manuais para o arquivo digital. O trabalho de Dutta et al. (2015) utilizou um software chamado *WhatISee*, com o objetivo de registrar a data, horário e um rótulo para cada observação realizada. Trata-se de um aplicativo com função genérica, não desenvolvido especificamente para observação do comportamento animal.

Um software para automatizar a rotulação dos dados foi desenvolvido, aperfeiçoando o trabalho de observação, importação e manipulação dos dados obtidos em campo. A ferramenta, além de reduzir o tempo de rotulação dos dados, também elimina a possibilidade de erros na transcrição das anotações manuais para o arquivo digital. Na Seção 3.2.1 são apresentados os detalhes do software.

3.2.1 Ferramenta para Auxílio na Rotulação dos Dados

A ferramenta desenvolvida permite que o usuário: (1) escolha a quantidade de animais que se deseja observar; (2) escolha a quantidade de comportamentos a serem observados; (3) personalize o rótulo dos comportamentos observados; e pode ser utilizada em qualquer dispositivo móvel que possua o sistema operacional Android. Na Figura 2 são apresentadas duas telas do aplicativo.

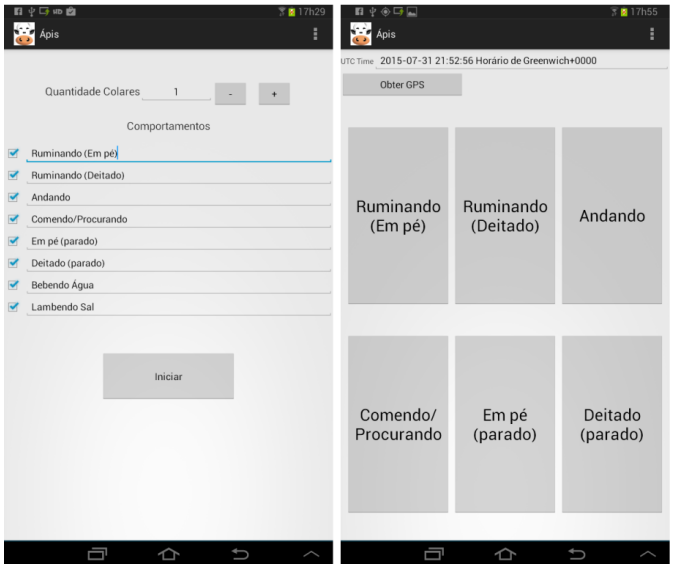


Figura 2: Telas do software para auxílio na rotulação dos dados.

Ao iniciar o software, ele sincroniza o aparelho com os satélites GPS, apresentando a data e horário do satélite. Obter o tempo diretamente do satélite é um fator importante para garantir a sincronização dos dados coletados pela aplicativo e os gerados pelos sensores que estão no colar acoplado no pescoço do animal. A data/horário é o elo que une a observação do analista e os dados gerados no colar.

Ao selecionar um animal e um comportamento, o aplicativo registra a data e horário (do satélite GPS), o ID do animal selecionado e o comportamento observado. Além de ser utilizada em conjunto com o colar, a ferramenta pode ser útil a outros trabalhos de observação animal, já que é possível a personalização dos comportamentos e da quantidade de animais observados, e permitir o acesso aos dados gravados em um arquivo de texto.

3.3 Software para Aferição dos Comportamentos

A interpretação manual dos dados obtidos nos colares é complexa, considerando o alto volume gerado. Os objetos (padrões) podem ser categorizados através de algoritmos computacionais, que se baseiam na extração de características que discriminam os objetos (THEODORIDIS; KOUTROUMBAS, 2006).

Foi utilizado o reconhecimento de padrões para aferir os comportamentos dos bovinos, baseado no modelo de aprendizagem supervisionada. Neste modelo, um rótulo da categoria é

fornecido para cada registro em um conjunto de treinamento e o algoritmo busca aprender uma solução para rotular instâncias desconhecidas (DUDA; HART; STORK, 2000). A classificação é definida como a tarefa de rotular cada registro em uma das categorias pré-definidas (TAN; STEINBACH; KUMAR, 2009). Cada registro é caracterizado por uma tupla (x,y), onde x é o conjunto de atributos e y o atributo especial, chamado de atributo alvo ou de categorização. Uma função alvo (modelo) mapeia cada conjunto de atributos x para um dos rótulos de classes y pré-determinados.

Para a construção e validação dos modelos de classificação foi utilizado o pacote WEKA (*Waikato Environment for Knowledge Analysis*), que é uma coleção de algoritmos de aprendizado de máquina, de código aberto, que pode ser aplicado para tarefas de mineração de dados. Ele é desenvolvido na linguagem Java, contém uma GUI (*Graphical User Interface*) para interagir com arquivos de dados e produzir resultados visuais. Também é possível usar as suas funcionalidades a partir da sua API (*Application Programming Interface*) em um código Java.

Os dados gerados nos colares foram utilizados na construção de dois modelos de predição do comportamento, utilizando os algoritmos de classificação *Random Forest* e o algoritmo SVM: o arquivo com os registros coletados possui 13.088 instâncias. A acurácia dos modelos foram avaliadas utilizando o método *cross-validation* com 10 *folds*. Neste método o conjunto de dados é dividido em 10 partes de tamanhos aproximados, nos quais 9 partes são utilizadas como conjunto de treinamento e a parte restante como conjunto de teste. O processo repete-se 10 vezes, até que todos os *folds* sejam utilizados como conjunto de teste (TAN; STEINBACH; KUMAR, 2009).

Para avaliar o impacto/variação que cada colar e animal possui no modelo construído, agrupou-se os dados em outros conjuntos, chamados de conjuntos combinados. Eles foram gerados, a partir dos conjuntos individuais, em dois cenários: conjuntos de dados agrupados por animal (três animais foram utilizados); e conjuntos de dados agrupados por colar (quatro colares foram utilizados).

Para avaliar os classificadores gerados com os conjuntos de dados combinados, foi utilizado o método *Holdout*, em um conjunto de teste formado por um extrato de 15% dos registros de cada conjunto (1.975 instâncias), selecionados aleatoriamente nos arquivos, com o cuidado apenas para que as classes não ficassem desbalanceadas.

4 RESULTADOS E DISCUSSÃO

Com o algoritmo *Random Forest*, o modelo de classificação obteve acurácia de 93,99%: em um universo de 13.088 instâncias, 12.301 foram classificadas corretamente e 787 incorretamente. As classes identificadas foram: Em Pé (acurácia de 88,6%), Deitado (acurácia de 93,4%), Andando (acurácia de 95%) e Pastando/Procurando (acurácia de 96,4%).

Com o algoritmo SVM, utilizando a configuração *default* do WEKA, o modelo apresentou 43,79% de acurácia. Considerando as possibilidades de ajustes nas configurações do kernel

utilizado no SVM, os parâmetros C (*cost*) e y (*gamma*) foram ajustados, utilizando o método *grid search*. Com o kernel *radial basis function* ajustado, o classificador apresentou acurácacia de 93,2%. Do total de 13.088 registros, 12.198 foram corretamente classificados, enquanto 890 foram classificadas incorretamente. A classe Pastando/Procurando obteve 94,3% de acurácia, enquanto Deitado obteve 95,5%, Em Pé 90,1% e Andando 87,1%.

O modelo obteve melhor resultado se comparado ao trabalho de Jesus (2014), que classificou o comportamento de vacas utilizando o algoritmo END. A tabela 2 compara o resultado de cada classe entre os trabalhos. Os resultados dos comparativos apresentados servem apenas como base para indicar os valores obtidos neste trabalho, já que os conjuntos de dados são diferentes, assim como o número de animais e o tempo de observação.

Tabela 2: Acurácia de cada classe comparada ao trabalho de Jesus (2014).

	Acurácia	
Classe	Jesus (2014)	Esta pesquisa
Pastando/Procurando	73,1%	96,4%
Andando	87,5%	95%
Em Pé	56,9%	88,6%
Deitado	84,6%	93,4%

Outra avaliação realizada, foi verificar o quanto um animal ou um colar influenciam na acurácia da classificação, o modelo foi aferido utilizando os dados individuais de cada animal e colar. O resultado mostra que ao retirar os dados de um animal ou de um colar há uma variação considerável na acurácia, conforme mostra a Tabela 3.

Os resultados apontam a possibilidade de generalização do uso do equipamento e do modelo de classificação, porém a acurácia precisa ser melhor avaliada, em um experimento que utilize um número maior de colares e animais.

Tabela 3: Resultado da validação dos conjuntos agrupados por animal utilizando o método *Holdout*.

Combinação	Acurácia (%)
Animal 1 + 2 + 3	94,7341
Animal 1 + 2	92
Animal 1 + 3	90,9367
Animal 1	87,8481
Animal 2 + 3	87,3417
Animal 2	83,3417
Animal 3	82,4810

5 CONCLUSÕES

O foco dos trabalhos tem sido identificar os comportamentos e indicar a localização do animal na área de pasto, buscando validar se é possível ou não aferir os comportamentos por meio de sensores. A proposta deste trabalho foi desenvolver uma solução de hardware e de software, para coletar os dados e aferir os comportamentos bovinos. Além deles, foi desenvolvida uma ferramenta para rotulação dos dados (em casos que há necessidade de registro manual dos comportamentos).

Os resultados da identificação automática dos comportamentos foram superiores aos apresentados por Jesus (2014), indicando que incluindo os dados de movimentação do animal (além dos dados de posicionamento) é possível melhorar a acurácia das classificações.

Quanto ao procedimento utilizado no experimento relatado, outros trabalhos podem monitorar um número maior de animais e colares, além de configurar a rede de sensores sem fio, para que os dados coletados possam ser enviados automaticamente a uma estação base.

Por fim, a classificação automática dos comportamentos combinada à outras variáveis, podem ser úteis na aferição de novos comportamentos, que indiquem variações nos comportamentos habituais do animal, como um animal potencialmente doente ou com problemas com a mosca-dos-chifres. Para entender o comportamento dos animais e sua interação com o ambiente é necessário monitorar outras variáveis - tais como, temperatura, umidade relativa, pressão barométrica, radiação solar, velocidade do vento, direção do vento, precipitação - sendo possível adicioná-los ao sistema, seja diretamente no colar ou por meio de estação que se comunique com os colares pela rede sem fio. Além deles, sistemas já existentes que façam a captura da temperatura corporal, identificação RFID e mensuração automática do peso de cada animal, também podem ser agregados.

AGRADECIMENTOS

Os autores agradecem à FUNDECT (Termo de Outorga 102/2014), pelo apoio financeiro concedido ao trabalho.

REFERÊNCIAS

BRASIL. *Produto Interno Bruto da agropecuária deve ser de R$ 1,1 trilhão*. dec 2014. Disponível em: <http://www.agricultura.gov.br/comunicacao/noticias/2014/12/produto-interno-bruto-da-agropecuaria-deve-ser-de-rs-1-trilhao>. Acesso em: 27 abr. 2015.

CÁCERES, E. N. et al. Computational precision livestock - position paper. In: *II Workshop of the Brazilian Institute for Web Science Research*. Rio de Janeiro: [s.n.], 2011.

DIOSDADO, J. V. et al. Classification of behaviour in housed dairy cows using an accelerometer-based activity monitoring system. *Animal Biotelemetry*, v. 3, n. 1, p. 15, 2015. ISSN 2050-3385. Disponível em: <http://www.animalbiotelemetry.com/content/3/1/15>.

DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern Classification*. 2. ed. New York: Wiley-Interscience, 2000.

DUTTA, R. et al. Dynamic cattle behavioural classification using supervised ensemble classifiers. *Computers and Electronics in Agriculture*, v. 111, p. 18 – 28, 2015. ISSN 0168-1699. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0168169914003123>.

GONZÁLEZ, L. et al. Behavioral classification of data from collars containing motion sensors in grazing cattle. *Computers and Electronics in Agriculture*, v. 110, n. 0, p. 91 – 102, 2015. ISSN 0168-1699. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0168169914002798>.

GUO, Y. et al. Using accelerometer, high sample rate gps and magnetometer data to develop a cattle movement and behaviour model. *Ecological Modelling*, Elsevier, v. 220, n. 17, p. 2068 – 2075, 2009. Disponível em: <http://eprints.qut.edu.au/33751/>.

IICA. *New tools for precision livestock management*. [S.l.], 2009. Disponível em: <http://repiica.iica.int/docs/B1632I/B1632I.pdf>.

JESUS, L. *Identificação do Comportamento Bovino por meio do Monitoramento Animal*. Dissertação (Mestrado) — Universidade Federal de Mato Grosso do Sul, Campo Grande, 2014.

KILGOUR, R. J. In pursuit of "normal": A review of the behaviour of cattle at pasture. *Applied Animal Behaviour Science*, v. 138, n. 1–2, p. 1 – 11, 2012. ISSN 0168-1591. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0168159111004229>.

MARTISKAINEN, P. et al. Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. *Applied Animal Behaviour Science*, v. 119, n. 1–2, p. 32 – 38, 2009. ISSN 0168-1591. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0168159109000951>.

MOREAU, M. et al. Use of a tri-axial accelerometer for automated recording and classification of goats' grazing behaviour. *Applied Animal Behaviour Science*, v. 119, n. 3–4, p. 158 – 170, 2009. ISSN 0168-1591. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0168159109001324>.

NADIMI, E. et al. Monitoring and classifying animal behavior using zigbee-based mobile ad hoc wireless sensor networks and artificial neural networks. *Computers and Electronics in Agriculture*, v. 82, n. 0, p. 44 – 54, 2012. ISSN 0168-1699. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0168169911003206>.

OLIVEIRA, M. T. P. de. *Análise Comportamental de Bovinos Baseada em Trajetórias Semânticas Aplicada à Pecuária de Precisão*. Dissertação (Mestrado) — Universidade Federal de Mato Grosso do Sul, Campo Grande, 2013.

ROBERT, B. et al. Evaluation of three-dimensional accelerometers to monitor and classify behavior patterns in cattle. *Comput. Electron. Agric.*, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 67, n. 1-2, p. 80–84, jun. 2009. ISSN 0168-1699. Disponível em: <http://dx.doi.org/10.1016/j.compag.2009.03.002>.

RUIZ-GARCIA, L.; LUNADEI, L. The role of RFID in agriculture: Applications, limitations and challenges. *Computers and Electronics in Agri-*

culture, v. 79, n. 1, p. 42 – 50, 2011. ISSN 0168-1699. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0168169911001876>.

SCHEIBE, K.; GROMANN, C. Application testing of a new three-dimensional acceleration measuring system with wireless data transfer (WAS) for behavior analysis. *Behavior Research Methods*, Springer-Verlag, v. 38, n. 3, p. 427–433, 2006. ISSN 1554-351X. Disponível em: <http://dx.doi.org/10.3758/BF03192796>.

TAN, P.-N.; STEINBACH, M.; KUMAR, V. *Introdução ao DATAMINING Mineração de Dados*. Rio de Janeiro, RJ, Brasil: Editora Ciência Moderna, 2009. ISBN 9788573937619.

THEODORIDIS, S.; KOUTROUMBAS, K. *Pattern Recognition*. 3. ed. San Diego: Academic Press, 2006.

WATANABE, S. et al. A new technique for monitoring the detailed behaviour of terrestrial animals: A case study with the domestic cat. *Applied Animal Behaviour Science*, v. 94, n. 1–2, p. 117 – 131, 2005. ISSN 0168-1591. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0168159105000353>.

Javalix: um sistema para acompanhamento de malhadas de javalis

Diego de Carvalho Neves da Fontoura, Diego Mengarda, Fabiane Nunes Prates Camargo, Leandro da Silva Camargo*, Leonardo Bidese de Pinho.
Programa de Pós-graduação em Computação Aplicada, Universidade Federal do Pampa e Embrapa Pecuária Sul, Bagé, Rio Grande do Sul, Brasil
*e-mail: leandrocamargo@ifsul.edu.br

Definição do Problema

A presença de javalis nas propriedades rurais localizadas nas mais diversas regiões do Brasil tem sido alvo de preocupação para agricultores e organizações ligadas ao meio ambiente. O animal está entre as cem piores espécies invasoras do mundo, onde sua agressividade e facilidade de adaptação, associadas à reprodução descontrolada e à ausência de predadores naturais, resultam em uma série de impactos ambientais e socioeconômicos. Dentre os problemas ocasionados pela presença dos javalis, estão a predação de animais de pequeno porte, a devastação de lavouras e a transmissão de doenças.

Justificativa

O Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (Ibama), desde 2013, autoriza a caça desses animais como forma de controle populacional. De acordo com o Art. 37, Inciso II, da Lei n° 9.605 (Brasil, 1998), não é crime o abate do animal, quando realizado para proteger lavouras, pomares e rebanhos da ação predatória ou destruidora de animais, desde que legal e expressamente autorizado pela autoridade competente. Assim, o registro de ocorrências, levantados através de um aplicativo, poderá contribuir para o controle de malhada de javalis, constituindo-se num importante instrumento de auxílio aos produtores das regiões afetadas.

Desenvolvimento

O Javalix trata-se de um aplicativo desenvolvido para dispositivos móveis capaz de registrar o movimento de malhadas de javalis a partir de registro voluntário de usuários. Dentre suas funcionalidades estão: inclusão de registros e locais favoritos e visualização de mapa de malhadas registradas no sistema. A partir do Sensor de Posicionamento Global (GPS) ativo, o aplicativo adquire de forma automática a localização por meio da latitude e longitude. Além disso, em caso de indisponibilidade de conexão com os satélites do GPS, o usuário tem acesso a um mapa, de forma desconectada (off-line), podendo apontar manualmente a posição do local. Para implementação do aplicativo utilizou-se o Ionic, um framework que possibilita a construção de aplicações móveis utilizando HTML, CSS e Javascript para criação das interfaces. A conexão entre aplicativo e servidor para a transferência de dados foi feita através do protocolo Hypertext Transfer Protocol (HTTP), utilizando uma RESP API para operações com o Banco de Dados. A estrutura suporta até 10.000 requisições simultâneas que atenderão a necessidade inicial da implementação do projeto. O servidor web foi desenvolvido em PHP com o framework Laravel e o aplicativo acessa esse serviço utilizando Ajax, uma tecnologia que torna o aplicativo bastante dinâmico e com maior capacidade de respostas.

Considerações Finais

O aplicativo foi desenvolvido levando em consideração as limitações técnicas de disponibilidade de sinal em áreas afastadas do perímetro urbano, bem como, das restrições de uso da tecnologia apresentadas por uma parcela de usuários. Trata-se de uma aplicação simples e prática, capaz de disponibilizar informações relevantes acerca da presença de javalis e contribuir no controle populacional desses animais.

Referências

BRASIL. Lei n° 9.605 de 12 de fevereiro de 1998. Dispõe sobre as sanções penais e administrativas derivadas de condutas e atividades lesivas ao meio ambiente, e dá outras providências.1998. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l9605.htm. Acesso em 10 de Julho de 2017.

Aplicativo móvel para análise de conforto térmico de ambientes

Arilson José de Oliveira Júnior, Silvia Regina Lucas de Souza
Faculdade de Ciências Agronômicas, FCA, UNESP, Botucatu/SP
*e-mail: arilsonjr@outlook.com

Definição do Problema

Análises de confortabilidade térmica são essenciais para projetos de instalações nos dias atuais. A adequação de ambientes a aspectos climáticos são de extrema importância, pois, de fato, há uma grande relação entre a variabilidade térmica dos ambientes (galpões, indústrias, residências e escritórios) e a eficiência e qualidade produtiva tanto dos animais, como dos seres humanos. O conforto térmico e a ambiência são áreas de estudo que tem por objetivo conceder o bem-estar para indivíduos, sejam eles animais ou pessoas. Nos dias atuais grande parte das análises de conforto térmico no meio agrícola, ressaltando às em ambientes de produção animal, utilizam índices que possibilitam a avaliação do nível de estresse térmico de um determinado local por meio de variáveis climatológicas, como o Índice de Temperatura e Umidade – ITU e o Índice de Temperatura de Globo Negro e Umidade – ITGU. Do mesmo modo, em locais de trabalho humano há um limite de tolerância para exposição ao calor que deve ser monitorado mediante o chamado Índice de Bulbo Úmido Termômetro de Globo – IBUTG. Para a resolução desses índices, não há atualmente um sistema computacional específico que forneça *in loco* a condição térmica de um ambiente.

Justificativa

O aplicativo móvel desenvolvido permite que pequenos produtores, pesquisadores e demais profissionais avaliem (*in loco*) as condições térmicas de ambientes mediante o cálculo dos índices de conforto térmico específicos para seres humanos e animais. Tal ferramenta torna possível a realização de ações imediatas a fim de mitigar o estresse térmico. Esse tipo de tecnologia proporciona um nível de portabilidade e autonomia ainda não presentes nas avaliações de condições térmicas de ambientes no mercado, que atualmente são baseadas no armazenamento de medidas em *data loggers* e posterior análise em *softwares* proprietários.

Desenvolvimento

O aplicativo foi desenvolvido para o sistema operacional *Android* (a partir da versão 4.4) em conjunto com a criação de um dispositivo portátil para coleta e transmissão de variáveis climatológicas, como temperatura do ar, temperatura de globo negro e umidade relativa. Para o dispositivo portátil utilizou-se um microcontrolador padrão *Arduino*, em conjunto com um sensor de temperatura e umidade relativa do ar *AM2302* e um módulo *Bluetooth BC 417 HC-06*. Para o desenvolvimento do aplicativo foi utilizado a linguagem de programação orientada a objetos *Java* em ambiente de desenvolvimento integrado *Android Studio 2.3.3*. O aplicativo permite estabelecer uma conexão *Bluetooth* com o dispositivo móvel, assim o dispositivo portátil envia as medidas climatológicas para o aplicativo no qual são definidos previamente os intervalos de aquisição dos dados, possibilitando assim a escolha de um indivíduo para análise de conforto térmico. O aplicativo também permite que o usuário insira os dados de forma manual com o uso de outros instrumentos de medida presentes no local, ou mediante a obtenção de dados coletados previamente.

Considerações Finais

Resultados positivos já foram obtidos por meio de testes de validação do dispositivo portátil e das funcionalidades presentes no aplicativo em ambiente real de criação animal e em escritórios. Vale considerar que o aplicativo vem sendo aprimorado de acordo com as mais recentes pesquisas sobre conforto térmico de animais e seres humanos.

ClimaZone: Sistema integrado de zoneamento climático

Bruno José Alves, Elvis William de Oliveira Barbieri
Grupo de Sistemas de Informação, Embrapa Informática Agropecuária
e-mail: jalvesbruno@gmail.com, elviswob@gmail.com

Definição do Problema

A agricultura é uma das atividades econômicas mais dependentes das condições do tempo e do clima. Diversas práticas, desde a escolha do tipo de solo e do ciclo utilizado no mesmo, entre outras, tem influência em períodos que são considerados favoráveis para o plantio de uma determinada cultura.

Sendo assim, o uso de tecnologias que possibilitem o acesso a essas informações e que possam auxiliar o agricultor em suas decisões tornam-se cada vez mais necessárias.

Justificativa

Tendo em vista o desenvolvimento das tecnologias de informação e comunicação, e a forma na qual essas desempenham papel transformador na sociedade, levando informação em tempo real, simplificando processos, diminuindo custos e tornando mais eficiente diversos segmentos da indústria e do comércio, torna-se cada vez mais evidente a necessidade de integrarmos essas tecnologias aos setores que agregam grande valor econômico ao país, como é o caso da agricultura e pecuária.

Desta forma, processos e informações fundamentais para tomada de decisão podem ser consultados a qualquer momento por meio de dispositivos móveis conectados através da internet, trazendo assim confiabilidade e segurança para aqueles que desejam empreender nesses setores.

Desenvolvimento

O ClimaZone é um aplicativo mobile que oferece aos usuários informações sobre o zoneamento de risco climático para diversas culturas. Tal zoneamento, mostrado por meio de um gráfico, ocorre quando o usuário informa o estado, a cidade, a cultura desejada, o ciclo e o solo.

Neste gráfico, cada período é organizado em um intervalo de dez dias (decêndio). Os decêndios marcados com verde indicam que o período considerado é favorável para o plantio, enquanto que os decêndios marcados em vermelho, indicam um período desfavorável para o plantio.

O aplicativo conta com uma base de dados extraída do sistema Web Agritempo[1], com relação ao zoneamento agrícola por tabelas. Este sistema foi desenvolvido e mantido pela Embrapa Informática Agropecuária, empresa que atua no desenvolvimento e pesquisa agropecuária. O aplicativo foi desenvolvido para a plataforma Android, utilizando a linguagem de programação Java e a linguagem de marcação XML, na IDE Android Studio.

Considerações Finais

O aplicativo móvel ClimaZone realiza a consulta de períodos favoráveis e desfavoráveis para o plantio de aproximadamente 100 culturas.

Referências

[1] http://www.agritempo.gov.br/agritempo/index.jsp

SBIAgro 2017
02 a 06 de outubro 2017
Campinas - SP

Fruit Fly Classification

Matheus M. Leonardo, Carlos A. B. Ribeiro, Fabio A. Faria
Universidade Federal de São Paulo, Grupo para Inovação baseada em Imagens e Sinais
*e-mail: ffaria@unifesp.br

Problema

Moscas-de-frutas são de grande importância biológica e econômica para a agricultura de diferentes países no mundo, especialmente para o Brasil. O Brasil é o terceiro maior produtor de frutas do mundo, com 44 milhões de toneladas em 2016. As perdas diretas e indiretas causadas por moscas-de-frutas podem exceder dois bilhões de dólares por ano, colocando essas pragas como uns dos maiores problemas na agricultura mundial. No Brasil, estima-se que as perdas econômicas diretamente relacionadas à produção, custo de controle de praga e de exportação, seja entre 120 e 200 milhões de dólares por ano.

Justificativa

Entre as moscas-de-frutas de maior importância econômica nas Américas, está às espécies de *Anastrepha*. Esse gênero é mais diversificado nos trópicos e subtrópicos da América, com aproximadamente 300 espécies conhecidas, sendo 120 registradas no Brasil. Entretanto, poucas espécies possuem importância econômica, como a mosca-de-fruta sul-americana *Anastrepha fraterculus* (Wiedemann), a mosca de fruta do oeste da Índia *Anastrepha obliqua* (Macquart), e a mosca-de-fruta da goiaba *Anastrepha striata* Schiner. Essas três espécies são consideradas pragas de grande importância para muitas agências reguladoras. A identificação das espécies é algo crucial, tanto para o desenvolvimento de estudos na biologia quanto para o controle de pragas e redução de danos, quarentenas e distribuição de frutas. A identificação das espécies é baseada na morfologia das asas e na maior parte dos acúleos (a extremidade penetrante do órgão reprodutor feminino). Contudo, a distinção entre algumas espécies de moscas-de-frutas é bem difícel de ser encontrada. *Anastrepha Fraterculus* é o caso mais emblemático de um complexo de espécies crípiticas registradas nas Américas, pois ela é a praga principal somente em algumas áreas de sua ocorrência, que abrange do México ao norte da Argentina. O erro na identificação pode ser um problema sério para a realização de restições de quarentena, gerencimanto integrado de pragas e outros programas de controles de praga. Novos procedimentos para a identificação tem sido abordados, tais como análises morfométricas e moleculares, visando uma identificação mais precisa para as moscas-de-frutas do gênero *Anastrepha*. No entanto, poucos trabalhos na literatura foram propostos para identificar o gênero *Anastrepha* através de técnicas de processamento de imagem e aprendizado de máquina. Talvez esse fato esteja relacionado a alta similaridade entre espécies pertencentes ao gênero *Anastrepha*.

Desenvolvimento

Em nosso aplicativo para Android *mobile*, nós implementamos um sistema com a linguagem Python e adotamos uma biblioteca *open-source* de aprendizado de máquina, chamada TensorFlow, que foi originalmente desenvolvida por pesquisadores e engenheiros da Google (Google Brain Team, dentro da Google's Machine Intelligence research organization). Posteriormente, nós treinamos uma arquitetura de aprendizagem profunda (Google's Inception) com imagens de asas de moscas-de-frutas (301 amostras), de três espécies do gênero Anastrepha.

Considerações finais

Neste trabalho, nós propomos aplicar aprendizagem profunda para a tarefa de reconhecimento de moscas-de-frutas. Entretanto, nosso aplicativo é uma versão inicial, e em nossos experimentos a abordagem proposta alcançou excelentes resultados de eficácia, quando comparado com técnicas mais recentes da literatura. Para trabalhos futuros, poderemos realizar experimentos com outras espécies de moscas-de-frutas e usar uma maior base de imagens. No futuro, nós pensamos em aprimorar a interface, para que ela se torne mais amigável ao usuário e adicionando técnicas de processamento de imagens.

Material Adicional

http://www.ict.unifesp.br/ffaria

SBIAgro 2017
02 a 06 de outubro 2017
Campinas - SP

Avaliação do Bem-estar Animal no Transporte de Gado de Corte

Marcel da Silva Camargo*[,1], Ana Paula Lüdtke Ferreira[1], Naylor Bastiani Perez[1,2]
[1]Programa de Pós-graduação em Computação Aplicada, Universidade Federal do Pampa - Bagé, Rio Grande do Sul, Brasil
[2]EMBRAPA Pecuária Sul – Bagé, Rio Grande do Sul, Brasil
*e-mail: mscamargo@gmail.com

Definição do Problema

A recente evolução na cadeia de abastecimento de carne tem sido bastante dinâmica em termos de melhorias na segurança alimentar, qualidade e ajustes em sistemas de produção alternativos. Os consumidores agora estão mais interessados sobre a origem animal, sistemas ecológicos, alimentação animal e manuseio correto durante o transporte e o abate. Vários relatórios confirmam que os consumidores estão muito preocupados com o bem-estar dos animais (LAMA et al., 2013).

A principal contribuição cientifica esperada deste trabalho é a possibilidade de predizer através do método proposto, que determinada forma de manejo, estrutura das instalações da propriedade rural ou a forma de transporte, pode afetar na qualidade do produto final. Gerando para os consumidores um produto que atenda processos mais éticos na cadeia produtiva. Exibindo a correlação entre variáveis fisiológicas e de ambiente com o bem-estar animal, que possam ser avaliadas e classificadas para determinar o rendimento e qualidade da carne.

Justificativa

Grande parte das perdas na cadeia produtiva de carne bovina, está relacionada com a falta de comprometimento com o bem-estar animal nos processos que envolvem o transporte dos animais para o abate. Assim, pretende-se investigar as possíveis consequências usando um modelo de avalição das etapas envolvidas no transporte dos animais.

Desenvolvimento

As variáveis operacionais obtidas durante o processo de transporte são: técnicas de manejo utilizadas e instalações de embarque do produtor, tipo do caminhão usado para o transporte, densidade de carga do caminhão, tempo de transporte da propriedade rural até o matadouro e tempo e condições de desembarque no matadouro. A variável densidade de carga é dependente do tipo de caminhão utilizado.

Para o desenvolvimento do método de avaliação, um conjunto de dados obtidos foi submetido à função naive-bayes, onde os preditores métricos são tratados assumindo a distribuição normal, dado o rótulo de cada coluna do conjunto de dados. Foi utilizada a função *predict*, a qual calcula as probabilidades condicionais posteriores para cada classe usando a regra de Bayes, projetada para análise das probabilidades marginais de variáveis preditoras, dada a classe.

O estudo revela que o classificador probabilístico baseado no teorema de Bayes fornece informações relevantes ao produtor rural, auxiliando na identificação de possíveis causas de baixo rendimento de sua produção.

Considerações Finais

Comparando com todas as etapas da cadeia produtiva de bovinos de corte, o transporte é uma das principais causas de perda de rendimento da produção. Assim, a identificação antecipada de problemas que podem causar estas perdas, faz-se necessária para o aumento de produtividade e consequente ganho financeiro. Neste trabalho, foram aplicadas técnicas de discretização de dados e classificação, com a intenção de predizer o rendimento da carcaça baseado em um conjunto de dados, coletados no processo de transporte.

Referências

GRANDIN, T. Animal welfare and society concerns finding the missing link. Meat Science, v. 98, p. 461–469, 2014.

LAMA, G. C. M. D. L. et al. Attitudes of meat retailers to animal welfare in spain. Meat Science, v. 95, p. 569–575, 2013.

MENDONÇA, F. S. et al. Pre-slaughtering factors related to bruises on cattle carcasses. Animal Production Science, 2016.

Cyber Leaf

Renan de Paula Rosa
Departamento de Informática -- Universidade Estadual de Ponta Grossa (UEPG) Campus Uvaranas -- Av. Carlos Cavalcanti -- 4748 -- 84.030-900 -- Ponta Grossa -- PR -- Brazil
e-mail: renanrosa@outlook.com

Definição do Problema

O aplicativo Cyber Leaf visa reconhecer lagartas que atacam lavouras, auxiliando assim tanto profissionais especializados na área, como produtores rurais, etc. Com o Cyber Leaf, qualquer pessoa poderá utilizar o aplicativo para reconhecer lagartas, sem a necessidade de conhecimento prévio na área ou de acompanhamento técnico.

Justificativa

O reconhecimento de lagartas normalmente é feito por profissionais da área, com o Cyber Leaf as demais pessoas interessadas poderão utilizar o aplicativo a fim de reconhecer lagartas. O Cyber Leaf visa o reconhecimento de lagartas por meio de técnicas de deep learning.

Desenvolvimento

A solução Cyber Leaf foi desenvolvida na arquitetura cliente-servidor. Através do aplicativo os usuários poderão enviar as imagens de lagartas para o servidor. O servidor irá processar as imagens e responder ao aplicativo as informações sobre a lagarta, caso ela tenha sido identificada. As imagens são processadas por técnicas de deep learning, o processamento é baseado na extração de características da imagem como cor, textura e forma.

Considerações Finais

O Cyber Leaf possui uma taxa de acerto elevada no reconhecimento de lagartas, isto faz dele uma solução confiável, embora a base de dados seja pequena.

SBIAgro 2017
02 a 06 de outubro 2017
Campinas - SP

TomBanana – Um aplicativo para previsão de colheita da bananeira cv. Tropical

Bruno Queiroz Santos, Franciscone Luiz de Almeida Junior, Guilherme da Silva Teixeira, Matheus Galvão Ferreira, Gabriel Marques de Melo
Departamento de Ciência da Computação
*e-mail: emakers.nemesis@gmail.com

Definição do Problema

A bananeira está entre as frutíferas tropicais mais consumidas no Brasil e representa grande importância para o Brasil em termos culturais, econômicos e sociais. Porém, um dos obstáculos referentes à cultura de banana é a ausência de variedades comerciais resistentes a pragas e doenças, adaptáveis a diferentes ambientes e de porte adequado [1]. O melhoramento genético traz uma solução viável se comparado com a cultura tradicional. Dentre os cultivares geneticamente melhorados podem ser observadas características – tanto da bananeira quanto do fruto em si – que influenciam a qualidade da safra. O objetivo deste trabalho é apresentar um aplicativo (TomBanana) que, por meio de técnicas de inteligência computacional (Redes Neurais Artificiais – RNAs [2]), é capaz de prever a colheita da bananeira cv. Tropical para auxiliar o produtor a realizar o melhoramento genético e previsão de faturamento.

Justificativa

Em geral, a previsão da colheita depende de conhecimento de especialistas. A principal funcionalidade do TomBanana é realizar a previsão do peso do cacho de banana, viabilizando, portanto, a previsão da rentabilidade da safra de forma rápida e acessível. Dessa forma, o aplicativo fornece um auxílio ao produtor na tomada de decisões referentes à viabilidade da bananeira, além de possibilitar a aquisição de financiamentos bancários com maior embasamento científico.

Desenvolvimento

O *design* do aplicativo foi elaborado de forma intuitiva onde, basicamente, é fornecida pelo usuário uma sequência de valores de determinadas características da bananeira e, por fim, a RNA estima o peso do cacho de banana, que é exibido ao usuário. A arquitetura da RNA foi feita baseada em [1] e foi implementada considerando a matriz de pesos da RNA já treinada. Sendo assim, neste trabalho até o presente momento, foi implementada somente a etapa *forward* da RNA. As variáveis de entrada consideradas são: peso da ráquis, comprimento do engaço, diâmetro do engaço, peso da segunda penca, número total de pencas por cacho, número de frutos por cacho, peso do fruto, comprimento externo do fruto, diâmetro ou calibração lateral do fruto e número de folhas vivas na colheita. E a variável de saída é o peso do cacho. O aplicativo foi desenvolvido em Java utilizando o *AndroidStudio* e foi compilado para a versão 4.4 do Android.

Considerações Finais

Pode-se ressaltar que não foram encontradas soluções comerciais parecidas com a proposta, e que o principal concorrente a essa aplicação são os próprios profissionais especialistas na área. Porém, o aplicativo visa oferecer uma solução mais econômica e acessível ao produtor para auxiliar e não para substituir o profissional na área.

Referências

[1] SOARES, J. D. R. Modelagem matemática na previsão de colheita da bananeira: regressão linear múltipla x redes neurais artificiais. Tese de doutorado. UFLA, Lavras, 2011.

[2] BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDERMIR, T. B. Redes Neurais Artificiais: Teoria e Aplicações. 2ª Edição. Rio de Janeiro, LTC, 2007.

Análise multitemporal do arroz usando NDVI de superfície

Cassiane Jrayj de Melo Victoria Bariani[2] *,Nelson Mario Victoria Bariani*[1]*, José Rodrigo Fernandes Caresani*[3]*, Roberto Dutra de Felice*[1]

[1] Universidade Federal do Pampa, Itaqui, Rio Grande do Sul, vbariani@gmail.com

[2] Universidade Federal do Pampa, Itaqui, Universidade Federal de Santa Maria, Cachoeira do Sul, Rio Grande do Sul, cassiane.victoria@gmail.com

[3]Universidade Estadual do Rio Grande do Sul, Campus São Borja, jose-caresani@uergs.edu.br

RESUMO

A data de plantio e a cultivar selecionada são meios usados pelo agricultor para manipular a duração do ciclo e o cronograma das fases fenológicas da lavoura orizícola, sendo assim variáveis importantes na caracterização dos ciclos de lavoura em cada região em determinada safra, podendo ser adequadamente determinados por sensoriamento remoto. Foram usadas imagens de satélite para uma análise multitemporal de lavouras de arroz, safras 1984/85, 2006/07 e 2016/2017. As imagens foram processadas independentemente para cada data no Spring SIG do INPE v.5.2., no qual os ND (Números Digitais) foram transformados em FRB (Fatores de Reflectância Bidirecional) aparente usando um algoritmo escrito em linguagem LEGAL; também foi realizada a correção atmosférica das imagens. O NDVI foi calculado a partir dos FRB de superfície corrigidos para as imagens Landsat, e a partir dos ND originais para outros satélites, para fins de comparação da definição temporal do ciclo. Os talhões de arroz correspondentes à órbita 224/80 da série Landsat foram identificados por classificação supervisionada, assistida por informações complementares locais para confirmação; os limites foram posteriormente vetorizados. Estatísticas descritivas das imagens NDVI foram obtidas para cada talhão e os valores médios foram graficados em função do tempo. Métodos alternativos para geração dos perfis de NDVI em função do tempo para os talhões selecionados também foram utilizados: Multispec 3.1 e o pacote dtwSat do R. Os perfis multitemporais foram agrupados em classes: i) Semeadura e colheita precoce (ciclo precoce e muito precoce); ii) Semeadura precoce e colheita tardia (ciclo médio e longo); iii) semeadura tardia (ciclo precoce ou médio). As diferenças constatadas entre os perfis multitemporais leva a entender que a introdução de novas cultivares de arroz inseridas desde meados da década de 1980 até 2016 provocou mudanças na proporção de cada tipo de arroz presente na região. As variedades identificadas como de semeadura e colheita precoce está presentes em maior proporção na safra 2006/07 comparado com as outras. Na safra 2016/17 percebeu-se um aumento significativo das cultivares de ciclo médio. Informações da industria arrozeira regional corroboram e justificam esta tendência.

PALAVRAS-CHAVE: Sensoriamento Remoto Multiespectral, Perfis temporais, Índice de Vegetação.

ABSTRACT

The duration of the cycle of rice cultivation is a variable of interest in the characterization and quantification of the most used cultivars in each region in a given crop year, and can be adequately performed by remote sensing. Satellite images were used for a multitemporal analysis of rice crops, 1984/85, 2006/07 and 2016/17. The images were processed independently for each date in the Spring GIS software of INPE, v.5.2., in which the Digital Numbers (DN) were transformed into FRB (Apparent Bidirectional Reflectance Factors) using an algorithm written in LEGAL language. The atmospheric correction of the images was also performed. The NDVI was calculated from the corrected surface FRBs for the Landsat satellites, and from the original NDs for other satellites, in order to compare the temporal length of the cycle. Rice plots corresponding to Landsat series 224/80 orbit were identified by supervised classification assisted by additional confirmatory information . Descriptive statistics of the vectorized plot areas in the NDVI images were obtained for each plot and the mean values were plotted as a function of time. Alternative methods for generation of NDVI profiles as a function of time for the selected stands were also used: Multispec 3.1 and the dtwSat package of R statistical software. The multitemporal profiles were grouped into classes: early planting and cropping (early cycle); early planting and late cropping (medium or long cycle); late cropping and late cycle. The differences observed among the multitemporal profiles suggest that the introduction of new rice cultivars inserted from the mid 1980s to 2016 caused changes in the proportion of each type of rice present in the region. The varieties identified as precocious are present in greater proportion in the 2006/07 crop compared to the others. In the 2016/17 season, there is a significant increase in medium-cycle cultivars. Information from the regional rice industry corroborates and justifies this trend.

KEYWORDS: Remote Sensing, Radiometric Transformation, Vegetation Index.

INTRODUÇÃO

Segundo Bazgeer *et al*. (2006), as imagens de satélite têm surgido como uma importante e promissora fonte de dados para a estimativa da área cultivada e rendimento de grãos, em função de gerarem informações objetivas a respeito das culturas agrícolas, com adequada caracterização espacial e temporal.

O NDVI (*Normalized Difference Vegetation Index*), proposto por Rouse *et al*. (1973), é uma operação aritmética simples e muito utilizada atualmente para estudos do comportamento espectral, caracterização e monitoramento da vegetação bem como detecção e avaliação de desmatamentos. O NDVI é obtido através da Equação1:

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}} \quad (1)$$

ρ_{nir} = reflectância da banda no infravermelho próximo;

ρ_{red} = reflectância da banda no vermelho.

O NDVI está correlacionado com o índice de área foliar (IAF) nas lavouras de arroz, pode apresentar algumas limitações quanto a saturação em alguns dosséis e sensibilidade as condições atmosféricas e do solo do fundo (XIAO *et al*., 2003). Diversos algoritmos tem sido desenvolvidos para identificar lavouras de arroz usando perfis multitemporais obtidos de LSWI, NDVI, e EVI (XIAO *et al*., 2002, 2005; CHEN *et al*., 2011). O NDVI pode ser usado em modelos de rendimento de acordo com as estimativas realizadas por Nuarsa *et al.* (2011). Novos índices de vegetação tem sido propostos na literatura científica, um exemplo é o trabalho de Wang *et al.* (2007).

EVI (*Enhanced Vegetation Indice*) e NDVI oriundas do sensor MODIS foram usadas para a identificação das áreas orizícolas no RS no trabalho de D'Arco *et al*. (2005) e Klering *et al.* (2009). Imagens NDVI do SPOT foram usadas em mapeamentos de lavouras de arroz no trabalho de Nguyen *et al*. (2012). Rizzi & Rudorff (2005) usaram o NDVI para estimar a área da cultura de soja no RS.

Os estádios fenológicos do arroz são descritos pela SOSBAI (2016), o arroz após a emergência e formação das primeiras folhas inicia o perfilhamento, quando a panícula começa a se formar ocorre um rápido alongamento dos entre-nós até a florescência da planta, o arroz se autofecunda, se inicia a formação dos grãos até atingir a maturação fisiológica, após isso, através da perda de umidade ocorre a maturação da colheita.

Após os estágios iniciais de desenvolvimento quando a cultura cobre totalmente a lâmina da água, a resposta espectral está associada diretamente as variações de biomassa e no vigor do dossel (XIAO *et al*., 2006).

O objetivo deste trabalho foi realizar uma análise multitemporal das lavouras de arroz na área de estudo, usando o NDVI determinado a partir de FRB de superfície para gerar os perfis multitemporais, buscando uma caracterização temporal dos diferentes estádios de desenvolvimento do ciclo fenológico desta cultura. A comparação das safras 1984/85 e 2006/07 permite verificar si a introdução de novas cultivares de arroz inseridas nesse período acarretou em alguma mudança relevante no comportamento espectral/temporal das lavouras de arroz na região de estudo.

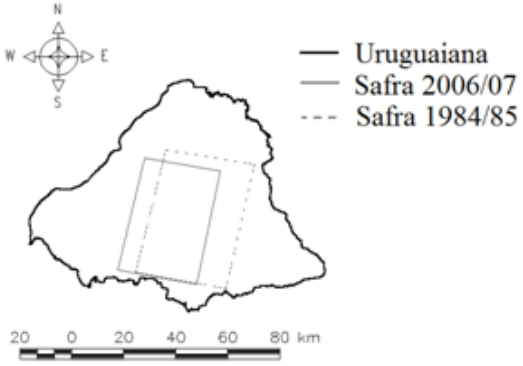
MATERIAL E MÉTODOS

A área de estudo é uma zona tradicionalmente arrozeira, incluída nas cenas 224/80, 225/80, 224/81 e 225/81 dos satélites Landsat 5 e 8, incluindo o município de Itaqui e parte dos

municípios limítrofes, localizada na Fronteira Oeste do Rio Grande do Sul, a região de maior produção de arroz no Estado e em todo o Brasil.

Para este estudo foram utilizadas imagens do sensor TM/L5 e adquiridas no final de 1983 e início de 1984 para analisar a safra 1984/85, e imagens de período análogo para 2006 e 2007 para a safra 2006/07, todas obtidas na galeria de imagens do INPE (Instituto Nacional de Pesquisas Espaciais). Para a safra 2016/17 foram selecionadas cenas do sensor OLI/L8. Dentro da área estudada, existem regiões que aparecem em cenas vizinhas, cuja sobreposição permite uma maior resolução temporal, favorecendo a análise multitemporal de arroz. Estas áreas foram usadas com vantagem em procedimentos de classificação e verificação de conclusões sobre os ciclos multitemporais. Na região encontra-se o município de Uruguaiana, o maior produtor de arroz do Brasil (IBGE). Para a análise das imagens foram utilizados os programas Spring versão 5.2 (CÂMARA et al., 1996), desenvolvido e distribuído pelo INPE, o Multispec (BIEHL and LANDGREBE, 2002), e o pacote dtwSat do R (Maus *et al*, 2017). Inicialmente, as imagens foram registradas a partir dos mosaicos de imagens Geocover, disponível em http://zulu.ssc.nasa.gov/mrsid/, as imagens registradas foram importadas no Spring 5.2 usando o sistema UTM/WGS84. A Figura 1 delimita a área de estudo em ambas as safras.

Figura 1 - Limite do município de Uruguaiana com a área de estudo das safras estudadas.



As imagens foram transformadas de ND a FRB aparente dentro do Spring 5.2 usando um algoritmo escrito em linguagem Legal. Neste algoritmo, são considerados o ângulo zenital e os índices da calibração (Lmin e Lmax) para cada banda espectral, e os valores de irradiância solar (Esun) estimados no topo da atmosfera.

A série temporal de imagens convertidas em FRB aparente foi submetida ao processamento do modelo 6S (*Second Simulation of the Satellite Signal in the Solar Spectrum*), o processamento inclui a caracterização da atmosfera da qual se pretende minimizar seu efeito sobre a reflectância dos alvos, o objetivo é minimizar o efeito

atmosférico sobre os alvos. Como resultado do processamento foi obtido um conjunto de imagens FRB de superfície, são as imagens com correção atmosférica.

Após a transformação dos FRB aparente em FRB de superfície, as imagens foram recortadas usando um arquivo vetorial correspondente a área de estudo, foi selecionada uma área de estudo para cada safra analisada. As imagens foram usadas para gerar o NDVI de cada data analisada, o qual foi gerado a partir de FRB de superfície. As imagens foram exportadas em Geotif e importadas no Multispec 3.1, onde ocorreu a seleção de amostras para gerar os perfis multitemporais apresentados neste trabalho. No processo de geração dos mapas de cobertura vegetal, o arroz foi classificado de acordo com a análise dos perfis realizada, para esta finalidade foi usado o classificador supervisionado de máxima verossimilhança implementado no Spring 5.2.

RESULTADOS E DISCUSSÃO

Foram analisadas aproximadamente 30 lavouras de arroz para cada safra, a seleção das amostras foi realizada buscando homogeneidade espectral/temporal dos pixels internos de cada amostra coletada e diversidade entre as diferentes amostras selecionadas, buscando assim obter perfis multitemporais que representem as diferentes datas de semeadura, duração de ciclo da lavoura e colheitas, realizadas em distintos períodos. Os perfis multitemporais obtidos foram analisados e classificados em: i) Semeadura e colheita precoce; ii) Semeadura precoce e colheita tardia; iii) Semeadura tardia.

As Figuras de 2 a 3 apresentam uma seleção dos perfis obtidos, a partir dos quais é realizada a discussão.

De acordo com a Figura 2a, os talhões apresentam um NDVI próximo de 0,3 na imagem de 6 de outubro de 1984, a análise da Figura 3a mostra um NDVI em 11 de novembro de 2006 próximo de 0,43. Portanto percebe-se que outubro é o período mais representativo de semeadura do arroz para as lavouras em que ela é realizada cedo, pois foi o que apresentou valores de NDVI inferiores, uma semeadura nos meses de outubro e novembro é coerente com os perfis obtidos por Klering *et al.*(2009) no RS.

Os valores máximos de NDVI no ano de 1985 estão entre 0,75 e 0,9, obtidos a partir da imagem de 1 de janeiro de 1985, os valores máximos de NDVI para 2007 estão entre 0,87 e 0,95 na imagem de 14 de janeiro de 2007, valores coerentes com o máximo obtido por Xiao *et al.* (2005). Deste modo, nos perfis analisados (Figuras 2a e 3a) o máximo da safra 1984/85 ocorreu no início de janeiro e para a safra 2006/07 no meio de janeiro.

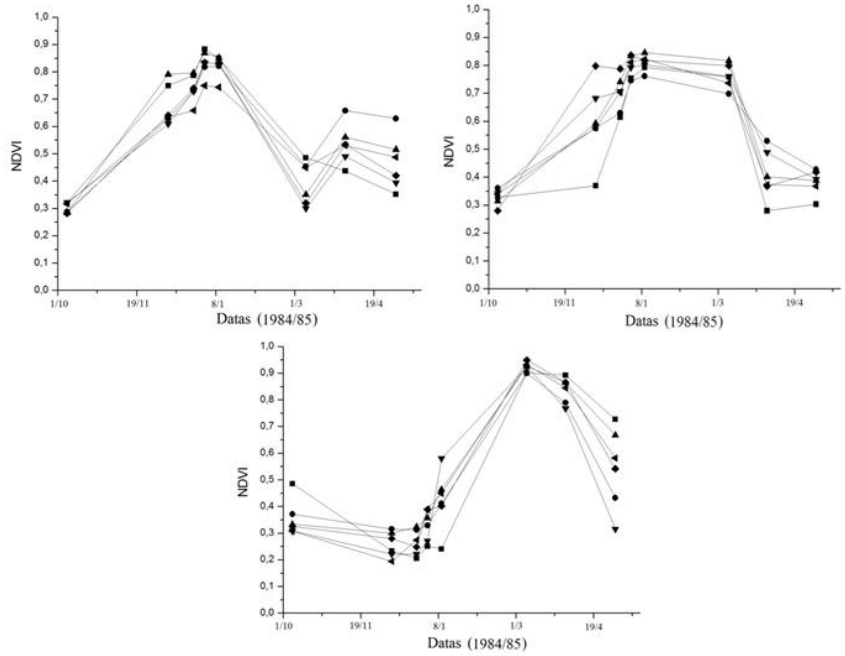


Figura 2 - Perfis multitemporais obtidos na safra 1984/85: a) semeadura e colheita precoce, b) semeadura precoce e colheita tardia, c) semeadura tardia.

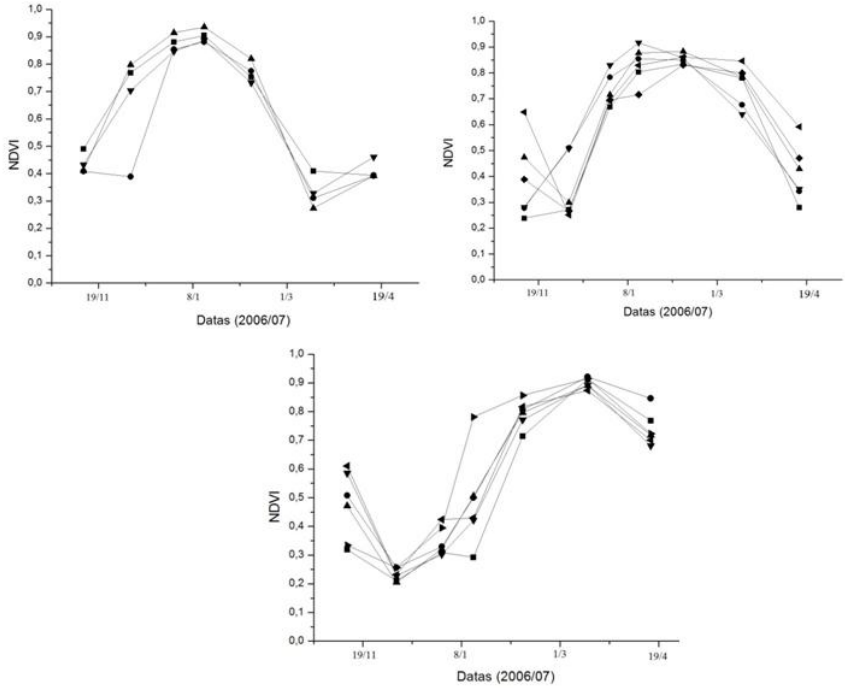


Figura 3 - Perfis multitemporais obtidos na safra 2006/07: a) semeadura e colheita precoce, b) semeadura precoce e colheita tardia, c) semeadura tardia.

A forte queda de NDVI registrada em 6 de março de 1985 indica que a colheita ocorreu em alguma data entre 10 de janeiro e 6 de março de 1985, comportamento parecido se visualiza na safra 2006/2007, o baixo valor de NDVI registrado em 12 de março de 2012,

indica que a colheita aconteceu entre 8 de fevereiro e 12 de março de 2007, valores coerentes com o mínimo obtido por Xiao *et al.* (2005). Estes resultados permitem estimar que a colheita das lavouras classificada como semeadura e colheita precoce ocorre na segunda quinzena de fevereiro e inicio de março.

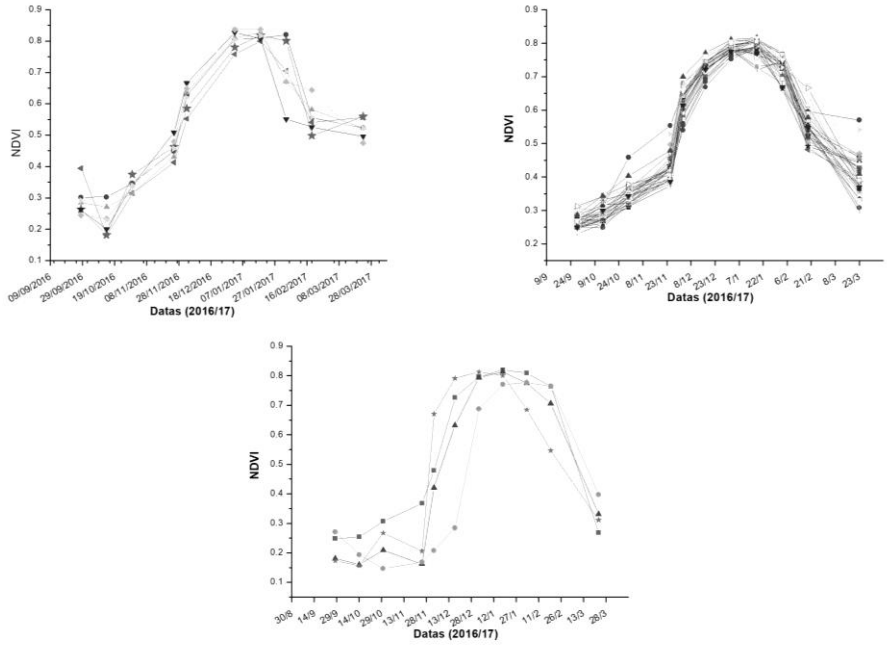


Figura 4 - Perfis multitemporais obtidos na safra 2016/17 com imagens Landsat 8: a) semeadura e colheita precoce, b) semeadura precoce e colheita tardia, c) semeadura tardia.

Prosseguindo com a análise das Figuras 2a e 3a, após a colheita, uma pequena elevação de NDVI é verificada em 31 de março e uma queda ocorre novamente em 2 de maio no ano de 1985, uma elevação semelhante é observada para 13 de abril de 2007. A pequena elevação de NDVI mensurada em ambas as safras após a colheita pode ser conseqüência do rebrote do arroz que em alguns casos é aproveitada como pastagem para o gado, introduzindo o sistema integração lavoura-pecuária. Este pequeno pico visualizado após a colheita não é verificado no perfil obtido no trabalho de Xiao *et al*. (2005), tampouco no trabalho de Klering *et al.* (2009), uma possível explicação para isso pode ser os costumes agrícolas/pecuários diferenciados nas lavouras cultivadas na China, analisadas no artigo de Xiao *et al*. (2005) e o uso de uma "mascara geral" no caso do trabalho de Klering *et al.* (2009), a qual não permite identificar peculiaridades da safra plantada precocemente.

Os perfis da Figura 2b e 3b consistem em lavouras de longo ciclo, existentes em praticamente todo o período de verão do RS. No início de março, ambas as safras apresentaram valores de NDVI um pouco inferiores ao seu auge (que ocorreu no mês janeiro), provavelmente esta queda se deve ao período de maturação do arroz, no qual as folhas

adquirem coloração amarelada e diminuem a atividade fotossintética, o que afeta o contraste das bandas do vermelho e infravermelho, diminuindo os valores de NDVI, a queda de NDVI no mês de março é coerente com os resultados de Klering *et al.* (2009).

De acordo com a Figura 2c, a maioria dos talhões apresentou perfis com valores baixos de NDVI em 10 de janeiro de 1985, o mesmo ocorre para 14 de janeiro de 2007, o que caracteriza uma semeadura tardia comparado com os perfis discutidos anteriormente, o que discrepa dos valores obtidos por Klering *et al.* (2009), esta divergência provavelmente se deve ao fato de que no presente trabalho as lavouras foram classificada em três categorias, seria o equivalente a possuir três mascaras de mapeamento, e não apenas 1.

Prosseguindo com a análise das Figuras 2c e 3c, o NDVI máximo ocorre em 6 de março de 1985, comportamento parecido ocorre na safra 2006/07, o NDVI máximo ocorre em 12 de março de 2007. Portanto em ambas as safras o NDVI máximo ocorreu em março, em datas que podem ser consideradas equivalentes, pois a pequena diferença temporal (poucos dias) ocorre em função da aquisição em diferentes datas das imagens analisadas de uma safra para a outra, os valores da safra 1984/85 são um pouco superiores.

Quedas de NDVI são observadas em 2 de maio de 1985, algumas acentuadas e outras mais leves, os valores baixos de NDVI próximos de 0,4 consistem em talhões que anteriormente passaram pelo processo de colheita, já as quedas mais leves, provavelmente se devem ao período de maturação do arroz. A comparação das safras leva a crer que a colheita das lavouras de arroz classificadas como semeadura tardia ocorre majoritariamente na segunda quinzena de abril.

Após a análise multitemporal das curvas de NDVI as lavouras de arroz foram classificadas de acordo com os perfis analisados, foi usado o classificador supervisionado de máxima verossimilhança. A Tabela 1 apresenta os resultados dos tipos de arroz classificados para as safras analisadas.

De acordo com a Tabela 1, a proporção de arroz classificada como semeadura e colheita precoce é maior na safra 2006/07 que na safra 1984/85, a proporção do arroz de semeadura tardia é maior na safra 1984/85 comparado com a safra 2006/07. A proporção de arroz classificado como semeadura precoce e colheita tardia resultou bem próxima em ambas as safras. Estimativas semelhantes não foram encontradas na literatura cientifica, pois os mapeamentos realizados em outros trabalhos acadêmicos não costumam descriminar estes aspectos temporais da cadeia produtiva do arroz. Para a safra 2016/17 foi observado um aumento na classe de semeadura precoce e colheita tardia, que corresponde a cultivares de ciclo médio, mas que não conseguiu ser analisado mais detalhadamente neste trabalho.

Tabela 1- Resultados da classificação do arroz.

Classe	1984/85		2006/07	
	ha	%	ha	%
Semeadura e colheita precoce	943	10,1	3494	35,3
Semeadura tardia	3900	41,7	1565	15,8
Semeadura precoce e colheita tardia	4503	48,2	4846	48,9
Total	9346		9905	

CONCLUSÕES

De uma forma geral, a comparação dos perfis multitemporais obtidos das lavouras das safras 1984/85, 2006/07 e 2016/17 demonstram uma mudança nas cultivares de arroz predominantes na matriz produtiva da região desde meados da década de 80 até 2017 com um aumento das cultivares de ciclo precoce evidenciada em 2007 e um predomínio das de ciclo médio em 2017.

Entre os métodos utilizados, a análise temporal produzida pelo processamento individual dos NDVI dos talhões mediante obtenção da estatística descritiva e gráfico ao longo do ciclo resultou mais confiável, porém mais demorado que os outros métodos.

ACKNOWLEDGEMENTS

The authors acknowledge UNIPAMPA, UFSM and UERGS for institutional support to this work.

REFERÊNCIAS

BAZGEER, S.; MAHEY, R.K.; SHARMA, P.K.; SOOD, A.; SIDHU, S.S. Pre-Harvest wheat yield prediction using agromet-spectral-trend-yield models for Hoshiarpur and Rupnagar district of Punjab. Journal of the Indian Society of Remote Sensing, v.34(3), p.269-277, 2006.

BIEHL, L., LANDGREBE, D. MultiSpec—a tool for multispectral–hyperspectral image data analysis. Computers & Geosciences, 28(10), pp.1153-1159, 2002.

CÂMARA, G.; SOUZA, R.C.M.; FREITAS, U.M.; GARRIDO, J.; MITSUO, F.I. Integrating remote sensing and GIS by object-oriented data modeling. Computers & Graphics, v.20(3), p.395-403, 1996.

CHEN, J.; HUANG, J.; HU, J. Mapping rice planting areas in southern China using the China Environment Satellite data. Mathematical and Computer Modelling, v.54, p.1037-1043, 2011.

D'ARCO, E.; RIZZI, R.; RUDORFF, B.F.T. Análise de combinações dos índices de vegetação EVI e NDVI derivadas do sensor MODIS para identificação de áreas de arroz irrigado. Anais: XII Simpósio Brasileiro de Sensoriamento Remoto, Goiânia, Brasil, 16-21 abril 2005, INPE, p. 91-93.

KLERING, E.L.; GUTERRES, D.; FONTANA, D.C.; BERLATO, M.A.. Comparação entre os perfis temporais de arroz irrigado das coleções MODIS V004 e V005 para a região orizícola do Rio Grande do Sul. Anais: XIV Simpósio Brasileiro de Sensoriamento Remoto, Natal, Brasil, 25-30 abril 2009, INPE, p. 239-245.

MAUS, V.; CÂMARA, G.; APPEL, M.; PEBESMA, E. dtwSat: Time-Weighted Dynamic Time Warping for Satellite Image Time Series Analysis in R. The software is available from the IIASA repository PURE pure.iiasa.ac.at/14514/. Precompiled binary available from CRAN at cran.r-project.org/web/packages/dtwSat/index.html

NGUYEN, T.T.H.; BIE DE, C.A.J.M.; ALI, A.; SMALING, E.M.A.; CHU, T.H. Mapping the irrigated rice cropping patterns of the Mekong delta, Vietnam, through hyper-temporal SPOT NDVI image analysis. International Journal of Remote Sensing, v.33(2), p.415-434, 2012.

NUARSA, I.W.; NISHIO, F.; HONGO, C. Relationship between Rice Spectral and Rice Yield Using Modis Data. Journal of Agricultural Science, v.3(2), 2011.

RIZZI, R.; RUDORFF, B.F. Estimativa da área de soja no Rio Grande do Sul por meio de imagens Landsat. Revista Brasileira de Cartografia, v.57(3), p.226-234, 2005.

ROUSE, J.W.; HAAS, R.H.; SCHELL, J.A.; DEERING, D.W. Monitoring vegetation systems in the great plains with ERTS. Earth Resources Technology Satellite-1 Symposium, Washington: Procedings, v.1, sec. A, p.309- 317, 1973.

SOSBAI, 2016. Sociedade Sul-Brasileira de Arroz Irrigado. Recomendações técnicas da pesquisa para o Sul do Brasil. Reunião Técnica da Cultura do Arroz Irrigado. Agosto de 2016.

WANG, F.; HUANG, J.; TANG, Y.; WANG, X. New Vegetation Index and Its Application in Estimating Leaf Area Index of Rice. Rice Science, v.14(3), p.195-203, 2007.

XIAO, X.; BOLES, S.; FROLKING, S.; SALAS, W.; MOORE, B.; LI, C. Observation of flooding and rice transplanting of paddy rice fields at the site to landscape scales in China using vegetation sensor data. International Journal of Remote Sensing, v.23, p.3009– 3022, 2002.

XIAO, X.; BRASWELL, B.; ZHANG, Q.; BOLES, S.; FROLKING, S.; MOORE, B. Sensitivity of vegetation indices to atmospheric aerosols: Continental-scale observations in Northern Asia. Remote Sensing of Environment, v.84, p.385– 392, 2003.

XIAO, X.M.; BOLES, S.; LIU, J.Y.; ZHUANG, D.F.; FROLKING, S.; LI, C.S. Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. Remote Sensing of Environment, v.95(4), p.480– 492, 2005.

XIAO, X.M.; BOLES, S.; FROLKINGA, S.; LI, C.; BABUA, J.; SALAS, W.; MOORE, B. Mapping paddy rice agriculture in South and Southeast Asia using multitemporal MODIS images. Remote Sensing of Environment, v.100, p.95-113, 2006.

Análise comparativa entre técnicas de fusão de imagens Landsat-8 para delimitação de áreas de cana-de-açúcar

Gisela Maiara Silva[1,2], Ana Cláudia dos Santos Luciano[2], Karina Maria Berbert Bruno[2], Daniel G. Duft[2], Matheus Phelipe Miller Dias da Silva[2], Guilherme Martineli Sanches[2]

[1]Graduanda, Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas, Campinas, São Paulo, Brasil, gisela.silva@feagri.unicamp.br

[2]Laboratório Nacional de Ciência e Tecnologia do Bioetanol (CTBE/CNPEM), Campinas, São Paulo, Brasil, gisela.silva@bioetanol.org.br, ana.luciano@bioetanol.org.br, karina.bruno@bioetanol.org.br, daniel.duft@bioetanol.org.br, matheus.dias@bioetanol.org.br, guilherme.sanches@bioetanol.org.br

RESUMO

Este trabalho tem por objetivo analisar o efeito da utilização de técnicas de fusão de imagens para delimitação de áreas de cana-de-açúcar a partir do processo de segmentação. A área de estudo está localizada na região oeste do Estado de São Paulo, uma das principais regiões produtoras de cana-de-açúcar. Foram utilizadas imagens do sensor OLI do satélite LANDSAT 8, referentes as bandas 6, 5 e 4 (multiespectral) e a banda 8 (pancromática). As técnicas de fusão *Gram-Schmidt* (GS) e HSV foram aplicadas às imagens multiespectral em conjunto com a imagem pancromática. A avaliação dos resultados foi feita com base na análise visual e estatística. Na análise estatística foi avaliada a fidelidade espectral das imagens fusionadas em relação à imagem multiespectral original. Para isso, foi utilizado a média, o desvio padrão e o coeficiente de correlação. Por fim, para delimitação das áreas de cana-de-açúcar e outros usos foi feita a segmentação da imagem multiespectral e das imagens fusionadas. Os resultados indicaram que a técnica GS apresentou melhor preservação do contraste e brilho da imagem em comparação a técnica HSV. Além disso, as áreas de cana-de-açúcar foram melhor delimitadas pela segmentação quando utilizadas as imagens com fusão GS. A técnica de fusão, aplicada em imagens do sensor OLI, auxiliou na melhor delimitação das áreas de cana-de-açúcar. No entanto, análises adicionais são necessárias a fim de melhor avaliar os métodos de fusão e suas implicações na delimitação e classificação da cana-de-açúcar e outros usos e cobertura da terra.

PALAVRAS-CHAVE: Processamento de imagens, Segmentação de imagens, OBIA.

ABSTRACT

This work aims to analyze the effect of image fusion techniques to identify sugarcane areas from the segmentation process. The study area is located in the western region of Sao Paulo State, one of the main sugarcane producer region. Images from LANDSAT 8 OLI sensor were used for bands 6, 5 and 4 (multispectral) and band 8 (panchromatic). *Gram-Schmidt* (GS) and HSV fusion techniques were applied to the multispectral images together with the panchromatic image. The results were then evaluated based on visual and statistical analysis. In the statistical analysis, the spectral fidelity of the fused images was evaluated in relation to the original multispectral image. For this, we considered the mean, standard deviation and the correlation coefficient of the images. Finally, to delimit sugarcane areas and other uses, the segmentation of the multispectral image and the fused images was done. The results indicated that the GS technique presented better contrast preservation and image brightness compared to the HSV technique. In addition, sugarcane areas were better delimited by segmentation when using GS fused images. The fusion technique, applied in images of the OLI sensor, helped in the better delimitation of sugarcane areas. However, further analysis is needed in order to better evaluate the methods of fusion and their implications in the delimitation and classification of sugarcane and other uses and land cover.

KEY WORDS: Image processing, Image segmentation, OBIA.

INTRODUÇÃO

O Brasil é o maior produtor mundial de cana-de-açúcar e também um dos maiores produtores de etanol do mundo (OECD e FAO, 2015). O país produziu na última safra 657,18 milhões de toneladas de cana-de-açúcar em uma área correspondente de 8,65 milhões de hectares (CONAB, 2017). Atualmente, devido à importância da cana-de-açúcar no setor sucroenergético brasileiro, a previsão de safra e, portanto, a estimativa de área e produção é essencial para o planejamento do setor, uma vez que tal informação afeta uma série de atividades das operações nas usinas (PICOLI et al., 2009).

A previsão de safras tem sido feita ao longo dos anos por órgãos oficiais como IBGE, CONAB e MAPA através de levantamentos em campo. Este processo demanda tempo e recursos, além de ser muitas vezes subjetivo. Com o intuito de obter informações de forma rápida e precisa, as imagens de satélite tem favorecido a previsão de safras, especificamente no que diz respeito à quantificação e localização de área da cultura (RIZZI; RUDORFF, 2005;

SCHULTZ et al., 2016). Para isso, diversas técnicas de classificação de imagens têm sido utilizadas na identificação das áreas de cana-de-açúcar.

A classificação de imagens a partir do processo de Análise de Imagens Orientada a Objetos (OBIA), nas últimas décadas, tem se mostrado como um método eficiente para extração de informações como área e localização (BELGIU; DRĂGUT, 2014). O processo de OBIA aplicado à imagens de sensores ópticos, como os do satélite Landsat, apresenta melhor desempenho do que a classificação por pixel (FERNANDES et al., 2012; GAROGALO et al., 2015).

Um dos requisitos para aplicação da classificação baseada em objetos é a criação dos objetos representativos nas imagens a serem analisadas. Estes objetos, em geral, são criados a partir de algoritmos de segmentação (DRĂGUT et al., 2014). A técnica de segmentação de imagens é eficiente em delimitar diferentes alvos da cobertura da terra, possibilitando a quantificação de área plantada de culturas, como por exemplo (QUEVEDO et al., 2009).

O processo de segmentação e classificação de áreas de cana-de-açúcar tem sido melhorado com a utilização de fusão de imagens (JOHNSON et al., 2014; MULIANGA et al., 2016). Segundo Wald (1999), a fusão de dados consiste em uma estrutura formal, onde são expressos conceitos e ferramentas para junção de diferentes dados e obtenção de informações de melhor qualidade, ou seja, a fusão melhora a delimitação de objetos na imagem, a classificação e os processos matemáticos entre bandas.

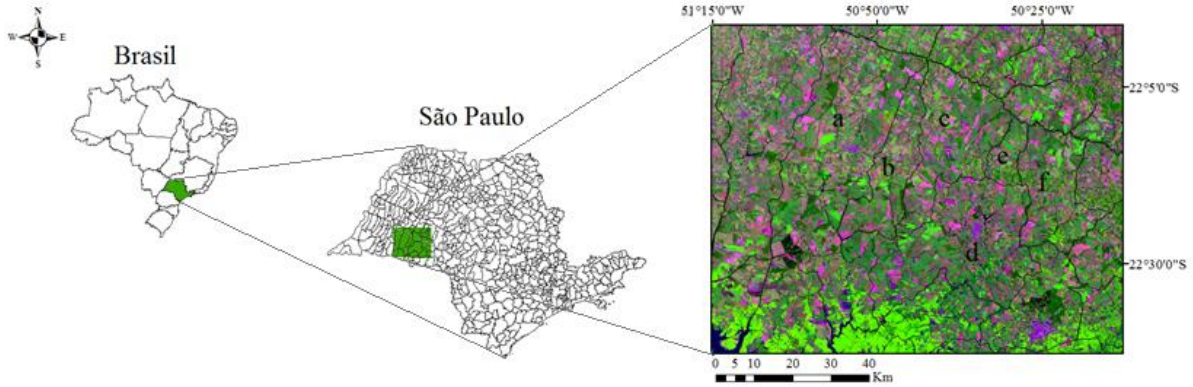
As técnicas de fusão são divididas em três grupos, as que utilizam um modelo de domínio espacial, as de domínio espectral e as que trabalham com operações algébricas (PINHO et al., 2005). Como exemplo das de domínio espacial tem-se a *Wavelets* e a técnica de *High-Pass Filter*. Já as de operação algébricas, a *Brovey* é um exemplo (LEONARDI et al., 2009*).* No grupo de modelos de domínio espectral, a fusão é feita através de processos que realizam uma transformação na imagem multiespectral para gerar um conjunto novo de bandas. Como exemplo de fusão dos modelos de domínio espectral encontram-se as de componentes principais, a HSV também denominada IHS (I-Intensidade, M-Matiz e S-Saturação) e a *Gram-Schmidt*. De modo geral, as técnicas de domínio espectral têm apresentado resultados satisfatórios quando aplicadas às imagens Landsat para delimitação de diferentes objetos (ADAMI et al., 2007; SCHUNDERS; BACKER, 2001; SOARES et al., 2015).

Neste contexto, este trabalho tem por objetivo analisar o efeito da utilização de técnicas de fusão de imagens para delimitação de áreas de cana-de-açúcar a partir do processo de segmentação.

MATERIAL E MÉTODOS

A área de estudo está localizada na região oeste do Estado de São Paulo, entre as latitudes 21º56'9,91''S e 22º42'51,01''S e longitudes 51º15'18,59''O e 50º12'36,37''O, e apresenta área total de 929.836,0 ha (Figura 1). Os principais municípios da região são: Rancharia, João Ramalho, Quatá, Paraguaçu Paulista, Borá e Lutécia. Nesta região, a área média de cana-de-açúcar na última safra foi de aproximadamente 181.487 ha (IBGE, 2015).

Figura 1 – Localização da área de estudo a) Rancharia, b) João Ramalho, c) Quatá, d) Paraguaçu Paulista, e) Borá e f) Lutécia.



Para delimitação das áreas de cana-açúcar foram selecionadas imagens do sensor OLI do satélite Landsat-8 referentes às bandas 4 (vermelho: 0,64 – 0,67µm), 5 (NIR: 0,85 – 0,88 µm) e 6 (SWIR 1: 1,57 – 1,65µm), com resolução espacial de 30 m, e à banda 8 (pancromática: 0,50 – 0,68 µm) com resolução espacial de 15 metros. As imagens foram obtidas do banco de dados da USGS (*United States Geological Survey*), disponível no site (https://earthexplorer.usgs.gov/), para as cenas de órbita/ponto 222/75 e 222/76, nas datas de 28/01/2014 e 15/01/2015.

O processo metodológico consistiu na fusão de imagens, segmentação e análise dos dados (Figura 2). Inicialmente, para cada uma das datas selecionadas, foi realizada a fusão das bandas multiespectrais 4, 5 e 6 com a banda 8. O processo de fusão foi feito no software ENVI 5.1, a partir das técnicas "*Gram-Schmidt Spectral Sharpening*" e HSV. A técnica "*Gram-Schmidt Spectral Sharpening*" simula uma banda pancromática com resolução semelhante a multiespectral e, em seguida aplica a transformação *Gram-Schmidt* na imagem multiespectral e pancromática (RSI, 2003). Com isso, a banda simulada é substituída pela imagem pancromática. A técnica "HSV" consiste na transformação da imagem de baixa resolução espacial do espaço RGB de cores para o espaço HSV (SCHNEIDER et al., 2003). Para isso, a componente V é substituída pela imagem pancromática, e em seguida é efetuada a operação inversa, retornando para o espaço RGB.

Para analisar o resultado da fusão foram feitas comparações visuais entre as imagens fusionadas e a imagem multiespectral original. Além disso, foram feitas avaliações da qualidade espectral das imagens fusionadas a partir de técnicas estatísticas referentes a média, desvio padrão (DP), diferença do desvio padrão (DDP) e coeficiente de correlação (WALD et al., 1997). A média e desvio padrão das imagens estão relacionadas com o brilho e contraste, respectivamente (PINHO et al., 2005). A avaliação baseada na diferença do desvio padrão (Equação 1) indica se houve perda ou ganho de informações na imagem durante o processo de fusão (WALD et al., 1997). No caso de valores positivos há perda de informação, e valores negativos ganho de informação.

$$DDP = \frac{\sigma_o - \sigma_f}{\sigma_o} \qquad \text{Eq. (1)}$$

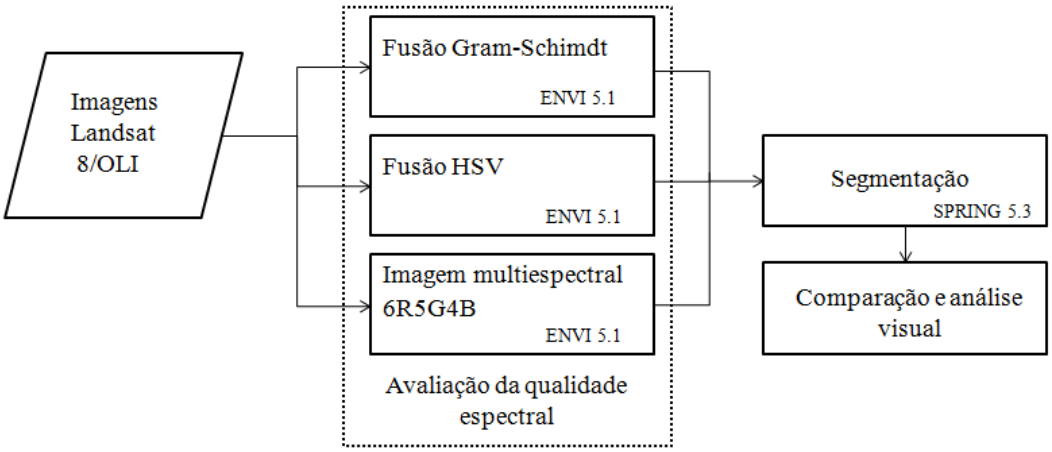
Onde σ_o é o desvio padrão da imagem original (multiespectral) e o σ_f é o desvio padrão da imagem fusionada. O coeficiente de correlação (Cor) (Equação 2) indica o grau de similaridade de pequenas estruturas na imagem e deverá ser em condições ideais igual a 1 (SCHEUNDERS & DE BACKER, 2001; WALD et al., 1997).

$$Cor\ (x,y) = \frac{[(x-[x])][(y-[y])]}{\sqrt{[(x-[x])^2][(y-[y])^2]}} \qquad \text{Eq. (2)}$$

Onde x/y é o valor do pixel na imagem original/fundida e [.] denota a média ao longo de todos os pixels.

Após o processo de fusão de imagens foi feita a segmentação para as imagens resultantes de cada fusão. Esta etapa foi realizada no software SPRING 5.3. As segmentações foram feitas com base nas imagens das bandas 4 e 5 do ano de 2014 e 2015, pois assim é possível delimitar as mudanças dos polígonos de cana-de-açúcar ao longo do período de crescimento da cultura, antes de se iniciar a colheita em março de 2015. Para o processo de segmentação foi utilizado o algoritmo de crescimento de regiões com parâmetros de área de 20 e similaridade de 200. A avaliação da segmentação para cada imagem fusionada foi feita a partir de análises visuais e quantitativas. Nestas análises foram avaliados o número e o tamanho dos polígonos gerados com base na identificação dos alvos presentes nas imagens de satélite. Além disso, foi utilizado os dados do CANASAT 2015, disponibilizados pela UNICA (União da Indústria de Cana-de-Açúcar) para análise da delimitação dos polígonos de cana-de-açúcar.

Figura 2 – Fluxograma metodológico



RESULTADOS E DISCUSSÃO

A comparação visual das imagens fusionadas com a imagem multiespectral original mostrou melhor preservação de cores quando foi aplicada a técnica de fusão *Gram-schimdt* (GS) em comparação a técnica HSV (Figura 3). No entanto, foi possível verificar que em áreas de cana-de-açúcar as linhas entre os talhões foram melhor delimitadas com a técnica HSV (Figura 3c).

Figura 3: Comparação entre as imagens a) multiespectral original, b) *Gram-Schmidt* e c) HSV para o ano de 2015. Composição de bandas 6R5G4B.



A variação dos valores de média e desvio padrão entre as imagens fusionadas e a imagem multiespectral foi menor para a técnica GS do que para a técnica HSV (Tabela 1). Isto demostra que a técnica HSV alterou o brilho da imagem enquanto que a técnica GS preservou o brilho. Quanto a preservação do contrate das imagens, foi verificado que a técnica GS também apresentou melhores resultados do que a técnica HSV. Isto porque, há menores variações de desvio padrão para os resultados da técnica GS entre as imagens fusionadas e a multiespectral original (Tabela 1).

A imagem GS apresentou os maiores valores do coeficiente de correlação para a maioria das bandas (Tabela 1). No ano de 2015, foi possível verificar que o coeficiente de correlação para a banda 4 foi maior para a técnica HSV. Além disso, a técnica GS apresentou valores de

correlação mais próximos de 1 (Tabela 1). Estes valores indicaram que há um maior grau de similaridade dos pixels da imagem multiespectral original (MO) com a imagem GS (WALD et al., 1997). Isto indica a preservação da informação espectral da imagem com a utilização da técnica de fusão GS.

Tabela 1 – Valores de média, desvio padrão (DP) e coeficiente de correlação (Cor) das imagens fusionadas pelas técnicas GS e HSV e original (MO).

Método	Banda 6			Banda 5			Banda 4		
	Média	DP	Cor	Média	DP	Cor	Média	DP	Cor
				2014					
MO	13780,0	2795,6	1	19805,5	3439,0	1	8487,1	2092,2	1
GS	13862,6	2502,9	0,95	19548,9	2394,4	0,90	8422,9	1331,9	0,83
HSV	104,6	59,9	0,86	110,7	37,3	0,43	82,2	61,8	0,69
				2015					
MO	14447,6	2538,5	1	20335,9	2700,2	1	8395,3	1216,9	1
GS	14509,2	2582,0	0,88	20104,9	2394,9	0,98	8382,6	1198,0	0,85
HSV	95,9	62,5	0,83	93,8	34,3	0,40	73,1	56,7	0,87

A técnica de fusão GS aplicada nas imagens do ano de 2015 apresentou valores da DDP mais próximos de zero do que a HSV (Tabela 2). Por isso, a análise mostrou que as perdas de informação espectral foram menores com a técnica GS. Para o ano de 2014 nenhum ganho de informação espectral foi identificado, pois os valores de DDP foram positivos. Já para o ano de 2015, foi verificado ganhos de informações apenas na banda 6 para a imagem GS.

Tabela 2 – Valores de diferença de desvio padrão (DDP) para as bandas 6, 5 e 4.

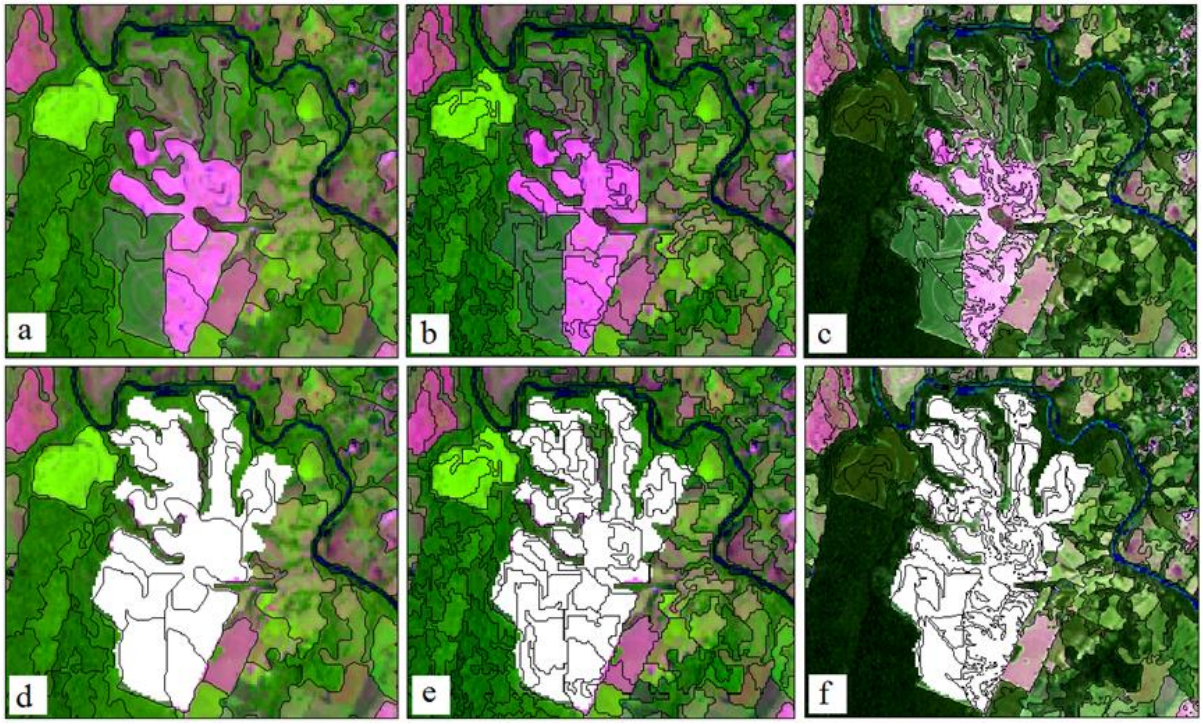
	2014		
	Banda 6	Banda 5	Banda 4
GS	0,1047	0,3037	0,3633
HSV	0,9786	0,9891	0,9704
	2015		
	Banda 6	Banda 5	Banda 4
GS	-0,0171	0,1131	0,0155
HSV	0,9785	0,9857	0,9527

A segmentação das imagens multiespectral original, GS e HSV delimitou 21.486, 85.229 e 99.903 polígonos, respectivamente. A maior quantidade de polígonos para as imagens fusionadas ocorreu devido ao aumento de resolução espacial com o processo de fusão. Quanto a delimitação das áreas de cana-de-açúcar, a segmentação aplicada à imagem multiespectral original apresentou polígonos com limites similares ao do CANASAT, porém foi verificado

que algumas áreas de cana-de-açúcar não foram bem delimitadas (Figura 4a e 4d). Já a segmentação da imagem GS apresentou melhores delimitações dos polígonos do que a segmentação aplicada a imagem multiespectral original e HSV (Figura 4).

Neste sentido, as segmentações com base nas imagens fusionadas mostraram melhor delimitação das áreas de cana-de-açúcar. No entanto, a segmentação da imagem HSV mostrou uma maior quantidade de polígonos, o que pode limitar o tempo de processamento da classificação de imagem orientada a objetos. Além disso, a análise dos resultados estatísticos de fidelidade espectral mostrou que a técnica HSV apresentou menor preservação da informação espectral das imagens em comparação a técnica GS. Tal fato pode ter influenciado a criação de polígonos desnecessários na delimitação de áreas de cana-de-açúcar e, portanto, na identificação dos objetos da imagem.

Figura 4 – Análise das imagens do ano de 2015 a) multiespectral original com sobreposição da segmentação, b) *Gram-Schmidt* com sobreposição da segmentação, c) HSV com sobreposição da segmentação, d) multiespectral original com sobreposição da segmentação e dados do CANASAT em branco, e) *Gram-Schmidt* com sobreposição da segmentação e dados do CANSAT em branco, f) HSV com sobreposição da segmentação e dados do CANASAT em branco



CONCLUSÕES

A aplicação da técnica de fusão *Gram-Schimdt* em imagens do sensor OLI do satélite Landsat-8 apresentou melhores resultados de preservação de contraste e brilho em comparação a aplicação da técnica HSV. A fusão de imagens influenciou na formação dos polígonos

formados a partir da segmentação e, portanto, na delimitação das áreas de cana-de-açúcar. A técnica de fusão *Gram- Schmidt* apresentou melhores delimitações de polígonos de cana-de-açúcar em comparação a utilização de imagens sem fusão e fusionadas pela técnica HSV. As técnicas de fusão de imagens auxiliam na criação de mapas indicativos das áreas de cana-de-açúcar e, portanto, impactam no planejamento e gestão do setor sucroenergético.

AGRADECIMENTOS

Ao Laboratório Nacional de Ciência e Tecnologia do Bioetanol (CTBE) e ao projeto SUCRE (CTBE/PNUD Project BRA/10/G31 - Sugarcane Renewable Electricity).

REFERÊNCIAS

ADAMI, M. et al. Fusão de imagens por IHS para melhorar a identificação de uso do solo em elementos amostrais. Eng. Agríc., Jaboticabal , v. 27, n. 2, p. 529-536, 2007.

BELGIU, M.; DRĂGUŢ, L. Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery. ISPRS Journal of Photogrammetry and Remote Sensing, v. 96, p. 67-75, 2014.

CONAB – Companhia Nacional de Abastecimento. Cana-de-açúcar, Safra 2016/2016. Segundo levantamento, Agosto 2016. Disponível em: < http://www.conab.gov.br/OlalaCMS/uploads/arquivos/16_08_17_10_07_35_boletim_cana_portugues_-_2o_lev_-_16-17.pdf > Acesso em: 17 de abril de 2017.

DRĂGUŢ, L. et al. Automated parameterisation for multi-scale image segmentation on multiple layers. ISPRS Journal of Photogrammetry and Remote Sensing, v. 88, p. 119-127, 2014.

FERNANDES, R. R.; NUNES, G. N.; SILVA,T. S. F. Classificação orientada a objetos na caracterização da cobertura da terra no Araguaia. Pesquisa Agropecuária Brasileira, v.47, n. 9, p. 1251-1260, 2012.

GAROFALO, D. F. T.; MESSIAS, C. G.; LIESENBERG, V.; BOLFE, E. L.; FERREIRA, M. C. Análise comparativa de classificadores digitais em imagens do Landsat-8 aplicados ao mapeamento temático. Pesquisa Agropecuária Brasileira, v. 50, n. 7, p. 593-604, 2015.

IBGE, Instituto Brasileiro de Geografia e Estatística. Disponível em < https://sidra.ibge.gov.br/tabela/5457#resultado > Acesso em 09/05/2017.

JOHNSON, B. A.; SCHEYVENS, H.; SHIVAKOTI, B.R. An ensemble pansharpening approach for finer-scale mapping of sugarcane with Landsat 8 imagery. International Journal of Applied Earth Observation and Geoinformation, v.33, p.218-255, 2014.

LEONARDI, F. et al. Fusão de Imagens CBERS 2B: CCD-HRC. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 14., 2009, Natal. Anais…Natal: INPE, 2009. p. 6951-6958.

MULIANGA, B.; BÉGUÉ, A.; CLOUVEL, P.;TODOROFF, Mapping Cropping Practices of a Sugarcane-Based Cropping System in Kenya Using Remote Sensing, Remote Sensing, v.7, n.11, p.14428-14444, 2016.

OECD e FAO. Perspectivas Agrícolas no Brasil: desafios da agricultura brasileira 2015-2024. FAO, Roma, 2015.

PICOLI, M. C. A.; RUDORFF, B. F. T.; RIZZI, R.; GIAROLLA, A. Índice de vegetação do sensor MODIS na estimativa da produtividade agrícola da cana-de-açúcar. Revista Bragantia, Campinas, v.68, n.3, p.789-795, 2009.

PINHO, C. M. D.; RENNÓ, C. D.; KUX, H. J. H. Avaliação de técnicas de fusão aplicadas à imagem Quickbird. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 12., 2005, Goiânia. Anais...Goiânia: INPE, 2005. p. 4225- 4232.

Plano Agrícola Pecuário 2006/07. Disponível em: http://www.agricultura.gov.br/. Acesso em: 13 jan. 2007.

QUEVEDO, E. R. et al. Aplicação da segmentação (SPRING) de imagens para avaliação ambiental multitemporal da bacia do Córrego Ceroula–MS, no período 1985-2007. In: SIMPÓSIO DE SENSORIAMENTO REMOTO, 14.,2009, Natal. Anais...Natal: INPE, 2009. p. 1481-1488.

RIZZI, R.; RUDORFF, B. F. T. Estimativa da área de soja no Rio Grande do Sul por meio de imagens Landsat. Revista Brasileira de Cartografia, v.3; n.57, p.226-234, 2005.

RSI (Research Systems, Inc). ENVI User'sGuide. ENVI Version 4.0 September, 2003.

SCHEUNDERS, P.; DE BACKER, S. Multispectral image fusion and merging using multiscale fundamental forms. In: IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, 1., 2001. Proceedings… Thessaloniki: IEEE, 2001. p. 902-905.

SCHNEIDER M. J.; BELLON O. R. P.; ARAKI H. Experimentos em fusão de imagens de alta resolução. Boletim de Ciências Geodésicas, v. 9, n.1, p.75-88, 2003.

SCHULTZ, B. et al. Classificação orientada a objetos em imagens multitemporais landsat aplicada na identificação de cana-de-açúcar e soja. Revista Brasileira de Cartografia, v.1, n. 68, p.131-143, 2016.

SOARES, R. B. Aplicação de técnica de fusão em imagens Landsat 8/ OLI. In: Simpósio Brasileiro de Sensoriamento Remoto, 17., 2015, João Pessoa. Anais…João Pessoa: INPE, 2015.p. 4836-4833.

WALD, L.; RANCHIN, T.; MANGOLINI, M. Fusion of satellite images of different spatial resolution: assessing the quality of resulting images. Photogrammetric Engineering & Remote Sensing, v. 63, n. 6, p. 691-699, 1997.

WALD, L. Some terms of reference in data fusion. IEEE Transaction on Geoscience and Remote Sensing. v. 37, n. 3, p.1190-1193, 1999.

Séries temporais de índices de vegetação para identificação de quebra de safra

Daniel de Castro Victoria[1], *Santiago Viana Cuadra*[2], *Aryeverton Fortes de Oliveira*[1]

[1] Embrapa Informática Agropecuária, Campinas, São Paulo, Brasil,

daniel.victoria@embrapa.br, ary.fortes@embrapa.br

[2] Embrapa Clima Temperado, Pelotas, Rio Grande do Sul, Brasil,

santiago.cuadra@embrapa.br

RESUMO

A agricultura está sujeita a intempéries que afetam a produção negativamente. Monitorar quebras causadas por clima adverso, em termos de quantidade perdida e área afetada, é de interesse para produtores, mercado consumidor e empresas do agronegócio. O sensoriamento remoto tem se mostrado útil para o monitoramento agrícola, principalmente com o uso de séries temporais. Esse trabalho utilizou séries de EVI/MODIS para calcular o Índice de Vegetação Padronizado (IVP), os quais foram comparados com a produtividade de soja nos municípios da região Sul do País. Os dados de sensoriamento remoto foram processados na plataforma de computação em nuvem *Google Earth Engine*, que se mostrou apropriada para manejar grandes volumes de dados. O IVP médio dos municípios mostrou boa relação com a produtividade de soja. Dessa forma, esse índice poderá ser utilizado para identificação de áreas afetadas por quebra de safra e para o monitoramento da produção agrícola.

PALAVRAS-CHAVE: Soja, Produtividade, Índice de vegetação, Sensoriamento remoto.

ABSTRACT

Agricultural production can be negatively affected by a range of climatic conditions. Monitoring production losses due to severe weather, both in quantity and spatial extent, is of interest to farmers, consumers and agribusiness industry in general. Remote sensing is a valuable asset in terms of agriculture monitoring, specially with the use time series. This work utilized EVI/MODIS time series to derive Standard Vegetation Index (SVI), which were

compared to municipal soy productivity in the south of Brazil. Google Earth Engine cloud platform was used to process all remote sensing data, and it proved to be an appropriate tool to handle large amounts of data. The average municipal SVI was related to soy productivity and helped identify municipalities that suffered crop failures in the last years. Thus, the SVI could be used to identify areas affected by droughts and monitor the agriculture production.
KEYWORDS: Soy, Productivity, Vegetation index, Remote sensing.

INTRODUÇÃO

A atividade agrícola está sujeita as mais variadas intempéries, que potencialmente podem causar redução ou perda quase completa da produção. Além dos prejuízos causados ao agricultor, problemas que afetam a produtividade podem ter consequências mais amplas, a depender do nível e da extensão geográfica da quebra. Problemas climáticos sistêmicos, que atingem grande parte das áreas de produção, alteram expectativas dos produtores, a oferta global de produtos agrícolas e levam a flutuações sazonais ou permanentes dos preços. A Política Agrícola Brasileira incorporou, desde a década de 1970, mecanismos de proteção da renda contra situações desfavoráveis criadas por oscilações de preços ou da produção.

Visando o abastecimento regular e a promoção da competitividade do agronegócio, a Política Agrícola contou com aperfeiçoamentos significativos, com instituição de programas de seguro rural e subvenção ao prêmio de seguro rural. Esses se basearam em avaliações de risco sobre a época de plantio, a partir do Zoneamento Agrícola de Risco Climático (ZARC), promovendo o uso eficiente dos recursos públicos envolvidos. Tais iniciativas aliviam os efeitos econômicos danosos de um evento de quebra para o produtor. No entanto, as perdas ainda poderão refletir em outros setores da economia, como preços ao consumidor, cotações no mercado internacional, inflação sazonal, a depender da extensão do dano. Dessa forma, a quantificação das quebras e a extensão geográfica das áreas afetadas é de grande interesse.

O monitoramento agrícola é uma etapa de gestão em processos decisórios distintos, subsidiando o planejamento e o desenvolvimento de atividades de transferência de riscos no setor privado e a reação do governo em eventos catastróficos. Interfere na avaliação privada de seguradoras, resseguradoras e agentes do agronegócio. Ademais, relações estabelecidas nos contratos de seguro, resseguro e apoio governamental efetivo em ocorrência de sinistros podem ser impactados significativamente com o monitoramento. Neste sentido, o sensoriamento remoto deve ser continuamente investigado e usado para prover informações ao público. Essas tecnologias vem sendo empregadas para tais fins, como o mapeamento da cobertura vegetal (LUNETTA et al., 2006; MORTON et al., 2006; VICTORIA et al., 2012) e

para o monitoramento das condições da vegetação e sua relação com precipitação (LEIVAS et al., 2014; SOUSA JÚNIOR; SAUSEN; LACRUZ, 2011; WANG; RICH; PRICE, 2003). Tais trabalhos focaram no monitoramento da estiagem. Esse trabalho vai uma etapa além e avalia a relação entre o Índice de Vegetação Padronizado (IVP) (PARK; KIM; CHOI, 2008), de séries temporais do *Enhanced Vegetation Index* (EVI/MODIS), com a produtividade da cultura da soja em municípios produtores da região sul do Brasil, entre os anos de 2004 e 2014.

MATERIAL E MÉTODOS

A área de estudo foi a região sul do País, considerando apenas os maiores produtores de soja, representados pelos municípios responsáveis por 95% da produção no ano de 2014. Ao todo foram analisados 201 municípios de um total de 1191. A análise utilizou dados tabulares de produção agrícola e área colhida por município, e imagens de sensoriamento remoto.

Dados tabulares

A produtividade municipal de soja nos anos de 1990 a 2014 foi calculada a partir da quantidade produzida (t) e área colhida (ha), com dados da Produção Agrícola Municipal (IBGE, [s.d.]). Informações da região Sul foram separadas e os municípios ranqueados de acordo com a produção em 2014. A produção acumulada foi utilizada para identificar os responsáveis por 95% da produção, limitando o número de municípios em 201. Somente séries com comprimento mínimo de 10 anos sem falhas foram consideradas.

Dados de sensoriamento remoto

O IVP calcula a anomalia de um índice de vegetação pela diferença em relação a média de longo prazo, ponderado pelo desvio padrão (Equação 1). Esse pode ser classificado em: muito abaixo do normal ($\leq$ -2,0), abaixo do normal (-2,0 < IVP $\leq$ -1,5), ligeiramente abaixo do normal (-1,5 < IVP $\leq$ -1,0), normal (-1,0 < IVP $\leq$ 1,0), ligeiramente acima do normal (1,0 < IVP $\leq$ 1,5), acima do normal (1,5 < IVP < 2,0) e muito acima do normal ($\geq$ 2,0) (SOUSA JÚNIOR; SAUSEN; LACRUZ, 2011).

$$IVP_{p,a} = \frac{EVI_{p,a} - média\left(EVI_p\right)}{sd\left(EVI_p\right)} \quad \text{Eq. 1}$$

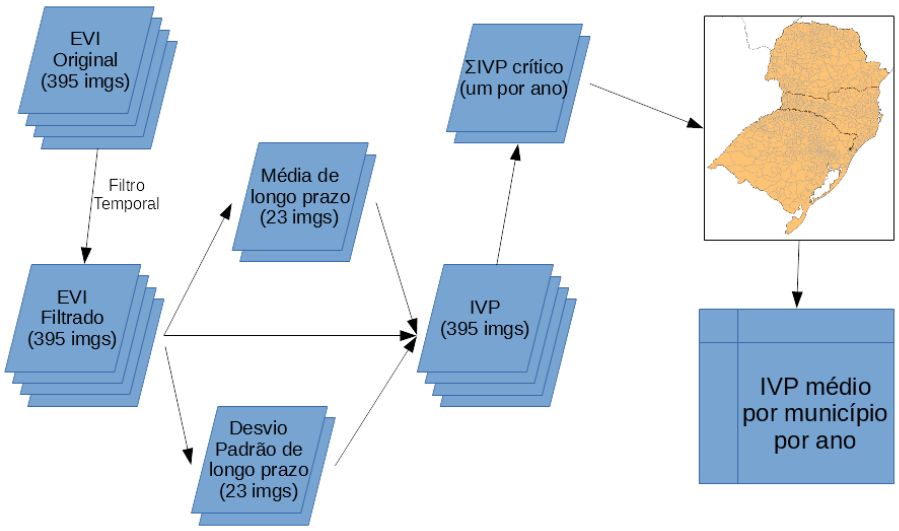
Onde *p* e *a* indicam o período e ano das imagens (IVP e EVI). *Média* e *sd* são a média e desvio padrão, respectivamente.

O IVP foi calculado a partir do EVI do sensor MODIS, de fevereiro de 2000 a maio de 2017, do produto MOD13Q1. Esse produto contém imagens compostas temporais de índices de vegetação (NDVI e EVI), a cada 16 dias (23 imagens/ano), com resolução espacial de 250 metros. Mais detalhes sobre o MODIS podem ser encontrados em Rudorff (2007).

São necessários dois *tiles* MODIS para recobrir a área de estudo. Cada *tile* abrange 10° x 10° de lat/long, um raster com 4800 x 4800 colunas/linhas e 270 Mb. A análise envolve 790 imagens (2 *tiles* x 395 passos de tempo), totalizando 208 Gb em dados brutos. Além disso, o processamento exige organização, armazenamento de produtos intermediários e *softwares* específicos para tratamento de imagens. Para contornar essas dificuldades, as imagens foram processadas na plataforma *Google Earth Engine* ("Google Earth Engine", [s.d.]). A vantagem está no fato da plataforma disponibilizar grande quantidade de dados já organizados (MODIS, Landsat, TRMM, etc), uma *API* (*application programming interface*) de processamento de imagens e dados vetoriais, além da infraestrutura computacional para a execução das rotinas. Assim, utilizando a *API* e linguagem *JavaScript*, é possível preparar e executar uma rotina nos servidores do Google. Outros trabalhos utilizaram plataforma, como o mapeamento global da dinâmica de florestas (HANSEN et al., 2013), de áreas alagadas (PEKEL et al., 2016) e do uso e cobertura das terras no Brasil ("MapBiomas", [s.d.]).

As etapas do processamento (Figura 1) compreenderam: i) seleção das imagens de EVI; ii) filtragem temporal; iii) cálculo da média e desvio padrão em cada período; iv) cálculo da série temporal de IVP; v) somatório do IVP no período crítico e vi) extração do valor médio do somatório de IVP em cada município.

Figura 1. Procedimento utilizado para cálculo do IVP médio por município por ano.



Fonte: autores

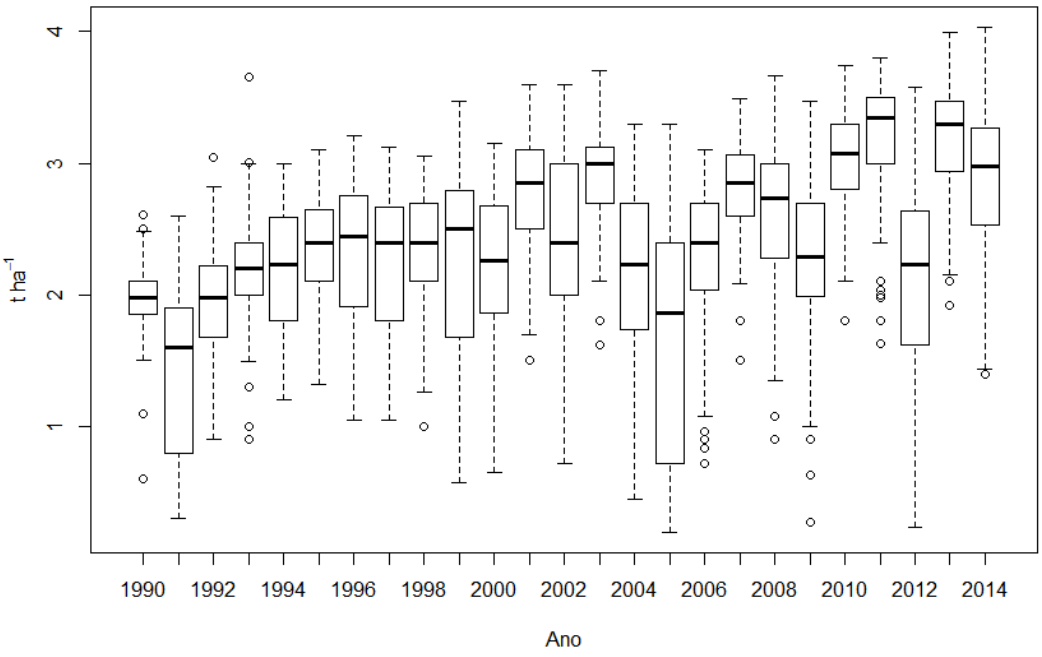
As imagens do produto MOD13Q1 são uma composição temporal onde grande parte influência das nuvens é retirada. No entanto, o produto ainda apresenta alguma interferência, a qual a filtragem temporal visa remover. O método utilizado, de remoção de mínimos locais (VICTORIA et al., 2012), preenche a série com um mínimo de alteração. O procedimento consiste em percorrer a série e, caso o EVI no tempo *t* seja menor que os valores em *t-1* e *t+1*, esse é considerado mínimo local e substituído pelo mínimo entre *t-1* e *t+1*. Em seguida são calculadas as imagens de EVI médio, do desvio padrão para cada período e do IVP.

A semeadura da soja na região sul ocorre nos meses de outubro a dezembro, com o período crítico para o desenvolvimento da cultura entre janeiro e fevereiro. Dessa forma, para cada ano foram somadas as 4 imagens de IVP do período crítico (1 e 17 de janeiro, 2 e 18 de fevereiro) e obtidos os valores médios em cada município da região sul, para cada ano avaliado. Por fim, esses valores foram comparados com a produtividade média municipal.

RESULTADOS E DISCUSSÃO

A produtividade média da soja nos municípios estudados aumentou entre os anos de 1990 e 2014 (Figura 2). No ano de 1990, a média municipal foi de 1,98 t ha^{-1}, alcançando 2,78 t ha^{-1} em 2001. A maior produtividade foi de 3,24 t ha^{-1} no ano de 2010 e a menor, 1,42 t ha^{-1} em 1991. É possível perceber anos em que a produtividade apresentou forte redução ou variação entre os municípios, com destaque para 1991, 2005, 2009 e 2012. Nas safras de 1990/91, 2004/05 e 2011/12 ocorreram perdas extremamente elevadas em diversos municípios da região Sul associadas a baixos índices pluviométricos. Na safra de 2004/05, o déficit de precipitação no Sul do Brasil causou grandes prejuízos à produção agrícola, que, de acordo com a Companhia Nacional de Abastecimento (Conab), foi 13 milhões de toneladas inferior ao esperado. Os maiores danos ocorram no Rio Grande do Sul, onde a produção total reduziu de 5,5 milhões para 2,2 milhões de toneladas, uma queda de 60% em relação ao esperado.

Figura 2. Produtividade de soja nos principais municípios produtores da região Sul de 1990 e 2014.



Fonte: autores

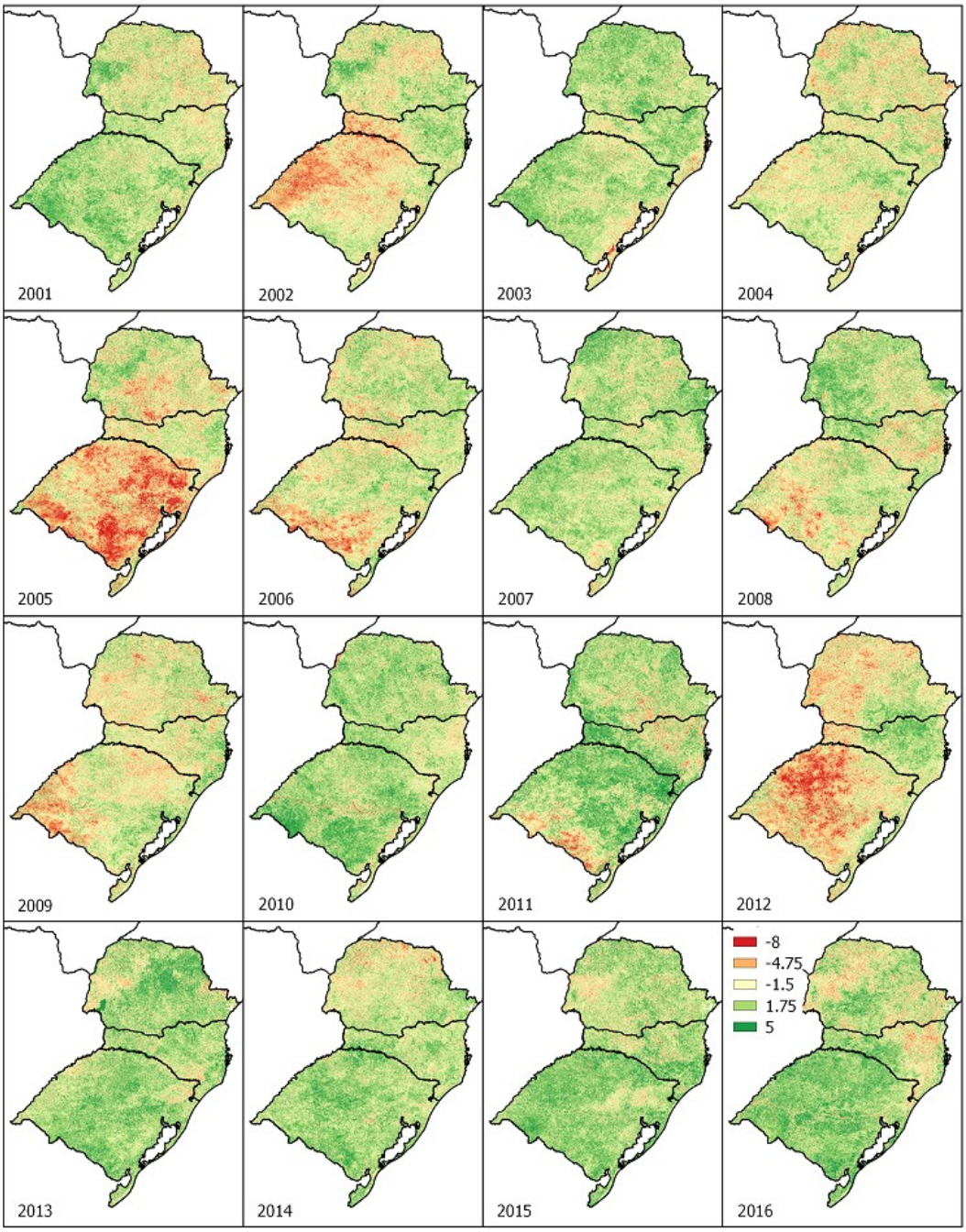
O *Google Earth Engine* se mostrou muito eficiente no processamento de grande volume de dados. Existe uma barreira inicial a sua utilização, principalmente para pessoas com pouco conhecimento de programação. No entanto, a *API* conta com boa documentação, tutoriais com exemplos práticos, além de uma ativa comunidade de usuários que auxilia na resolução de problemas. No que diz respeito a capacidade computacional, uma vez desenvolvido o *script*, todo o processamento para obtenção dos dados médios de IVP acumulado para os municípios da região sul, para um ano, foi executado em poucos minutos. Para fins de comparação, somente a etapa final do processamento, a extração de valores médios da imagem em diferentes polígonos, pode levar horas em um computador pessoal nos dias atuais, a depender da resolução espacial adotada e do número de polígonos.

O IVP acumulado indica a severidade da anomalia do EVI durante o período crítico da cultura. Normalmente o IVP de uma única imagem não atinge valores que excedam ±2. Dessa forma, valores muito baixos do IVP acumulado (i.e. -8) significam que mais de uma imagem no período apresentou forte anomalia, indicando condições adversas por período prolongado.

Os anos de 2002, 2005, 2009 e 2012 (Figura 3) apresentam extensas áreas na região sul com valores baixos de IVP acumulado, sendo 2005 e 2012 os mais críticos. Esses também são os anos em que a produtividade média municipal atingiu valores muito reduzidos, indicando relação entre o IPV acumulado e a produtividade. Para os anos de 2002 e 2009, os

mapas de IVP não apresentam condições tão negativas quanto os demais. Porém, esses também são anos em que a produtividade média municipal apresentou valores reduzidos.

Figura 3. IVP acumulado para a região sul, nos anos de 2001 a 2016

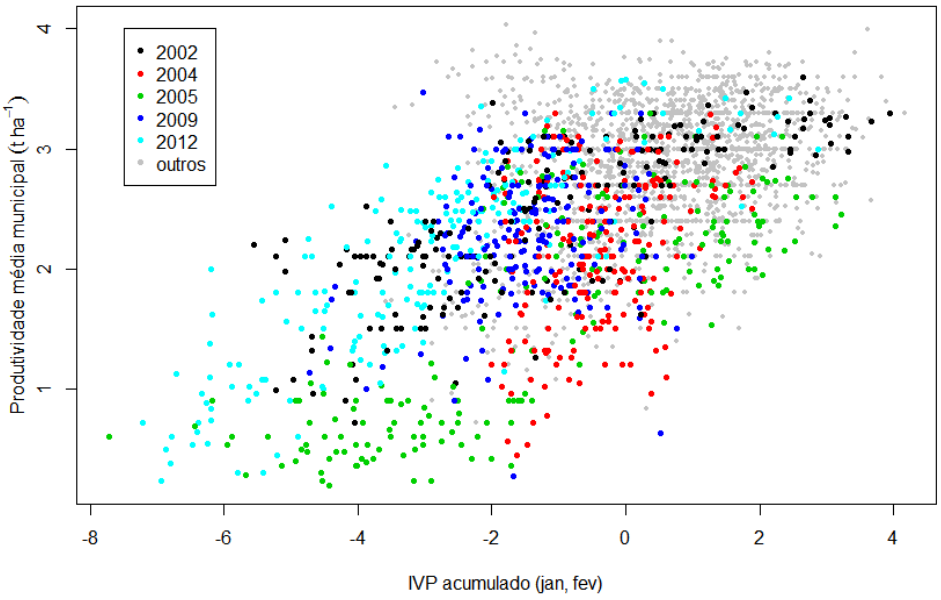


Fonte: autores

O IVP acumulado apresentou relação com a produtividade média (Figura 4). A grande maioria dos municípios com IVP < -4 tiveram produtividade < 2 t ha^{-1}. A regressão linear foi significativa ($p < 0,001$), apesar da alta dispersão ($r^2 = 0,37$). Em 2005 e 2012 é possível notar

mais facilmente a relação entre IVP e produtividade. Já em 2002, 2004 e 2009, a dispersão dos dados é maior. Isso indica que o método é mais útil para detectar grandes quebras.

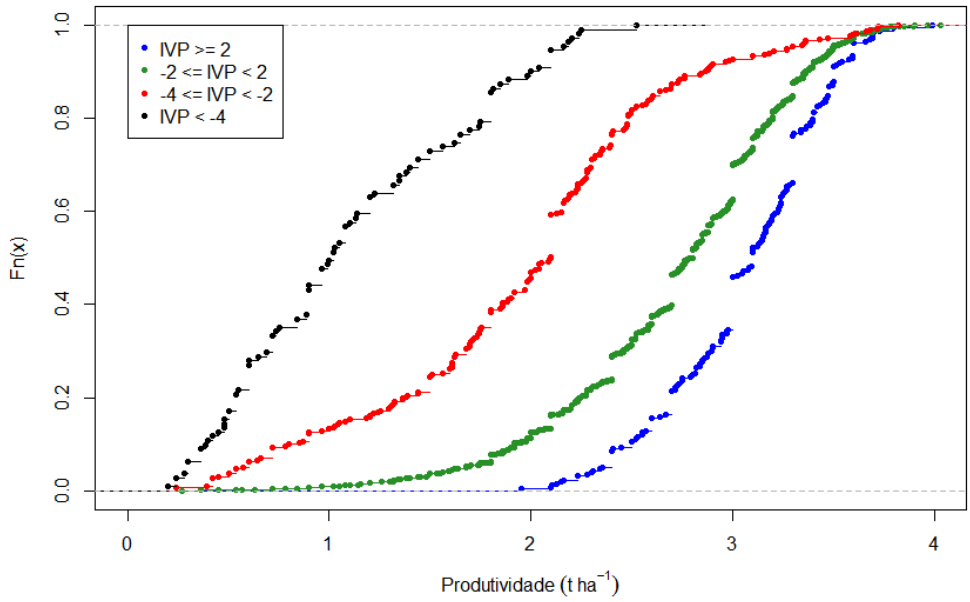
Figura 4. IVP acumulado e produtividade da soja para municípios da região Sul. Anos com quebras em destaque.



Fonte: autores

A distribuição cumulativa das produtividades municipais, em diferentes faixas de IVP, também mostrou a importância desse indicador (Figura 5). Para municípios e anos em que o IVP >= 2, a produtividade mínima foi de 1,95 t ha^{-1} e 75% dos dados apresentaram produtividade até 3,36 t ha^{-1}. No outro extremo, com IVP <= -4, a mínima foi de 0,2 t ha^{-1} e 75% das observações foram inferiores a 1,70 t ha^{-1}. As distribuições nas diferentes faixas de IVP são estatisticamente distintas pelo teste de Kolmogorov–Smirnov.

Figura 5. Função de distribuição cumulativa das produtividades municipais de soja, em diferentes faixas de IVP.



Fonte: autores

A resolução espacial de 250 m do MODIS possibilita identificar com maior precisão os municípios afetados por condições adversas. Mesmo em anos com quebras muito fortes, é possível identificar municípios que apresentaram elevado IVP acumulado, ou seja, estão localizados em áreas distantes das regiões críticas. Dessa forma, é possível avaliar, de maneira rápida e precisa, a extensão geográfica das áreas afetadas por eventos de quebra de safra.

CONCLUSÕES

A plataforma *Google Earth Engine* se mostrou uma alternativa viável e prática para o processamento de grande volume de dados geoespaciais.

Nas últimas décadas, a região Sul passou por eventos de quebra de safra, relacionados a condições climáticas adversas, afetando a produtividade média municipal. O IVP identificou as regiões e épocas mais afetadas, mostrando potencial para monitoramento em larga escala das condições gerais de desenvolvimento das culturas, promissor para avaliações municipais e regionais. Anos mais críticos apareceram com sinais claros de perda no desenvolvimento da biomassa, altamente correlacionados com oscilações significativas de produtividade.

A relação entre produtividade primária de biomassa e IVP pode ser maior que a relação com a produtividade final de grãos. Efeitos tecnológicos podem favorecer a produção e enchimento de grãos de maneira não aparente no sensoriamento remoto, o que pode levar o IVP a ser um indicador associado mais fortemente com modelos agroclimáticos ajustados em condições atuais de produção e utilizados para simulação do crescimento de culturas sobre

séries históricas. Este resultado pode ser uma importante informação para validação e melhoria da qualidade dos resultados de modelos utilizados no Zoneamento Agrícola.

REFERÊNCIAS

Google Earth Engine. Disponível em: <https://earthengine.google.com/>. Acesso em: 19 maio. 2017.

HANSEN, M. C. et al. High-Resolution Global Maps of 21st-Century Forest Cover Change. **Science**, v. 342, n. 6160, p. 850–853, 15 nov. 2013.

IBGE. **Sistema IBGE de Recuperação Automática - SIDRA**. Disponível em: <http://www.sidra.ibge.gov.br/>. Acesso em: 5 nov. 2014.

LEIVAS, J. F. et al. AVALIAÇÃO DO ÍNDICE DE VEGETAÇÃO PADRONIZADO NO MONITORAMENTO INDICATIVO DE ESTIAGENS EM PERÍODOS CRÍTICOS DA SOJA NO SUL DO BRASIL. **Revista Brasileira de Cartografia**, v. 5, n. 66/5, 4 nov. 2014.

LUNETTA, R. S. et al. Land-cover change detection using multi-temporal MODIS NDVI data. **Remote Sensing of Environment**, v. 105, n. 2, p. 142–154, 30 nov. 2006.

MapBiomas. Disponível em: <http://mapbiomas.org/>. Acesso em: 19 maio. 2017.

MORTON, D. C. et al. Cropland expansion changes deforestation dynamics in the southern Brazilian Amazon. **Proceedings of the National Academy of Sciences**, v. 103, n. 39, p. 14637–14641, set. 2006.

PARK, J.-S.; KIM, K.-T.; CHOI, Y.-S. **Application of Vegetation Condition Index and Standardized Vegetation Index for Assessment of Spring Drought in South Korea**. IEEE, 2008. Disponível em: <http://ieeexplore.ieee.org/document/4779463/>. Acesso em: 19 maio. 2017

PEKEL, J.-F. et al. High-resolution mapping of global surface water and its long-term changes. **Nature**, v. 540, n. 7633, p. 418–422, 7 dez. 2016.

RUDORFF, B. F. T. R. **Sensor Modis e Suas Aplicações Ambientas no Brasil**. [s.l.] Editora Parêntese, 2007.

SOUSA JÚNIOR, M. DE A.; SAUSEN, T. M.; LACRUZ, M. S. P. **Monitoramento de estiagem an região sul do Brasil utilizando dados EVI/Modis no período de Dezembro de 2000 a Junho de 2009**. XV SBSR. **Anais**... In: SBSR. Curitiba, PR, Brasil: INPE, 2011. Disponível em: <http://mtc-m16d.sid.inpe.br/col/sid.inpe.br/mtc-m19@80/2010/05.07.16.56/doc/publicacao.pdf>. Acesso em: 25 maio. 2017

VICTORIA, D. DE C. et al. Cropland area estimates using Modis-NDVI times series in the state of Mato Grosso, Brazil. **Pesquisa Agropecuária Brasileira**, v. 47, n. 9, p. 1270–1278, 9 nov. 2012.

WANG, J.; RICH, P. M.; PRICE, K. P. Temporal responses of NDVI to precipitation and temperature in the central Great Plains, USA. **International journal of remote sensing**, v. 24, n. 11, p. 2345–2364, 2003.

Aplicação web para criação de mapas temáticos em agricultura de precisão

Ligia Francielle Borges[1], Claudio Leones Bazzi[2], Kelyn Schenatto[3], Eduardo Godoy de Souza[4], Fernando Nakayama de Queiroz[5]

[1] Discente do programa em Tecnologias Computacionais para o Agronegócio, Universidade Tecnológica Federal do Paraná, Medianeira, Paraná, Brasil, ligia.lfb@gmail.com

[2] Departamento de Ciência da Computação, Universidade Tecnológica Federal do Paraná, Medianeira, Paraná, Brasil, bazzi@utfpr.edu.br

[3] Departamento de Ciência da Computação, Universidade Tecnológica Federal do Paraná, Santa Helena, Paraná, Brasil, kschenatto@utfpr.edu.br

[4] Centro de Ciências Exatas e da Terra, Universidade Estadual do Oeste do Paraná, Cascavel, Paraná, Brasil, Eduardo.Souza@unioeste.br

[5] Mestre em Engenharia Agrícola, Universidade Estadual do Oeste do Paraná, Cascavel, Paraná, Brasil, fernandonakayama@gmail.com

RESUMO

A análise de mapas temáticos, no contexto da agricultura de precisão (AP), é um recurso utilizado para compreender a variabilidade espacial dos atributos do solo em áreas de cultivo, podendo trazer importantes subsídios para a realização de um manejo racional, promovendo aumento da produtividade e redução do impacto ambiental. Neste contexto, o objetivo desse trabalho foi desenvolver e disponibilizar uma ferramenta computacional em ambiente web, que permite a geração de mapas temáticos fazendo uso de métodos tradicionais de interpolação de dados sugeridos na literatura (inverso da distância elevado à uma potência, média móvel e vizinho mais próximo). Os resultados obtidos mostraram que o uso do software permitiu que fossem gerados mapas temáticos de variáveis normalmente utilizadas em agricultura de precisão, possibilitando avaliar a distribuição de cada atributo na área a ser trabalhada e identificar possíveis causas de variação na produtividade. Por estar instalado em servidor com recursos sob demanda, o software possui desempenho satisfatório do ponto de vista funcional e pode ser acessado de qualquer ambiente web.

PALAVRAS-CHAVE: Software, Variabilidade espacial, Tecnologia agrícola.

ABSTRACT

The analysis of thematic maps, in the precision agriculture (AP) context, is a resource used to understand the spatial variability of soil attributes in cultivated areas, which can bring important subsidies for rational handling, improving yield and reducing environmental impact. Thus, the main goal of this research was to develop and make available a web-based software, which provides the ability to draw thematic maps according to well-known methods of data interpolation (inverse of distance weighted, moving average, nearest-neighbor). The use of the software made possible generating thematic maps of variables usually used in precision agriculture, allowing to evaluate the distribution of each attribute in the area and identify possible causes of variations in yield. As the software is installed in a on-demand server, it provides good performance from a functional point of view and can be accessed from any web environment.

KEYWORDS: Software, Spatial variability, Agricultural technology.

INTRODUÇÃO

A agricultura de precisão (AP) pode ser definida como uma forma moderna de gerenciamento da atividade agrícola que objetiva aplicar os insumos de produção no local correto, no momento adequado e nas quantidades necessárias para áreas cada vez menores e homogêneas, tanto quanto a tecnologia e os custos envolvidos permitam manter a rentabilidade (DOBERMANN; PING, 2004), uma vez que, dentro de uma mesma lavoura é possível encontrar subáreas com valores e qualidades químicas diferenciados, o que pode ocasionar diferentes possibilidades de respostas à aplicação de insumos (AMADO et al., 2005).

Segundo Demattê et al. (2007), os sistemas de informações geográficas (SIGs) são sistemas gerenciadores de informações referentes a dados de solos, culturas, climas e topografia. Estas informações ao serem armazenadas no sistema, são aliadas aos objetivos do usuário que gerencia e cruza dados para obter informações. Na AP, a utilização de um SIG é imprescindível para gerenciar e analisar dados georreferenciados relativos as propriedades do solo e a produtividade das culturas (SRINIVASAN, 2006).

De acordo com Blackmore e Larscheid (1997), na AP o manejo da variabilidade pode ser classificado como variabilidade espacial, adotada para medir as mudanças dos atributos na área de plantio, geralmente é apresentado por meio de mapas temáticos. Para Molin (2002), o mapeamento da produtividade é uma parte essencial da agricultura de precisão, o mapa é a

informação completa para se visualizar a variabilidade espacial e auxiliar no diagnóstico da lavoura, mas para que isso seja possível é necessário a utilização de equipamentos de GPS, que o produtor agrícola tenha acesso a um software com um banco de dados que gere relatórios, mapas de produtividade e que seja capaz entender esses mapas para uma forma que possa ser usada no controle e monitoramento do implemento, ou seja, a AP possui uma forte relação com a tecnologia da informação.

Os softwares utilizados para auxiliar nos ciclos da AP, desde a navegação até os sistemas de controle, demandam conhecimento especifico da ferramenta, uma vez que para cada marca, o formato dos arquivos, instruções de equipamentos e registro de informação são únicos. Como geralmente esses programas são desenvolvidos para uma única tarefa, existe a necessidade de utilizar ferramentas complementares. Na maioria dos casos, os softwares para geração de mapas temáticos não foram especificamente desenvolvidos para a AP, mas sim para manipulação de dados genéricos, o que pode tornar mais desafiador o uso deste programa no lugar de uma ferramenta dedicada para a AP (WHELAN; TAYLOR, 2013).

O objetivo desse trabalho foi desenvolver e disponibilizar gratuitamente uma ferramenta computacional para criação de mapas temáticos que possibilite aos produtores rurais fazer uso de técnicas de AP visando o aumento da produtividade e redução de impactos ambientais e de custos de produção, realizando a gestão das áreas produtoras de forma otimizada.

MATERIAL E MÉTODOS

Para o desenvolvimento do projeto buscou-se utilizar ferramentas livres, para viabilizar o projeto sem necessidade de custos adicionais com licença ou suporte. O sistema gerenciador de banco de dados (SGBD) PostgreSQL foi utilizado devido a sua flexibilidade e sua extensão para dados georreferenciados por sua extensão PostGIS que mantém compatibilidade com linguagens de programação tradicionais e possui capacidade ilimitada para base de dados.

A linguagem de marcação HTML (Hyper Text Markup Language) foi utilizada na camada de estruturação de conteúdo da página Web e para a camada de apresentação que é constituída pelas folhas de estilos, foi utilizado à linguagem CSS (Cascading Style Sheets), que permite definir como os elementos que compõem uma página, um documento ou aplicação Web serão exibidos. Na camada de comportamento constituída pelos scripts, a linguagem de programação JavaScript, foi utilizada para manipulação e processamento de dados.

Para a comunicação entre a linguagem de desenvolvimento e o banco de dados, foi utilizado o framework de licença livre Hibernate, que permite o desenvolvimento de

aplicativos com mais facilidade pois converte as classes Java diretamente em tabelas no banco de dados. Para a manipulação de eventos como animações, foi utilizado a biblioteca JQuery, que é desenvolvida em JavaScript e o framework brasileiro VRaptor 3, foi utilizado para o desenvolvimento de aplicações Java Web por ser considerado de fácil utilização e por proporcionar o desenvolvimento mais ágil quando comparado com métodos tradicionais.

Para o desenvolvimento da aplicação foi utilizado o ambiente de desenvolvimento integrado (IDE) Eclipse, que possui licença livre e que reúne características e ferramentas de apoio ao desenvolvimento de softwares com o objetivo de agilizar este processo. O servidor Web Java Tomcat foi utilizado em conjunto com o IDE Eclipse para compilar o programa.

O sistema foi hospedado em servidor próprio, localizado no departamento de computação da Universidade Tecnológica Federal do Paraná - câmpus Medianeira, disponibilizado e mantido com recursos do próprio departamento, para fins de pesquisa.

As imagens de satélite exibidas no software e utilizadas no ambiente de desenvolvimento são de licença livre e disponibilizadas pela biblioteca OpenLayers 3. Essa ferramenta, desenvolvida em linguagem JavaScript, possui integração com serviços online de mapeamento, possibilitando aos desenvolvedores a criação de aplicações geográficas para o ambiente Web com capacidade de exibir e manipular dados geoespaciais em navegadores Web.

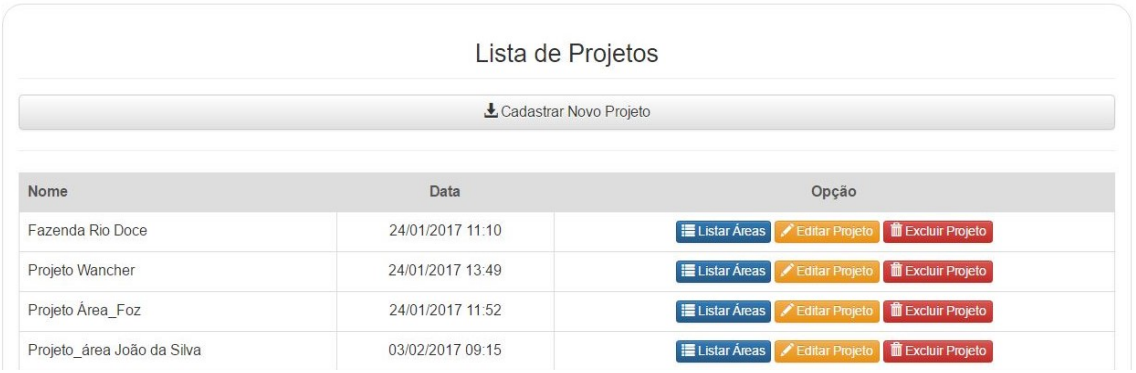
A arquitetura de software para geração de mapas temáticos (Figura 3), foi construída com base nos conceitos de REST (get; post; put; delete), para auxiliar na integração com diferentes sistemas utilizando a API REST (interface entre dois programas diferentes que permite a comunicação entre eles).

Após a conclusão da etapa de programação, o software passou por testes de desempenho e acurácia. Para cada tela e funcionalidade específica foram realizados casos de testes visando minimizar resultados indesejados e corrigir erros. Além dos testes de conteúdo, navegação, interface e compatibilidade, foram desenvolvidos testes que permitem identificar erros na comunicação entre a WebApp e o banco de dados. Também foram realizados testes da camada WebApp do lado do servidor para assegurar que os dados do usuário, são extraídos e transmitidos corretamente.

RESULTADOS E DISCUSSÃO

A primeira etapa de utilização do software envolve a criação de projetos onde são gerenciados os talhões e grades amostrais de cada propriedade (Figura 1). Após a criação dos projetos é possível que o usuário acesse a lista de todos os seus projetos cadastrados.

Figura 1 – Lista de projetos.



Conforme fluxograma (Figura 2), após a criação do projeto são iniciados os procedimentos de inserção das áreas de estudo e importação dos dados amostrais referentes aos atributos químicos e físicos do solo e produtividade da cultura que foram coletados e analisados previamente. Para o cadastro das áreas de estudo e amostras são utilizados arquivos do tipo texto contendo as coordenadas geográficas do contorno da área, dos pontos amostrais e os valores dos atributos (Figuras 3 e 4).

Figura 2 – Fluxograma de funcionamento do software.

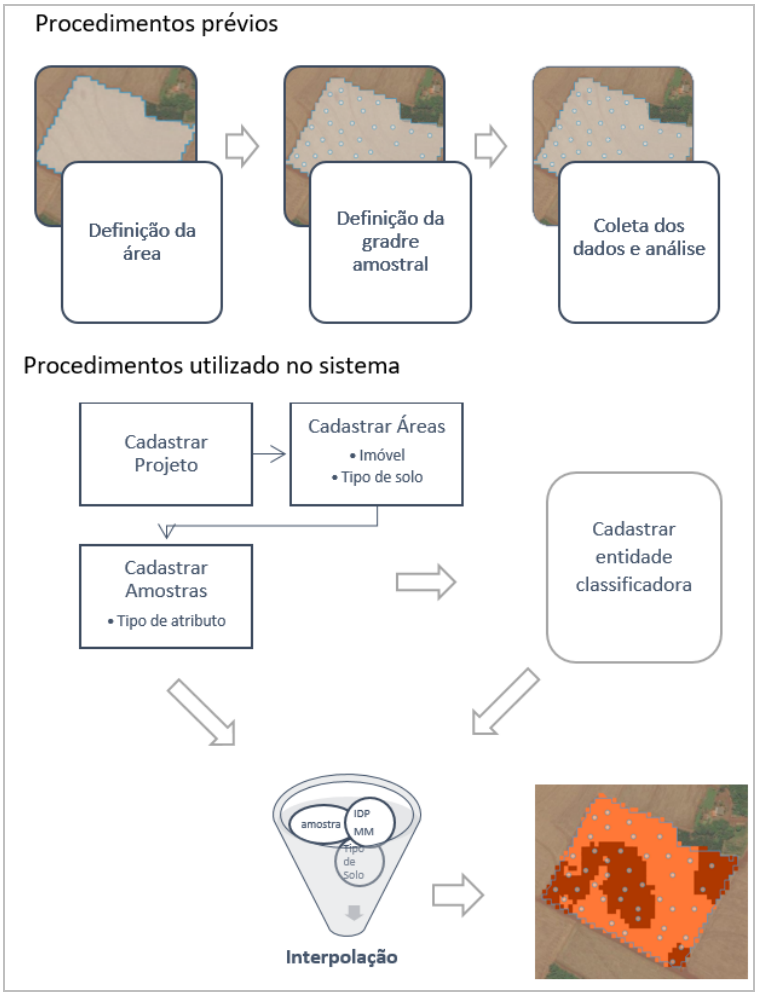


Figura 3 – Lista de atributos.

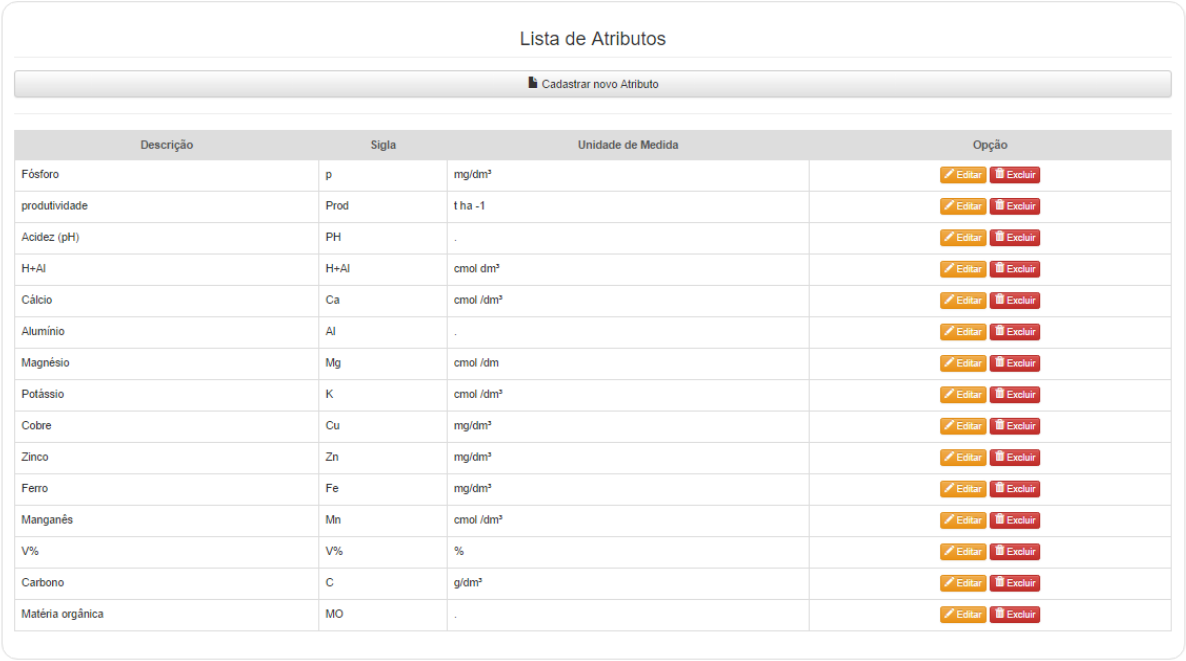
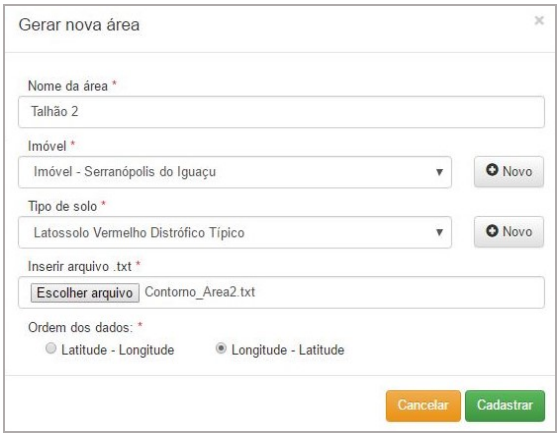
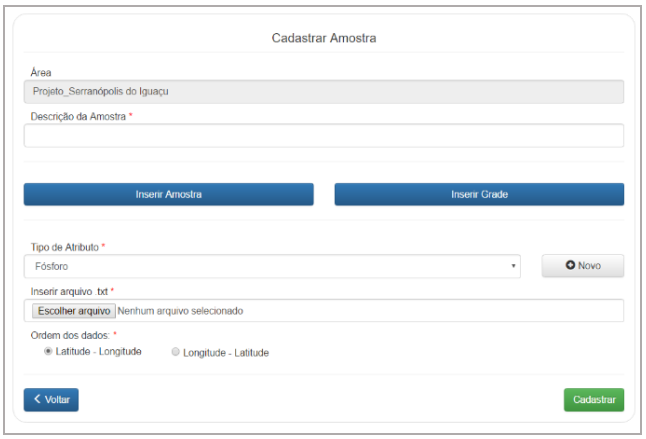


Figura 4 – Módulos para o cadastro de área por arquivo texto e módulo para cadastro de amostras.





Após o cadastro das áreas e seus atributos, o sistema permite a edição do contorno da área e dos valores dos pontos amostrais, assim como sua exclusão, localização ou informações (Figura 5). Não é permitido o trabalho simultâneo de duas áreas, mas duas áreas ou amostras podem ser adicionadas a mais de um projeto. Após os demais dados serem cadastrados, é possível realizar a interpolação dos dados para a geração de mapas temáticos. Para essa etapa recomenda-se primeiramente realizar o cadastro de uma entidade classificadora corresponde a uma empresa, órgão ou pesquisador de referência sobre classificações de atributos como Mg, P, Ca, Mn, Fe, etc, em forma tabelada que se pretende gerar mapas. O cadastro pode ser acessado por meio do menu principal em entidade classificadora, que permite realizar o cadastro dos intervalos e cor para cada nível de classificação (Figura 6).

Figura 5 – Módulo para edição de uma amostra.

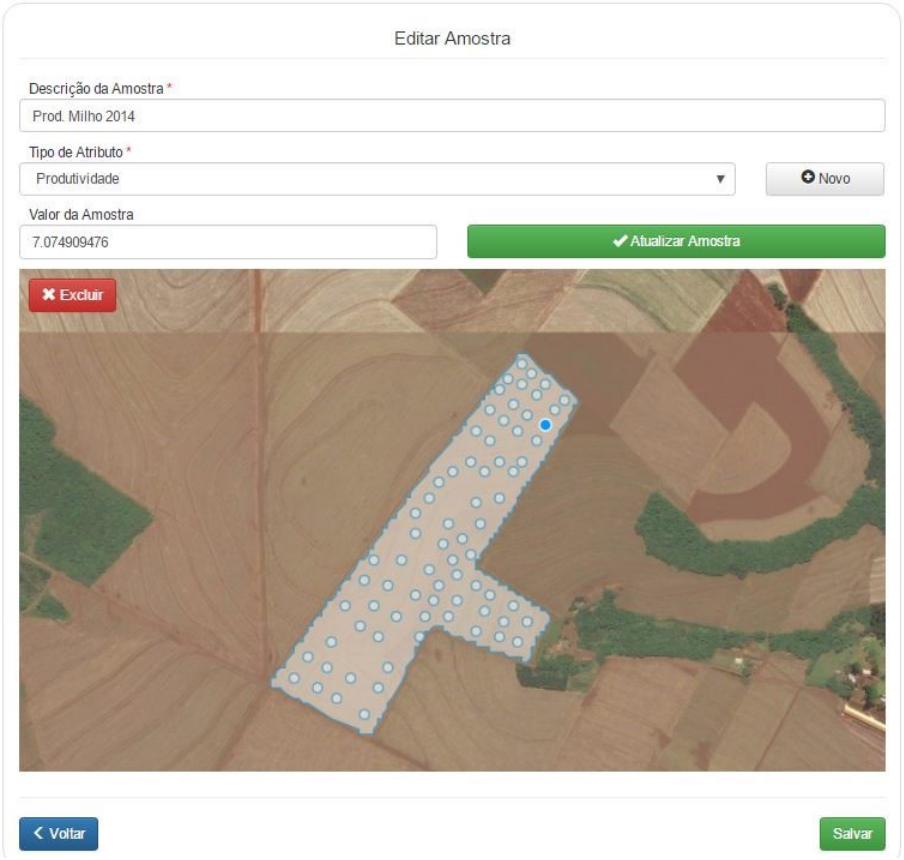
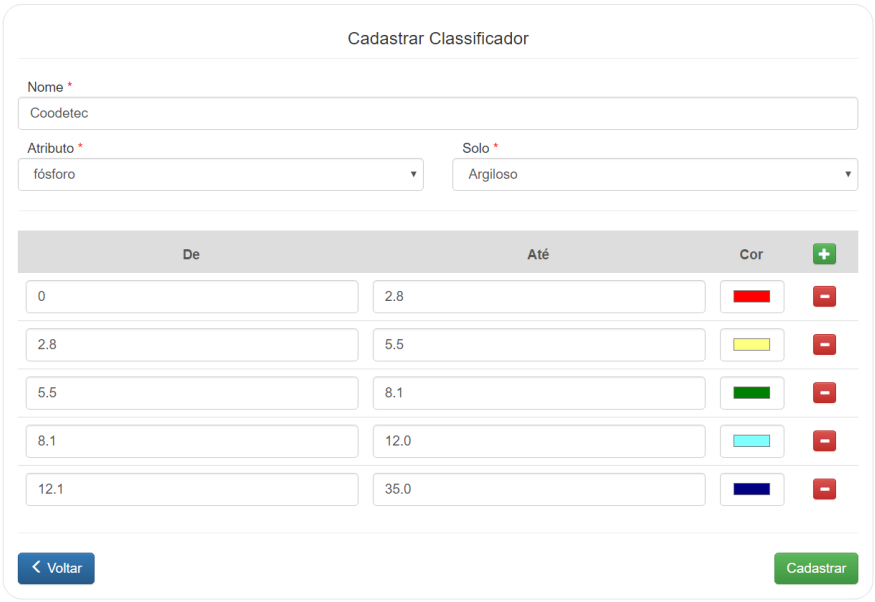


Figura 6 – Cadastro de entidade classificadora.



O ambiente permite a geração de mapas temáticos por meio do submenu gerar mapas interpolados disponível no menu principal. Na tela de cadastro o usuário tem a possibilidade de configurar as opções conforme sua necessidade, considerando parâmetros referentes ao tipo de interpolador (Inverso da Distância, Inverso da Distância Elevado a uma Potência e

Média Móvel), ao tamanho do pixel do mapa (resolução espacial), os valores do expoente (somente para o IDP), número de vizinhos e raio. Um nome para o mapa deve ser preenchido para que o mesmo seja gerado, sendo que recomenda-se utilizar nomes sugestivos em relação ao período e atributo que se está trabalhando (Figura 7).

Figura 7 – Mapa gerado utilizando o método de interpolação IDP com pixel de 5x5m.



CONCLUSÕES

Em relação à plataforma desenvolvida pode se concluir que o sistema se mostrou uma opção para geração de mapas temáticos, utilizando métodos de interpolação inverso da distância elevado a uma potência e média móvel.

Através dos mapas temáticos gerados no sistema Web, foi possível identificar a variabilidade dos atributos estudados.

AGRADECIMENTOS

À Universidade Tecnológica Federal do Paraná, à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) e ao Ministério de Agricultura Pecuária e Abastecimento – MAPA, pelo apoio financeiro.

REFERÊNCIAS

AMADO, T.J.C.; NICOLOSO, R.; LANZANOVA, M.; SANTI, A.L.; LOVATO, T. A compactação pode comprometer os rendimentos de áreas sob plantio direto. Revista Plantio Direto, v. 1, n. 89, p.34-42, 2005.

BLACKMORE, B.S.; LARSCHEID, G. Strategies for managing variability. In: Europe an Conference on Precision Agriculture, Warwick, Proceedings... London: BIOS Scientific, p. 851-859, 1997.

DEMATTÊ, J.A.M.; GALDOS, M.V.; GUIMARÃES, R.; GENÚ, A.M.; NANNI, M.R.; ZULLO J, J. Quantification of tropical soil attributes from ETM+/Landsat-7 data. International Journal of Remote Sensing, v. 28, n. 17, p. 3813-3829, 2007.

DOBERMANN, A.; PING, J.L. Geostatistical integrativon of yield monitor data and remote sensing improves yield maps. Agronomy Journal, v. 96, p. 285-297, 2004.

MOLIN, J. P. Agricultura de Precisão: O gerenciamento da variabilidade. Engenharia Agrícola, v. 22, n. 1, p. 83-92, 2002.

SRINIVASAN, A. Hand book of precision agriculture: principles and applications. Haworth Press, cap.1, New York. 2006.

WHELAN, B.; TAYLOR, J. Precision Agriculture for Grain Production Systems. Collingwood, Australia: CSIRO Publishing Determination of management zones for a tobacco field based on soil fertility, 2013.

BDGF: um sistema web para recuperação de informação de genótipos e fenótipos

Fábio Danilo Viiera[1], Danilo Gomes de Moura[1], Diego Félix da Silva[2], Roberto Hiroshi Higa[1], Adhemar Zerlotini[1]

[1] Embrapa Informática Agropecuária, Campinas, São Paulo, Brasil

[2] Programa Geneplus-Embrapa, Campo Grande, Mato Grosso do Sul, Brasil

fabio.vieira@embrapa.br, danilo.moura@colaborador.embrapa.br, dfelixsilva@hotmail.com, roberto.higa@embrapa.br, adhemar.zerlotini@embrapa.br

RESUMO

Nos últimos anos, o uso de genotipagem em grande escala de dezenas ou centenas de milhares de polimorfismos de nucleotídeo único (SNP) para estimar o perfil genômico de animais permitiu o desenvolvimento de estudos de associação genótipo-fenótipo em escala genômica (GWAS) e a introdução da tecnologia de seleção genômica em programas de melhoramento genético. No entanto, esta situação implica na necessidade de armazenamento de grande volume de dados de genotipagem, fenotipagem e pedigree de um elevado número de animais. Para integrar esse volume de dados distintos, é primordial utilizar uma estrutura de armazenamento robusta, como um SGBD. Assim, uma questão importante a considerar é o equilíbrio entre normalização e desempenho durante o estágio de modelagem de banco de dados, pois isso terá um impacto direto na usabilidade e na experiência do usuário. Buscando resolver esse problema de armazenamento eficiente e consultas rápidas num grande volume de dados, este trabalho apresenta o sistema BDGF (Banco de Dados de Genótipos e Fenótipos). Seu modelo de dados permite a implementação do tipo JSON em campos de tabelas relacionadas a fenótipos e do tipo texto em campos da tabela genótipos. BDGF foi projetado para dar suporte a projetos de criação de animais da Embrapa, mas pode ser facilmente ajustado para armazenar dados de diversas fontes, tais como dados de plantas. Além disso, o sistema implementa políticas de acesso e segurança para fenótipos, genótipos e pedigree dos animais.

PALAVRAS-CHAVE: Tecnologias Java EE, Sistema Gerenciador de Banco de Dados, PostgreSQL, Polimorfismo de Nucleotídeo Unico, JSON.

ABSTRACT

In recent years, the use of large scale genotyping of tens or hundreds of thousands of Single Nucleotide Polymorphisms (SNP) to estimate the genomic profile allowed the development of both genotype-phenotype association studies in genomic scale (genome-wide association studies-GWAS) and the introduction of genomic selection technology in breeding programs. However, this situation implies the need of storing large volumes of genotyping, phenotyping and pedigree data from large numbers of animals. In order to effectively integrate such amount of distinctive datasets, it's advisable to use a robust storage structure, such as a DBMS. Therefore, a major issue to consider is the trade off between normalization and performance during the database modeling stage, as this will have a direct impact on the usability and user experience. In order to get efficient storage and fast queries in this high volume of data, in this work we present the BDGF system (Genotypes and Phenotypes Database). Its data model allows the implementation of JSON type in table fields related to phenotypes and character varying in table fields related to genotypes. BDGF is designed to support the animal breeding projects of Embrapa, but can be easily adjusted to store data from diverse sources, such as plant data. Furthermore, the system implements access and security policies to phenotypes, genotypes and pedigree of the animals.

KEYWORDS: Java EE Technologies, Database Management System, PostgreSQL, Single Nucleotide Polymorphism, JSON.

INTRODUÇÃO

Nos últimos anos, a utilização da tecnologia de genotipagem em larga escala de milhares de marcadores moleculares do tipo *Single Nucleotide Polymorphisms* (SNP) para estimar o perfil genômico de animais permitiu o desenvolvimento tanto de estudos de associação genótipo-fenótipo em escala genômica (do inglês *genome-wide association studies* - GWAS) quanto a introdução da tecnologia de seleção genômica em programas de melhoramento genético. As tecnologias atuais para geração de dados moleculares são capazes de genotipar de dezenas até centenas de milhares de marcadores SNP em um único ensaio para cada indivíduo, com enorme velocidade e automação (CAETANO, 2009).

Contudo, essa situação implica na necessidade de armazenamento de um grande volume de dados, não somente de genótipos, mas também de fenótipos e pedigree de um

número cada vez maior de animais. Dessa forma, realizar o armazenamento adequado e a extração de conhecimento útil a partir dessa quantidade de dados torna-se um grande desafio. Dado o volume de dados considerado (centenas de milhares de animais fenotipados e, possivelmente, genotipados para centenas de milhares de marcadores do tipo SNPs), uma questão importante a se considerar no desenvolvimento de uma solução computacional é a adequabilidade da modelagem do banco de dados à aplicação desejada, pois esta terá impacto direto nos tempos de consulta e escrita em sistemas gerenciadores de bancos de dados relacionais (RDBMS – do inglês *Relational Database Management System*) onde essa informação estará armazenada.

Diante disso, com o objetivo de fornecer uma solução que fosse eficiente tanto no armazenamento quanto na consulta desse alto volume de dados, o sistema Web Banco de Dados de Genótipos e Fenótipos (BDGF) foi desenvolvido utilizando um diagrama de dados inicialmente proposto por (HIGA e OLIVEIRA, 2015). Esse diagrama foi redesenhado de forma que possibilitasse a implementação do tipo *JavaScript Object Notation* (JSON) em campos de tabelas relacionadas a fenótipos e do tipo texto (*character varying*) em campos de tabelas relacionadas a genótipos do diagrama. Com a implementação do tipo JSON e do tipo texto nessas tabelas, foi possível o uso da abordagem *Not Only SQL* (NoSQL - http://nosql-database.org) para armazenar parte dos dados sem que seja necessário se importar com a normalização dos mesmos, agilizando consultas que necessitariam realizar junções (*joins*) com outras tabelas.

MATERIAL E MÉTODOS

Material

O sistema está hospedado em computadores dedicados ao desenvolvimento e servidores (virtuais) para testes e homologação. Para o desenvolvimento do sistema Web, foram escolhidos componentes de tecnologia de informação (TI) disponível no mercado, ou seja, dentro da filosofia do uso de *software* livre. Primeiramente, como sistema operacional, optou-se pelo Linux Ubuntu (https://www.ubuntu.com) em todos os equipamentos.

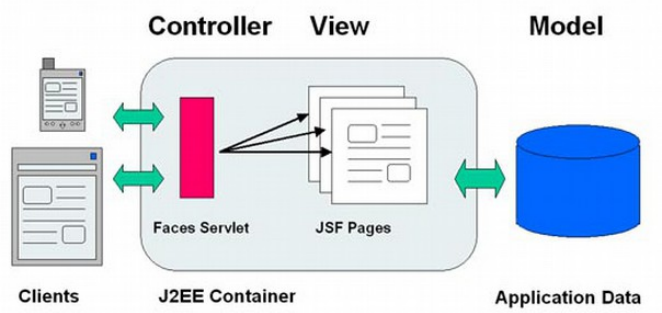
O sistema gerenciador de banco de dados (SGBD) escolhido foi o PostgreSQL (https://www.postgresql.org), versão 9.5. O PostgreSQL foi selecionado por ser um SGBD confiável, amplamente utilizado no mercado e pelo qual a equipe de desenvolvimento tem conhecimentos mais avançados. Além disso, oferece o formato JSON para tipificar os campos de tabelas. O tipo JSON é oferecido pelo PostgreSQL em suas últimas versões, a partir da 9.2. A tecnologia JSON consiste num formato de padrão aberto que consiste de conjuntos de pares

na forma "chave:valor". Ela foi utilizada para que fosse possível viabilizar o uso da abordagem NoSQL para armazenar parte dos dados sem que seja necessário se importar com a normalização dos mesmos.

O *software* utilizado para controle de versão foi o Subversion (SVN), versão 1.8.8. Já a linguagem de programação escolhida foi Java (https://www.oracle.com/br/java/), versão 8, e seus componentes da tecnologia *Java Enterprise Edition* (Java EE), que consiste de um conjunto de serviços e de interfaces de programa de aplicação (API) para o desenvolvimento de sistemas corporativos.

Dentre as tecnologias Java EE disponíveis e utilizadas pelo BDGF destaca-se, entre outras, a estrutura *Java Server Faces* (JSF). A arquitetura do *framework* JSF emprega o modelo MVC (*Model, View, Controller*), que faz a separação entre as camadas de apresentação e de aplicação. Na sua implementação como modelo MVC, o JSF possui uma camada de visualização bem distinta do conjunto de classes de modelo (Fig. 1). O JSF ainda se destaca por ser uma especificação do Java EE, isto é, todo servidor de aplicações Java tem que vir com uma implementação do *framework*.

Figura 1: Ilustração do modelo MVC (Model View Controller) adotado pelo JSF (Fonte: http://luissoares.com/jsf-parte-3-backing-beans/).



Todo sistema Web desenvolvido em Java que segue a especificação Java EE deve ser instalado e executado dentro de um servidor de aplicação. O servidor escolhido para abrigar o sistema BDGF foi o WildFly (http://wildfly.org/downloads/), versão 10, que é um dos servidores de aplicação mais seguros do mundo.

A aplicação foi desenvolvida na IDE (*Integrated Development Environment*) NetBeans (versão 8.1), que pode ser executada em diversas plataformas, como Windows, Linux, Solaris e MacOS. Essa versão do NetBeans já oferece o pacote JDK, versão 8, como linguagem Java padrão, além de variados recursos para se criar aplicativos profissionais para Web.

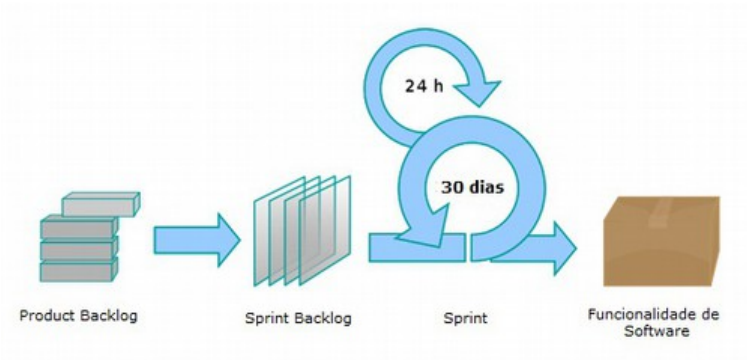
Métodos

Com base na lista de requisitos obtida para desenvolvimento do sistema de banco de dados, elaborada por meio da identificação do problema e das necessidades dos usuários, casos de uso e diagramas de sequência foram desenhados. A partir desse ponto, o projeto de desenvolvimento do sistema utilizou alguns conceitos do Scrum, que é um *framework* ágil para a realização de projetos complexos. O Scrum destaca-se dos demais métodos ágeis pela maior ênfase dada ao gerenciamento do projeto. Reúne atividades de monitoramento e *feedback*, em geral, por meio de reuniões rápidas e diárias com toda a equipe, procurando identificar e corrigir quaisquer deficiências no processo de desenvolvimento (SCHWABER, 2004).

Além disso, o método Scrum baseia-se em fundamentos como: equipes pequenas (no máximo, sete pessoas); requisitos desconhecidos e iterações curtas, dividindo o desenvolvimento em intervalos de tempos pequenos (no máximo, trinta dias), também chamados de *Sprints*. Ademais, existem três personagens importantes nesse processo: *Product Owner* (descreve os interesses de todos no projeto), *Time* (desenvolve as funcionalidades do produto) e *ScrumMaster* (responsável por garantir que todos sigam as práticas do Scrum).

O desenvovimento de um projeto Scrum inicia-se com uma visão do produto, a qual contém todas as características e restrições do produto determinadas pelo cliente (SCHWABER, 2004). Em seguida, cria-se o *Product Backlog* contendo a lista de todos os requisitos conhecidos, sendo então priorizados e divididos em tarefas, ou *sprints* (Fig. 2).

Figura 2: Representação geral do funcionamento Scrum (Fonte: http://www.devmedia.com.br/desenvolvimento-agil-com-scrum-uma-visao-geral/26343)



Dessa forma, a fase de construção compreendeu varias iterações (*sprints*), nas quais, em cada iteração, procurava-se desenvolver e corrigir pequenas partes do sistema, sendo esses testados e integrados no final, procurando satisfazer um subconjunto de requisitos do projeto. Tendo completado todas as tarefas do *backlog*, o desenvolvimento encontra-se agora em fase

de homologação pelos usuários em que, por meio de interações diretas com os usuários, o código fonte e a estrutura do sistema estão sendo ajustados e estabilizados.

Como ponto inicial do desenvolvimento do sistema, partiu-se pelo diagrama de dados inicialmente proposto por (HIGA e OLIVEIRA, 2015). Depois de diversas análises e levando-se em consideração que o SGBD utilizado seria o PosgreSQL e que a quantidade de dados a ser armazenada seria sempre crescente, optou-se por utilizar o formato JSON (*jsonb*) em campos de tabelas relacionadas a fenótipos e indivíduos (animais), e formato texto (*character varying*) em campos da tabela de genótipos do diagrama.

RESULTADOS E DISCUSSÃO

O sistema BDGF já possui muitos recursos implementados e está em processo de homologação pelos usuários. Por meio de sua interface Web é possível realizar consultas e importações de dados fenotípicos, genotípicos e de pedigree de diversas espécies de animais e, futuramente, de plantas também.

O usuário deve acessar o sistema BDGF por meio de um navegador Web tradicional. Ao acessar o endereço do sistema, a página de login será exibida (Fig. 3):

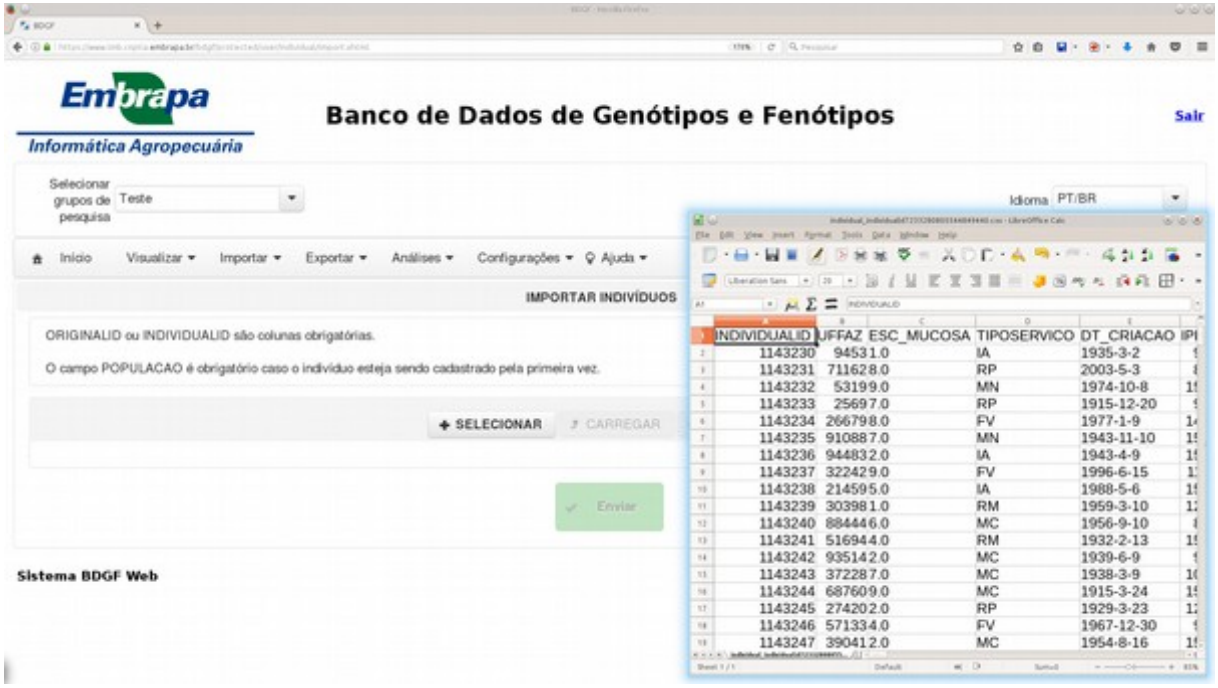
Figura 3: Tela de login do sistema BDGF.



Entre suas funcionalidades, destaca-se a função de importação de dados de animais. É possível importar dados de arquivos com colunas separadas por tabulações (TSV), que podem ser abertos e manipulados por *softwares* como o LibreOffice (Fig. 4). Esses arquivos precisam

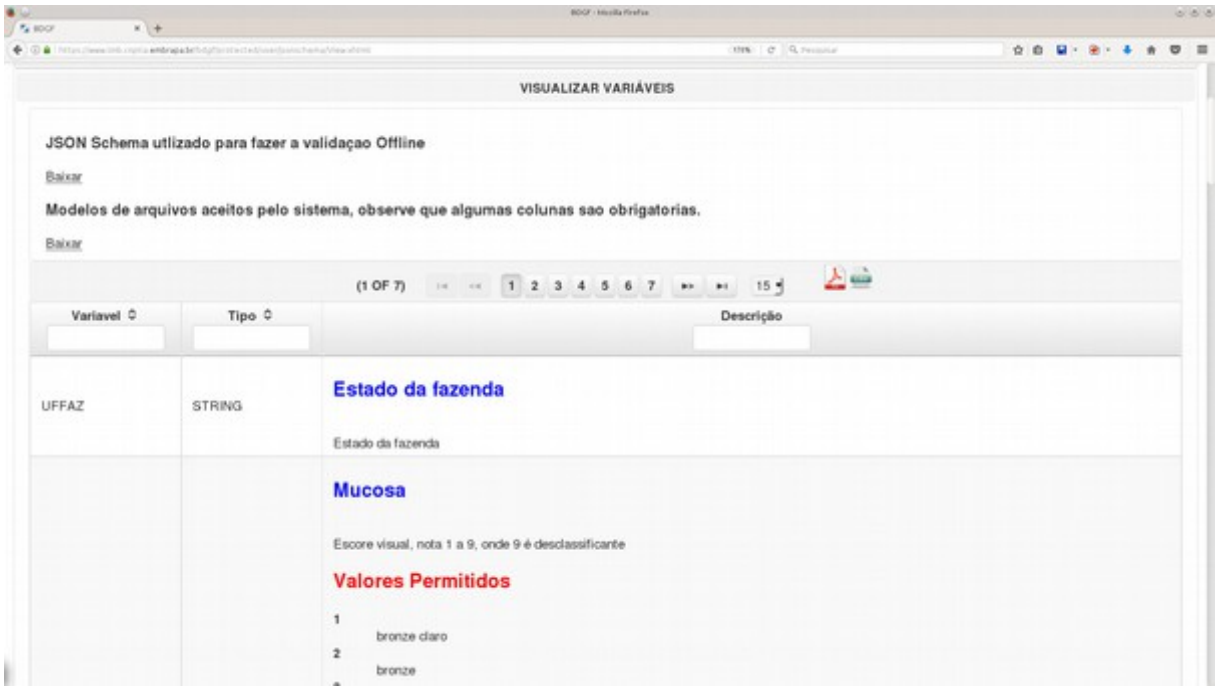
seguir um formato padronizado. O sistema fornece arquivos de exemplo para guiar o usuário a usar o formato apropriado.

Figura 4: Tela para importação de arquivo CSV com dados de animais



Contudo, cabe ressaltar que as variáveis dos fenótipos relacionados às espécies consideradas pelo sistema devem ser previamente registradas (Fig. 5). Essas variáveis devem ser importadas do Sistema de Experimentos da Embrapa - SIEXP (APOLINÁRIO *et al.*, 2016), onde foram definidas para a espécie com a qual o usuário trabalhará no seu grupo de usuários (ex: bovinos, suínos, etc.).

Figura 5: Tela para visualização das variáveis fenotípicas aceitas para um organismo importadas do SIExp.



Depois de importar os dados, é possível visualizar o pedigree de um animal listado na página de visualização de animais clicando em um ícone ao lado de cada registro (Fig. 6). A janela de pedigree pode ser expandida para facilitar a visualização dos animais e dos seus antepassados.

Figura 6: Janela de visualização de pedigree do animal selecionado.



Outra característica importante do BDGF é a capacidade de se exportar os genótipos dos animais para arquivos nos formatos MAP e PED (Fig. 7).

Figura 7: Tela para exportação de genótipos nos formatos MAP e PED.



Avaliando outros softwares com interface web desenvolvidos pela Embrapa Informática Agropecuária (VIEIRA, 2010; 2012a; 2012b) com funcionalidade de armazenamento de genótipos e fenótipos, e que contemplam algumas consultas básicas ao conjunto de dados moleculares (SNP), uma consulta simples em cerca de 800 animais e 700 mil marcadores SNP cada um demora pelo menos uma hora para ser processada. Uma consulta semelhante (utilizando a mesma máquina) feita no banco BDGF leva pouco menos de um minuto, pois a utilização de campos dos tipos JSON e texto nas tabelas de fenótipos e genótipos, respectivamente, retira parte da normalização necessária do desenho tradicional de bancos de dados, agilizando as consultas.

O sistema BDGF está sendo documentado e testado e espera-se que, em breve, esteja totalmente operacional.

CONCLUSÕES

O sistema BDGF foi projetado para apoiar programas de melhoramento animal da Embrapa, mas pode ser facilmente ajustado para armazenar dados de diversas fontes, como dados de plantas. Ele permite o armazenamento e acesso rápido a dados de pedigree, fenótipos e genótipos, utilizados para realização de avaliações genéticas. Além disso, o sistema

implementa políticas de acesso e segurança para fenótipos, genótipos e pedigree dos animais. Como trabalhos futuros, pretende-se integrar o BDGF com sistemas de avaliação genética, bem como construir Web Services para que outros sistemas possam obter informações da base de dados BDGF.

REFERÊNCIAS

APOLINÁRIO, D. R. de F.; QUEIROS, L. R.; VACARI, I.; CRUZ, S. A. B. da SIExp - Sistema de Informação de Experimentos da Embrapa. Versão v. 1.7.6. Campinas: Embrapa Informática Agropecuária, 2016.

CAETANO, A. R. Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. R. Bras. Zootec., Viçosa , v. 38, n. spe, p. 64-71, 2009.

HIGA, R. H.; OLIVEIRA, G. B. Banco de Dados de Genótipos e Fenótipos (BDGF) para suporte a estudos de associação genômica ampla e seleção genômica em programas de melhoramento animal. Campinas: Embrapa, 2015 (Série Documentos). Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/138127/1/Doc133.pdf>. Acesso em: 19 de Jul. 2017.

SCHWABER, Ken. Agile project management with Scrum, Microsoft Press. 2004.

VIEIRA, F. D. Sistema Consulta Dados de Ovinos. Versão 1.0. Campinas: Embrapa Informática Agropecuária, 2010. 1 CD-ROM.

VIEIRA, F. D. Sistema Bife de Qualidade. Versão 1.6. Campinas: Embrapa Informática Agropecuária, 2012a. 1 CD-ROM.

VIEIRA, F. D. Sistema Suínos. Versão 1.1. Campinas: Embrapa Informática Agropecuária, 2012b. 1 CD-ROM.

FERRAMENTA WEB PARA ACESSO AOS PERÍODOS DE SEMEADURA DE ARROZ DE SEQUEIRO E FEIJÃO COMUM

Marcelo Gonçalves Narciso[1], Silvando Carlos da Silva[2], Fábio Fernandes Noleto[3], Márcia Gonzaga de Castro Oliveira[4], Luciene Fróes Camarano de Oliveira[5], Elcio Perpétuo Guimarães[6]

[1] Pesquisador Doutor Embrapa Arroz e Feijão, Santo Antônio de Goiás, Goiás, Brasil, marcelo.narciso@embrapa.br

[2] Pesquisador Mestre Embrapa Arroz e Feijão, Santo Antônio de Goiás, Goiás, Brasil, silvando.silva@embrapa.br

[3] Analista de Comunicação, Especialista, Embrapa Arroz e Feijão, Santo Antônio de Goiás, Goiás, Brasil fabio.noleto@embrapa.br

[4] Analista de Transferência de Tecnologia, Mestre, Embrapa Arroz e Feijão, Santo Antônio de Goiás, Goiás, Brasil marcia.gonzaga@embrapa.br

[5] Analista de Transferência de Tecnologia, Mestre, Embrapa Arroz e Feijão, Santo Antônio de Goiás, Goiás, Brasil luciene.camarano@embrapa.br

[6] Pesquisador Doutor Embrapa Arroz e Feijão, Santo Antônio de Goiás, Goiás, Brasil, elcio.guimaraes@embrapa.br

RESUMO

É sabido que existem aplicativos que auxiliam os usuários no monitoramento remoto das máquinas agrícolas, outros fazem análises sobre nutrição dos animais, outros têm foco na fertilização, geolocalização, sementes, enfim, tudo para facilitar e aperfeiçoar o setor agrícola. Nessa mesma linha, o Ministério da Agricultura, Pecuária e Abastecimento (MAPA) possui uma série de informações sobre zoneamento de riscos climáticos, resultando em definição dos períodos de semeadura mais apropriados ao cultivo das principais espécies de importância econômica no Brasil. Com a utilização destas informações, foi possível reduzir os riscos

climáticos para as culturas, o retorno de capitais aplicados em operações de crédito agrícola, a redução das taxas de sinistralidade e, de modo geral, a diminuição do número de indenizações pagas pelo Programa de Garantia da Atividade Agropecuária (PROAGRO) e por seguradoras privadas. O MAPA armazena e disponibiliza em seu sitio web as informações sobre zoneamento. Dada a importância dessas informações, foi desenvolvida uma ferramenta web, que é o foco deste trabalho, que fornece, de maneira rápida e objetiva, os períodos de semeadura com menor risco climático para as culturas de arroz de terras altas e feijão comum no País, considerando tipo de solo e as cultivares de acordo com o grupo de ciclo a que pertencem.

PALAVRAS-CHAVE: Clima, Solo, Software, Cultivares, Plantio.

ABSTRACT

It is known that there are applications that assist users in remote monitoring of agricultural machinery, others do reviews on animal nutrition, others focus on fertilization, geo-location, seeds, everything to facilitate and perfect the agricultural sector. Along the same lines, the Ministries of Agriculture, Livestock and Food Supply (MAPA) has a series of information on climatic risk zoning, resulting in the definition of the most appropriate sowing periods for the cultivation of the main species of economic importance in Brazil. Using this information, it was possible to reduce the climatic risks to crops, the return of capital invested in agricultural credit operations, the reduction of claims rates and, in general, the reduction in the number of indemnities paid by the Guarantee Program Of the Agricultural Activity (PROAGRO) and by private insurers. MAPA stores and makes zoning information available on its website. Given the importance of this information, a web tool was developed, which is the focus of this work, which provides, in a fast and objective way, the sowing periods with lower climatic risk for upland rice and common bean crops in the country, Considering soil type and the cultivars according to the cycle group to which they belong.

KEYWORDS: Climate, Soil, Software, Cultivars, Planting

INTRODUÇÃO

O crescimento ocorrido na década de 2000 nas áreas de telefonia, redes locais sem fio e serviços via satélite permitem que informações e recursos possam ser acessados e utilizados em qualquer lugar e em qualquer momento (MONTEIRO, 2006). Nesse sentido, aplicativos móveis, inclusive aqueles voltados para o agronegócio, são uma tendência mundial demandada pela sociedade em geral. Os dispositivos como smartphones, tablets, câmeras

digitais, relógios, etc. estão se tornando padrão de acesso a sites para fazer compras, pagar contas, entretenimento e informação. Neste cenário, a Embrapa deverá estar preparada para o desenvolvimento, publicação e distribuição de aplicativos móveis, nas diversas plataformas existentes, para ampliar a oferta de produtos e tecnologias para produtores, pesquisadores e a sociedade em geral. A quantidade de tecnologias embarcadas em um smartphone como câmeras, GPS, acelerômetro são um potencial enorme na criação de aplicativos para reconhecimento de imagens, georeferenciamento, reconhecimento de padrões, leitura de códigos, entre outras.

Embora os dispositivos da computação móvel estejam cada vez mais presentes, dentro de sistemas produtivos agrícolas, esse tema ainda não possui quantidade de publicações dentro dos meios de divulgação científica que reflitam este avanço tecnológico. No entanto, já existem no mercado diversos aplicativos que vem sendo utilizados pelos agricultores brasileiros.

Estudos sobre a variabilidade climática tem sido tema de destaque no cenário agrícola brasileiro. Esse tema sempre foi a principal fonte de risco à atividade agrícola. A elevada taxa de sinistralidade na agricultura brasileira, na maioria devido a causas de adversidade climática, atesta esse fato. Então, foi nesse ambiente que o MAPA adotou o zoneamento de risco climático (definição de períodos de semeaduras) como política agrícola e solicitou a Embrapa que colaborasse no desenvolvimento de estudos que apoiassem tecnicamente essa decisão. Assim, os estudos foram realizados integrando modelos de simulação de crescimento e de desenvolvimento de culturas, bases de dados de clima e de solo, e ferramentas de geoprocessamento.

Esse trabalho permitiu a geração de uma série de produtos sobre zoneamento de riscos climáticos, resultando em definição dos períodos de semeadura mais apropriados ao cultivo das principais espécies de importância econômica no Brasil, assim surgiram as “Portarias Segmentadas por Unidade da Federação (UF)”, disponíveis no site do MAPA. Com a utilização desta técnica foi possível reduzir os riscos climáticos para as culturas, o retorno de capitais aplicados em operações de crédito agrícola, a redução das taxas de sinistralidade e, de modo geral, a diminuição do número de indenizações pagas pelo PROAGRO e por seguradoras privadas.

Na prática, o acesso às informações constantes nas portarias do MAPA se mostrou bastante complexo, pois as tabelas disponíveis são grandes e contém muitas informações, o que as torna de difícil compreensão. Isto tem gerado uma demanda muito grande pelo Serviço

de Atendimento ao Cidadão (SAC) da Embrapa, por agricultores, técnicos e instituições financeiras, que não conseguem extrair a informação de que necessitam das tabelas disponíveis nas portarias.

Pensando nisto e pela importância das informações para os usuários, a Embrapa Arroz e Feijão desenvolveu uma ferramenta web de acesso simplificado e de interface amigável (fácil navegabilidade e legibilidade) para consultas na página principal da unidade Arroz e Feijão, em “Produtos, Processos e Serviços”, com o nome “Infocult”.

Esta ferramenta de acesso facilitado fornece, de maneira rápida e objetiva, os períodos de semeadura com menor risco climático para as culturas de arroz de terras altas, conhecido também como arroz de sequeiro (o arroz cultivado sem irrigação, sujeito exclusivamente ao regime de chuvas) e feijão comum no País, considerando tipo de solo e as cultivares de acordo com o grupo de ciclo a que pertencem. Com efeito, há maior probabilidade de redução das perdas das lavouras por veranico, frio, granizo, geada e outras intempéries, com possibilidade de menor demanda por indenizações do seguro agrícola oferecido pelo Governo Federal, bem como das seguradoras privadas. Entretanto, é de suma importância citar que, na geração dessa tecnologia, não estão considerados os manejos de cultivos.

O objetivo deste trabalho é facilitar o acesso pelos usuários às informações do zoneamento de risco climático e cultivares por meio do desenvolvimento de um aplicativo computacional, cujo nome é Infocult, considerando o arroz de terras altas (sequeiro) e o feijoeiro comum. -

MATERIAL E MÉTODOS

O sistema web foi construído usando-se HTML5, em conjunto com as linguagens javascript (SILVA, 2017) e PHP (PHP, 2017) para manipulação de dados nas máquinas cliente e server, respectivamente, e uma base de dados foi construída usando-se o sistema gerenciador de banco de dados Mysql (MILANI, 2017).

Uma interface gráfica intuitiva foi elaborada com o objetivo de facilitar a navegabilidade dentro do sistema e também facilitar a usabilidade, permitindo que o usuário tenha acesso à informação, executando com sucesso as tarefas desejadas. Para isto, foi usada a teoria descrita em (NIELSEN, 2000) e (Rocha & Baranauskas, 2003). A ferramenta usada para a geração das páginas em HTML estático foi o Adobe Dreamweaver e para tratamento ou geração de imagens que foram usadas nas páginas do Infocult foi o Adobe Photoshop.

A validação do sistema realizada foi feita por parceiros das cadeias produtivas dos produtos foco desse trabalho. Foram enviadas solicitações de análises para 12 grupos diferentes de técnicos (Emater-PR, Semear Agrícola/Sementes Basso, Cabeça Branca Sementes, Ruraltins-TO, GTEC Feijão, Empaer-MT, Emdagro–SE, Emater-AL, Codevasf–AL, Emater–GO, UEMA–MA, Coopervid–MA), indicados pelos Setores de Agricultura Familiar e de Negócios e Mercados da Embrapa Arroz e Feijão. Estes parceiros puderam compartilhar informações da ferramenta com seu grupo de trabalho. Eles receberam o link do software preliminar e avaliaram num prazo de 15 dias. Posteriormente enviaram sugestões de melhorias que foram acatadas para o aperfeiçoamento da ferramenta.

RESULTADOS E DISCUSSÃO

Ao acessar o sistema, existem três informações de entrada para serem dadas para o sistema: a cultura, o estado e o município. Na base de dados do sistema já existem todos os estados e os municípios de cada estado. Quando da escolha da cultura, aparecem três opções, para a escolha de uma delas, as quais são: Arroz de terras altas (sequeiro), Feijão comum, primeira e segunda safra. Assim, ao se escolher uma das três opções acima, em conjunto com um estado e um município deste, é mostrado, como resultado, o período de semeadura e outras informações, tais como as ilustradas nas Figuras 1 e 2.

Figura 1 – Formulário para escolha do período de semeadura de uma dada cultura, em um dado município.

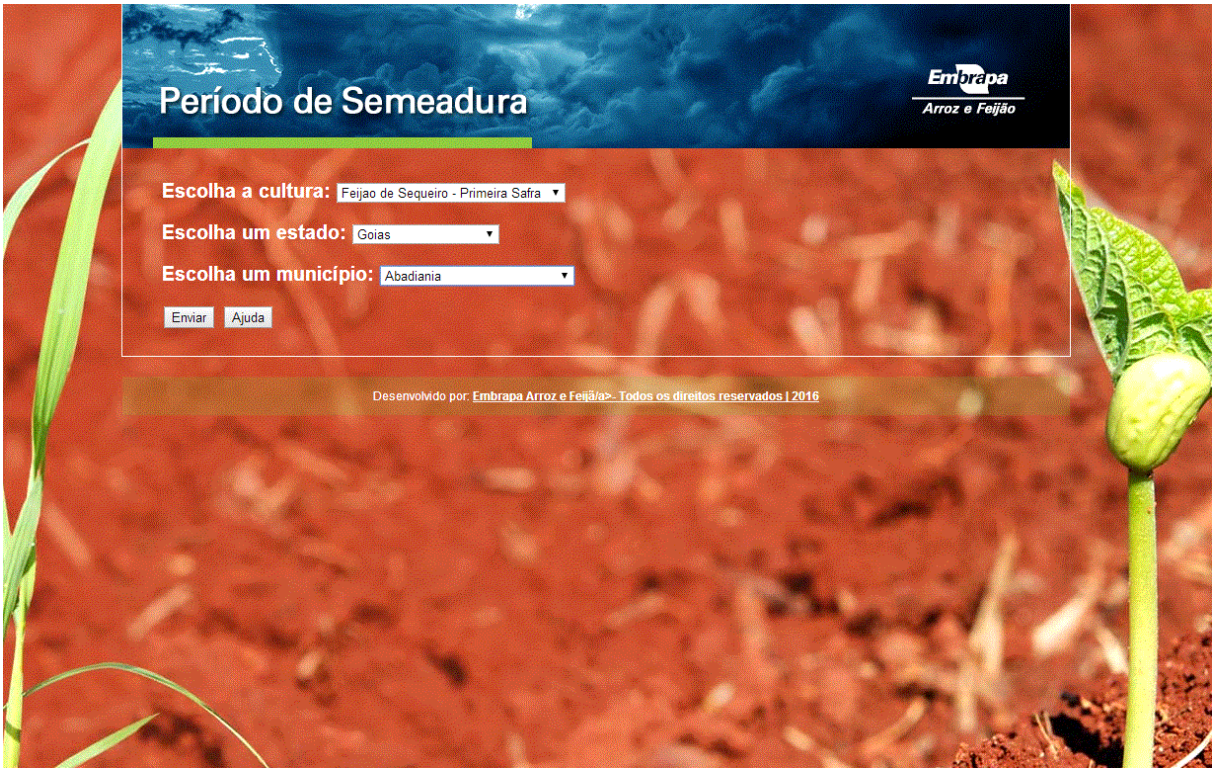
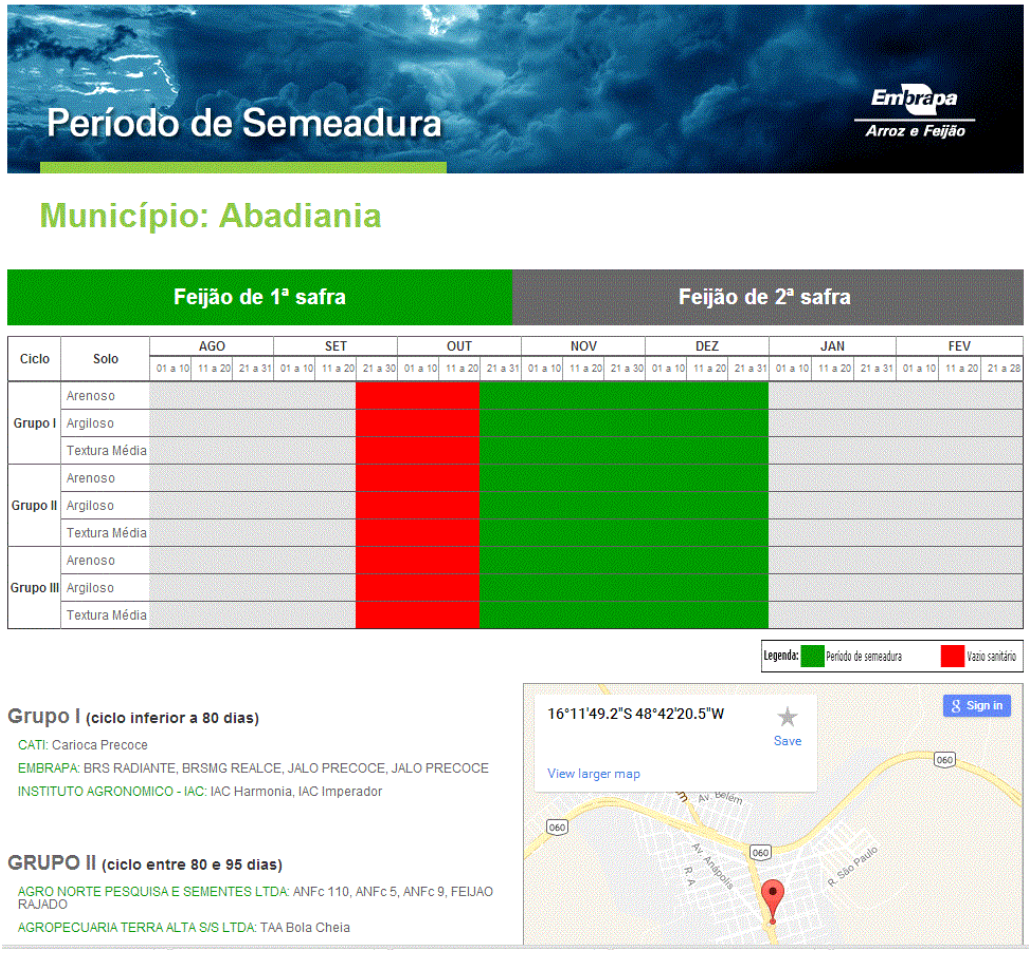
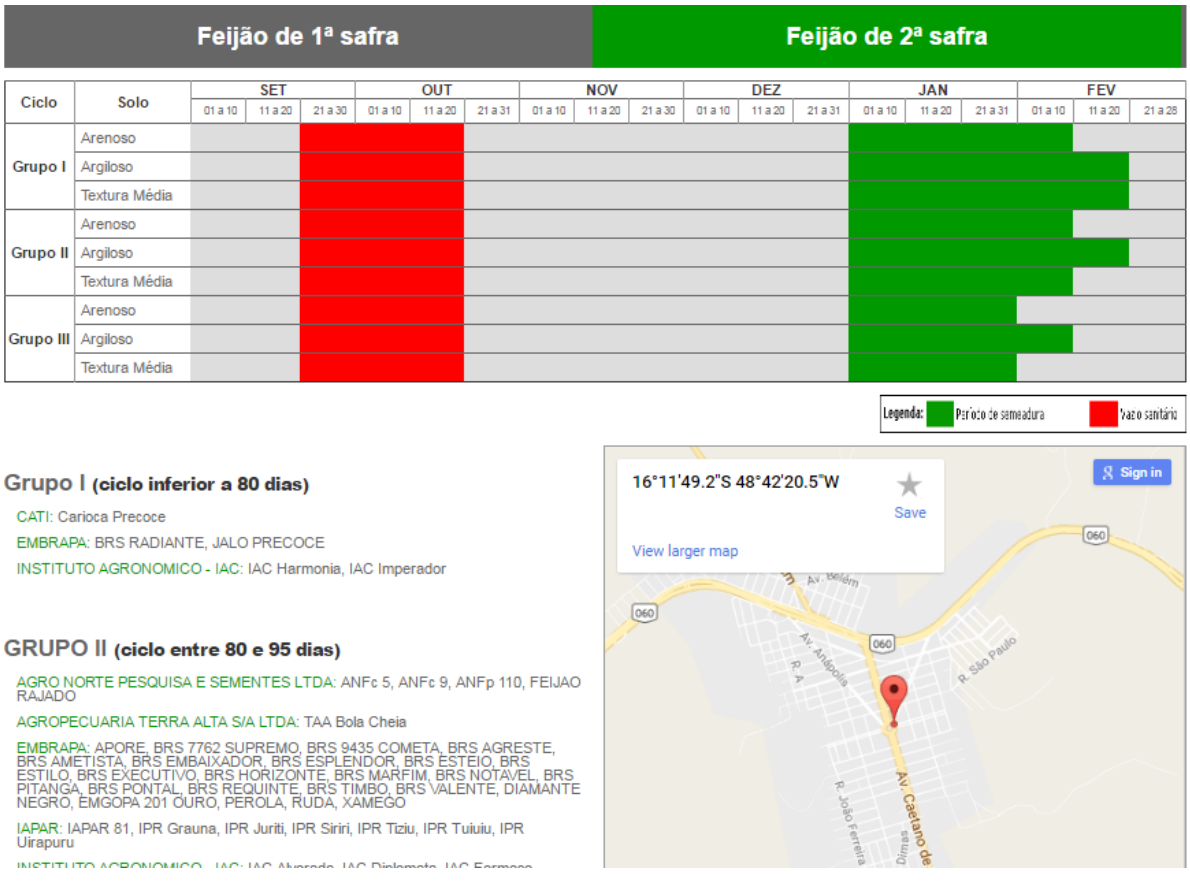


Figura 2 – Resultados para Abadiânia-GO (feijão, primeira safra)



Na Figura 2 observa-se que existe um período, em vermelho, denominado “vazio sanitário”, no qual não se pode plantar o feijão comum, para evitar ao máximo a infestação por adulto da mosca-branca, inseto transmissor do vírus do Mosaico Dourado do Feijoeiro, cujos danos impedem o desenvolvimento da planta de feijão comum. O vazio sanitário é uma medida legislativa de controle de pragas, no caso, mosca-branca, que visa impedir a “ponte verde”, que se cria quando não há impedimento de semeadura da cultura, no caso feijão-comum numa determinada época do ano. Em verde está a época de plantio para a primeira safra. Para o caso da segunda safra, basta ao usuário clicar na aba “Feijão de segunda Safra” para visualizar a época de plantio recomendada, conforme ilustra a Figura 3.

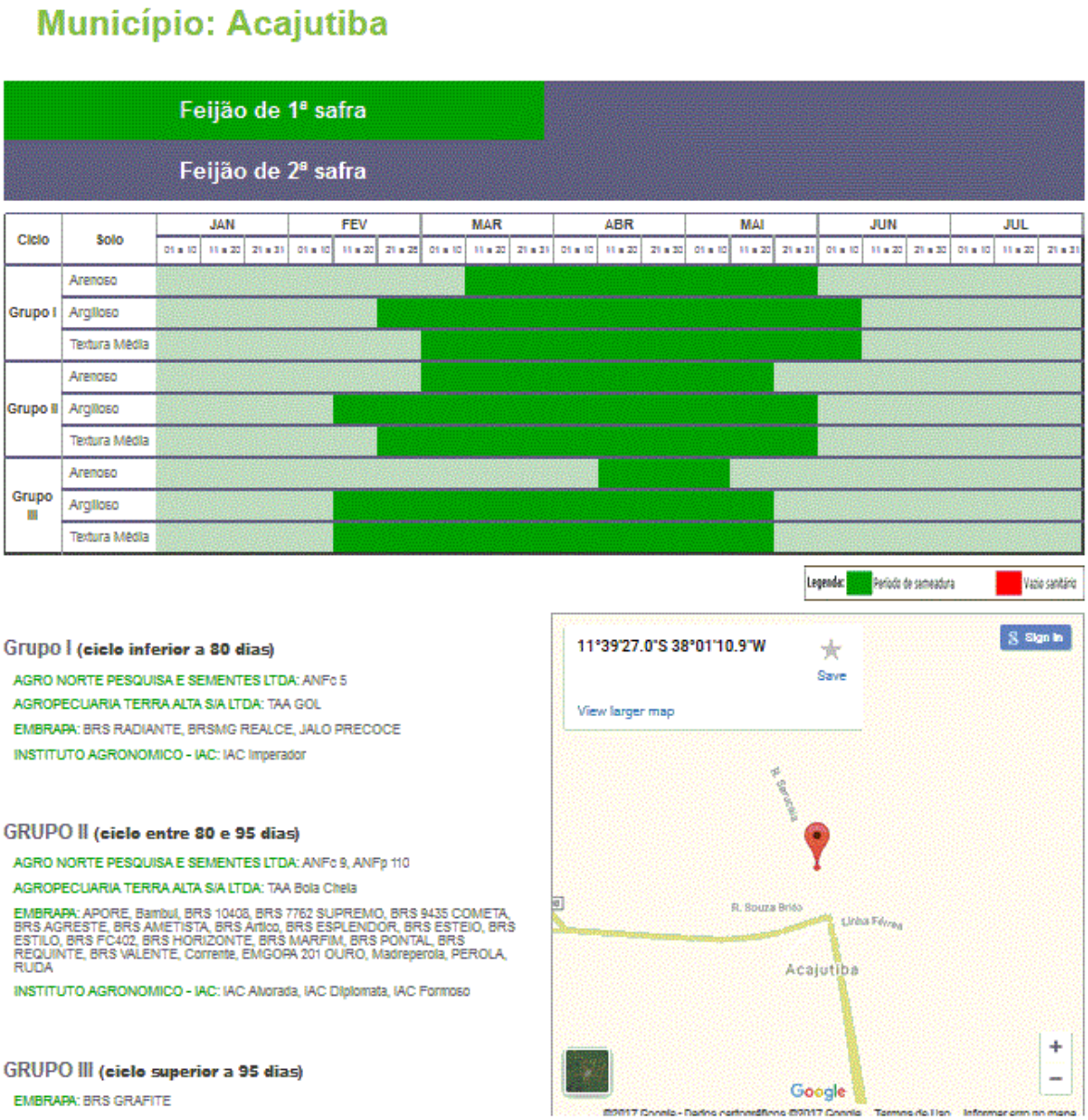
Figura 3 – Resultados para Abadiânia-GO (feijão, segunda safra)



Na Figura 3 acima, observa-se que a época de plantio muda conforme o solo e também conforme o Grupo. Cada Grupo contém as cultivares que o produtor poderá escolher para o plantio. Assim, o produtor poderá escolher a cultivar que melhor lhe convier e plantar as mesmas nas épocas mencionadas pelo sistema.

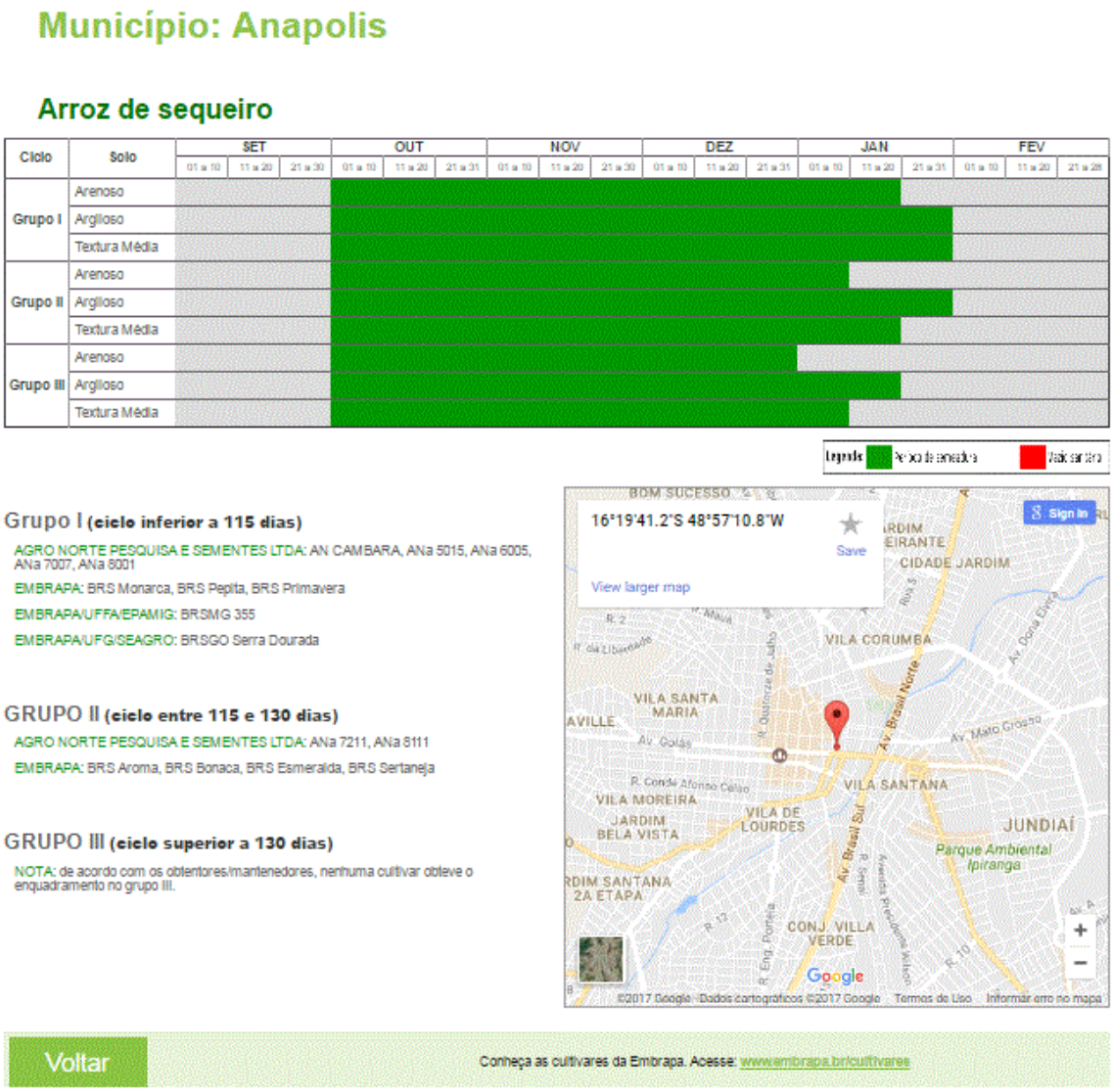
Os estados nos quais existem o “vazio sanitário” são Goiás e Minas Gerais. Os demais estados não tem este período. A maioria dos estados possui a época de plantio de feijão para primeira e segunda safra. Para o caso da Bahia, existe apenas a primeira safra para feijão, como pode ser visto na Figura 4, com o exemplo para a cidade de Acajutiba, ilustrado a seguir.

Figura 4 – Época de plantio de feijão para o Estado da Bahia



Para o caso do arroz de sequeiro (terras altas), o período de plantio varia de setembro a fevereiro do próximo ano, conforme a região e o tipo de solo, conforme ilustrado para o município de Anápolis, na Figura 5.

Figura 5 – Época de plantio de arroz de terras altas para a cidade de Anápolis



CONCLUSÕES

O software INFOCULT foi desenvolvido para facilitar o acesso dos usuários as informações de plantio do arroz de Terras Altas (sequeiro) e do feijoeiro comum, fortalecendo as cadeias produtivas dos respectivos cultivos. Este sistema encontra-se hospedado na página da Embrapa Arroz e Feijão, na aba "Produtos, Processos e Serviços". Está disponível ao público em geral e pode ser acessada no link: www.cnpaf.embrapa.br/semeadura/.

AGRADECIMENTOS

Aos técnicos das instituições que participaram da validação, pela avaliação da ferramenta e pelas sugestões para melhoria: Emater-PR, Semear Agrícola/Sementes Basso, Cabeça Branca

Sementes, Ruraltins-TO, GTEC Feijão, Empaer-MT, Emdagro–SE, Emater-AL, Codevasf–AL, Emater–GO, Agrodefesa-GO, UEMA–MA, Coopervid–MA,

REFERÊNCIAS

MILANI, A. Mysql – **Guia do Programador**. Editora Novatec, São Paulo, 2017.

MONTEIRO, J.D.A. **Desenvolvimento de aplicações multi-plataformas para dispositivos móveis**. 2006. 148 p. Dissertação (Mestrado em Ciências Matemáticas e de Computação) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2006.

NIELSEN, J. Projetando Websites: designing web usability. Tradução de Ana Gibson. Rio de Janeiro: Elsevier, 2000a.

PHP. Sítio disponível em www.php.net. Sítio visitado em 14/05/2017.

ROCHA, H.V.; BARANAUSKAS, M.C. Design e Avaliação de Interfaces Humano-Computador. Editora Unicamp, Campinas-SP, 2003.

SILVA, M. S. JavaScript – Guia do Programador. Editora Novatec, São Paulo, 2017.

Realidade Aumentada como Ferramenta de Apoio a Estudos Pedológicos

Allan Silva Salles [1], Hugo Portocarrero [2], Aluísio Granato de Andrade [3], Rulian Rodrigues [4]

[1]Graduando do curso de Geografia, UERJ, Rio de Janeiro, RJ, allansalles96@gmail.com

[2]DSC. Engenharia Civil, UERJ, Rio de Janeiro, RJ, hportocarrero@gmail.com

[3]PHD. Engenharia Agronômica, Embrapa Solos, Rio de Janeiro, RJ, aluisio.andrade@embrapa.br

[4]Graduando do curso de Geografia, UERJ, Rio de Janeiro, RJ, rulianrodrigues62@gmail.com

RESUMO

Neste trabalho a realidade aumentada é utilizada como uma ferramenta de apoio a estudos pedológicos, mediante o cadastro de informações georreferenciadas formando um banco de dados acessível por dispositivos móveis. O acervo bibliográfico da Embrapa Solos serviu de subsídio para a criação do Cenário Solos do Brasil e a catalogação de diversos perfis de solos da região Metropolitana do Rio de Janeiro no aplicativo móvel de realidade aumentada *Paprikamix*, desenvolvido pela empresa *Eyllo*. O trabalho foi realizado no contexto da Oficina de Aplicativos Geoespaciais aplicados ao Ensino e a Pesquisa, um projeto de extensão da Universidade do Estado do Rio de Janeiro, e, posteriormente, inserido como atividade ao Projeto Embrapa de Pesquisa e Inovação para Aprimoramento da Taxonomia de Solos Brasileiros. Trinta e dois pontos foram cadastrados visando a sua utilização por profissionais das ciências do solo, bem como em trabalhos de campo das disciplinas Pedologia e Geomorfologia na Universidade do Estado do Rio de Janeiro. O resultado deste trabalho possibilitará o acesso a diversas informações sobre classificação e análises físicas e químicas do solo, que se encontram dispersas na literatura, de modo a ampliar o conhecimento agrícola aprimorando a tomada de decisões no campo.

PALAVRAS-CHAVE: Realidade Aumentada, Dispositivos Móveis, Pedologia

ABSTRACT

In this work, augmented reality is used as a tool to support pedological studies, through the registration of georeferenced information forming a database accessible by mobile devices.

The bibliographic collection of Embrapa Solos served as a subsidy for the creation of the Brazilian Soils Scenario and the cataloging of several soils profiles of the Rio de Janeiro Metropolitan Region in the mobile augmented reality application Paprikamix developed by Eyllo. The work was carried out in the context of the Workshop on Geospatial Applications applied to Teaching and Research, an extension project of the State University of Rio de Janeiro, and later inserted as an activity to the Embrapa Research and Innovation Project to Improve Taxonomy of Brazilian Solos. Thirty-two points were registered for use by soil science professionals, as well as in field studies of Pedology and Geomorphology at the State University of Rio de Janeiro. The result of this work will allow the access to diverse information on classification and physical and chemical soil analysis, which are dispersed in the literature, in order to increase the agricultural knowledge improving the decision making in the field.
KEYWORDS: Augmented Reality, Mobile Devices, Pedology

INTRODUÇÃO

Desde a difusão da internet móvel, gadgets têm se tornado cada dia mais acessíveis e funcionais graças aos avanços da informática e das ciências tecnológicas. Atualmente, o Brasil possui aproximadamente quatro smartphones para cada cinco habitantes: são cerca de 168 milhões de aparelhos em uso no país segundo pesquisa realizada pela FGV-SP (PORTAL LINK, ESTADÃO, 2016). Dentre as inovações tecnológicas mais recentes merece destaque a Realidade Aumentada. Segundo Azuma (1997), a Realidade Aumentada é uma tecnologia que permite ao usuário ver o mundo real e o mundo virtual simultaneamente, com objetos sobrepostos ou compostos com o mundo real. Assim, *gadgets* atuam como agentes intermediários para o entrelaçamento de informações virtuais e dados georreferenciados reais.

Tratando sobre os aspectos cognitivos da Realidade Aumentada, Engberg e Bolter (2014) a caracterizam como Poliestética, uma vez que opera com múltiplos sentidos humanos. Assim, esta tecnologia possibilita ocupar dois lugares ao mesmo tempo, mantendo o senso de localização física e cultural do usuário e disponibilizando-lhe o acesso ao mundo virtual da internet. Ao proporcionar esta associação, a Realidade Aumentada tem o potencial de alterar a relação da sociedade com a informação. Através dela, o usuário pode perceber a informação como parte agregada ao seu ambiente visual, e não apenas de maneira estática e muitas vezes desassociada de seu contexto original. Desta maneira, ao combinar a informação simbólica (digital) e o mundo físico, a tecnologia oferece ao usuário uma experiência de percepção diferenciada (BOLTER *et al.*, 2013).

Neste sentido, o presente trabalho foi realizado dentro do contexto da Oficina de

Aplicativos Geoespaciais aplicados ao Ensino e a Pesquisa da Universidade do Estado do Rio de Janeiro. Na oficina são realizadas atividades de treinamento, aprendizado e desenvolvimento tecnológico na área de aplicativos geoespaciais voltados para plataformas móveis. Em 2016 foi formalizada uma atividade sob responsabilidade desta Oficina dentro do contexto do Projeto Embrapa - Pesquisa e Inovação para Aprimoramento da Taxonomia de Solos Brasileiros, que visa induzir e realizar pesquisas básicas e inovadoras em química, física e mineralogia de solos, direcionadas à revisão e atualização do Sistema Brasileiro de Classificação de Solos (SiBCS).

Neste trabalho foi desenvolvido o Cenário de Realidade Aumentada Solos do Brasil, tendo sido para tal formada uma parceria entre o Instituto de Geografia da UERJ, a empresa *Eyllo* e a Embrapa Solos. Dessa forma tornou-se possível utilizar a plataforma de Realidade Aumentada *Paprikamix* e o vasto acervo de levantamentos de solos da Embrapa Solos para o encaminhamento do projeto.

O *Paprikamix* é um aplicativo de realidade aumentada para dispositivos móveis, desenvolvido pela empresa *Eyllo*, no qual o usuário pode utilizar seu smartphone ou tablet para localizar pontos de interesse e informações detalhadas sobre esses pontos. Em sua plataforma é possível navegar por diversos cenários, cada um destes com propostas diferentes, como localizar pontos turísticos do Rio de Janeiro, áreas de comércio, pontos de ônibus e etc. O aplicativo utiliza a base de mapas do aparelho: *Google Maps* em dispositivos *Android* e Mapas em dispositivos *Apple*. Além de poder visualizar os dados virtuais sobrepostos às imagens capturadas pela câmera do dispositivo, é possível também navegar com as informações dispersas de acordo com suas localizações sobre o mapa (Figura 1). O aplicativo encontra-se disponível gratuitamente nas lojas virtuais *Play Store* e *Apple Store*, e para sua utilização é necessário que o dispositivo móvel possua acesso a GPS, câmera e giroscópio.

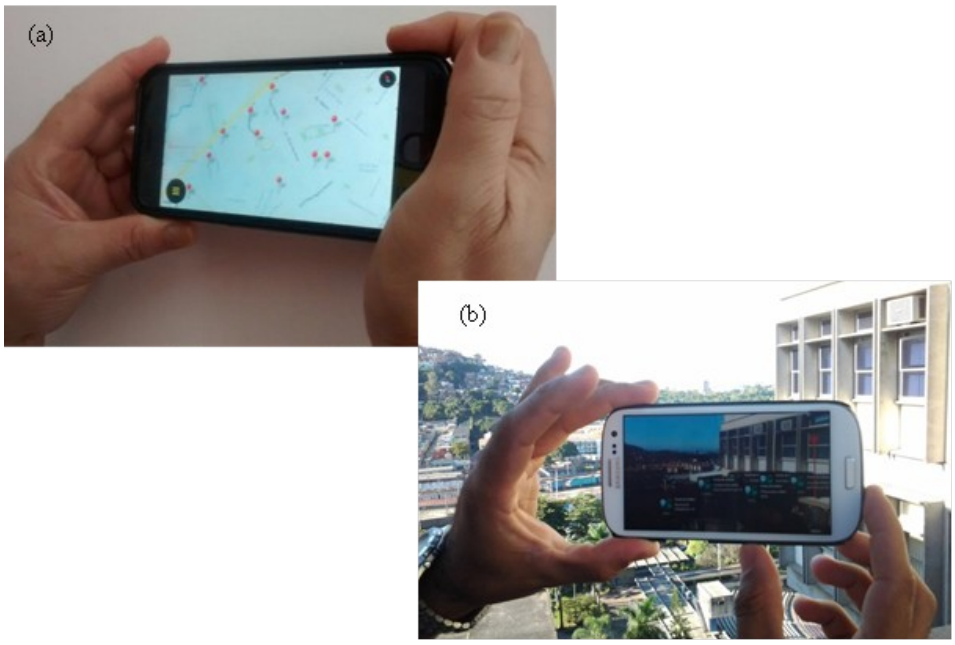


Figura 1- Realidade Aumentada: (a) informações digitais pontuadas no mapa e (b) informações sobrepostas à imagem da câmera.

Através do cenário desenvolvido, o usuário tem a sua disposição diversas informações relacionadas à classificação e análise de solos, sinalizadas em pontos georreferênciados denomiados "*geotags*". Para compor este cenário foram levantadas informações constantes no acervo de dados relacionados ao levantamento, classificação e mapeamento dos solos brasileiros da Embrapa Solos, que se encontram disponíveis em bibliotecas (EMBRAPA, 1980, 2013; LUMBRERAS, J. F. *et al*., 2001; Embrapa Solos, 2003; LUMBRERAS, J. F., 2011) e em parte na internet (https://www.bdsolos.cnptia.embrapa.br/consulta_publica.html). Através do aplicativo, foi possível apresentar não somente a caracterização da classe de solo, mas também anexar imagens, vídeos, teses e dissertações, consistindo assim em uma ótima ferramenta, ampliando o conhecimento agrícola e aprimorando a tomada de decisões no campo.

MATERIAL E MÉTODOS

Para o desenvolvimento deste trabalho foi elaborado o Cenário Solos do Brasil dentro da plataforma *Paprikamix*, no qual a equipe vem inserindo os diferentes perfis de solos catalogados pela Embrapa Solos na Região Metropolitana do Estado do Rio de Janeiro, que já conta com trinta e dois perfis referenciados.

A cada *geotag* inserida no Cenário foram associadas a descrição morfológica do solo e fotos do perfil em questão. É possível, ainda, anexar vídeos e áudios, para fazer da experiência mais didática para usuários que não trabalhem diretamente com ciências do solo,

ou ainda dissertações e teses, aprofundando o detalhamento da informação exposta de forma a possibilitar aos profissionais debaterem sobre as classificações e abrir uma discussão científica sobre os métodos de classificação, descrição, etc. O cenário também pode servir como guia de campo, havendo a projeção de criar novas *geotags* com roteiros pré-definidos, a exemplo dos trabalhos de campo feitos pelos alunos de graduação de Geografia da UERJ.

Visando o setor agrícola, é viável a disponibilização de informações pertinentes à fertilidade e aptidão agrícola ao pequeno e médio agricultor de forma simples e acessível.

Inserção de geotags

O aplicativo é atualizado por meio de um sistema *online* de uso exclusivo da equipe de desenvolvedores. Através do *login* do desenvolvedor, este é redirecionado à página de ferramentas e configurações do sistema. Seguidamente, são apresentadas, no sistema, duas opções: autoria e análises. Na primeira opção, o programador pode criar cenários de seus próprios conteúdos multimídias e, na segunda, analisar e descobrir padrões significativos nos dados já inseridos nos cenários (SANTANA, L. S. & PORTOCARRERO, H., SIMPURB 2015). A inserção de *geotags* no sistema ocorre de forma bem simples, através da ferramenta de autoria. Primeiramente, é preciso localizar a área do perfil de solo a ser cadastrado no mapa. Ao ser localizado, a criação da *geotag* se dá em um clique. Em seguida, em uma caixa de ferramentas, são colocadas as atribuições e informações necessárias àquele ponto, como fotos, vídeos, áudios, textos e, futuramente, outros tipos de mídias. Ao fechar a caixa, a *geotag* é salva automaticamente tornando-se disponível para o usuário comum. A seguir são apontadas algumas etapas da criação de *geotags* para o Cenário Solos do Brasil:

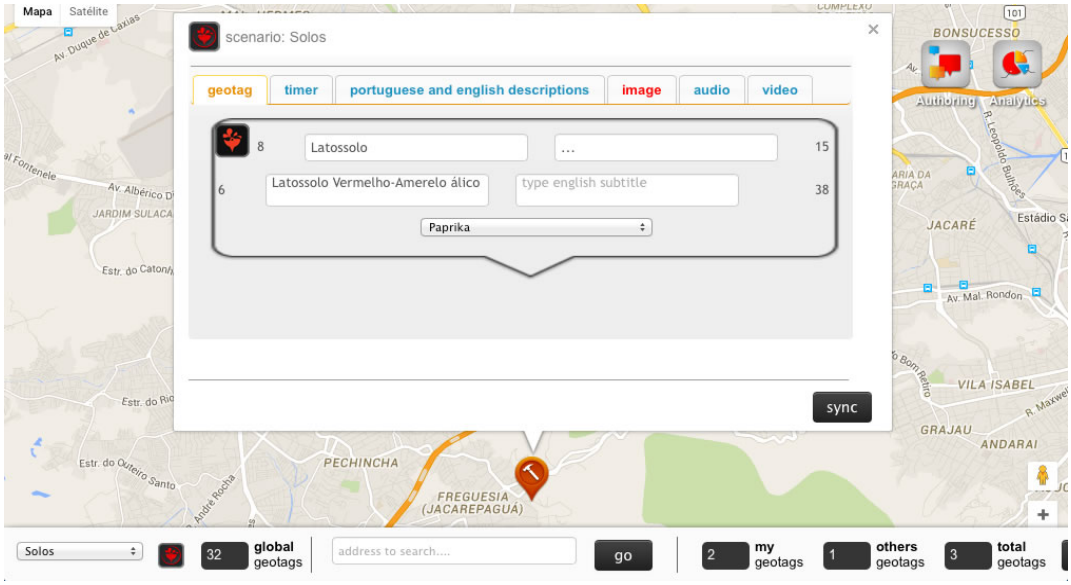


Figura 2- Na aba "*geotag*" é feito o registro da localização da *geotag* e a denominação da classificação do solo em português, segundo o Sistema Brasileiro de Classificação de Solos (EMBRAPA, 2006), e em inglês, segundo a norma universal *Soil Taxonomy* (*USDA*, 1999);

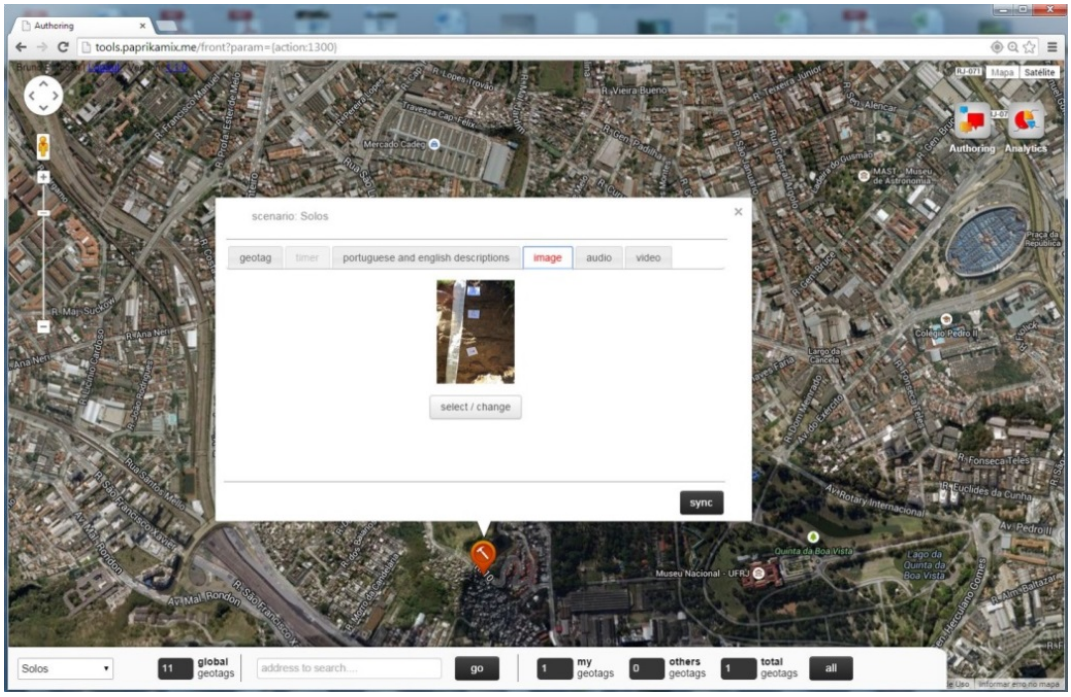


Figura 3 – Na aba "*image*", é anexada a foto do perfil do solo com descriminação dos horizontes.

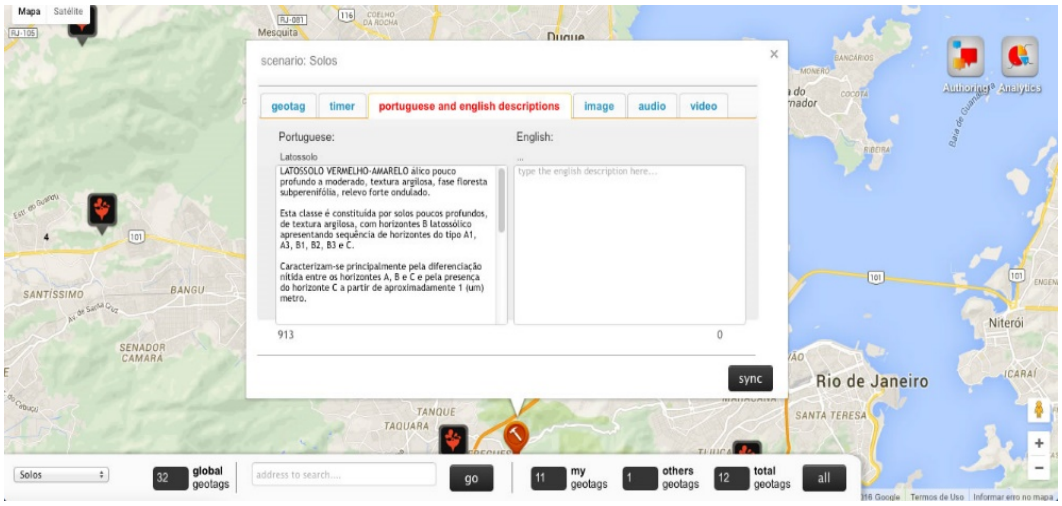


Figura 1 – Na aba "*portuguese and english descriptions*", são adicionadas descrições mais detalhadas dos horizontes do solo, como resultados de análises químicas e físicas. Também é possível inserir artigos a respeito da região em questão. Os textos podem ser exibidos em português e inglês.

Diferentes geotags

A equipe vem desenvolvendo diferentes *geotags* para diferenciar os níveis de credibilidade da determinação da classe de solo, baseado na fonte da análise. Desta maneira, profissionais das ciências do solo poderão averiguar a necessidade da realização de uma nova análise de acordo com o rigor estabelecido para sua pesquisa ou trabalho. As *geotags* a serem adicionadas ao sistema são: Perfil Embrapa, Perfil Tese, Perfil Dissertação, Informação de Interesse Geotécnico, etc.

Usuários Especiais/ Colaboradores

Com a criação de Usuários Especiais e/ou Colaboradores, destinado a profissionais da área de ciências do solo, será possível a expansão do projeto a nível nacional, uma vez que este poderá criar remotamente *geotag*s no Cenário Solos do Brasil. A *geotag* recém-criada será temporária e estará sujeita a avaliação técnica e a um prazo para que esta ocorra. Após verificação, se aprovada, a *geotag* será fixada definitivamente e sinalizada com o respectivo nível de credibilidade. Desta maneira, será possível que a plataforma alcance boa parte do território nacional, fazendo desta uma ferramenta mais completa.

RESULTADOS E DISCUSSÃO

Resultados alcançados

O principal resultado alcançado foi a criação do Cenário Solos do Brasil na plataforma e o cadastro dos pontos do levantamento dos solos do estado do Rio de Janeiro, conforme mostra a Figura 4, utilizadas em trabalhos de campo da disciplina Pedologia de cursos de graduação em Geografia, Engenharia Agronômica e áreas afins. Ademais, a reunião do material bibliográfico dos levantamentos de solos cedidos pela Embrapa foi outra fase fundamental para a realização do projeto.

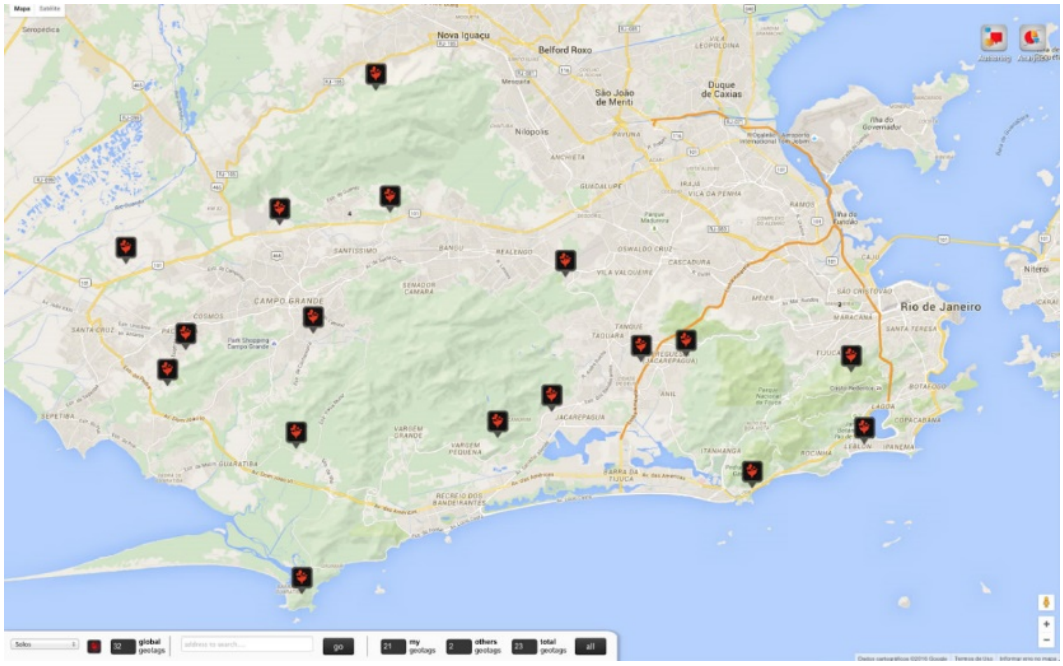


Figura 2 - *Geotag*s do Cenário Solos do Brasil na região Metropolitana do Rio de Janeiro.

Resultados almejados

O projeto encontra-se em um processo de expansão progressiva. Estão em fase de planejamento a incorporação de informações de diferentes bases cartográficas para consulta, como mapas de solos e de aptidão agrícola, a associação de usuários especiais ou colaboradores, que poderão adicionar novas *geotags* remotamente, e a criação de diferentes tipos de *geotags*, significando diferentes níveis de credibilidade de acordo com a fonte da informação.

Com a aplicação desses aprimoramentos, a equipe almeja, posteriormente, transformar o Cenário Solos do Brasil em um Aplicativo para Cadastro e Disponibilização de Informações Pedológicas, à parte da plataforma *Paprikamix*. Esta mudança trará maior autonomia e praticidade, fazendo do aplicativo uma ferramenta específica para solos, o que propiciará

suporte para os trabalhos de campo, solucionando possíveis dúvidas e garantindo que sejam seguidas as normas de qualidade adotadas pela Embrapa e Sociedade Brasileira de Ciência do Solo. Aliado a isso, deverá ser adicionado ao aplicativo uma versão digital do Manual de Descrição e Coleta de Solo no Campo (SANTOS et. al., 2013), documento oficial da SBCS que define e normatiza as características morfológicas normalmente utilizadas na descrição e coleta de perfis de solos no campo.

Através da expansão da ferramenta, a equipe tem como objetivo torná-la de acesso público para o estudo de solos, servindo de fonte de informação para pesquisadores, agrônomos, engenheiros, pedólogos, geógrafos e agricultores. Graças às inúmeras possibilidades disponíveis no aplicativo, faz-se possível reunir informações que dialoguem com todas estas ciências e com o público em geral.

CONCLUSÕES

O desenvolvimento deste trabalho resulta em facilidade e praticidade para o alcance de informações e análises referentes a solos, tanto para o meio acadêmico e técnico nas áreas de agronomia e engenharias que dão suporte à agricultura de grande porte, quanto para a agricultura de médio e pequeno porte, graças a acessibilidade de informações.

Algumas ferramentas do aplicativo *Paprikamix* ainda se encontram em fase de aprimoramento, como o raio para a obtenção de informações (atualmente 2 quilômetros partindo do posicionamento do usuário). Outro limitante, também em fase de aprimoramento, diz respeito a forma de inserção de *geotags*, que é realizada pelo endereço do mesmo e não através de coordenadas, isto é, o sistema usa como base cartográfica o *Google Maps*. Dentre as projeções de melhorias em futuro próximo estão a alternativa de inserção de localidades a partir de coordenadas geográficas, o que permitirá uma precisão maior ao referenciamento geoespacial. Mesmo com estas limitações, a plataforma mostrou grande potencial devido ao seu caráter inovador, a possibilidade de cadastro de diversas mídias e a facilidade de inserção de novas *geotags*. A praticidade e mobilidade do banco de dados gerado pelo Cenário Solos do Brasil farão do *Paprikamix* um recurso capaz de aprofundar e expandir o acesso a informações sobre o solo e suas particularidades, ampliando o conhecimento agrícola e aprimorando a tomada de decisões no campo.

REFERÊNCIAS

AZUMA, R. A. 1997. Survey of Augmented Reality. In.: *Presence: Teleoperators and Virtual Environments*. Vol. 6 (4). pp. 355-385. MIT Press.

BOLTER, J. D; ENGBERG, M.; MACINTYRE, B., 2013 – Media Studies, Mobile Augmented Reality, and Interaction Design. In.: *Interactions*, Vol 20 (1). pp. 36-45. Association for Computing Machinery.

EMBRAPA SOLOS. 2003. *Levantamento de reconhecimento de baixa intensidade dos solos do Estado do Rio de Janeiro*. Rio de Janeiro, 245 p. Boletim de pesquisa e desenvolvimento, 32.

EMBRAPA. 1980. *Serviço Nacional de Levantamento e Conservação de Solos. Levantamento semidetalhado e aptidão agrícola dos solos do município do Rio de Janeiro, RJ*. Rio de Janeiro, EMBRAPA-SNLCS, 1980. 389 p. (EMBRAPA-SNLCS. Boletim Técnico, 66).

EMBRAPA. 2013. *Sistema Brasileiro de Classificação de Solos*. 3ª Edição. Brasília. 353 p.

ENGBERG, M. & BOLTER, J. D., 2014. Cultural expression in augmented and mixed reality. In.: *Convergence: The International Journal of Research into New Media Technologies*. Vol. 20(1). pp. 3–9. SAGE.

LUMBRERAS, J. F.; CARVALHO FILHO, A.; CALDERANO FILHO, B.; SANTOS, R. D. dos. 2001 – *Levantamento pedológico, vulnerabilidade e potencialidade ao uso das terras: quadrículas de Silva Jardim e Rio das Ostras, Estado do Rio de Janeiro*. In.: Congresso Brasileiro de Ciência do Solo, 28, 2001, Londrina: Sociedade Brasileira de Ciência do Solo; Embrapa Soja; IAPAR; UEL; UEM, 2001. Rio de Janeiro, Embrapa Solos.

PORTAL LINK, ESTADÃO – *Brasil chega a 168 milhões de smartphones em uso*. Disponível em: <http://link.estadao.com.br/noticias/gadget,brasil-chega-a-168-milhoes-de-smartphones-em-uso,10000047873>. Acesso em 23 de maio de 2016.

SANTANA, L. S. & PORTOCARRERO, H. 2015. – *Potencialidades e Limitações da Plataforma Paprikamix na Análise Multiescalar dos Espaços Urbanos*. In.: XIV Simpósio Nacional de Geografia Urbana: Perspectivas e Abordagens da Geografia Urbana no Século XXI. Fortaleza.

SANTOS, R.D.; LEMOS, R. C.; SANTOS, H. G.; KER, J. C.; DOS ANJOS, L. H. C., 2013 – *Manual de Descrição e Coleta de Solo no Campo*. Sociedade Brasileira de Ciência de Solo, 6º ed. revista e ampliada. Viçosa, SBCS/EMBRAPA-CNPS.

UNITED STATES DEPARTMENT OF AGRICULTURE (USDA). 1999 – *Soil Taxonomy: A Basic System of Soil Classification for Making and Interpreting Soil Surveys*. 2ª Ed. Washington. U.S. Government.

Dados da agricultura brasileira: apoio ao planejamento estratégico

Edson Bolfe[1], *Marcos A G Pena Jr.*[1], *Elisio Contini*[1], *Fabiano Mariath*[1], *Giani Silva*[1]

[1]Embrapa, Distrito Federal, Brasil, edson.bolfe@embrapa.br, marcos.pena@embrapa.br, elisio.contini@embrapa.br, fabiano.mariath@embrapa.br, giani.tavares@embrapa.br

RESUMO

Nas últimas décadas, o Brasil se tornou um dos líderes globais na produção agrícola. Fatores que contribuem para tais resultados incluem disponibilidade de recursos naturais, condições climáticas favoráveis, pesquisa agrícola, o empreendedorismo da sua agroindústria e dos seus agricultores e políticas governamentais para o setor agrícola. Atualmente, várias organizações públicas fornecem dados e informações alfanuméricas sobre a agricultura brasileira. Crítica para entender o conteúdo da informação, visualização de dados é uma ferramenta indispensável para examinar, analisar e validar a análise, bem como dar suporte às pessoas na tomada de decisões. Visualização de dados é a ação, entregue muitas vezes por ferramentas digitais, de esclarecer as informações dentro de uma coleção de dados através de diferentes formas gráficas. O Sistema de Inteligência Estratégica (Agropensa) da Empresa Brasileira de Pesquisa Agropecuária (Embrapa) tem como um dos seus objetivos apoiar o público em geral a construir sua própria visão estratégica sobre o setor agropecuário, dentre outras ferramentas, se utilizando para tal de *business intelligence* (BI) e painéis baseados nos dados do Instituto Brasileiro de Geografia e Estatística (IBGE). A Embrapa incorporou o uso do BI e a da visualização de dados para a agricultura, possibilitando ganhos analíticos em suas ações de inteligência estratégica (IE); permitindo um planejamento mais eficaz em pesquisa, desenvolvimento e inovação (PD&I).

PALAVRAS-CHAVE: *Business intelligence* (BI), Visualização de dados, Inteligência Estratégica.

ABSTRACT

Over the last decades, Brazil has become one of the global leaders in agriculture production. Factors contributing to such results include natural resources availability, favorable climate conditions, agricultural research, farmers and agroindustry entrepreneurship and governmental policies towards the agricultural sector. Currently, several public organizations provide data and alphanumeric information on Brazilian agriculture. Critical to understand the content of information, data visualization is an indispensable tool to examine, scrutinize, and validate analysis and upholds people making decisions. Data visualization is the action, delivered often by digital tools, of clarifying the information within a collection of data through different graphic shapes. The Strategic Intelligence System (Agropensa) of Brazilian Agricultural Research Corporation (Embrapa) aims at supporting the general public to construct their own strategic view on the agricultural sector by means of developed business intelligence (BI) and dashboards based on Brazilian Institute of Geography and Statistics (IBGE) data. Embrapa incorporated the use of BI and data visualization for agriculture, enabling analytical gains in its strategic intelligence (IE) actions; enabling more effective planning in research, development and innovation (PD&I).

KEYWORDS: Business intelligence (BI), Data visualization, Strategic intelligence.

INTRODUÇÃO

O Brasil se tornou um dos líderes globais na produção agrícola. Dentre os principais fatores que contribuem para esses resultados destacam-se a disponibilidade de recursos naturais, as condições climáticas favoráveis, a pesquisa agrícola, o empreendedorismo e as políticas governamentais para o setor agrícola.

Várias organizações públicas geram dados e informações sobre a agricultura brasileira. Um aspecto crítico para a melhor compreensão das informações geradas é a forma de visualização de dados. Este processo torna-se indispensável para examinar, analisar e validar estudos e apoia o planejamento estratégico na área agrícola. Cada tomada de decisão exige análises de diferentes formatos e ferramentas de visualização apoiam estas operações (MYATT; JOHNSON, 2009). Visualização de dados é a ação organizada por ferramentas digitais, visa gerar informações a partir de banco de dados de diferentes formas gráficas. Dados sobre o setor agrícola no Brasil, entre outros, são coletados, organizados e fornecidos pelo Instituto Brasileiro de Geografia e Estatística (IBGE, 2017). O IBGE fornece dados agrícolas através do Sistema de Recuperação Automática (SIDRA), mas a base de dados da Pesquisa Agropecuária Municipal (PAM, 2015) é apenas uma interface para obter os dados do

IBGE. Esta ferramenta não permite que as pessoas interajam com os dados e construam diferentes formas de visualização.

O Sistema de Inteligência Estratégica da Embrapa (AGROPENSA, 2017 a) atua: (i) na captura e na prospecção de tendências, para a identificação de futuros possíveis; e, (ii) no mapeamento e no apoio à organização, à integração e à disseminação de base de dados e de informações agrícolas. Esses dois grandes focos têm como principal objetivo a elaboração de estudos de futuro, os quais contenham e difundam conhecimentos e informações em apoio à formulação de estratégias de Pesquisa, Desenvolvimento e Inovação (PD&I) para a própria Embrapa e para organizações parceiras.

Nesse sentido, este trabalho objetiva apresentar ações do Agropensa vinculadas à coleta, à organização e à disponibilização de dados da agricultura brasileira por meio de produtos de *Business Intelligence* (BI) como *dashboards* (painel de controle ou interativo) para apoiar a tomada de decisão e o planejamento estratégico do setor rural.

MATERIAL E MÉTODOS

O painel de controle de *business intelligence* é um conjunto de visões analíticas de dados, de diversas fontes, transformados, integrados e consolidados a fim de fornecer informações imediatas e relevantes por meio de indicadores apresentados em relatórios e *dashboards* com gráficos e tabelas, além dos demais componentes de análises com um elevado nível de interação. “Um painel é uma exibição visual das informações mais importantes necessárias para alcançar um ou mais objetivos; consolidado e organizado em uma única tela para que as informações possam ser monitoradas de relance” (PEW; STEPHEN, 2006, p. 26).

A ferramenta de BI utilizada na Embrapa para a construção dos *dashboards* é a SAP *Business Objects Dashboards* (Xcelsius). Outras também foram essenciais na coleta e transformação dos dados como o *Data* Service no processo de ETL (extração, transformação e carga dos dados), *Web Intelligence* (relatórios gráficos e tabelas) e Ferramenta de *Design* de Informações (Universos – camada semântica). Os dados foram armazenados em um *data warehousing* aplicando modelagem dimensional e permitindo o processamento analítico on-line (OLAP) através de ferramentas *self services/data discovery*. O SGBD (Sistema de Gerenciamento de Banco de Dados) ainda é relacional padrão (*PostgreSQL*), porém está previsto a utilização de um SGBD orientado a coluna (*Sybase IQ*), mais apropriado para ambientes analíticos. Os painéis interativos foram desenvolvidos a partir da necessidade desafiadora de coletar, organizar, analisar e disseminar os crescentes dados agrícolas (como produção por culturas, área colhida e rendimento médio do país por produtos) com uma

tecnologia inteligente e ágil e gerar resultados úteis por modelagem multidimensional, análise estatística, mineração de dados e *business intelligence* (BI) (HIRA; DESHPANDE, 2015). O processo de implementação do projeto de BI incluiu as seguintes fases: i) Planejamento e Elaboração; ii) Construção; iii) Apresentação e Validação.

Planejamento e Elaboração

Nessa fase foram levantadas informações gerenciais e de apoio à decisão, as fontes de dados, os riscos, as responsabilidades e as premissas para definição do escopo do projeto, planejamento e sua aprovação. Ela se subdividiu em três etapas: levantamento de requisitos; mapeamento das fontes de dados; e, elaboração do documento de visão.

O levantamento de requisitos tem o objetivo de buscar e registrar quais as informações necessárias para análise de acordo com o objetivo do negócio, quais indicadores e seus níveis de detalhamentos seriam apresentados, bem como a política de segurança para o acesso. Nessa fase definiu-se iniciar o projeto com dados agropecuários da pesquisa do IBGE: Produção Agrícola Municipal (PAM); em seguida, foi realizado o mesmo para a Pesquisa Pecuária Municipal (PPM); e, por último, a Pesquisa Trimestral do Abate (PTA) (todas pesquisas do IBGE). Para cada pesquisa, identificou-se onde, como e quais dados e informações coletar conforme necessidade. Os dados disponibilizados da PAM consistem em informações estatísticas de lavouras pecuárias e permanentes com as variáveis: área colhida, produção, rendimento médio e valor da produção. As variáveis da PPM são: efetivo por tipos de rebanho, produção na aquicultura, vacas ordenhadas e ovinos tosquiados. A série temporal para os dados da PAM e da PPM são de 1990 a 2015 e estão apresentadas em todos os níveis territoriais. Os dados da Pesquisa Trimestral do Abate (PTA) se compõe dos tipos de abate de animais, pesquisa trimestral do couro e leite e produção de ovos de galinha. Todos possuem variáveis específicas.

Na sequência foi realizado o mapeamento das fontes de dados, para os quais, relativamente a cada requisito levantado, identificou-se suas fontes diretas, autorizações para sua utilização e forma de uso pelos gestores.

A última etapa foi a elaboração do documento de visão, o qual traz um panorama geral das necessidades e características dos produtos de BI a serem implementados, a fim de mitigar riscos de entendimento e facilitar análise pelo gestor do negócio do que serão apresentados como produtos.

Construção

Fase de modelagem dos dados conforme requisitos e implementação dos processos de ETL, que automatizam a coleta e a carga das diversas fontes no banco de dados do *business intelligence*. O objetivo foi preparar os dados no ambiente de BI para desenvolver seus produtos. A construção se subdivide em: modelagem dos dados (elaborar conforme requisitos e documento de visão), elaboração dos processos de ETL e elaboração de mecanismos de acesso aos dados no ambiente de BI, que são os universos e/ou serviços *webservices*. A criação dos universos consiste numa camada semântica que permite acesso fácil aos dados, se autorizados, para elaboração de produtos. O serviço *webservice* permite que produtos como *dashboards* acessem diretamente os dados no ambiente de BI, aumentando assim a interoperabilidade.

i) **Apresentação e Validação**

A fase de apresentação inclui a elaboração dos produtos de BI conforme os requisitos definidos no documento de visão. Entre aqueles utilizados no projeto tem-se universos, *dashboards*, relatórios *web intelligence*. Após coleta, transformação e disponibilização dos dados no ambiente de BI desenvolveu-se os painéis interativos PAM, PPM, PTA. Todos eles possuem relatórios *web intelligence* em seus detalhamentos ou para exportação de dados. No da PAM, por exemplo, há um relatório *web intelligence* comparativo de produção entre níveis territoriais iguais (exemplo comparativo de UF com UF: área colhida de milho entre BA e MG).

A validação tem o objetivo de validar os produtos sob a perspectiva do usuário/negócio e conforme os requisitos especificados no documento de visão. São avaliadas a existência de falhas ou interpretação errada com relação ao requisito funcional. O usuário tem um prazo para testar a funcionalidade do *dashboard* no ambiente de homologação e conferir se os dados estão corretos. Atendendo ao que foi proposto, é dado o aceite, se não, os mesmos são reajustados e validados novamente.

RESULTADOS E DISCUSSÃO

O sistema Agropensa disponibilizou várias formas para analisar dados de agricultura usando diferentes parâmetros para gerar resultados de interesses específicos dos usuários. Nas tabelas 1, 2 e 3 estão apresentados os dados e as variáveis obtidos e disponibilizados na forma de consultas interativas via três *dashboards*: Produção Agrícola Municipal, Pesquisa Pecuária Municipal e Pesquisa Trimestral do Abate. Esta ação permitiu ao público interagir e construir diferentes formas de visualização de dados agrícolas. Abaixo seguem informações e variáveis obtidas e disponibilizadas via Dashboards.

Tabela 1 – Base de Dados: Produção Agrícola Municipal (PAM) – período de 1990 a 2015

Tabelas	Culturas	Quant.	Variáveis	Unidade Territorial
Lavouras Permanente, Temporária e Safras	Abacate, Abacaxi, Algodão arbóreo, Algodão herbáceo, Alho, Amendoim, Arroz, Aveia, Azeitona, Banana, Batata-doce, Batata-inglesa, Borracha, Cacau, Café Arábica, Café Canephora, Café Total, Cana-de-açúcar, Caqui, Castanha de caju, Cebola, Centeio, Chá-da-índia, Coco-da-baía, Dendê, Erva-mate, Ervilha, Fava, Feijão, Figo, Fumo, Girassol, Goiaba, Guaraná, Juta, Laranja, Limão, Linho, Maçã, Malva, Mamão, Mamona, Mandioca, Manga, Maracujá, Marmelo, Melancia, Melão, Milho, Noz, Palmito, Pera, Pêssego, Pimenta-do-reino, Rami, Sisal ou agave, Soja, Sorgo, Tangerina, Tomate, Trigo, Triticale, Tungue, Urucum, Uva Safras (feijão, batata-inglesa, amendoim e milho)	66 culturas e 4 safras	Área Colhida; Produção; Rendimento médio; Valor da Produção	Brasil, Grandes regiões, UF, Mesorregião, Microrregião e Município

Fonte: PAM/Instituto Brasileiro de Geografia e Estatística (IBGE).

Tabela 2 – Base de Dados: Pesquisa Pecuária Municipal (PPM) - – período de 1990 a 2015 para todas as tabelas, exceto a de Produção de Aquicultura (período de 2013 a 2015)

Tabelas	Tipos	Variáveis	Unidade Territorial
Ovinos tosquiados (efetivo)	Ovinos tosquiados (em mil cabeças) cruzando com lã (tabela produtos de origem animal em quilogramas)	Número de cabeças X Produção de lã; Rendimento médio (kg/cabeça)	Brasil; Grandes Regiões; UF; Mesorregião; Microrregião e Município
Efetivo de rebanhos	Bovino, bubalino, equino, asinino, muar, suíno, suíno, caprino, ovino, galos-frangas-frangos-pintos, galinhas, codornas, coelhos	Número de cabeças (Em mil cabeças)	
Produtos de origem animal	Leite (mil litros); Ovos de galinha, ovos de codorna (em mil dúzias); Mel de abelha, casulos do bicho-da-seda e lã (em quilogramas)	Produção de origem animal; Valor da produção Preço médio (em mil reais); Percentual do valor da produção	
Vacas ordenhadas (efetivo)	Vacas ordenadas cruzando com Leite (tabela produtos de origem animal)	Número de cabeças X Produção de leite, rendimento médio	
Produção da aquicultura	Carpa, Curimatã, Curimbatá, Dourado, Lambari, Jatuarana, Matrinxã, Pacu, Piau, pintado, Pirapitinga, Pirarucu, Tambacu, Tambaqui, Tilápia, entre outros (em quilogramas)	Produção Valor da produção Preço médio	

Fonte: Instituto Brasileiro de Geografia e Estatística (IBGE).

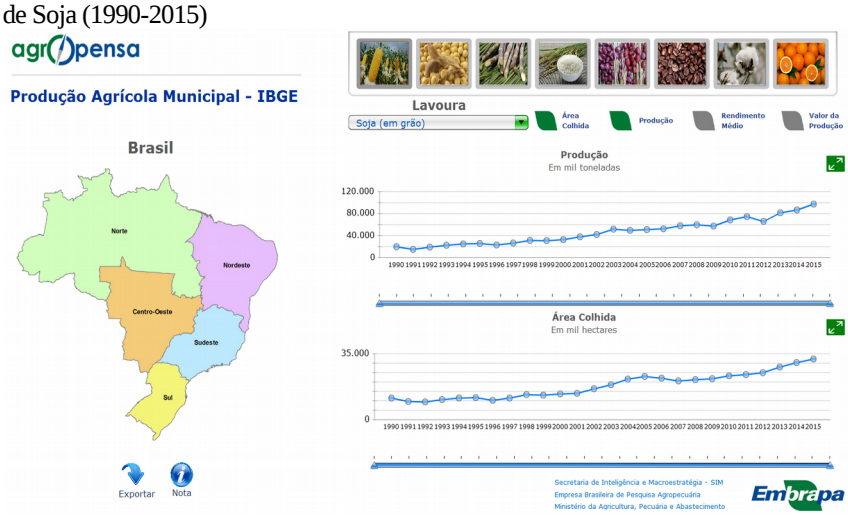
Tabela 3 – Base de Dados: Pesquisa Trimestral do Abate (PTA), das unidades territoriais Brasil, Grandes Regiões e Unidades Federativas (UF)

Tabelas	Tipos	Variáveis	Inspeção; Procedência do couro	Anos
Produção trimestral do abate de animais	Bovinos Abatidos: (Bois, vacas, novilhos, novilhas, vitelos e vitelas) Suínos Abatidos Frangos Abatidos	Número de informantes; Número de animais abatidos; Peso total das carcaças	Inspeção: Federal Estadual e Municipal	1997 a 2015
Produção de ovos de galinha	-	Número de informantes; Número de galinhas poedeiras; Quantidade de ovos produzidos		1995 a 2015
Pesquisa Trimestral do Couro	Adquirido para curtimento Adquirido de terceiros Curtido	Número de informantes Couro Bovino para curtimento	Procedência: Matadouro frigorífico; Matadouro municipal; Intermediários (salgadores) Outros curtumes Outras origens	1999 a 2015
Leite	-	Número de informantes; Leite adquirido; Leite industrializado	Inspeção: Federal Estadual e Municipal	1997 a 2015

Fonte: Instituto Brasileiro de Geografia e Estatística (IBGE).

As figuras 1, 2 e 3 representam as telas de entrada de cada um dos dashboards gerados (https://www.embrapa.br/agropensa/bases-de-dados). Esse método de visualização de dados pode apoiar a tomada de decisão gerencial (TELEA, 2015). Ao estudar esses fatores de uso da terra com outros conjuntos de dados geoespaciais é possível obter uma melhor compreensão dos processos relacionados com a dinâmica da agricultura no Brasil, como expansão, retração, transição e intensificação, ajudando a estabelecer políticas públicas para a região.

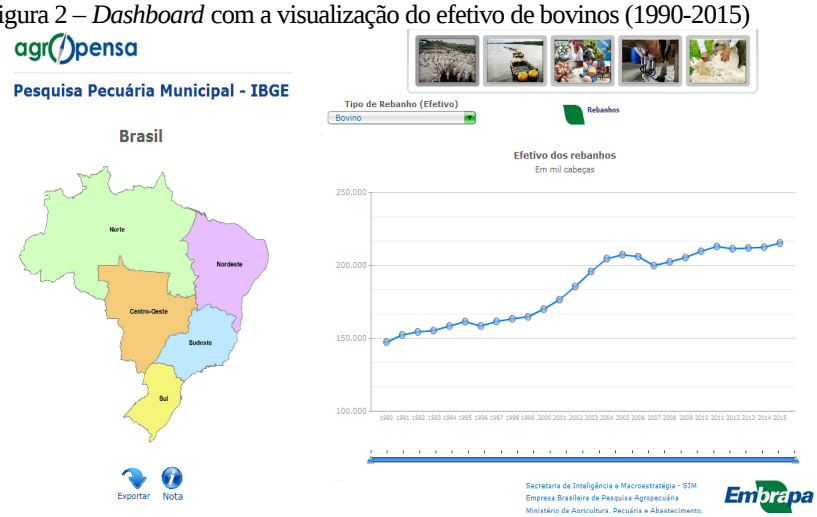
Figura 1 – *Dashboard* com a visualização dos dados da produção (em mil toneladas) e da área colhida (em mil hectares) de Soja (1990-2015)



Fonte: AGROPENSA, 2017 b.

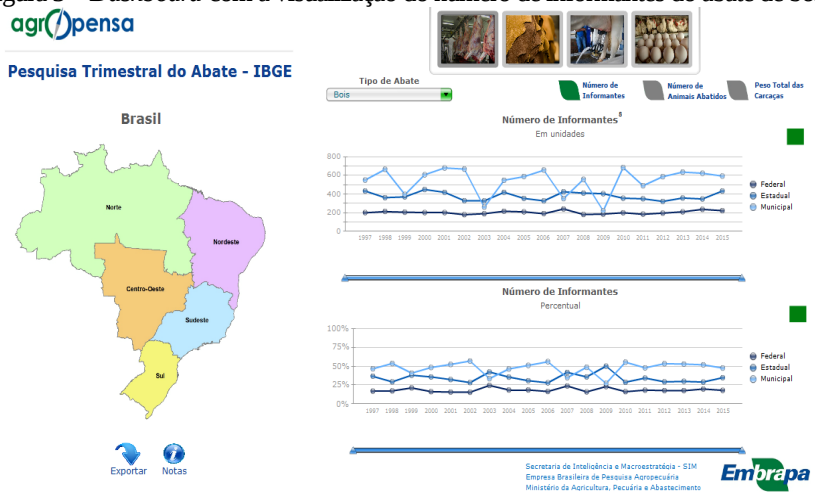
A Embrapa incorporou a plataforma de BI com capacidade de transformação e organização de grandes volumes de dados para potencializar o negócio e para uma inovadora visualização de dados em dados de agricultura de modo amigável e interativo. Com isso, ganha-se automaticamente uma vantagem competitiva em seu trabalho devido a dados precisos para análise; permite pré-planejamento dentro de ações de pesquisa, desenvolvimento e inovação.

Figura 2 – *Dashboard* com a visualização do efetivo de bovinos (1990-2015)



Fonte: AGROPENSA, 2017 b.

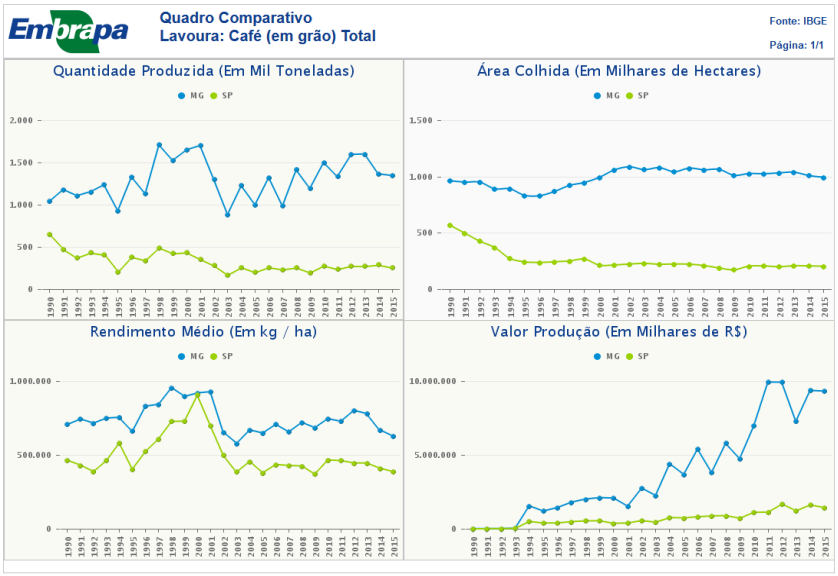
Figura 3 – *Dashboard* com a visualização do número de informantes do abate de bois (1997-2015)



Fonte: AGROPENSA, 2017 b.

Pode-se ainda gerar inúmeras análises *on-line* gerando gráficos, mapas e relatórios *web intelligence*, relógios ou marcadores com visão de desempenhos e que geram alertas de acordo com a escala definida e tabelas de indicadores. A ferramenta possibilita desenvolver diversas visualizações de dados com informações de alta qualidade para gestão do negócio e tomada de decisões. Na figura 4, é apresentada um relatório que compara as variáveis referentes a cultura de café (em grão) total entre os estados de Minas Gerais e São Paulo.

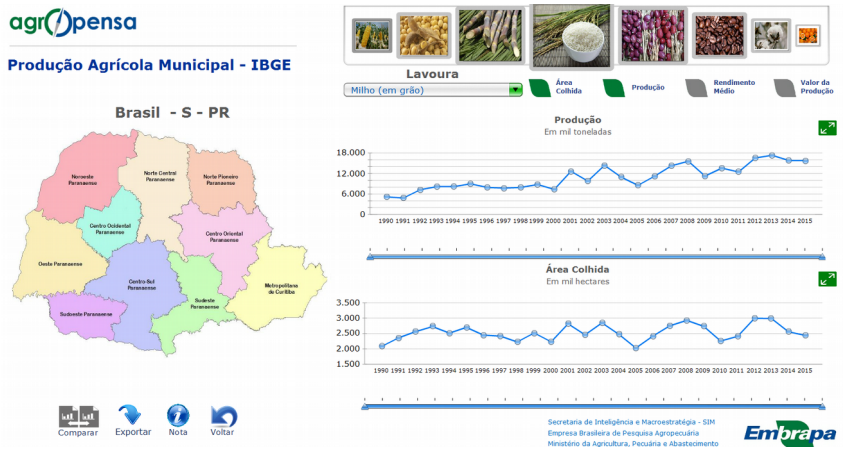
Figura 4 – *Dashboard* com a visualização comparativa das variáveis referentes à cultura de Café entre os estados de Minas Gerais e São Paulo (1990-2015)



Fonte: AGROPENSA, 2017 b.

A figura 5 mostra a evolução do milho no estado do Paraná de 1990 a 2015. A navegação dos níveis territoriais é feita clicando no mapa até a mesorregião. Microrregião e municípios estão apresentadas em lista numa caixa de texto para melhor visualização. Esses últimos níveis territoriais se subdividem em muitos subníveis, o que dificultaria a navegação devido a diversidade de divisões no mapa.

Figura 5 – *Dashboard* com a visualização da produção e área colhida de milho no estado do Paraná (1990-2015)



Fonte: AGROPENSA, 2017 b.

CONCLUSÕES

As informações organizadas estão disponíveis via Agropensa na forma de painéis interativos (*dashboards*) com acesso a dados da produção agrícola, pecuária e abates brasileiros dos últimos anos. Essa gama de dados disponibilizados apoia análises setoriais indicando tendências e trazendo vantagem competitiva na tomada de decisão estratégica. São ferramentas valiosas para a consulta e análise por agricultores, empresários, pesquisadores,

analistas, gestores e membros de cooperativas, associações, sindicatos, bancos, ministérios, estados e prefeituras. Dessa forma, torna-se um instrumental relevante para apoiar o planejamento estratégico público e privado da agricultura brasileira.

Tanto a visualização de dados como o BI são ferramentas que tornam o acesso e a análise de dados mais diretos e eficazes. Como um dos principais focos de atuação do Agropensa é mapear e apoiar a organização, a integração e a disseminação de base de dados e de informações agropecuárias, os esforços envidados e condensados nos painéis desenvolvidos e apresentados no presente artigo concorrem para a consecução do seu principal objetivo, qual seja: a elaboração de estudos de futuro, os quais contenham e difundam conhecimentos e informações em apoio à formulação de estratégias de Pesquisa, Desenvolvimento e Inovação (PD&I) para a própria Empresa e organizações parceiras.

REFERÊNCIAS

AGROPENSA. Sistema de Inteligência Estratégica da Embrapa (*home page*). Disponível em: www.embrapa.br/en/agropensa. Acesso em: 10 mar. 2017 a.

AGROPENSA. Página de painéis interativos do Agropensa. Disponível em: www.embrapa.br/en/agropensa/bases-de-dados. Acesso em:13 mar. 2017 b.

EMBRAPA. **Processo de Implementação de Projetos de *Business Intelligence* da Embrapa** - Anexo I. Brasília, DF, 2014.

GONZATTO, Jordano. **Gestão de desempenho e a tecnologia de Business Intelligence** – Vistra Tec. Disponível em: < https://mireporting.wordpress.com/2012/02/16/gestao-de-desempenho-e-a-tecnologia-de-business-intelligence/ >. Acesso em: 17 abr. 2017.

HIRA, S; Deshpande, P.S. **Data Analysis using Multidimensional Modeling, Statistical Analysis and Data Mining on Agriculture Parameters**, Procedia Computer Science, vol. 54, pp. 431-439, 2015. Disponível em: http://www.sciencedirect.com/science/article/pii/S1877050915013745?via%3Dihub. Acesso em: 15 mar. 2017.

IBGE. Instituto Brasileiro de Geografia e Estatística (IBGE) (*home page*). Disponível em: www.ibge.gov.br/english/disseminacao/eventos/missao/default.shtm. Acesso em 07 mar. 2017.

MYATT, G.J.; JOHNSON, W.P. **Making sense of data II: a practical guide to data visualization, advanced data mining methods, and applications**. Hoboken, New Jersey: John Wiley & Sons, 2009.

PAM. **Produção Agrícola Municipal** (*home page*). Disponível em: https://sidra.ibge.gov.br/pesquisa/pam/tabelas. Acesso em: 10 mar. 2017.

PEW, Stephen. **Information Dashboard Design**: The effective Visual Communication of Data. O'Really Media, 1 edition, 2006.

TELEA, A.C. **Data Visualization: Principles and Practice.** 2nd Ed. Boca Raton, FL: CRC Press, 2015.

Integração de Dados Geoespaciais na Rastreabilidade de Grãos

Silvia Ribeiro Mantuani [1], *Mônica Cristine Scherer Vaz* [2], *Maria Salete Marcon Gomes Vaz* [1]

[1] Programa de Pós Graduação em Computação Aplicada, Universidade Estadual de Ponta Grossa, Ponta Grossa, Paraná, Brasil, silviamantuani@gmail.com, salete@uepg.br.

[2] Programa de Pós Graduação em Engenharia da Produção, Universidade Tecnológica Federal do Paraná, Brasil, monicacsvaz@yahoo.com.br.

RESUMO

A segurança, qualidade e a origem dos alimentos são o foco dos consumidores atuais, que buscam informações relacionadas ao sistema de produção e cuidados com o ambiente. Sistemas de rastreabilidade convencionais apresentam-se como uma tecnologia adequada para analisar a situação de um produto em qualquer etapa de seu histórico na cadeia produtiva, porém não informam ao decorrer da cadeia de produção, os dados específicos em relação ao meio ambiente onde um lote do produto foi produzido. A geo-rastreabilidade mostra-se como uma forma de complementar essa carência dos sistemas de rastreabilidade, possibilitando abranger informações geográficas sobre o produto. Para integrar as informações geográficas já existentes é necessário utilizar a Infraestrutura de Dados Espaciais (IDE), para combinar diversas fontes de dados, originando informações sobre a área analisada. O objetivo desse artigo é especificar as tecnologias para integração de dados geoespaciais para vinculação da geoinformação ao *Framework* de Rastreabilidade de Grãos – RastroGrão. Com a integração dos dados geoespaciais, os agentes da cadeia produtiva e os consumidores têm informações seguras e precisas sobre os produtos que consomem, com verificação do local e entorno, onde foi produzido, transportado e armazenado o produto, além das práticas envolvidas na produção de determinado produto.

PALAVRAS-CHAVE: Segurança, RastroGrão, Geo-Rastreabilidade, Cadeia produtiva.

ABSTRACT

The safety, quality and origin of food are the focus of today's consumers, who seek information related to the production system and care for the environment. Conventional

traceability systems present themselves as a suitable technology to analyze the situation of a product at any stage in its history in the production chain, but do not inform the production chain of specific environmental data where a Was produced. Geotraceability is a way of complementing this lack of traceability systems, making it possible to cover geographic information about the product. In order to integrate the existing geographic information, it is necessary to use the Spatial Data Infrastructure (SDI) to combine several data sources, giving information about the analyzed area. The objective of this article is to specify the technologies for the integration of geospatial data to link geoinformation to the Grain Traceability Framework - RastroGrão. With the integration of geospatial data, the agents of the production chain and consumers have accurate and reliable information about the products they consume, with verification of the location and environment, where the product was produced, transported and stored, as well as the practices involved in the production of product
KEYWORDS: Security, Tracking, Geot raceability, Production Chain.

INTRODUÇÃO

A origem e a trajetória de produtos podem ser atestadas por meio de sistemas de rastreabilidade convencional, garantindo a qualidade e segurança alimentar. A desvantagem disso é que o sistema não informa, nas etapas da cadeia de produção, os dados específicos sobre influência geográfica dos produtos, dificultando a aplicação da rastreabilidade em cadeias produtivas. A geoinformação pode comprovar a qualidade dos produtos, possibilitando indicar padrões de produção relativos ao meio ambiente onde o lote do produto foi produzido, transportado ou armazenado.

A geoinformação, em geral, não fica disponível para os agentes da cadeia produtiva e para o consumidor final, porém podem fornecer informações adicionais e relevantes em relação à qualidade e à segurança do produto para o consumo, como especificidades sobre o local de produção com ponto de captação de água, rede rodoviária, área protegida, zona industrial, dados de meteorologia ou agroambiental, mapas de solo, mapas geológicos, hidrografia, modelo digital de elevação, entre outros (OGER et al., 2010).

Para complementar os sistemas de rastreabilidade convencionais surge a geo-rastreabilidade, possibilitando aplicar a geoinformação na rastreabilidade de cadeias produtivas, por meio de tecnologia de informação, sensoriamento remoto vinculando coordenadas geográficas com as informações sobre o produto e o entorno ambiental (EMBRAPA, 2010). A geo-rastreabilidade implica em vantagem competitiva, permitindo a

inclusão de informações específicas inerentes ao fluxo de dados de rastreabilidade, com dados geográficos agregando valor aos produtos (OGER et al., 2010).

A integração de dados geoespaciais favorece a tomada de decisão em relação localização geográfica para os agentes envolvidos na cadeia produtiva, auxiliando na garantia de origem de produtos, além de proporcionar maior segurança às informações levadas aos consumidores, e permitindo encontrar respostas a possíveis problemas relacionados ao produto devido ao meio em que foi produzido. Para integrar as informações geográficas já existentes é necessário utilizar a Infraestrutura de Dados Espaciais (IDE), para combinar diversas fontes de dados, originando informações sobre a área analisada.

O *Framework* de Grãos, RastroGrão, refere-se a um sistema web de rastreabilidade, podendo ser customizado de acordo com o perfil do usuário e permitindo o registro de dados dos agentes da cadeia de produção(VAZ, 2014). O RastroGrão não foi modelado para disponibilizar informações geográficas relacionadas à cadeia produtiva de grãos.

Este artigo objetiva especificar as tecnologias para integração de dados geoespaciais para vinculação da geoinformação ao *Framework* de Rastreabilidade de Grãos – RastroGrão.

INTEGRAÇÃO DE DADOS GEOESPACIAIS

Segundo Borges (1997), os dados espaciais são todos os tipos de dados que caracterizam fenômenos associados a qualquer aspecto espacial. A medida observada de um fenômeno sobre/sob a superfície terrestre é chamado dado geográfico. Para Câmara et. al., (2005), os dados geográficos ou georreferenciados são dados espaciais com dimensão espacial associada à sua localização na superfície da Terra, num determinado instante ou período de tempo.

Para integrar e compartilhar dados geoespaciais são utilizados padrões definidos pelo *Open Geospatial Consortium* (OGC). Para definir a Infraestrutura de Dados Espaciais (IDE) são utilizados os padrões geográficos definidos pelo OGC, *Web Feature Service* (WFS) (ISO/TC211, 2010), *Web Map Service* (WMS) (ISO/TC211, 2005) e *Web Coverage Service* (WCS) (ISO/TC211, 2010b).

No Brasil, a infraestrutura de dados integra os dados geoespaciais existentes nas diversas instituições do Governo, facilitando o compartilhamento e disseminação dos dados. Essa infraestrutura corresponde à coleção de bases de tecnologias, políticas, critérios, padrões e pessoas para promover o compartilhamento de dados geoespaciais por meio dos níveis de governo, setores privados, órgão sem fins lucrativos e universidades, com metadados sobre origem, qualidade e descrição semântica (FEDERAL GEOGRAPHIC DATA COMMITTEE, 2017).

Com o uso da infraestrutura é possível combinar dados de diversas fontes para originar informações sobre a área analisada, compreendida como a junção entre diversos provedores de dados geográficos, cada qual promove o acesso aos dados por meio de serviços web específicos e tecnologias para a integração de dados geoespaciais (DAVIS; ALVES, 2006).

Tecnologias para integração de dados geoespaciais

Metadados

Segundo o *Federal Geographic Data Committee* (FGDC), os metadados são dados que descrevem dados em relação ao conteúdo, a qualidade, aos procedimentos de geração, localização, entre outras características. Os metadados geoespaciais incluem os mais diversos tipos de dados georreferenciados como imagem de satélite, fotografias aéreas, amostragem geológica, observações biológicas, dados de redes de infraestrutura, dados censitários, entre outros (Decreto nº 6.666, de 27 de novembro de 2008).

Para utilizar metadados, na integração de dados geoespaciais, é necessário analisar os elementos usados para descrever o conteúdo do dado, sua finalidade e domínio. Segundo Kashyap, Shah e Sheth (1995), os metadados são classificados em metadados independente de conteúdo, capturando as informações e não dependendo do conteúdo dos dados associado; metadados dependentes do conteúdo dos dados e metadados descrevendo o conteúdo, possibilitando o entendimento do conteúdo sem acessar os dados diretamente.

A escolha do tipo de metadados implica diretamente no tipo de integração de dados geoespaciais a ser realizada. Se for integração semântica é necessário o uso do metadados descrevendo o conteúdo, o significado do dado, além de fazer uso da linguagem compartilhada *eXtensible Markup Language (XML)* e suas extensões.

XML

A *eXtensible Markup Language (XML)* é uma linguagem de marcação padrão estabelecida pelo *World Wide Web Consortium* (W3C,2003), independente de plataforma e com estrutura hierárquica. Os metadados escritos em XML favorecem a integração de dados, pois são compreendidos e lidos por diferentes sistemas, devido à escrita realizada explicitamente, por meio de *tags*.

A XML fundamentou *Geography Markup Language* (GML) (ISO/TC211, 2007) para o compartilhamento de informações geográficas e estabelecida pelo OGC (2012). Essa linguagem, também, possui estrutura hierárquica e utiliza tags para identificar seus elementos.

A *Scalable Vectorial Graphics* (SVG*),* baseada em XML, serve para descrever gráficos bidimensionais. Essa linguagem dá suporte a três tipos de objetos gráficos, vetores, imagem e

texto (W3C, 2003). Os gráficos gerados podem ser renderizados em diversos tamanhos sem perder a qualidade. Os dados são carregados de forma rápida, aumentando o desempenho e ampliando as possibilidades de integração de dados para visualização geométrica.

Ontologias

Segundo Gruber (1993), as ontologias são conjunto de conceitos padronizados, termos e definições para determinado domínio. Ontologia é uma especificação explícita de uma conceituação. Ontologias podem ser utilizadas para recuperação de informação na internet, processamento de linguagem natural, gestão do conhecimento, web-semântica e educação, oferecendo vantagens como comunicação, formalização, representação de conhecimento e reutilização. Cada ontologia tem sua finalidade, seu contexto compartilhado e permite a reutilização dos componentes, sendo uma forma de registrar e documentar conceitos e seus relacionamentos.

Para a integração dos dados é utilizada a geo-ontologia, ontologia com conceitos geoespaciais e seus relacionamentos, auxiliando o estabelecimento de vocabulário comum na integração de dados geoespaciais. Segundo CÂMARA et al. (2005), geo-ontologia possui dois tipos básicos de conceitos, o físico representando entidades ou fenômenos físicos do mundo real e conceitos sociais, representando entidades ou fenômenos sociais.

A geo-ontologia também classifica os dados geoespaciais, com uso de geo-objeto, elementos que possui fronteira bem definida associada a entidades identificáveis e geo-campos, elementos que possui variação contínua no espaço (CÂMARA et al., 2005).

FRAMEWORK DE RASTREABILIDADE DE GRÃOS – RASTROGRÃO

O Framework RastroGrão (VAZ, 2014) foi especificado para atender a demanda por integração dos dados e disponibilidade da informação para o cliente final, por meio da internet, buscando auxiliar o processo de rastreamento do processo de produção. Com a sua estrutura, é possível rastrear os produtos, as fases dos processos de produção de cada produto, além dos dados a serem rastreados em cada fase. O processo pode auxiliar na rastreabilidade de qualquer tipo de grãos, já que o sistema pode ser gerenciado pelo usuário de acordo com as suas necessidades.

O processo de rastreabilidade não é estático e deve abranger todos os agentes da cadeia produtiva, o usuário pode personalizar o processo, de acordo com cada necessidade. As regras do agronegócio e as normas da área estão em constantes pesquisas/atualizações, podendo resultar em novas necessidades em um processo de rastreabilidade.

O *Framework* foi definido com uma interface amigável, tanto para a criação da estrutura de rastreabilidade como para o registro das informações pelo usuário. O acesso é

realizado pela internet, permitindo que as informações estejam disponíveis para todos os agentes da cadeia. Desta forma, é possível atender às necessidades de cada empresa, visto que não se trata de um sistema com campos pré-definidos, mas sim, customizável e de fácil adaptação às regras do agronegócio.

MATERIAIS E MÉTODOS

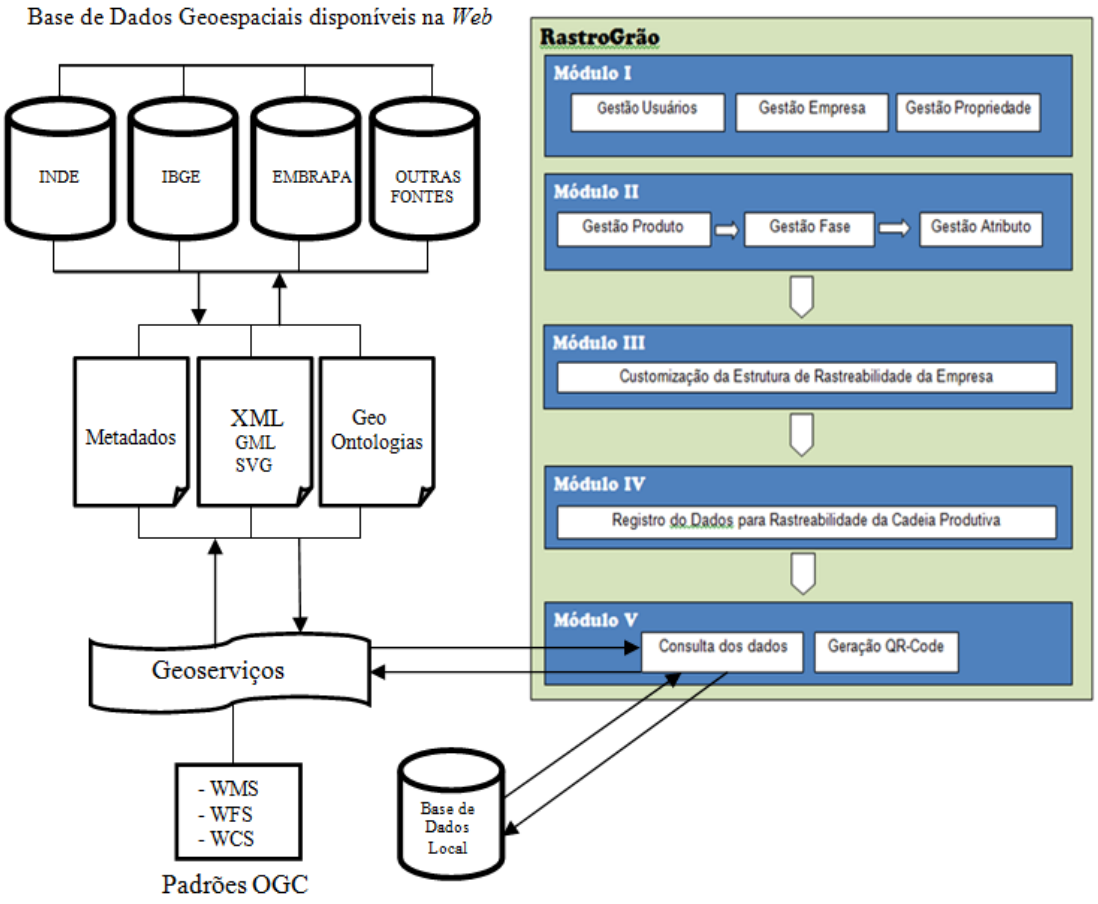
A estrutura especificada foi definida utilizando como base o *Framework* de Rastreabilidade RastroGrão (Vaz, 2014), considerando a análise de seus 5 (cinco) módulos para definir em qual desses será realizado a extensão com integração de forma automática dos dados geoespaciais disponibilizados por órgãos governamentais, por meio de geoserviços, garantindo segurança às informações acessadas pelos consumidores finais. Os módulos analisados são os que seguem. 1) Cadastro Geral, compreendendo a Gestão de Usuários para acesso ao *Framework*; Gestão das Empresas, compreendendo as entidades com o processo de rastreabilidade estruturado e serão as gestoras dos dados; e Gestão da Propriedade, compreendendo os locais onde os grãos são produzidos e de onde inicia o processo de rastreabilidade. 2) Gestão da Estrutura da rastreabilidade, relacionada à gestão dos produtos, fases e atributos que farão parte da biblioteca, podendo ser customizada para cada empresa. 3) Customização, módulo onde cada empresa seleciona os produtos, fases e respectivos atributos, fazendo parte da estrutura a ser rastreada. 4) Registro dos Dados, módulo onde a empresa fará a inserção dos dados da produção, a rastreabilidade efetiva. 5) Consulta dos dados e Geração das etiquetas de QRCode, contendo as informações que cada empresa inseriu no banco de dados e disponibilizadas para consulta da empresa e/ou usuário final.

RESULTADOS E DISCUSSÕES

A estrutura especificada por Vaz (2014) possibilita a extensão para gestão de geoinformação, integrando dados geoespaciais. O módulo definido para a integração dos dados geoespaciais foi o Módulo V(Figura 1), já que é no momento da consulta dos dados, que tanto empresa e/ou usuário final poderão solicitar informações detalhadas em relação ao produto adquirido.

A estrutura permite a captura de dados geográficos disponibilizados pelos órgãos públicos. Na captura, os dados serão solicitados por geoserviços, utilizando padrões de acordo com o tipo de informação desejada pelo usuário. Na sequência são utilizadas tecnologias para integração de dados geoespaciais, tais como o uso de metadados, Geo-Ontologia e as Linguagens *XML* e suas extensões *GML* e *SVG*. Os geoserviços permitirão a busca de informação em bases geoespaciais públicas disponíveis na *WEB*, tais como INDE, IBGE e EMBRAPA.

Figura 1 - Extensão da estrutura RastroGrão para integração de dados geoespaciais



Fonte: O autor

O Padrão *WMS* (ISO/TC211, 2005) especifica um serviço que produz dinamicamente mapas georreferenciados, disponibilizando o mapa solicitado na forma de imagem, nos Formatos *PNG*, *GIF* ou *JPEG* ou de forma vetorial utilizando a Linguagem *SVG*. Este padrão funciona com o uso de três funções principais: (i) *GetCapabilities*, retornando os metadados do serviço; (ii) *GetMap* retornando mapas com base em parâmetros geográficos e de representação estabelecidos; e (iii) *GetFeatureInfo* retornando atributos de feições específicas apresentadas no mapa.

O Padrão *WFS* (ISO/TC211, 2010) permite acesso à informação em nível de feições, nos mapas produzidos pelo Padrão WMS, por meio da Linguagem *GML*. As funções usadas nesse serviço são: (i) *GetCapabilities* retornando metadados do serviço; (ii) *DescribeFeatureType* recuperando a estrutura *XML* da feição; e (iii) *GetFeature* obtém os dados em si, de acordo com os parâmetros solicitados, retornando o arquivo dos dados em formato *GML*.

O Padrão *WCS* (ISO/TC211, 2010b) estabelece o acesso a elementos em formato matricial, referindo-se a conteúdos que são do tipo geocampo, tais como, imagens de satélites e dados de elevação digital, permitindo a interoperabilidade entre os dados geoespaciais.

As geo-ontologias são consideradas para a integração de dados geoespaciais, pela existência de ontologias especificadas formalmente, permitindo aos especialistas traduzirem as ontologias para classes por meio da orientação a objeto e possibilitando encontrar classes que servirão para integração de dados geoespaciais. Também, oferece ao usuário final metadados referentes às informações disponíveis através do uso de *XML* (SOARES; TANAKA; BAIÃO, 2010).

Com isso espera-se que o dado disponível permaneça de forma colaborativa e relevante para a sociedade. Giannopoulos et al. (2014) desenvolveram um sistema que reconhece diferentes representações de coordenadas geográficas com a aplicação de expressões regulares e ontologias de alinhamento, e usando a Linguagem *XML*. O sistema faz o reconhecimento de coordenadas e transforma o dado, por meio do processamento e análise dos dados geoespaciais de maneira uniforme. Antonialli (2015) propôs um *Framework* para integração de dados geoespaciais com o intuito de integrar dados geológicos com a utilização de ontologias, usando alinhamento de ontologias e a Linguagem *XML*.

Além de trabalhos envolvendo ontologias para integração de dados geoespaciais, há também trabalhos usando metadados e a Linguagem *XML* e suas extensões – *GML* e *SVG*, como o trabalho desenvolvido por Nakamura (2010), que elaborou uma proposta de IDE, visando o compartilhamento de dados geográficos para o Parque Estadual de Intervales – SP. Speranza et. al. (2012) utilizaram padrões e softwares para catalogar metadados geoespaciais de dados temáticos, para melhoria da organização e da qualidade dos dados, evitando duplicidade, e para incentivar a interoperabilidade entre sistemas.

Este artigo apresenta uma estrutura de integração de dados geoespaciais disponíveis na web, para oferecer informações geográficas, agregando valor ao produto, maior segurança e qualidade aos produtos permitindo aos agentes da cadeia produtiva e consumidores finais a obtenção das informações do local e entorno, onde o produto adquirido foi produzido, transportado e armazenado.

CONCLUSÕES E PERSPECTIVAS DE PESQUISAS FUTURAS

Nesse artigo foram pesquisadas as principais tecnologias que integram dados geoespaciais, visando utilizar dados disponíveis, gratuitamente, para adição de informações geográficas na rastreabilidade da cadeia produtiva de grãos. Além disso, foi realizada a extensão do *Framework* de Rastreabilidade de Grãos – RastroGrão (VAZ, 2014), com acréscimo da geo-

rastreabilidade, possibilitando aos agentes da cadeia e consumidores finais, a disponibilidade de informações precisas sobre os produtos que consomem, com verificação do local e entorno do local, onde foram produzidos, transportados e armazenados, além das práticas envolvidas na produção dos produtos. As informações disponibilizadas caracterizam as influências do meio ambiente e a percepção de como determinados ambientes influenciam as características dos produtos. Como perspectivas de trabalhos futuros sugere-se o desenvolvimento da estrutura especificada utilizando as tecnologias de integração de dados geoespaciais.

REFERÊNCIAS

ANTONIALLI, R. C. Framework para integração de dados geoespaciais: Integração de dados geológicos. 2015,126 f. Dissertação (Mestrado em Ciência da Computação) – UNESP, Rio Claro, 2015.

BORGES, K. A. V. Modelagem de dados geográficos: Uma Extensão do Modelo OMT para Aplicações Geográficas. 1997,139 f. Dissertação(Mestrado em Administração Pública) - Fundação João Pinheiro, Belo Horizonte, 1997.

CÂMARA, G.; DAVIS, C. A.; VINHAS, L.; QUEIROZ, G. R.; CASANOVA, M. Bancos de dados geográficos. Curitiba: MundoGeo, 2005.

DAVIS Jr., C. A,; ALVES, L. L. Interoperability through Web Services: Evaluating OGC Standards in Client Development for Spatial Data Infrastructures. In: PROCEEDINGS OF BRAZILIAN SYMPOSIUM ON GEOINFORMATICS, 2006, Minas Gerais. Anais.

DECRETO nº 6.666 de 27 de novembro de 2008. Institui a Infra Estrutura Nacional. Diário Oficial [da] República Federativa do Brasil, Poder Executivo, Brasília, DF, 28 nov. 2008. Seção 1, p.57.

EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária. Projeto internacional de Geo-Rastreabilidade participa de Expoagro na Argentina - Portal Embrapa. 2010 [acesso em 3 Jan 2017]. Disponível em: https://www.embrapa.br/busca-de-noticias/-/noticia/18117989/projeto-internacio nal -de-Geo-Rastreabilidade-participa-de-expoagro-na-argentina.

FEDERAL GEOGRAPHIC DATA COMMITTEE (FGDC). Content Standard for Digital Geospatial Metadata WorkBook. Reston, VA. Disponível em: <https://www.fgdc.gov/metadata>. Acesso 16 de Março de 2017.

GIANNOPOULOS, G. et al. FAGI-tr: A tool for aligning geospatial RDF vocabularies. In: ESWC 2014. 2014, London: IMIS Institute, “Athena” Research Center, 2014. p.5. Disponível em: < http://2014.eswc-conferences.org/sites/default/files/eswc2014pd_submission_35.pdf>. Acesso em: 15 de Março 2017.

GRUBER, T. R. A translation approach to portable ontologies. Knowledge Acquisition, Califórnia, v. 5, n. 2, p. 199-220, 1993.

ISO/TC211. ISO 19128:2005 Geographic information -- Web Map Server Interface. ISO/TC211, 2005

ISO/TC211. ISO 19136:2007 Geographic information -- Geography Markup Language (GML). ISO/TC211, 2007.

ISO/TC211. ISO 19142:2010 Geographic information -- Web Feature Service. ISO/TC211, 2010.

ISO/TC211. ISO 19143:2010 Geographic information -- *Web Coverage Service* ISO/TC211, 2010b.

KASHYAP, V.; SHAH, K.; SHETH, A. Metadata for building the MultiMedia Patch Quilt. In: Jajodia S, Subrahmanian V S (eds) Multimedia Database Systems: Issues and Research Directions. Springer-Verlag: 297-319. 1995.

NAKAMURA, E.T. Infraestrutura de Dados Espaciais em Unidades de Conservação: uma proposta para disseminação da informação geográfica do Parque Estadual de Intervales – SP. 2010. 142 f. Dissertação (Mestrado em Geografia Física). Universidade de São Paulo, São Paulo, 2010.

OPEN GEOSPATIAL CONSORTIUM – OGC (2012). Disponível em: <http://www.opengeospatial.org/standards/gml> Acesso: 08 de Março de 2017.

OGER, R. et al. Geotraceability: an innovative concept to enhance conventional traceability in the agri-food chain. Biotechnology Agronomy Society Environment. França, v. 14, 633-642, 2010.

SOARES, P.; TANAKA, A.; BAIÃO, F. Estudo dos principais conceitos sobre integração de dados geoespaciais. Relatório Técnico do Departamento de Informática, 2010, 38 f. Universidade Federal do Estado do Rio de Janeiro, 2010.

SPERANZA, E. A. et al. Metadados geoespaciais do Estado de Mato Grosso do Sul. In: 4º SIMPÓSIO DE GEOTECNOLOGIAS NO PANTANAL, 2012, Bonito. Anais. Bonito: Embrapa Informática Agropecuária/INPE, 2012, p. 395-404.

VAZ, M. C. S. Especificação de um Framework para Rastreabilidade da Cadeia Produtiva de Grãos. 2014, 87f. Dissertação de Mestrado em Computação Aplicada, Universidade Estadual de Ponta Grossa, Ponta Grossa, 2014.

W3C. Scalable Vector Graphics (SVG) 1.1 Specification. Versão 1.1. 2003.

Indicadores Geográficos para o Framework de Rastreabilidade de Grãos

Silvia Ribeiro Mantuani[1], Denise do Rocio Maciel[1], Monica Cristine Scherer Vaz[2], Maria Salete Marcon Gomes Vaz[1].

[1] Programa de Pós Graduação em Computação Aplicada, Universidade Estadual de Ponta Grossa, Ponta Grossa, Paraná, Brasil, silviamantuani@gmail.com, dnise_maciel@hotmail.com, salete@uepg.br.

[2] Programa de Pós-Graduação de Engenharia de Produção da Universidade Tecnológica Federal do Paraná, Paraná, Brasil, monicacsvaz@yahoo.com.br.

RESUMO

O consumidor está mais exigente em relação à segurança, qualidade e a origem dos alimentos que consome, buscando saber as condições ambientais em que o alimento é produzido. O uso de sistemas de rastreabilidade de alimentos é essencial, porém para garantir a disponibilidade de todas essas informações é necessário agregar a geoinformação nas etapas possíveis da cadeia produtiva. Esse processo é conhecido como Geo-Rastreabilidade. A associação de indicadores geográficos e demais informações resulta na melhoria da segurança do produto rastreado. O objetivo deste artigo é identificar os indicadores geográficos para integrar a Geo-Rastreabilidade a *Framework* de rastreabilidade de grãos, RastroGrão. Foi realizado levantamento de requisitos a partir de revisão de literatura, análise de sistemas de gestão de geoinformação e análise de softwares que implementam geoinformação integrada a dados de rastreabilidade de cadeias produtivas. Foram especificados os indicadores geográficos e sua aplicabilidade nas fases apropriadas existentes no RastroGrão. Os geoindicadores favorecem a tomada de decisão tanto para os agentes da cadeia produtiva quanto para os produtores, além disso, dispõem de informações detalhadas em relação ao ambiente entorno de um local de produção.

PALAVRAS-CHAVE: Geoinformação, Cadeia produtiva, RastroGrão, Geoindicadores, Geo-rastreabilidade.

ABSTRACT

The consumer is more demanding regarding the safety, quality and origin of the food that consumes, seeking to know the environmental conditions in which the food is produced. The use of food traceability systems is essential, but to ensure the availability of all this information, it is necessary to aggregate geoinformation in the possible steps of the production chain. This process is known as Geotraceability. The association of geographic indicators and other information results in improved security of the crawled product. The objective of this article is to identify the geographic indicators to integrate Geotraceability into the Grain Traceability Framework, RastroGrão. It was carried out requirements survey from literature review, analysis of geoinformation management systems and analysis of software that implement geoinformation integrated to traceability data of productive chains. Geographical indicators and their applicability were specified at appropriate stages in the RastroGrão. Geoindicators favor decision making both for the agents of the productive chain and for the producers, in addition, they have detailed information regarding the environment surrounding a production site.

KEYWORDS: Geoinformation, Production chain, RastroGrão, Geoindicators, Geotraceability.

INTRODUÇÃO

Com o aumento de importações e exportações de alimentos, o consumidor tornou-se mais exigente em relação à segurança, qualidade e a origem dos alimentos que consome, além de buscar saber quais as condições ambientais em que tal alimento foi produzido. Dessa forma, o uso de sistemas de rastreabilidade de alimentos é essencial. Para garantir a disponibilidade de todas essas informações é necessário agregar a geoinformação em todas as etapas possíveis da cadeia produtiva, aplicando o conceito de Geo-Rastreabilidade. A União Européia - UE foi pioneira em integrar dados de rastreabilidade com a geoinformação (PIZZIOL; LOUDJANI; GRANDGIRARD, 2007)

A geoinformação permite análise espacial de fenômenos geográficos, sua eficiência está relacionada a disponibilidade de dados geoespaciais. Geo-Rastreabilidade é a aplicação de geoinformação, através de tecnologia da informação e sensoriamento remoto na rastreabilidade de cadeias produtivas. Associa atributos espaciais a informações do produto (EMBRAPA, 2010). O *Framework* de Grãos,

RastroGrão, refere-se a um sistema web de rastreabilidade que pode ser customizado de acordo com o perfil do usuário ele permite o registro de dados de todos os agentes da cadeia de produção(VAZ, 2014). O *Framework* não foi modelado para disponibilizar informações em relação ao geoinformação dos grãos

O objetivo desse artigo é identificar os indicadores geográficos para integrar a Geo-Rastreabilidade a *Framework* de rastreabilidade de grãos, RastroGrão (VAZ, 2014).

FRAMEWORK DE RASTREABILIDADE DE GRÃOS

A identificação da origem e do percurso realizado pelo produto é requisito fundamental para sistemas de gestão de qualidade e processos de normalização e certificação. Além das exigências legais em torno da rastreabilidade, o consumidor final tem acesso às informações a respeito dos produtos a serem consumidos, motivando o desenvolvimento de soluções tecnológicas (DA COSTA, 2015).

O Framework RastroGrão (VAZ, 2014), realiza rastreabilidade do sistema produtivo de grãos. O mesmo visa o registro de dados de todos os agentes da cadeia de produção, assim como, a posterior disponibilização desses dados entre os agentes participantes e o cliente final(VAZ *et al.*, 2014).Os dados a serem rastreados são informados pelo próprio administrador do sistema e podem ser alterados conforme a necessidade de cada agente, com a definição dos Produtos, Fases e Dados a serem rastreados.

GEOINFORMAÇÃO E GEO-RASTREABILIDADE

A geoinformação é o produto do processamento e análise dos dados geoespaciais, quando comunicada, interpretada e aplicada, resulta na construção de conhecimento (HUBNER *et al.*, 2008). A Geo-Rastreabilidade é a aplicação de geoinformação por meio de tecnologia da informação e sensoriamento remoto, na rastreabilidade de cadeias produtivas, ela associa indicadores espaciais com informações do produto. Na Geo-Rastreabilidade estão definidas as coordenadas geográficas, compondo uma análise integrada dos processos de produção, unindo os dados da rastreabilidade convencional com as visões espacial e temporal desses sistemas (TÔSTO *et al*., 2014).

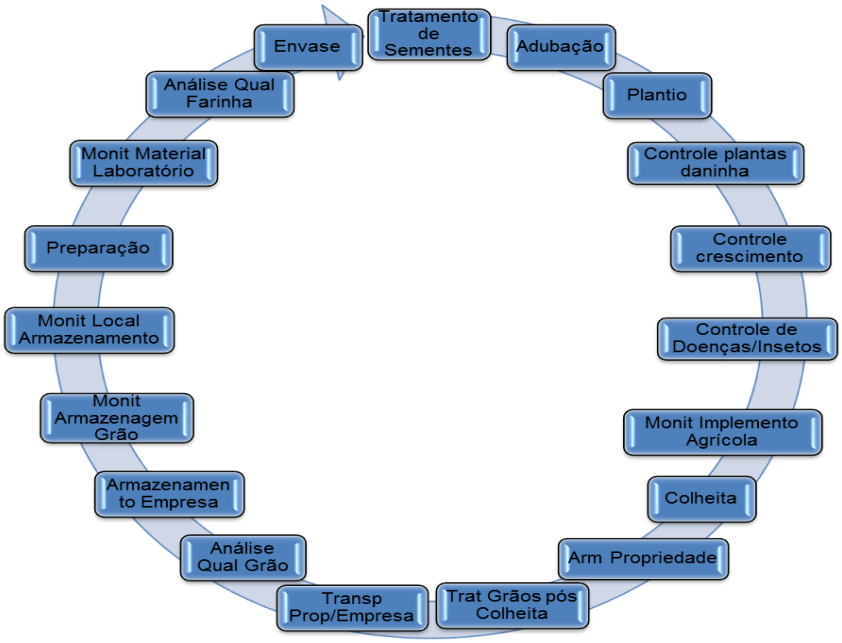
MATERIAL E MÉTODOS

A metodologia proposta para a pesquisa foi a análise de softwares que integram manualmente as informações de geo-rastreabilidade e a análise do Framework

RastroGrão para definição da geoinformação inerente as fases pertinentes para integração da geoinformação.

Para definir os indicadores geográficos para o RastroGrão foram escolhidas em quais fases (Figura 1) da cadeia produtiva, levantadas no trabalho de Vaz (2014), poderá ser inserida a geoinformação. Foi estabelecido como critério de escolha, o local onde a fase é realizada, se a fase é realizada dentro de um local fechado, por exemplo, em laboratório, essa fase não será usada para a integração do processo de geo-rastreabilidade.

Figura 1 - Fases da cadeia produtiva de grãos



Fonte: Adaptado de Vaz (2014)

A estrutura especificada foi definida utilizando como base o *Framework* de Rastreabilidade RastroGrão (Vaz, 2014), considerando os conceitos de rastreabilidade e geoinformação, de maneira separada, visando a união desses para gerar a geo-rastreabilidade e definir os indicadores geográficos para o *Framework*.

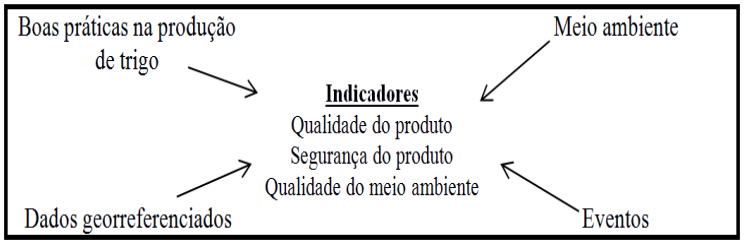
RESULTADOS E DISCUSSÃO

Nos sistemas de rastreabilidade, a informação de um determinado produto está associada a um código exclusivo, permitindo que possa ser consultada a qualquer tempo na cadeia produtiva. Para sistemas de rastreabilidade com a geoinformação, o código deve conter informações geográficas e administrativas para caracterizar e identificar, em espaço e tempo, o produto que será consultado. Para garantir a disponibilidade de tais informações, a Geo-Rastreabilidade usa indicadores geográficos, em conformidade com

as normas definidas, para integrar informações de fonte, qualidade e escala de observação (OMETTO et al., 2007).

Para estabelecer os indicadores de Geo-Rastreabilidade no RastroGrão é necessário obedecer às categorias (Figura 2):

Figura 2 - Categorias para compor indicadores de geo-rastreabilidade



Fonte: Otag (2008)

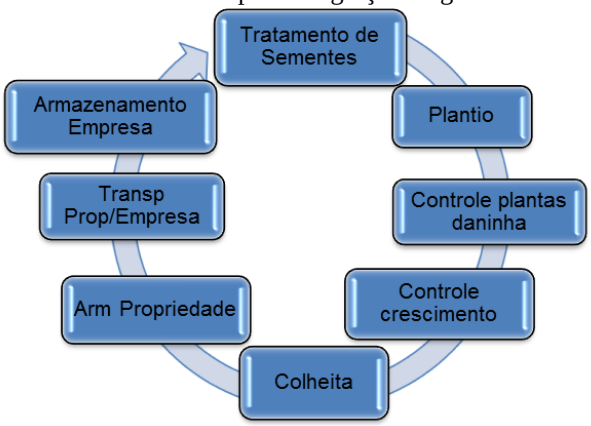
Os indicadores de Geo-Rastreabilidade envolvem a correlação entre um produto, uma parcela e seu ambiente. As incidências das práticas agrícolas no meio ambiente, incidências do ambiente na produção de grãos e influência das características do local de produção são analisadas. Assim, são considerados os aspectos do ambiente, como climáticos, topográficos e pedológicos, e os locais vizinhos aos dados, envolvendo os arredores da lavoura, como florestas, hidrografia e estradas, e os dados de eventos com a perspectiva temporal.

Para estabelecer os indicadores de Geo-Rastreabilidade, essa tecnologia faz uso dos conceitos de geoidentificador e de geoindicador. Um geoidentificador é definido com base em dois componentes. O primeiro componente contém informações sobre as características gerais do objeto geográfico e o segundo é o componente avançado referente ao editor de código ou aos dados de rastreabilidade inseridos pelo agente, para a unidade de produção.

O geoindicador é definido como um parâmetro quantitativo ou qualitativo, que fornece uma descrição geral do estado do ambiente ou de um produto, podendo ser usado para fins de inspeção (MAURIZI; VERREL, 2002).

Para definir os indicadores geográficos para o RastroGrão foram consideradas análises de trabalhos correlatos e as fases da cadeia produtiva escolhidas do trabalho de Vaz (2014). A Figura 3 apresenta as fases definidas para a integração do processo de geo-rastreabilidade, conforme os critérios pré-estabelecidos.

Figura 3 - Fases definidas para integração de geo-rastreabilidade



Fonte: Os autores

As fases da cadeia de produção, consideradas como pertinentes para a aplicação da Geo-Rastreabilidade são: tratamento de sementes, plantio, colheita, armazenamento propriedade, controle plantas daninhas e crescimento, transporte propriedade/empresa e armazenamento empresa. Cada fase será complementada com o indicador geográfico adequado a suas características.

No Quadro 2, são apresentados os geoidentificadores e os geoindicadores assinalando com "x" a fase que cada indicador geográfico será integrado.

Quadro 1 - Indicadores geográficos para o RastroGrão

Atributos			**Fases**							
			Tratamento de Sementes	Plantio	Colheita	Armazenamento Propriedade	Controle plantas daninhas	Controle crescimento	Transporte Propriedade/ Empresa	Armazenamento Empresa
Geoidentificador	1º componente	- longitude - latitude - altitude	x	x	x	x	x	x	x	x
	2º componente	- código	x	x	x	x	x	x	x	x
Geoindicadores	- Informações Meteorológicas		x	x	x	x	x	x	x	x
	- Hidrografia		x	x	x			x		
	- Florestas			x				x		
	- Estradas								x	
	- Perspectiva temporal		x	x	x	x	x	x	x	x
	- Topografia			x	x	x	x	x	x	

Fonte: O autor

Os geoidentificadores definem a localização e identificação do produto. O geoindicador de informações meteorológicas está relacionado às condições do tempo, o

geoindicador de hidrografia traz dados relacionados a irrigação e qualidade da água. O geoindicador florestas refere-se à existência de reservas florestais em volta da lavoura. O geoindicador estradas opera sobre o planejamento logístico. O geoindicador perspectiva temporal contextualiza acontecimentos no tempo e o geoindicador topografia determinar o relevo da plantação. Os geoindicadores favorecem a tomada de decisão tanto para os agentes da cadeia produtiva quanto para os produtores, além disso, dispõem de informações detalhadas em relação ao ambiente entorno de um local de produção.

O Quadro 2 apresenta uma análise comparativa entre os softwares de integração de Geo-Rastreabilidade estudados, focando o objetivo do software, as fases em que aplicam a geoinformação e as tecnologias envolvidas. Também é possível verificar as fases da rastreabilidade que podem ser integradas a geo-rastreabilidade no RastroGrão.

Quadro 2 - Comparação entre os softwares analisados que integram a geo-rastreabilidade

Software	**Objetivo**	**Fases de aplicação da geoinformação**						**Tecnologias envolvidas**
				Pós-colheita				
		Produção	Colheita	Transporte	Armazenamento	Processamento	Distribuição	
GeoRastro	Controlar a qualidade da carne bovina	x	*	*	*	*	*	Equipamentos eletrônicos
GeoTraceAgri	Definir uma metodologia para a amostragem, aquisição, utilização e tratamento de dados georreferenciados que serão usados para gerar indicadores agroambientais em diversas escalas geográficas.	x	*	*	*	*	*	Geoprocessamento Sensoriamento Remoto
GeoWine	Definir vinhos de qualidade por meio de indicadores	x	*	*	*	*	*	QR-code Geoprocessamento
GTIS CAP	Definir e criar um sistema integrado de informação que auxilie tanto os órgãos administrativos europeus como nacionais responsáveis pela Política Agrícola Comum, e aos produtores de produtos vegetais e animais.	x	*	*	*	*	*	Sensoriamento Remoto Imagens de satélites
GeoFairTrade	Desenvolver novos indicadores sociais, econômicos e ambientais baseados em transparência para melhorar comércio justo.	x	*	*	*	*	*	GPS Mapas
RastroGrão	Rastrear o processo produtivo de grãos	x	x	x	x	x	x	Bases de dados geoespaciais públicas

* Não mencionada na literatura analisada.

Fonte: O autor

O Software GeoTraceAgri estabelece o padrão para aplicação da Geo-Rastreabilidade em softwares para agricultura. O Software GeoWine foi desenvolvido usando a metodologia proposto no Software GeoTraceAgri, assim como o GTIS CAP, ambos com finalidades diferentes, embora voltados para agricultura.

O Software GeoFairTrade, foi desenvolvido para apoiar o agricultor em relação a venda no comércio justo. A geoinformação aplicada nos Softwares analisados cobre, principalmente, a fase de produção.

Os Softwares apresentados utilizam diversas tecnologias para aquisição de dados, entre elas, dispositivos eletrônicos para armazenamento da movimentação de cada bovino no pasto, permitindo verificar as condições de produção bovina, no GeoRastro.

O GeoWine faz uso da tecnologia QR-Code para que os consumidores finais possam acessar via internet todas as informações de vinho presentes na garrafa que o mesmo adquiriu. O sensoriamento remoto e geoprocessamento são usados nos Softwares GeoTraceAgri, GTIS CAP e GeoWine. No GeoFairTrade são utilizados GPS e Mapas.

No RastroGrão a geo-rastreabilidade é realizada nas fases de produção, colheita, transporte, armazenamento, processamento e distribuição. A informação integrada ao RastroGrão através dos órgãos públicos, contempla os agentes da cadeia produtiva em qualquer uma das fases, ou seja, em algum momento do trajeto de determinado agente, a informação auxilia na tomada de decisão. O consumidor final tem informações precisas e detalhadas em relação ao produto que adquiriu.

A geo-rastreabilidade foi usada em todos os Softwares na fase de produção. Para a validação de cada modelo, foi realizada a implantação piloto em áreas pertinentes de cada região, possibilitando a aquisição de dados para cada sistema.

O RastroGrão apresenta geo-rastreabilidade para as etapas de produção, colheita, transporte, armazenamento, processamento e distribuição, integrando os dados geoespaciais e visando a disponibilidade de informações inerentes aos indicadores levantados.

A geoinformação nos softwares analisados é aplicada de forma manual e somente na fase de produção, para o RastroGrão é possível realizar a integração da Geo-Rastreabilidade em todas as fases da rastreabilidade na cadeia produtiva com informações advindas de fontes confiáveis.

CONCLUSÕES E PERSPECTIVAS FUTURAS

Os indicadores geográficos contribuem para facilitar a comunicação entre os diversos agentes da cadeia produtiva, possibilitando avaliar melhoras na produção e práticas agrícolas, eficiência na gestão de riscos sanitários e da qualidade do produto, auxiliar na tomada de decisão, como a escolha das melhores práticas agrícolas, além de facilitar a inspeção para empresas de certificação proporcionando a supervisão da produção agrícola sustentável. Contribuem para que os cidadãos interessados na segurança alimentar e ao mesmo tempo na proteção ambiental, obtenham maiores informações a respeito dos aspectos do local onde foi produzido o produto adquirido.

Essa tecnologia é essencial, visto que o componente espacial aumenta o valor dos produtos no mercado, sendo utilizado como uma maneira de acrescer a confiança nos produtos adquiridos pelos consumidores, que terão o conhecimento em relação à trajetória, segurança e qualidade da produção ao consumo.

Como perspectivas futuras pretende-se especificar a estrutura de integração de dados geoespaciais no RastroGrão, utilizando informações de órgão governamentais seguros e disponíveis na *web*.

AGRADECIMENTOS

Os autores agradecem o suporte financeiro concedido pela Comissão de Aperfeiçoamento do Pessoal do Nível Superior – CAPES.

REFERÊNCIAS

CÂMARA G., DAVIS C., MONTEIRO A.M.V. Introdução a ciência da Geoinformação, Disponível em: <http://www.dpi.inpe.br/Gilberto /livro/ introd/>. Acesso em: 3 jan 2017.

DA COSTA, J., VAZ, M.S.M.G., VAZ, M.C.S. Geração de QR-Code para rastreabilidade da produção de grãos em dispositivos móveis sem acesso à internet. In: X CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 2015, Ponta Grossa. Anais. Ponta Grossa: UEPG, 2015, p. 21-23.

DEBORD M. Consumers' New Demand on Sustainable Traceability. In: WORLD CONFERENCE ON AGRICULTURAL INFORMATION AND IT; 2008, Japan. Anais. Japan, 2008, p. 23-26.

EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária -. Embrapa Monitoramento por Satélites. 2014. Disponível em: <https://www.cnpm.embrapa.br/projetos/georastro/conteudo/resumo.html>. Acesso em: 23 dez 2016.

EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária -. Projeto internacional de Geo-Rastreabilidade participa de Expoagro na Argentina - Portal Embrapa. Disponível em:< https://www.embrapa.br/busca-de-noticias/-/noticia/18117989/projeto-internacional -de-Geo-Rastreabilidade-participa-de-expoagro-na-argentina>. Acesso em: 3 jan 2017.

HUBNER C., OLIVEIRA F.H. Gestão da Geoinformação em Implementações Multiusuários. In: CONGRESSO BRASILEIRO DE CADASTRO TÉCNICO MULTIFINALITÁRIO, 2008,Florianópolis. Anais. Florianópolis, UFSC, 2008, p. 19-23;... Disponível em: <http://www.geolab.faed. udesc.br/publicacoes/Cleice/cobraco2008_1.pdf>. Acesso em: 4 jan. 2017.

MAURIZI B., VERREL, J.L. Des indicateurs pour les actions de maîtrise des pollutions d'origine agricole. 2002, p. 13-14.

OMETTO A.et al. Geotraceability and life cycle assessment in environmental life cycle management: towards sustainability[Internet]. Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA.2007. Disponível em: <https://www.embrapa.br/busca-geral/-/busca/Geotraceability?buscaPortal= Geotraceability>. Acesso em 1 fev 2017.

OTAG REPORT, 2008. Traceability in beef production and crisis management in bovine sector: state of art. OTAG Project Report. 2008.

PIZZIOL, P.; LOUDJANI, P.; GRANDGIRARD, D. On-farm geo-traceability as advanced tool for a competitive and sustainable agriculture. In: International Symposium on Food Traceability, European Commission Agriculture and Fisheries Unit Institute for Protection and Security of citizens Joint Research Centre 2007 Maio 10-11; República da Coreia. 2007. Disponível em: < https://www.researchgate.net/publication/274077288> Acesso em: 08 jan 2017.

TELEPARC: Geomatics centre - European Geomatics center of Gers - European projects GeoTraceAgri – Geotraceability; Geo Fair Trade: Geomatics & The Fair Trade Industry; GTIS CAP: Geotraceability, Common Agricultural Policy [Internet]. 2016. Disponível em: <http://www.teleparc.net/index.php?option=com_projets&task=%20view_projet_tab&id=%2013&%20toc=1&toc_type=%203&toc_val=4>. Acesso em: 3 jan 2017.

TÔSTO S.G., et al. Geotecnologias e Geoinformação: O produtor pergunta, a Embrapa responde. Empresa Brasileira de Pesquisa e Agropecuária. 2014, 256 p.

VAZ M.C.S., et al. Geração de QR-Code para Acesso aos Dados Rastreados na Cadeia Produtiva de Grãos. Revista Espacios, v.35, n. 2, p. 12, Jan 2014.

VAZ, M. C. S. Especificação de um Framework para Rastreabilidade da Cadeia Produtiva de Grãos. 2014, 87f. Dissertação de Mestrado em Computação Aplicada, Universidade Estadual de Ponta Grossa, Ponta Grossa, 2014.

ST-Vis: Uma ferramenta para visualização espaço-temporal em regiões de plantio de cana-de-açúcar

Ana Paula dos Santos Braatz Vieira[1], *Rafael Stoffalette João*[1], *Luciana Alvim Santos Romani*[2], *Marcela Xavier Ribeiro*[1]

[1]Departamento de Computação, Universidade Federal de São Carlos (UFSCar), São Carlos, Brasil, {ana.vieira, rafael.joao, marcela}@dc.ufscar.br

[2]Embrapa Informática Agropecuária, Campinas, Brasil, luciana.romani@embrapa.br

RESUMO

Interpretar e compreender dados espaço-temporais e suas variações ao longo do tempo é, de fato, uma tarefa complexa, pois na grande maioria das vezes os dados são dispostos em forma de tabela e texto. A ferramenta ST-Vis (*Spatio-Temporal Visualization*), apresentada neste trabalho, visa auxiliar o especialista do domínio de plantio de cana-de-açúcar a compreender os dados espaço-temporais. A ST-Vis fornece uma representação visual de dados textuais e tabulares, combinando o gráfico de Coordenadas Paralelas e o Sistema de Informação Geográfica (SIG), obtendo-se uma forma de visualizar dados espaço-temporais com interação por meio de um mapa. Além disso, a ST-Vis aplica a Textura Temporal mapeando de forma linear a variação do Índice de Vegetação por Diferença Normalizada (NDVI) de áreas produtoras de cana-de-açúcar. Foi realizada uma avaliação preliminar com especialistas de domínio, verificando-se que a visualização proposta na ferramenta permite uma melhor compreensão dos dados e possibilitando a identificação de padrões.

PALAVRAS-CHAVE: Dados Espaço-Temporais, Coordenadas Paralelas, NDVI.

ABSTRACT

Interpreting and comprehending spatio-temporal data and their variations over time is indeed a complex task, since the vast majority of cases the data are available in table and text form. The tool ST-Vis (Spatio-Temporal Visualization), presented in this work, aims to help the expert on sugarcane planting do comprehend the spatio-temporal data. ST-Vis provides a texture representation of tabular data, combining the Parallel Coordinates graph and the Geographic Information System (GIS), obtaining a way to visualize spatio-temporal data with interaction through a map. In addition, ST-Vis applies the Temporal Texture by linearly mapping the variation of the Normalized Difference Vegetation Index (IVDN or NDVI) of sugarcane producing areas. In this way, all this visualization helps the data comprehension by the expert, where he/she can identify patterns and rules in the information.

KEYWORDS: Spatio-Temporal data, Parallel Coordinates, NDVI.

INTRODUÇÃO

A análise de dados complexos como os dados espaço-temporais tem sido importante aos especialistas, e está sendo um desafio na área de geoinformação. Essa área utiliza representações estáticas de fenômenos espaciais dinâmicos que não conseguem capturar todas as informações e idealizá-las de uma forma adequada, havendo perdas de informações temporais na representação visual (CASANOVA et al., 2005). Assim, existe a necessidade de desenvolver modelos espaço-temporais para conseguir representar de forma adequada os fenômenos que variam tanto no espaço como no tempo, uma vez que, com essas análises pode-se detectar padrões e regras nos dados.

Existem poucas ferramentas que tratam os dados espaciais que variam ao longo do tempo, como a ferramenta VIS-STAMP (*Visualization System for Space-Time and Multivariate Patterns*) (GUO et al., 2006) que utiliza pequenos mapas cartográficos auto organizados e coloridos em uma tabela para uma visualização dos dados espaciais e temporais. Já a ferramenta GeoVISTA *CrimeViz* (ROTH et al, 2010) possui um mapa do Google no qual marca os focos de regiões com os tipos de criminalidade na representação espacial e possui uma visualização temporal por meio de histograma de frequência. A ferramenta GeoSTAT (*GEOgraphic SpatioTemporal Analysis Tool*) (DE OLIVEIRA, M. G.; DE SOUZA B. C. 2012) possui um mapa interativo do Google para a representação espacial e um gráfico bidimensional para a representação temporal, e atua com dados sobre focos de

incêndio. A ferramenta InfoScope (INFOSCOPE, 2015) tem um mapa cartográfico interativo para a visualização espacial, gráfico de Coordenadas Paralelas e algumas técnicas para visualização de informação, porém apenas aceita arquivos com extensão .mis (*Macrofocus InfoScope*) próprios da ferramenta.

Contudo, a ST-Vis visa representar os dados espaciais que variam ao longo do tempo por meio de uma visualização simultânea e de fácil compreensão. A ferramenta fornece visualizações interativas por meio da representação espacial por meio de mapa, da representação temporal por meio da Textura Temporal e por fim, da representação espaço-temporal por meio do gráfico de Coordenadas Paralelas.

Entre as ferramentas que tratam dados espaço-temporais, pouquíssimas focam na análise de informações sobre o cultivo da cana-de-açúcar ou no NDVI, foco da ferramenta ST-Vis apresentada neste artigo. Um exemplo de ferramenta que trata o NDVI é a SATVeg (Embrapa) (SATVEG, 2014), que é um sistema web para a visualização de séries temporais com o auxílio de um mapa.

A cana-de-açúcar tem uma grande importância na economia brasileira, pois o Brasil é o maior produtor de cana-de-açúcar do mundo. Além da produção de açúcar, se destaca principalmente a produção do etanol, uma importante fonte de substituição de combustíveis fósseis (AGRIC, 2016; MINISTÉRIO DA AGRICULTURA, 2016; NOVACANA, 2016).

Neste trabalho, para analisar as informações do cultivo da cana-de-açúcar são utilizados dados relacionados ao Índice de Vegetação por Diferença Normalizada (NDVI - *Normalized Difference Vegetation Index*). Esse índice é um indicativo do estado da planta, que pode ser aplicado ao monitoramento e mapeamento de culturas, detecção de secas, localização de áreas com pragas, estimação de produtividade e modelagem hidrológica (DUFT, D., 2014).

O NDVI permite fazer análises em diversas escalas sobre a cobertura vegetal de uma determinada região e é obtido por meio da captação de imagens de sensores a bordo de satélites como o AVHRR/NOAA e o MODIS/Terra.

O valor do NDVI varia de -1 a 1, de modo que quanto mais próximo do 1, maior indício da presença de vegetação, e quanto mais próximo do -1, maior indício da presença de solos descobertos, rochas ou água. Teoricamente o valor 0 é referente à uma vegetação sem folhas, submetido a condição de estresse hídrico por déficit de água no solo (INSA, 2014).

O objetivo da ferramenta ST-Vis é visualizar dados espaço-temporais de forma que a informação espacial seja exibida simultaneamente com a temporal, permitindo que a análise

espaço-temporal por parte do usuário seja realizada em uma única visualização interativa e de forma compreensível ao especialista. Para gerar as representações visuais na ST-Vis, foram utilizadas técnicas de visualização espacial e visualização de informação.

Este artigo está descrito da seguinte forma: primeiramente, em materiais e métodos são apresentadas as etapas para o desenvolvimento da ferramenta, detalhando cada visualização utilizada. Na sequência, em resultados e discussões é apresentada uma avaliação da ferramenta, e por fim são apresentadas as principais conclusões deste trabalho.

MATERIAL E MÉTODOS

Para facilitar a visualização e análise dos dados espaço-temporais, foi desenvolvida a ferramenta ST-Vis para a visualização espaço-temporal simultânea e interativa. Proporcionando a representação dos dados através de mapas e gráficos que sejam compreensíveis aos especialistas.

Para gerar a visualização espacial é comum a utilização de um mapa onde os dados espaciais podem ser plotados. Na ST-Vis as informações no mapa interagem com o gráfico de Coordenadas Paralelas, descrito mais adiante, que proporciona uma visualização espaço-temporal de múltiplas dimensões em um plano bidimensional, e o mesmo interage com a Textura Temporal. A Textura Temporal mapeia de forma linear valores numéricos para variações de cores (usadas pelo especialista) ao longo do tempo, gerando uma forma de visualização temporal.

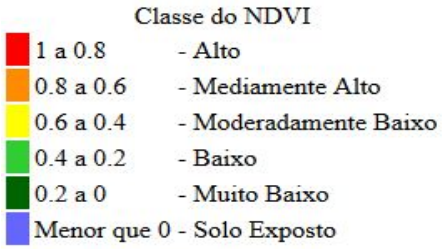
Textura Temporal

A Textura Temporal é uma abordagem para a representação temporal dos dados que variam ao longo do tempo na ferramenta ST-Vis. Essa textura mapeia os dados temporais de uma única latitude e longitude de forma linear, associando os valores numéricos a cores proporcionais à variação da classe do NDVI segundo as cores utilizadas pelos especialistas da área, como apresenta a Figura 1.

Devido a associação dos valores de NDVI com cores, a ST-Vis representa os dados temporais de forma linear, onde cada linha representa o período de um ano e o especialista pode escolher a granularidade dos seus dados, sendo quinzenal ou mensal. Deste modo, ao escolher a granularidade quinzenal, a ST-Vis exibe a Textura Temporal com *n* linhas

contendo 24 colunas, assim cada linha representa o período de um ano, como mostra a Figura 2.

Figura 1 - Variação da classe de NDVI da cana-de-açúcar.



Fonte: Adaptado de (AQUINO, C. M. S.; OLIVEIRA, J. G. B. 2012).

Figura 2 - Textura Temporal com granularidade quinzenal.



O Mapa

A ST-Vis utiliza um mapa para gerar uma visualização espacial, onde marcadores são plotados, ou seja, os pontos verdes no mapa, como pode ser visto na Figura 3. Esses pontos são dados espaciais referentes às variáveis de latitude e longitude dos dados. É possível analisar focos de interesse de uma região, como mostra a Figura 3. O mapa é gerado por meio de uma API do Google Maps.

Figura 3 – Visualização espacial com marcadores plotados.

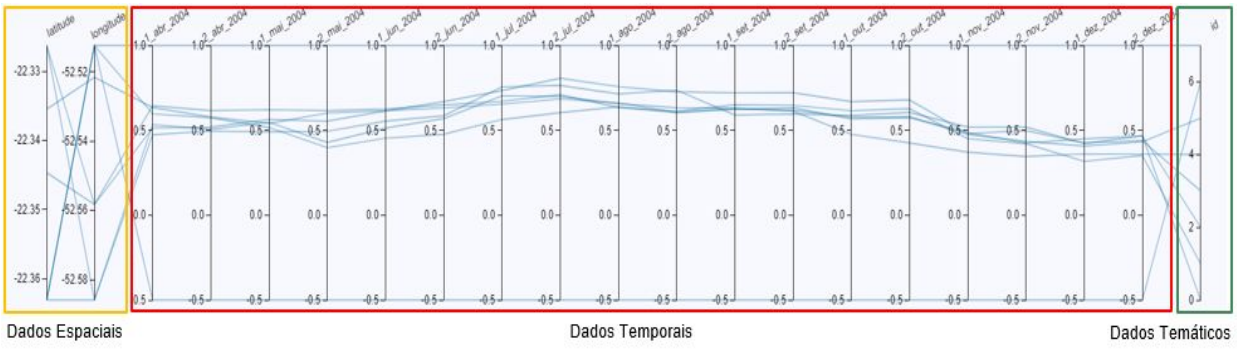


O Gráfico de Coordenadas Paralelas

A técnica de Coordenadas Paralelas, introduzida por Inselberg (1985) consegue representar dados multidimensionais, utilizando linhas verticais (eixos) e horizontais, exibindo suas dimensões de forma paralela umas com as outras em um plano. E permite relacionar as informações dos dados por meio de linhas horizontais (linha azul) que percorrem todos os eixos na representação visual, como mostra a Figura 4.

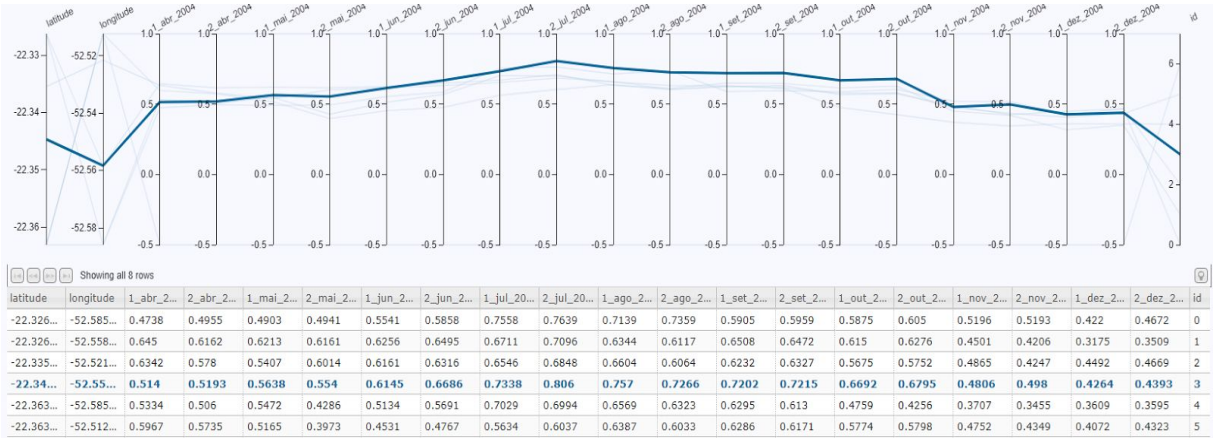
A ST-Vis gera uma visualização espaço-temporal simultânea, associando cada eixo a uma dimensão temporal, incluindo dois eixos verticais para a informação espacial (latitude e longitude), além de adicionar outros eixos para dados temáticos (dados não espaciais e não temporais, exemplo: id, nome, cidade) como mostra a Figura 4. O gráfico de Coordenadas Paralelas tende facilitar a exploração e a compreensão dos dados, e a representar os dados espaciais e temporais de forma simultânea.

Figura 4 – Visualização espaço-temporal gerada a partir do gráfico de Coordenadas Paralelas.



Para melhorar na interatividade da ferramenta ST-Vis, o gráfico está associado a uma tabela que exibe os dados carregados pelo especialista, juntamente com o mapa e a Textura Temporal. Deste modo, ao percorrer a tabela, destaca-se uma linha horizontal no gráfico de Coordenadas Paralelas (Figura 5), o marcador no mapa e a Textura Temporal.

Figura 5 - Destaque da linha horizontal nos eixos paralelos do gráfico.



Showing all 8 rows

latitude	longitude	1_abr_2...	2_abr_2...	1_mai_2...	2_mai_2...	1_jun_2...	2_jun_2...	1_jul_20...	2_jul_20...	1_ago_2...	2_ago_2...	1_set_2...	2_set_2...	1_out_2...	2_out_2...	1_nov_2...	2_nov_2...	1_dez_2...	2_dez_2...	id
-22.326...	-52.585...	0.4738	0.4955	0.4903	0.4941	0.5541	0.5858	0.7558	0.7639	0.7139	0.7359	0.5905	0.5959	0.5875	0.605	0.5196	0.5193	0.422	0.4672	0
-22.326...	-52.558...	0.645	0.6162	0.6213	0.6161	0.6256	0.6495	0.6711	0.7096	0.6344	0.6117	0.6508	0.6472	0.615	0.6276	0.4501	0.4206	0.3175	0.3509	1
-22.335...	-52.521...	0.6342	0.578	0.5407	0.6014	0.6161	0.6316	0.6546	0.6848	0.6604	0.6064	0.6232	0.6327	0.5675	0.5752	0.4865	0.4247	0.4492	0.4669	2
-22.34...	**-52.55...**	**0.514**	**0.5193**	**0.5638**	**0.554**	**0.6145**	**0.6686**	**0.7338**	**0.806**	**0.757**	**0.7266**	**0.7202**	**0.7215**	**0.6692**	**0.6795**	**0.4806**	**0.498**	**0.4264**	**0.4393**	**3**
-22.363...	-52.585...	0.5334	0.506	0.5472	0.4286	0.5134	0.5691	0.7029	0.6994	0.6569	0.6323	0.6295	0.613	0.4759	0.4256	0.3707	0.3455	0.3609	0.3595	4
-22.363...	-52.512...	0.5967	0.5735	0.5165	0.3973	0.4531	0.4767	0.5634	0.6037	0.6387	0.6033	0.6286	0.6171	0.5774	0.5798	0.4752	0.4349	0.4072	0.4323	5

RESULTADOS E DISCUSSÃO

A ST-Vis diferentemente das ferramentas citadas na introdução, gera uma visualização simultânea e interativa dos dados espaço-temporais. A ferramenta também analisa regiões de plantio com o foco em representar por meio da forma de Textura Temporal, as variações do NDVI. A ST-Vis carrega arquivos em formato CSV (*Comma-Separated Values*), exibindo-os simultaneamente no mapa, no gráfico de Coordenadas Paralelas, na forma de tabela e na Textura Temporal. Os dados da tabela interagem com as demais visualizações na ferramenta.

A ST-Vis é focada na análise temporal da variação do NDVI, e ao passar o mouse em uma instância da tabela, alguns eventos são habilitados para facilitar a análise do especialista, como o destaque na linha da tabela com a cor azul (Figura 6d); o destaque da linha horizontal entre os eixos paralelos do gráfico referente a linha selecionada (Figura 6c); o marcador no mapa referente a latitude e longitude da linha destacada salta no mapa (Figura 6a); e por fim, é exibida a Textura Temporal (Figura 6b). A ferramenta também permite a realização de duas consultas simultâneas em diferentes regiões. Podendo o especialista comparar as regiões entre si (Figura 7), relacionando a variação do NDVI de acordo com o foco de interesse.

Os valores dos dados a serem carregados podem ser normalizados (Figura 8) ou não (Figura 6), dependendo de como o especialista de domínio requer a visualização. Existe uma função nos eixos do gráfico de Coordenadas Paralelas que o especialista pode selecionar alguns focos de interesse, como mostra o quadrado vermelho na Figura 8. Essa seleção permite a filtragem de instâncias na visualização facilitando assim a análise dos dados. De acordo com os especialistas, a ferramenta apresenta uma visualização diferenciada da informação. Há ganhos na execução da tarefa pois permite a visão das séries tanto no mapa

(localização/espaço) quanto pelas Coordenadas Paralelas (espaço-temporal). No entanto, novas validações serão feitas com um número maior de especialistas a fim de aprimorar a ferramenta.

Figura 6 (a) - Mapa; (b) - Textura Temporal; (c) - Gráfico de Coordenadas Paralelas; (d) – Tabela.

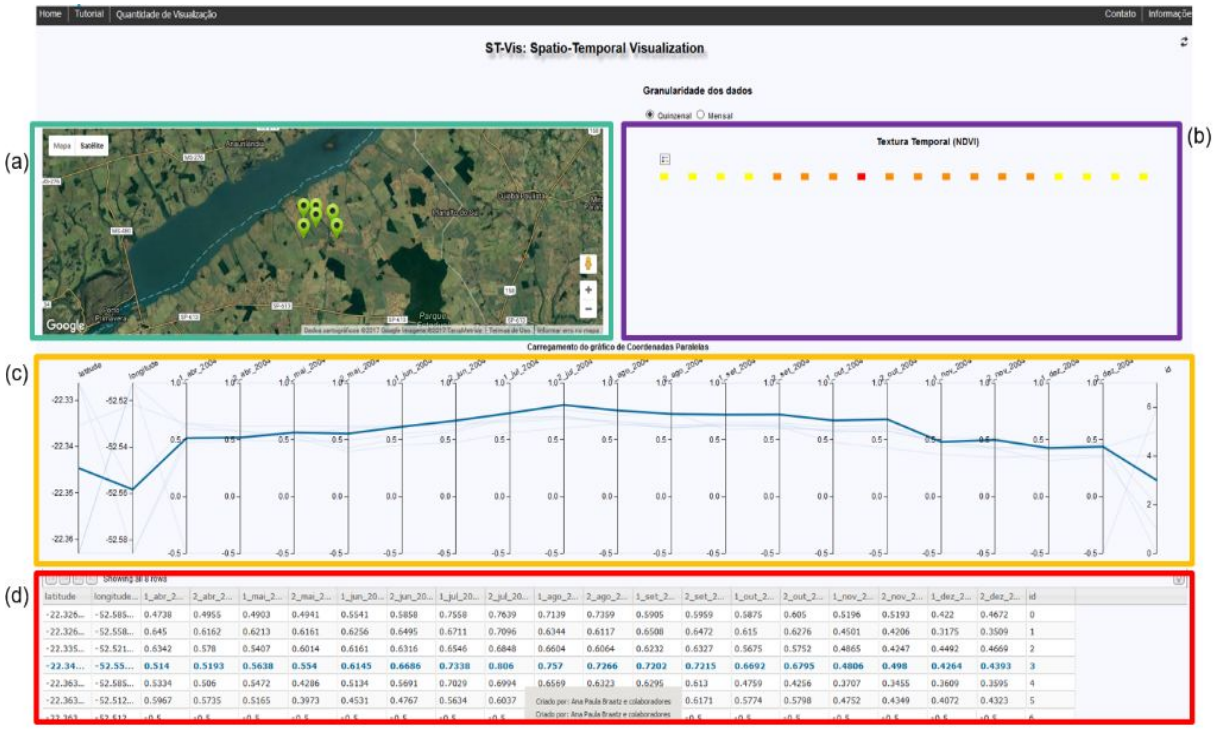


Figura 7 - Realização de consultas simultâneas

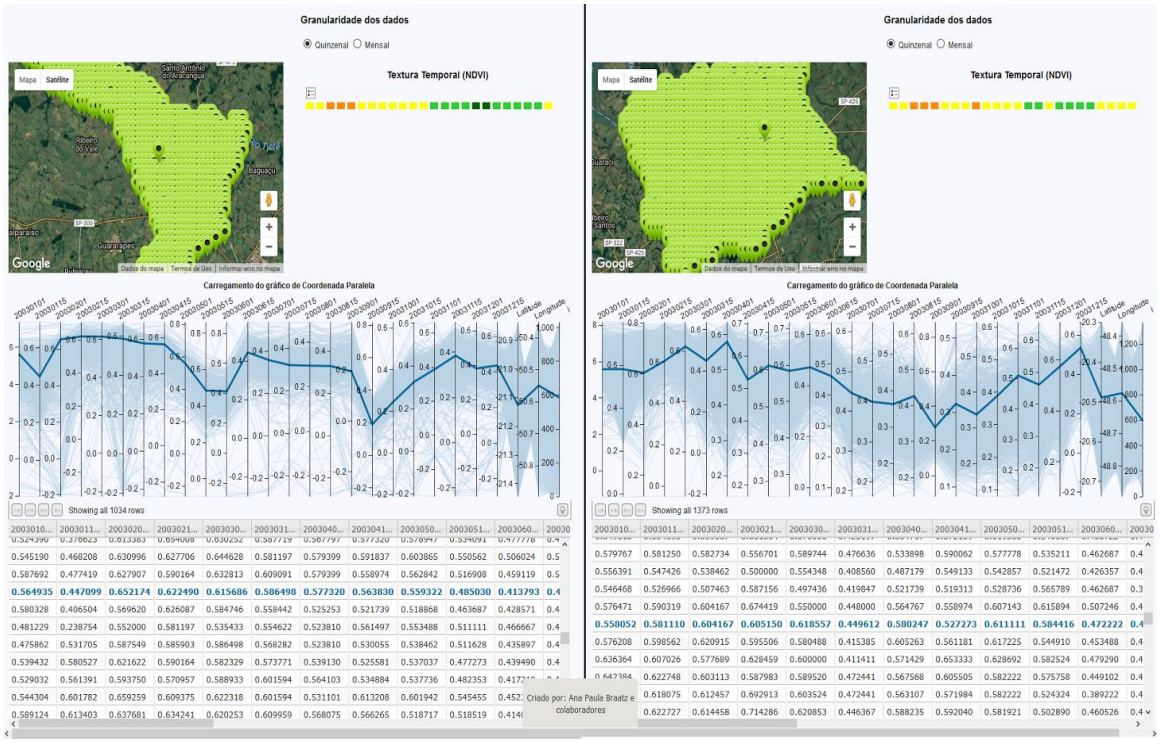
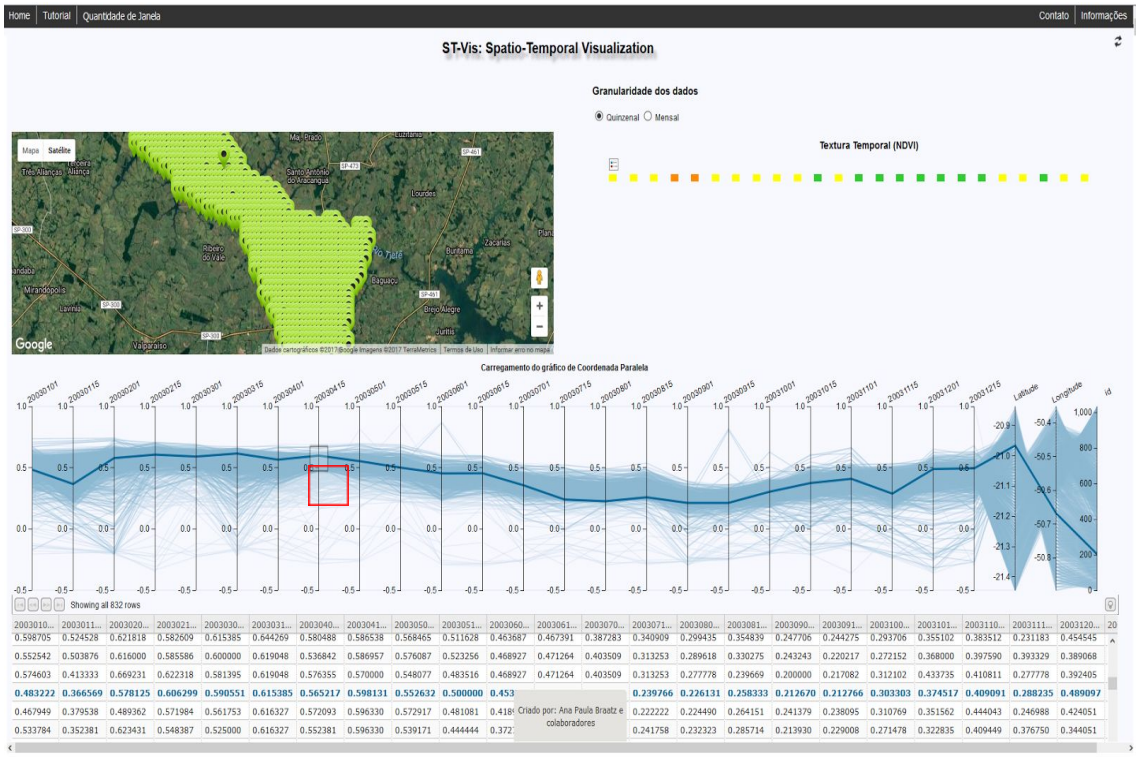


Figura 8 - Visualização com dados normalizados.



CONCLUSÕES

A visualização de dados espaço-temporais de maneira simultânea ainda é um desafio na área de visualização devido à difícil tarefa de adicionar a variação temporal a dados georeferênciados. Neste sentido, este artigo apresentou a ferramenta ST-Vis que fornece uma representação visual de dados textuais e tabulares, combinando o mapa com um gráfico de Coordenadas Paralelas e a Textura Temporal. Os resultados desta visualização auxiliam os especialistas de domínio na compreensão das informações espaço-temporais, permitindo uma visualização espaço-temporal interativa. Assim, o especialista consegue identificar algumas regras e padrões de interesse em áreas de plantio da cana-de-açúcar compreendendo as informações de forma mais intuitiva. De modo que possa visualizar a variação de uma região ao longo do tempo a partir dos valores do NDVI, sendo possível analisar os resultados deste índice para todas as aplicações que utilizam essa informação, aprimorando o monitoramento das culturas agrícolas, o que leva há uma redução das perdas no campo.

REFERÊNCIAS

AGRIC. Produção de Cana-de-açúcar. [2016?]. Disponível em:< http://www.agric.com.br/producoes/cultivo_da_cana.html>. Acesso: 10 Maio, 2017.

AQUINO, C. M. S.; OLIVEIRA, J. G. B. Estudo da dinâmica do índice de vegetação por diferença normalizada (NDVI) no núcleo de São Raimundo Nonato-PI. **GEOUSP: Espaço e Tempo (Online)**, São Paulo, n. 31, p. 157-168, 2012. ISSN 2179-0892. Disponível em: <http://www.revistas.usp.br/geousp/article/view/74261/77904>. Acesso em: 28 Mar. 2017.

CASANOVA, A. M. et al. **Banco de dados geográficos**. Curitiba: Mundogeo. 2005.

DE OLIVEIRA, M. G.; DE SOUZA B. C. GeoSTAT - A system for visualization, analysis and clustering of distributed spatiotemporal data. **GEOINFO**. 2012. p. 108-119.

DUFT, D. O que significa NDVI e qual a sua relação com agricultura. InteliAgro. 2014. Disponível em: < http://inteliagro.com.br/o-que-significa-ndvi-e-o-qual-sua-relacao-com-agricultura/>. Acesso em: 11 Maio, 2017.

GUO, D. et al. A Visualization System for Space-Time and Multivariate Patterns (VIS-STAMP). **IEEE Transactions on Visualization and Computer Graphics**, v. 12, n. 6, p. 1461-1474, 2006.

INFOSCOPE, Versão 3.6.1. Macrofocus GMBH, 2015. Disponível em: <https://www.macrofocus.com/public/products/infoscope/>. Acesso em: 20 Abr. 2017.

INSELBERG, A. The plane with parallel coordinates. **The Visual Computer**, v. 1, n. 2, p. 69-91, 1985.

INSTITUTO NACIONAL DO SEMIÁRIDO - INSA. **Índice de Vegetação por Diferença Normalizada.** 2014. Disponível em: <http://www.insa.gov.br/ndvi/#.V6fKRTWWmTA>. Acesso em: 5 Maio, 2017.

MINISTÉRIO DA AGRICULTURA. Cana-de-açúcar. [2016?]. Disponível em:< http://www.agricultura.gov.br/vegetal/culturas/cana-de-acucar>. Acesso: 10 Maio, 2016.

NOVACANA. A produção de cana-de-açúcar no brasil (e no mundo). [2016?]. Disponível em:< https://www.novacana.com/cana-de-acucar/producao-cana-de-acucar-brasil-e-mundo/>. Acesso: 10 Maio, 2016.

ROTH, R. E. et al. A user-centered approach for designing and developing spatiotemporal crime analysis tools. **Proceedings of GIScience**. 2010. Disponível em: < https://www.geovista.psu.edu/publications/2010/RothEtAl_2010_AUserCenteredApproach.pdf>. Acesso em: 2 Maio 2017.

SATVEG. Embrapa. [2014?]. Disponível em: <https://www.satveg.cnptia.embrapa.br/satveg/login.html>. Acesso em: 15 Abr. 2017.

Identificação de falhas no plantio de soja através da utilização de veículos aéreos não-tripulados (VANTs)

Renan Cunha[1], *Luan Casagrande*[1], *Yuri Crotti*[1], *Rodrigo Maciel*[1], *Roderval Marcelino*[2]

[1]Laboratório de Pesquisa Aplicada, Universidade Federal de Santa Catarina
Araranguá, Santa Catarina, Brasil
(renan.cs, luan.casagrande, yuri.crotti, rodrigo.maciel)@grad.ufsc.br

[2]Laboratório de Pesquisa Aplicada, Universidade Federal de Santa Catarina - DEC
roderval.marcelino@ufsc.br

RESUMO

A utilização de VANTs (Veículos Aéreos Não-Tripulados) na área de sensoriamento remoto voltado para a agricultura, já é uma realidade tanto no Brasil como na maior parte dos outros países. Os VANTs podem oferecer aos gerentes de recursos uma rota promissora para um monitoramento responsivo, oportuno e econômico em diferentes escalas espaciais e temporais. Considerando que a soja tem um papel de destaque na produção agrícola brasileira, este artigo propõe a aplicação de técnicas de processamento de imagens e reconhecimento de padrões em imagens obtidas através de VANTs, para fazer a identificação automática de falhas em plantações de soja. O objetivo final do software proposto foi mapear a centroide (latitude e longitude) de cada falha encontrada e mensurar o tamanho da mesma em m^2. Devido à ausência de dados de validação, obtidos em solo, para se fazer uma comparação com os dados de saída do software, a validação do sistema foi feita de forma visual. Através da inspeção visual dos resultados, foi possível verificar que o sistema foi capaz de atingir os objetivos propostos em geral. É importante destacar o uso de processamento em blocos na solução proposta, o que viabiliza a paralelização do problema aqui tratado.

PALAVRAS-CHAVE: Processamento de imagem, Sensoriamento remoto, Agricultura de precisão.

ABSTRACT

The use of UAVs (Unmanned Aerial Vehicles) in the remote sensing area applied to the agriculture is already a reality in Brazil and in almost all countries in the world. UAVs can offer to the resource managers a promising route for responsive, timely and cost-effective monitoring at different spatial and temporal scales. Considering that soybean crop plays an important role

in Brazil's agricultural production, this article proposes the application of image processing and pattern recognition techniques in imagery acquired by UAVs in order to automatically identify faults in Soybean plantations. The final objective of the proposed software was to map the centroid (latitude and longitude) of each fault identified and measure its size in m^2. Due to the absence of validation data obtained in soil to make a comparison with the software's output, the system validation was done visually. Through this visual inspection of the results, it was possible to verify that the system was able to reach the objectives proposed in general. It is important to highlight the use of block processing in the proposed solution, which makes possible the parallelization of the solution proposed in this paper.

KEYWORDS: Image processing, Remote sensing, Precision agriculture.

INTRODUÇÃO

De acordo com Pederi e Cheporniuk (2015), há pouco tempo atrás a população mundial atingiu a marca dos sete bilhões de pessoas. Esse número tende a crescer e estima-se que em 2050 a população mundial seja de aproximadamente 10 bilhões de pessoas. Este crescimento traz a tona diversas preocupações, dentre elas, as relacionadas a produção de alimentos para suprir as necessidades de todos.

Em Busemeyer et al. (2013) e em Navia et al. (2016), os autores destacam que esta demanda por alimentos cada vez maior se tornará um grande desafio para os agricultores, agrônomos e cientistas em geral, envolvidos na produção dos mesmos. Fica evidente a necessidade do uso de meios visando otimizar os métodos atuais de manejo afim de se obter uma maximização na produção e minimização nas eventuais percas que possam ocorrer durante o plantio.

A agricultura de precisão surgiu justamente para auxiliar na otimização da produção. O uso de informações diversas, obtidas através das tecnologias dispostas no campo, podem auxiliar significativamente o produtor nas tomadas de decisões durante uma safra (GEORGE et al., 2013).

Os recentes avanços tecnológicos tornaram os VANTs (Veículos Aéreos Não-Tripulados) em um dos meios de obtenção de informações mais eficientes na área da agricultura. Através de câmeras acopladas a estas aeronaves, é possível fazer um imageamento completo de uma área plantada de forma rápida, precisa, sem riscos e com custos relativamente baixos (BUSEMEYER et al., 2013). No Brasil esta tecnologia está sendo implantada aos poucos e pretende ser uma ferramenta aliada a produção das diversas culturas aqui existentes.

Conforme é discutido em Hirakuri (2015), a soja é a cultura agrícola mais relevante para o agronegócio nacional, sendo fundamental para o desenvolvimento socioeconômico de vários polos produtores e também por ser uma cultura que serve como base para uma infinidade de outros produtos alimentícios. Tendo isto em vista, é de suma importância que seja feita cada vez mais a integração de tecnologias como por exemplo, o uso de VANTs no cultivo de soja.

Durante todo o ciclo de cultivo da cultura de soja, a mesma encontra-se suscetível a ataques de diferentes pragas, capazes de causarem perdas significativas a produção (EMBRAPA, 2003). Identificar estes ataques para que ações corretivas sejam feitas é uma tarefa complicada, principalmente em grandes plantações. Uma solução que pode vir a sanar este problema relacionado ao acompanhamento constante de uma área de cultivo, é o uso de sensoriamento remoto através de imagens aéreas captadas por VANTs.

Porém, apenas captar as imagens a partir de VANTs não constitui um acompanhamento eficaz. Visando prover ferramentas de agricultura de precisão que possam contribuir no processo de cultivo de soja, este trabalho propõe o desenvolvimento de um *software* que, a partir das imagens coletadas por VANTs e do uso de técnicas de processamento de imagens, seja capaz de identificar, de forma autônoma, falhas no plantio de soja.

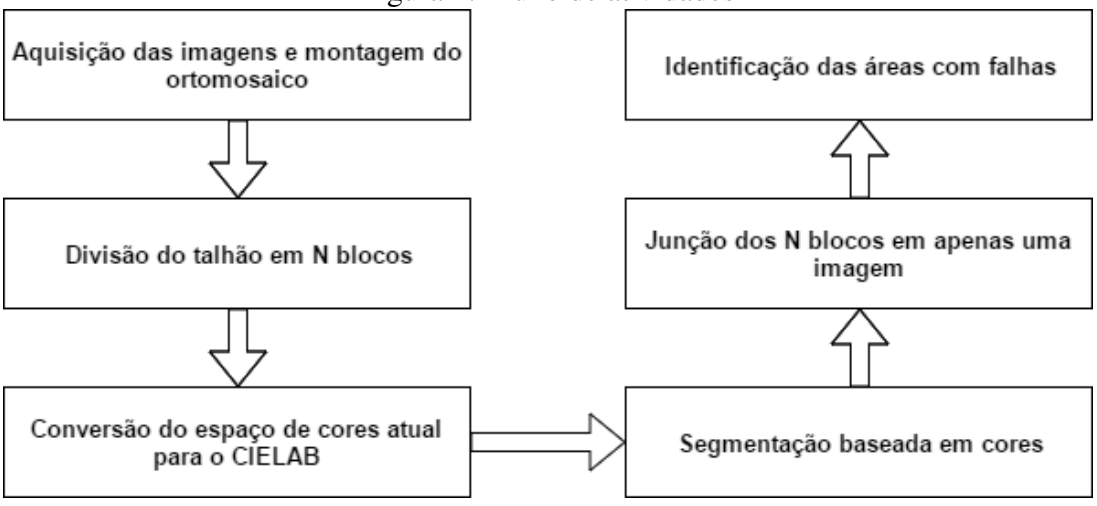
MATERIAL E MÉTODOS

Nesta seção serão abordados os materiais e a metodologia que foram aplicados no desenvolvimento da ferramenta proposta. Resumidamente, o fluxo de atividades é dado de tal maneira que primeiramente é feita a aquisição das imagens através de VANTs, em seguida é montado o ortomosaico, que nada mais é do que a junção das diversas imagens captadas pela aeronave em apenas uma imagem. A partir do ortomosaico é feito uma série de processamentos visando identificar as áreas com falhas.

A Figura 1 mostra um diagrama que representa a ordem de execução de cada tarefa.

É importante destacar que, os algoritmos utilizados em todas as etapas elucidadas na Figura 1, foram implementados pelos presentes autores através o software MATLAB. No caso do algoritmo de segmentação k-means, utilizou-se uma implementação já existente na *toolbox* "*Statistics and Machine Learning*".

Figura 1: Fluxo de atividades



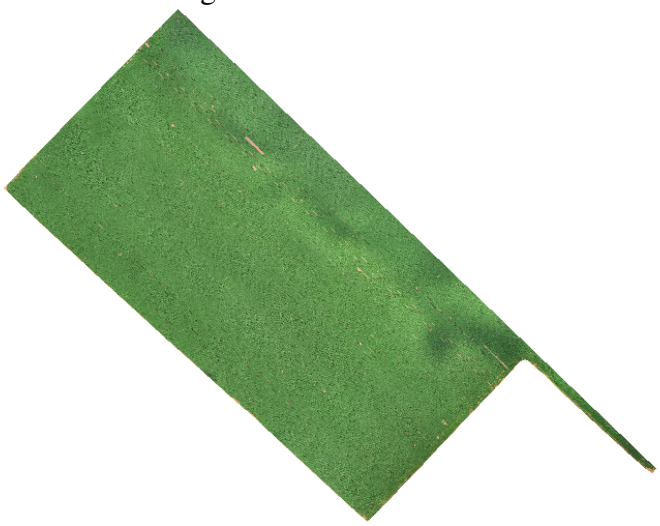
Fonte (próprios autores)

Aquisição das imagens e ortomosaico

Os autores deste trabalho não realizaram os voos para obtenção das imagens da plantação de soja aqui analisada. Como o foco da proposta é a aplicação de técnicas de processamento de imagens visando identificar falhas, optou-se por utilizar um mosaico já pronto, fornecido por terceiros e que atendiam as especificações técnicas necessárias. A soja do talhão analisado já se encontrava em um estágio avançado no plantio, estando disposta de modo consideravelmente denso.

O ortomosaico utilizado estava no formato GeoTIFF, com GSD de 6,19cm/pixel e cobria uma área total de aproximadamente 760ha. Visando facilitar a demonstração dos resultados, neste trabalho será utilizado apenas o Talhão 1 da plantação analisada, que está disposto na Figura 2 e contém uma área próxima de 90 hectares.

Figura 2: Talhão analisado



Fonte (próprios autores)

Divisão do talhão em N blocos

Haja vista que geralmente as imagens a serem analisadas são muito grandes, é comum se fazer o uso da técnica chamada de processamento em blocos. Basicamente a imagem analisada é dividida em N blocos de modo que cada bloco é processado individualmente, evitando assim um consumo exacerbado de recursos físicos da máquina que está realizando o processamento da imagem. Além disso, é importante ressaltar que o processamento em bloco possibilita o uso de processamento paralelo, o que reduz significativamente o tempo computacional necessário para a extração dos resultados (HEARTZ, 1989).

No presente trabalho optou-se por dividir o talhão analisado de 22000x28000 pixels em 5 blocos de 4400x5600 pixels. Esta escolha foi arbitrária e definida pelos autores do estudo.

Conversão do espaço de cores atual para o CIELAB

Visando facilitar o processo de segmentação da imagem que será descrito na próxima sub-seção, é feito nesta etapa uma conversão do espaço de cores atual para o espaço de cores CIE L*a*b*. De acordo com TAKATSUI (2011), o sistema de cores CIE L*a*b* é constituído por três parâmetros L*, a* e b*, conforme descrito abaixo:

- L*: referente à luminosidade do objeto a ser avaliado, podendo variar do preto ao branco;
- a*: medida do croma no eixo vermelho-verde;
- b*: medida do croma no eixo amarelo-azul.

Para se fazer esta conversão da espaços, primeiro transforma-se a imagem do espaço RGB para XYZ, vide equação 1.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} * \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1}$$

Após, converte-se de XYZ para Lab (Equação 2).

$$\begin{aligned} L &= 116 * \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16 \\ a &= 500 * \left(\left(\frac{X}{X_n}\right)^{\frac{1}{3}} - \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}}\right) \\ b &= 200 * \left(\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - \left(\frac{Z}{Z_n}\right)^{\frac{1}{3}}\right) \end{aligned} \tag{2}$$

Onde X_n, Y_n e Z_n são os valores de X, Y e Z para o ponto branco de referência. Neste trabalho seguiu-se a recomendação BT.709 por ITU-R, onde [X_n Y_n Z_n] = [0.95045 1 1.088754]. Após, remove-se as bandas ”a”e ”b”e reestrutura-se a matriz obtida em L para um vetor que será utilizado na próxima etapa do processo que é a segmentação baseada em cores.

Segmentação baseada em cores

A partir do vetor resultante obtido na sub-seção anterior é utilizada uma técnica de *Clustering* visando agrupar áreas similares encontradas na imagem obtida através dos VANTs. No presente trabalho foi utilizado o algoritmo *k-means clustering*, que é um algoritmo de aprendizado não-supervisionado muito utilizado na área de *data mining*.

Nesta etapa é feita a isolação dos diferentes objetos na imagem. Um ponto muito relevante a se destacar é que a técnica de segmentação utilizando este método de *clustering* independe de limiares, ou seja, não faz-se necessário a definição de parâmetros de coloração definidos pelo usuário, todo o processo de descoberta dos objetos na imagem é feito de modo automático e não-supervisionado.

Neste caso utilizou-se o k do k-means igual a 2. Este parâmetro indica a quantidade de *clusters* que serão gerados. Como uma plantação de soja geralmente tem uma densidade foliar

alta, fica isolado em um dos *clusters* toda a área relativa a plantação, ou seja, a área verde. A partir deste *cluster*, é montado uma imagem binária de tal modo que toda a área verde isolada no *cluster* é transformado em branco na imagem binária e todo o resto continua preto. Essa imagem binária é o resultado final desta etapa.

No presente trabalho os autores optaram por utilizar Squared Euclidean distance (Equação 3) para divisão dos *clusters* no algoritmo *k-means clustering*.

$$d(x, c) = (x - c)(x - c)' \tag{3}$$

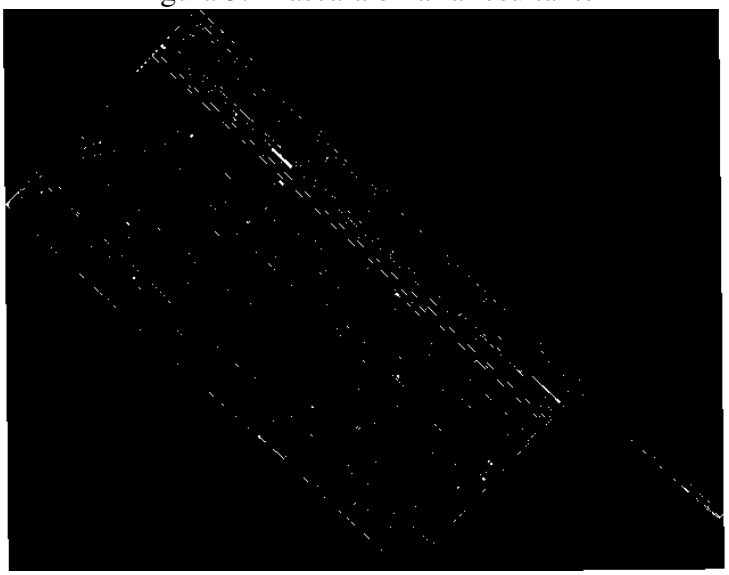
Onde x é o vetor de dados a ser analisado e c é a centroide desse vetor.

Junção dos N blocos em apenas uma imagem

Conforme foi descrito anteriormente, a conversão do espaço de cores e a segmentação baseada em cores foram etapas cujo processamento foi realizado em blocos. Porém, para se fazer a identificação das falhas é preciso analisar a imagem completa da plantação, uma vez que uma falha contínua pode estar contida em dois blocos distintos, o que exigiria um controle muito mais complexo na hora de fazer a localização das irregularidades no plantio.

Nesta etapa todos os blocos que foram criados na etapa de divisão do ortomosaico, convertidos para o espaço de cores CIE Lab e segmentados através do *k-means clustering* são novamente justapostos de tal modo que é formada novamente uma nova imagem binaria com a mesma dimensão do ortomosaico original, demonstrada na Figura 3.

Figura 3: Máscara binária resultante



Fonte (próprios autores)

Identificação das áreas com falhas

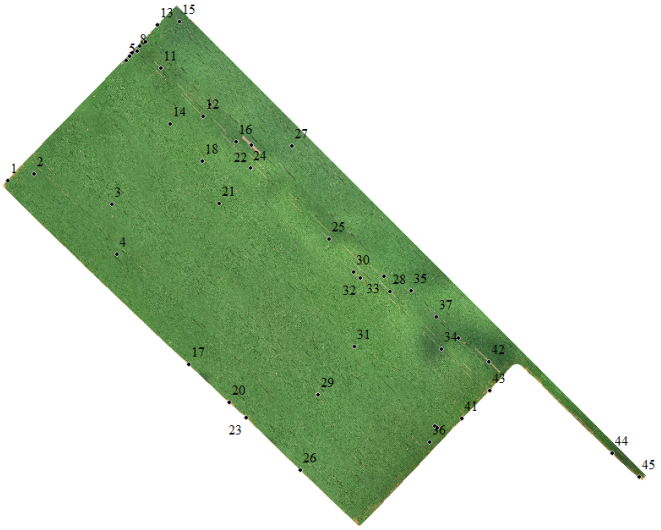
Na etapa anterior foi montada uma imagem binaria final onde tudo que não fazia parte da plantação, ou seja, tudo que não seguia o padrão de tonalidade da soja foi transformado em branco e todas as áreas verdes foram transformadas em preto.

A partir da imagem binária, as vezes também chamada de máscara binária, é feita a localização efetiva das áreas consideradas como falhas no meio da plantação. Utilizando uma técnica de análise de vizinhança é feito operações sobre a máscara binária afim de se agrupar os *pixels* brancos destacados na Figura 3. Cada um dos grupos criados pode ser considerado como uma falha na plantação, sendo que através do GSD da imagem é possível inferir qual o tamanho (área em m^2) de cada um destes grupos (falhas). Neste trabalho foi considerado como falhas as áreas superiores a $2m^2$.

RESULTADOS E DISCUSSÃO

Como resultado final o *software* proposto acaba gerando para o usuário um arquivo do tipo *shapefile* contendo as centroides de todas as falhas encontradas na plantação, conforme pode-se ver na Figura 4.

Figura 4: *Shapefile* indicando localização das falhas



Fonte (próprios autores)

Como esta *shapefile* possui informações de georreferencia, é possível identificar as posições (latitude e longitude) de cada centro de falha.

Além da *shapefile*, também foi possível gerar uma planilha contendo todas as áreas identificadas como falhas, sua área em m^2 e a posição (latitude e longitude) da centroide da mesma. Esta planilha serviria de apoio ao profissional responsável por analisar as irregularidades no plantio, sugerindo as posições das áreas onde provavelmente há uma falha.

CONCLUSÕES

A partir do desenvolvimento deste trabalho ficou evidente o grande potencial do uso de processamento digital de imagens em análises voltadas a agricultura, mais especificamente na cultura de soja. Um dos pontos mais relevantes observados na etapa de processamento foi a eficiência do método de clusterização escolhido *k-means clustering* para isolar as áreas plantadas. Além disso, é de suma importância destacar a utilização do processamento em blocos, que é uma técnica que torna viável o uso de computadores convencionais para se fazer o processamento das imagens, uma vez que os blocos são imagens das quais máquinas comuns são capazes de interpretar e rodar os algoritmos necessários.

Apesar de não haver um mapeamento feito a campo de forma manual por um profissional qualificado, para se fazer a validação estatística do sistema e medir a acurácia do mesmo, os autores avaliaram visualmente que os resultados obtidos atenderam de forma satisfatória o objetivo principal que era identificar as falhas no plantio.

REFERÊNCIAS

BUSEMEYER, L. et al. BreedVision — a multi-sensor platform for non-destructive field-based phenotyping in plant breeding. *Sensors*, MDPI AG, v. 13, n. 3, p. 2830–2847, feb 2013. Disponível em: <https://doi.org/10.3390

EMBRAPA. *Tecnologias de produção de soja região central do Brasil*. 2003. Disponível em: <http://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Soja/SojaCentralBrasil2003/importancia.htm>.

GEORGE, E. A. et al. Uav systems for parameter identification in agriculture. In: *2013 IEEE Global Humanitarian Technology Conference: South Asia Satellite (GHTC-SAS)*. [S.l.: s.n.], 2013. p. 270–273.

HEARTZ, R. *Data block processing for fast image generation*. Google Patents, ago. 8 1989. US Patent 4,855,937. Disponível em: <https://www.google.com/patents/US4855937>.

HIRAKURI, A. A. B. J. M. H. *Análise econômica de diferentes densidades de semeadura na cultura da soja*. [S.l.], 2015.

NAVIA, J. et al. Multispectral mapping in agriculture: Terrain mosaic using an autonomous quadcopter uav. In: *2016 International Conference on Unmanned Aircraft Systems (ICUAS)*. [S.l.: s.n.], 2016. p. 1351–1358.

PEDERI, Y. A.; CHEPORNIUK, H. S. Unmanned aerial vehicles and new technological methods of monitoring and crop protection in precision agriculture. In: *2015 IEEE International Conference Actual Problems of Unmanned Aerial Vehicles Developments (APUAVD)*. [S.l.: s.n.], 2015. p. 298–301.

TAKATSUI, F. *SISTEMA CIE LAB: ANÁLISE COMPUTACIONAL DE FOTOGRAFIAS*. 2011.

Sistema de sugestão de densidade para plantações de banana utilizando veículos aéreos não tripulados

Luan Casagrande[1], *Yuri Crotti*[1], *Renan Cunha*[1], *Roderval Marcelino*[2], *Rodrigo Maciel*[1]

[1]Laboratório de Pesquisa Aplicada, Universidade Federal de Santa Catarina
Araranguá, Santa Catarina, Brasil
(luan.casagrande,yuri.crotti,renan.cs, rodrigo.maciel)@grad.ufsc.br

[2]Laboratório de Pesquisa Aplicada, Universidade Federal de Santa Catarina - DEC
roderval.marcelino@ufsc.br

RESUMO

Este artigo apresenta um sistema para analisar imagens de alta resolução obtidas através de veículos aéreos não tripulados (VANTs) para geração de mapas sugestivos de densidade em plantações de banana. O principal objetivo do sistema é auxiliar o agricultor no gerenciamento e planejamento da população de plantas, visando manter a densidade de plantação ideal. A solução aqui proposta pode ser dividida em duas etapas distintas, sendo: Análise de textura através da técnica *segmentation based fractal texture analysis* (SFTA) e classificação dos blocos através de *support vector machine* (SVM). para validar a solução proposta, uma aplicação foi feita em uma plantação de bananas na região de Turvo - SC. Através de uma análise visual, foi possível verificar que o resultado obtido na aplicação real representou de forma adequada a realidade. Entende-se que o sistema proposto ainda demanda novas validações em solo, porém foi possível perceber que o algoritmo possui potencial, pois atendeu os objetivos do trabalho.
PALAVRAS-CHAVE: Processamento de imagem, Aprendizagem de Máquina, Agricultura de Precisão.

ABSTRACT

This paper presents a system that aims to analyze high-resolution images acquired by unmanned aerial vehicles (UAVs) for the generation of density suggestive maps for banana plantations. The main system's objective is to assist the farmer in the management and planning of the population, in order to maintain optimal planting density. The solution proposed can be divided into two steps, being: texture analysis through segmentation based fractal texture analysis (SFTA) and block classification through support vector (SVM). In order to validate the proposed solution, an application was made in a plantation in Turvo - SC. Through a visual evaluation, it is

possible to verify that the result obtained in the real application adequately represented the reality. It is known that the proposed system demands new validations in soil, but it is visible that the algorithm has potential since the main objectives of this work were complete successfully.
KEYWORDS: Imagtioe Processing, Machine Learning, Precision agriculture.

INTRODUÇÃO

A banana, atualmente, é uma das frutas mais produzidas e consumidas no mundo. Em 2011, 107 milhões de toneladas métricas foram produzidas em mais de 130 países, o que representa aproximadamente 0,1% de toda a área agrícola do mundo (FAO, 2013). Ainda segundo FAO (2013), toda esta produção chegou a um valor comercial de 9 biliões de dólares. O Brasil encontra-se em uma posição de destaque na produção de bananas. Segundo Potts et al. (2014), em 2011 o Brasil estava entre os cinco maiores produtores com 7% da produção mundial. A expectativa de produção no Brasil para 2017 é de aproximadamente 6.984.637 toneladas com um rendimento médio de 14.364 kg/ha (IBGE, 2017).

O espaçamento utilizado para estruturar o bananal influência diretamente no ciclo vegetativo, e por consequência, no rendimento médio da área. Segundo Harry e Willison (1987) e Borges e Souza (2004), a rentabilidade de um bananal tende a aumentar na mesma proporção da densidade do plantio até determinado ponto. Outro ponto observado pelos autores é que a maioria dos plantios comerciais se desenvolve abaixo da densidade recomendada (HARRY; WILLISON, 1987; BORGES; SOUZA, 2004), reduzindo então o potencial da área plantada. Por conta da criticidade deste aspecto, o espaçamento ideal e suas consequências estão sendo estudadas por diversos autores, como em: (BORGES; SOUZA, 2004; ZONETTI et al., 2002; VANHOUDT, 2009), dentre outros.

Outro aspecto importante, é o controle da população da plantação através do desbaste. Segundo Borges e Souza (2004), o desbaste é essencial por diversos fatores, como: Manter o número de plantas por hectare, manter o padrão e o tamanho do cacho, incrementar os ciclos produtivos, regular produção e momento de colheita, aumentar rendimento, dentre outros. Ou seja, existe uma clara necessidade de se planejar e controlar a população de uma plantação de bananas visando o melhor aproveitamento da área.

Porém, na medida em que as propriedades crescem de tamanho, esse detalhamento necessário de densidade de plantio é reduzido, pois geralmente faz-se uma amostragem e o resultado de uma amostra é estabelecida para uma área considerável ou, às vezes, até mesmo para um talhão inteiro. Consequentemente, a tomada de decisão na plantação, como a necessidade de desbaste para ajuste da densidade, fica substancialmente prejudicada.

Considerando-se tal problemática, este trabalho propõe uma nova abordagem para geração de mapas de densidade através de técnicas de processamento de imagem e aprendizagem de máquina em imagens de alta resolução obtidas por Veículos Aéreos Não Tripulados (VANT). Tal solução visa possibilitar ao responsável pela área um maior nível de detalhamento da den-

sidade da plantação em intervalos de tempo flexíveis. Visando a validação visual do sistema proposto, uma aplicação foi realizada em Turvo - SC.

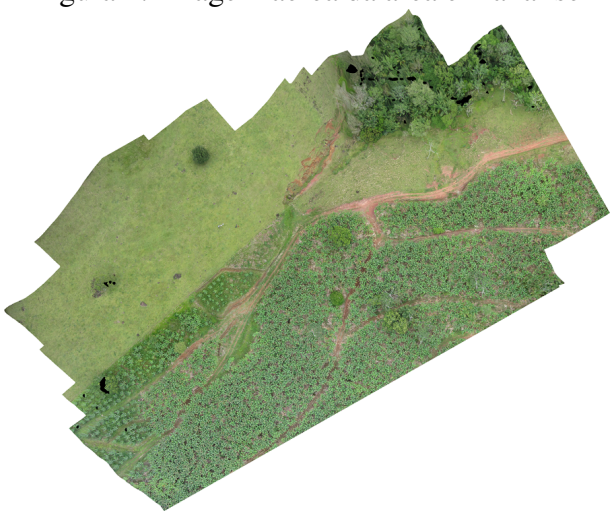
A primeira seção irá descrever o VANT e a metodologia utilizada para aquisição das imagens e geração do mosaico. Em seguida irá se especificar o algoritmo desenvolvido para geração de mapas de densidade. Após, irá se detalhar e discutir os resultados obtidos na plantação em análise. Então, irá se concluir o trabalho e apresentar futuras implementações.

VANT e AQUISIÇÃO DE IMAGEM

Para aquisição das imagens da área proposta, utilizou-se um hexacoptero desenvolvido pela empresa SkyDrones chamado de microVANT SD6 Spyder XL. O VANT possui capacidade de carga de até 1650g que deve ser dividida entre o peso da bateria e da câmera embarcada. Considerando uma bateria 4 células com 14,8V, o VANT consegue atingir uma autonomia de até 15 minutos.

A câmera utilizada para aquisição das imagens foi uma Canon PowerShot ELPH 300HS. O *sensor width* da câmera é de aproximadamente 6,16 mm e o *focal length* definido foi de 4,3 mm. O tamanho das imagens obtidas foi 4000 x 3000 pixeis. Considerando-se as especificidades da câmera, obteve-se uma resolução espacial de 5cm/pixel a 140 metros de altura, conforme demonstrado na Figura 1.

Figura 1: Imagem aérea da área em análise



Fonte: Autores

Conforme demonstrado nos trabalhos (SILVA et al., 2014; JúNIOR et al., 2015; HENRIQUES; MANTA; MARNOTO, 2015), os melhores resultados ocorrem onde a sobreposição é de cinco imagens ou mais. Consequentemente, utilizou-se uma sobreposição lateral de 70% e longitudinal de 60%.

O software escolhido para geração do mosaico foi o Agisoft Photoscan Pro 2015. O VANT armazena a latitude, longitude e altitude de cada imagem tirada e esta informação foi

posteriormente carregada no software para definir as posições iniciais das imagens. Após gerou-se o mosaico (Figura 1), considerando a configuração de "alta" qualidade no Photoscan Pro. O georreferenciamento foi realizado usando apenas os dados obtidos através dos dados de telemetria do voo. Pontos de controle em solo e correções pós voo não foram usadas. Neste trabalho, utilizou-se apenas as três bandas espectrais (vermelho, verde e azul).

SISTEMA PARA GERAÇÃO DE MAPAS DE DENSIDADE EM PLANTAÇÕES DE BANANA

Antes de se iniciar o sistema, deve-se definir alguns parâmetros que serão utilizados durante o processamento. O primeiro parâmetro é a área em análise para delimitar a ação do sistema. O segundo parâmetro representa os blocos que serão utilizados para treinamento e teste da máquina. Este parâmetro pode ser identificado visualmente ou através de coordenadas geográficas. Faz-se importante destacar que o tamanho dos blocos deve ser padronizado. Outro fator essencial é a necessidade de se definir 2 classes distintas, sendo: não densa e densa. Tal fator é necessário, pois o classificador utilizado é um classificador binário fundamentado em uma função de base radial, e consequentemente, a definição das classes nas extremidades permite melhor aproveitamento do sistema por definir o espaçamento máximo entre as classes. Na sequencia, o algoritmo proposto inicia o processo que pode ser dividido em duas etapas distintas: extração de características fundamentadas na textura e classificação dos blocos. A Figura 2 representa parte dos blocos utilizados para treinamento e teste do sistema. Para esta imagem em análise, identificou-se visualmente 104 blocos.

Figura 2: (A) Classe não densa, (B) Classe densa



Fonte: Autores

Extração de Características

Um método eficaz de representar uma região em forma de imagem é quantificar seu conteúdo de textura. Este descritor desempenha um papel essencial no processamento de imagens, porque muitas propriedades podem ser medidas por ele. Considerando que neste trabalho foi utilizado

apenas o espectro visível, os recursos extraídos da textura foram utilizados para compensar a falta de riqueza nos dados coletados. Neste caso, considerou-se especificamente a técnica SFTA, descrita por (COSTA; HUMPIRE-MAMANI; TRAINA, 2012).

SFTA pode ser descrita através de dois métodos. No primeiro, a imagem modelo é decomposta em imagens binárias através da técnica *two-Threshold Binary Decomposition* (TTBD) (COSTA; HUMPIRE-MAMANI; TRAINA, 2012). Os conjuntos de limiares (T) que minimizam a variância entre as classes da imagem são calculados através do algoritmo de Otus em vários níveis (LIAO; CHEN; CHUNG, 2001). As imagens binárias que representam cada T são obtidas aplicando a imagem original na Equação 1.

$$L_b(x,y) = \begin{cases} 1, se\ t_l < L(x,y) \leq t_u \\ 0, caso\ contrário \end{cases} \quad (1)$$

Onde t_l e t_u são o menor e maior valor de limiar e $L_b(x,y)$ é a imagem binária.

Após aplicar TTBD no modelo em escala de cinza, inicia-se o segundo método, onde define-se o vetor de característica através de três valores distintos. O primeiro é descrito pela complexidade das bordas do objeto $L_e(x,y)$. Este valor é obtido através da Equação 2.

$$L_e(x,y) = \begin{cases} 1, se\ \exists(x^{'},y^{'}) \in N_8[(x,y)], \\ L_b(x^{'},y^{'}) = 0 \wedge L_b(x,y) = 0, \\ 0, caso\ contrário \end{cases} \quad (2)$$

Onde $N_8[(x,y)]$ representa a região de pixeis que estão conectadas nas 8 direções a (x,y). $L_e(x,y)$ recebe 1 se o (x,y) no correspondente $L_b(x,y)$ tem valor 1 e $(\wedge)$ tendo no mínimo um vizinho pixel com valor 0. Caso contrário recebe 0. A dimensão fractal é calculada a partir de $L_e(x,y)$ usando o algoritmo *box counting* descrito em (JR et al., 2010).

A segunda e terceira característica representam o valor médio e o tamanho (contagem de pixeis) das imagens binárias. Para mais detalhes sobre o método, referir ao trabalho de (COSTA; HUMPIRE-MAMANI; TRAINA, 2012). Neste caso, primeiro dividiu-se a imagem em análise em blocos com o mesmo tamanho dos blocos de treinamento/teste. Após, todos estes blocos passaram pelo processo de extração de informação. Para este problema, definiu-se um T, de forma experimental, como 4.

Classificação dos Blocos

Nesta etapa utilizou-se SVM para classificação dos dados extraídos. SVM é uma técnica de classificação que é fundamentada em *kernel* (BURGES, 1998). Escolheu-se esta técnica por apresentar uma performance melhor em um maior número de problemas, levando em conta que o SVM é eficiente e estável. Além disso, segundo (AKANDE et al., 2014), o SVM apresenta boa capacidade de generalização em cenários reais, onde ela geralmente supera outras técnicas também utilizadas em previsões e classificações.

Support Vector Machine baseia-se na noção da margem entre diferentes classes (KOTSIANTIS, 2007). Seu funcionamento está fundamentado na definição de um hiperplano que visa separar o espaço entre as classes das observações do conjunto de treinamento. Maximizar a margem, criando assim a maior distância possível entre as classes, resulta em um resultado mais aproximado, pois consegue então reduzir o erro de generalização.

Para treinar o classificador, os dados extraídos dos modelos foram normalizados. Inicialmente, definiu-se de forma randômica a separação de metade dos dados para treinamento e o restante, para teste do sistema. Considerando-se a possibilidade de grande variação dos dados extraídos, escolheu-se a *Radial Basis Function* como *kernel* para o SVM aqui proposto. Sendo assim, buscou-se encontrar o Gamma (γ) e a penalidade ótima para o *kernel* escolhido utilizando a técnica *grid search algorithm*, apresentado em (CHANG; LIN, 2011). Essa etapa é essencial para a técnica proposta, pois a escolha inadequada do γ ou da penalidade pode reduzir drasticamente a acurácia do sistema. Neste trabalho, os índices estatísticos usados para definir o Gamma e a penalidade foram Acurácia (Equação 3), sensibilidade Equação 4 e especificidade Equação 5).

$$Acurácia = \frac{VC1 + VC2}{VC1 + VC2 + FC1 + FC2} \tag{3}$$

$$Sensibilidade = \frac{VC2}{VC2 + FC1} \tag{4}$$

$$Especificidade = \frac{VC1}{VC1 + FC2} \tag{5}$$

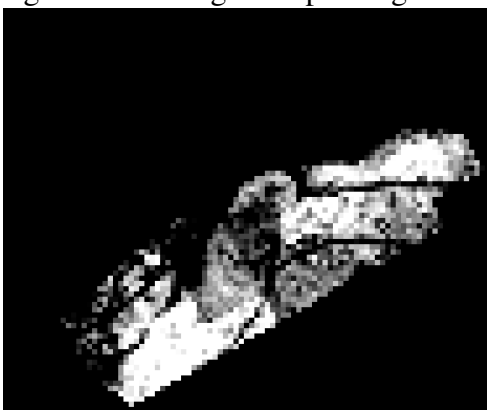
Onde VC1 representa os blocos que foram identificados corretamente como a classe 1, VC2 como verdadeiro classe 2, FC1 como falso classe 1 e FC2 como falso classe 2.

Os blocos definidos como teste foram utilizados para identificar o (γ) e a penalidade ideal para o problema proposto. Com o classificador corretamente definido, todos os blocos da imagem em análise receberam uma classe e uma distância que representam quão longe o bloco está da classe.

RESULTADOS E DISCUSSÕES

A classe e a distância que cada bloco recebeu do classificador binário foram utilizados para geração dos mapas. A classe definida como não densa recebeu um índice que variou de 0 até 0,5. Este valor depende diretamente da distância que o bloco recebeu. Quando mais longe da classe, maior o valor do índice. Já a classe definida como densa recebeu um índice variante entre 0,5 e 1. Quanto mais longe, menor o índice. Esse índice, por sua vez, foi utilizado para gerar uma imagem em escala de cinza, onde os blocos mais próximos do branco representam uma área muito densa enquanto os mais próximos do preto descrevem uma área não densa. O resultado final está apresentado na Figura 3.

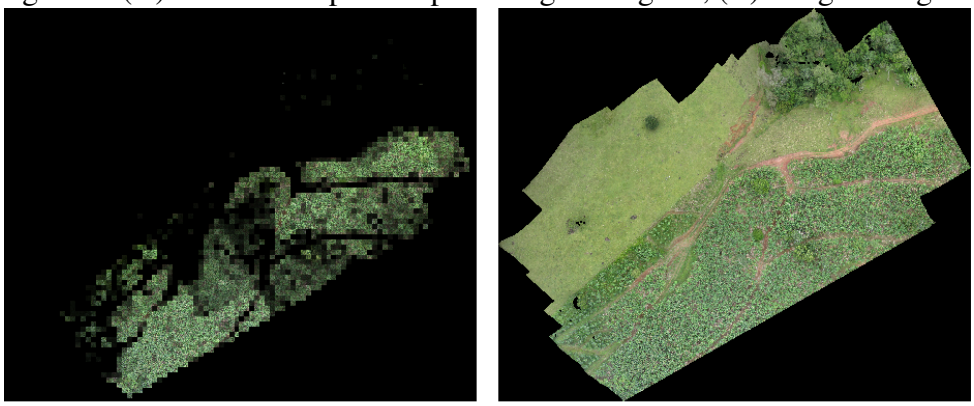
Figura 3: Índice gerado pelo algoritmo



Fonte: Autores

Visando uma comparação visual do resultado final com a imagem original, multiplicou-se este índice nos blocos da imagem original. Com isso, os blocos com uma coloração mais forte representam uma área mais densa enquanto os blocos com uma coloração mais fraca representam uma área menos densa. O resultado visual desta multiplicação e a imagem em análise estão apresentados na Figura 4.

Figura 4: (A) Índice multiplicado pela imagem original, (B) Imagem original



(A) (B)

Fonte: Autores

Através de uma análise visual é possível verificar que o resultado demonstra representar a realidade. É importante destacar que o algoritmo aqui proposto tem o objetivo de sugerir com maior detalhamento a densidade da plantação em análise, ou seja, o índice gerado visa ser uma aproximação da realidade através da análise da estrutura e de coloração dos blocos. A Figura 5 apresenta um segmento de cada imagem apresentada na Figura 4.

Tendo por exemplo uma comparação entre os segmentos (A) e (B) apresentados na Figura 5, considerando o padrão de amostras já descrito, fica visível que quando o bloco em análise possuí solo exposto e/ou vegetação rasteira, o sistema atribui um valor muito baixo ou nulo para a sua representação. Além disso, a palha de bananeira contribui diretamente para a redução do

índice. Ou seja, o sistema foi capaz de segmentar claramente grande parte dos blocos em que não haviam folhagens da planta em análise.

Figura 5: (A) Segmento da imagem apresentada com o índice multiplicado pela imagem original, (B) Segmento da imagem original



(A) (B)

Fonte: Autores

Porém, um fator negativo da solução proposta é a definição de valores medianos ou altos do índice para blocos que compõem alguns tipos específicos de plantas, como por exemplo os blocos que representam a árvore presente na Figura 5. Considerando-se que nesta solução se utilizou apenas imagens no espectro do visível, o índice de densidade pode sofrer influência de plantas com textura próxima da planta aqui analisada. Entende-se que o uso de dados coletados em outros espectros poderiam resolver o problema aqui descrito.

CONCLUSÕES

Neste trabalho foi proposto uma solução para geração de mapas de densidade em plantações de banana. Esta solução foi fundamentada em análise de textura através de SFTA e em aprendizagem de máquina, utilizando SVM.

Conforme demonstrado nas seções anteriores, pode-se verificar visualmente que, o sistema apresentou resultados satisfatórios nas imagens coletadas em uma plantação em Turvo - SC. Apesar de não haver um mapeamento feito a campo por um profissional responsável visando a validação estatística do método gerado, é possível concluir que o algoritmo proposto possui um potencial para esta aplicação. Como trabalhos futuros, pretende-se testar a solução proposta em novas áreas e validar os resultados obtidos em conjunto com profissionais qualificados.

REFERÊNCIAS

AKANDE, K. O. et al. Performance comparison of SVM and ANN in predicting compressive strength of concrete. *IOSR Journal of Computer Engineering*, IOSR Journals, v. 16, n. 5, p. 88–94, 2014.

BORGES, A. L.; SOUZA, L. da S. *O Cultivo da Bananeira*. O Cultivo da Bananeira, Bahia: Embrapa Mandioca e Fruticultura, 2004.

BURGES, C. J. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, Springer Nature, v. 2, n. 2, p. 121–167, 1998.

CHANG, C.-C.; LIN, C.-J. LIBSVM. *ACM Transactions on Intelligent Systems and Technology*, Association for Computing Machinery (ACM), v. 2, n. 3, p. 1–27, apr 2011.

COSTA, A. F.; HUMPIRE-MAMANI, G.; TRAINA, A. J. M. An efficient algorithm for fractal analysis of textures. *2012 25th SIBGRAPI Conference on Graphics, Patterns and Images*, p. 39–46, Aug 2012. ISSN 1530-1834.

FAO. *FAO Statistical Yearbook*. Roma, Itália: Food and Agriculture Organization of the United Nations, 2013. Http://www.fao.org/docrep/018/i3107e/i3107e.PDF.

HARRY, R.; WILLISON, N. Book. *Bananas*. 3rd ed. ed. [S.l.]: Harlow, Essex, England : Longman Scientific Technical, 1987. Rev. ed. of: Bananas / N.W. Simmonds. 1982, c1966. ISBN 0470206845 (Wiley).

HENRIQUES, M. J.; MANTA, V.; MARNOTO, J. Avaliação da qualidade posicional. aplicação a um levantamento realizado em coimbra utilizando vant. In: *Atas das I Jornadas Lusófonas de Ciências e Tecnologias de Informação Geográfica*. [S.l.: s.n.], 2015. p. 723–747.

IBGE. *Levantamento Sistemático da Produção Agrícola*. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística, 2017. Ftp://ftp.ibge.gov.br/Producao_Agricola/Fasciculo_Indicadores_IBGE/estProdAgr_201703.pdf.

JR, C. T. et al. Fast feature selection using fractal dimension. *Journal of Information and Data Management*, v. 1, n. 1, p. 3–16, 2010.

JúNIOR, L. R. A. et al. Validação de ortomosaicos e modelos digitais de superfície utilizando fotografias obtidas com câmera digital não métrica acoplada a um vant. In: *XVII Simpósio Brasileiro de Sensoriamento Remoto - SBSR*. [S.l.: s.n.], 2015. p. 2157–2164.

KOTSIANTIS, S. B. Supervised machine learning: A review of classification techniques. In: *Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*. Amsterdam, The Netherlands, The Netherlands: IOS Press, 2007. p. 3–24. ISBN 978-1-58603-780-2. Disponível em: <http://dl.acm.org/citation.cfm?id=1566770.1566773>.

LIAO, P.-S.; CHEN, T.-S.; CHUNG, P.-C. A fast algorithm for multilevel thresholding. *Journal of Information Science and Engineering*, p. 713–727, 2001.

POTTS, J. et al. *The State of Sustainability Initiatives Review*. [S.l.]: International Institute for Environment and Development, 2014.

SILVA, D. C. da et al. Qualidade de ortomosaicos de imagens de vant processados com os softwares aps, pix4d e photoscan. In: *V Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação*. [S.l.: s.n.], 2014. p. 747–754.

VANHOUDT, N. *On-farm Assessment of Banana Plant Density in Rwanda*. Tese (Doutorado) — Katholieke Universiteit Leuven, jul 2009.

ZONETTI, P. D. C. et al. AnÁlise de custo de produÇÃo e lucratividade de bananeira 'nanicÃo jangada' sob duas densidades de cultivo em ilha solteira-sp. *Revista Brasileira de Fruticultura*, v. 24, n. 2, p. 406–410, aug 2002.



Análise de viabilidade no uso de Multi-VANTs em ambientes agrícolas através da avaliação de protocolos de roteamento

Ederval Pablo Ferreira da Cruz[1], *Rodolfo Oliveira*[2]

[1] Instituto Federal do Espírito Santo (IFES) – Campus Itapina, Colatina, Esp. Santo, Brasil, ederval.cruz@ifes.edu.br

[2] FCT/Universidade Nova de Lisboa, Caparica, Portugal

rado@fct.unl.pt

RESUMO

Uma das áreas que recentemente tem chamado a atenção da comunidade científica é a das Flying Ad Hoc Networks (FANETs) ou Redes Ad Hoc Voadoras. As FANETs são formadas por Veículos Aéreos Não Tripulados (VANTs) ou popularmente conhecidos como drones que possuem a capacidade de trocar informações entre si. Vários artigos que abordam VANTs aplicados à agricultura adotam o uso de um único VANT. Porém, é vislumbrado em um futuro próximo que vários VANTs estarão nos céus trocando informações entre si, auxiliando as mais diversas aplicações. O presente artigo busca avaliar o desempenho de três clássicos protocolos de roteamento (AODV, GPSR e OLSR) variando as velocidades dos VANTs, em um cenário usando uma área real. Resultados mostram que os protocolos avaliados não lidam muito bem com situações onde os VANTs necessitam trocar dados entre si em alta velocidade. Porém, em cenários com baixa mobilidade dos VANTs, o protocolo GPSR mostra-se viável para determinadas aplicações agrícolas.

PALAVRAS-CHAVE: FANET, Roteamento, Protocolos, VANT, Drones.

ABSTRACT

One of the areas that recently has attracted attention of the research community is the Flying Ad Hoc Networks (FANET). FANETs are formed by Unmanned Air Vehicles (UAVs) or more known as drones that have ability to exchange information among them. Several papers that approach UAVs in applied agriculture adopt only one UAV. However, it is glimpsed in a near future that several UAVs will be flying in the sky exchanging among them, helping in

various kinds of applications. This paper investigates the performance of three classic routing protocols (AODV, GPSR and OLSR) varying UAVs' speed in a scenario using a real area. Results show that the protocols evaluated do not cope well in situations where UAVs need exchange data among them in high speed. However, in scenarios with low mobility of the UAVs, GPSR protocol seems to be viable to be adopted in certain agricultural applications.
KEYWORDS: FANET, Routing, Protocols, UAVs, Drones.

INTRODUÇÃO

Quando o assunto envolve tecnologia, certamente é visto que a sociedade está passando por transformações cada vez mais profundas. Graças ao advento de novos equipamentos com maior poder de computação, miniaturizados, novas formas de comunicação, entre outras características, permitem vislumbrar diversos tipos de aplicações. Graças a esses avanços, tem surgido o aumento no uso dos Veículos Aéreos Não Tripulados (VANTs) ou comumente chamados de drones. Um *hot topic* que está se destacando e chamando a atenção da comunidade científica ao redor do mundo, em relação ao uso de VANTs, é conhecido na literatura como *Flying Ad-Hoc Networks* (FANETs) ou Redes Ad Hoc Voadoras (BEKMEZCI; SEN; ERKALKAN, 2013).

As FANETs são uma subcategoria das Redes Móveis Ad Hoc (MANETs). As MANETs são um tipo de rede móvel na qual, basicamente, os nós podem se movimentar e que não dependem de nenhuma infraestrutura para entrar em operação. Existem trabalhos que abordam o uso de um único VANT (Single-VANT) para a realização de tarefas específicas (TORRES-SÁNCHEZ, 2013). Contudo, é visto que com a possibilidade de novas aplicações com maiores níveis de complexidade, a adoção de sistemas Multi-VANT torna-se real e mais vantajosa. Na arquitetura Multi-VANT, os nós trabalham de forma coordenada e cooperativa. O sistema Multi-VANT possui diversas vantagens em relação ao sistema Single-VANT, das quais podem ser destacadas: (1) Maior tempo de vida da rede; (2) Realização da missão designada pelo usuário em menor tempo; (3) Maior escalabilidade; (4) Minimização dos custos de operação e manutenção (YANMAZ et al., 2010).

O uso de VANTs também tem ganhado força na área agrícola e motivando a comunidade científica a investigar o seu uso. No trabalho de Chaves e La Scalea (2015), os autores usam VANTs no processamento de imagens para quantificar as áreas de solo e vegetação de uma região. O trabalho de Cerbaro (2016) demonstra, através de uma plataforma de baixo custo, a coleta e processamento de imagens de áreas agrícolas para a geração de um índice de vegetação muito conhecido, o NDVI, que é utilizado por possuir forte correlação

com o crescimento das plantas. Sendo assim, é possível vislumbrar diversas aplicações no campo, desde o monitoramento do crescimento de plantas, análise da umidade do solo, monitoramento de animais em um terreno, entre outras.

Porém, um dos maiores desafios a serem enfrentados em qualquer rede de comunicação, e nas FANETs não é diferente, é o roteamento dos dados, de forma confiável, entre um par origem-destino. Características das FANETs como alta mobilidade com diferentes velocidades, consumo de energia, interferências, condições ambientais, entre outras, podem levar a comunicações instáveis com taxas de erros significativas, particionamento da rede, mudanças na topologia e tempo limitado de vida da rede, devido às restrições energéticas dos VANTs, levando a constantes atualizações das tabelas de roteamento, impactando diretamente no desempenho da rede. Por isso, os protocolos de roteamento nas FANETs devem lidar com as características existentes de forma a conseguir um desempenho satisfatório de acordo com a métrica desejada (atraso, taxa de entrega, tempo de comunicação, entre outros).

Em relação aos protocolos de roteamento em FANETs, podem ser citadas duas categorias: roteamento baseado na topologia e roteamento geográfico.

O roteamento baseado na topologia busca reduzir o consumo de energia, além de minimizar o uso da largura de banda com mensagens de controle. Por isso, os nós da rede encaminham os pacotes baseados nas informações dos links existentes na rede. Em relação aos protocolos baseados na topologia, destacam-se duas categorias: protocolos pró-ativos e protocolos reativos.

O protocolo AODV (PERKINS, 2008) é um protocolo reativo onde as rotas não são atualizadas com a mudança da topologia, mas somente quando um nó-origem deseja enviar um pacote para um nó-destino. O protocolo OLSR (JACQUET, 2001) é um protocolo pró-ativo em que cada nó mantem uma ou mais tabelas de roteamento contendo informações de rotas para todos os nós da rede.

No roteamento geográfico os nós da rede devem possuir a capacidade de determinar a sua posição e, além disso, conhecer a posição dos outros nós. De forma geral, o conhecimento dessas posições pode ser obtido através de serviços de posicionamento baseados em satélites como o GPS. O protocolo GPSR (KARP, 2000) é um tipo de protocolo que se encaixa nesta categoria. O protocolo utiliza-se do encaminhamento guloso, onde é escolhido o nó vizinho mais próximo do nó de destino para encaminhar os pacotes.

No trabalho de Singh e Verma (2015), os autores realizam uma avaliação de protocolos de roteamento em FANETs, porém o trabalho não deixa claro qual modelo de mobilidade foi utilizado, fator que impacta diretamente no desempenho dos protocolos de roteamento, assim

como dimensões da área avaliada. Em Thounhom e Anusas-Amornkul (2016) os autores realizam uma avaliação de três protocolos de roteamento em diferentes altitudes em uma FANET. Porém, a avaliação não analisa a performance de nenhum protocolo geográfico e nem do protocolo OLSR, que é largamente adotado nas MANETs.

Na área agrícola, diversos trabalhos já vislumbram o uso de VANTs. Porém, muitos trabalhos adotam o sistema Single-VANT (CARDOSO et al, 2015). Por isso, o objetivo do presente artigo, visto que a adoção de sistemas Multi-VANT apresenta diversas vantagens em comparação com o sistema Single-VANT, é analisar a performance de três protocolos de roteamento, sendo dois deles baseados na topologia e um deles baseado no roteamento geográfico e assim verificar a viabilidade no uso de sistemas Multi-VANT, como por exemplo, na área agrícola.

MATERIAL E MÉTODOS

Nesta seção, serão apresentados detalhes sobre a avaliação de desempenho dos protocolos avaliados. Na Figura 1 são mostradas, em destaque, as dimensões da área implementada no simulador.

Figura 1 – Área destacada, extraída da ferramenta Google Earth, implementada no simulador



Fonte: (Google Earth, 2017)

Com o objetivo de fornecer maiores detalhes sobre a área simulada, na Tabela 1 são informadas as latitudes e longitudes da área extraída no Google Earth, destacadas na Figura 1. A área simulada possui as dimensões de 1.158 x 497 metros.

Tabela 1 – Coordenadas da área usadas na simulação

Pontos da imagem	Latitude	Longitude
A	19°29'35.91"S	40°45'59.30"O
B	19°29'35.91"S	40°46'38.94"O
C	19°29'52.08"S	40°45'59.30"O
D	19°29'52.08"S	40°46'38.94"O

Foi utilizado como simulador de redes para avaliar o comportamento dos protocolos de roteamento avaliados em uma FANET, o Network Simulator 2, mais conhecido pela comunidade científica como NS-2 (NS2). Devem ser levadas em consideração algumas características adotadas na simulação realizada. É considerado que os VANTs voam em uma altitude fixa. Considerando que os VANTs podem alcançar velocidades consideráveis, no presente trabalho foram adotadas diferentes velocidades para os VANTs de forma a verificar o impacto no comportamento dos protocolos de roteamento. O modelo de mobilidade adotado pelos VANTs foi o Gauss-Markov, onde basicamente consiste em determinar a próxima localização de um nó, baseada na sua localização e velocidades anteriores. Sendo assim, um dado VANT com velocidade V_t em um instante de tempo t é dependente da velocidade V_t em um instante de tempo t-1. A ferramenta BonnMotion (ASCHENBRUCK; ERNST; GERHARDS-PADILLA, 2010) foi utilizada para gerar a mobilidade dos VANTs.

A geração dos caminhos de comunicação foi dividida em duas fases. Na primeira fase, os caminhos entre os VANTs foram criados de origens aleatórias para destinos aleatórios, a uma taxa de 5 caminhos/segundo e seguindo uma distribuição exponencial. Se o caminho é criado, pode-se dizer que ele está disponível. Na segunda fase, o nó de origem periodicamente envia pacotes pelo caminho disponível até o momento em que este é quebrado.

Foram realizadas 10 rodadas de simulações para cada protocolo e para cada velocidade adotada na simulação. De forma a obter resultados estatisticamente mais confiáveis, os resultados mostrados possuem um nível de confiança de 95%.

Na Tabela 2 são mostrados mais alguns parâmetros utilizados na simulação. O modelo de propagação escolhido foi devido ao fato de não somente considerar o caminho direto entre dois nós, mas também o reflexo causado pela onda eletromagnética que incide o solo, atingindo o receptor, causando uma perda de sinal.

Tabela 2 – Parâmetros usados nas simulações

Protocolo MAC	802.11 com 11Mbps (tráfego unicast) e 2 Mbps (tráfego broadcast)
Modelo de Propagação de Rádio	Two Ray Ground
Protocolos avaliados	AODV, GPSR, OLSR
Tempo de Simulação	1800 segundos (30 minutos)
Quantidade de VANTs	10
Raio de transmissão	100 metros
Velocidades dos VANTs	10, 20, 30, 40 e 50 metros/segundo

De forma a caracterizar o desempenho de cada protocolo em cada cenário, as métricas avaliadas foram as seguintes:

- **Caminhos disponíveis/criados** – Um caminho disponível é a situação em que um nó de origem consegue enviar um pacote de *ping*, responsável por originar o caminho, e este é recebido pelo nó de destino. Caso haja uma comunicação entre os dois nós (origem-destino), logo é considerado um caminho disponível/criado.
- **Tempo de duração dos caminhos** – Considerando um caminho criado com sequência de nós *z* chamado de $C = \{v_1, v_2, ..., v_z\}$ em um instante t_1, a duração dos caminhos é considerado o maior intervalo de tempo $[t_1, t_2]$ para a sequência C.
- **Atraso na entrega dos pacotes** - Considerando um caminho criado com sequência de nós *z* chamado de $C = \{v_1, v_2, ..., v_z\}$ onde v_1 é o nó de origem e v_z é o nó de destino, o atraso do pacote é a diferença entre o tempo final t_f (o pacote chega ao nó de destino v_z) e o tempo inicial t_i (o pacote sai do nó de origem v_1).

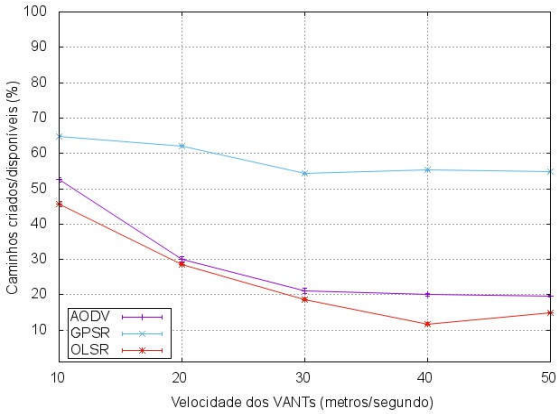
RESULTADOS E DISCUSSÃO

Nas figuras a seguir são mostrados os resultados acerca da performance dos protocolos de roteamento para as três métricas de desempenho explicadas na seção anterior.

Na Figura 2 é possível verificar que os protocolos OLSR e AODV têm suas performances impactadas diretamente conforme a velocidade dos VANTs aumenta, mostrando que em situações de alta mobilidade, mostram-se deficientes devido a pequena porcentagem de caminhos criados, especialmente no cenário de maior velocidade dos VANTs. O protocolo GPSR de acordo com o aumento da velocidade não é tão impactado quanto os demais protocolos avaliados. A queda ocorre devido à desconexão da rede e instabilidade dos

caminhos. A velocidade dos VANTs influencia diretamente na recepção das informações geográficas dos nós, impactando suavemente no desempenho do protocolo GPSR.

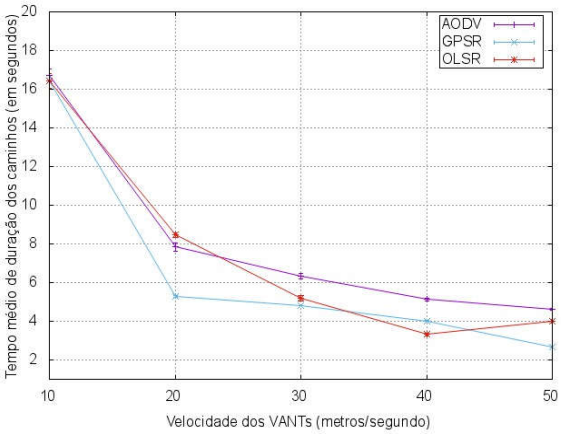
Figura 2 – Taxa de caminhos criados/disponíveis pelos protocolos avaliados



Fonte: Do próprio autor

Na Figura 3 os resultados da duração média dos caminhos criados são mostrados. Todos os protocolos sofrem conforme os VANTs aumentam a sua velocidade. O protocolo GPSR cria mais caminhos que os demais protocolos avaliados, conforme visto na Figura 2, porém são caminhos de baixa qualidade, com pouca duração, mostrando que, dependendo do tipo de aplicação, o desempenho pode não ser satisfatório. O protocolo OLSR, mesmo com o recurso de Multi-Point Relay (MPR), que reduz a quantidade de mensagens de controle, consumindo menos banda, assim como os demais protocolos, não obtém resultados satisfatórios nos cenários de maior velocidade dos VANTs.

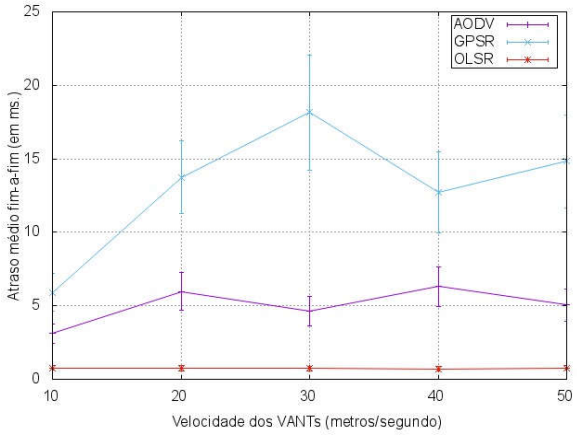
Figura 3 – Duração média dos caminhos criados



Fonte: Do próprio autor

Na Figura 4 é possível ver que o protocolo GPSR, devido ao fato de entrar em situações de máximo local, que é quando um dado nó da rede não consegue encaminhar um pacote para o nó mais próximo do destino, leva ao aumento do atraso na entrega do pacote. O protocolo OLSR, graças ao mecanismo de MPR, consegue reduzir o número de mensagens de controle, sendo sua performance melhor em relação ao atraso quando comparado com os outros protocolos avaliados. O protocolo AODV perde na comparação devido a sua técnica de inundação que necessita atualizar as tabelas de roteamento constantemente graças as quebras de rotas que ocorrem por causa da alta mobilidade dos nós implicando em constantes mudanças na topologia da rede.

Figura 4 – Atraso médio fim-a-fim



Fonte: Do próprio autor

CONCLUSÕES

Neste trabalho foi abordada uma área que tem chamado muita atenção por parte da comunidade científica: o roteamento nas FANETs. . Protocolos tais como o AODV, GPSR e OLSR, não foram concebidos originalmente para ambientes dinâmicos, com freqüentes mudanças na topologia da rede, que é o caso das FANETs. Estes protocolos foram concebidos originalmente para MANETs, redes estas com menor grau de mobilidade, o que implicava em um menor número de mudanças na topologia, levando a um número reduzido de mensagens de controle na rede.

Os resultados mostram que os protocolos avaliados no cenário do presente trabalho não são muito viáveis em situações onde os VANTs possuem alta velocidade, alcançando taxas reduzidas de caminhos criados. Mesmo o protocolo GPSR superando os outros protocolos nos

cenários com maior velocidade dos VANTs, o tempo médio de duração dos caminhos criados é bem reduzido, o que para aplicações críticas na agricultura, como por exemplo a transmissão de vídeos ou fotos de alta qualidade de uma propriedade obtida pelos VANTs, pode não ser viável. Contudo, no cenário onde os VANTs se locomovem com a menor velocidade (10 m/s = 36km/h), os resultados mostram que aplicações como leitura de sensores de solo para transmissão entre os VANTs até um centro de controle onde um usuário pode tratar esses dados recebidos, ou até mesmo transmissão de fotos e vídeos (dependendo da qualidade), pode ser viável. Para aplicações onde o atraso não é fator determinante o protocolo GPSR se mostrou, de forma geral, ter melhor performance em relação aos demais protocolos avaliados no presente trabalho.

O presente trabalho mostra que ainda tem muito a se investigar em busca de protocolos de roteamento que lidem com as características dos VANTs que fazem das FANETs um subtipo de rede MANET complexa de se lidar.

Como trabalhos futuros, os autores investigarão o impacto de outros modelos de mobilidade nos protocolos avaliados, em busca de verificar qual tem melhor desempenho, além de, ir em busca de um protocolo de roteamento que lide com as diversas características das FANETs, em especial que consiga aumentar o tempo de comunicação entre os nós e na integração de novas tecnologias tais como Redes Definidas por Software (SDN) e Computação em Névoa, que podem ajudar a melhorar a performance das FANETs.

REFERÊNCIAS

ASCHENBRUCK, N., ERNST, R, GERHARDS-PADILLA E., and SCHWAMBORN, M.: "BonnMotion - a Mobility Scenario Generation and Analysis Tool**,"** in Proc. of the 3rd International ICST Conference on Simulation Tools and Techniques (SIMUTools '10), Torremolinos, Malaga, Spain, 2010.

BEKMEZCI, I., SEN, I., ERKALKAN, E., "Flying ad hoc networks (FANET) test bed implementation", Comput. Eng. Dept., Turkish Air Force Acad., Istanbul, Turkey, 10.1109/RAST.2015.7208426, 665-668, 2015

CARDOSO, Alexandre et al. Técnicas de Engenharia de Software e Visualização da Informação de dados capturados por aeronaves remotamente pilotadas associadas na Pecuária de precisão. X Simpósio Brasileiro de Agroinfomática - SBIAGRO, Ponta Grossa, 2015.

CERBARO, Vinícius Andrei. CROP DRONE: UMA SOLUÇÃO PARA COLETA, ARMAZENAMENTO E DISPONIBILIZAÇÃO DE DADOS AGRÍCOLAS. 2016. Dissertação de mestrado. Universidade de Passo Fundo.

CHAVES, Arthur Avelar; LA SCALEA, Rodrigo Antonio. Uso de VANTs e processamento digital de imagens para a quantificação de áreas de solo e de vegetação. Anais XVII Simpósio Brasileiro de Sensoriamento Remoto-SBSR, João Pessoa-PB, Brasil, v. 25, 2015.

JACQUET, P.; MUHLETHALER, P.;CLAUSEN, T.et al., "Optimized link state routing protocol for ad hoc networks". In: IEEE International Multi Topic Conference, pp. 62–68, Pakistan, Dec. 2001.

KARP, B.; KUNG, H.T. "GPSR: greedy perimeter stateless routing for wireless networks". In:ACM Proceedings of the 6th Annual International Conference on Mobile Computing and Networking, pp. 243–254, New York, Aug. 2000

NS2. Versão 2.33. Disponível em: http://nsnam.isi.edu/nsnam/index.php/Main Page

PERKINS, C.; ROYER, E. M.; DAS, S. ―Ad hoc On-Demand Distance Vector (AODV) Routing‖. Disponível em: http://www.faqs.org/rfcs/rfc3561.html. Outubro de 2008.

SINGH, Kuldeep; VERMA, Anil Kumar. Experimental analysis of AODV, DSDV and OLSR routing protocol for flying adhoc networks (FANETs). In: Electrical, Computer and Communication Technologies (ICECCT), 2015 IEEE International Conference on. IEEE, 2015. p. 1-4.

THOUNHOM, Somsong; ANUSAS-AMORNKUL, Tanapat. The Study of Routing Protocols for UAVs using Paparazzi Mobility Model with Different Altitudes. In: Proceedings of the 2016 International Conference on Communication and Information Systems. ACM, 2016. p. 106-111.

TORRES-SÁNCHEZ, Jorge et al. Configuration and specifications of an unmanned aerial vehicle (UAV) for early site specific weed management. PLoS One, v. 8, n. 3, p. e58210, 2013.

YANMAZ, E.; COSTANZO, C; BETTSTETTER, C.; ELMENREICH, W. A discrete stochastic process for coverage analysis of autonomous UAV networks, in: Proceedings of IEEE Globecom-WiUAV, IEEE, 2010.



Um Panorama da Colaboração Científica na Pesquisa Agroinformática Brasileira

Sandro da Silva Camargo[1], *Rebeca Einhardt Fiss*[2], *Fabiane Nunes Prates Camargo*[1,2], *Leandro Bochi da Silva Volk*[1], *Jose Pedro Pereira Trindade*[1]

[1]Programa de Pós-Graduação em Computação Aplicada,
Universidade Federal do Pampa & Embrapa Pecuária Sul
Bagé, Rio Grande do Sul, Brasil
sandro.camargo@unipampa.edu.br

[2]Instituto Federal Sul-Riograndense
Rio Grande do Sul, Brasil
rebeca.fiss@ifsul.edu.br,fabiane.camargo@ifsul.edu.br

RESUMO

A colaboração em pesquisa tem se tornado fundamental para o progresso científico nas mais diversas áreas do conhecimento humano. Mensurar o nível destas colaborações é uma tarefa complexa, cujo custo computacional cresce exponencialmente em função do escopo que se deseja analisar. O presente artigo visa apresentar uma abordagem para identificar e analisar a rede de colaboração entre instituições no escopo da pesquisa em informática agropecuária no país, a partir da análise dos 139 artigos publicados nos anais do X Congresso Brasileiro de Agroinformática, ocorrido em 2015. A abordagem utilizada envolveu a aplicação de técnicas de visualização de dados e Análise de Redes Sociais. A rede de colaboração gerada permitiu identificar as instituições mais participativas no contexto da Agroinformática, enfatizando a relevância da Empresa Brasileira de Pesquisa Agropecuária como catalizadora da pesquisa nesta área. Espera-se que os resultados aqui apresentados possam fornecer elementos para auxiliar os órgãos governamentais a estabelecerem políticas para fortalecimento da pesquisa e desenvolvimento na área de Agroinformática, bem como enfatizar o papel da Embrapa na rede de pesquisa agropecuária brasileira.

PALAVRAS-CHAVE: Redes, Ciência, SBIAgro, Empresa Brasileira de Pesquisa Agropecuária.

ABSTRACT

Collaboration in research has become central to scientific progress in the most diverse areas of human knowledge. Measuring the level of these collaborations is a complex task, which computational cost grows exponentially according the scope to be analyzed. This paper aims to present an approach to identify and analyze the collaboration network among institutions involved in agricultural computing research based on the analysis of the 139 articles published in the annals of the X Brazilian Congress of Agroinformatics, which took place in 2015. The approach used has involved the application of data visualization techniques and social network analysis. The collaboration network allowed to identify the most participatory institutions in the context of Agroinformatics, emphasizing the relevance of the Brazilian Agricultural Research Corporation as a catalyst for research in this area. We hope results presented here may provide elements to help government agencies to establish policies to strengthen agricultural research and development, as well as emphasize Embrapa's role in the Brazilian agricultural research network.
KEYWORDS: Networks, Science, SBIAgro, Brazilian Agricultural Research Corporation.

INTRODUÇÃO

A colaboração em pesquisa tem se tornado uma atividade fundamental para o progresso da ciência por melhorar a comunicação entre grupos de pesquisadores, por permitir o compartilhamento de competências e por facilitar a produção de novos conhecimentos científicos. Mensurar o nível de colaboração entre membros de uma comunidade é uma tarefa complexa, que pode ser feita de várias maneiras distintas. A forma mais evidente de colaboração é a coautoria em publicações, que tem sido frequentemente utilizada como um parâmetro confiável para a mensuração dos níveis de colaboração (SU, 2010).

A fim de avaliar a colaboração na área de informática aplicada à agricultura no país, foi delimitado o contexto do Congresso Brasileiro de Agroinformática (SBIAgro), que é o principal fórum científico nacional a respeito deste tema[1]. Este evento constitui-se em um ambiente de compartilhamento de resultados de pesquisas, discussão de trabalhos em andamento e disseminação de informações sobre inovação na área de Agroinformática, reunindo pesquisadores, estudantes e interessados no assunto. O SBIAgro é um evento bianual que está na sua décima primeira edição, sendo promovido pela Associação Brasileira de Agroinformática[2], a qual é uma sociedade sem fins lucrativos com caráter autônomo e multi-institucional. Desde sua criação, em 1997, o SBIAgro foi sediado em dez diferentes cidades, apresentadas na Figura 1a, sendo elas: Belo Horizonte-MG (1997), Campinas-SP (1999), Foz do Iguaçú-PR (2001), Porto Seguro-BA (2003), Londrina-PR (2005), São Pedro-SP (2007), Viçosa-MG (2009), Bento Gonçalves-RS (2011), Cuiabá-MT (2013) e Ponta Grossa-PR (2015).

Tomando-se por base os dados estatísticos de produção de cereais, leguminosas e oleagi-

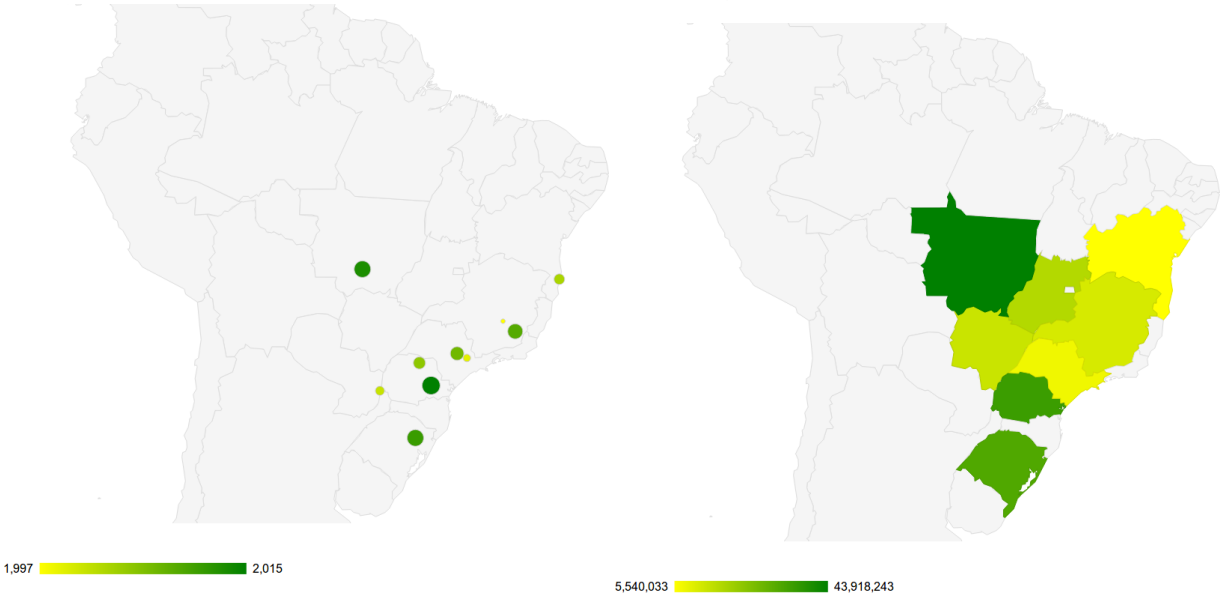
[1] https://www.sbiagro.org.br/sbiagro2017
[2] https://www.sbiagro.org.br/

nosas por Unidade da Federação (IBGE, 2017), verifica-se que o evento tem uma forte ligação com estados de notória vocação agrícola, visto que o evento já foi sediado nos dois maiores produtores de grãos da região Sudeste (São Paulo e Minas Gerais), dois maiores da região Sul (Rio Grande do Sul e Paraná), maior do Centro Oeste (Mato Grosso) e maior produtor do Nordeste (Bahia). Os oito estados com maior produção de grãos são apresentados na Figura 1b, com sua respectiva produção em toneladas, onde pode ser identificada a relação entre a alta produção de grãos do estado e seu histórico de acolhimento a alguma edição anterior do SBIAgro. Dada a importância do Congresso Brasileiro de Agroinformática no contexto nacional, é esperado que ele reflita um panorama da colaboração científica no âmbito da pesquisa em informática agropecuária sendo, por este motivo, escolhido como escopo de aplicação da abordagem aqui proposta.

Figura 1: Relação entre as sedes e produção de grãos dos estados

(a) Cidades sedes do SBIAgro entre 1997 e 2015.

(b) Oito estados com maior produção de grãos, em toneladas, em 2016.



Fonte: Elaborada pelo autor

Na literatura, podem ser encontrados diversos trabalhos de aplicação da abordagem de Análise de Redes Sociais, ou *Social Network Analysis* (SNA), para a identificação das redes de colaboração em pesquisa a partir de dados de publicações. Yu e Wang (2016) construíram e analisaram uma rede de colaboração, a partir de mais de 10.000 publicações anteriores a 2015. O foco deste trabalho foi restrito a dispositivos médicos regenerativos implantáveis, derivados de animais, baseados em tecnologia de engenharia de tecidos. Foi analisado o processo de evolução da rede e o estado corrente da colaboração. Como resultado, o trabalho evidenciou a escassa colaboração entre instituições, o potencial de fortalecimento da interação entre diferentes países, e que as instituições nas extremidades da rede poderiam aprimorar sua colaboração e compartilhar suas descobertas com outras instituições mais centrais. Já Wu e Duan (2015) mediram as atividades de colaboração científica em pesquisa psiquiátrica a nível de autores,

instituições e países. Para isso, foram consultados mais de 36.000 artigos. Como resultado, além da descrição do comportamento colaborativo na área de psiquiatria, também foi sugerido que a colaboração pode oferecer evidências científicas e sugestões razoáveis para embasarem a criação de políticas para guiarem o financiamento da pesquisa psiquiátrica no futuro. Racherla e Hu (2010) exploraram os padrões de colaboração na comunidade de pesquisa em turismo, tomando por base os três principais periódicos da área. Como resultado, verificou-se que a rede de colaboração é grande e complexa, e está dispersa na forma de diversos grupos bem delimitados de pesquisadores. Andrade et al. (2014) discutiram como os processos de colaboração científica podem ser identificados e caracterizados através da abordagem de Análise de Redes Sociais e redes complexas.

Também é discutido que a colaboração em comunidades científicas tem sido estudada como um dos elementos básicos para explicar a difusão de conhecimento. Por fim, Cechinel (2013) fez uma análise da evolução da colaboração científica entre países durante os primeiros sete anos da Conferência Latino-Americana de Objetos e Tecnologias de Aprendizagem. O autor sugeriu que os resultados poderiam auxiliar a comunidade a estabelecer políticas focadas na integração de países que ainda não estavam bem conectados dentro da comunidade.

Estimulado pela análise dos resultados de trabalhos correlatos, o presente estudo visa identificar e apresentar a rede de colaboração entre as instituições de pesquisa brasileiras na área da informática agropecuária, tomando-se por base os artigos publicados nos anais do SBIAgro 2015. O estudo se propõe a apresentar, a partir de evidências de coautorias de artigos, quais instituições exercem papel de destaque na área. Como impactos esperados, almeja-se que a rede de colaboração aqui identificada subsidie a criação de políticas governamentais de fomento à pesquisa e desenvolvimento em informática agropecuária, que a própria comunidade possa conhecer-se melhor a partir da identificação de seus membros e colaborações, e que outras instituições interessadas em começar a desenvolver pesquisas nesta área possam, a partir do conhecimento de quem são os membros chave, buscar parcerias para ingresso na rede de pesquisa.

O restante do artigo está organizado da seguinte forma: A Seção Material e Métodos apresenta uma breve descrição das características da base de dados utilizada para este estudo e a abordagem de Análise de Redes Sociais aplicada. A Seção Resultados e Discussão apresenta os resultados obtidos a partir da análise dos dados e a rede de colaboração identificada. A Seção Conclusões apresenta um resumo das descobertas, as restrições da abordagem utilizada e as perspectivas de trabalhos futuros.

MATERIAL E MÉTODOS

O presente trabalho utilizou como base os 139 artigos publicados nos Anais do X Congresso Brasileiro de Agroinformática, que estão disponíveis no site do evento[3]. Neste conjunto de arti-

[3]http://eventos.uepg.br/sbiagro/2015/anais/SBIAgro2015/index.htm

gos, foram identificados através de uma abordagem não automática: 531 autores, 65 instituições ou empresas sediadas em 15 estados brasileiros, além de duas instituições do Canadá e uma dos Estados Unidos. Um artigo com autores de duas ou mais instituições diferentes foi considerado como produto de uma colaboração, de acordo com a abordagem utilizada em (NEWMAN, 2004). Tendo-se um artigo qualquer com n instituições, onde n$\geq$ 2, para a abordagem de Redes Sociais foram consideradas como colaboração todas as combinações das n instituições, agrupadas 2 a 2, totalizando $C_{n,2}$ colaborações.

Para fins de simplificação de visualização, também foi utilizada a abordagem de *Edge Bundle* (TELEA; ERSOY, 2010). Neste caso, foram consideradas como colaborações somente as relações entre a instituição do autor principal e cada uma das instituições dos colaboradores. Assim, havendo autores de n instituições no artigo, foram consideradas $n - 1$ colaborações. Para a visualização interativa de dados georeferenciados foi utilizada a API(*Application Programming Interface*) Google Charts[4], que é um serviço gratuito oferecido pela Google (ZHU, 2012). Para a representação das colaborações na forma de uma rede social, foi utilizada a ferramenta Gephi 0.9.1[5], que é uma ferramenta gratuita e de código aberto para criação, análise e exploração de redes complexas (BASTIAN; HEYMANN; JACOMY, 2009).

A aplicabilidade das Redes Sociais para análise de colaboração em uma rede de pesquisa deve-se ao fato de elas serem uma metáfora amplamente utilizada para representar as relações entre membros de uma comunidade. A abordagem de análise de Redes Sociais é um conjunto de técnicas focada no estudo de uma estrutura social, considerando primariamente os dados dos relacionamentos e o contexto social dos membros, em detrimento das propriedades dos próprios membros (OTTE; ROUSSEAU, 2002).

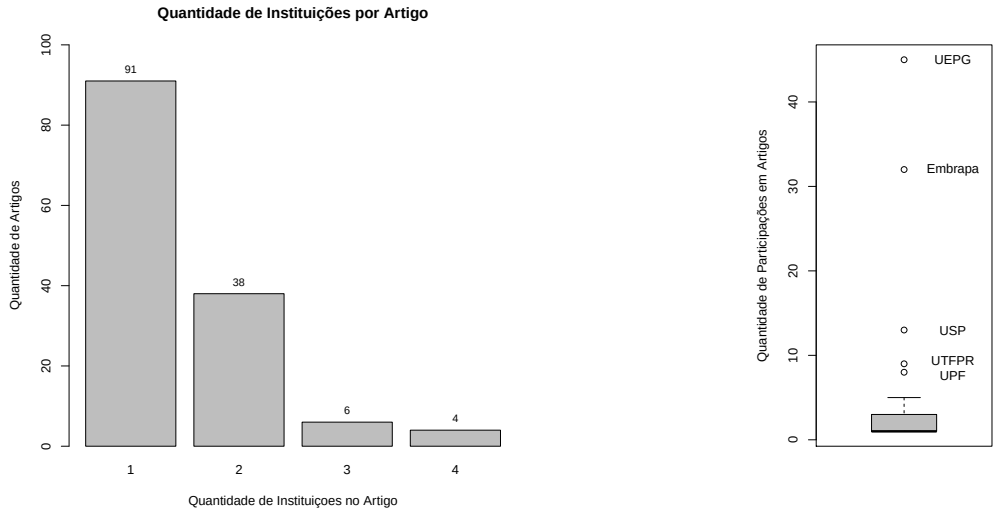
Seguindo a abordagem de Análise de Redes Sociais, uma rede de colaboração pode ser representada na forma de um Grafo $G(N, A)$, onde N é o conjunto de nós, que aqui representam as instituições de ensino ou pesquisa, e A é o conjunto de arestas, tendo-se que cada aresta a_z é um subconjunto de N com apenas dois elementos (NEWMAN, 2010). O tamanho do nó é diretamente proporcional ao número de artigos em que algum filiado da instituição foi coautor. As arestas do grafo foram definidas por $A(n_x, n_y)$ onde n_x e n_y são instituições parceiras em um artigo publicado. Para cada vez em que um par de instituições parceiras se repete, a espessura desta aresta é incrementada no grafo.

A Figura 2a apresenta uma análise preliminar da quantidade de colaborações identificadas nos 139 artigos publicados nos Anais do SBIAgro 2015. Neste contexto, 91 dos artigos foram escritos por coautores pertencentes a uma única instituição. Com participação de coautores de duas, três e quatro instituições foram identificados 38, 6 e 4 artigos, respectivamente. Assim, verifica-se que em aproximadamente 67% dos 139 artigos publicados não houve colaboração entre instituições, mostrando o grande potencial para aumento de interação nas pesquisas na área de Agroinformática.

[4]https://developers.google.com/chart/interactive/docs/
[5]https://gephi.org/

Figura 2: Dados dos trabalhos publicados no SBIAgro 2015

(a) Quantidade de instituições por trabalho (b) Quantidade de trabalhos por instituição



Fonte: Elaborada pelo autor

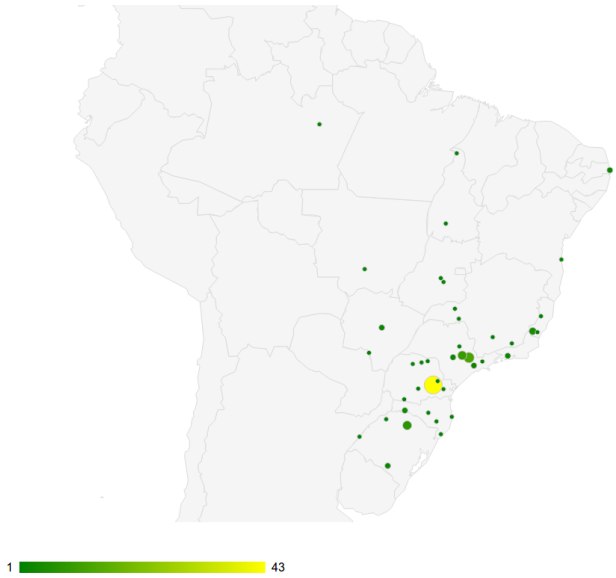
A partir da análise da filiação de cada um dos coautores dos artigos, foi gerada a Figura 2b, que apresenta a frequência de participação das instituições nos artigos publicados no SBIAgro 2015. A figura mostra que a abordagem estatística considerou, como dados discrepantes, as situações da Universidade Estadual de Ponta Grossa (UEPG), Empresa Brasileira de Pesquisa Agropecuária (Embrapa), Universidade de São Paulo (USP), Universidade Tecnológica Federal do Paraná (UTFPR) e Universidade de Passo Fundo (UPF) tendo participado de 45, 32, 13, 9 e 8 artigos, respectivamente. Tomando-se Q_1 como o quartil inferior, Q_3 como o quartil superior e d como a distância interquartílica, dada por $d = Q_3 - Q_1$, foram considerados como dados discrepantes aqueles valores menores que $Q_1 - 1,5d$ ou maiores que $Q_3 + 1,5d$. A participação das demais instituições estaria em um patamar estatisticamente aceitável, sendo uma instituição envolvida com cinco artigos, quatro instituições com quatro trabalhos, sete instituições que estiveram envolvidas com três trabalhos publicados, e dez instituições que colaboraram em dois trabalhos. Foram ainda identificadas 32 instituições que estiveram envolvidas em um único artigo, sendo importante mencionar a presença de coautores de três instituições estrangeiras: *Canadian Institute of Technology, University of Guelph* e *University of Florida*.

Assim, meramente a análise estatística da quantidade de participações, já indicaria evidências das instituições que mais se sobressaem no contexto da pesquisa na área de Agroinformática. A expressiva participação da UEPG deve ter sido potencializada pelo fato da instituição ter hospedado a edição 2015 do SBIAgro, visto que na edição anterior, ocorrida em 2013, esta Universidade colaborou em sete artigos.

Foi também realizada uma análise sobre as cidades dos autores principais dos artigos, a fim de buscar evidências georeferenciadas sobre a pesquisa na área de Agroinformática. Neste sentido, a Figura 3 permite identificar uma concentração no desenvolvimento de artigos nas

regiões sul e sudeste do país. Artigos produzidos por autores principais das regiões Norte e Nordeste foram raros. Corroborando com os resultados apresentados na Figura 2b, as cidades com maior contribuição na produção de artigos para o SBIAgro2015 foram: Ponta Grossa-PR (43), Campinas (13), Piracicaba (10) e Passo Fundo (8 artigos). As demais cidades eram representadas por autores principais em cinco ou menos artigos.

Figura 3: Cidades do autor principal dos trabalhos publicados no SBIAgro 2015

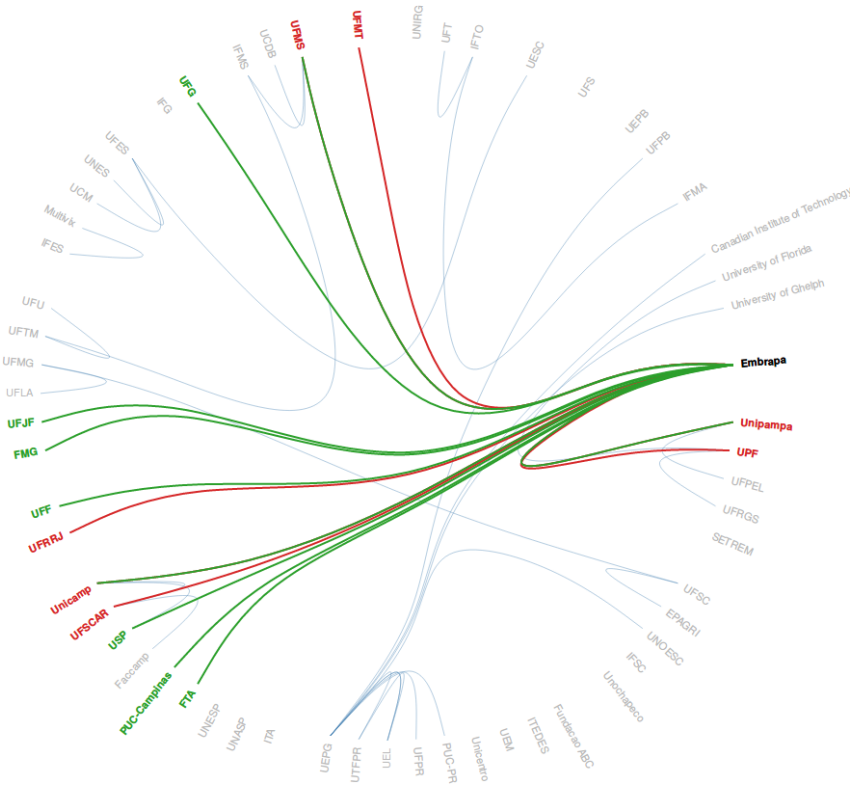


Fonte: Elaborada pelo autor (http://www.sandro.pro.br/pub/sbiagro2017/cidades-autores.html)

RESULTADOS E DISCUSSÃO

A Figura 4 mostra a rede de colaboração existente nos 139 trabalhos publicados nos Anais do SBIAgro 2015, tomando por base as instituições mencionadas na filiação dos autores de cada artigo. A figura enfatiza a importância da Embrapa como catalizador da pesquisa no âmbito da Agroinformática brasileira, em virtude de ter apresentado, no contexto analisado, a rede de colaboração mais abrangente envolvendo participação como autor principal ou como coautor em trabalhos de outras 14 instituições brasileiras de sete estados diferentes: Rio Grande do Sul, São Paulo, Rio de Janeiro, Minas Gerais, Goiás, Mato Grosso e Mato Grosso do Sul. Tal rede ainda envolve a participação indireta de várias outras instituições. Arestas em verde ligadas ao nome da Embrapa indicam que as instituições na outra extremidade da linha colaboraram em um ou mais artigos onde o autor principal era filiado da Embrapa. Já as arestas em vermelho, indicam que pesquisadores filiados à Embrapa participaram como coautores em artigos onde o autor principal era da instituição na outra extremidade da linha. A predominância das arestas verdes indica que, na maioria dos trabalhos em colaboração, o autor principal era filiado da Embrapa. A Universidade Estadual de Ponta Grossa, que sediou o SBIAgro 2015, apresentou a segunda maior rede de colaboração, envolvendo autoria ou coautoria em trabalhos com outras oito instituições, sendo a metade delas sediadas no próprio estado do Paraná.

Figura 4: *Edge Bundle* mostrando a rede de colaboração identificada nos trabalhos publicados no SBIAgro 2015



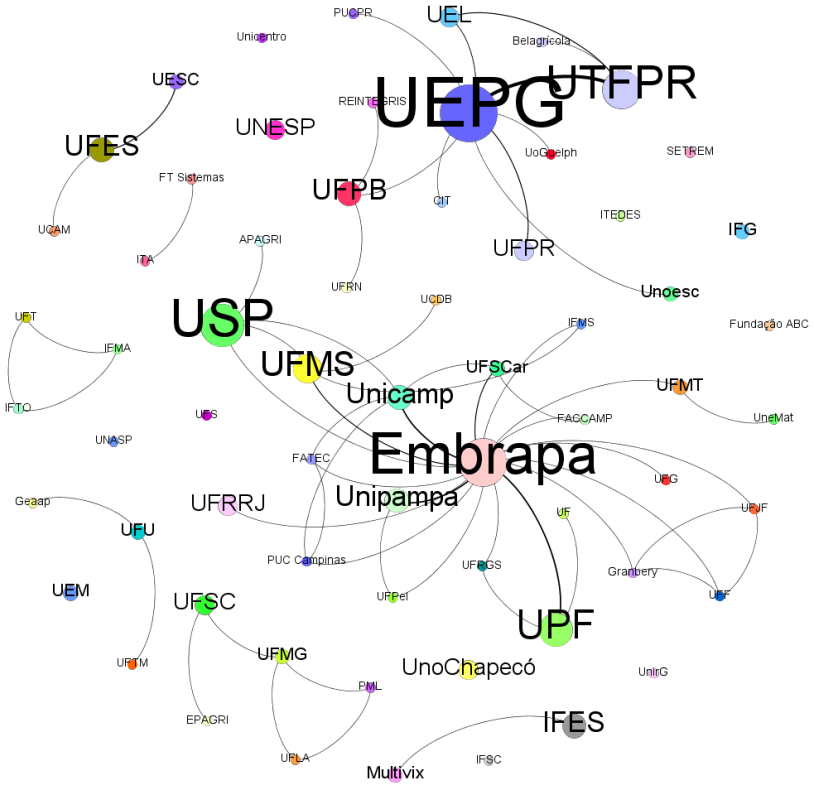
Fonte: Elaborada pelo autor (http://www.sandro.pro.br/pub/sbiagro2017/rede.html)

Já a rede social apresentada na Figura 5 mostra a existência de uma grande subrede de colaboração com forte participação da Embrapa (como elemento mais central, comprovado pelas principais métricas de centralidade, tais como *Betweenness* e *Pagerank*), USP e UPF, além da participação de outras 20 instituições. Uma segunda subrede pode ser identificada na figura, com a maioria das instituições sediadas no estado do Paraná, na qual a UEPG tem um papel de destaque, contando com outros onze colaboradores. Diversas outras subredes menores de colaboração existem, tipicamente formadas por duas ou três instituições, além de serem identificadas várias outras instituições que não estão ligadas a nenhuma subrede em virtude da falta de artigos em colaboração. Tal situação mostra um claro potencial de incremento dos trabalhos em parceria com uma das grandes subredes já consolidadas. A análise da rede social de pesquisa também permite evidenciar que a maioria das colaborações tem peso baixo, tipicamente geradas por um único artigo em parceria. No entanto, fica evidente a existência de algumas colaborações mais consistentes, notoriamente entre UEPG-UTFPR com 7 artigos em parceria; e Embrapa-UPF, Embrapa-Unicamp e Embrapa-Unipampa, todas com 3 colaborações.

CONCLUSÕES

Este trabalho apresentou um panorama da colaboração científica em informática agropecuária no Brasil a partir da identificação da rede de colaboração científica entre as instituições que

Figura 5: Rede Social das colaborações identificadas nos artigos publicados no SBIAgro 2015



Fonte: Elaborada pelo autor

tiveram artigos publicados nos anais do X Congresso Brasileiro de Agroinformática, em 2015. A abordagem aplicada permitiu identificar as instituições e as subredes mais relevantes neste contexto, e mostrou que há um grande potencial para o aumento de parcerias de pesquisa entre as instituições. A rede de colaboração aqui apresentada permite evidenciar a importância da Empresa Brasileira de Pesquisa Agropecuária como uma das instituições chave para o fortalecimento da pesquisa e desenvolvimento na área de Agroinformática. Além disso, também é possível identificar uma grande subrede secundária, envolvendo a Universidade Estadual de Ponta Grossa, que sediou o evento. Análises preliminares mostram que um fator potencializador para as colaborações é a existência de programas de pós-graduação com docentes de várias instituições, notoriamente na área de Computação Aplicada, mas também em outras áreas do conhecimento. Como restrição da abordagem aplicada, destaca-se que parte do trabalho de identificação das instituições envolvidas nos artigos foi realizada através de uma atividade não automatizada.

Como trabalho futuro, pretende-se utilizar a biblioteca OpenNLP para superar esta restrição. Também se pretende realizar a criação de um banco de talentos, que poderia ser gerenciado pela Associação Brasileira de Agroinformática, com áreas de interesse e competências dos pesquisadores da área, para aprimorar o potencial de colaboração, de parcerias e a qualidade dos trabalhos desenvolvidos.

REFERÊNCIAS

ANDRADE, M. T. T. et al. Contextualized analysis of social networks: Collaboration in scientific communities. *Social Networking*, v. 3, n. 2, Fevereiro 2014.

BASTIAN, M.; HEYMANN, S.; JACOMY, M. Gephi: an open source software for exploring and manipulating networks. In: *International AAAI Conference on Weblogs and Social Media*. Menlo Park, California: The AAAI Press, 2009. v. 8, p. 361–362.

CECHINEL, C. Scientific collaboration between countries in laclo from a social network analysis perspective. In: *Oitava Conferência Latino-Americana de Objetos e Tecnologias de Aprendizagem*. Valdivia, Chile: [s.n.], 2013. (LACLO 2013).

IBGE. *Indicadores IBGE: Estatística da Produção Agrícola*. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística, 2017.

NEWMAN, M. E. J. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National academy of Sciences of the United States of America*, v. 101, n. suppl 1, p. 5200–5205, 2004.

NEWMAN, M. E. J. *Networks: An Introduction*. New York, NY, USA: Oxford University Press, 2010.

OTTE, E.; ROUSSEAU, R. Social network analysis: a powerful strategy, also for the information sciences. *Journal of information Science*, v. 28, p. 441–453, 2002.

RACHERLA, P.; HU, C. A social network perspective of tourism research collaborations. *Annals of Tourism Research*, v. 37, n. 4, Outubro 2010.

SU, Y. F. *Study on the Cooperation Network of Spory Research Papers*. Shanghai: Shanghai University of Sport, 2010.

TELEA, A.; ERSOY, O. Image-based edge bundles: Simplified visualization of large graphs. In: *Proceedings of the 12th Eurographics / IEEE - VGTC Conference on Visualization*. Chichester, UK: The Eurographs Association & John Wiley & Sons, Ltd., 2010. (EuroVis'10), p. 843–852.

WU, Y.; DUAN, Z. Social network analysis of international scientific collaboration on psychiatry research. *Int J Ment Health Syst*, v. 9, n. 2, Janeiro 2015.

YU, S.-Y.; WANG, H.-M. Scientific collaboration: a social network analysis based on literature of animal-derived regenerative implantable medical devices. *Regen Biomater*, v. 3, n. 3, p. 197–203, Maio 2016.

ZHU, Y. Introducing google chart tools and google maps api in data visualization courses. *IEEE Computer Graphics and Applications*, v. 32, n. 6, p. 6–9, Novembro 2012.

Metodologia para a construção de portfólios tecnológicos agrícolas a partir de publicações técnico-científicas

Maria Fernanda Moura[1], Celina Maki Takemura[2], Ivan Lucas Cabral Silva[3], Luisa Miyashiro Tápias[3], Carolina Tavares de Oliveira[3], Luis Henrique Bassoi[4], Stanley Robson de Medeiros Oliveira[1]

[1] *Pesquisador(a) da Embrapa Informática Agropecuária, Campinas, São Paulo, Brasil, {maria-fernanda.moura, stanley.oliveira}@embrapa.br*

[2] *Pesquisadora da Embrapa Monitoramento por Satélite, Campinas, São Paulo, Brasil, celina.takemura@embrapa.br*

[3] *Estagiário da Embrapa Informática Agropecuária, Campinas, São Paulo, Brasil, mtapias.luisa@hotmail.com, caroli.agro@gmail.com*

[4] *Pesquisador da Embrapa Instrumentação, SãoCarlos, São Paulo, Brasil, luis.bassoi@embrapa.br*

RESUMO

Neste trabalho é apresentada a proposta uma metodologia semi-automática para construir portfólios tecnológicos, a partir de um grande número de publicações técnico-científicas. Para delimitar o conjunto de dados foi utilizado um sistema de busca confiável sobre uma base validada de publicações, selecionadas do Sistema Aberto e Integrado de Informação em Agricultura (SABIIA), que cobre todo o domínio de interesse. Para a construção do portfólio foram adotados: alguns recursos linguísticos - como dicionários construídos pelos especialistas no domínio; técnicas de extração de informações nos textos por similaridade; técnicas de análise de estatística descritiva (percentis, gráficos); e, técnicas de reconhecimento de padrões associativos (análise de regras de associação). O experimento conduzido mostrou a aplicabilidade da metodologia, tendo resultado em um portfólio de tecnologias do uso da água na agricultura.

PALAVRAS-CHAVE: mineração de textos, extração de informação em textos, dicionários, regras de associação, cobertura geoespacial de textos, desambiguação de topônimos.

ABSTRACT

In this paper, a semi-automatic methodology to build technological portfolios is proposed from on a large number of technical-scientific publications. To delimit the data set, a reliable search system was used on a validated database of publications, selected from the Open and Integrated Information System in Agriculture (SABIIA), which covers the whole domain of interest. For the construction of the portfolio some linguistic resources were adopted, to wit: dictionaries built by specialists in the field; techniques for extracting information in texts by similarity; techniques of descriptive statistics analysis (percentiles, graphs); and, associative pattern recognition techniques (association rule analysis). The experiment conducted showed the applicability of the methodology, resulting in a portfolio of technologies associated with water use in agriculture.

KEYWORDS: text mining, information extraction, dictionaries, association rules, geospatial coverage of texts, toponym disambiguation

INTRODUÇÃO

Neste trabalho, entende-se portfólios tecnológicos como planilhas com a relação de tecnologias, localidades de aplicação, tipo de solo, culturas e outros atributos de interesse para caracterizar o uso da tecnologia. Esses portfólios, após identificados, devem ser analisados por especialistas nas tecnologias, a fim de verificar a importância de cada relação encontrada. Um exemplo é o portfólio construído com o uso da metodologia aqui proposta, no qual o interesse é listar as tecnologias agrícolas, utilizadas no Brasil, e relativas ao uso da água na agricultura, com as regiões brasileiras e culturas agrícolas (OLIVEIRA et al., 2016). Após a identificação desse portfólio e análise das tecnologias de uso mais significativo (neste caso, as que mais aparecem nas publicações técnico-científicas de agricultura no Brasil), os especialistas de domínio verificaram quais dessas tecnologias podem ser soluções para problemas decorrentes da avaliação e da adaptação de tecnologias para os biomas brasileiros, frente às mudanças climáticas. Neste caso, a construção do portfólio é o passo inicial e fundamental para as demais análises.

Considerando-se que a construção de um portfólio é a base de dados para outras análises, é interessante que ele seja construído a partir de um grande volume de referências bibliográficas, com ampla cobertura espacial, que tenham sido validadas e componham uma fonte confiável de informações. Dado que o domínio do trabalho é agrícola, optou-se por buscar informações via o Sistema Aberto e Integrado de Informação em Agricultura (SABIIA) (VACARI et al., 2011). A vantagem do SABIIA é reunir dados de vários provedores

(repositórios institucionais e temáticos, e periódicos científicos) previamente coletados e selecionados de acordo com as áreas de interesse da Empresa Brasileira de Pesquisa Agropecuária (Embrapa); o que nos oferece uma variada fonte de publicadores e de material altamente confiável. Essa variedade resulta em grandes conjuntos de resposta de busca, o que inviabiliza a análise humana desses resultados, em um curto intervalo de tempo.

Assim, neste trabalho, apresenta-se uma metodologia semi-automatizada com base em processos de mineração de dados textuais. Semi-automatizada porque há a dependência humana para: a) a definição das expressões de busca; b) construção de um vocabulário e sua sinonímia, caso não exista um disponível na literatura; e, claro, c) a avaliação e seleção dos resultados – tarefa de pós-processamento. A automatização do processo se dá na construção das planilhas a partir dos textos, extração/identificação de geolocalidades e extração de relações, por meio de regras de associação. As regras de associação buscam encontrar o relacionamento entre itens de dados que ocorram com uma certa frequência, ou seja, identificar padrões em dados históricos (AGRAWAL et al., 1994). Por exemplo, uma regra encontrada para o portfólio de tecnologias agrícolas ligadas ao uso da água foi "se (região=nordeste) e (cultura=uva) então (tecnologia=irrigação por gotejamento)". Os resultados obtidos com as regras de associação resumem a compreensão das principais relações entre os atributos (que no experimento são tecnologias, culturas, solos e regiões) e essas são as relações que interessam para indicar a versão final do portfólio. Deve-se observar que, o contexto semântico das palavras e suas sinonímias não são considerados na extração das relações, o que pode resultar em falsas relações associativas, mesmo que elas apresentem uma alta frequência.

Nos próximos itens, detalha-se a metodologia proposta, discutem-se os principais resultados obtidos com o experimento das tecnologias agrícolas relacionadas ao uso da água na agricultura brasileira e apresentam-se as conclusões.

MATERIAL E MÉTODOS

A Figura 1 ilustra as etapas de construção do portfólio, cujos detalhes são apresentados nas próximas subseções.

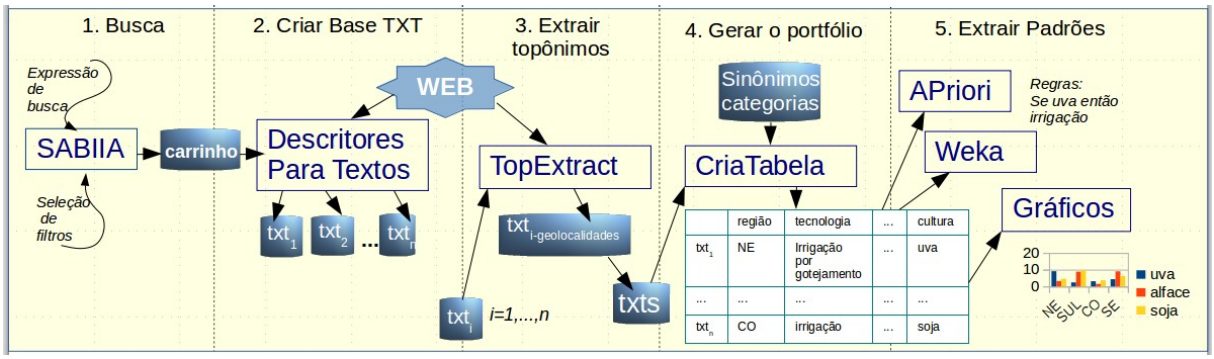
Busca

A busca é realizada pelo SABIIA, devido à variedade de publicações de acesso livre, validação dos domínios (publicadores) pelas bibliotecárias da Embrapa e, consequentemente, confiança nos dados.

O primeiro passo nesta etapa é construir as expressões de busca. Essa tarefa é realizada por especialistas do domínio das tecnologias junto aos mineradores de textos. Dessas expressões depende o resultado da análise, dado que elas delimitam o conjunto de dados sobre o qual os resultados serão obtidos. Logo, pode ser que se volte a esta etapa, se nas próximas etapas os dados não fornecerem uma cobertura de documentos aceitável.

Os filtros, "seleção de filtros" na Figura 1, são recursos que o SABIIA oferece, tais como filtrar os resultados de busca de acordo com: provedores de dados, autores, palavras-chave presentes nos resultados, tipo de publicação (relatórios, artigos, etc), ano de publicação e idioma. Esses filtros ajudam a delimitar os resultados de uma busca.

Figura 1: Metodologia semi-automática para a construção de portfólios tecnológicos



Criar base TXT

Quando o minerador e o especialista estiverem satisfeitos (este é um julgamento subjetivo) com o resultado da busca, eles devem salvar os documentos correspondentes no "carrinho do SABIIA". O carrinho é um arquivo RTF, que pode ser convertido para txt plano, e que carrega todos os metadados de cada publicação recuperada. Nesta etapa, o carrinho é convertido para *n* arquivos textos.

A ferramenta de software[1] que faz essa conversão, permite que, se for interesse da análise de dados, sejam incluídos os textos completos na base TXT, delimitados por descritores XML e se, e somente se, os textos completos forem de acesso livre e, estiverem facilmente disponíveis nas urls indicadas em seus metadados.

Extrair Topônimos

Normalmente, é de interesse nesse tipo de análise, que se saiba os locais onde essas tecnologias são utilizadas. Para isso tem sido usada a ferramenta TopExtract[2], que utiliza uma biblioteca Python poliglota - que identifica entidades nomeadas como topônimos nos textos.

[1] Nomeada decritores2txt.py, em python, que pode ser requisitada aos autores.

[2] A ferramenta TopExtract também é desenvolvida em Python e pode ser requisitada aos autores.

Após identificar essas entidades, a ferramenta executa um processo de desambiguação, que leva em conta todo o conjunto de topônimos encontrados em cada texto. Por exemplo, se o Rio São Francisco aparece no texto, é necessário saber qual o polígono envolvente do trecho presente no texto (TAKEMURA et al., 2016), que pode ser considerado como o mais próximo a um outro topônimo também presente no texto, por exemplo, Petrolina. A partir do polígono encontrado, delimita-se a que polígono das regiões brasileiras ele pertence (Norte, Sul, etc).

Gerar o Portfólio

Essa etapa envolve: i) extração de termos dos textos, por similaridade e sinonímia; ii) criação das transações presentes nos textos; e iii) eliminação de dados redundantes ou que não contribuem para os resultados.

Na i) extração de termos dos textos, primeiro resolve-se a questão de quais são os termos de interesse e suas sinonímias. Normalmente essa tarefa é manual, a menos que já exista a relação de termos e seus sinônimos para o domínio em questão. Após essa identificação escolhe-se um grau de similaridade que se ache adequado à aplicação, por exemplo, 75% de coincidência - utilizando-se “edit distance” para calcular a similaridade. Por exemplo, ao comparar palavras como “alho” e “milho”, é conveniente que se utilize um grau de similaridade acima de 80%. Também é aconselhável que se construam dicionários de categorias; por exemplo, para a categoria “controle fitossanitário” pode-se ter um conjunto de termos tais como “{pulverização, pesticida, agrotóxico}” e outros de interesse. As categorias também são colocadas no portfólio e podem ser objeto de análise.

Para a ii) criação das transações geram-se combinações dos atributos. Por exemplo, se um documento apresenta para o atributo “cultura” os valores “alface” e “uva” e, para o atributo “tecnologia” os termos “pulverização” e “irrigação”, criam-se as combinações: (“alface”, “pulverização”); (“alface”, ”irrigação”); (“uva”, “pulverização”); e, (“uva”, ”irrigação”). Cada combinação dá origem a uma transação, ou seja, uma linha do portfólio.

Para a etapa de iii) eliminação de dados redundantes ou desnecessários, considera-se, primeiramente, colunas (atributos) que estejam todos vazios. A eliminação de linhas, isto é, de documentos, segue a ideia de que o número de combinações para cada documento não pode ser muito discrepante entre os demais documentos. Para isso, após observações empíricas de alguns casos, decidiu-se eliminar os documentos de acordo com os percentis das frequências de combinações que eles geram. Um número razoável a ser eliminado, de acordo com os experimentos realizados, é do percentil 75% ou superior - porém, manteve-se a escolha do percentil como parâmetro do processo dado que essa escolha depende de cada resultado de

busca. Uma outra forma de eliminar dados é considerar apenas as combinações que obrigatoriamente contenham uma instância de um determinado atributo, por exemplo, "tecnologia" não pode estar vazia na combinação. Esses filtros podem ser aplicados de forma independente e ao longo das análises, sem que seja necessário gerar um novo portfólio.

Extrair Padrões

Nesta etapa analisa-se o portfólio obtido, gerando-se regras de associação entre os seus atributos e gráficos descritivos da distribuição de seus atributos, de modo a visualizar um resumo dos dados. As regras de associação ajudam a identificar quais são as relações estatisticamente mais significantes entre regiões, tecnologias e culturas, e, também as relações mais raras, que, em muitos casos, podem ser as mais importantes. De acordo com a análise dessas relações, mantém-se ou não linhas no portfólio encontrado. Essas decisões são dos especialistas nas tecnologias junto aos mineradores de dados.

A opção pelo uso de regras de associação é a sua facilidade de interpretação, e a existência de algoritmos bastante eficientes frente a um grande conjunto de dados. As regras de associação podem ser representadas da forma $X \rightarrow Y$, em que X representa o antecedente e Y o consequente, tal que $X \cap Y = \emptyset$. Para cada regra estão associadas duas medidas tradicionais: confiança (Conf) e suporte (Sup). Sup representa o número de tuplas que contêm X e Y, ao passo que Conf constitui a razão entre o número de tuplas que contêm X e Y sobre o número de tuplas que contêm X e: Sup=P(X U Y); Conf=P(X/Y). Do ponto de vista conceitual, o suporte representa a significância estatística dos itens (termos) nas tuplas, ao passo que a confiança determina a força da regra; um confiança alta implica em regras mais comuns (fortes) e uma confiança baixa em regras mais raras.

Para a geração das regras de associação foi utilizado o algoritmo Apriori (LIU et al., 1998), disponível no software Weka (HALL et al., 2009). A versão permite ao usuário definir o consequente das regras que serão geradas.

RESULTADOS E DISCUSSÃO

O experimento aqui mostrado é o resultado da construção de um portfólio de tecnologias agrícolas relacionadas ao uso da água na agricultura. A fonte de dados são os documentos técnico-científicos disponibilizados no SABIIA. As expressões de busca foram inicialmente propostas por especialistas do domínio (empregados da Embrapa) e, ao longo das análises, foram sendo aperfeiçoadas e evoluídas (OLIVEIRA et al., 2016). Ainda, os especialistas trabalharam na construção de dicionários de sinônimos para as tecnologias e para as culturas,

a partir dos termos dos thesaurus agrícolas Thesagro[3] e do Agrovoc[4] combinados aos termos presentes nas expressões de busca e retornados junto aos filtros de assunto pelo SABIIA. Por exemplo, para cultura "MANDIOCA" o dicionário construído apresenta as opções "mandioca, cassava, manioc", e, para tecnologia "IRRIGACAO POR GOTEJAMENTO" o dicionário construído apresenta "irrigacao por gotejamento, irrigação por gotejamento, drip irrigation", ou seja, inclusive as variações produzidas pela retirada de caracteres latinos (ç, ...).

Após a realização de algumas análises e observação da distribuição de frequências das culturas e tecnologias, os especialistas identificaram categorias de culturas e tecnologias presentes nos resultados de busca. Um exemplo é a classe "Base alimentar e commodities" que engloba as culturas "arroz, feijao, milho, soja, cana-de-acucar, limao, laranja, algodao, trigo, cafe, sorgo, mandioca, chuchu, pepino, batata" e, um outro, a classe de tecnologia "Melhoramento genetico" que engloba "cultivar, metodos, propagacao vegetativa, hibridacao, variacao genetica, variabilidade genetica, micropropagacao, germoplasma, melhoramento genetico ".

A base de dados final, após várias consultas e reformulações e filtros de consultas, foi formada por 2.185 documentos recuperados do SABIIA. Para esses documentos foram geradas 126.565 combinações (linhas do portfólio consideradas como transações). Como o objetivo inicial do portfólio era identificar tecnologias que envolvam o uso da água na agricultura utilizadas nas regiões do Brasil, foram eliminadas as linhas do portfólio que não apresentavam informação sobre o atributo tecnologia, resultando em 99.668 transações. Em seguida, foi observado que algumas linhas do portfólio eram constituídas de várias combinações de tecnologias com culturas, geradas de uma mesma publicação (mesmo documento txt). Para lidar com esse problema, foi aplicado um filtro que eliminou as linhas do portfólio cujas frequências foram superiores ao percentil 95%. Com isso, o portfólio foi reduzido a 84.204 transações. Por último, foram eliminadas do portfólio as linhas que não apresentaram informações sobre regiões do Brasil, pois tanto tecnologia como região eram objeto de estudo desse trabalho, resultando em 2.185 combinações (linhas).

Deve-se salientar que os procedimentos adotados acima foram realizados depois de várias tentativas empíricas de se obter um recorte razoável dos dados. Esse recorte razoável

[3] Disponível em: http://snida.agricultura.gov.br:81/binagri/html/Cen_Thes1.html. Consultado em 23 de maio de 2017.

[4] Disponível em: http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus. Consultado em 23 de maio de 2017.

deveria ser capaz de apresentar regras interessantes e também regras raras, envolvendo tecnologias, culturas e regiões. Essas decisões, embora baseadas em medidas, são subjetivas.

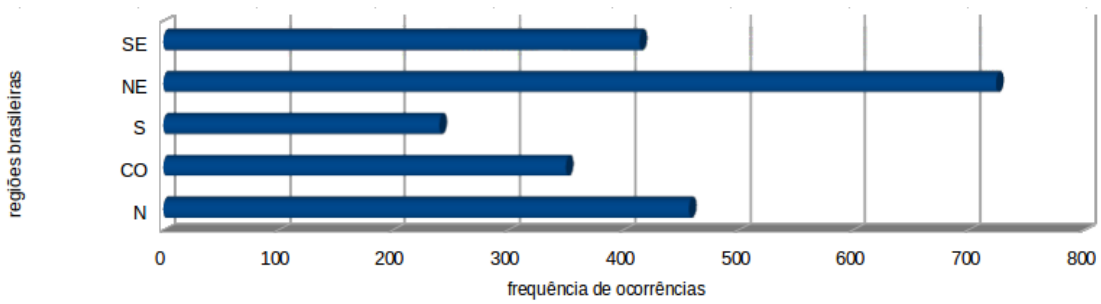
Dessa forma, obteve-se um portfólio, para o qual algumas linhas são ilustradas na Figura 2. No portfólio guardou-se a informação do arquivo (documento) que gerou cada transação, pode-se observar que o A_54904.201406150000.txt gerou duas transações, pois contém a tecnologia - "plantio direto" relacionada às classes ("solo" e "sistemas de producao"), ambas atreladas à cultura da soja.

Figura 2: Visão parcial do portfólio de tecnologias.

Arquivo	tecnologia	tecnologiaClasse	cultura	culturaClasse	solo	regiao	uf
A_13595.2015061500000.txt	irrigacao	'engenharia agricola'	uva	frutiferas		NE	BA
A_43443.2013061500000.txt	'manejo fitotecnico'	'controle fitossanitario'	alho	olericulturas		NE	BA
A_54904.2014061500000.txt	'plantio direto'	solo	soja	'base alimentar e commodities'		NE	BA
A_54904.2014061500000.txt	'plantio direto'	'sistemas de producao'	soja	'base alimentar e commodities'		NE	BA
A_20638.1982061500000.txt	germoplasma	'melhoramento genetico'	arroz	'base alimentar e commodities'		SE	MG
A_20638.1982061500000.txt	'variabilidade genetica'	'melhoramento genetico'	couve	olericulturas		SE	MG
A_20638.1982061500000.txt	'variabilidade genetica'	'melhoramento genetico'	arroz	'base alimentar e commodities'		SE	MG
A_61753.2015061500000.txt	manejo	solo	mandioca	'base alimentar e commodities'		CO	GO
A_44806.2000061500000.txt	'manejo do solo'		alho	olericulturas	latossolo	CO	GO
A_51410.2000061500000.txt	manejo	solo	feijao	'base alimentar e commodities'		CO	GO
A_55263.1984061500000.txt	fertilizante	solo	seringueira	extrativas		CO	GO
A_55263.1984061500000.txt	adubacao	solo	seringueira	extrativas		CO	GO
A_53347.2004061500000.txt	'plantio direto'	solo				S	RS
A_53347.2004061500000.txt	'plantio direto'	'sistemas de producao'				S	RS
A_33803.2014061500000.txt	manejo	solo			argissolo	S	RS
A_55657.2011061500000.txt	fertilizante	solo	alho	olericulturas	chernossolo	S	RS
A_23013.1998061500000.txt	irrigacao	'engenharia agricola'	pimenta	olericulturas		N	RR
A_23013.1998061500000.txt	adubacao	solo	pimentao	olericulturas		N	RR
A_25681.2010061500000.txt	'melhoramento genetico'	'melhoramento genetico'	alho	olericulturas		N	RR
A_25681.2010061500000.txt	'melhoramento genetico'	'melhoramento genetico'	arroz	'base alimentar e commodities'		N	RR

Para a geração das regras, foram conduzidos vários experimentos em que a tecnologia foi fixada como o consequente da regra (Y) e os demais atributos foram combinados no antecedente da regra (X). Desta forma, as regras identificaram qual o grau de associação entre tecnologias, tipos de solo culturas e UF em cada região do Brasil. No entanto, devido ao grande desbalanço entre os dados de cada região, como mostrado no gráfico da Figura 3, as 2.185 linhas do portfólio (instâncias) foram segmentadas por região, e cada região analisada individualmente. Dessa forma, a segmentação permitiu que regras de associação das regiões com menor frequência (Sul e Centro-Oeste) fossem identificadas no processo de mineração. As regras geradas apresentaram baixos valores de suporte e altos valores de confiança.

Figura 3: Dispersão das regiões brasileiras nos dados obtidos



Após a segmentação, para a região Norte foram encontradas 32 regras, considerando o suporte de 2% e a confiança de 90%, sendo parte delas redundantes ou não apresentava novidade para o especialista do domínio. Um dos exemplos de uma regra relevante foi *(Se tecnologiaClasse = planta & culturaClasse = base alimentar e commodities & solo = latossolo ==> tecnologia = fertilizante)*. Outro exemplo foi *(Se tecnologiaClasse = sistemas de producao & culturaClasse = base alimentar e commodities ==> tecnologia = plantio direto)*.

Para dados disponíveis das demais regiões do Brasil foram geradas regras similares. Por exemplo, para a região Nordeste, com suporte de 2% e confiança de 90%, foram geradas 40 regras. Exemplos dessas regras incluem *(Se tecnologiaClasse = engenharia agricola & cultura = uva & culturaClasse = frutiferas & regiao = NE ==> tecnologia = irrigacao)* e *(Se cultura = maca & culturaClasse = frutiferas & regiao = NE & uf = BA ==> tecnologia = manejo dos recursos hidricos)*.

Para a região Centro-Oeste foram geradas 18 regras, considerando suporte igual a 3% e confiança igual a 90%. Exemplos dessas regras incluem: *(Se cultura = trigo & culturaClasse = base alimentar e commodities & uf = DF ==> tecnologia = fertilizante)* e *(Se tecnologiaClasse = solo & culturaClasse = olericulturas & solo = latossolo ==> tecnologia = plantio direto)*.

Por fim, para a região Sul foram geradas 36 regras de associação para o suporte igual a 2% e confiança igual a 90%, com destaque para as regras *(Se cultura = trigo & culturaClasse = base alimentar e commodities & uf = SC ==> tecnologia = adubacao)* e *(Se culturaClasse = base alimentar e commodities & solo = gleissolo ==> tecnologia = fertilizante)*.

CONCLUSÕES

Este trabalho apresentou uma metodologia semi-automatizada para a construção de portfólios tecnológicos a partir de um grande volume de publicações técnico-científicas com o uso de ferramentas de mineração de textos. A metodologia é baseada em: i) busca e delimitação do domínio de resultados, o que depende de um trabalho minucioso com auxílio dos especialistas de domínio; ii) criação automática de base de dados; iii) criação do portfólio, com decisões de filtros a serem aplicados dependentes dos mineradores de dados; e, iv) análise dos dados a partir de regras de associações automaticamente geradas. O processo é retroalimentável, pois observações em uma etapa podem sugerir aos analistas o retorno a uma ou mais etapas.

Deve-se observar que, nesta metodologia não foram considerados os contextos semânticos nos quais os termos dos dicionários aparecem. Os termos são considerados apenas

por presença, frequência e combinações destas no mesmo texto. Isto permite construir uma representação *bag of words* das transações, que pode resultar em regras não verdadeiras, que devem ser avaliadas pelos especialistas do domínio. Sendo que, o experimento ilustrativo mostra que a metodologia serviu ao seu propósito, tendo colaborado na construção de um portfólio de tecnologias relacionadas ao uso da água na agricultura.

Adicionalmente, a proposta da metodologia permite que ela seja adaptada a vários outros problemas semelhantes, desde que se trabalhe os recursos linguísticos (dicionários de sinonímias e categorias) dependentes do domínio e do conhecimento dos especialistas.

REFERÊNCIAS

AGRAWAL, R.; SRIKANT, R. Fast algorithms for mining association rules in large databases. In: INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, 20., 1994, Santiago. Proceedings... San Francisco: Morgan Kaufmann, 1994. p. 478-499.

HALL, M. A.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. **The WEKA Data Mining Software: An Update.** SIGKDD Explorations, v. 11, n. 1. 2009.

LIU, B., HSU, W., MA,Y. Integrating Classification and Association Rule Mining. In: Fourth International Conference on Knowledge Discovery and Data Mining, 80-86, 1998.

OLIVEIRA, S. R. de M.; MOURA, M. F.; TÁPIAS, L. M.; OLIVEIRA, C. T. de; BASSOI, L. H. Técnicas de mineração de textos para auxiliar a construção de portfólio de tecnologias de irrigação a partir de publicações científicas. Campinas: Embrapa informática Agropecuária, 2016. 96 p. il. (Embrapa informática Agropecuária. Documentos, 142).

TAKEMURA, C. M.; SILVA, G. B. S. da; OLIVEIRA, S. R. de M.; MOURA, M. F. Desambiguação de topônimos usando dicionários geográficos. In: SEMINÁRIO DA REDE AGROHIDRO, 4., 2016, Brasília, DF. Água e agricultura: incertezas e desafios para a sustentabilidade frente às mudanças do clima e do uso da terra: anais. Planaltina, DF: Embrapa Cerrados, 2016. p. 67-74.

VACARI, I.; VISOLI, M. C.; GONZALES, L. E. Acesso aberto a informação científica agropecuária na internet: caso do sistema aberto e integrado de informação em agricultura (Sabiia). In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 8., 2011, Bento Gonçalves. Anais... Florianópolis: UFSC; Pelotas: UFPel, 2011. Não paginado.

OpenSLID: Uma ferramenta capaz de tornar mais eficientes os estudos sobre a dinâmica dos campos

César Huegel Richa[1], *Leandro B. S. Volk*[2,3], *José Pedro P. Trindade*[2,3], *Leonardo B. Pinho*[3]

[1] Programa de Pós-Graduação em Ciência da Computação, UFSC, Florianópolis-SC, Brasil, cesar.huegel@posgrad.ufsc.br

[2]Embrapa Pecuária Sul, Bagé-RS, Brasil, leandro.volk@embrapa.br, jose.pereira-trindade@embrapa.br

[3]Programa de Pós-Graduação em Computação Aplicada, UNIPAMPA/Embrapa, Bagé-RS, Brasil, leonardopinho@unipampa.edu.br

RESUMO

A busca por práticas pecuárias mais eficientes tem impulsionado o desenvolvimento e uso de técnicas de sensoriamento remoto (SR), inclusive, para um melhor entendimento das relações solo-planta-animal, dependentes do conhecimento da dinâmica dos campos em regiões específicas. Apesar de eficazes, métodos convencionais de classificação de cobertura vegetal baseados em SR possuem potencial de incremento na automatização, mas também em relação a eficiência, em função da grande quantidade de imagens e metadados a serem avaliados. Este trabalho adota elementos de pesquisa descritiva para apresentar a ferramenta OpenSLID, desenvolvida em conjunto pela EMBRAPA e pela UNIPAMPA , capaz de agilizar a manipulação de metadados e de realizar, de forma mais eficiente, a busca por imagens de interesse. Após sintetizar os seus requisitos funcionais e os principais aspectos da sua arquitetura, são expostos resultados preliminares do seu uso, tendo como estudo de caso imagens de uma região do Rio Grande do Sul historicamente associada à pecuária, permitindo concluir que aumenta a eficiência do processo de classificação ao maximizar o conjunto de imagens que atendem ao padrão de interesse estabelecido, bem como por evitar que o usuário busque manualmente, de forma repetitiva e suscetível a erro, as imagens do Projeto Landsat.

PALAVRAS-CHAVE: Ciência da Informação, Pecuária de Precisão, Sensoriamento Remoto, Classificação de Cobertura Vegetal, Ferramentas de Software Aberto.

ABSTRACT

The search for more efficient livestock practices has promoted the development and use of remote sensing (RS) techniques, including a better understanding of soil-plant-animal relationships, dependent on the knowledge of the dynamics of the fields in specific regions. Although effective, conventional methods of vegetation coverage classification based on RS have potential of increment in the automation, but also in relation to the efficiency, due to the great amount of images and metadata to be evaluated. This work adopts elements of descriptive research to present the OpenSLID tool, developed jointly by EMBRAPA and UNIPAMPA , able to speed up the manipulation of metadata and to perform, more efficiently, the retrieval of images of interest. After synthesizing its functional requirements and the main aspects of its architecture, preliminary results of its use, taking as a case study images referring to a region of Rio Grande do Sul historically associated with livestock, allow us to conclude that OpenSLID increases the efficiency of the classification process by maximizing the set of images that meet the established interest pattern, as well as by preventing the user from manually retrieving Landsat Project images repetitively and susceptible to error.

KEYWORDS: Information Science, Precision Livestock Farming, Remote Sensing, Vegetal Coverage Classification, Open source Software Tools.

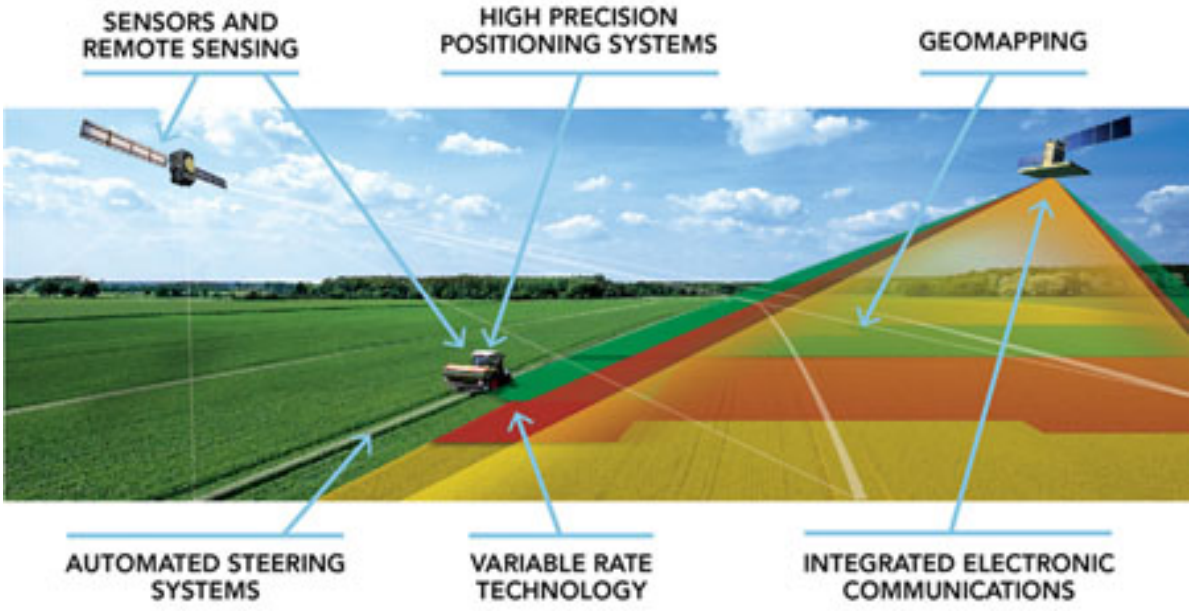
INTRODUÇÃO

Dados atuais apresentados pela CEMA (2017), Associação Europeia que representa a Indústria de Máquinas Agrícolas, apontam que a população mundial cresce em cerca de 240 mil pessoas por dia, podendo chegar a 9,6 bilhões em 2050, o que implica a necessidade de aumentar em 70%, de forma sustentável, a produção de alimentos. Neste contexto, aprimorar as atividades agrícolas é essencial para suprir este significativo crescimento na demanda por alimentos.

Em resposta a essa necessidade, surgiram, na comunidade científica e na indústria, diferentes ações de pesquisa, desenvolvimento e inovação envolvendo a aplicação de novas tecnologias em processos e produtos capazes de aumentar a eficiência dos sistemas de produção agrícola, alinhadas com o que se convencionou denominar por Agricultura de Precisão (AP). A Figura 1 ilustra um cenário de AP que combina conceitos e tecnologias de sensores e Sensoriamento Remoto (SR), sistemas de posicionamento de alta precisão, geoprocessamento, sistemas de colheita automatizada, tecnologias de taxa variável e de comunicações eletrônicas integradas.

Conforme a CEMA (2017), os principais desafios atuais da pecuária são os de (i) atender o crescimento da produção (estimulada pelo aumento no número de pessoas adotando dietas ricas em proteínas), (ii) aliviar a carga de trabalho dos envolvidos no manejo dos animais e assegurar a viabilidade econômica da propriedade rural (por meio de soluções escaláveis e custo-efetivas para o gerencimento de propriedades rurais), e (iii) promover a sustentabilidade e conforto animal (não apenas aumentar a produção, mas também encontrar maneiras de minimizar o impacto ambiental da pecuária e de assegurar altos níveis de bem-estar e de saúde para os animais).

Figura 1 – Agricultura de Precisão: tecnologias e conceitos-chave.



Fonte: (CEMA, 2017)

Neste contexto, Berckmans (2008) destaca o impacto rápido e crescente da Ciência da Informação e a Tecnologia da Informação nos métodos usados na pecuária, fazendo surgir uma subárea de AP, denominada por Pecuária de Precisão, ou *Precision Livestock Farming* (PLF). Na PLF são aplicados princípios de engenharia de controle na otimização da produção e dos processos de gerenciamento amparados por transferência de informação eletrônica. Assim, técnicas potencialmente inovadoras, simples e de baixo custo podem ser empregadas para coleta de dados dos animais e do ambiente, seguidas pela avaliação destes por modelos computacionais baseados em conhecimento. Para Melo, Silva e Guimarães (2013), existem diversas ferramentas que podem ser utilizadas para AP, as quais combinam sensores, sistemas de informação, máquinas adaptadas e gestão de conhecimento. Entre elas, destaca-se o SR, utilizado para obter informações de objetos sem entrar em contato com os mesmos, sendo capaz de detectar e armazenar radiação eletromagnética em determinada faixa do espectro eletromagnético (intervalo com todas as possíveis frequências da radiação eletromagnética), e ainda gerar informações que possam ser interpretadas, seja na forma de imagem, gráfica ou tabelas. Em função das suas características, o SR se apresenta como possível solução para a gestão de um dos elementos principais das relações solo-planta-animal que se estabelecem em um sistema de produção pecuária: os campos.

Entender a dinâmica dos campos na paisagem pode ser estratégica para diferentes objetivos, podendo impactar na definição de políticas públicas, em ações duráveis de manejo para a conservação da pecuária de campo e/ou intensificação sustentável da produção animal, bem como na definição de práticas de manejo que possam inverter o atual processo de conversão dos Campos (OVERBECK et al., 2007). O conhecimento sobre a dinâmica dos campos em regiões específicas mostra-se fundamental para as práticas de extensão rural, permitindo que se

estabeleça um processo de educação informal que objetive o uso racional dos recursos naturais e a prestação da assistência técnica pertinente, para a consolidação das práticas conservacionistas que garantirão a continuidade e a sustentabilidade do processo produtivo da propriedade (DIAS et al., 1999). Como exemplo, um método baseado em SR foi usado por Trindade et al. (2016), para descrever, a partir de imagens de satélite disponibilizadas gratuitamente pelo Projeto Landsat, a dinâmica dos Campos – das regiões da Campanha, Fronteira Oeste e Missões do Rio Grande do Sul – com potencial de utilização pastoril na Produção Pecuária, independente de seu estado de conservação, desde a década de 1970 até o Outono de 2015. Com este método, foi possível evidenciar que a cobertura vegetal campestre nesta área poderá ser fortemente reduzida em se mantendo as mesmas condições que levaram ao padrão de substituição da cobertura vegetal campestre nas três regiões.

Embora seja eficaz, o método baseado em SR usado por Trindade et al. (2016) para o estudo da dinâmica dos campos envolve diversas etapas manuais e repetitivas, com potencial de aprimoramento. Considerando-se a grande quantidade de imagens a serem processadas e a necessidade de um estudo contínuo, a automatização do método de classificação, ou de certas etapas deste, foi identificada como uma característica desejável e relevante para tornar o método mais eficiente. Neste contexto, este trabalho apresentada a ferramenta *Open source Smart Landsat Imagery Downloader* (OpenSLID), desenvolvida como parte das ações de cooperação entre a EMBRAPA e a UNIPAMPA , capaz de agilizar a manipulação de metadados e de realizar, de forma mais eficiente, a busca por imagens de interesse.

O restante do trabalho está dividido em três seções: na seção de Material e Métodos é apresentado o material utilizado para a condução do trabalho, bem como a metodologia empregada; na seção de Resultados e Discussão são relatados os resultados obtidos, discutidos em relação a outros trabalhos e resultados associados; e na seção de Conclusões são sintetizadas as principais conclusões.

MATERIAL E MÉTODOS

Este trabalho faz uso de elementos de pesquisa descritiva para apresentar a ferramenta de *software* OpenSLID, resultante de um projeto amparado por atividades de pesquisa exploratória, realizadas por uma equipe interdisciplinar, sob a coordenação de pesquisadores das áreas de Agrárias da EMBRAPA e de Computação da UNIPAMPA . Esta foi desenvolvida tendo como principal requisito tornar mais eficiente, com base em ferramentas *open source*, o método de classificação da cobertura de vegetação baseado em SR proposto por Trindade et al. (2016). Após um estudo inicial sobre o método e as oportunidades para torná-lo mais eficiente, optou-se por concentrar o processo de desenvolvimento em requisitos relacionados à pesquisa, *download* e pré-processamento de imagens de satélite Landsat.

Com a ferramenta OpenSLID, a pesquisa por imagens pode ser realizada a partir de diferentes filtros, o que possibilita a recuperação de cenas que contemplem parâmetros específicos

que as classifiquem como de interesse. O *download* de cenas pode ser realizado, tanto por meio de filtros quanto por meio de uma lista contendo os identificadores das cenas de interesse. Já o pré-processamento de imagens deve ser realizado a partir de cenas obtidas previamente e que estejam disponíveis na árvore de subdiretórios da ferramenta.

A utilização desta ferramenta pode ocorrer de forma supervisionada ou automatizada. No caso de supervisionada, o utilizador interage com a ferramenta realizando a pesquisa, o *download* e o pré-processamento. No caso de automatizada os utilizadores podem definir um catálogo, através de um arquivo de entrada, contendo *datasets*, para os quais deverão constar filtros de pesquisa e parâmetros de *download* e pré-processamento. Além da automatização dos processos supracitados, existe o conceito de execução programada, que permite manter um banco de imagens de interesse atualizado de acordo com o catálogo de *datasets*. O sistema em questão possui como principais funcionalidades a pesquisa e o *download* de imagens de satélite Landsat. Adicionalmente, o sistema provê um módulo de pré-processamento, onde diferentes conjuntos de operações sobre imagens de satélite podem ser implementadas de acordo com a necessidade da aplicação. É valido ressaltar que o sistema não se propõe a realizar processamento de imagens de aplicações específicas, tão pouco a disponibilizar ferramentas de visualização e edição das imagens de satélite.

Dentre os sistemas relacionados estão: QGIS (antigo Quantum GIS), que provê funcionalidades, através de sua API (*Application Programming Interface*), para implementação do conjunto de operações de pré-processamento de imagens; *Landsat Bulk Metadata Service*, serviço de disponibilização dos metadados referentes ao inventário Landsat disponibilizado pela *United States Geological Survey* (USGS). Existem, pelo menos, outras duas ferramentas similares, Landsat Util (2017) e Landsat-Download (2017) , ambas open source, desenvolvidas para busca e *download* de imagens de satélite, no entanto não possuem grande parte das características da OpenSLID, cujo desenvolvimento teve partes inspiradas nestas.

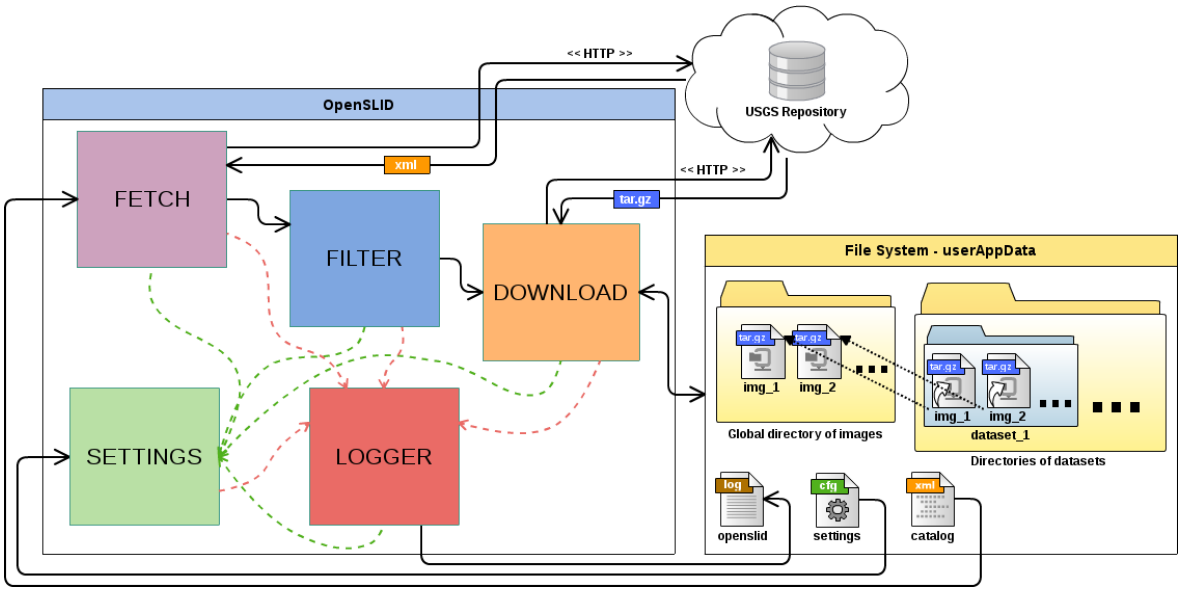
OpenSLID é destinada a usuários que utilizam massivamente imagens de satélite em seus projetos, principalmente quando existe a necessidade de análise temporal destas. Para estes usuários, em especial, a busca por imagens que satisfaçam determinados requisitos para utilização é uma tarefa custosa, pois demanda tempo, além da dificuldade de se avaliar os inúmeros metadados vinculados às inúmeras cenas de interesse e o tempo dispendido para download das imagens. Outro fator que intensifica a demanda por soluções que automatizem este processo é a necessidade de manter uma base de imagens de interesse atualizada, ou seja, a necessidade de realizar pesquisas periódicas ao repositório de imagens. A ferramenta viabiliza o uso compartilhado da base de imagens entre diferentes equipes e projetos por meio de um esquema de sincronização entre estações de trabalho e a base de imagens.

RESULTADOS E DISCUSSÃO

O protótipo da solução *open source* desenvolvida, OpenSLID, contempla atualmente a primeira etapa do processo de classificação de imagens que compreende a entrada (aquisição de imagens de satélite). Além de automatizar esta etapa, torna-a mais eficiente em termos de seus resultados (descoberta de imagens úteis para o estudo).

A arquitetura da solução desenvolvida, ilustrada na Figura 2, está dividida em cinco blocos de funcionalidades específicas: FETCH, responsável por realizar a leitura do arquivo "catalog.xml" (detalhado a seguir) e realizar a busca por metadados no repositório da USGS de acordo com os datasets e parâmetros de filtragem definidos no catálogo; FILTER, responsável por aplicar filtros específicos aos metadados resultantes da consulta ao repositório da USGS; DOWNLOAD, responsável por realizar o download das imagens que foram aceitas pelo filtro, salvá-las no diretório de imagens global e criar links simbólicos no diretório de imagens do dataset correspondente; SETTINGS, responsável por carregar as configurações setadas no arquivo "setting.cfg" e definir outras configurações padrões da ferramenta; LOGGER, responsável por gerenciar os níveis de logs (*debug, info, warning, error e critical*).

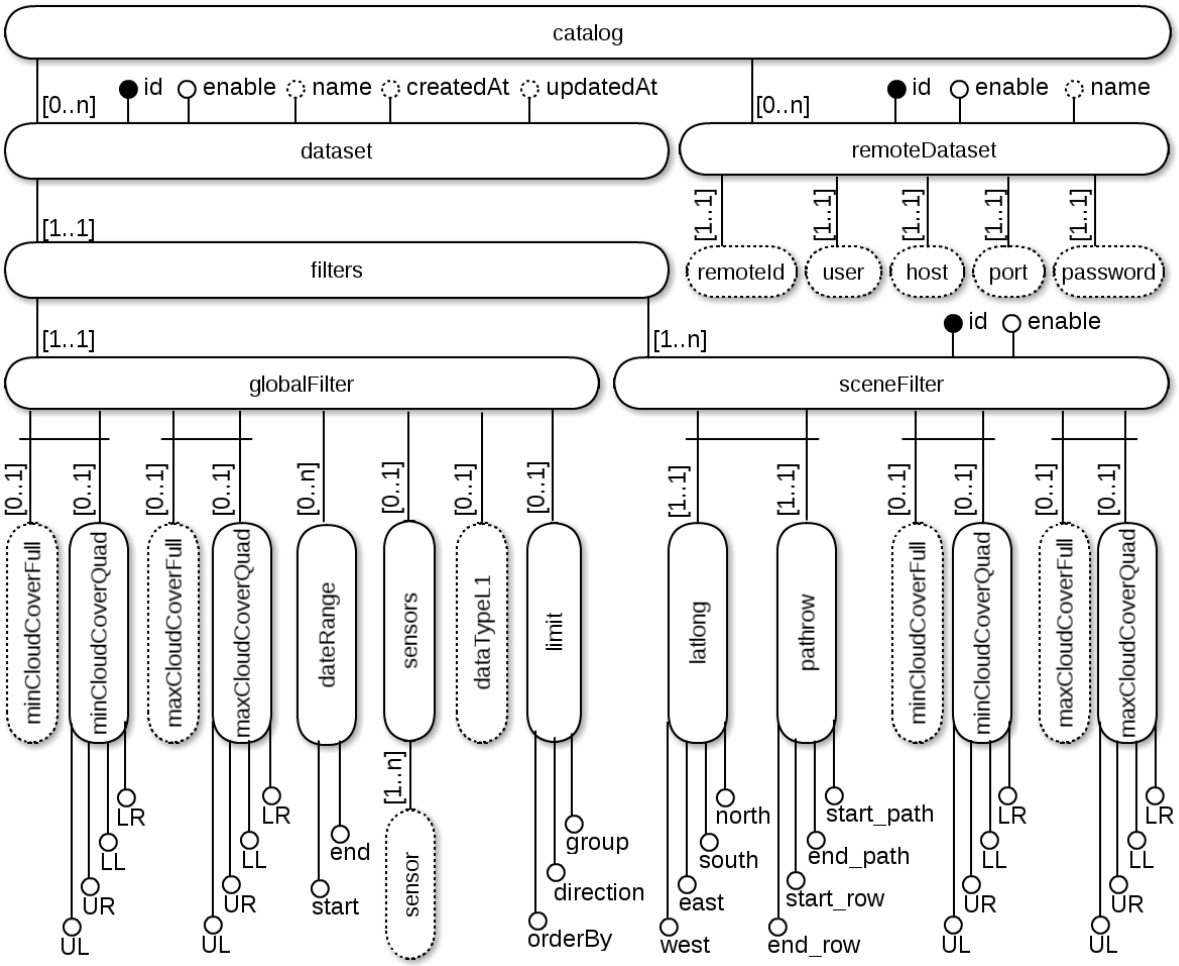
Figura 2 – Diagrama geral da arquitetura da solução.



Fonte: Próprios autores

A entrada da ferramenta consiste em um arquivo XML contendo um catálogo de datasets, onde estão definidos filtros específicos. Cada dataset corresponde a um agrupamento de cenas definidas por meio destes filtros, seguindo um modelo lógico hierárquico em XML que permite a representação de diferentes características referentes a estrutura e restrições. Dentre estas características estão: elementos simples e complexos; atributos simples; atributos identificadores; restrições de composição e cardinalidade entre elementos pai-filho; e ocorrência de disjunção entre elementos. A Figura 3 apresenta o modelo lógico XML destes dados e permite compreender seu nível de detalhamento.

Figura 3 – Modelo lógico XML.



Fonte: Próprios autores

Descreve-se a seguir os principais elementos do modelo XML:

- ***catalog***: elemento raiz do modelo. É um elemento complexo que pode conter até n ocorrências do elemento "dataset";
- ***dataset***: representa um conjunto de imagens de trabalho definidas por meio de filtros. Também é um elemento complexo e é composto por um único elemento "filters". Além disso, este elemento possui os seguintes atributos: "id", atributo único e obrigatório que identifica o elemento em relação aos demais; "enable", atributo obrigatório que indica se o *dataset* está ativo ou não; "name", atributo opcional arbitrário para determinar um nome ao *dataset*; "createdAt", atributo opcional que indica a data de criação do *dataset*; "updatedAt", atributo opcional que indica a data da última atualização das definições do *dataset*;
- ***filters***: elemento complexo que agrupa os filtros de um *dataset*. Este é composto por um único filtro global (globalFilter) e pelo menos um filtro de cena (sceneFilter). Foi criado pois o modelo será expandido e novas informações serão agregadas ao *dataset*;

- ***globalFilter***: elemento complexo onde são definidos parâmetros relacionados à filtragem global das imagens de um *dataset*. Sua composição é dada em função das seguintes restrições: pode conter um único elemento de mínima cobertura de nuvens (minCloudCoverFull ou minCloudCoverQuad); pode conter um único elemento de máxima cobertura de nuvens (maxCloudCoverFull ou minCloudCoverQuad); pode conter um único elemento simples de data inicial (startDate); pode conter um único elemento simples de data final (endDate); pode conter um conjunto de filtro para sensores (sensors); e um elemento filtro simples para o tipo de dado (dataTypeL1);
- ***sceneFilter***: elemento complexo onde são definidos parâmetros relacionados à filtragem de cena. Sua composição é dada em função das seguintes restrições: deve conter um elemento relacionado ao posicionamento geográfico das imagens (latlong ou pathrow); pode conter um único elemento para de mínima cobertura de nuvens (minCloudCoverFull ou minCloudCoverQuad); pode conter um único elemento de máxima cobertura de nuvens (maxCloudCoverFull ou minCloudCoverQuad);
- ***latlong***: elemento complexo que indica, por meio de seus atributos obrigatórios (north, south, east, west), um retângulo cuja função é delimitar as imagens cobertas pelo respectivo filtro para compor as imagens de um *dataset*;
- ***pathrow***: elemento complexo que indica, a partir de atributos obrigatórios (start_path, end_path, start_row, end_row), um conjunto de imagens por meio de suas referências de path e row.
- ***min,maxCloudCoverFull***: elementos simples que contém, respectivamente, um valor percentual mínimo e máximo de cobertura de nuvens em toda a imagem;
- ***min,maxCloudCoverQuad***: elementos complexos que possuem os seguintes atributos: UL (upper left), indica o quadrante superior esquerdo; UR (upper right), indica o quadrante superior direito; LL (lower left), indica o quadrante inferior esquerdo; e LR (lower right), indica o quadrante inferior direito. Estes dois elementos indicam, respectivamente, um valor percentual mínimo e máximo de cobertura de nuvens em cada quadrante da imagem;
- ***remoteDataset***: representa um *dataset* remoto, localizado em outra máquina (e.g. base compartilhada de imagens). É um elemento complexo composto por um elemento que identifica o *dataset* remoto e outros cujos dados são utilizados para efetuar a sincronização via protocolo *Secure Copy* (SCP). Seus atributos seguem o mesmo padrão do elemento "dataset".

Cabe destacar que o formato XML foi escolhido com base nas seguintes características: apresentar alto nível de padronização; ser baseado em texto simples, o que o torna de fácil leitura e edição; permitir validação da sua estrutura e de seus dados por meio de um XML Schema. Esta validação ocorre no início da execução do módulo FETCH, garantindo sua estrutura correta e a consistência dos dados contidos, já que foi definido em conformidade com o mesmo XML Schema dos metadados da USGS.

A principal contribuição da ferramenta está na precisão dos filtros (e.g. filtro de cobertura de nuvem por quadrante) aplicados aos metadados, maximizando o conjunto de imagens úteis resultante para um determinado estudo. Além disso, a ferramenta permite a definição de um *dataset* remoto, o que permite construir, dentro de uma organização, um servidor de imagens, contendo diferentes *datasets* para diferentes estudos, eventualmente de equipes distintas, mas otimizando a utilização de espaço de armazenamento e da largura de banda de rede com a execução agendada em horários pré-determinados e um único *download* para cenas coincidentes em diferentes *datasets*.

Tabela 1 – Levantamento de imagens Landsat 7 (CC <10% e tipo de correção L1T).

Path/Row	Período							
	1999-2000	2001-2002	2003-2004	2005-2006	2007-2008	2009-2010	2011-2012	2013-2014
223/81	10 / 15	7 / 11	0 / 6	0 / 7	1 / 3	2 / 3	0 / 0	0 / 0
223/82	10 / 16	4 / 8	1 / 4	3 / 5	2 / 4	4 / 5	1 / 1	0 / 0
224/80	2 / 10	5 / 12	8 / 9	2 / 3	2 / 3	3 / 6	2 / 3	0 / 0
224/81	3 / 9	3 / 9	1 / 10	1 / 2	0 / 2	5 / 6	2 / 3	0 / 0
224/82	3 / 7	1 / 6	1 / 6	2 / 4	2 / 1	4 / 6	1 / 2	0 / 0
225/80	2 / 12	5 / 12	1 / 10	1 / 6	2 / 2	5 / 7	0 / 2	0 / 0
225/81	7 / 11	4 / 8	0 / 8	1 / 6	2 /2	6 / 6	0 / 2	0 / 0

A Tabela 1 apresenta o número de imagens úteis identificadas manualmente e por meio da ferramenta, tendo como filtros a data, o *Path/Row* da cena, o percentual de cobertura de nuvem e o tipo de correção. Conforme pode ser observado nesta tabela, no período 1999-00 foram identificadas 37 imagens manualmente, enquanto que a ferramenta permitiu o uso de 80 imagens. Ou seja, neste período, o método manual usado originalmente descartava 43 imagens que poderiam sem usadas nas etapas posteriores no método de classificação. De forma geral, pode-se perceber que, com exceção de 2013-14 e de um possível erro humano identificado no período 2007-2008 para *Path/Row* 224/82, a quantidade de imagens que deixaram de ser descartadas foi significativa a partir do emprego da ferramenta desenvolvida.

Complementarmente, notou-se que o uso de XML foi um facilitador, prevenindo erros e simplificando a sua identificação por parte dos usuários. Isso se deve à existência do arquivo de validação, conforme o padrão XSD (*XML Schema Definition*), o qual descreve a estrutura do arquivo XML determinando quais elementos são obrigatórios, qual o número mínimo e máximo de ocorrências de determinados elementos, tipos de dados, dentre outras definições.

Por fim, destaca-se que o método usado por Trindade et al. (2016) é um caso de aplicação da ferramenta. Outras aplicações de SR que envolvam pesquisa, *download* e pré-processamento de imagens Landsat podem igualmente se beneficiar de suas características.

CONCLUSÕES

O emprego crescente da Ciência da Informação em soluções de Pecuária de Precisão se mostra cada vez mais promissor, contribuindo para o entendimento dos processos que interferem nas relações solo-planta-animal. Neste contexto o estudo da dinâmica dos Campos se mostra

fundamental, podendo este ser realizado a partir de métodos de Sensoriamento Remoto que envolvem a análise de grandes quantidades de imagens de satélite, passíveis de se tornarem mais eficientes com o uso de uma ferramenta computacional. A ferramenta OpenSLID, apresentada neste trabalho, demonstrou ser capaz de maximizar os resultados em termos do número de imagens úteis para o estudo, evitando erros humanos envolvidos no processo e otimizando o uso de espaço de armazenamento e da largura de banda de rede.

REFERÊNCIAS

BERCKMANS, D. Precision livestock farming (PLF). *Computers and Electronics in Agriculture*, v. 62, n. 1, p. 1, 2008.

CEMA. *Precision Farming: key technologies & concepts*. 2017. Acesso em: 11 mai. 2017. Disponível em: <http://www.cema-agri.org/page/precision-farming-key-technologies-concepts>.

DIAS, M. et al. *Manual de impactos ambientais: orientações básicas sobre aspectos ambientais de atividades produtivas*. Fortaleza, CE: Banco do Nordeste, 1999.

Landsat-Download. *Automated download of LANDSAT data from USGS website*. 2017. Acesso em: 11 mai. 2017. Disponível em: <https://github.com/olivierhagolle/LANDSAT-Download>.

Landsat Util. *A utility to search, download and process Landsat 8 satellite imagery*. 2017. Acesso em: 11 mai. 2017. Disponível em: <https://github.com/developmentseed/landsat-util>.

MELO, G. de A.; SILVA, S. F. da; GUIMARãES, A. M. Perspectivas de utilização do sensoriamento remoto na agricultura de precisão. In: *IX Congresso Brasileiro De Agroinformática*. Cuiabá, MT: [s.n.], 2013. Acesso em: 11 mai. 2017. Disponível em: <http://200.129.241.80/sbiagro/sbianais/paginas/trabalhos/118801_1.pdf>.

OVERBECK, G. E. et al. Brazil's neglected biome: the South Brazilian Campos. *Perspectives in Plant Ecology, Evolution and Systematics*, Elsevier, v. 9, n. 2, p. 101–116, 2007.

TRINDADE, J. P. P. et al. Dinâmica dos Campos da Campanha, Fronteira Oeste e Missões do Rio Grande do Sul. In: *XXVI Congresso Brasileiro de Zootecnia (ZOOTEC 2016)*. Santa Maria, RS: [s.n.], 2016. v. 26.

Resolution of NDVI for phenology monitoring of irrigated rice in Southwestern Brazil

Nelson Mario Victoria Bariani[1], *Cassiane Jrayj de Melo Victoria Bariani*[2]

[1] Universidade Federal do Pampa, Itaqui, Rio Grande do Sul, vbariani@gmail.com

[2] Universidade Federal do Pampa, Itaqui, Universidade Federal de Santa Maria, Cachoeira do Sul, Rio Grande do Sul, cassiane.victoria@gmail.com

ABSTRACT

Rice crop phenology derived from well established models and field observations was correlated, in this study, to the normalized difference vegetation index (NDVI) generated from satellite images. The crop cycle in 28 irrigated rice areas, in the southwestern region of Rio Grande do Sul, Brazil, was analyzed. The NDVI was computed from surface reflectance data of Landsat 8/OLI images for several dates throughout the crop cycle. A relation between NDVI and phenology was established, in which the resolution of the NDVI scale was determined as the Tukey test difference (95% confidence). The results show 0.04 NDVI units as a confident value for plant growth discrimination, corresponding to an interval of 7 days between images. This value was considered adequate for the remote discrimination of the main phenological stages of rice with NDVI in the development and late season periods.

KEYWORDS: NDVI calibration, rice phenology, irrigated rice monitoring, Landsat 8

INTRODUCTION

Agricultural management and climatic change are increasingly recognized as the major drivers of biodiversity change. Action to reduce anthropogenic impact on the environment and species within it will be most effective when targeted towards activities that have the greatest impact on natural lands (BURNS et al., 2016). Irrigated agriculture with high level management is seen as a way of optimizing productivity, and reducing the pressure for new areas.

Approaches based on remote sensing (RS) to support agricultural management allow the ability to monitor crop development over time and space, and have the potential of improving the estimation of the timing for phenology-dependent management actions which impact crop yield and indirectly determine the efficiency of water and land use for agriculture (LU, 2006; MATEOS et al., 2013).

There is an increasing tendency to perform detailed mapping of phenology and stages of vegetation growth by remote sensing, which demands spatialized and georreferenced data of appropriate quality to the needs of crop management, monitoring of natural resources and water use regulation, among other applications (CALERA et al., 2005; LU, 2006; ANDERSON et al., 2012). One of the most widely used remote sensing approaches to retrieve information related to the structural properties of vegetation and dynamics of plant growth is through the use of spectral vegetation indices (JONES & VAUGHAN, 2010), in particular the Normalized Difference Vegetation Index (NDVI; ROUSE et al., 1974).

One strategic value to be considered is the resolution of NDVI – the minimum detectable difference - which influences the expected sensitivity of the index to correlated operational variables, e.g. phenological stage. In the present study, we analyze the resolution of NDVI for different groups of rice cultivars and its influence in the sensitivity of different stages of the crop cycle.

Moreover, monitoring phenology and crop growth of different cultivars allows understanding of the actual crop development patterns and the identification of critical stages of crop development and its relation to various practices of crop management (VIÑA et al., 2004), including irrigation, nutrition and phytosanitary interventions. In general, these monitoring tasks are costly and time-consuming or frequently subject to a great uncertainty due to the extrapolation of point observations to large crops (BASTIAANSSEN, 2000).

Although there are many studies that correlate the least expensive radiometric information of RS with the more expensive field measurements of biophysical variables, there is a lack of studies reporting a numerical limit or resolution for the RS variables as NDVI. This is probably due to the intrinsically high random error present in most data from low spatial/high temporal resolution sensors (MODIS, AVRHS, SPOT Vegetation) generally used and also due to the natural variability of most crop or vegetation lands, which somehow concealed this limit.

Baghzouz et al. (2010) reported complete phenological phases recognition for semi-arid vegetation cover throughout the year using ground-based NDVI and compared those profiles with Landsat-NDVI profiles. The ground NDVI data was reported with centesimal precision

(0.01), although the satellite data was considered affected by bare soil influence. Water limitation and stress phases were detected. They also find highly significant linear relationships between NDVI and evapotranspiration (ET).

Scanlon et al. (2002) used monthly NDVI data at 8-km resolution and reported centesimal precision NDVI and millesimal precision sensitivity to precipitation, when analyzing multi-year AVHRR-NDVI data from southern Africa.

The analysis of the literature shows that there is a lack of information about the order of magnitude of NDVI absolute or relative resolution and the consequent sensitivity to vegetation changes in different stages of different cultivars.

In addition, there is some concern related to the independent variable used to describe the growing season, as it is the other component that determines the sensitivity. The day of the year is a natural and intuitive unit for management, but sometimes misleading in terms of profile recognition or direct correlation with biophysical variables as phenology. The length of the different periods of the crop cycle may vary due to uncertainties in sowing date, the time needed for emergency and climate conditions. Considering that limitation, Steinmetz et al. (2014) developed a model and corresponding program for calculation of the GDD in a given season based on meteorological data, returning as output the predicted dates for the appearance of phenological stages for several rice cultivars. The intervals defined by this model are well accepted by the scientific community and were used as a time reference for the chronology and duration of the phenological stages of rice.

This paper analyzes the application of NDVI calculated from Landsat 8 images to detect the phenological stages of rice in irrigated areas. This particular configuration of satellite and vegetation characteristics was studied to define critical operational parameters: the absolute resolution and its effect in the relative sensitivity of NDVI to the phenology changes during the crop stages. Hence, the objective of this work was to establish the value of resolution of NDVI for rice cultivars and to discuss its impact in the sensitivity for detection of phenology in different stages, through medium space resolution orbital remote sensing detectors, for its potential application to calibrate a locally adjusted curve for NDVI – phenology correlation.

MATERIALS AND METHODS

The study areas are located in the southwestern region of Rio Grande do Sul State, Brazil, mainly in the Municipality of Itaqui and neighboring ones, where rice crop is

predominant (Figure 1). Weather data is available from the National Institute of Meteorology (INMET), which has stations in Uruguaiana and São Borja. A typical soil in the region can be described as: "Brunizém Hidromórfico cálcico textura argilosa" ou "Chernossolo Ebânico Carbonático vértissólico", according to Brazilian Soil Classification (SCHNEIDER, 2008). Irrigated areas are dominant in the region, twenty-eight of which were selected for this work due to the presence of rice crops.

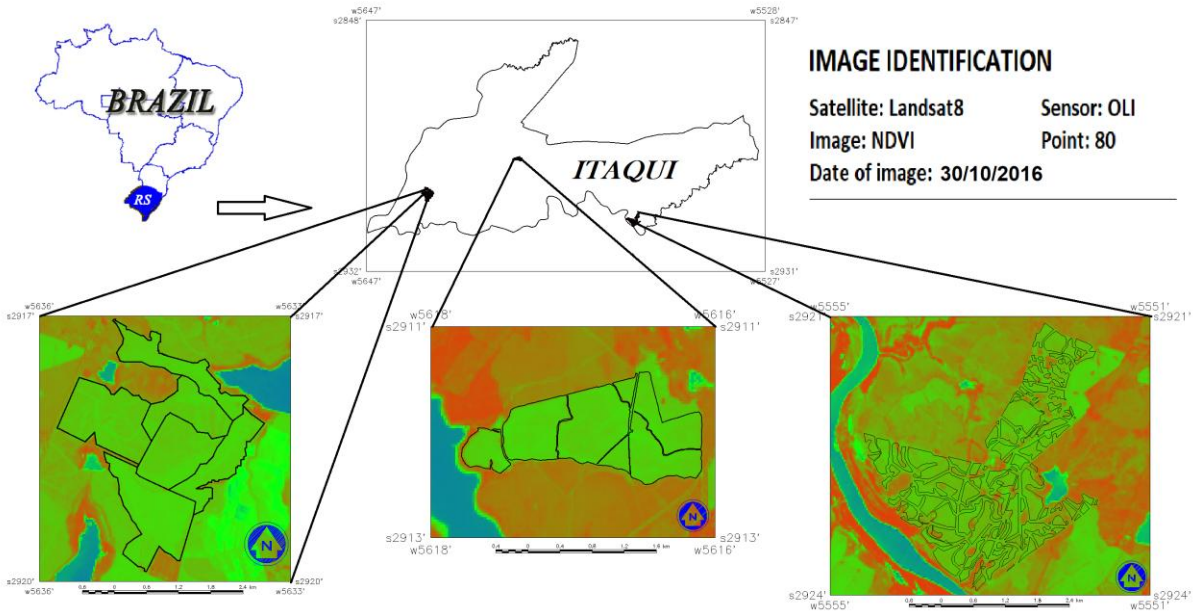


Figure 1- Municipal limits mesh of the study area (RS) and the maps of two regions with irrigated rice crops analyzed in this study.

Remote sensing data and products

The images used in this study were obtained from the OLI (Operacional Terra Imager) sensor on board of Landsat 8 satellite, for path/row 224/80 and 225/80, which in some cases overlap in the study area. The dates of image acquisition considered in this study cover the period from September 2016 to March 2017, totalizing 11 images as can be seen in Table 1.

Table 1- Information of the date (day/month) and path (for row 80) of the Landsat8/OLI images analyzed relative to 2016/2017.

Date	28/9	14/10	30/10	25/11	03/12	17/12	02/01	18/01	03/02	19/02	23/03
Path	225	225	225	224	225	225	224	224	225	225	224
Days after 1/1/2017	271	287	303	329	337	351	367	383	399	415	447

The digital processing of the bands of Landsat 8 images for each date was performed using the SPRING software, version 5.2.6 (CÂMARA et al., 1996), with projection set to UTM system with reference to WGS84 datum. Image processing included geometric

correction, geo-referencing adjustment to less than 0.5 pixel error, radiometric calibration, creation of image compositions for false color RGB543 combination, vector editing of irrigated area's limits, inspection of occurrence of clouds or image failures in the irrigated area areas, atmospheric correction, vectorization, dimensional reduction of the crop areas and finally NDVI calculation. Some routines in the LEGAL internal processing language have been adapted for the execution of image processing in SPRING GIS.

The conversion of the raw images pixel data (digital numbers) into radiance and reflectance was done using the procedure and calibration coefficients described in the L8 Handbook (USGS, 2016). The atmospheric correction of the images was made according to procedures adapted from Chavez (1996) and Hall et al. (1991). The NDVI was computed according to Rouse et al. (1974).

The average NDVI for each irrigated area was plotted against time for the visualization of the crop cycle, called hereafter NDVI curve.

Phenology

The phenological stages considered for rice were those proposed by Counce et al. (2000). They were grouped in stages with similar slope in the growth cycle, as proposed by FAO 56 paper (ALLEN et al., 1998) for irrigation management. In FAO-56, the crop season curve is divided into four growth stage periods: the initial, the development, the midseason and the late season. The initial and midseason periods are characterized by horizontal line segments and the development and late season periods are characterized by rising and falling line segments (Fig. 2). Additionally, the analysis of the features of the NDVI curve profile along the crop cycle (based on the red points of Fig.2) was considered for identification of the crops and groups of cultivars. The rice cultivars were characterized according to the duration of the cycle in the groups early (100-120 d), middle (121-150 d) and late (>150 d), adapted from SOSBAI (2016).

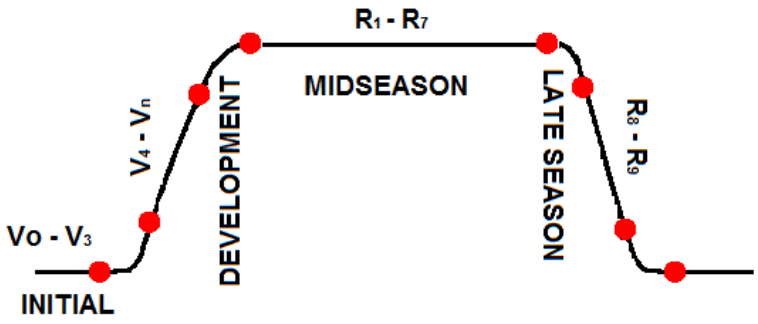


Figure 2- Periods defined in the NDVI curve, following the FAO 56 paper recommendations for curve. Phenology was considered according to GDArroz model and field observations.

The chronology of the phenological stages was obtained from the GDArroz program, developed by Steinmetz et al. (2014), based in an accumulated growing degree days (AGDD) scale that was correlated to days after emergence (DAE) or sowing (DAS) for easy interpretation. The exact timing of the phenology and its comparison with field information is not the focus of this work, but field information about the cultivars and phenology was obtained for some areas, directly from the technical staff involved in production, or through the field monitoring data of the Instituto Riograndense do Arroz (IRGA).

Resolution and sensitivity of NDVI to phenological stage identification

The resolution of NDVI - minimum significative detected difference in NDVI as the plants grow – is a key parameter to assess the applicability of NDVI for phenology monitoring. In this work the resolution of NDVI, calculated from L8/OLI images, for the discrimination of plant growth, was calculated using the Tukey's honestly significant difference (from Tukey´s HSD test) after an analysis of variance (ANOVA), to compare the average NDVI value of the pixels inside the cropland perimeters. The p-value, defined as the probability that the difference in NDVI could be originated by random variation and not due to real differences in plant growth (for 95% confidence interval) was considered as a confidence indicator for the results. The R statistical software (R D. C. T., 2011) was used for ANOVA and Tukey test execution, computing the significative differences between NDVI average values considering the date and pivot number as factors.

RESULTS AND DISCUSSION

ANOVA and Tukey Honestly Significant Difference Test

The results of the ANOVA test suggest that the NDVI average values for the irrigated rice areas were significantly different between successive image dates, with 95% confidence, and p-values ≈ 0. The Tukey HSD test allowed determining the lowest significative difference between irrigated area's average NDVI values. A minimal time interval of 8 days was observed between satellite images of 23/03/2017 and 30/03/2017; this interval corresponded to the minimum difference observed in this study between irrigated area's average NDVI values, around 0.04 NDVI units with p-value < $1x10^{-5}$. These results indicate that a resolution of 0.04 units can be considered as an achievable limit when analyzing the sensitivity of NDVI for the differentiation of biophysical variables in rice irrigated areas in a

sub-tropical climate. This can be used to determine the phenological stage sensitivity of NDVI. If the difference between two phenological stages of crops grown in irrigated lands is traduced in a NDVI difference greater than 0.04 units, the stages could be remotely discriminated through their NDVI values. This important quantitative basis for the use of NDVI was not found clearly established in the literature, but it is coherent with information by other authors as Hatfield & Prueger (2010) or Scanlon et al. (2002) or Wang et al. (2014). This sensitivity was favored by the good uniformity of crop development in irrigated fields, and by the great number of pixels that could be averaged for the calculations.

NDVI curves - Profile and sub-division of the curves

As an illustration of the NDVI resolution determination importance for remote sensing monitoring, we analyze here the NDVI curves, i.e., the pixel averaged NDVI values for each irrigated area at different image dates. The curves were grouped according to the typical profile of different rice cultivars (Fig.3). The characteristic profiles obtained for different crops or cultivars have also been recognized in previous studies e.g. Calera et al. (2004), Wang et al.(2016), Tasumi and Allen (2007), Martin et al. (2007).

Considering the NDVI resolution of 0.04 determined, we can conclude that the rapid growth edge (development period) of the NDVI curve and the senescence-maturation edge (late season period) have considerable greater sensitivity for phenological stages discrimination than the initial and middle period, where the NDVI variations are small. The development period generally includes stages V4-V13, which means 9 different phenological stages; the NDVI goes from around 0.4 to 0.8 for all groups. This means that we need (0.8-0.4)/9 ~0.04 as a NDVI resolution enough for the discrimination of all the phenological stages in the development period, which is equal to the significant resolution determined in this work. Similar conclusions arise for the late season period, with only two relevant phenological stages (R8 and R9) that can be easily discriminated by the NDVI difference. This is valid for all the cultivars, but the magnitude and rate of change of the sensitivity varies between different groups of cultivars. For the initial period the influence of the soil is dominant and makes it difficult to monitor the vegetative characteristics. For the midseason period, the discrimination of the phenological stages cannot be performed with the NDVI variable alone, but could be performed considering the temporal variable from the beginning of the period. The profiles of the curves for the various irrigated areas show similar characteristics for the same crop cultivar, as it was also observed by Tasumi and Allen (2007) and Hatfield & Prueger (2010) for other crops. Evident differences are visible between

different rice groups (Fig.3) where the time scale for each plot can be adjusted to DATE, DAS, DAE or GDD for comparison purposes.

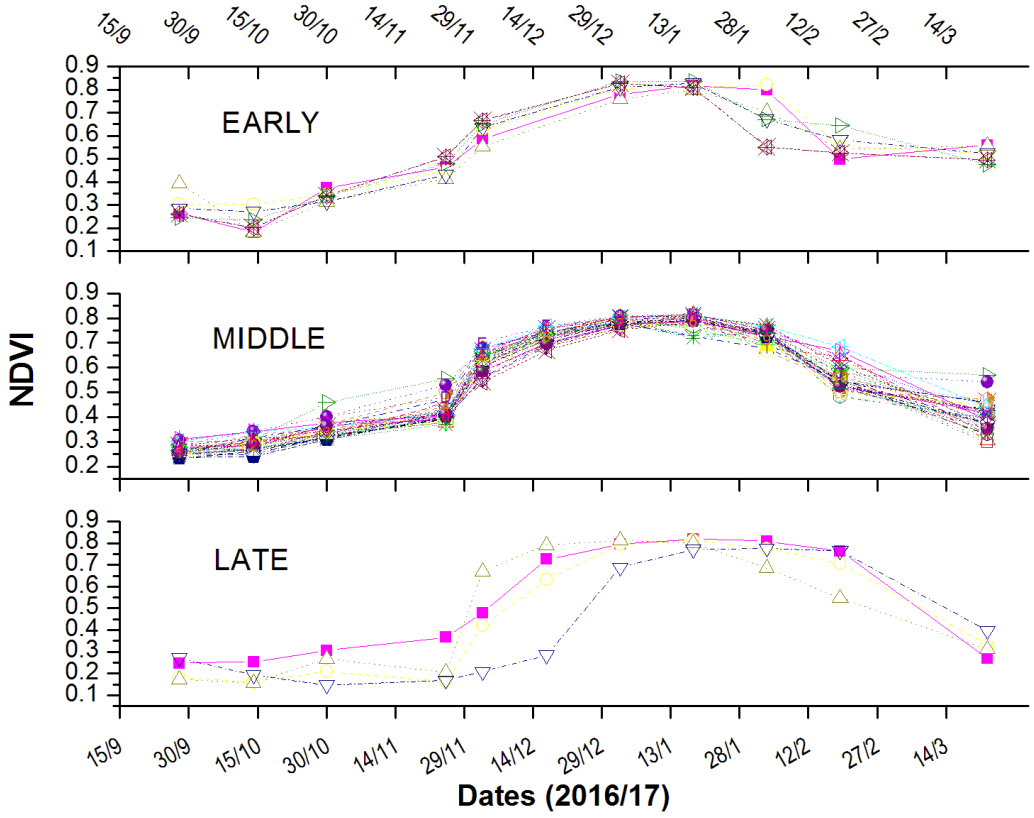


Figure 3- NDVI curves for rice maturation groups (early, middle and late) showing the NDVI variation with time for irrigated areas in the 2016/17 agricultural season. The graph symbols correspond to Landsat 8 images dates.

CONCLUSIONS

1) A local average NDVI vs. Time curve for each irrigated area considered in the study region was determined with a resolution of 0.04 NDVI units, considered as the Tukey significative difference for the NDVI sets of pixels of different Landsat 8/OLI images along the crop cycle. The curves were able to show the crop growth differences among cultivars.

2) The NDVI curves have enough sensitivity for phenological stages discrimination in the development and late periods. In the initial and midseason periods, the time variable would be necessary to predict the phenology, with reference to the beginning of each period. These observations can be valuable for a potential calibration of NDVI curves for the recognition of phenology.

ACKNOWLEDGEMENTS

The authors acknowledge UNIPAMPA, UFSM and IRGA for institutional support to this work. They also acknowledge the local enterprises Agropecuária Gomes, Pitangueira, Agrosul and CV Monitoramento for field information on rice crops.

REFERENCES

ALLEN, R. G.; PEREIRA, L.; RAES, D.; SMITH, M.. **Crop Evapotranspiration. Guidelines for Computing Crop Water Requirements.** FAO Irrig. Drain. Pap. 56, FAO, Rome, 1998.

ANDERSON, M.C., ALLEN, R.G., MORSE, A., KUSTAS W.P. Use of Landsat thermal imagery in monitoring evapotranspiration and managing water resources. **Remote Sensing of Environment,** v.122, p.50–65, 2012.

BAGHZOUZ, M., DEVITT D. A., FENSTERMAKER , L. F., YOUNG M. H. Monitoring Vegetation Phenological Cycles in Two Different Semi-Arid Environmental Settings Using a Ground-Based NDVI System: A Potential Approach to Improve Satellite Data Interpretation. **Remote Sensing**, v.2, p.990-1013, 2010. doi:10.3390/rs2040990

BASTIAANSSEN, W.G.M.; MOLDEN, D. J.; MAKIN, I. W. Remote sensing for irrigated agriculture: examples from research and possible applications. **Agricultural Water Management**, v. 46, p. 137-155. 2000.

BURNS F.; EATON M.A.; BARLOW K.E.; BECKMANN B.C.; BRERETON T., BROOKS D.R.; et al. (2016) Agricultural Management and Climatic Change are the Major Drivers of Biodiversity Change in the UK. PLoS ONE 11(3): e0151595. https://doi.org/10.1371/journal.pone.0151595

CALERA, A.; GONZÁLEZ-PIQUERAS, J.; MELIA, J. Monitoring barley and corn growth from remote sensing data at field scale. **International Journal of Remote Sensing**, v.25 p. 97-109, 2004.

CÂMARA, G.; SOUZA, R. C.; FREITAS, U. M.; GARRIDO, J. C. Integrating remote sensing and GIS by object-oriented data modeling. **Computers & Graphics**, v. 20 (3), 395-403, 1996.

CHAVEZ, J. Image-based atmospheric corrections – revisited and improved. **Photogrammetric Engineering and Remote Sensing**, v.62 (9), p. 1025-1036, 1996.

COUNCE, P.A. et al. A uniform, objective, and adaptative system for expressing rice development. **Crop Science**, Madison, v.40, n.2, p.436-443, 2000.

HALL, F.G., STREBEL, D.E., NICKESON, J.E., GOETZ,S.J. Radiometric rectification: toward a common radiometric response among multidate, multisensory images. **Remote Sensing of Environment**, v.35, n.1, p. 11-27, 1991.

HATFIELD, J. L.; PRUEGER, J. H. Value of using different vegetative indices to quantify agricultural crop characteristics at different growth stages under varying management practices. **Remote Sensing,** v. 2, p. 562-578, 2010.

JONES, H.G., VAUGHAN, R.A., 2010. **Remote sensing of vegetation. Principles, techniques, and applications.** Oxford University Press, Oxford.

LU, D. The potential and challenge of remote sensing-based biomass estimation. International Journal of Remote Sensing, Vol. 27, No. 7, 1297–1328, 10 April 2006.

MARTIN, K. L.; GIRMA, K.; FREEMAN, K. W.; TEAL, R. K.; TUBANA, B.; ARNALL, D. B.; CHUNG, B.; WALSH, O.; SOLIE, J. B.; STONE, M. L. Expression of variability in corn as influenced by growth stage using optical sensor measurements. Agronomy Journal, v. 99, p. 384-389. 2007

MATEOS, L.; GONZÁLEZ-DUGO, M. P.; TESTI, L.; VILLALOBOS, F. J. Monitoring evapotranspiration of irrigated crops using crop coefficients derived from time series of satellite images. I. Method. **Agricultural Water Management**, v.125, p. 81-91, 2013.

R DEVELOPMENT CORE TEAM (2011). **R: A language and environment for statistical computing.** R Foundation for statistical computing. Vienna, Austria, 2011.

ROUSE, J. W.; HAAS, R. H.; SCHELL, J. A.; DEERING, D. W. Monitoring vegetation systems in the Great Plains with ERTS. In: **Third Earth Resources Technology Satellite-1 Symposium. Proceedings.** Greenbelt: NASA, 1974.

SCANLON, T. M.; ALBERTSON, J. D; CAYLOR, K. K.; WILLIAMS, C. A. Determining land surface fractional cover from NDVI and rainfall time series for a savanna ecosystem. **Remote Sensing of Environment,** v. 82, p. 376-388. 2002.

SCHNEIDER, P. **Solos do Rio Grande do Sul.** Porto Alegre: EMATER/RS, 2008a. 222p. (UFRGS. Departamento de Solos. Faculdade de Agronomia).

SOSBAI, R. T. D. C. D. A. I. **Arroz irrigado:** recomendações técnicas da pesquisa para o Sul do Brasil. ISBN 978-85-69582-02-1. ed. Pelotas: [s.n.], 2016. 200 p. From: http://www.sosbai.com.br/docs/Boletim_RT_2016.pdf . Accessed in 21/06/2017.

STEINMETZ, S. et al. **GD arroz: programa baseado em graus-dia para estimar a data de diferenciação da panícula visando a adubação nitrogenada em cobertura**. Pelotas: Embrapa Clima Temperado, 2014. 12p. (Circular Técnica, 155).

TASUMI, M.; ALLEN, R. G. Satellite-based ET mapping to assess variation in ET with timing of crop development. **Agricultural Water Management**, v.88, p. 54-62, 2007.

USGS. Department of the Interior. U.S. Geological Survey. **Landsat 8 (L8) Data Users Handbook. Version 2.0.** EROS, Sioux Falls, South Dakota. March 29, 2016.

VIÑA, A.; GITELSON, A.; RUNDQUIST, D.; KEYDAN, G.; LEAVITT, B.; SCHEPER, J. Monitoring maize (Zea mays L.) phenology with remote sensing. **Agronomy Journal**, v. 96, p. 1139-1147, 2004

WANG, R.; CHERKAUER, K.; BOWLING, L. Corn response to climate stress detected with satellite-based NDVI time series. **Remote Sensing**, v. 8, p. 1-22. 2016.

WANG, L. et al. Multi-Temporal Detection of Rice Phenological Stages Using. **Rice Science**, 21, n. 2, 2014. 108-115.

Sensitivity of NDVI for remote sensing monitoring of soybean stages under irrigation in Southern Brazil

Cassiane Jrayj de Melo Victoria Bariani[2] ,*Nelson Mario Victoria Bariani*[1]

[1] Universidade Federal do Pampa, Itaqui, Rio Grande do Sul, vbariani@gmail.com

[2] Universidade Federal do Pampa, Itaqui, Universidade Federal de Santa Maria, Cachoeira do Sul, Rio Grande do Sul, cassiane.victoria@gmail.com

ABSTRACT

In this study, the normalized difference vegetation index (NDVI) generated from satellite images was applied to study the cycle of growth of soybean crop in 10 irrigated perimeters in the southern region of Rio Grande do Sul, Brazil. Surface reflectance data of LANDSAT8/OLI images was used to monitor the crop cycle. Well established published information about the regional timing of soybean phenology together with the NDVI vs time curve were used to define a relation between NDVI and phenology. The variation of 0.02 NDVI units during plant growth was found significative in this study. That sensitivity was enough for the correlation of NDVI to the phenological stages of soybean during the crop cycle, defining an accessible boundary for soybean monitoring by remote sensing in irrigated lands.

KEYWORDS: NDVI calibration, soybean phenology, agriculture monitoring, Landsat8.

INTRODUCTION

Remote sensing can give support to the management of irrigation through the spatialization of crop development over time and space, with reduced field monitoring (KUENZER and KNAUER, 2013; MATEOS et al., 2013; OZDOGAN et al., 2010). From this point of view, there is an increasing interest in mapping land surface phenology (LSP) and growth stages of vegetation by remote sensing (GAO et al., 2017; ZHANG et al., 2017; WANG et al., 2014; HENEBRY and BEURS, 2013). One of the most widely used spectral

vegetation indices (JONES and VAUGHAN, 2010) is the Normalized Difference Vegetation Index (NDVI; ROUSE et al., 1974).

Monitoring phenology and crop growth allows understanding of the actual crop development patterns and identifying critical stages of crop development in relation to various practices of crop management (VIŇA et al., 2004; WANG et al., 2014), including irrigation, nutrition and phytosanitary interventions. In general, these tasks are costly and time-consuming when based in field monitoring, although still associated with a great uncertainty due to the extrapolation of point observations to large crops (BASTIAANSSEN, 2000).

Concerning the monitoring time scale, the growing degree days (GDD), which consider differences in growing season temperature, have been often used as an auxiliary variable for the definition of calendar days from planting or emerging date, conditioned by the availability of a network of suitable meteorological stations or alternative source for temperature measurements.

The GDD is useful for better adjustment of biophysical or operational variables timing, e.g., phenology time intervals, maturity length, duration of crop cycle, crop coefficients and evapotranspiration, between others (STRECK et al., 2006; STEINMETZ et al., 2010; NIELSEN and HINKLE, 1996; HUNSAKER et al., 2003; FARAHANI et al., 2008).

This paper presents a case study of the application of NDVI calculated from Landsat images to detect the phenological stages of soybean in center-pivot irrigated areas and defines the duration time of the crop growth stages. This can be used to define the locally adjusted curve of the crop coefficient. The ultimate goal is to contribute to provide an operational basis for space-assisted local agricultural management related to support for irrigation modeling and management (BARIANI et al, 2016).

MATERIALS AND METHODS

The study area is located in Cruz Alta, in the northwestern region of Rio Grande do Sul State, Brazil , between 28° 34' to 28° 45" S and 53° 30' to 53° 14' W (Figure 1). This is a subtropical humid region, with mean annual precipitation of 1755 mm. Rainfall is distributed across the four seasons. Considering a 30 years series of data, the average monthly rainfall is 135 mm with standard deviation of 20 mm. Water stress occurs in the summer due to increased evaporative demand. The maximum temperature (37-38°C) occurs from December to March and the minimum (0 to -4.5 °C) occurs from May to September. Weather data is available from the National Institute of Meteorology (INMET), which has a station located inside the

study site (28.63°S, 53.6°W and 472.5 m altitude). Center-pivots irrigated areas are dominant in the region, twenty-eight of which were selected for this work due to the presence of soybean or maize crops. The area of the irrigated fields varied between 10 and 100 hectares and the plants density ranged from 260.000 to 340.000 plants ha^{-1} for soybean.

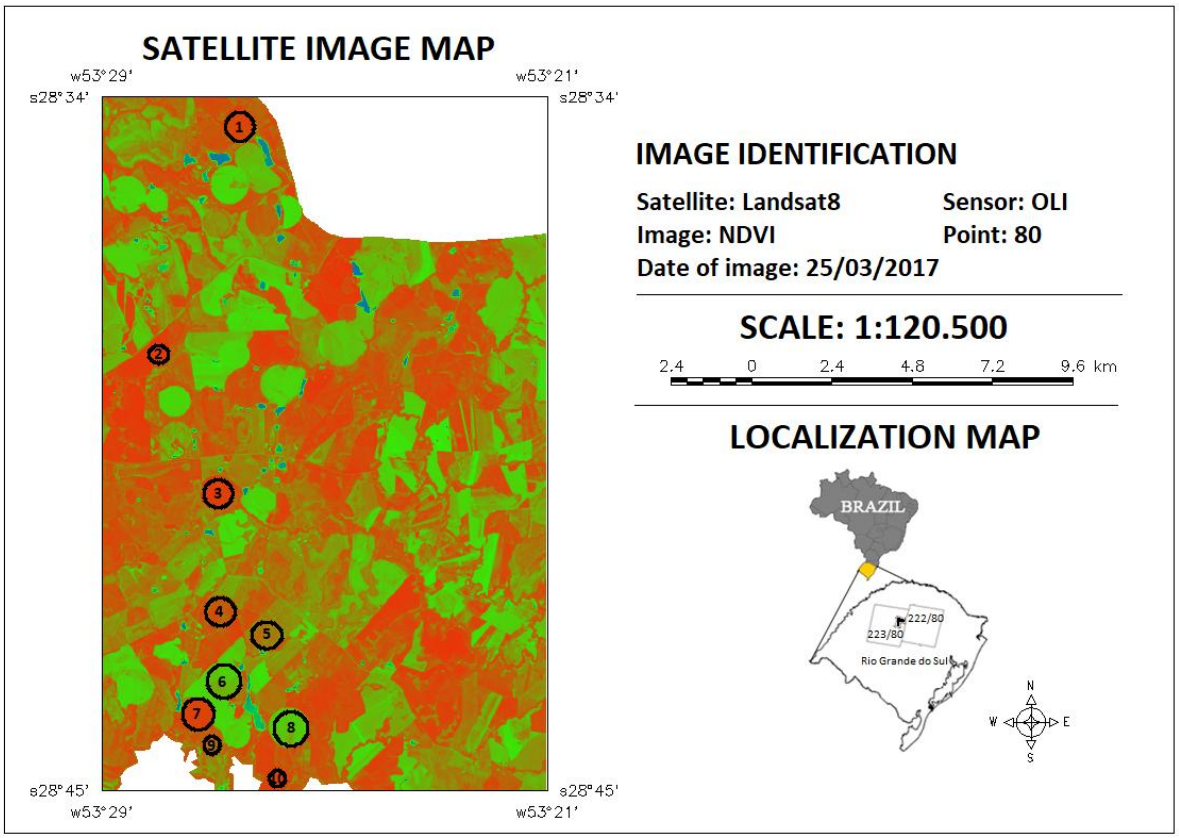


Figure 1. Location and numbering of the center-pivots in Cruz Alta, Rio Grande do Sul. The location map on left includes the Landsat paths 222 and 223, row 80, which cover the Cruz Alta area.

Remote sensing data and products

The images used in this study were obtained from the OLI (Operational Land Imager) sensor aboard Landsat 8 for path/row 222/80 and 223/80, which overlap in the study area. The dates of image acquisition considered in this study cover the period from September 2016 to June 2017, totalizing 15 images as can be seen in Table 1.

Table 1. Information of the path and date (for row 80) of the Landsat8/OLI images analyzed relative to 2016/2017. *Day of Year

Date	14/09	07/10	23/10	24/11	26/12	4/1	11/1	20/1	27/1	28/2	9/3	25/3	1/4	17/4	13/6
Path	222	222	223	223	223	222	223	222	223	223	222	222	223	223	222
DOY*	258	274	281	329	361	4	11	20	27	59	68	84	91	107	164

The digital processing of the bands of Landsat 8 images for each date was performed using the SPRING software, version 5.2.6 (CÂMARA et al., 1996), with projection set to UTM system with reference to WGS84 datum. Image processing included geometric correction, geo-referencing adjustment to less than 0.5 pixel error, radiometric calibration, vector editing of irrigated area's limits, inspection of occurrence of clouds or image failures in the irrigated area areas, atmospheric correction, vectorization and clipping of the crop areas and finally NDVI calculation. Some routines in the LEGAL internal processing language have been adapted for the execution of image processing in SPRING GIS.

The conversion of the raw images pixel data (digital numbers) into radiance was done using the procedure described in the L8 handbook (USGS, 2016). The atmospheric correction of the images was made according to Chavez (1996) or Hall (1991). The NDVI was computed according to Rouse et al. (1974).

The average NDVI for each irrigated area was plotted against time for the visualization of the crop cycle, called hereafter NDVI curve. Additionally, the qualitative analysis of the NDVI curve profile along the crop cycle showed a characteristic "fingerprint" that allowed identification of the crop.

Cultivars and Phenology

In Brazil, the cultivar classification based on the maturation cycle is based in the accumulated GDD until flowering. The GDD concept is related to the accumulation of temperature (average between maximum and minimum daily temperature) over a base value that is considered of 14°C for soybean. For soybean the accumulated GDD for maturation is considered higher than 1000 (CAMARGO et al. ,1987).

The phenological stages considered for soybean were those proposed by Fehr & Canivess (1977). The stages considered were the vegetative and reproductive stages, represented by V and R, followed by numerical indices that identify the successive stages. Emergency was designated as VE, and cotyledon stages for soybean were designated VC. V0 was used for pre-emergence stages.

NDVI curve profile analysis

Derivative analysis of the NDVI vs. DAS curve was also used as a measure of the variation of the NDVI with time and as a help for the determination of growth periods and phenological stages with mathematical support. The derivative was calculated using the approximation $(y_n-y_{n-1})/(x_n-x_{n-1})$ for daily NDVI data obtained by interpolation. A positive/negative value indicated the amount of increase/decrease in NDVI per day, and the variability in this rate was analyzed in the time or GDD plot of the derivative.

The accumulated growing degree days (AGDD), which consider cumulative differences in growing season temperature, has been often considered for better adjustment of operational variables like Kc to actual conditions (NIELSEN & HINKLE, 1996; HUNSAKER et al., 2003; FARAHANI et al., 2008).

RESULTS AND DISCUSSION

Geographic data base and image processing

The complete set of NDVI thematic images (Figure 2) shows the temporal evolution of the NDVI values inside the center-pivots' areas. The dates correspond to the satellite imaging day. The label soybean appears in the dates in which the crop was present in the center-pivot, giving an idea of the extent of the crop cycle. The presence of sectors in some center-pivots was generally associated to management processes, e.g., sowing or harvesting, that were in progress in that date. These images were useful for checking the causes of anomalous behavior in individual NDVI curves (pixels average) along time, and contributed to the adjustment of representative vectorial limits of the crop inside the center-pivot area, which were essential for a precise calculation of the NDVI statistical parameters.

The radiometric classification procedure and the NDVI vs. DAS profile recognition procedure allowed the identification of soybean crop in 10 center-pivots. The cross-linking of the information from both procedures allowed a lower uncertainty for the crop recognition.

The descriptive statistics applied to the NDVI values of the pixels in each center-pivot showed a reasonable Gaussian behavior and a maximum value of 0.1 for the standard deviation, compatible to results by Kamble et al. (2013) or Maxwell & Sylvester (2012).

ANOVA and Tukey Honestly Significant Difference Test

The results of the ANOVA test allowed indicating that the NDVI average values for the irrigated areas were significantly different between successive image dates, with 95% confidence, and p-values ≈ 0. The Tukey HSD test allowed determining the lowest significative difference between irrigated area's average NDVI values. A minimal time interval difference was observed in this study between irrigated area's average NDVI values ranging from 0.01 to 0.03 NDVI units with p-value of $3x10^{-7}$. These results indicate that a value of sensitivity of 0.02 units can be considered as an achievable limit of sensitivity for the differentiation of NDVI values in irrigated area irrigated areas for soybean crops in a sub-tropical climate.

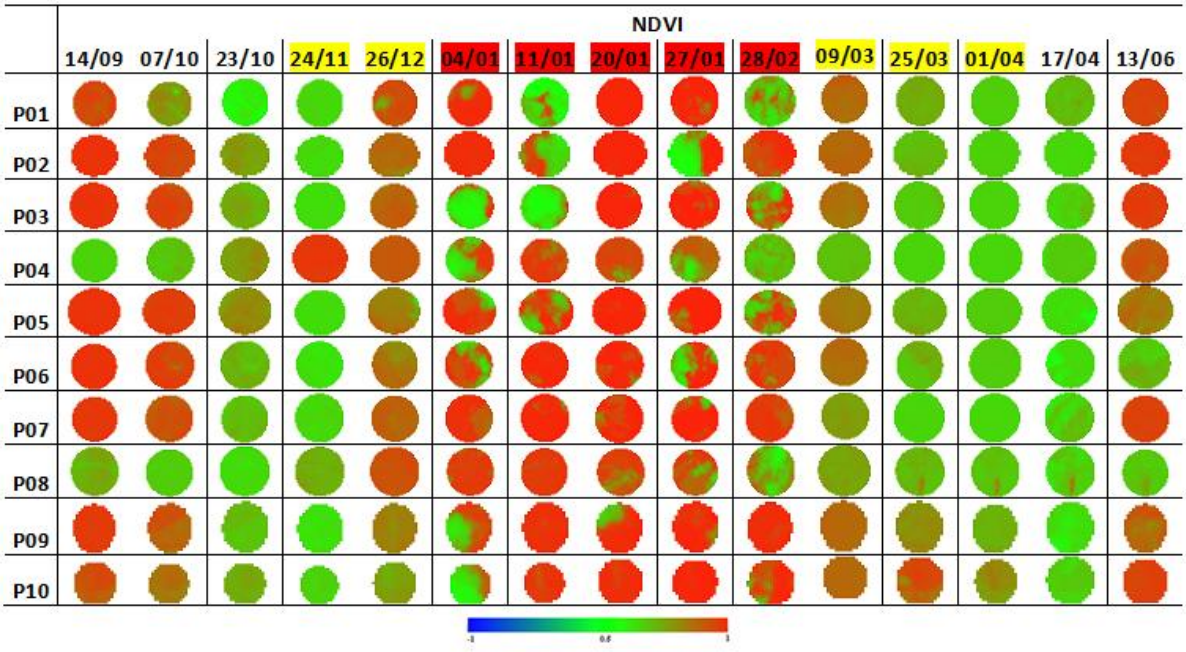


Figure 2. Graphical representation of the NDVI values for all considered center-pivots cropped with soybean, ranging from green (vegetation under active growth) to red (full development and maturation) for various satellite path dates.

This can be related to the phenological stage sensitivity. If the difference between two phenological stages of crops grown in central pivots is traduced in a NDVI difference greater than 0.02 units, the stages could be remotely identified through their NDVI values. This important result was not found clearly established in the literature, but it is coherent with information by other authors as Hatfield & Mayers (2010) or Scanlon et al. (2002). This good sensitivity was favored by the greater uniformity of crop development in irrigated fields, and by the great number of pixels that could be averaged for the calculations.

NDVI vs. DAS curves

Profile of the curves

Using the pixel averaged NDVI values for each irrigated area at different image dates, individual NDVI curves were constructed for soybean (Fig. 3). The characteristic profiles seen for each crop in this study have also been observed in previous studies e.g. Calera et al. (2004), Wang et al.(2016), Tasumi and Allen (2007), Martin et al. (2007).

Phenology and NDVI

The single NDVI curves of the soybean cultivars cropped in each specific center-pivot were averaged to obtain a representative curve in a DAS (or GDD) time scale. The mean of the multiple curves, after adjusting for coincidence of the fast growing edge, produced an NDVI curve representative of the regional agricultural practice (Figure 4). This curve can be used for modeling or monitoring purpose, e.g., as a reference for NDVI-Phenology calibration, or

Table 2. Output of Tukey's HSD test for the identification of significative differences between irrigated areas' average NDVI in different dates. The complete sets of crop pixels for each pivot were used. A sensitivity of 0.02 was the value identified. The p value indicates that all dates showed significant difference in NDVI with 95% confidence.

Dates		Difference	Lower Limit	Upper Limit	p value
24/11/2016	25/03/2017	-0.08	-0.09	-0.07	0
26/12/2016	09/03/2017	-0.15	-0.16	-0.14	0
24/11/2016	09/03/2017	-0.17	-0.18	-0.16	0
27/01/2017	25/03/2017	0.46	0.45	0.47	0
24/11/2016	04/01/2017	-0.39	-0.40	-0.38	0
27/01/2017	04/01/2017	0.54	0.53	0.55	0
28/02/2017	04/01/2017	0.39	0.38	0.40	0
27/01/2017	24/11/2016	0.63	0.62	0.64	0
28/02/2017	24/11/2016	0.48	0.47	0.49	0
20/01/2017	27/01/2017	-0.02	-0.03	-0.01	0

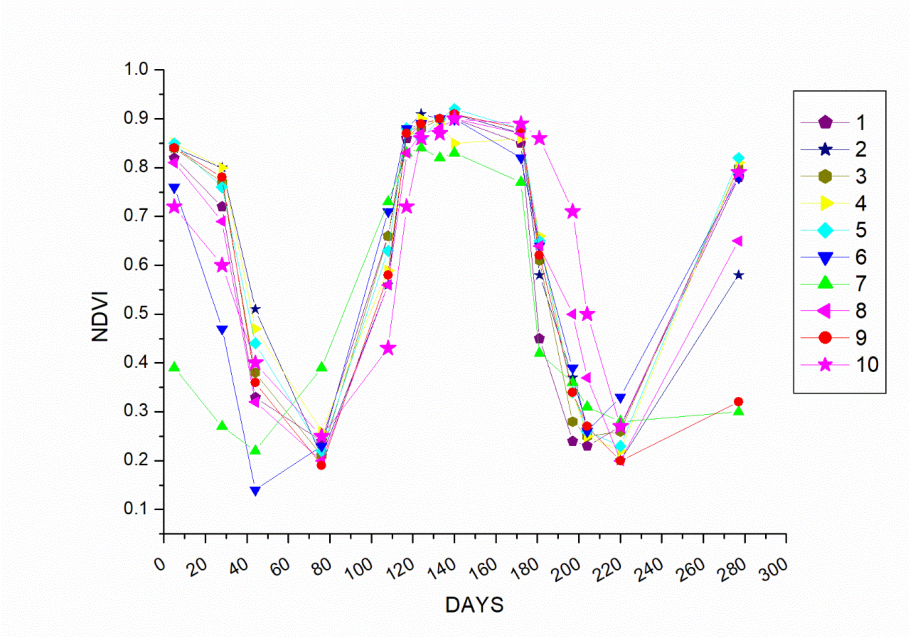


Figure 3. NDVI curves for soybean showing the NDVI variation with time (Days after 09/09/2016) for 10 irrigated areas in the 2016-2017 agricultural season.

to identify different growing patterns or stress that could affect productivity (Wang et al., 2016). In this work the curve was associated with the phenological information obtained from the field monitoring database of the irrigation management system that operates in the study region. Results shown in Fig. 4 for soybean are compatible with those reported by Hatfield & Prueger (2010), Martin et al. (2007) and Raun et al. (2005). Important to notice that those

curves are smoothed and monotonic, without noise in all cases. Tasumi and Allen (2007) found similar curves from 3095 irrigated fields. Other studies used filters with monotonic criteria to eliminate low quality data (WANG et al., 2016).

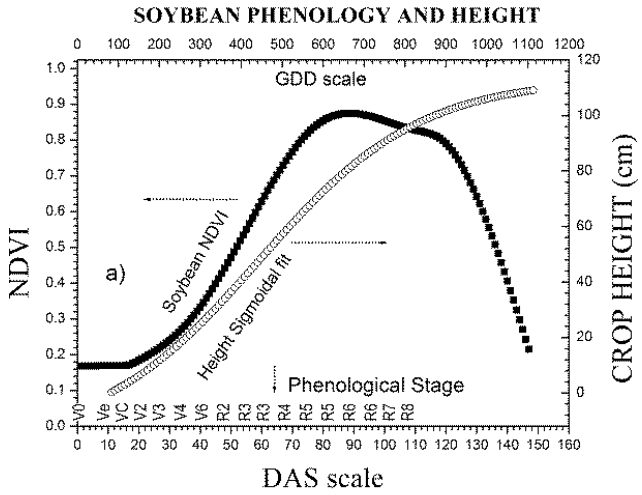


Figure 4 - Relation between NDVI, Phenology and Days after Sowing for Soybean. The height data was adjusted to a sigmoid, and consisted with the phenology and NDVI as a function of DAS and GDD.

CONCLUSIONS

1) A local average NDVI vs. DAS or AGDD curve for each irrigated area considered in the study region was determined, showing the crop cycle. These curves showed a kind of fingerprint of the local crop cultivars and agricultural management habits.

2) A locally calibrated phenology (NDVI, DAS or GDD) function, with sensitivity of 0.02 NDVI units was obtained for the recognition of phenology, allowing the phenological stage monitoring.

3) The determination of the periods of the crop development stages recommended by FAO56 irrigation management schedule was obtained based on NDVI-DAS (or GDD) curve. The intervals of NDVI values per period were obtained.

ACKNOWLEDGEMENTS

The authors acknowledge UNIPAMPA and UFSM for institutional support to this work.

They also acknowledge IRGA and the local enterprises Agropecuária Gomes, Pitangueira, Agrosul and CV Monitoramento for field information on rice crops.

REFERENCES

BASTIAANSSEN, W. G. M. et al. **Remote sensing for irrigated agriculture:** examples from reserch and possible applications. International Water Management. Sri Lanka, Colombo: [s.n.]. 2000. p. 137-155.

BARIANI, C. J. M. V. Combinação de modelos de balanço hídrico no solo e sensoriamento remoto para o monitoramento de áreas irrigadas. **Tese**. 143p.

CALERA, A. B. et al. Irrigation management from space: towards user-friendly products. **Irrig. Drain.**, Syst., v. 19, p. 337-353, 2005.

CALERA, A.; GONZÁLEZ-PIQUERAS, J.; MELIA, J. Monitoring barley and corn growth from remote sensing data at fild scale. **International Journal of Remote Sensing**, 2004. 97-109.

CÂMARA, G. et al. Intergrating remote sensing and GIS by object-oriented data modeling. **Computers & Graphics**, v. 20, n. 3, p. 395-403, Maio 1996.

CHAVEZ, J. Image-based atmospheric corrections – revisited and improved. **Protogrammetric Engineering and Remote Sensing**, 62, n. 9, 1996. 1025-1036.

FARAHANI, H.J., OWEIS, T.Y., IZZI, G. Crop coefficient for drip-irrigated cotton in a Mediterranean environment. **Irrigation Science** v. 26, p.375-383, 2008.

FEHR, W. R. B.; NEUMAIR, N.; NEPOMUCENO, A. L. **Stages of soybean development**. [S.l.]: Ames: Lowa State University, 1977.

GAO, F.; ANDERSON, M.C.; ZHANG X.; YANG Z.; ALFIERI, J.G.; KUSTAS, W.P.; MUELLER, R.; JOHNSON, D.M.; PRUEGER, J.H. Toward mapping crop progress at field scales using Landsat and MODIS imagery. **Remote Sensing of Environment,** 188 (2017) 9–25.

HALL, F.G., STREBEL, D.E., NICKESON, J.E., GOETZ,S.J. Radiometric rectification: toward a common radiometric response among multidate, multisensory images. **Remote Sensing of Environment**, v.35, n.1, p. 11-27, 1991.

HATFIELD, J. L.; PRUEGER, J. H. Value of using different vegetative indices to quantify agricultural crop characteristics at different growth stages under varying management practices. **Remote Sensing,** v. 2, p. 562-578, 2010.

HUNSAKER, D. J. et al. Estimating cotton evapotranspiration crop coefficients with a multispectral vegetation index. **Irrigation Science**, v. 22, p. 95-104, 2003.

JONES, H.G., VAUGHAN, R.A., 2010. **Remote sensing of vegetation. Principles, techniques, and applications.** Oxford University Press, Oxford.

KAMBLE, B.; IRMAK, A.; HUBBARD, K. Estimating Crop Coefficients Using Remote Sensing-Based Vegetation Index. **Remote Sensing**, 2013. 1588-1602.

KUENZER, C ; KNAUER, K. Remote sensing of rice crop areas, International Journal of Remote Sensing, 34:6, 2101-2139, 2013. DOI: 10.1080/01431161.2012.738946

LU, D. The potential and challenge of remote sensing-based biomass estimation. International Journal of Remote Sensing, Vol. 27, No. 7, 1297–1328, 10 April 2006.

MARTIN, K. L.; GIRMA, K.; FREEMAN, K. W.; TEAL, R. K.; TUBANA, B.; ARNALL, D. B.; CHUNG, B.; WALSH, O.; SOLIE, J. B.; STONE, M. L. Expression of variability in corn as influenced by growth stage using optical sensor measurements. Agronomy Journal, v. 99, p. 384-389. 2007

MATEOS, L. et al. Monitoring evapotranspiration of irrigated crops using crop coefficients derived from time series of satellite images.I. Method. **Agriculture Water Mamagement**, 125, 2013. 81-91.

MAXWELL, K.S.; SYLVESTER, M. K. Identification of "ever-cropped" land (1984–2010) using Landsat annual maximum NDVI image composites: Southwestern Kansas case study. **Remote Sensing of Environment**, v. 121, p. 186-195, 2012.

NIELSEN, D.C., HINKLE, S.E. Field evaluation of basal crop coefficients for corn based on growing degree days, growth stage, or time. **Transactions of the ASAE. American Society of Agricultural Engineers**, v39, p. 97-103, 1996.

OZDOGAN, M.; YANG, Y.; ALLEZ, G.; CERVANTES, C. Remote sensing of irrigated agriculture: Opportunities and challenges. *Remote sensing*, *2*(9), 2274-2304, 2010.

RAUN, W.R., SOLIE, J.B.; MARTIN, K.L.; FREEMAN, K.W.; STONE, M.L.; JOHNSON, G.V.; MULLEN, R.W. Growth stage, development, and spatial variability in corn evaluated using optical sensor readings. **J. Plant Nutrition**, v. 28, p.173-182. 2005.

ROUSE, J. W. et al. **Monitoring Vegetation Systems in the Great Plains with ERTS**. Third Earth Resources Technology Satellite-1 Symposium. Greenbelt: NASA. 1974.

SCANLON, T. M.; ALBERTSON, J. D; CAYLOR, K. K.; WILLIAMS, C. A. Determining land surface fractional cover from NDVI and rainfall time series for a savanna ecosystem. **Remote Sensing of Environment,** v. 82, p. 376-388. 2002.

STEINMETZ, S.; MAGALHÃES JÚNIOR, A.M.; FAGUNDES, P. R. R.; SCIVITTARO, W. B.;ALMEIDA, I. R.; REISSER JÚNIOR, I.; DEIBLER, A.N.; MATZENAUER, R.; RADIN, B.; PRESTES, S.D.; SILVA, M. F. Uso de graus-dia para estimar a data de diferenciação da panícula (DP) de seis subgrupos de cultivares de arroz irrigado visando à adubação nitrogenada em cobertura no Rio Grande do Sul. Pelotas: Embrapa Clima Temperado, 2010. 75 p. (Embrapa Clima Temperado. Boletim de Pesquisa e Desenvolvimento, 121).

STRECK, N. A.; BOSCO, L. C.; MICHELON, S.; ROSA,H. T.; WALTER, L. C.; PAULA, G. M. de; CAMERA, C.;LAGRO, I.; MARCOLIN, E. Avaliação da resposta ao fotoperíodo em genótipos de arroz irrigado.Bragantia, Campinas, v. 65, n. 4, p. 533-541, 2006.

TASUMI, M.; ALLEN, R. G. Satellite-based ET mapping to assess variation in ET with timing of crop development. **Agricultural Water Management**, 88, 2007. 54-62.

USGS. Department of the Interior. U.S. Geological Survey. **Landsat 8 (L8) Data Users Handbook. Version 2.0.** EROS, Sioux Falls, South Dakota. March 29, 2016.

VIÑA, A.; GITELSON, A.; RUNDQUIST, D.; KEYDAN, G.; LEAVITT, B.; SCHEPER, J. Monitoring maize (Zea mays L.) phenology with remote sensing. **Agronomy Journal**, v. 96, p. 1139-1147, 2004

WANG, L., ZHANG, F. C., JING, Y. S., JIANG, X. D., YANG, S. B., & HAN, X. M. Multi-Temporal Detection of Rice Phenological Stages Using Canopy Spectrum. Rice Science, 21(2): 108−115 (2014). DOI: 10.1016/S1672-6308(13)60170-5

WANG, R.; CHERKAUER, K.; BOWLING, L. Corn Response to Climate Stress Detected with Satellite-Based NDVI Time Series. **Remote Sensing**, 269, 2016. 2-22.

ZHANG, X., FRIEDL, M., SCHAAF, C., 2009. Sensitivity of vegetation phenology detection to the temporal resolution of satellite data. **Int. J. Remote Sens**. 30, 2061–2074.

ZHANG, X.Y., TAN, B., YU, Y.Y., 2014. Interannual variations and trends in global land surface phenology derived from enhanced vegetation index during 1982–2010. Int. J. Biometeorol. 58, 547–564.



A Importância do *Ground Sample Distance* na Construção do Mosaico de Imagens Aéreas

Leandro Rosendo Candido[1], *Lúcio André de Castro Jorge*[2], *Maximiliam Luppe*[1]

[1] Dep. Eng. Elétrica e Computação, Universidade de São Paulo
São Carlos, São Paulo, Brasil
l.candido@usp.br, maxluppe@sc.usp.br

[2]Embrapa Instrumentação
São Carlos, São Paulo, Brasil
lucio.jorge@embrapa.br

RESUMO

A Agricultura de Precisão (AP) vem ganhando notoriedade nos dias atuais fornecendo tecnologias com a finalidade de aumentar o lucro e a produtividade do agricultor. Entretanto, trabalhar com imagens ainda é um desafio na AP, pois dependendo do tipo de processamento pode levar horas para obter-se um resultado. Esse artigo tem como objetivo apresentar uma maneira diferente de calcular a translação de imagens áereas utilizando os dados fornecidos pelo *Global Positioning System* (GPS) e o *Ground Sample Distance* (GSD) com a finalidade de obter-se um mosaico de imagens aéreas. Após o cálculo a translação é passada para o modelo de câmera *pinhole* fornecido pela biblioteca OpenCV e então as imagens são unidas. Ao unir as imagens utilizando o método proposto neste artigo, o agricultor pode tomar decisões ou visualizar o comportamento de sua lavoura.

PALAVRAS-CHAVE: Câmera Pinhole, Intrínseco, Extrínseco, OpenCV, *Stitching*;

ABSTRACT

Precision Agriculture (AP) is gaining notoriety these days by providing technologies to increase farmer profitability and productivity. However, working with images is still a challenge in AP because depending on the type of processing it may take hours to get a result. This article aims to present a different way of calculating the translation of aerial images using data providing by Global Positioning System (GPS) and Ground Sample Distance (GSD) in order to obtain a mosaic of aerial images. After the calculation the translation is passed to the pinhole camera model provided by the OpenCV library and then the images are merged. By linking the images using the method proposed in this article, the farmer can make decisions or visualize the behavior of

his crop.
KEYWORDS: Pinhole Camera, Intrinsic, Extrinsic, OpenCV, *Stitching*;

INTRODUÇÃO

A Agricultura de Precisão (AP) está cada vezes mais presente na vida dos agricultores com a finalidade de facilitar as inspeções nas culturas e tomar decisões com o intuito de maximizar os lucros e minimizar as perdas em sua produção. Essa área tem ganhado destaque nos jornais brasileiros como escrito por Gomes (2016) e Papp (2016) que publicaram anúncios de parcerias entre grandes empresas de tecnologia (Qualcomm e Microsoft) e algumas empresas agrícolas (Embrapa e Monsanto).

É comum o uso de imagens aéreas em AP para se ter uma visão mais ampla da área cultivada, possibilitando que se tomem decisões visando melhorar a identificação de pragas e deficiências no cultivo. Entretanto, dependendo do tipo de processamento a ser realizado nas imagens, será necessário o uso de um computador de alto desempenho. A construção do Mosaico de Imagens Aéreas (MIA) é uma das aplicações, envolvendo imagens, que exige um alto desempenho computacional. Estudos como os apresentados por Candido, Luppe e Jorge (2016) e Candido, Jorge e Luppe (2016) tem sido feito para diminuir o esforço computacional ao construir um MIA e assim trazer os resultados ao agricultor mais rápido e este, baseado no resultado fornecido, começar uma ação em sua cultura para aumentar os lucros.

Este trabalho tem como propósito apresentar uma maneira de calcular a translação de cada imagem que deve compor o modelo de câmera *pinhole* fornecido pelo OpenCV. O cálculo é baseado nos dados de *Global Positioning System* (GPS) e no *Ground Sample Distance* (GSD) e com isso obter um MIA próximo do esperado pelos agricultores.

MATERIAL E MÉTODOS

Nesta sessão são apresentados os conceitos de câmera pinhole, os passos para se obter um MIA e uma breve explicação sobre o que é e como calcular o GSD. Tambem é descrito como os dados de GPS e GSD devem ser manipulados afim de que as imagens se encaixem utilizando o modelo de câmera *pinhole* oferecido pela biblioteca OpenCV.

Modelo de Câmera Pinhole

Esse modelo de câmera mapeia o mundo 3D, que é onde vivemos, para o mundo 2D (mundo das imagens).

Esse modelo divide-se em duas partes chamadas de Parâmetros Intrínsecos (PI) e Parâmetros Extrínsecos (PE).

Parâmetros Intrínsecos

Os PIs relacionam as referências da imagem com as referências da câmera. Essas referências são:

- **Câmera**: distância focal (f) e tamanho do pixel (s_x e s_y);
- **Imagem**: ponto central da imagem (O_x e O_y);

Trucco e Verri (1998) equacionam os PIs da seguinte forma:

$$M_{int} = \begin{bmatrix} -f/s_x & 0 & o_x \\ 0 & -f/s_x & o_y \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

Parâmetros Extrínsecos

Os PEs descrevem as transformações entre a imagem e os pontos 3D do mundo.

Para obter-se a matriz dos PEs é necessário a matriz 3x3 de rotação (R), que descreve os ângulos que a imagem foi registrada, e da matriz 3x1 de translação (T), que descreve a posição onde a imagem foi registrada.

Trucco e Verri (1998) equacionam os PEs da seguinte forma:

$$M_{ext} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & -R_1^T * T \\ r_{21} & r_{22} & r_{23} & -R_2^T * T \\ r_{31} & r_{32} & r_{33} & -R_3^T * T \end{bmatrix} \quad (2)$$

onde,

- r_{11} ... r_{33} representa a matriz de rotação aplicado nas imagens;
- T representa a matriz de translação e;
- R_n^T representa a rotação aplicado a translação;

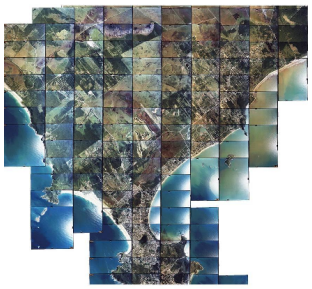
Mosaico de Imagens Aéreas (MIA)

O conceito geral de MIA é unir, de forma artística ou sistêmica, duas ou mais imagens que foram registradas através de um veículo aéreo. As imagens a serem unidas devem possuir uma área de sobreposição. Está área é utilizada quando deseja-se construir o MIA de forma automática, pois assim o computador consegue encontrar quais são as partes das imagens que se conectam. A Figura 1 mostra um MIA feito com a junção de várias imagens.

Joshi e Sinha (2013) descrevem os prinicais passos computacionais para construir um MIA. Esses passos são:

- **Entrada**: imagens que irão compor o mosaico;
- **Extração de Características**: é o primeiro passo para construção do MIA, pois ele extrai informações (características) das imagens;

Figura 1: Exemplo de um MIA



Fonte `http://aerorobot.mercadoshops.com.br/vant-aeronave-automatica-para-aerofotogrametria-23xJM`

- **Combinação de Características**: as áreas de sobreposição das imagens que estão conectadas compartilham a mesma quantidade de características e com isso é necessário dizer de um forma computacional que determinado ponto em uma imagem é o mesmo ponto em outra imagem, porém em posição diferente;
- **Homografia usando *Random Sample Consensus* (RANSAC)**: nem todas as características combinadas são combinações perfeitas, com isso é necessário removê-las e esse é o papel do RANSAC. Ele remove as características *outliers* deixando apenas às *inliers*. Após essa remoção, é calculado a matriz de homografia das imagens que se conectam e com base nessa matriz os parâmetros de câmera (intrínsecos e extrínsecos) são estimados;
- ***Warping e Blending***: o *warping* manipula as imagens para que elas se ajustem de acordo com os parâmetros de câmera estimado no passo anterior. E o *blending* deixa o moasico com uma coloração uniforme, borrando suavemente;
- **Saída**: junção das imagens formando uma unica imagem;

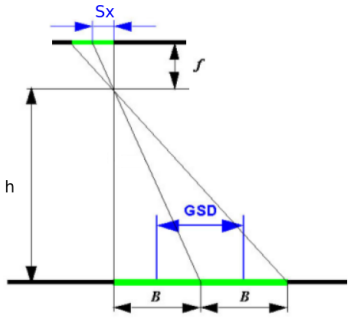
Na AP, a finalidade de unir as imagens aéreas é para que o agricultor tenha uma visão ampliada de sua lavoura e possa tomar decisões afim de corrigir falhas na plantação, identificar variações de solo através de técnicas de processamento de imagens ou identificar pragas e doenças que estão começando a atacar a plantação.

Ground Sample Distance (GSD)

O GSD é uma pequena distância, em unidades do terreno, representado por um *pixel*. Ele está diretamente relacionado com a altura em que o veículo aéreo estava no momento do registrou da imagem, com a distância focal da câmera utilizada e com o tamanho físico do *pixel* no sensor. A Figura 2 apresenta como as variáveis estão distribuida quando deseja-se registrar uma imagem e apartir dessa representação é possível extrair equação para calcular o GSD (equação 3) onde o sx representa o tamanho físico do *pixel* no sensor, o h a altura e o f a distância focal da câmera.

$$B = GSD = \frac{sx \times h}{f} \tag{3}$$

Figura 2: Representação do GSD



Fonte autor

É importante ter em mente que quanto maior a resolução da imagem, menor é o valor do GSD representado naquela imagem.

Modelo de Câmera Pinhole no OpenCV

O Modelo de Câmera Pinhole fornecido pelo OpenCV segue a mesma descrita na sessão acima.

Porém ao inserir a translação como parâmetro é necessário antes calcular qual é o deslocamento que uma imagem está distante das outras. A equação 4 para calcular o deslocamento das imagens envolvidas na construção do MIA é:

$$D = \frac{m \times a}{f \times sx} \tag{4}$$

onde,

- m = longitude/latitude convertidos para metros;
- a = 1 metro dividido pelo GSD (tomando cuidado com as unidades);
- f = distância foacal;
- sx = tamanho físico do *pixel*;

Esse passo deve ser repetido para as coordenadas de latitude e longitude para obter a translação (x, y) de cada imagem, onde x é representado pela longitude e y é representado pela latitude. A translação em z representa uma escala na imagem e deve seguir a razão: altura de vôo programada subtraida pela altura de vôo real (registrada pelo GPS).

RESULTADOS E DISCUSSÃO

Os dados de GPS registrado no momento de captura de cada imagem estão representados na Tabela 1. Os dados da câmera utilizada na execução deste trabalho são:

- Distância Focal (f) = 3.6 mm
- Tamanho físico do pixel (b) = 0,00154 mm

A partir dos dados calculou-se qual é o GSD e converteu (em metros) os dados de latitude e longitude para cada imagem. Como a diferença entre os valores de altura são muito pequenos,

Tabela 1: Dados de registro das imagens utilizadas nesse trabalho.

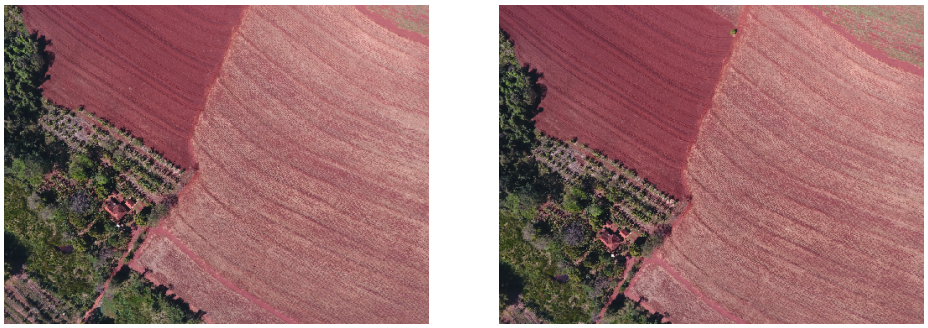
	Longitude			**Latitude**			**Altura**
Nome Img	*Grau*	*Minutos*	*Segundos*	*Grau*	*Minutos*	*Segundos*	*Metros*
DJI_0187.JPG	46	56	36.538	22	32	51.939	100.5
DJI_0188.JPG	46	56	36.682	22	32	52.882	100.5

foi assumido como se todas as imagens foram registradas a uma altura de 100 m e a altura de vôo deprogramada também foi 100 m. Aplicando a equação 3 obtem-se um GSD de 4 cm.

A Tabela 2 apresenta os valores de Translação de cada imagem após a aplicação da equação 4. Esses valores são as entradas de translação esperado pelo OpenCV.

Tabela 2: Valores de translação esperado pelo OpenCV

Nome Img	**X**	**Y**	**Z**
DJI_0187.JPG	557860,702	267950,074	0
DJI_0188.JPG	557861,177	267953,187	0



(a) Image 1 (b) Imagem 2

Figura 3: Imagens que serão unidas.

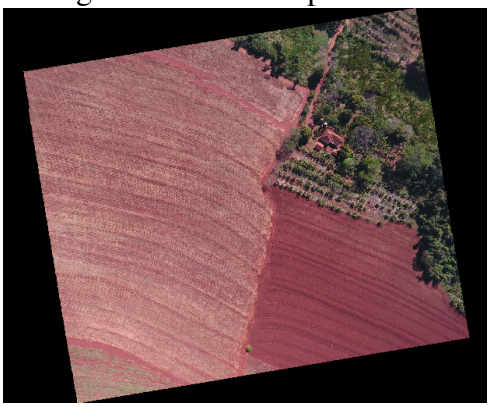
A Figura 3 apresenta as imagens que serão unidas e a Figura 4 mostra o resultado de como as imagens ficaram encaixadas após essas transformações. Esta Figura está inclinada, porque foi aplicado a rotação nas duas imagens.

Olhando o resultado da união o agricultor consegue ter uma noção de como sua plantação está distrubuida e como ela está se comportando.

CONCLUSÕES

Este trabalho apresentou uma outra maneira de realizar a translação utilizando o modelo de câmera *pinhole* fornecido pela biblioteca OpenCV. Essa maneira utiliza o GSD para calcular o deslocamento de uma imagem para a outra e obter os valores a serem passados como parâmetro para o modelo do OpenCV. Com a imagem resultante é possível visualizar como está distribuído a lavoura e o terreno fazendo com que o agricultor tome as decisões cabíveis afim de aumentar seus lucros e produtividade.

Figura 4: MIA de duas imagens construído a partir da abordagem desse trabalho.



Fonte autor

Os próximos passos serão de implementar os algoritmos que realizam a correção do encaixe nas imagens que compoem o mosaico de imagens aéreas.

REFERÊNCIAS

CANDIDO, L. R.; JORGE, L. A. de C.; LUPPE, M. Aerial image mosaics built using images with vegetation index pre-calculated. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *SPIE Remote Sensing*. [S.l.], 2016. p. 99980K–99980K.

CANDIDO, L. R.; LUPPE, M.; JORGE, L. A. d. C. *Aerial Image Mosaics Construction Using Heterogeneous Computing for Agricultural Applications*. Lisbon, Portugal: BIONATURE 2016 the Seventh International Conference on Bioenvironment, Biodiversity and Renewable Energies : ENVIROSENS 2016 : the International Symposium on Remote Sensing for Climate and Earth Monitoring : June 26-30, 2016, Lisbon, Portugal, 2016. 22–23 p. ISBN 9781612084893. Disponível em: <https://www.thinkmind.org/index.php?view=article&articleid=bionature_2016_2_20_77009>.

GOMES, J. R. *Monsanto anuncia parceria com Microsoft para startups no agronegócio*. O Estado de S.Paulo, 2016. Disponível em: <http://economia.estadao.com.br/noticias/geral,monsanto-anuncia-parceria-com-microsoft-para-startups-no-agronegocio,10000060928>. Acesso em: 22 maio 2017.

JOSHI, H.; SINHA, M. K. A survey on image mosaicing techniques. *international Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume*, v. 2, 2013.

PAPP, A. C. *Embrapa e Qualcomm querem potencializar drones no campo*. O Estado de S.Paulo, 2016. Disponível em: <http://economia.estadao.com.br/noticias/agro,embrapa-e-qualcomm-querem-potencializar-drones-no-campo,10000060122>. Acesso em: 22 maio 2017.

TRUCCO, E.; VERRI, A. *Introductory techniques for 3-D computer vision*. [S.l.]: Prentice Hall Englewood Cliffs, 1998.



A Low-Cost Six Band Multi-Spectral Camera Platform for In-Flight Near Real-Time Vegetation Index Computation and Delivery

Lucas Diogo de Mendonça[1,2], *Cesar G. Chaves*[1], *Luiz Eduardo Vicente*[3], *Eduardo Rodrigues de Lima*[1]
{*lucas.mendonca, cesar.arroyave, eduardo.lima*}*@eldorado.org.br*

[1] Instituto de Pesquisas Eldorado
Campinas, São Paulo, Brasil

[2]Faculdade de Engenharia Elétrica e Computação,Universidade Estadual de Campinas
Campinas, São Paulo, Brasil

[3]Empresa Brasileira de Pesquisa Agropecuária – Embrpa Meio Ambiente
Rodovia SP 340, KM 127,5 – 13820-000 – Jaguariúna - SP, Brasil

RESUMO

Ao longo dos últimos dez anos, o interesse no uso de Veículos Aéreos Não Tripulados (VANTs) para aplicações na agricultura tem crescido significativamente. Uma das áreas de maior interesse nesse campo é o monitoramento do crescimento e saúde das plantas, onde os VANTs podem ser utilizados em conjunto com câmeras multiespectrais, as quais excedem as capacidades das convencionais. Há alguns anos o custo de VANTs era muito alto, o que, juntamente com a necessidade de um operador especializado e elevado custo de manutenção tornavam sua aquisição inviável para a maioria dos agricultores. No entanto, o aumento do interesse em VANTs com uso recreacional possibilitou a redução destes custos, tornando este tipo de veículo uma ferramenta atraente e acessível para a inovação agrícola. Infelizmente, os equipamentos (e.g. câmeras multiespectrais) necessários para transformar VANTs recreativos em ferramentas de auxilio na agricultura não observaram tal redução de custos. Este trabalho apresenta um projeto que, usando principalmente materiais "de prateleira", visa criar uma câmera multiespectral de baixo custo, que além de funcionalidades similares às de câmeras comerciais, pode computar e transmitir índices de vegetação, durante o voo.

PALAVRAS-CHAVE: Câmera Multiespectral, Índice de Vegetação, Infravermelho

ABSTRACT

Over the last decade, the interest in using Unmanned Aerial Vehicles (UAVs) for agricultural applications has grown. One of the most appealing application, in this field, is plant and health

growth monitoring, where UAVs can work in unison with multi-spectral cameras, exceeding the capabilities of conventional ones. Some years ago, UAVs were very expensive, which together with the need of a specialized operator and high maintenance costs made their acquisition not feasible for most agriculturists. The increased interest in recreational UAVs lowered those costs, making these type of vehicles an attractive and accessible tool for agricultural innovation. Unfortunately, the equipment (e.g. multispectral camera) required to turn recreational UAVs into agricultural auxiliary tools did not observe such a cost reduction. This paper presents a project, which mainly uses over-the-shelf materials to create a low-cost camera, which in addition to the features of current commercial cameras, can compute and deliver vegetation indexes during flight.
KEYWORDS: Multispectral Camera, Vegetation Index, Infrared

Introduction

Remote sensors have been used, for many years, as tools for acquiring valuable information. As time goes by, sensing is done with higher precision, thus, increasing the volume of the acquired data. Moreover, such precision enables tackling even the most diverse problems each day. A few examples of these uses are: vegetation management (MORAN; INOUE, 1997), forestry , land use (VOGELMANN et al., 2001), and planning (VOGELMANN et al., 2001). Agriculture is no different, having key characteristics being identified (BASTIAANSSEN; MOLDEN; MAKIN, 2000; GITELSON, 2016), managing and preventing plant diseases (RAJI et al., 2016), infestation observation (NANSEN; ELLIOTT, 2016), and water stress (TILLING et al., 2007), among other uses. Perhaps the single most valued statistic is yield estimation (SHANAHAN et al., 2001). These are just a few examples of a multitude of possible uses enabled by remote sensing.

One way to contain plant diseases is by monitoring their health throughout its growing cycle and identifying deficits or problems at their earliest stages. This can heavily influence the need to use agrotoxics, fertilizers and even reduce the amount of irrigation.

Governmental institutes can rely on data acquired from satellites, which sometimes has low resolution and long date intervals, for studies such as ecosystem changes. On the other hand, agriculturist require data with high resolution and low date intervals in order to take prompt action and maintain their crops safe. Moreover, many visible signs are only presented when a disease is greatly spread, therefore it is better to use multispectral images, which help to identify some of these diseases before visual signals appear. Consequently, this lack of readily available data has drawn attention towards Unnamed Areal Vehicles (UAV) equipped with multispectral cameras.

Together with the growing interest of the general public towards using UAV as recreational devices, UAV prices have been drastically dropping, easier control have been implemented, and are now becoming easier to obtain. Unfortunately the same has not been true for multispectral cameras. A popular UAV-quad-copter can cost upwards of $R\$3.000$, such as the

DJI Phantom 3 (DJI, 2017b), and a higher end model such as the DJI Inspire 2 (DJI, 2017a) can cost around $R\$15.000$. Meanwhile, a multispectral camera such as the Tetracam uMCA 6 Snap (TETRACAM, 2017) is around $R\$79.000$, and the Tetracam MCAW 6 (TETRACAM, 2017) will cost $R\$89.700$, a hefty price to be put on a craft with a high risk of falling out of the skies.

As most regular cameras have recently seen price drops, the idea of creating a cost-viable multispectral camera based on regular cameras comes to mind. By applying specific filters to high-end quality photography cameras, the authors in (DEAN; WARNER; MCGRAW, 2000) and (SAKAMOTO et al., 2011) were able to capture infrared bands and produce good quality sensed data. The main goal of the project is to create, with an even lower price, a near-commercial quality multispectral camera using mostly low cost off-the-shelf materials available nationally. This paper presents the first results obtained with a prototype of a six band multispectral camera platform. Results show that even though the selected commercial cameras were designed for normal RGB photography, they were able to successfully generate images after being filtered to a narrow band. Additionally, the presented six band multispectral camera platform, built with parts that cost 7.14% of the price of a commercial multispectral camera, was able to capture information that allowed the generation of indexes with distinguishable values for the different types of tested vegetation.

Materials & Methods

The six band multispectral camera platform presented in this paper, was developed within a multi-prototype road-map which began with the creation of a four band multispectral camera platform. Some four band multispectral camera prototypes were tested and the one with the best results, regarding capability to capture infrared ranges, was presented in (MENDONCA et al., 2017). The prototype presented in this paper is drawn up in two distinct parts, i.e. hardware and software, improvements and/or implementations, which are detailed in this section.

Hardware

Regarding hardware, the six band multispectral camera platform is basically an extension of our four band multispectral camera platform presented in (MENDONCA et al., 2017). This prototype incorporates: 1) Four identical cameras; 2) Three narrow band filters (inside the cameras); 3) A Single Board Processing Unit (SBPU); 4) A connections-unifying PCB; 5) Project-specific designed WiFi antenna; 6) A high capacity Battery; And 7) An enclosing 3D-printed box. The components are mounted as depicted in Figure 1 and detailed thereafter.

Cameras with wide lens were used to increase the captured area in a single image, such cameras are high speed shooting capable, which allow the UAV to transit at higher rates while capturing the images. Moreover, three of the cameras had their infrared blocking filters replaced by narrow band filters and the four cameras were hardwired in order to be controlled by the single board computer.

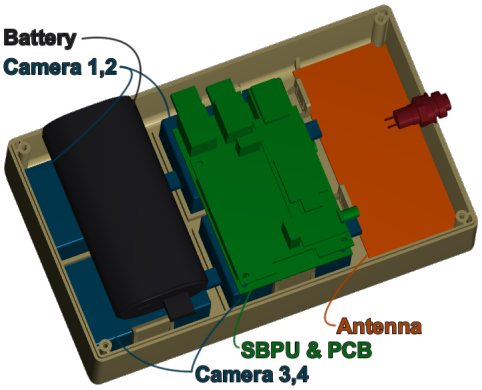


Figure 1: 3D View of the proposed multispectral camera.

Filters were added to narrow the sensed bands. The infrared blocking filter of three of the cameras were replaced by band pass filters to select the desired region of the spectrum. The response of the camera sensor, as presented by the manual (APTINA IMAGING, 2010), was taken into consideration when choosing the desired bands. The center wavelength of the chosen filters correspond to: 1) 675 ± 25nm; 2) 725 ± 25nm; And 3) 850 ± 25nm. A rough spectral response of the cameras with the tree filters can be seen in Figure 2.

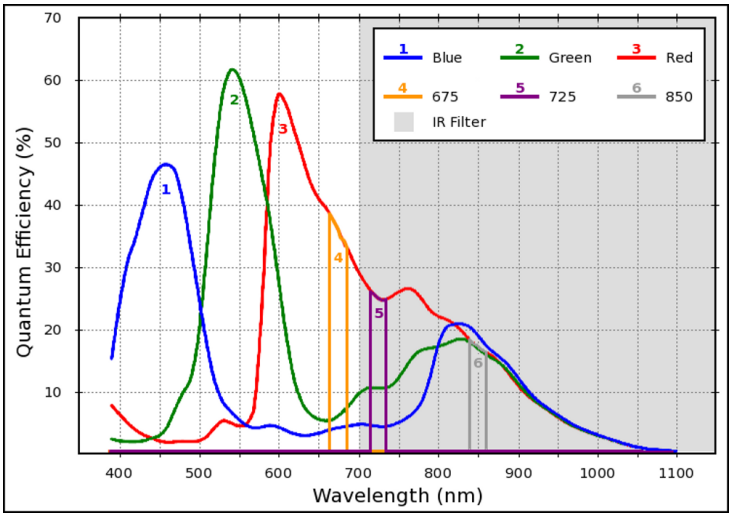


Figure 2: Expected Camera's Spectral Response.

Since the main purpose of the camera is agriculture mapping and anomaly detection, specifically the area around the vegetation slope (HUETE, 1988) within the near-infrared band was considered. The chosen values represent the narrow end of the red spectrum, the middle of the vegetation slope, and the top of the slope. These three bands are used to distinguish different plantations apart and are exceptional in displaying plant stress, as well as other characteristics.

A **Single Board Processing Unit (SBPU)** controls everything within the multispectral camera. With the cameras being electronically controlled, the SBPU was installed to allow varied functionalities to be programed into it, and have the cameras behave in unison. The SBPU capable of outputting simple controls signals has complete control over the cameras, allowing it to com-

mand the cameras as needed. The four cameras are commanded to shoot simultaneously and to transfer the captured images to the SBPU for processing. Also implemented within the SBPU are the protocols that turn it into an access point, making it possible for other devices, that are WiFi compatible to connect to it in order to access the generated vegetation indexes.

The **Unifying Printed Circuit Board (PCB)** simplified all the connections within the camera by having a central point of connection. All components that needed control or that gave commands were connected to this PCB. It simplified energy distribution, command connections, and data lines. The PCB was designed and printed internally specifically for the project.

A **Specific WiFi Antenna** was created for this project, due to the fact that common market antennas are designed to distribute their signal evenly in all directions. Since in the camera will be aboard a UAV, only one side of the antenna is needed for transmission/reception, thus allowing for a directed antenna to improve the communication range. The designed antenna is presented in Figure 3.b. Moreover, an over-the-air full anechoic chamber (Figure 3.e) was used to simulate a complex multi-path environment at the location of the camera (Figure 3.d) in a repeatable way by using radio channel emulators connected to a circular array of probes. The simulation of a realistic radio environment, was conducted in order to assess the communication performance of the antenna.

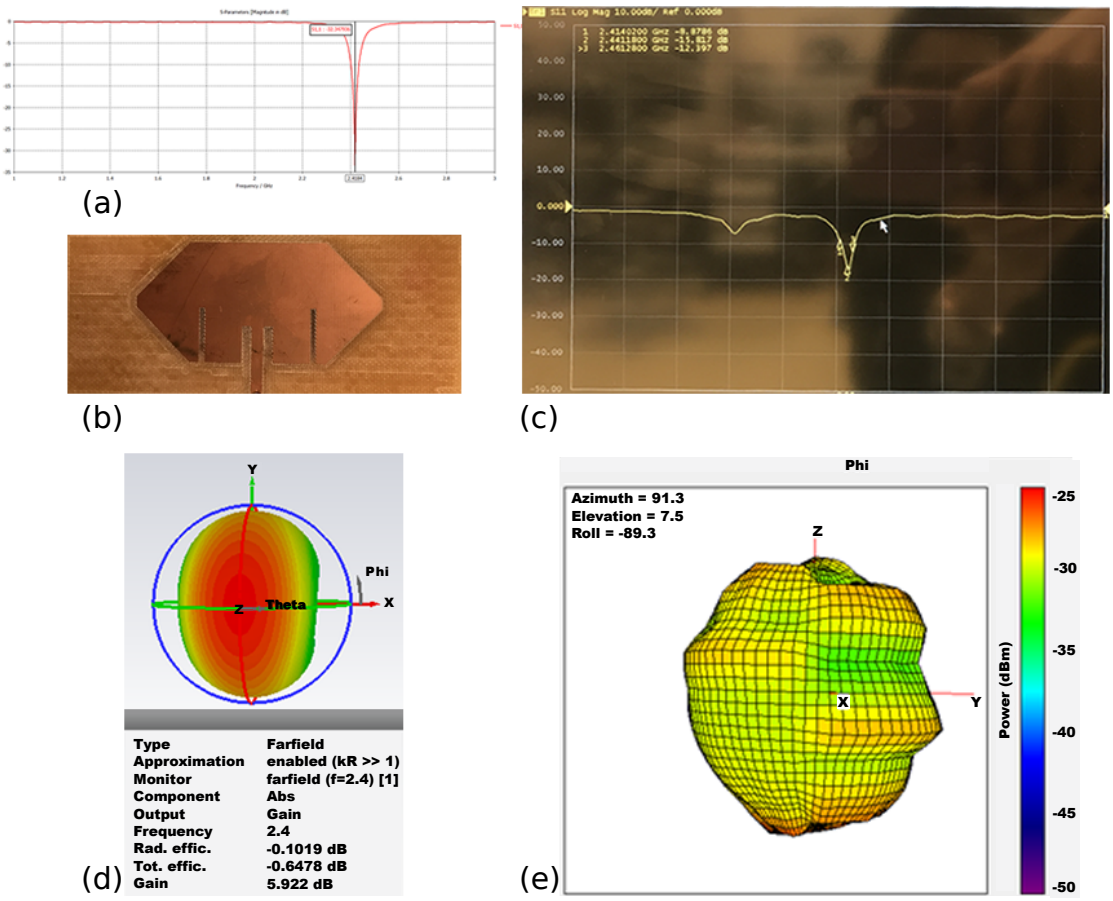


Figure 3: Designed Antenna: (a) Simulated result; (b) Printed Antenna; (c) Measured result; (d) Close up in the full anechoic chamber; (e) Wide view of the full anechoic chamber

A **Battery** was added to give the camera stand alone functionality, powering the SBPU and increasing the cameras capturing time during a single flight. There is not much to the chosen battery, but a few options were evaluated, the best by means of space/weight ratio was chosen.

A **3D-printed Case** was created to encapsulate the previously listed components and reduce possible vibrations during a flight.

Software

A three level software architecture was developed in order to control everything as a multispectral camera and not as four separate regular cameras connected to a SBPU, such architecture is depicted in Figure 4.

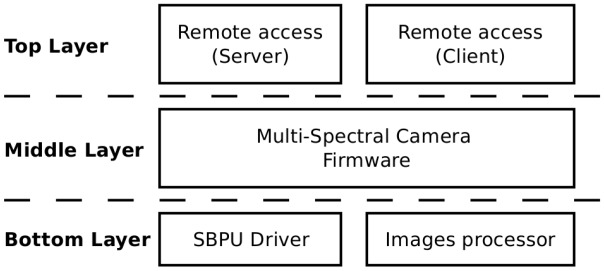


Figure 4: Software architecture of the proposed multispectral camera

The **SBPU Driver:** Implemented using the Python scripting language (ROSSUM, 1995), this Python library provides absolute control over the four cameras. Such control includes mainly: 1) Turning the cameras on and off; 2) Taking pictures simultaneously with all the cameras or a subset of them; and 3) Retrieving pictures from the camera's memory card.

The **Images Processor:** Implemented using C and OpenCV (BRADSKI, 2000), receives the pictures retrieved from the four cameras and corrects, aligns, and standardizes them. The wide lens distortion is estimated and the image undistorted. The images are then aligned to have the different bands match. Finally an RGB image displays the wide-band RGB captured image, while the second image is an RGB composed of the three narrow bands (i.e. 675nm,725nm, and 850nm respectively). The result of each capture after this process is two images: RGB and Infrared (multiple bands).

The **Multi-Spectral Camera Firmware:** Also implemented using the Python scripting language, orchestrates the entire data-acquisition/index-generation process.

The **Remote Access (Server):** Implemented using Bash scripting, interfaces the communication between the client and the *Middle Layer*. Such communication encompasses forwarding of commands from the client towards the *Multispectral Camera Firmware*, as well as the transmission of calculated indexes towards the client.

The **Remote Access (Client):** Implemented also using Bash scripting, enables a WiFi-compatible device to connect to the camera, and once connected, send commands and received computed

indexes. Currently the computed indexes are: 1) NDVI by (CARLSON; RIPLEY, 1997); 2) RDVI by (ROUJEAN; BREON, 1995); 3) EVI2 by (JIANG et al., 2008); 4) OSAVI by (RONDEAUX; STEVEN; BARET, 1996); And 5) MSAVI by (QI et al., 1994). Moreover, All image processing procedures are also implemented within the *Remote Access Client*, allowing different visualizations of the data to be calculated locally, avoiding the need to re-transfer all the images.

Results

Since no VANT is available at the current state of the project, the proposed multispectral camera (Figure 5) was tested by generating the vegetation indexes of a nearly horizontal picture of an area with different vegetation densities, depicted in Figure 6 with the Infrared composite.



Figure 5: Pictures of the camera.



(a) (b)

Figure 6: (a) Natural RGB; (b) Infrared composite.

All the resulting vegetation indexes are presented in Figure 8. Moreover, for a better understanding of the results, Figure 7 presents one of the indexes with arrows pointing at five areas o interest that correspond to: a) a well detailed border between three different responses, a building, pavement, and grass; b) an area that had its grass removed and that has been slowly growing again; c) higher vegetation, where its own leaves shadow each other; d) a path of trampled grass; and e) a well defined asphalt. A noted blurring in some strong edges was accounted and is believed to be less apparent on vertical images.

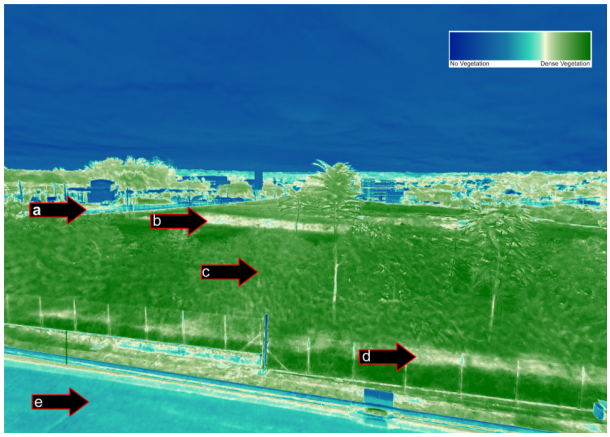


Figure 7: Comparative Result.

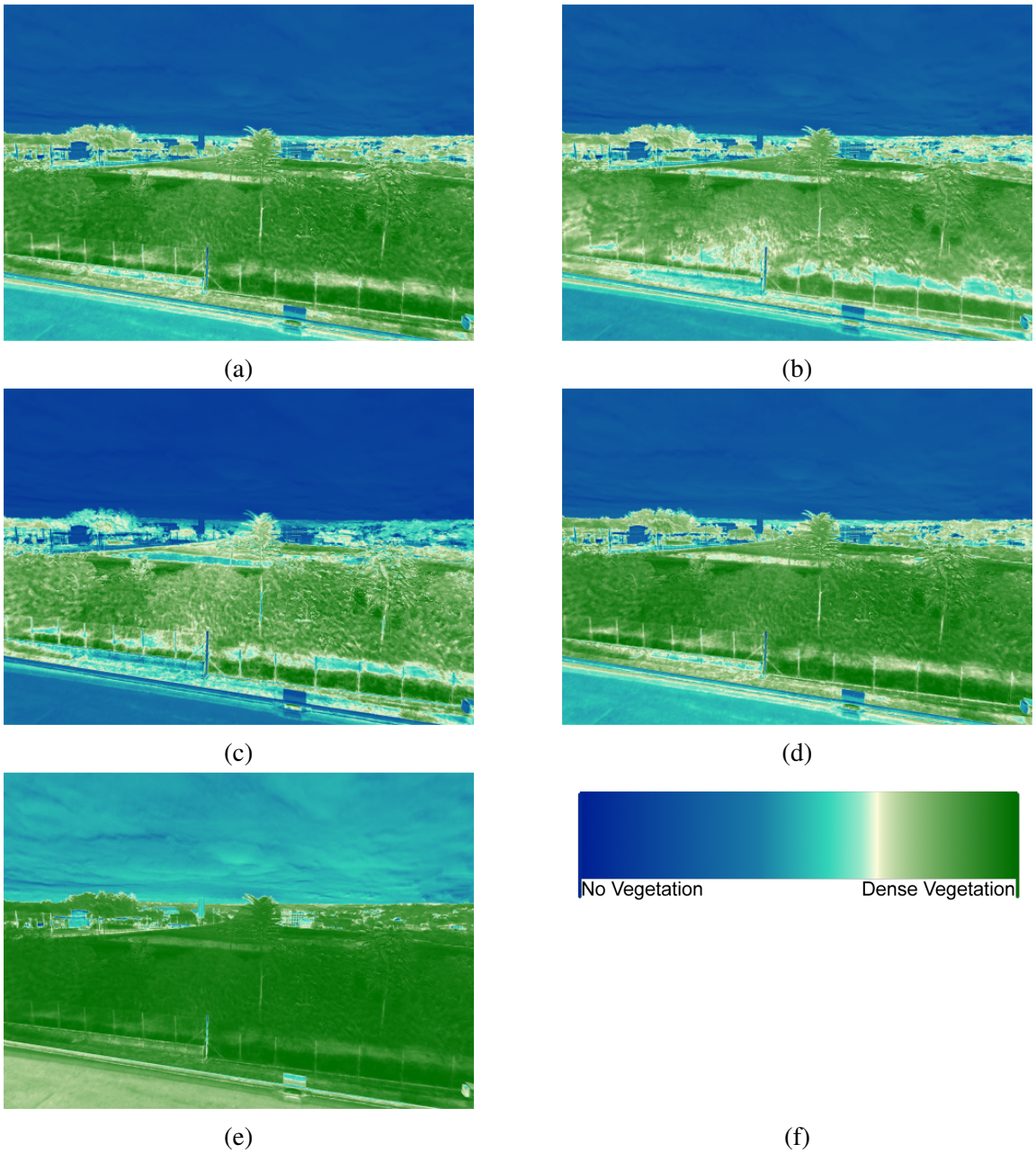


Figure 8: Resulting indexes: (a) NDVI; (b) RDVI; (c) EVI2; (d) OSAVI; (e) MSAVI; (f) Key Used in the Index Look-up-table

Conclusion

This paper presented a six band multispectral camera platform that, according to the presented test results, is capable of calculating indexes where different types of vegetation can be identified. Moreover, the materials used for building such camera cost $R\$5.000$, which can be considered low-cost knowing that the acquisition cost of a commercial grade multispectral camera in Brazil is roughly around $R\$70.000$. Additionally, the camera weight almost 800g, which is within the acceptable range for a medium sized UAV.

Future Work

Future work include: 1) testing the presented six band multispectral camera platform aboard a UAV so that perpendicular pictures can be acquired; 2) compare results with other equipment to obtain a quantitative accuracy; 3) the optimization of the used hardware in order to reduce the weight of the camera; and 4) an improvement on the software architecture so that a more flexible platform can be achieved, thus allowing researchers from universities and R&D institutes to customize it as needed and/or calculate other indexes.

Acknowledgements

The authors would like to acknowledge the help received during this project. Starting with Embrapa's researchers involved directly and indirectly so far. An immense support was also received from Eldorado's Flavia Cabral da Costa, in designing and testing the antenna.

REFERENCES

APTINA IMAGING. *AR0330: 1/3-Inch CMOS Digital Image Sensor Features*. [S.l.], 2010.

BASTIAANSSEN, W. G.; MOLDEN, D. J.; MAKIN, I. W. Remote sensing for irrigated agriculture: examples from research and possible applications. *A.W.M.*, Elsevier, 2000.

BRADSKI, G. Learning opencv. *Dr. Dobb's Journal of Software Tools*, 2000.

CARLSON, T. N.; RIPLEY, D. A. On the relation between ndvi, fractional vegetation cover, and leaf area index. *Remote sensing of Environment*, Elsevier, v. 62, n. 3, p. 241–252, 1997.

DEAN, C.; WARNER, T. A.; MCGRAW, J. B. Suitability of the dcs460c colour digital camera for quantitative remote sensing analysis of vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier, v. 55, n. 2, p. 105–118, 2000.

DJI, . *DJI Inspire 2*. 2017. [Online; accessed 24-May-2017]. Disponível em: <http://store.dji.com/product/inspire-2>.

DJI, . *DJI Phantom 3*. 2017. [Online; accessed 24-May-2017]. Disponível em: <http://store.dji.com/product/phantom-3-standard>.

GITELSON, A. A. 15 remote sensing estimation of crop biophysical characteristics at various scales. *Hyperspectral Remote Sensing of Vegetation*, CRC Press, p. 329, 2016.

HUETE, A. R. A soil-adjusted vegetation index (savi). *Remote sensing of environment*, Elsevier, v. 25, n. 3, p. 295–309, 1988.

JIANG, Z. et al. Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment*, Elsevier, v. 112, n. 10, p. 3833–3845, 2008.

MENDONCA, L. et al. Low-cost multi-spectral camera platform for in-flight near real-time vegetation index computation and delivery. 2017.

MORAN, M. S.; INOUE, Y. Opportunities and limitations for image-based remote sensing in precision crop management. *Remote sensing of Environment*, Elsevier, 1997.

NANSEN, C.; ELLIOTT, N. Remote sensing and reflectance profiling in entomology. *Annual review of entomology*, Annual Reviews, v. 61, p. 139–158, 2016.

QI, J. et al. A modified soil adjusted vegetation index. *Remote sensing of environment*, Elsevier, v. 48, n. 2, p. 119–126, 1994.

RAJI, S. N. et al. Detection and classification of mosaic virus disease in cassava plants by proximal sensing of photochemical reflectance index. *J. of the I. S. of R. S.*, Springer, 2016.

RONDEAUX, G.; STEVEN, M.; BARET, F. Optimization of soil-adjusted vegetation indices. *Remote sensing of environment*, Elsevier, v. 55, n. 2, p. 95–107, 1996.

ROSSUM, G. *Python Reference Manual*. Amsterdam, The Netherlands, 1995.

ROUJEAN, J.-L.; BREON, F.-M. Estimating par absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment*, Elsevier, 1995.

SAKAMOTO, T. et al. Assessment of digital camera-derived vegetation indices in quantitative monitoring of seasonal rice growth. *ISPRS J. of P. and R. S.*, Elsevier, 2011.

SHANAHAN, J. F. et al. Use of remote-sensing imagery to estimate corn grain yield. *Agronomy Journal*, American Society of Agronomy, v. 93, n. 3, p. 583–589, 2001.

TETRACAM, . *Ordering Products*. 2017. [Online; accessed 24-May-2017]. Disponível em: <http://www.tetracam.com/Store.htm>.

TILLING, A. K. et al. Remote sensing of nitrogen and water stress in wheat. *Field Crops Research*, Elsevier, v. 104, n. 1, p. 77–85, 2007.

VOGELMANN, J. E. et al. Completion of the 1990s national land cover data set for the conterminous united states from landsat thematic mapper data and ancillary data sources. *Photogrammetric Engineering and Remote Sensing*, v. 67, n. 6, 2001.

Amenização de ruídos de reflexão da radiação eletromagnética em imageamento aéreo

Alessandro Stachak[1], Jessyca Kawana Franquitto[1], Selma Regina Aranha Ribeiro[2], Alaine Margarete Guimarães[3], Antônio David Viniski[1], Eduardo Fávero Caires[4].

[1] Mestrando do Programa de Pós-Graduação em Computação Aplicada, Universidade Estadual de Ponta Grossa, Ponta Grossa, Paraná, Brasil, alessandro.sbk@gmail.com, jefranquitto@hotmail.com, davidviniski@gmail.com.

[2] Departamento de Geociências, Universidade Estadual de Ponta Grossa, Ponta Grossa, Paraná, Brasil, selmar.aranha@gmail.com.

[3] Departamento de Informática, Universidade Estadual de Ponta Grossa, Ponta Grossa, Paraná, Brasil, alainemg@uepg.br.

[4] Departamento de Ciência do Solo e Engenharia Agrícola, Universidade Estadual de Ponta Grossa, Ponta Grossa, Paraná, Brasil, efcaires@uepg.br.

RESUMO

As imagens adquiridas remotamente necessitam de correções de radiometria e geometria, devido a ruídos causados por ações alheias ao alvo. Os métodos de correção de radiometria melhoram a imagem para fins de fotointerpretação, realçando classes de uso da terra (ou do solo) de interesse. Entretanto, existem outros métodos, como técnicas de Processamento Digital de Imagens (PDI) que também realizam correções em imagens. Com isso, visando avaliar algumas técnicas alternativas, o objetivo desde trabalho foi aplicar técnicas de PDI, filtros e transformações do histograma, para reduzir a reflexão da radiação eletromagnética na cultura de trigo em uma imagem capturada por Aeronave Remotamente Pilotada (RPA, do inglês - *Remotely Piloted Aircraft*) na Fazenda Paiquerê, localizada no município de Piraí do Sul, Paraná. Utilizaram-se filtros passa-baixa Gaussiano, filtro Bartlett de amplitude de cor e transformação do histograma, onde foi aplicada a redução de intensidade em 10%, 20% e 30%. As transformações do histograma reduziram significativamente o ruído e ocasionaram menores alterações na cena não afetada pelo ruído, em comparação com as outras técnicas testadas.

PALAVRAS-CHAVE: Agricultura, Filtro, Histograma, Processamento Digital de Imagens RPA.

ABSTRACT

Interference induced by actions unrelated to the aerial imaging target requires radiometric and geometric corrections in the obtained image. The radiometry correction improves the image for photointerpretation purposes, highlighting classes of interest for land or soil uses. However, other methods, such as DIP (Digital image processing) techniques, also perform corrections in images. Aiming evaluate some alternative techniques, the goal of this paper

was to apply DIP, filters and histogram transformations techniques to reduce the reflection of electromagnetic radiation in wheat culture of an image captured by RPA (Remotely Piloted Aircraft) at Paiquerê Farm, at Piraí do Sul city, Paraná state. We compared the low-pass Gaussian and the Bartlett with color amplitude filters, along with histogram transformations with intensity reduction of 10%, 20% and 30%. The best result was obtained by histogram transformations, which significantly reduced the noise and caused minor changes in the scene not affected by noise.
KEYWORDS: Agriculture, Filter, Histogram, Digital image processing RPA.

INTRODUÇÃO

O imageamento aéreo com Aeronave Remotamente Pilotada (RPA - do *inglês Remotely Piloted Aircraft*) vem crescendo devido a possibilidade de obtenção de imagens com alta resolução espacial, baixo custo e facilidade de aquisição em curto espaço de tempo (alta resolução temporal) (KEDZIERSKI; WIERZBICKI, 2015). No entanto, no imageamento aéreo, principalmente com RPA, diversos fatores devem ser controlados: como o horário e condições atmosféricas propícias para o voo (KELCEY; LUCIEER, 2012). Mesmo assim, assumindo que as condições para o imageamento sejam ideais, a tonalidade (resposta espectral) das imagens podem não atender ao desejado (LANGHI; GARCIA, 2008).

A correção radiométrica é uma etapa do pós-processamento das imagens obtidas no voo, que procura reduzir os efeitos da atmosfera (KELCEY; LUCIEER, 2012) amenizando ou eliminando os ruídos presentes nas imagens (LANGHI; GARCIA, 2008). A correção e calibração radiométrica são essenciais para remover os efeitos de variáveis ambientais tais como iluminação, névoa, dentre outros, principalmente para sensoriamento remoto com RPA (KELCEY; LUCIEER, 2012).

A realização das correções nas imagens está diretamente ligada a finalidade de obtenção das mesmas. Por exemplo, as altas frequências correspondem às modificações abruptas dos níveis digital, enquanto as baixas frequências correspondem as variações suaves dos níveis de cinza. Quando há o interesse maior na forma da iluminação de fundo, deve-se utilizar filtros passa-baixa, técnica de processamento digital de imagens (PDI), que eliminam todas as altas frequências (correspondendo a borda dos objetos) e realçam a iluminação de fundo (ALBUQUERQUE, 2000). Sendo assim, os filtros passa-baixa são adequados para suavização de ruídos nas imagens.

Outras técnicas de PDI visam eliminar ruídos, dentre elas as transformações no histograma, tais como a intensidade. Essa técnica é conhecida como transformação ponto-a-ponto, onde seu valor final só depende do valor original, diferente dos filtros passa-baixa, os quais dependem da vizinhança (FILHO; NETO, 1999). Ambas as técnicas de passa-baixa,

passa-alta e transformação do histograma são focadas em melhorias visuais da imagem (FILHO; NETO, 1999). No sensoriamento remoto, as modificações do histograma são aprimoramentos no realce radiométrico, melhorando a nitidez das imagens orbitais, destacando as características desejadas (RICHARDS, 2006).

Assim, o objetivo deste trabalho foi aplicar técnicas de PDI, filtros e transformações do histograma, para reduzir o ruído nas imagens causado pela reflexão da radiação eletromagnética na cultura de trigo.

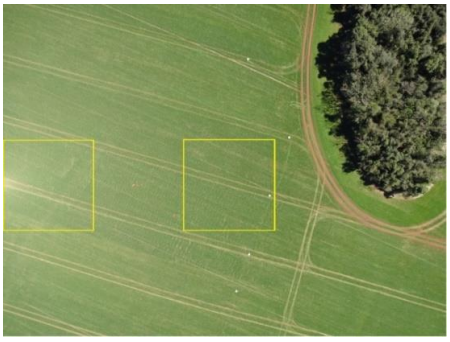
MATERIAL E MÉTODOS

Equipamentos e aquisição da área de estudo

A área experimental para captura da imagem está localizada na Fazenda Paiquerê, no município de Piraí do Sul, Paraná (50° 7’ 51.70” S 24° 21’ 23.48” W e altitude 970m), com cultivo de trigo e com quatro diferentes aplicações de adubação nitrogenada. A RPA utilizada foi o modelo *ebee* da *SenseFly* (Parrot Company, Suíça), com a câmera *Sony WX RGB* de 18.5 megapixels acoplada, a qual obtém imagens nos canais de cores do visível na resolução espectral de 660*nm* para o vermelho, 450*nm* no azul e 520*nm* para o verde.

A imagem selecionada foi do voo realizado em 22 de setembro de 2016, próximo às 12hrs00mim (horário de Brasília), com resolução espacial de 3,4cm e altura de tomada das imagens de 120m do solo. No voo selecionado foram obtidas 126 imagens da área experimental, de aproximadamente 10 hectares. Dentre essas imagens foi selecionada uma, na qual foram aplicados filtros e transformações (Figura 1). Essa imagem foi selecionada por apresentar nitidamente o reflexo da radiação, ou seja, ruído na cultura de trigo e as placas que separam os tratamentos, evitando assim realizar comparações entre tratamentos, que poderiam acarretar em diferenças devido aos tratamentos e não ao ruído. Os quadrados amarelos destacados na Figura 1, correspondem às áreas que foram comparadas nos histogramas.

Figura 1: Imagem Selecionada



Fonte: Os autores.

Filtros e transformações do histograma

Os filtros e transformações do histograma foram aplicados na imagem visando reduzir os ruídos da reflexão de energia, sendo eles: (i) Filtro *Gaussiano* passa-baixa, que por meio de uma máscara usada de tamanho 3x3 calcula o peso pela média dos pixels vizinhos (TOMASI; MANDUCHI, 1998), sendo usado na redução de ruídos (DIREKOGLU; NIXON, 2008); (ii) Filtro Barlett (*warping filters*), que altera a amplitude das cores (GOMES, 1997) diminuindo o efeito das altas frequências; e (iii) Transformação de intensidade mediante histograma, a qual adota a média de intensidade da imagem como limiar, e assim transforma o histograma de acordo com a aplicação, dado pela equação linear $g=c.f+b$ onde, c corresponde ao controle de contraste, f a escala de cinza e b controla o ajuste do brilho (FILHO; NETO, 1999). Neste trabalho foram aplicados as reduções de 10%, 20% e 30%.

Programas Utilizados

O plano de voo, assim como o acompanhamento de sua execução, foram feitos no *software eMotion,* versão 2, fornecido pelo fabricante da RPA utilizada. Para aplicar os filtros Gaussiano e Bartlett na imagem foi usada a ferramenta gratuita *ImLab*[1], e para redução de intensidade foi empregado o aplicativo gratuito GIMP[2], versão 2.8 (GIMP, 2008).

Nos pontos amostrais da imagem foi usado o programa QGIS[3] versão 2.18 de 64 bits (SHERMAN et al., 2004) para extrair os valores digitais dos canais de cores vermelho, verde e azul (RGB do inglês – *red, green e blue*). Os testes estatísticos de comparação entre as amostras extraídas no QGIS foram executados na ferramenta R[4] (R Core Team, 2014) versão 3.3.3.

Comparação dos resultados

Como não existe imagem da mesma área sem o ruído, foi realizada a análise visual das imagens geradas no mesmo voo, porém, não foi encontrada a parte da mesma imagem em nenhuma delas. Com isso, foram selecionadas dentro da própria imagem, áreas que não foram tão afetadas, por fotointerpretação, pelo reflexo de energia. Para tanto, foi destacado o ruído e selecionadas regiões afetadas e não afetadas, com as quais foram gerados dois histogramas, os quais foram devidamente comparados. Para a coleta de pontos amostrais, foram usados polígonos iguais (10mx17m) em todas as regiões.

[1] Documentação e download disponível em: http://imlab.sourceforge.net/.
[2] Documentação e download disponível em: http://gimp.org/.
[3] Documentação e download disponível em: https://www.qgis.org/.
[4] Documentação e download disponível em: https://www.r-project.org/.

Dentro de cada polígono, foram gerados 15 pontos aleatórios, sendo coletados os valores digitais nos três canais de cores RGB. Com esses valores realizou-se a soma dos valores digitais entre os canais, gerando um único valor. Para analisar se as amostras coletadas na imagem estavam se tornando mais parecidas estaticamente, realizou-se o teste ANOVA (análise de variância) (SALVATORE, 2001). Em seguida, aplicou-se o teste de Tukey com 95% de confiança, que relata a significância dessa diferença (TUKEY, 1949).

Para avaliar o quanto os filtros e transformações do histograma alteraram os valores digitais das regiões da imagem, foi empregada a técnica de detecção de mudanças (JENSEN et al., 2009). Assim, foi obtida a média da diferença absoluta entre os valores originais e os valores atribuídos pelos filtros e transformações, expressos pela Equação 1. Considerando que a diferença entre a imagem original e ela mesmo é zero, quanto maior essa diferença, maior será a alteração que os filtros e transformações realizaram na imagem.

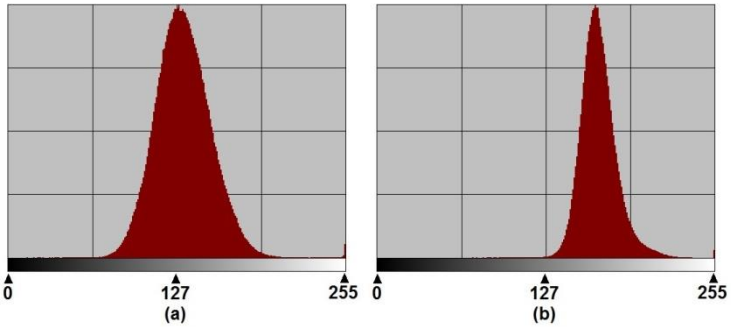
$$Diferença = \frac{\sum_{k=1}^{n} |Y_k - \mathrm{X_k}|}{N} \quad \text{Equação (1)}$$

A última etapa foi à comparação dos histogramas das regiões afetadas e não afetadas, entre a imagem original e as geradas após a aplicação dos filtros e transformações do histograma. Para tanto, novos polígonos foram utilizados, dos quais foram coletadas amostras de todas as áreas. Em cada polígono foram gerados 200 pontos aleatórios, e o coeficiente de variação (CV) (SALVATORE, 2001) foi calculado.

RESULTADOS E DISCUSSÃO

Comparando-se os histogramas das áreas afetadas e não afetadas da imagem original, verifica-se, na Figura 2 (b), que o pico da curva do nível digital é mais próximo a escala 255 do histograma, região de maior intensidade. O mesmo não se repetiu no histograma referente a região pouco afetada pelo ruído, conforme pode ser observado na Figura 2 (a).

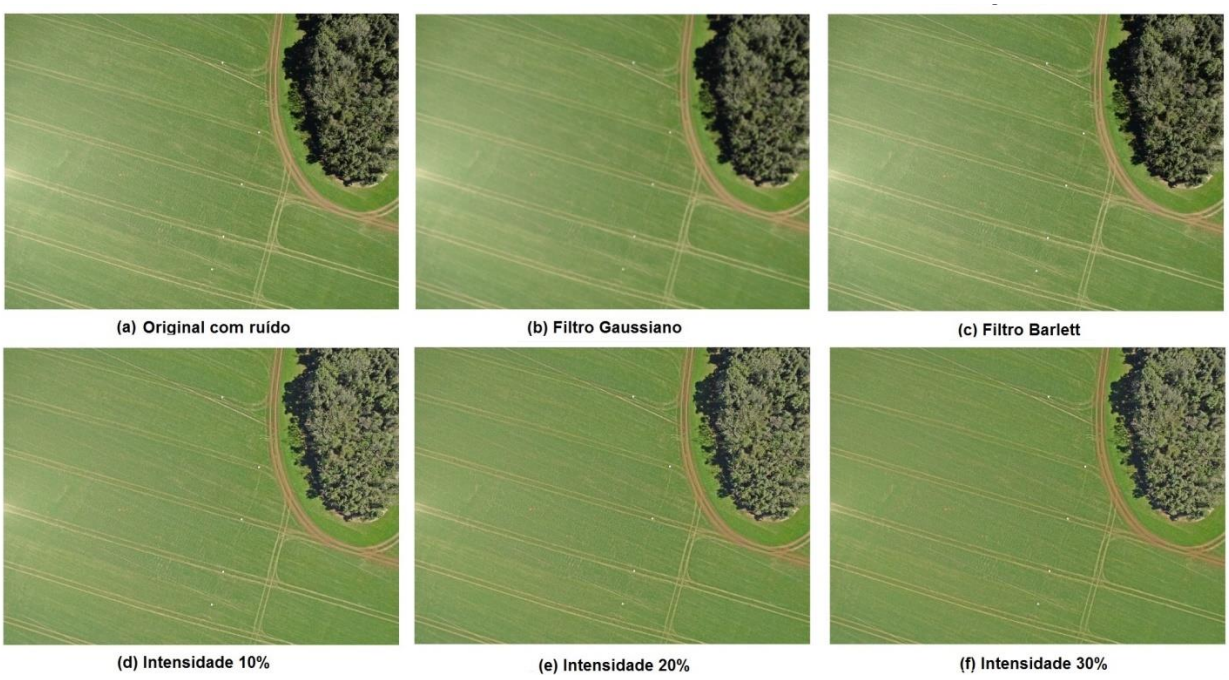
Figura 2: Histogramas, (a) região sem ruído, (b) região com ruído.



Fonte: Os autores.

Na Figura 3 verifica-se os resultados obtidos após a aplicação dos filtros e transformações do histograma. Visualmente o filtro de Barlett, Figura 3 (c), não apresentou alterações em relação a imagem original (Figura 3 (a)). Já as alterações de intensidade modificaram significativamente a imagem (Figura 3 (d), (e) e (f)). Para o filtro Gaussiano, Figura 3 (b), houve um desfoque na imagem, devido ao filtro ser do tipo passa-baixa, o qual suaviza altas frequências (YOUNG; VLIET, 1995; JANECKI, 2012).

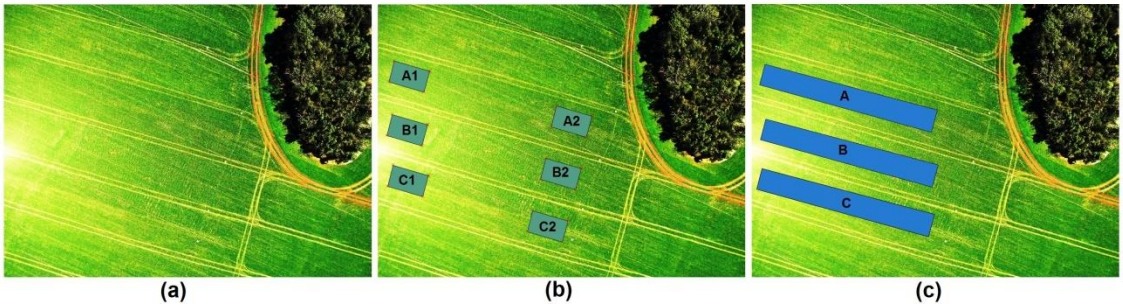
Figura 3: Imagens com filtros aplicados.



Fonte: Os autores.

A equalização do histograma foi feita com o *software* ImLab, realçando pontos de brilho na imagem (FILHO; NETO, 1999). Por consequência, destacou a área que apresenta ruído de luminosidade (Figura 4 (a)), possibilitando assim, selecionar as áreas a serem comparadas (Figura 4 (b)) e avaliar estatisticamente se a técnica reduziu o ruído. Foram usados seis polígonos de mesmas dimensões para a coleta das amostras, conforme demostrado na Figura 4 (b) e três polígonos para analisar o CV conforme Figura 4 (c).

Figura 4: Imagens com histograma equalizado e polígonos para amostra.



Fonte: Os autores.

Por apresentarem coloração normal, assumiu-se que as áreas das amostras A2, B2 e C2, não foram afetadas pelo ruído. Logo, essas regiões foram comparadas com as áreas mais afetadas (Figura 4 (b)).

As amostras foram analisadas na seguinte ordem, A1 com A2, B1 com B2 e C1 com C2 (Figura 4 (b)), devendo todas as áreas apresentar homogeneidade, por serem do mesmo tratamento e próximas entre si (70 metros).

A diferença entre as áreas na imagem original é maior de acordo com a proximidade do ponto de maior intensidade de brilho. A Tabela 1 mostra que a diferença entre as amostras B1 é maior que as demais, estando ela em um único grupo (diferença entre os grupos é expresso por letras distintas ao lado de cada média).

Tabela 1 – Teste de Tukey 5% sobre os filtros conforme divisão demonstrada na Figura 4(b).

Área	Original	Gaussiano	Barlett	Intens.10%	Intens.20%	Intens.30%
A1	415,07*bc*	416,07*b*	412,64*b*	396,14*abc*	377,57*abc*	361,86*ab*
A2	345,21*d*	349,86*d*	342,43*c*	349,57*c*	339,64*c*	330,71*b*
B1	477,43*a*	465,29*a*	473,79*a*	442,57*a*	414,64*a*	391,36*a*
B2	361,14*cd*	356,07*d*	361,79*c*	365,64*c*	349,00*bc*	343,21*ab*
C1	464,14*ab*	456,64*a*	460,93*a*	429,21*ab*	403,29*ab*	380,71*ab*
C2	382,71*cd*	373,43*c*	377,36*bc*	379,64*bc*	363,57*abc*	350,93*ab*

Para os filtros Gaussiano e Barlett, não houve alterações nas diferenças significativas entre as áreas, em relação as diferenças da imagem original. Conforme a Tabela 1, todos os valores resultantes após a aplicação dos filtros mantiveram as diferenças entre as amostras. Esse resultado também foi observado visualmente nas miniaturas correspondentes na Figura3.

A diferença entre os grupos foi reduzida quando a redução de intensidade foi aplicada, nas colunas *Intens. 10% e Intens. 20%* da Tabela 1, as amostras A1-A2 e C1-C2 passaram a ser estatisticamente iguais. Já com 30% de redução, a amostra B1-B2, mais afetada pelo ruído, passou a ser estatisticamente igual também. Isso demonstrou que aplicando a transformação do histograma para reduzir a intensidade se obteve os melhores resultados aparentes, tornando estatisticamente iguais áreas que antes eram distintas. Porém, as transformações podem alterar as informações das áreas que devem estar com coloração normal, A2, B2 e C2 (Figura 4 (b)). Com isso, visando determinar o quanto os filtros e transformações alteraram os valores digitais das áreas que não foram afetadas pelo ruído, essas amostras foram comparadas com a imagem original, usando técnicas de álgebra de mapas para detecção de diferenças.

O filtro gaussiano realizou maiores alterações nas regiões A2, B2 e C2 comparado com as regiões A1, B1 e C1, as mais afetadas pelo ruído, conforme Tabela 2, não sendo

desejada essa alteração, já que as áreas mais afetadas pelo ruído deveriam ser as mais alteradas. No filtro Barlett o mesmo se repete com menor intensidade, onde a imagem original (Figura 3(a)), e aquela onde foi aplicado o filtro de Barlett (Figura 3(c)) aparentam ser idênticas, o que foi reforçado pelo teste de Tukey, onde as diferenças significativas se mantiveram após aplicação o filtro, conforme Tabela 2.

Tabela 2 – Resultados das diferenças entre áreas originais e com filtro.

Área	Gaussiano	Barlett	Intens. 10%	Intens. 20%	Intens. 30%
A1	12,9	7,0	18,3	37,4	52,7
A2	60,6	21,6	53,8	12,2	17,3
B1	13,8	5,0	34,8	62,5	84,4
B2	58,0	85,0	11,0	12,2	19,0
C1	12,1	4,0	45,13	61,6	84,4
C2	35,8	9,8	57,4	19,4	31,9

Já com a redução de intensidade 20% e 30%, as áreas A1 e B1 são mais modificadas que as regiões A2 e B2, porém, a região C2 é mais alterada que a C1 em ambas as reduções citadas, conforme Tabela 2. Com a redução de intensidade em 30\%, as regiões com ruído foram as mais modificadas pela transformação, enquanto as áreas assumidas como corretas são menos modificadas pela redução, conforme Tabela 2.

Na comparação do CV, realizada entre os polígonos conforme Figura 4 (c), temos alguns resultados distintos, com o filtro gaussiano com a melhor redução de variação, conforme Tabela 3, deixando as áreas mais homogênias. Entretanto, conforme Tabela 2, ele realiza maiores alterações nas regiões menos afetadas pelo ruído, podendo causar a perda de informações. As reduções de intensidade e o filtro Barlett obtiveram CV menor que a imagem original, no entanto, a redução não foi tão expressiva quanto o gaussiano.

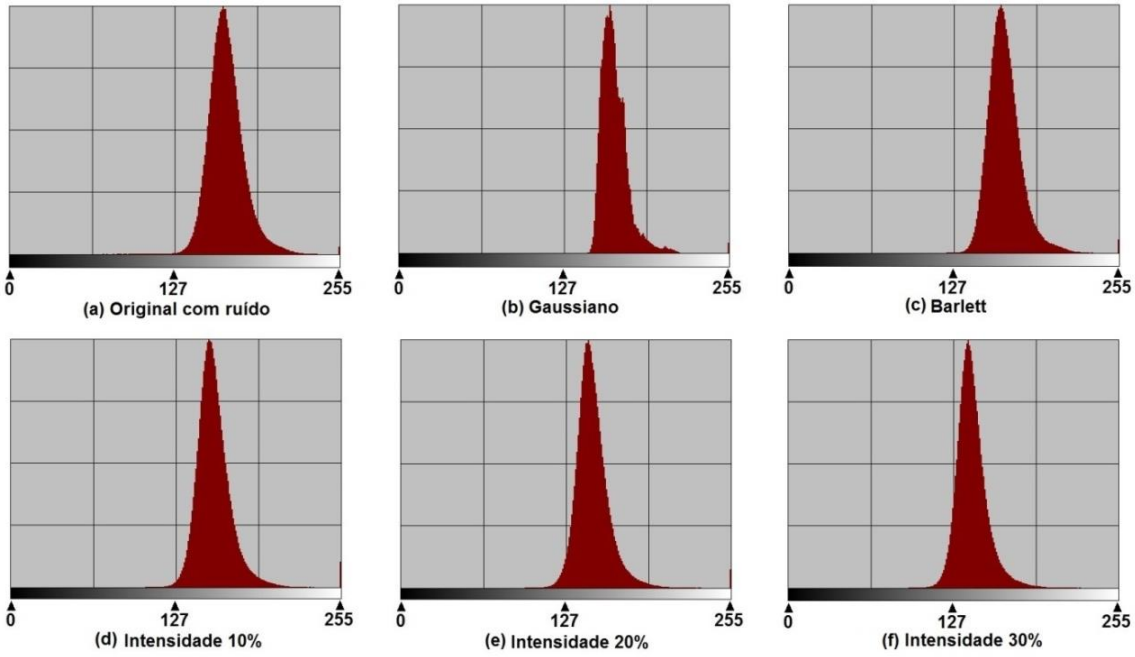
Tabela 3 – Comparativo do coeficiente de variação entre os filtros.

Área	Original	Gaussiano	Barlett	Intes.10%	Intes.20%	Intes.30%
A	12,6%	5,9%	10,5%	12,3%	12,5%	12,8%
B	14,4%	10,1%	13,0%	12,1%	12,2%	12,0%
C	10,8%	8,7%	10,1%	9,2%	9,2%	9,2%

Comparando os histogramas das regiões corrigidas pelos filtros e transformações, (Figura 6), observa-se que o filtro Barlett praticamente manteve o mesmo histograma da imagem original, refletindo os resultados observados na Figura 3, na qual não existe diferença visível entre a miniatura da imagem original e a miniatura da imagem em que o filtro foi aplicado. Já o filtro Gaussiano alterou significativamente o histograma, porém, manteve o pico (máximo da curva) do nível digital, na mesma região de intensidade da imagem original com o ruído. Para as transformações do histograma reduzindo a intensidade, o nível é gradativamente alterado, onde cada aplicação de redução trouxe o pico mais próximo da

região central do nível de intensidade (127), conforme Figura 6. Nesse processo, histogramas das regiões afetadas passam a ser mais semelhantes ao da região não afetada.

Figura 6 – Histogramas dos filtros e transformações.



Fonte: Os autores.

CONCLUSÕES

O estudo demonstrou que a transformação do histograma, aplicando uma redução de intensidade em 30% homogeneizou as áreas, efetuando maiores alterações nos valores digitais das regiões mais afetadas pelo ruído e menores alterações em regiões não afetadas.

As transformações, assim como os filtros, efetuam alterações no valor digital de cada pixel da imagem, no entanto, outras informações que a mesma expressa não foram mensuradas neste trabalho. Com isso, saber o quanto de informação real a imagem perdeu ou ganhou na aplicação dos filtros e transformações é de grande importância. Como trabalhos futuros, pode-se comparar os resultados de uma aplicação e avaliar se há algum ganho ou perda nos resultados com a aplicação das técnicas aqui apresentadas.

AGRADECIMENTOS

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo suporte financeiro.

REFERÊNCIAS

AL-AMRI, S. S.; KALYANKAR, N. V.; KHAMITKAR, S. D. A comparative study of removal noise from remote sensing image. arXiv preprint arXiv:1002.1148, 2010.

ALBUQUERQUE, M. P. Marcio Portes de. Processamento de imagens: métodos e análises. Centro Brasileiro de Pesquisas Físicas MCT, 2000.

DIREKOGLU, C.; NIXON, M. S. Image-based multiscale shape description using gaussian filter. In: IEEE. Computer Vision, Graphics & Image Processing, 2008. ICVGIP'08. Sixth Indian Conference on. [S.l.], 2008. p. 673–678.

FILHO, O. M.; NETO, H. V. Processamento digital de imagens. [S.l.]: Brasport, 1999. GIMP, G. Image manipulation program. User Manual, Edge-Detect Filters, Sobel, The GIMP Documentation Team, v. 8, n. 2, p. 8–7, 2008.

GOMES, L. V. a. J. Image Processing for Computer Graphics. [S.l.]: Springer New York, 1997. ISBN 978-1-4757-2747-0,978-1-4757-2745-6.

JANECKI, D. Edge effect elimination in the recursive implementation of gaussian filters. Precision Engineering, Elsevier, v. 36, n. 1, p. 128–136, 2012.

JENSEN, J. R. et al. Introductory digital image processing: a remote sensing perspective. [S.l.]: Prentice-Hall Inc., 1996.

KEDZIERSKI, M.; WIERZBICKI, D. Radiometric quality assessment of images acquired by uav's in various lighting and weather conditions. Measurement, Elsevier, v. 76, p. 156–169, 2015.

KELCEY, J.; LUCIEER, A. Sensor correction and radiometric calibration of a 6-band multispectral imaging sensor for uav remote sensing. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci, v. 39, p. 393–398, 2012.

LANGHI, P.; GARCIA, A. M. Correc¸ao radiom ˜ etrica de imagens a ´ ereas digitais por meio de ajuste polinomial. II Simposio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformacão , p. 000–000, 2008.

R Core Team. R: A Language and Environment for Statistical Computing. Viena, Áustria, 2014. Disponível em: <http://www.r-project.org>.

RICHARDS, X. J. a. J. A. Remote Sensing Digital Image Analysis: An Introduction. 4th ed. . [S.l.]: Springer Berlin Heidelberg, 2006. ISBN 978-3-540-25128-6,978-3-540-29711-6.

SALVATORE, D. R. D. Statistics and econometrics. 2. ed. [S.l.]: McGraw-Hill, 2001. (Schaum's Outline). ISBN 9780071348522,0071348522.

SHERMAN, G. E. et al. Quantum gis user guide. 2004. TOMASI, C.; MANDUCHI, R. Bilateral filtering for gray and color images. In: IEEE. Computer Vision, 1998. Sixth International Conference on. [S.l.], 1998. p. 839–846.

TUKEY, J. W. Comparing individual means in the analysis of variance. Biometrics, JSTOR, p. 99–114, 1949.

YOUNG, I. T.; VLIET, L. J. V. Recursive implementation of the gaussian filter. Signal processing, Elsevier, v. 44, n. 2, p. 139–151, 1995



Um modelo de Ontologia para o Agronegócio

Ana Carolina Clivatti Ferronato[1], André Gomes de Lima[1], Aline Pires Vieira de Vasconcelos[1], Henrique Rego Monteiro da Hora[1], Rogério Atem de Carvalho[1]

[1] Sistemas Aplicados à Engenharia e Gestão, Instituto Federal Fluminense, Campos dos Goytacazes, Rio de Janeiro, Brasil, accferronato@gmail.com, gomesdelimaandre@gmail.com, alineprofcefet@gmail.com, dahora@gmail.com, ratembr@gmail.com

RESUMO

A economia Brasileira está diretamente ligada ao agronegócio, sendo uma vasta área do país destinada para esses fins, neste cenário, entender, mapear e definir uma unidade agrícola é fundamental para que a mesma tenha todo o seu potencial produtivo explorado. Este trabalho tem como objetivo a criação de uma ontologia de domínio em *Unified Modeling Language* (UML) do agronegócio voltado para a produção de cultivos do Brasil. Esta pesquisa foi realizada utilizando a base *Scopus* por intermédio do portal de periódico da capes, com o intuito de selecionar artigos com temas similares e através de dados disponíveis nas bases de dados brasileiras relacionadas ao agronegócio. Mediante as pesquisas, foi possível identificar as interações entre os diversos setores da agronomia brasileira sendo possível a criação da ontologia. A inserção de novas tecnologias proporcionará uma maior produtividade à agronomia brasileira, proporcionando benefícios à economia e promovendo um maior reconhecimento deste setor em escala global.

PALAVRAS-CHAVE: Modelagem, Produção de Cultivos, Processo, UML.

ABSTRACT

The Brazilian economy is directly linked to agribusiness, it is known that a vast area in our country is destined for these purpose, in this scenario, the act of understanding, mapping and defining an agricultural unit is fundamental so that it has all its productive potential exploited. This work aims at the creation of an agribusiness ontology using the *Unified Modeling Language* (UML) and focusing on crop production in Brazil. This research was carried out using the *Scopus* database through the capes' journal portal to search and select articles with similar themes and through this available data in Brazilian databases related to agribusiness. Over the researches, it was possible to identify the interactions between the different sectors of the Brazilian agronomy, turning the creation of the ontology something possible. The insertion of new technologies will provide greater productivity to Brazilian agronomy, providing benefits to the economy and promoting greater recognition of this sector on a global scale.

KEYWORDS: Modeling, Crop Production, Proccess, UML.

INTRODUÇÃO

A agricultura tem experimentado mudanças significativas com o passar das últimas décadas, tanto mudanças estruturais quanto tecnológicas. As mudanças estruturais podem ser vistas no tamanho das áreas produtivas atuais. Uma vez que, a produção agrícola era organizada em torno de pequenas fazendas em sua maioria com o objetivo de uma produção familiar, apenas para sobrevivência dos envolvidos. De acordo com Papajorgji (2009) são justamente as grandes áreas produtivas as quais possuem as maiores possibilidades e melhores meios de aplicar novas tecnologias e, portanto, os avanços tecnológicos têm sido uma força motriz na mudança da estrutura agrícola.

A economia brasileira é fortemente baseada no agronegócio, que representa de forma direta, aproximadamente 27,48% (R$ 1,63 trilhões), do PIB brasileiro, que no ano 2016 foi de R$ 5,94 trilhões (CEPEA, 2017). A área cultivável no Brasil é de aproximadamente 305 milhões de ha de uso efetivo para o agronegócio, distribuídos em 56 milhões de unidades de agronegócio, fazendo o país ser o maior em área produtiva no mundo (IBGE, 2016). Este cenário pressupõe a existência de demanda da criação e implementação de mecanismos de melhoria da sustentabilidade e da rentabilidade do agronegócio. À vista disso o proposto trabalho possui um enfoque exclusivamente na produção de cultivos.

Uma vez que novas tecnologias continuem a emergir, os agricultores e pesquisadores agrícolas são confrontados com uma quantidade demasiada de informações que são utilizadas para analisar e combinar a fim de gerenciar com êxito todas as áreas da produção agrícola. Este cenário requer abordagens novas e complexas para o gerenciamento das unidades de agronegócio. Um primeiro passo para o desenvolvimento de um ambiente tecnológico ligado ao agronegócio consiste em entender, mapear e definir uma unidade agrícola. É neste âmbito que encontram-se as ontologias, que por sua vez, possuem o objetivo de definir as categorias gerais ou primitivas utilizadas para classificar todas as entidades do mundo, tais como seres humanos, animais, plantas, e assim por diante (PINET et al., 2009).

Atualmente, na literatura há trabalhos publicados restringindo as ontologias em partes específicas do domínio do agronegócio, como por exemplo, em Tao e Zhao (2012) onde é modelada uma ontologia para o comércio eletrônico de vegetais. As informações que compõem esse domínio estão ligadas a alguns tipos de classes de vegetais e a relação de hierarquia entre as mesmas, a partir desta modelagem é implementado um modelo baseado em ontologia de recuperação de informações, que é comparado aos mecanismos de recuperação de informações baseados em palavras-chave. Os resultados do proposto trabalho, mostram precisão da recuperação de informações, baseadas em ontologias é maior do que a

comparada. Enquanto em Ilic et al. (2006), é criada uma ontologia denominada Ontologia da Estrutura das Plantas (PSO), a qual é a primeira representação ontológica genérica da anatomia e morfologia de uma planta florida. A PSO é destinada a uma vasta comunidade de pesquisa de plantas, incluindo cientistas, responsáveis por bancos de dados genômicos e bioinformáticos. O PSO foi desenvolvido para ser utilizado em bancos de dados públicos especializados e possui mais de 10.000 anotações de genes e descrições de fenótipo dos bancos de dados participantes.

Além disso, em Santos Junior (2014) são mapeados os processos da produção de fibra de algodão no Brasil, onde é utilizada uma ontologia de domínio a qual resulta em conhecimento para a criação de um padrão de metadados, a fim de melhorar o processo de intercâmbio de dados no segmento estudado. Portanto, a partir dos trabalhos publicados na literatura é notório que não há um modelo genérico que aborde os processos do agronegócio, ou que modele as etapas deste segmento. O que é o objetivo do proposto trabalho, modelar uma ontologia para o agronegócio com o enfoque na produção de cultivos, e utilizando a UML (*Unified Modeling Language* – Linguagem de modelagem unificada), que representa uma notação padrão para a modelagem de sistemas de informação, permitindo que a partir dos conceitos modelados, outros modelos de sistemas possam ser gerados, como o modelo de software.

Além da linguagem UML ser utilizada com sucesso na modelagem de sistemas computacionais, e em função disto, ser amplamente conhecida por grande parte dos profissionais da área, há na literatura estudos que reconhecem os benefícios de se utilizar uma ferramenta de modelagem padronizada, como a UML, na construção de ontologias conforme apresentado em Pinet et al. (2009).

MATERIAL E MÉTODOS

Referencial Teórico

A inserção da agricultura nas novas tecnologias proporciona aos cientistas uma ampla e quase inexplorada área de estudo, além de auxiliar todos os agentes relacionados a sua prática, seja direta ou indiretamente. Uma melhoria em um processo de plantio traz consigo melhorias e lucros não só ao produtor, mas às cooperativas, fábricas de maquinários, empresas de transporte e até mesmo ao país exportador do cultivo produzido.

Nos dias atuais, a necessidade de integrar todo tipo de processo à uma moderna tecnologia gerou a E-agricultura, ou seja, a automatização e atualização dos processos do agronegócio. Segundo PANI; MISHRA (2015) a E-agricultura pode ser entendida como um

sistema de informação baseado na web que fornece aos agricultores, ou a qualquer usuário envolvido no processo produtivo, informações sobre culturas, recursos agrícolas, a nutrição das plantas, as condições climáticas para uma cultura em particular, várias tecnologias e informações de mercado, etc.

Os sistemas que vêm sendo desenvolvidos para a área da agricultura fornecem recursos sobre o domínio do agronegócio, sendo ele na produção de cultivos ou na pecuária, e trazem consigo a necessidade de um mapeamento detalhado da área para então se formar uma base consistente de dados para o domínio, tendo a certeza de que esta é formada apenas por dados significativos, o que torna a solução muito precisa. A criação de uma ontologia facilita a interação de agentes da área técnica e prática e possibilita o compartilhamento e padronização da estrutura das informações sobre o domínio em questão, possibilita o reuso do domínio de conhecimento modelado, torna inteligíveis as hipóteses sobre o domínio, além de separar o conhecimento operacional do conhecimento do domínio.

É dessa maneira então que, de acordo com PALMA et. al. (2016) os modelos de negócio e processo são aplicados, visando uma padronização pois estes sistemas de informação não possuem vocabulário especificado e não têm um significado formal para os termos. Para superar tal inconsistência, ontologias descritas em OWL (*Web Ontology language*), linguagem para definir e instanciar ontologias na Web, são utilizadas em aplicações baseadas na web que constroem dados semânticos para qualquer domínio. Porém, ao tratar-se de ontologias de domínio modeladas a fim de alcançar viáveis requisitos de software, há a necessidade de um padrão teórico-metodológico que é inexistente, segundo SALES; ALMEIDA; GOMES (2008), quando utilizadas as linguagens específicas para modelagem de ontologias, por exemplo OWL e RDF (*Resource Description Framework*) linguagem utilizada para representação de informação na Web .

É nesta lacuna então que, de acordo com KOGUT et al. (2002), a modelagem de ontologias em linguagens como UML é consistente e promissora. Uma vez que ontologias incluem hierarquias de classe e subclasse, relações entre classes, definições de atributos de classe e axiomas que especificam restrições, e todos estes conceitos podem ser representados em um diagrama de classes. Portanto, o uso da UML neste âmbito fundamenta-se no fato de que a UML é uma notação gráfica baseada em muitos anos de experiência, em análise de software e design, por uma variedade de empresas em um amplo espectro de indústrias e domínios, além de possuir mecanismos padrão para definir extensões para contextos específicos e ser amplamente adotada tanto na indústria, quanto no ensino.

A UML desenvolvida pelo OMG (*Objective Managmente Group*) é uma linguagem que, por intermédio de diagramas, auxilia na especificação e documentação tanto para modelos de sistemas de softwares como também para modelagem de modelos de negócios, permitindo uma ampla visualização do objeto desejado (OMG, 2017). Segundo as definições para a modelagem em UML descritas em (BOOCH; RUMBAUGH; JACOBSON, 2012), é determinado que diagramas de classes são utilizados para representar todos os tipos de objetos presentes no cenário e os vários relacionamentos estáticos entre eles, utiliza-se a perspectiva conceitual, na qual é possível representar apenas os conceitos do domínio em estudo sem entrar no mérito de interfaces, métodos ou implementação.

Trabalhos Relacionados

Para realização desta pesquisa foi realizado um levantamento bibliométrico utilizando a base *Scopus*, por intermédio do portal da Capes. Foi possível analisar a quantidade de trabalhos publicados com enfoque na área de agronegócio e ontologia utilizando a linguagem UML. A fim de buscar a maior quantidade de trabalhos na base de pesquisa, foram utilizados como palavras chaves as áreas de pesquisas e seus tesauros. Os critérios utilizados no levantamento podem ser observados na tabela.

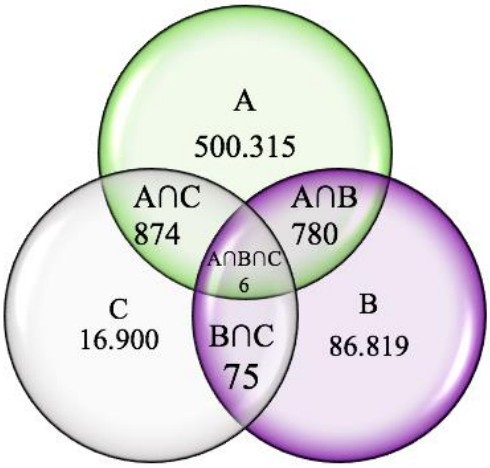
Tabela 1 - Critérios utilizados para o levantamento bibliométrico utilizando a base Scopus

Área da pesquisa e seus tesauros	Identificação
Title-abs-key (“agribuse*” or “agricultur*” or “agribusine*”)	A
Title-abs-key (“Ontolog*” or “antolog*”)	B
Title-abs-key (“UML” or “Unified Modeling Language”)	C
Title-abs-key (“agribuse*” or “agricultur*” or “agribusine*) and (“Ontolog*” or “antolog*”)	A∩B
Title-abs-key (“agribuse*” or “agricultur*” or “agribusine*) and (“UML” or “Unified Modeling Language”)	A∩C
Title-abs-key (“Ontolog*” or “antolog*”) and (“UML” or “Unified Modeling Language”)	B∩C
Title-abs-key (“agribuse*” or “agricultur*” or “agribusine*) and (“Ontolog*” or “antolog*”) and (“UML” or “Unified Modeling Language”)	A∩B∩C

Fonte: (O AUTOR)

A Figura 1 ilustra a quantidade de trabalhos que a busca retornou em cada área e nas intercessões quando comparado a outra. É possível observar que separadamente esses assuntos possuem uma quantidade relevante de pesquisas, demostrando que elas foram bastante estudadas pelo meio científico. No entanto, a interseção entre as três áreas de pesquisas revela a existência de poucos trabalhos com essas áreas em comum, revelando ser uma área de estudo promissora.

Figura 1 - Diagrama de Venn



Fonte: (O AUTOR)

Mediante o resultado obtido, foi realizada uma triagem com o intuito de selecionar os artigos mais pertinentes para o desenvolvimento desta pesquisa. A partir da tabela 2 podem ser observados esses trabalhos, assim como a contribuição que os mesmos forneceram ao meio científico.

Tabela 2 - Principais trabalhos encontrados

Autores	Contribuição
The Use of UML as a Tool for the Formalisation of Standards and the Design of Ontologies in Agriculture (PINET et al., 2009)	Os autores realizaram um estudo de caso baseado no agronegócio em animais e monitoramento sanitário destacando a relevância da ontologia para a integração entre os sistemas de informação e como a UML pode ser utilizada para descrever ontologia, possuindo semelhanças com a OWL.
Ontology-Based Simulation Applied to Soil, Water, and Nutrient Management (BECK et al., 2009a)	Destacaram a importância da ontologia como base para simulação. Foi realizado um estudo de caso utilizando a simulação baseada em ontologia para gestão de solo, água e nutrientes.
An ontology-based information retrieval model for vegetables e-commerce (TAO; ZHAO, 2012)	Um modelo baseado em ontologia de recuperação de informações foi implementado, e que possui mais funções do que a recuperação de informações por palavra-chave da web.
Ontologia aplicada ao mapeamento de classe de dados para a melhoria do intercâmbio de dados na cotonicultura no Brasil SANTOS JUNIOR (2014)	O autor apresenta o mapeamento dos processos da produção de fibra de algodão no Brasil, por meio de uma ontologia de domínio, viabilizando a criação de um padrão de metadados.
Building semantics of E-agriculture in India (PANI; MISHRA, 2015)	Os autores modelaram diferentes sistemas agrícolas da Índia em diagramas de classes utilizando UML. Criada uma ontologia para o sistema agrícola em OWL.

Fonte: (O AUTOR)

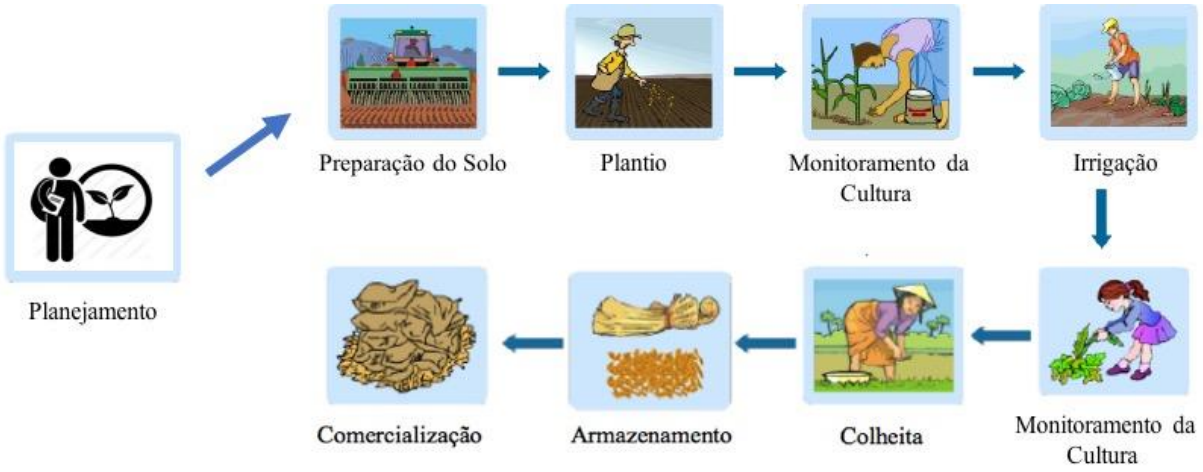
Em relação aos trabalhos previamente publicados nesta área de estudo, o presente artigo apresenta um modelo que abrange todo o processo de cultivos no agronegócio, sem especificar o produto cultivado ou técnicas utilizadas. São modeladas todas as etapas que compõem o processo, o que não é apresentado em trabalhos encontrados na literatura.

RESULTADOS E DISCUSSÃO

A Ontologia

A partir de informações levantadas em (SENAR, 2012) inicia-se o processo de identificação e mapeamento das etapas existentes na produção de cultivos. Foram definidas oito etapas, sendo a de monitoramento da cultura replicada durante a execução do processo, totalizando 9 etapas que garantem o funcionamento do processo inteiro, a Figura 2 apresenta um fluxo no qual as etapas são realizadas.

Figura 2 - Etapas do agronegócio.



Fonte: Adaptado de (GREEN SCHOOLS, 2016)

A primeira etapa que consiste no planejamento da produção do cultivo, é onde o engenheiro agrônomo, profissional responsável pelo estudo do solo e da cultura, realiza juntamente com o proprietário da unidade agrícola o planejamento da produção. Esta etapa está incluída no modelo devido a necessidade, identificada pelos autores, de levantar informações e realizar processos de tomada de decisões que são fundamentais para todo o gerenciamento da produção. Já as etapas 2 e 3, preparação do solo e plantio respectivamente, são realizadas pelo funcionário da unidade sob supervisão do engenheiro agrônomo seguindo as decisões do proprietário.

Na sequência, há a etapa de monitoramento da cultura onde o engenheiro agrônomo volta a estudar o cultivo para então apresentar soluções ao proprietário quanto a necessidade de aplicação de correções no solo, pesticidas e fertilizantes. Esta etapa é repetida entre o plantio e a colheita da cultura e é realizada pelo funcionário da unidade.

As etapas de irrigação, colheita e armazenamento acontecem sob os mesmos aspectos das etapas 2 e 3, onde o responsável por sua realização é o funcionário, mas sempre seguindo as análises do engenheiro e decisões do proprietário. Por fim, há a etapa de comercialização

cujo responsável é o proprietário. A Tabela 3 abaixo, criada pelos autores a partir de informações disponíveis em SENAR (2012), detalha e especifica as ações realizadas em cada etapa e seus respectivos envolvidos.

Tabela 3 - Descrição das etapas do Agronegócio com enfoque na produção de cultivos

Etapas	Responsáveis	Atividades
Planejamento	Eng. Agrônomo e Proprietário	• Realizar a análise do solo • Definir qual o sistema de irrigação será utilizado • Determinar quais máquinas, ferramentas e equipamentos serão mais adequados ao cultivo • Determinar quais correções devem ser feitas no solo antes do plantio
Preparação do Solo	Funcionário	• Limpar a área • Aplicar o corretivo de solo • Aplicar pesticida • Aplicar fertilizante
Plantio	Funcionário	• Plantar as sementes utilizando a semeadora • Aplicar fertilizante • Aplicar herbicidas
Monitoramento da Cultura	Eng. Agrônomo, Proprietário e Funcionário	• Aplicar herbicidas • Aplicar fungicidas • Aplicar praguicidas • Realizar capina manual
Irrigação	Funcionário	• Realizar irrigação
Colheita	Funcionário	• Utilizar colheitadeira para a colheita da cultura
Armazenamento	Funcionário	• Transportar a colheita para silos ou cooperativas
Comercialização	Proprietário	• Comercializar a colheita

Fonte: (O AUTOR)

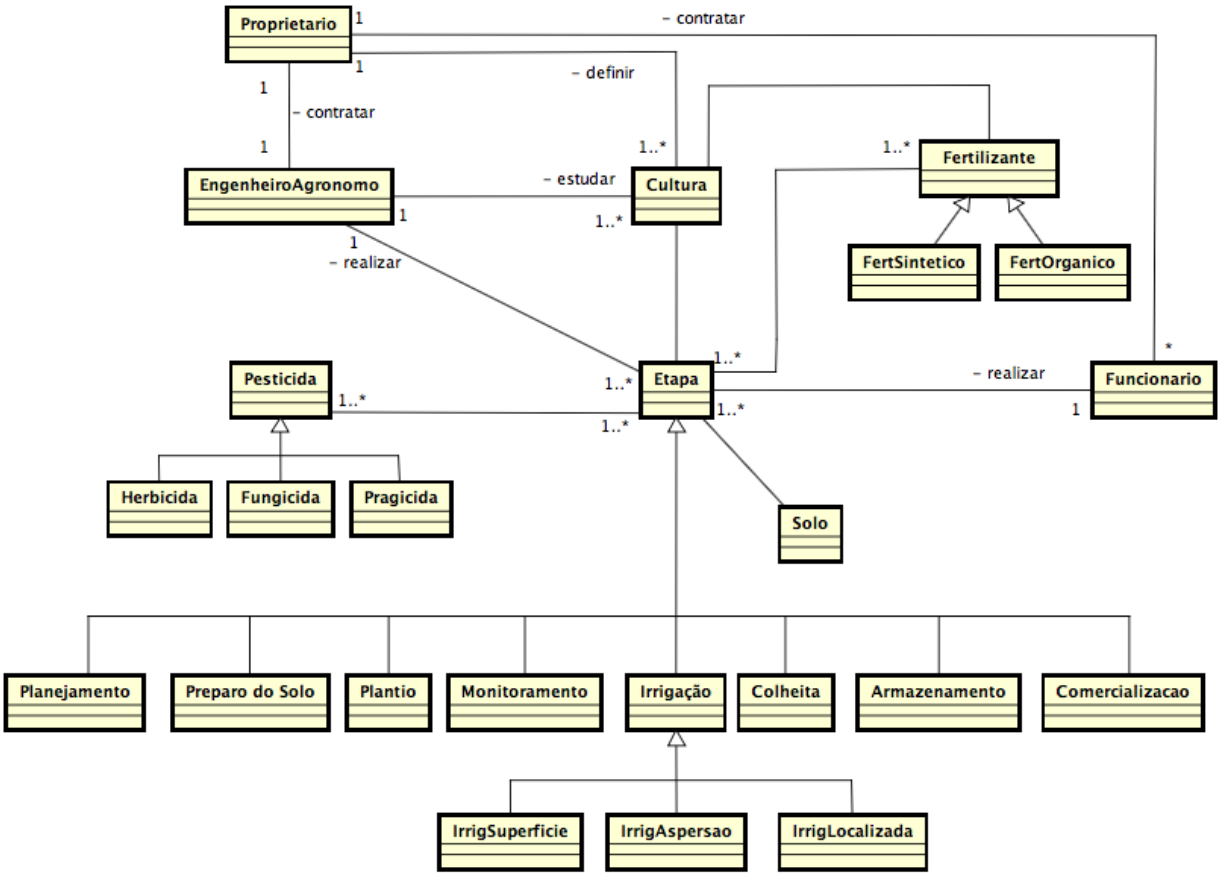
Com base no mapeamento das informações de cada etapa, é realizada a modelagem do cenário, gerando-se um diagrama de classes geral e um detalhado para cada etapa do processo, utilizando a linguagem de modelagem. Neste artigo, apenas o diagrama geral é apresentado, devido a questão de limitação de espaço e número de páginas. A Figura 3, apresenta a modelagem geral da ontologia criada, representada por um diagrama de classes, e define graficamente todos os conceitos descritos anteriormente na Tabela 3.

CONCLUSÕES

A economia brasileira é fortemente ligada ao agronegócio, por esse motivo uma melhoria neste setor, por meio de implementação de novas tecnologias torna viável uma melhor produtividade no mesmo, o que proporciona ao país retornos não só econômicos como também de investimento em tecnologia, possibilitando um maior reconhecimento do país em uma escala global. Neste cenário um setor de agronegócio planejado e minuciosamente executado é necessário com a inserção de tecnologias a fim de aproveitar melhor as áreas disponíveis e gerir os recursos necessários.

Neste presente trabalho foi possível realizar um estudo sobre o agronegócio voltado para produção de cultivos no Brasil. Foi realizada a análise de cada etapa desse processo, permitindo a modelagem de uma ontologia de domínio que padroniza a visualização do seus conceitos e seus relacionamentos estáticos, ou seja, a forma na qual elas interagem. Mediante essa ontologia, um conhecimento prévio desse setor é fornecido a futuros agentes que a utilizarão, facilitando um planejamento antecipado, uma vez que a ontologia serve de base para outros modelos de sistemas, como modelos de software o que proporciona uma solução de pesquisa melhor e mais eficiente na área estudada.

Figura 3 - Ontologia para o Agronegócio.



Fonte: (O AUTOR)

REFERÊNCIAS

An INSPIRE-Based Vocabulary for the Publication of Agricultural Linked Data. In: An INSPIRE-Based Vocabulary for the Publication of Agricultural Linked Data. [s.l: s.n.]. p. 124–133.

BECK, H. et al. Ontology-Based Simulation Applied to Soil, Water, and Nutrient Management. In: PAPAJORGJI, P. (Ed.). . Advances in Modeling Agricultural Systems. Boston, MA: Springer US, 2009a. v. 25p. 209–242.

BECK, H. et al. Ontology-Based Simulation Applied to Soil, Water, and Nutrient. Advances in Modeling Agricultural Systems, 2009b.

BECK, H. et al. Ontology-Based Simulation Applied to Soil, Water, and Nutrient Management. In: PAPAJORGJI, P. (Ed.). . Advances in Modeling Agricultural Systems. Boston, MA: Springer US, 2009c. v. 25p. 209–242.

BOOCH, G.; RUMBAUGH, J.; JACOBSON, I. UML: guia do usuário : o mais avançado tutorial sobre Unified Modeling Language (UML), elaborado pelos próprios criadores da linguagem. Rio de Janeiro: Campus : Elsevier, 2012.

CEPEA, 2016. PIB do Agronegócio Brasileiro - Centro de Estudos Avançados em Economia Aplicada - CEPEA-Esalq/USP. Desenvolvimento de Sites.

GREEN SCHOOLS. Food and Energy - Different ways to Farm. In: GREEN SCHOOLS - NOVA SCOTIA. Nova Scotia, 2016. Disponível em: <https://www.greenschoolsns.ca/feeling-resourceful>. Acesso em: 30 mar. 2017

IBGE. Anais Estatísticas Agropecuária Brasil, 2015. Disponível em: <www.ibge.gov.br>

ILIC, K. et al. The Plant Structure Ontology, a Unified Vocabulary of Anatomy and Morphology of a Flowering Plant. PLANT PHYSIOLOGY, v. 143, n. 2, p. 587–599, 22 dez. 2006.

KOGUT, P. et al. UML for ontology development. The Knowledge Engineering Review, v. 17, n. 1, mar. 2002.

OMG. What is UML | Unified Modeling Language. Disponível em: <http://www.uml.org/what-is-uml.htm>. Acesso em: 26 maio. 2017.

PANI, S.; MISHRA, J. Building semantics of E-agriculture in India: Semantics in e-agriculture. IEEE, dez. 2015

PAPAJORGJI, P. Advances in Modeling Agricultural Systems. Boston, MA: Springer US, 2009. v. 25

PINET, F. et al. The Use of UML as a Tool for the Formalisation of Standards and the Design of Ontologies in Agriculture. In: PAPAJORGJI, P. (Ed.). Advances in Modeling Agricultural Systems. Boston, MA: Springer US, 2009. v. 25p. 131–147.

SALES, L. F.; ALMEIDA, M. L. DE; GOMES, H. E. Ontologias de Domínio: Um estudo das relações conceituais. Perspectivas em Ciência da Informação, v. 13, n. 2, p. 62–76, 2008.

SANTOS JUNIOR, C. DOS. Ontologia aplicada no mapeamento de classe de dados para a melhoria do intercambio de dados na cotonicultura no Brasil. Tese (Doutorado)—São Paulo: Escola Politécnica da Universidade de São Paulo - Departamento de Engenharia de Computação e Sistemas Digitais, 2014.

SENAR. Hortaliças: cultivo de hortaliças raízes, tubérculos, rizomas e bulbos. Serviço Nacional de Aprendizagem Rural. Brasília, BR: SENAR, 2012. v. I

TAO, T.; ZHAO, M. An Ontology-Based Information Retrieval Model for Vegetables E-Commerce. Journal of Integrative Agriculture, v. 11, n. 5, p. 800–807, maio 2012.

Avaliação de métodos de detecção de tópicos em pré-processamento para classificação de textos agrícolas

Flavio M. M. Barros[1], *Stanley R. M. Oliveira*[2]

[1]Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas
São Paulo, Brasil
flavio.barros@feagri.unicamp.br

[2]Embrapa Informática Agropecuária,
Campinas, São Paulo, Brasil
stanley.oliveira@embrapa.br

RESUMO

A classificação de documentos é uma tarefa útil para agrupamento, organização e também para recomendação de textos. Em particular, na agricultura em que boa parte dos usuários (e.g., agricultores e produtores rurais) de sistemas de informações não têm grande experiência com a tecnologia da informação, a classificação de documentos pode ser utilizada para gerar recomendações de leitura. Neste trabalho, buscou-se construir e comparar modelos capazes de diferenciar textos sobre a cultura da cana-de-açúcar de outros textos relacionados a outras culturas ou criações. Para criar modelos de classificação de textos, os dados são transformados em matrizes termos-documentos, de forma que os dados apresentam alta dimensionalidade. Para construir melhores modelos de classificação de textos agrícolas foram testados: a) métodos de redução de dimensionalidade utilizando LDA (Latent Dirichlet Allocation) e PCA (Principal Component Analysis); b) número de tópicos/componentes principais; c) unigrama/bigrama; e d) algoritmos Random Forest, Gradiente Boosting e SVM (Support Vector Machine), de forma a determinar os fatores que mais impactam o AUC (Area Under the Curve). Os resultados demonstraram que os fatores estatisticamente significativos são o tipo de pré-processamento, com vantagem para LDA, e o tipo de algoritmo utilizado, com destaque para o SVM. O número de tópicos e de componentes principais e o uso de unigrama e bigrama não tiveram efeito estatisticamente significativo na performance dos modelos em termos de AUC.

PALAVRAS-CHAVE: Mineração de textos, Aprendizado de máquina, Redução de dimensionalidade, Sistema de informação agrícola.

ABSTRACT

Document classification is a useful task for grouping, organizing and also for text recommendation. In particular, in agriculture in which a large proportion of users (E.g., farmers and rural producers) of information systems do not have much experience with information technology, document classification can be used to generate reading recommendations. In this work, we tried to construct and compare models capable of differentiate texts about the culture of sugarcane from other texts related to other cultures or creations. To create text classification models, the data is transformed into term-documents matrices, so that the data present high dimensionality. For building better models for classification of agricultural texts, we have tested: a) methods for dimensionality reduction using LDA (Latent Dirichlet Allocation) and PCA (Principal Component Analysis); b) number of topics / main components; c) unigram / bigram; and d) algorithms Random Forest, Gradient Boosting, and Support Vector Machine (SVM) to determine the factors that most impact the AUC (Area Under the Curve). The results demonstrated that the statistically significant factors are the type of pre-processing, with advantage for LDA, and the type of algorithm used, highlighting the SVM one. The number of topics and principal components, the use of unigram and bigram did not have a statistically significant effect in the performance of the models in terms of AUC.

KEYWORDS: Text mining, Machine learning, Dimensionality reduction, Agricultural information systems.

INTRODUÇÃO

A agricultura é um dos setores protagonistas da economia brasileira cada vez mais dependente de informação. Essa dependência se dá pelas necessidades advindas das constantes mudanças tecnológicas, pela operação em mercados globais e pela extrema sensibilidade às mudanças ambientais (climática, hidrológica, solo e etc). Assim o processo de tomada de decisões de longo e médio prazos, em resposta a esses e outros fatores, requer um entendimento e o acesso rápido e fácil a uma grande quantidade de informações técnicas agrícolas (CASH, 2001). O acesso facilitado à informação de qualidade é particularmente crítico para pequenos produtores, que em geral têm maiores dificuldades de acesso à informação sobre as melhores práticas na agricultura (PARIKH; PATEL; SCHWARTZMAN, 2007).

No Brasil e no mundo há diversas iniciativas com o objetivo de ofertar informações técnicas agrícolas de qualidade. E especialmente no Brasil, destaca-se um projeto da EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária) chamado Agência Embrapa de Informação, um dos portais mais acessados da internet brasileira na área de agricultura e que compreende centenas de páginas web, artigos, planilhas e materiais multimídia.

A fonte primária de informação do portal são textos técnicos agrícolas que condensam o conhecimento de anos de pesquisa em diversas unidades da Embrapa em todo o Brasil, compreendendo mais de 33 cultivos e criações, como cana-de-açúcar, soja, agronegócio do leite,

dentre outros. Todo o material está organizado em uma estrutura hierárquica, tal que textos relacionados a culturas específicas, como cana-de-açúcar, são agrupados em tópicos dentro do portal. A organização do material é feita pelos próprios pesquisadores da Embrapa e não é feita de forma automática.

De acordo com Barros et al. (2013), após uma investigação sistemática dos dados de uso da Agência, os pesquisadores concluíram que os os usuários podem não estar fazendo uso da totalidade de informações técnicas disponíveis. Muitos usuários visualizam somente uma parte muito pequena do material disponível por causa da dificuldade de encontrar a informação desejada em um grande volume de páginas Web. Ainda que as páginas estejam organizadas em uma estrutura de tópicos, há a possibilidade de que os usuários não tenham conhecimentos técnicos para encontrar rapidamente a informação desejada. Outro desafio é que a organização humana dos conteúdos pode deixar de lado relações importantes entre assuntos disponíveis em páginas diferentes, de forma que mesmo que um usuário utilize a estrutura hierárquica de tópicos ainda pode ter dificuldades em encontrar um texto correspondente às suas necessidades.

Em cenários como esse, um sistema automatizado, capaz de identificar o assunto relacionado a uma página sendo visitada por um usuário, poderia utilizar esta informação para gerar recomendações e/ou entender o perfil dos usuários. As informações do perfil da comunidade e as recomendações podem facilitar a navegação e exploração dos conteúdos da Agência. Entretanto, esta tarefa coloca diversos desafios técnicos: a) como obter a melhor representação estruturada do texto?; b) como escolher técnicas de classificação com bom desempenho no contexto?; c) como usar a representação estruturada para exploração dos hábitos de uso do portal?

Para abordar esses desafios, este trabalho teve como objetivo criar representações dos textos da Agência, utilizando ferramentas para detecção automática de tópicos e modelos de aprendizado de máquina capazes de classificar textos agrícolas sem intervenção humana. Especificamente, foram avaliados vários métodos de detecção de tópicos utilizados na etapa de pré-processamento para classificação de textos agrícolas.

MATERIAL E MÉTODOS

A metodologia deste trabalho é composta de algumas etapas relacionadas à detecção de tópicos de interesse para classificação de textos agrícolas.

Origem dos Dados

A coleção de textos utilizada foi extraída da Agência Embrapa de Informação [1], constituída de mais de 30 culturas agrícolas, com abrangência nacional. O usuário tem acesso a todo o conteúdo do site na forma de textos, artigos, arquivos de imagem, arquivos de som e planilhas eletrônicas. Todo o conteúdo foi organizado para atender pesquisadores, produtores rurais, profissionais de assistência técnica e extensionistas. De acordo com (BERTIN; LEITE; PEREIRA,

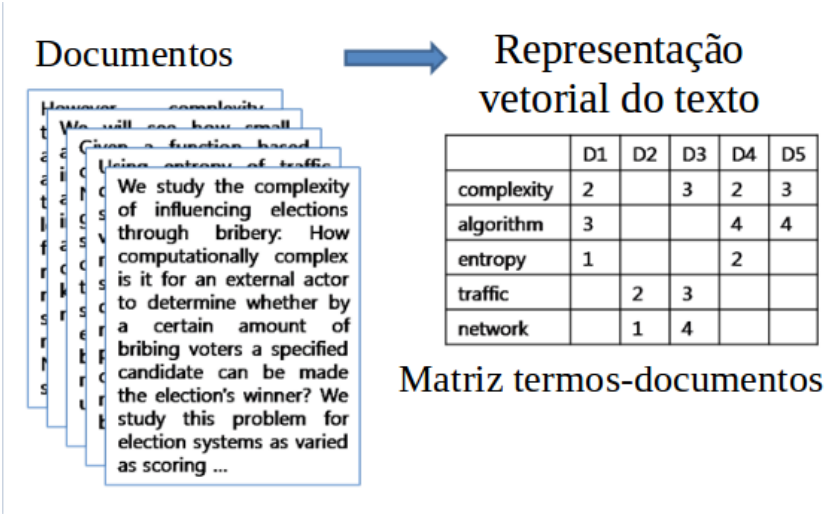
[1] http://www.agencia.cnptia.embrapa.br

2009), a EMBRAPA tem mantido uma política de oferta de informações técnicas para pesquisadores, produtores e a sociedade em geral, com a missão de fazer chegar aos agentes interessados os resultados da pesquisa científica, relativa a toda cadeia produtiva do agronegócio, com o objetivo de criar uma robusta infraestrutura social, técnica e econômica tão necessária ao processo de desenvolvimento. Todos os textos agrícolas utilizados neste artigo foram retirados de páginas web nesse portal.

Os dados utilizados neste artigo foram textos agrícolas completos das páginas da Agência Embrapa de Informação sobre culturas como, por exemplo, a cana-de-açúcar. Devido ao grande volume de dados disponível no portal foi utilizada uma amostra aleatória de 1.500 páginas diferentes, contando com 1.086 páginas com tema sobre a cana-de-açúcar e 414 páginas relacionadas a outros cultivos, como arroz, açaí dentre outros. Destaca-se que textos relacionados à cultura da cana-de-açúcar são mais numerosos e somente esta cultura corresponde a mais de 45% dos acessos ao portal (BARROS et al., 2013).

Processamento de Dados

O processamento dos dados foi constituído basicamente de duas etapas: 1) a transformação dos textos em uma matriz de termos-documentos e 2) extração de tópicos/componentes principais. A primeira etapa é característica de qualquer análise de mineração de textos e consiste basicamente em transformar dados não estruturados, nesse caso textos, em dados estruturados em uma tabela, onde as colunas (ou linhas) são os termos e as linhas (ou colunas) são documentos, uma representação conhecida na literatura como *bag-of-words* (FELDMAN; SANGER, 2006). Salienta-se que no caso específico deste trabalho, optou-se por usar a medida TF-IDF (*term frequency-inverse document frequency*) na construção da matriz de termos-documentos. Na Figura 1 é mostrado um exemplo desse tipo de processamento dos textos.



	D1	D2	D3	D4	D5
complexity	2		3	2	3
algorithm	3			4	4
entropy	1			2	
traffic		2	3		
network		1	4		

Figura 1: Transformação de textos em uma matriz termos-documentos. Os números representam a frequência com que cada termo aparece no respectivo documento, mas no trabalho foi utilizada a TF-IDF.

Utilizando a matriz de termos-documentos como entrada, foi utilizada a técnica de detecção de tópicos LDA (Latent Dirichlet Allocation) como descrita em Blei, Ng e Jordan (2003). Também, a partir da mesma matriz de termos-documentos foi utilizada a técnica PCA (Principal Component Analysis) para encontrar as componentes principais, como descrito em Hastie, Tibshirani e Friedman (2009). O número de tópicos/componentes principais é um parâmetro que deve ser escolhido pelo usuário, e optou-se neste trabalho em utilizar 30, 50 e 100 tópicos ou componente principais.

Modelos e Validação

A avaliação empírica neste artigo consistiu em um experimento que combinou dois tipos diferentes de técnicas de redução de dimensionalidade, LDA e PCA. Além disso também foram incluídos no experimento a avaliação do uso de unigramas e bigramas; a avaliação dos seguintes algoritmos (HAN, 2005): a) SVM b) Random Forest e c) Gradiente Boosting e o número de tópicos/componentes principais. Portanto, o experimento contou com a avaliação de quatro fatores diferentes: 1) a técnica de redução de dimensionalidade 2) bigramas ou unigramas 3) algoritmos e 4) número de tópicos/componentes principais. O objetivo principal do experimento foi comparar qual era o melhor método de redução de dimensionalidade. No entanto, apesar de se encontrar ampla literatura reportando performances de algoritmos, optou-se por incluir esses fatores no experimento deste artigo de forma a se obter a melhor configuração experimental para o contexto de textos agrícolas. Na avaliação foi utilizada a métrica Área sob a curva ROC (AUC), amplamente utilizada na literatura como métrica de comparação de classificadores (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

Para os experimentos, foi utilizada validação cruzada com 10 folds já que ajustes de hiperparâmetros foram realizados para indicar a melhor configuração entre classificadores e métodos de redução de dimensionalidade.

Ferramentas de Suporte

Todas as etapas do trabalho foram realizadas com a linguagem R[2]. Foram utilizados os pacotes `tm` para criação do Corpus de textos, stemização, limpeza, tokenização; o pacote `rvest` para web scrape; o pacote `topicmodels` para modelagem de tópicos com o LDA; o pacote `caret` para modelagem e teste; as funções `lm()` e `prcomp()` do pacote base para o ANOVA e PCA, respectivamente.

RESULTADOS E DISCUSSÃO

Foram realizados vários experimentos com o propósito de avaliar os melhores métodos para detecção e tópicos de interesse nos textos agrícolas. Quatro aspectos foram considerados:

[2]http://cran.r-project.org

redução de dimensionalidade com LDA/PCA, número de tópicos/componentes principais, uso de bigramas de unigramas e algoritmos. No total 36 configurações experimentais foram analisadas com a técnica ANOVA (DEVORE, 2008). De acordo com os resultados na Tabela 1, as técnicas de redução de dimensionalidade (Pré-processamento) e os algoritmos utilizados (Algoritmos) foram estatisticamente significativos (p-valor < 0.001); ambos, o número de tópicos/componentes principais (K) e bigrama/unigrama (ngrama), não apresentaram efeito significativo na métrica utilizada. Todos os modelos ficaram bem ajustados ao conjunto de dados ($R^2 > 0.6750$).

A Figura 2 mostra o histograma do AUC em relação aos algoritmos, usando todas as configurações experimentais consideradas nesse trabalho. Como pode ser observado, existe uma separação clara entre o desempenho dos algoritmos, com ligeira vantagem para o SVM. Percebe-se também que a distribuição do histograma para o SVM é bimodal, indicando que esse algoritmo tem efeito de interação com outro fator, no caso o uso do LDA ou do PCA.

Tabela 1: Tabela ANOVA do experimento.

	Mean Sq	F value	Pr(>F)
Pré-processamento	0.00155612	32.0432	$< 4 \times 10^{-6} ***$
K	0.00003526	0.7261	0.4009
ngrama	0.00005975	1.2304	0.2761
Algoritmos	0.00068758	14.1584	$< 5 \times 10^{-5} ***$
Residuals	0.00004856		
R^2	0.6750		
Adj. R^2	0.6209		
Num. obs.	36		

***p-value< 0.001

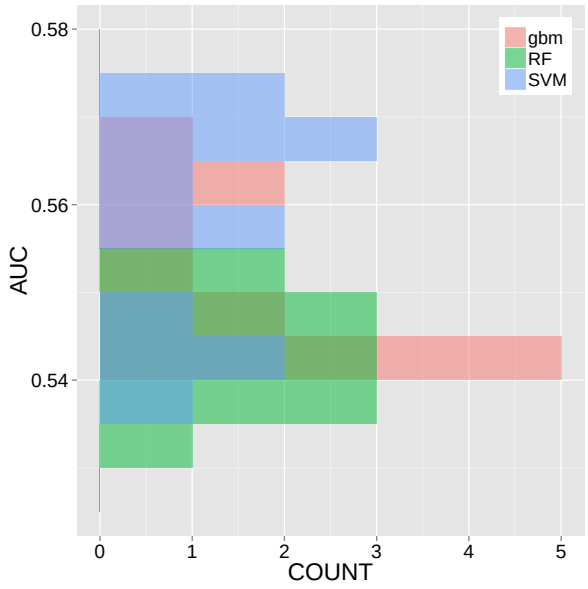


Figura 2: Histograma de AUC em relação aos algoritmos.

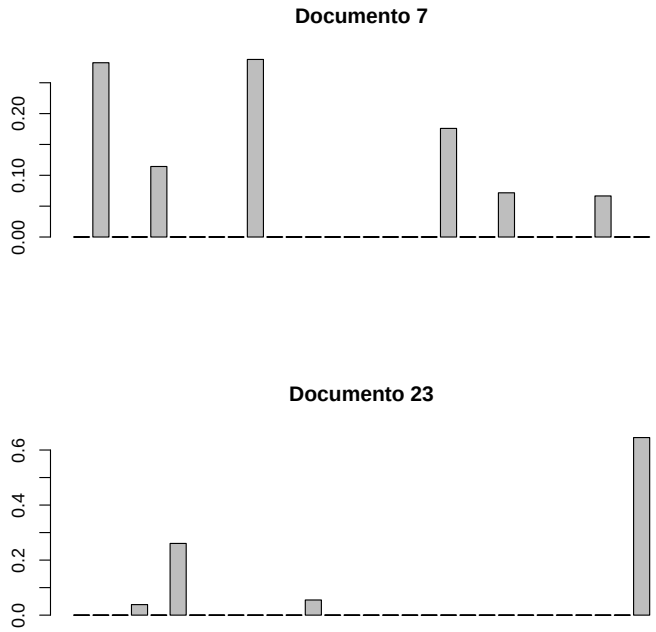


Figura 3: Exemplo de distribuição da probabilidade de tópicos para dois documentos da base de dados.

Na Figura 3 são apresentados dois exemplos da distribuição da probabilidade de tópicos em dois documentos da coleção, o Documento 7 e o Documento 23. Os valores representados nas barras são probabilidades tal que o valor total somado de todas as barras deve ser 1. Ele indica a probabilidade da presença de um certo tópico (ou conjunto de tópicos) em um certo documento. Assim, a estrutura de tópicos em um documento pode ser utilizada como um representação do texto, sendo uma alternativa a representação da matriz termos-documentos, onde cada termo em específico no texto é utilizado na representação. Utilizando tópicos, ao invés dos termos, como há menos tópicos, a representação do texto fica mais enxuta. Além disso, analisando as palavras que em conjunto formam cada tópico é possível fazer uma descrição qualitativa, mas sucinta, dos textos.

Na Figura 4 observa-se o histograma da métrica AUC em relação à técnica de redução de dimensionalidade utilizada. Oberva-se claramente que o LDA tem um desempenho superior na maioria das condições experimentais. Tanto na Figura 2 quanto na Figura 4, histogramas relativos aos dois fatores estatisticamente significativos da Tabela 1, observa-se que a configuração que apresenta melhores resultados é a configuração com LDA + SVM.

Como o teste estatístico da Tabela 1 rejeita a hipótese nula de que os fatores do experimento não tem efeito significativo, mas alguns fatores apresentam mais de um nível, como o fator algoritmos, para diferenciar em quais níveis há diferença estatisticamente significativa é necessário utilizar um teste *posthoc* como o Teste de Tukey (DEVORE, 2008). De acordo com os intervalos de confiança para as diferenças entre as médias para os algoritmos, no gráfico

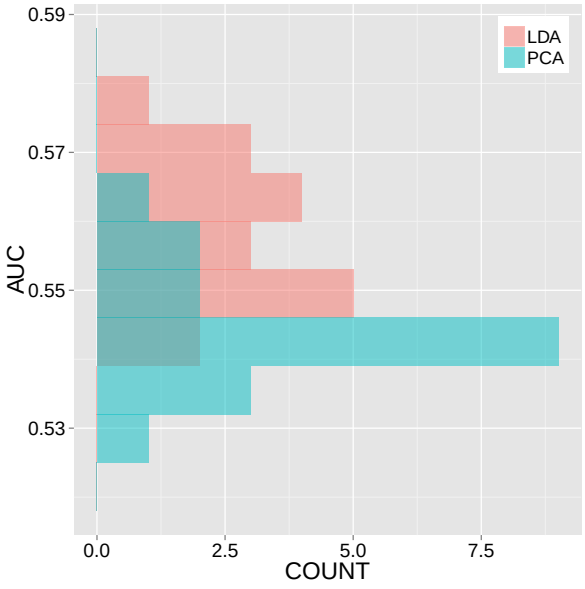


Figura 4: Histogram de AUC em relação à técnica de redução de dimensionalidade.

apresentado na Figura 5, o SVM apresenta desempenho superior aos dois outros algoritmos. Nota-se entretanto, que devido à bimodalidade da distribuição dos valores de AUC em relação ao algoritmo SVM, a diferença de desempenho entre os algoritmos SVM e Gradiente Boosting foi pequena e próxima do limite da margem de erro.

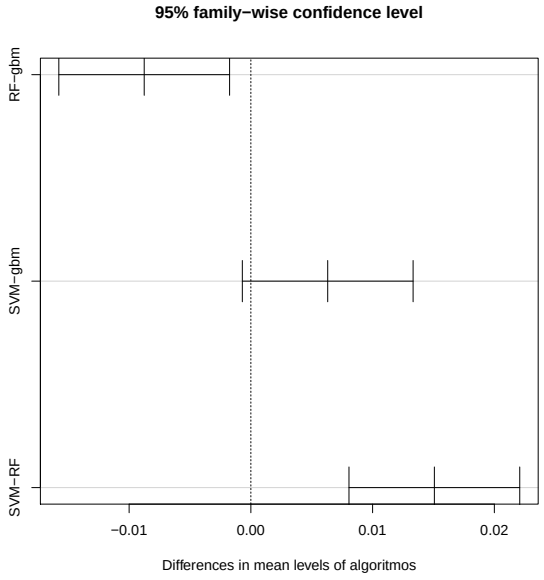


Figura 5: Intervalos de confiança para a diferença entre as médias do AUC para cada algoritmo.

Em resumo, a avaliação mostrou que o LDA usado como técnica de redução de dimensionalidade tem um desempenho superior ao PCA. Também mostrou-se que o SVM associado à técnica LDA tem um bom desempenho em termos de classificação de documentos. O uso de bigramas, pelo menos na tarefa de classificação de textos, não foi um fator significativo, tal

que pode-se optar por utilizar somente unigramas para reduzir a dimensão das matrizes termos-documento. Outra vantagem do LDA em relação ao PCA é que a técnica também permite analisar o significado de cada atributo utilizado após a redução de dimensionalidade, isto é, uma vez que cada documentos apresenta a prevalência de alguns tópicos, como apresentado na Figura 3, além da redução de dimensionalidade, esta informação pode ser utilizada de forma exploratória e para determinar os assuntos mais importantes para a comunidade de usuários.

CONCLUSÕES

Esse trabalho apresenta uma avaliação de métodos de detecção de tópicos, um procedimento utilizado no pré-processamento de classificação de textos. Vários experimentos foram realizados para avaliar os desempenhos das técnicas LDA e PCA para redução de dimensionalidade em matrizes de termos-documentos. Os resultados evidenciaram o potencial da técnica LDA para essa tarefa. O método proposto é capaz de melhorar o desempenho de classificadores em relação a técnicas tradicionais de redução de dimensionalidade como o PCA, especialmente quando se considera o algoritmo SVM na etapa principal do processo de mineração de dados.

A técnica LDA também tem mostrou grande potencial como ferramenta exploratória, podendo ser utilizada para caracterizar o interesse dos usuários em determinados assuntos. No futuro planeja-se avaliar outras técnicas de redução de dimensionalidade em comparação ao LDA, como *bag-of-related-words* e Lasso (least absolute shrinkage and selection operator).

REFERÊNCIAS

BARROS, F. d. F. M. M. de et al. Desenvolvimento e validação de um sistema de recomendação de informações tecnológicas sobre cana-de-açúcar. *Bragantia*, Instituto Agronômico, v. 72, n. 4, p. 1–9, 2013. ISSN 1678-4499. Disponível em: <http://www.scielo.br/pdf/brag/v72n4/aop_bragncea2061.pdf http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0006-87052013000400010&lng=pt&nrm=iso&tlng=en>.

BERTIN, P.; LEITE, F.; PEREIRA, F. Embrapa Technological Information: a bridge between research and society. *Agricultural Information Worldwide*, v. 2, n. 1, p. 10–18, 2009. Disponível em: <http://repositorio.bce.unb.br/handle/10482/4935>.

BLEI, D. M.; NG, A. Y.; JORDAN, M. I. Latent dirichlet allocation. *J. Mach. Learn. Res.*, JMLR.org, v. 3, p. 993–1022, mar. 2003. ISSN 1532-4435. Disponível em: <http://dl.acm.org/citation.cfm?id=944919.944937>.

CASH, D. W. In order to aid in diffusing useful and practical information: Agricultural extension and boundary organizations. *Science, Technology & Human Values*, v. 26, n. 4, p. 431–453, out. 2001. ISSN 0162-2439, 1552-8251. Disponível em: <http://sth.sagepub.com/content/26/4/431>.

DEVORE, J. *Probability and Statistics for Engineering and the Sciences, Enhanced Review Edition*. Cengage Learning, 2008. (Available 2010 Titles Enhanced Web Assign Series). ISBN 9780495557449. Disponível em: <http://books.google.co.uk/books?id=Wbym40WgsXMC>.

FELDMAN, R.; SANGER, J. *Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. New York, NY, USA: Cambridge University Press, 2006. ISBN 0521836573, 9780521836579.

HAN, J. *Data Mining: Concepts and Techniques*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2005. ISBN 1558609016.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning*. 2. ed. New York, NY: Springer New York, 2009. 745 p. (Springer Series in Statistics). ISBN 978-0-387-84857-0. Disponível em: <http://www.springerlink.com/index/10.1007/978-0-387-84858-7>.

PARIKH, T. S.; PATEL, N.; SCHWARTZMAN, Y. A survey of information systems reaching small producers in global agricultural value chains. In: *2007 International Conference on Information and Communication Technologies and Development*. IEEE, 2007. p. 1–11. ISBN 978-1-4244-1990-6. Disponível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4937421>.

Uso de ontologias e técnicas de visualização no apoio às pesquisas em eficiência alimentar de gado de leite

Heitor Magaldi[1], Regina Braga[1], Wagner Arbex[1,2,], Mariana Magalhães Campos[2], Carlos Cristiano Hasenclever Borges[1], José Maria N. David[1], Fernanda Campos[1]*

[1] Programa de Pós-Graduação em Ciência da Computação, Universidade Federal de Juiz de Fora, Juiz de Fora, Minas Gerais, Brasil, heitor.magaldi@gmail.com,{regina.braga, wagner.arbex, fernanda.campos, jose.david}@ufjf.edu.br, cchborges@ice.ufjf.br

[3] Embrapa Gado de Leite, Empresa Brasileira de Pesquisa Agropecuária – Embrapa, Juiz de Fora, Minas Gerais, Brasil, {wagner.arbex, mariana.campos}@embrapa.br

RESUMO

Com mercado atual cada vez mais competitivo, produtores leiteiros precisam reduzir custos e tornar seus rebanhos competitivos. Nesse sentido, o apoio computacional vem fornecendo alternativas a identificação de animais mais eficientes e, por consequência, proporcionando ganhos econômicos e ambientais. Este artigo apresenta uma arquitetura para apoio as pesquisas de eficiência alimentar desenvolvidas pela Embrapa Gado de Leite e, com o objetivo de, a partir do uso de ontologias e análise de dados, descobrir novo conhecimento e novas relações nas diversas bases de experimentos coletados em campo. Considerando os resultados preliminares, obtivemos indícios das vantagens do uso de ontologias e técnicas de visualização na análise e comparação dos índices de eficiência alimentar. Consideramos que o uso de ontologias e técnicas de visualização podem contribuir para o avanço das pesquisas em eficiência alimentar.

PALAVRAS-CHAVE: Ontologia, SWRL, Eficiência Alimentar, Gado de Leite.

ABSTRACT

Today's with increasingly competitive market, dairy farmers need to cut costs and make their herds competitive. In this sense, the computational support has provided alternatives to the identification of more efficient animals and, consequently, providing economic and environmental gains. This paper presents an architecture to support food efficiency research, developed by Embrapa Dairy Cattle, with the aim of discovering new knowledge and new

* Autor correspondente.

relationships in experiments, using ontologies and data analysis. Considering the preliminary results, we obtained evidence of the advantages of the use of ontologies and visualization techniques in the analysis and comparison of the indices of feed efficiency. We consider that the use of ontologies and visualization techniques can contribute to the advancement of research in feed efficiency.

KEYWORDS: Ontology, SWRL, Feed Efficiency, Dairy Cattle.

INTRODUÇÃO

Atualmente, a pecuária leiteira nacional vem lidando com novos desafios. O aumento dos custos de produção, da preocupação dos consumidores com a qualidade e segurança alimentar, do bem-estar animal e impactos ambientais da agropecuária ocasionam margens de lucro menores aos produtores (Campos et al., 2015). Diante disso, a busca por um sistema de produção de leite eficiente se torna necessário para vencer os desafios impostos.

Campos et al. (2015) destacam que os gastos com a alimentação representam o principal custo da atividade pecuária. Assim, a identificação dos animais mais eficientes no aproveitamento do alimento é alternativa econômica a ser adotada. Além disso, com a eficiência alimentar, além de impactos econômicos, impactos ambientais são observados, pois animais eficientes produzem menor desperdício de nutrientes e excreções. Um animal é classificado como eficiente quando atinge o mesmo nível de produção consumindo menos alimento que os demais.

Nesse sentido, com o objetivo de identificar animais com menor exigência energética, foram propostos diversos índices, tais como: Eficiência Alimentar Bruta (EA), Eficiência de Conversão Alimentar (ECA), Consumo Alimentar Residual (CAR), Ganho de Peso Residual (GPR), Taxa Relativa de Crescimento (TRC), Razão Keiber (RK), dentre outros (CROWLEY et al., 2010; GOMES et al., 2012; BERRY e CROWLEY, 2013; GRION et al., 2014; BONILHA et al., 2015).

Entretanto, Crowley et al. (2010) relatam que ainda não se sabe qual índice é o mais adequado, quando se trata de produção de leite. Montanholi et al. (2010) dizem que, através do entendimento dos fatores que regulam a eficiência alimentar, será possível predizer os animais mais eficientes. Assim, dificuldades na seleção e entendimento desses fatores se tornam desafiadores.

Hoje, os pesquisadores analisam esses dados de forma manual, utilizando de ferramentas estatísticas para seleção dos animais eficientes e, a partir disso, analisarem em que os animais eficientes diferem dos não eficientes. Dessa forma, uma arquitetura para o

apoio das pesquisas em eficiência alimentar, capaz de uma análise computacional dos dados, é de grande importância para os avanços das pesquisas nessa área.

O objetivo desse artigo é apresentar uma arquitetura de apoio as pesquisas em eficiência alimentar desenvolvidas pela Embrapa Gado de Leite que, a partir do uso de ontologias e análise de dados, permite a descoberta de novo conhecimento e relações nas bases de experimentos relacionados.

O artigo está organizado em quatro seções. Na seção 2, apresenta-se os materiais e métodos adotados para o trabalho. A seção 3 traz as discussões sobre os resultados; e, por fim, a seção 4, as conclusões.

MATERIAL E MÉTODOS

A grande competitividade do mercado atual exige que os produtores leiteiros busquem meios para reduzir seus custos e tornar seus rebanhos competitivos. Diante disso, surge o termo "Agricultura de precisão", definido como práticas e ferramentas de gerenciamento, compostas de sensores, atuadores e suporte à decisão, com objetivo de otimizar o processo produtivo (McBratney; Whelan; Ancev, 2005) (Bewley, J.M. 2013).

Gruber (1993) definiu ontologia como uma "especificação explicita de um conceito". Guarino, em 1998, por sua vez, descreveu que a ontologia pode ser descrita nas mais variadas formas, dependendo de seu campo de aplicação. Por exemplo, no senso filosófico, como "um sistema particular de categorias representando uma certa visão do mundo". Por outro lado, ontologias computacionais são meios para modelar formalmente a estrutura de um sistema, isto é, as entidades relevantes e suas relações que emergem a partir da sua observação e que são úteis para os objetivos desejados (Guarino et al., 2009).

Em 1999, Baker et al. destacam o crescimento do uso de ontologias no domínio de engenharia de software e aplicações web, com o intuito de promover a integração, interoperabilidade e visualização dos dados. Posteriormente, Miah, Gammack e Kerr (2007) desenvolveram um modelo ontológico capaz de centralizar o acesso aos dados de diversas bases, com objetivo de facilitar a consulta e descoberta de informações de forma mais simplificada aos usuários. Em 2015, Verhoosel, Bekkum e Evert propuseram uma abordagem com uso de ontologias com o propósito de unir bases distintas, proporcionando a análise de um grande volume de dados. Os autores desenvolveram uma ontologia composta de 28 conceitos relacionados a eficiência alimentar.

Tomic et al. (2015), em sua arquitetura *agriOpenLink*, utilizam ontologia para centralizar as consultas e acesso aos dados produzidos nos diversos serviços relacionados a agropecuária, tais como dados climáticos, eólicos, pragas, raças, entre outros.

Nesse sentido, torna-se indispensável o apoio computacional para os avanços das pesquisas agropecuárias. Assim, uma arquitetura que proporcione a padronização e interoperabilidade de dados dos experimentos, facilidade de acesso e avaliação através do uso de ontologias e técnicas de inteligência computacional é o objetivo desse artigo.

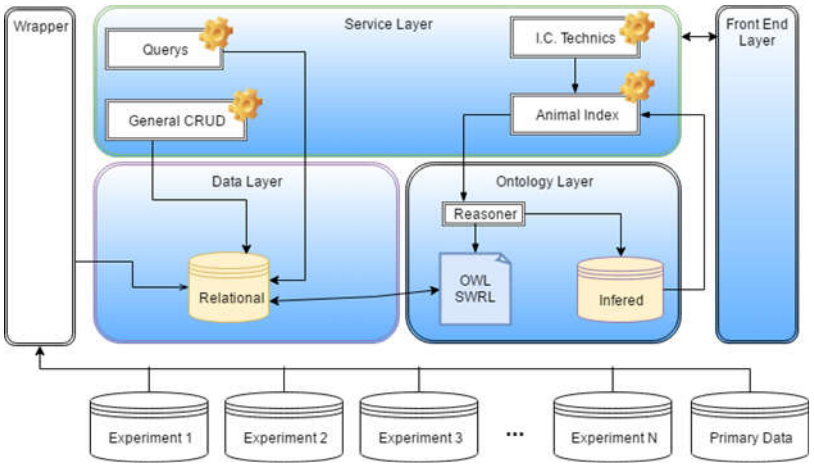
Como prova de conceito, foram utilizados dados da Embrapa Gado de Leite, relacionados a experimentos conduzidos com um grupo de 36 novilhas F1 Girolando, obtidas através do cruzamento das raças Gir e Holandesa. Os animais foram acompanhados nas principais fases da vida (Tabela 1) e tiveram seus desempenhos avaliados através de dados sanguíneos, morfométricos, consumo alimentar, entre outros. Os resultados obtidos nesses experimentos foram utilizados para verificar a viabilidade de uso da arquitetura.

Tabela 1 – Experimentos de eficiência alimentar desenvolvidos na Embrapa Gado de Leite

Fase	Idade dos Animais	Descrição do Experimento	Status
1	0 a 3 meses	Cria - Aleitamento	Concluída
2	3 a 9 meses	Recria – em confinamento	Concluída
3	9 a 15 meses	Ensaio de Metabolismo – em câmaras respirométricas	Concluída
4	15 a18 meses	Prova de eficiência – a pasto	Concluída
5	18 a 24 meses	Prova de eficiência – em confinamento	Concluída
6	-	Reprodução	Concluída
7	-	Lactação	Em curso

Inicialmente, devido aos diversos experimentos já conduzidos estarem armazenados em bases de dados heterogêneas, foi desenvolvido uma camada *wrapper* que permite a interoperabilidade dos experimentos já conduzidos para o modelo de dados da arquitetura, conforme demonstrado na Figura 1.

Figura 1 – Arquitetura proposta.



Fonte: (Autor, 2017)

Data Layer

Para o armazenamento dos experimentos, um modelo de dados baseado no modelo relacional foi desenvolvido, com a responsabilidade agora de recepcionar os dados dos novos experimentos. O banco de dados adotado foi o PostgresSQL[1].

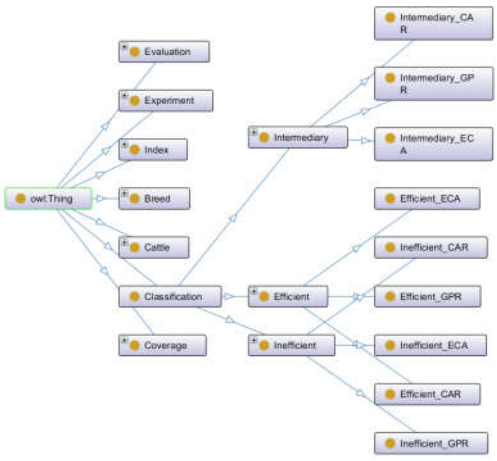
O modelo relacional é composto por informações relativas aos animais, experimentos e informações climáticas. Essas informações possibilitam a observação de características comuns aos grupos eficientes.

Ontology Layer

Para apoio a integração e análises dos dados dos experimentos, foi especificada uma ontologia, denominada Ontologia de eficiência alimentar (OEA) (Figura 2). A ontologia permite a integração semântica entre os experimentos relacionados a eficiência alimentar, permitindo aos pesquisadores a classificação dos animais nos experimentos e a interoperabilidade entre os dados, com vistas a realizar análises cruzadas e descobertas de novas conexões entre experimentos.

A linguagem adotada para implementação da ontologia foi o OWL (*Web Ontology Language*), recomendada pelo W3C (Bechhofer et al.,2004; Hitzler, Krotzsch e Rudolph, 2009). Além disso, a ontologia foi especificada para a classificação dos índices de eficiência alimentar. Assim, devido à necessidade de classificar os animais sobre os índices de CAR, GPR e ECA foram criadas classes específicas e regras específicas para realizar a classificação, considerando os três possíveis níveis: eficiente, intermediário e ineficiente (Figura 2).

Figura 2 – Ontologia de eficiência alimentar (OEA).



Fonte: (Autor, 2017)

[1] https://www.postgresql.org/

A classe cattle possui instâncias de todas as novilhas relacionadas ao experimento, correlacionando com raças, cobertura e avaliações. A classificação dos animais é realizada a partir do cálculo dos índices, cujas informações foram obtidas junto aos pesquisadores da Embrapa Gado de Leite. Para os cálculos, a arquitetura recupera dados alimentares do experimento, tais como, Ganho de Peso Diário (GPDkg), Ingestão de Matéria Seca (IMS), Peso Vivo Médio (PVmediokg) e Peso Médio (PMmediokg) de cada animal participante do experimento.

Tabela 1 – Base para apuração dos índices

Animal	GPDkg	IMS	PVmedio	PMmediokg
4065	1,07432	5,576804	228,675	58,80496964
4066	0,82611	4,509963	200,5218	53,2869988

Os três índices adotados para o experimento, apesar de fórmulas distintas, baseiam-se nos dados alimentares GPDkg, IMS e PMmediokg para suas apurações, demonstrado na Tabela 2.

Tabela 2 – Índices de eficiência alimentar

Índice	Fórmula
Consumo Alimentar Residual (CAR)	IMS Observado - IMS Esperado
Eficiência de Conversão Alimentar (ECA)	IMS / GMD
Ganho de Peso Residual (GPR)	GMD Observado - GMD Esperado

A apuração dos valores IMS esperado e GMD esperado é obtida através de regressões lineares múltiplas, conforme apresentado nas equações a seguir:

- IMS esperado

$$y = IMS \quad x1 = GPDkg \quad x2 = PMmediokg$$
$$\beta 0 + \beta 1 * GPDkg + \beta 2 * PMmediokg$$

- GMD esperado

$$y = GPDkg \quad x1 = IMS \quad x2 = PMmediokg$$
$$\beta 0 + \beta 1 * IMS + \beta 2 * PMmediokg$$

Os animais, sob os índices de CAR e ECA, são considerados eficientes quando menores forem os valores obtidos, ao contrário do índice GPR que quanto maior, melhor. O critério adotado pela Embrapa Gado de Leite para classificar animais intermediários foi o de 0.5 desvio padrão para baixo e para cima dos índices apurados.

Para a classificação automática dos animais segundo os índices e posterior análise dos resultados, foram criadas regras lógicas na ontologia, que permitem a classificação e a identificação de novas relações ontológicas. Para isso, regras especificadas em SWRL (*Semantic Web Rule Language*) foram criadas, considerando os índices CAR, GPR e ECA (Figura 3).

Figura 3 – Regras SWRL utilizadas para classificação dos animais na arquitetura.

Name	Rule
S1	Cattle(?cattle) ^ isEvaluationOf(?cattle, ?evaluation) ^ Experiment CAR(?evaluation, ?EvaluationCAR) ^ swrlb:lessThan(?EvaluationCAR, -0.13) -> Efficient CAR(?cattle)
S2	Cattle(?c) ^ isEvaluationOf(?c, ?y) ^ Experiment CAR(?y, ?EvaluationCAR) ^ swrlb:greaterThan(?EvaluationCAR, 0.13) -> Inefficient CAR(?c)
S3	Cattle(?c) ^ isEvaluationOf(?c, ?y) ^ Experiment CAR(?y, ?EvaluationCAR) ^ swrlb:lessThanOrEqual(?EvaluationCAR, 0.13) ^ swrlb:greaterThanOrEqual(?EvaluationCAR, -0.13) -> Intermediary CAR(?c)
S4	Cattle(?cattle) ^ isEvaluationOf(?cattle, ?evaluation) ^ Experiment GPR(?evaluation, ?EvaluationGPR) ^ swrlb:greaterThan(?EvaluationGPR, 0.0422) -> Efficient GPR(?cattle)
S5	Cattle(?cattle) ^ isEvaluationOf(?cattle, ?evaluation) ^ Experiment GPR(?evaluation, ?EvaluationGPR) ^ swrlb:lessThan(?EvaluationGPR, -0.0422) -> Inefficient GPR(?cattle)
S6	Cattle(?c) ^ isEvaluationOf(?c, ?y) ^ Experiment GPR(?y, ?EvaluationGPR) ^ swrlb:lessThanOrEqual(?EvaluationGPR, 0.0422) ^ swrlb:greaterThanOrEqual(?EvaluationGPR, -0.0422) -> Intermediary GPR(?c)
S7	Cattle(?cattle) ^ isEvaluationOf(?cattle, ?evaluation) ^ Experiment ECA(?evaluation, ?EvaluationECA) ^ swrlb:lessThan(?EvaluationECA, -0.3685) -> Efficient ECA(?cattle)
S8	Cattle(?c) ^ isEvaluationOf(?c, ?y) ^ Experiment ECA(?y, ?EvaluationECA) ^ swrlb:greaterThan(?EvaluationECA, 0.3685) -> Inefficient ECA(?c)
S9	Cattle(?c) ^ isEvaluationOf(?c, ?y) ^ Experiment ECA(?y, ?EvaluationECA) ^ swrlb:lessThanOrEqual(?EvaluationECA, 0.3685) ^ swrlb:greaterThanOrEqual(?EvaluationECA, -0.3685) -> Intermediary ECA(?c)

Fonte: (Autor, 2017)

As regras SWRL (Figura 3) permitem a composição de associações em busca de um novo conhecimento. Por exemplo, a regra S1 tem o papel de classificar os animais eficientes sobre o índice CAR. Para tal, ela utiliza informações previamente conhecidas, tais como: ser uma instância de Cattle, possuir uma avaliação no índice CAR e ter uma avaliação inferior a -0.13. Assim, um animal que possua essas combinações é classificado como CAR eficiente.

Service Layer

Foi implementado um serviço web RESTful em JAVA, responsável por disponibilizar serviços para o armazenamento, gerência e consulta aos dados. E, através desse, permitir a interoperabilidade com outras aplicações e serviços.

As seguintes tecnologias foram utilizadas: servidor de aplicação Tomcat[2], *reasoner* Pellet[3], Jena[4], Math[5], JSON(*JavaScript Object Notation*)[6], Weka[7] e a JDBC PostgreSQL[8]. As bibliotecas Jena e Pellet facilitam uso de ontologias OWL e máquinas inferência em JAVA. A bilbioteca Weka foi utilizada para análise dos dados, utilizando técnicas de agrupamento e classificação, entre outras.

Com objetivo de ilustrar a interação entre as tecnologias, um cenário de uso foi elaborado para um melhor entendimento dessa. Assim, a Figura 4 ilustra os passos de uma requisição do pesquisador até o retorno do serviço web com a visualização desejada.

2 http://tomcat.apache.org/

3 https://mvnrepository.com/artifact/net.sourceforge.owlapi/pellet-owlapi-ignazio1977

4 https://jena.apache.org/download/index.cgi

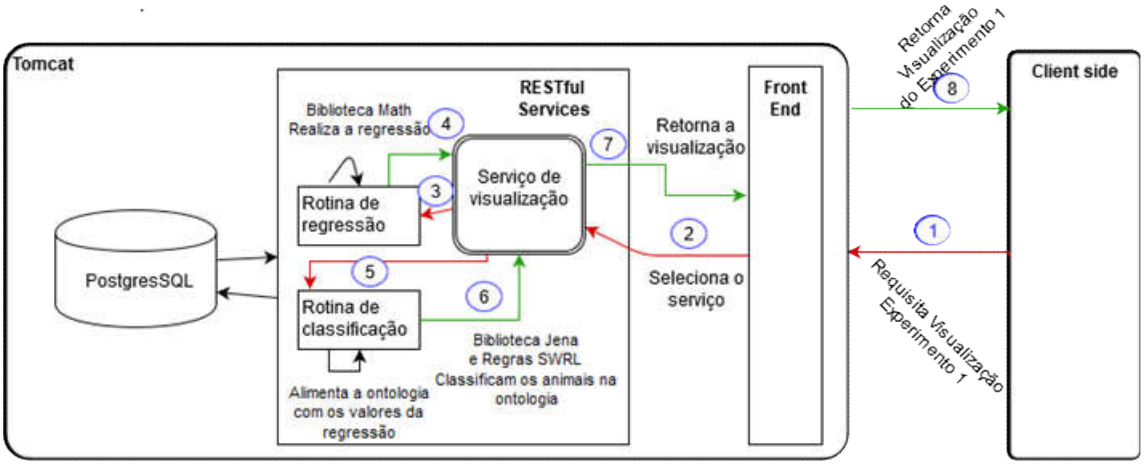
5 http://commons.apache.org/proper/commons-math/download_math.cgi

6 http://www.json.org/json-pt.html

7 http://www.java2s.com/Code/Jar/w/Downloadwekajar.htm

8 https://jdbc.postgresql.org/download.html

Figura 4 – Exemplo de interação entre as tecnologias, a fim de retornar a classificação dos animais no experimento 1.



Fonte: (Autor, 2017)

Front End Layer

Como projeto piloto foi especificada uma aplicação web para acesso aos serviços disponibilizados pelo *Service Layer*. Através da aplicação web, o pesquisador tem acesso as análises dos experimentos, além de obter informações dos animais e as classificações nos experimentos. Assim, a partir dos dados processados pela ontologia, são gerados gráficos que facilitam o entendimento dos dados e análises necessárias.

Na próxima seção, serão apresentados alguns resultados gerados a partir do uso da arquitetura, considerando um conjunto de dados reais, utilizados em experimentos na Embrapa Gado de Leite.

RESULTADOS E DISCUSSÃO

Com objetivo de ilustrar os ganhos obtidos com a adoção da arquitetura, foi elaborado um diagrama de sequência com os passos anteriores, ou seja, sem o uso da arquitetura (Figura 5) e os atuais, ou seja, com o uso da arquitetura (Figura 6), para a classificação dos animais em um experimento. A constante necessidade da intervenção do pesquisador e a inexistência de automação para a transcrição e a análise dos dados produzidos estão presentes no modelo anterior. Nesse sentido, um comparativo entre os experimentos, os animais e os índices era uma tarefa complexa, pois os comparativos ocorriam em planilhas e as análises ficavam a

cargo do pesquisador. Assim, quanto maior o volume dos experimentos, animais e índices, mais complexas eram as análises.

Através da utilização da arquitetura, o pesquisador deixa de lado a utilização de inúmeras ferramentas de terceiros e passa a utilizar apenas a arquitetura. Essa terá o papel de centralizar todas as informações dos experimentos e permitir o acesso distribuído e remoto aos dados. Para o acesso às classificações, o pesquisador escolhe apenas o experimento ou os experimentos desejados e repassa à arquitetura, que irá encarregar-se de classificar os animais e realizar as análises dos dados.

Figura 5. Diagrama de sequência do para classificação dos animais, anterior a arquitetura.

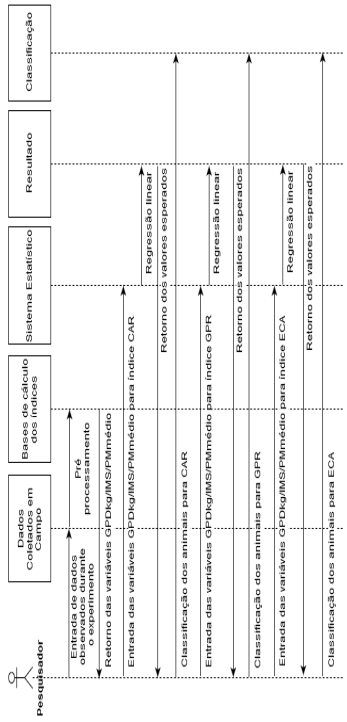
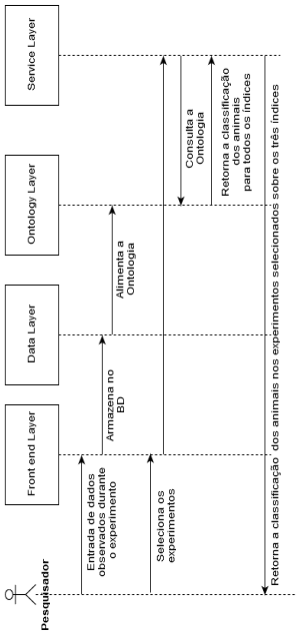


Figura 6. Diagrama de sequência do para classificação dos animais utilizando a arquitetura.



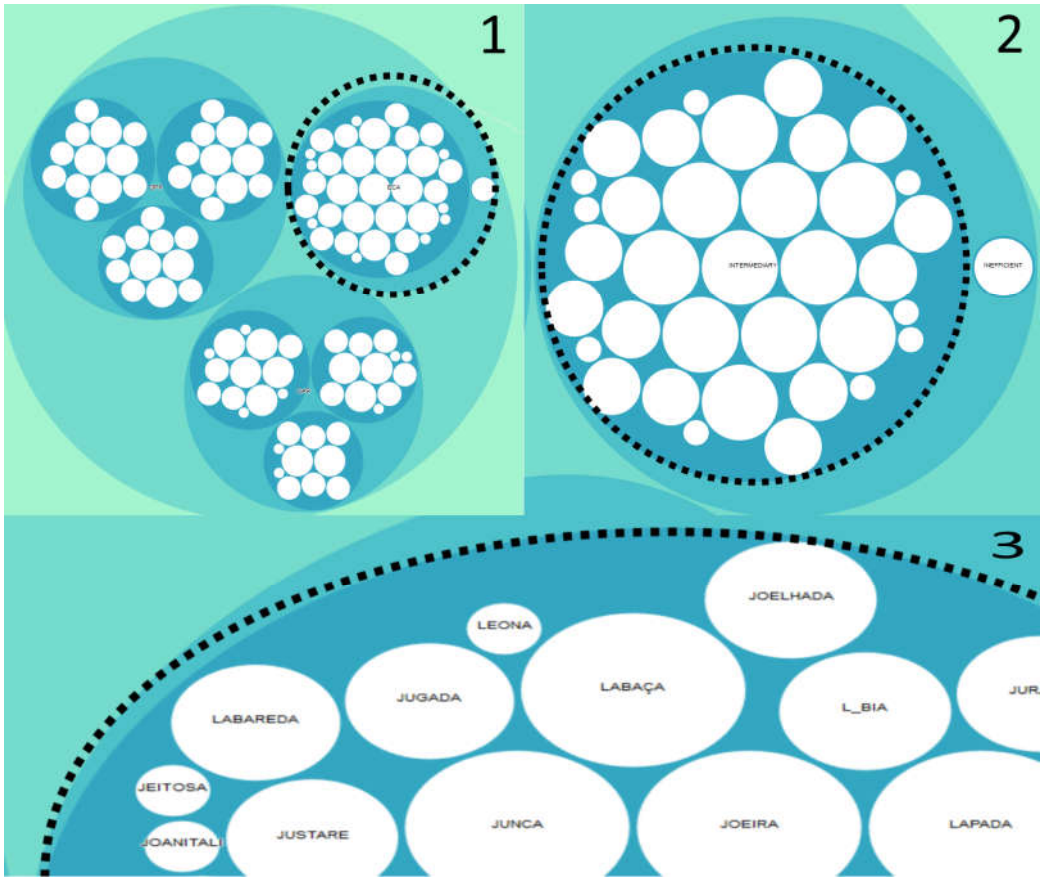
Como prova de conceito da viabilidade da arquitetura, utilizou-se dados de experimentos relacionados com eficiência alimentar. Mais especificamente, dados do experimento fase 2 - Recria em confinamento, com animais entre 3 e 9 meses, conforme especificado na Tabela 1.

O experimento foi conduzido com 36 novilhas F1 Girolando. Para isso, os serviços do Service Layer foram utilizados a partir do Front End Layer. Os dados do experimento foram carregados, segundo o modelo de dados, e instanciados na ontologia. A partir do uso do reasoner e regras lógicas definidas, estes dados relativos a animais do experimento foram

classificados como: eficientes, intermediárias e ineficientes, considerando os índices de eficiência alimentar.

A partir dos dados classificados é possível gerar visualizações para facilitar as análises necessárias por parte do pesquisador. Foram propostas duas formas de visualizações. A primeira aborda o agrupamento dos dados sobre os índices de eficiência alimentar, permitindo uma visão comparativa da classificação, interações e aprofundamento no conjunto ou subconjunto de interesse (Figura 7, Quadro 1). Já a segunda, de forma mais descritiva, apresenta o impacto do ponto de vista desejado sobre o todo (Figura 4, Quadro 2). Através dessa visualização é possível observar o comportamento dos conjuntos, obter as médias e descobrir padrões.

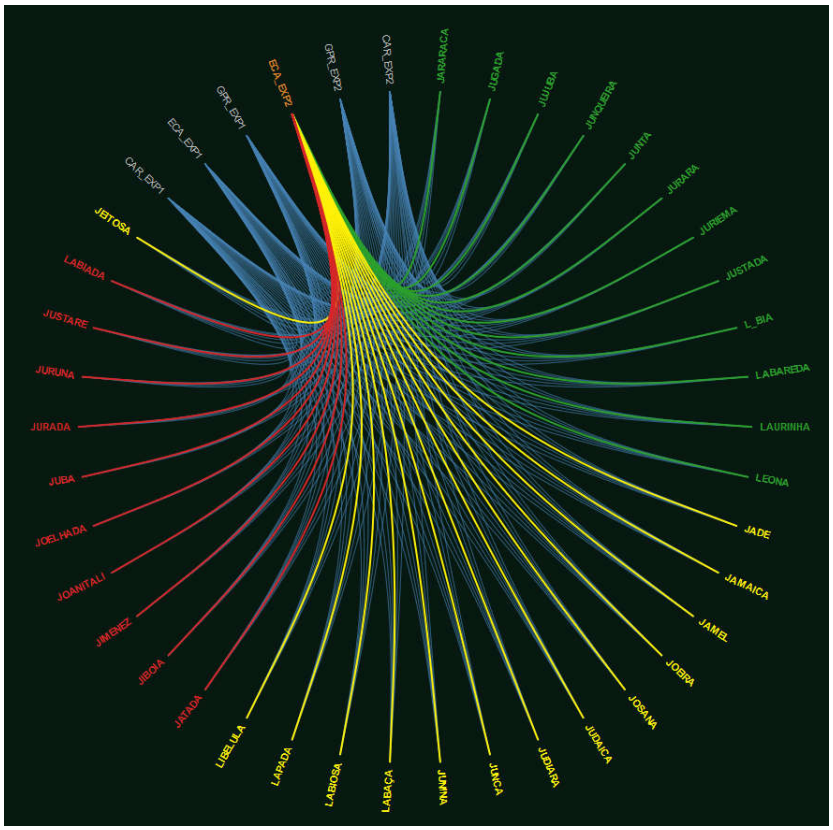
Figura 7 – Visualização de agrupamento, produzidas através de informações do experimento, agrupa os animais eficientes, intermediários e ineficientes sobre os três índices de eficiência alimentar.



Fonte: (Autor, 2017)

A Figura 5 traz informações sobre impactos e relacionamentos entre os indivíduos. Sob o ponto de vista do "ECA_Exp2", que é o índice de eficiência alimentar ECA no experimento 2, a visualização destaca os animais classificados como eficientes através da cor verde, intermediários na cor amarela e ineficientes na cor vermelha.

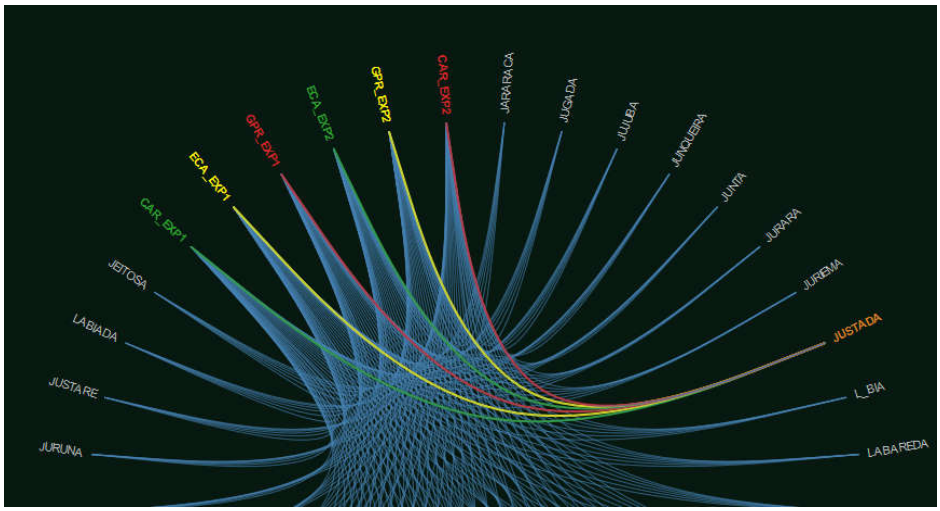
Figura 8 – Visualização do experimento 2, classificando os animais sob o índice ECA eficientes, intermediários e ineficientes.



Fonte: (Autor, 2017)

Por outro lado, sob o ponto de vista do animal, é possível observar que a novilha JUSTADA, por exemplo, é classificada como intermediária no índice ECA, ineficiente no índice GPR e eficiente no índice CAR.

Figura 9 – Visualização do animal Justada e sua classificação em dois experimentos, utilizando os índices de eficiência alimentar ECA, GPR e CAR.



Fonte: (Autor, 2017)

As vantagens obtidas com essa visualização são: análise de um grade volume de dados, interações intuitivas com a visualização que permitem a produção de diferentes resultados, a personalização dos experimentos relevantes a análise, entre outros. Os resultados e análises descritas foram obtidos através de um estudo observacional com pesquisadores da Embrapa Gado de Leite.

CONCLUSÕES

A grande competitividade do mercado atual exige que os produtores leiteiros busquem meios para reduzir seus custos e tornar seus rebanhos competitivos. Nesse sentido, o apoio computacional vem fornecendo alternativas a identificação de animais mais eficientes e, por consequência, proporcionando ganhos econômicos e ambientais.

Observa-se, pelos autores citados, a utilização da ontologia para fins exclusivos de mapeamento. Vale frisar, até o presente momento, nas buscas realizadas em bases indexadas da área da computação, não houve retornos similares a abordagem descrita. Dessa forma, o presente trabalho buscou apresentar uma arquitetura capaz de apoiar as pesquisas em eficiência alimentar, a partir do uso de ontologias e análise de dados, utilizando bases de experimentos coletados em campo.

Este trabalho apresentou uma arquitetura capaz de apoiar as pesquisas em eficiência alimentar, a partir do uso de ontologias, visualizações e análise de dados, utilizando bases de experimentos coletados em campo. A arquitetura visa promover maior poder de análise, acessibilidade aos experimentos, aumentando a confiabilidade e o reuso desses dados por outros pesquisadores. A partir de uma prova de conceito, podem-se observar indícios de que o uso de ontologias pode contribuir para o avanço das pesquisas em eficiência alimentar na Embrapa - Gado de Leite. No entanto, experimentos formais devem ser conduzidos, de forma a comprovar estes indícios. Como trabalho futuro, pretende-se desenvolver uma rede de ontologias, relacionando todos os experimentos descritos na Tabela 1, com o objetivo de observar a manutenção das classificações de eficiência nas fases de avaliação. Outro ponto a ser pesquisado é a busca por padrões nos dados brutos, através de algoritmos de classificação e agrupamentos.

REFERÊNCIAS

MIAH, Shah J.; GAMMACK, John; KERR, Don. Ontology development for context-sensitive decision support. In: **Semantics, Knowledge and Grid, Third International Conference on**. IEEE, 2007. p. 475-478.

BORST, Willem Nico. **Construction of engineering ontologies for knowledge sharing and reuse**. Universiteit Twente, 1997.

LAMBRIX, Patrick; EDBERG, Anna. Evaluation of ontology merging tools in bioinformatics. In: **Pacific Symposium on Biocomputing**. 2003. p. 589-600.

GUBER, T. A Translational Approach to Portable Ontologies. **Knowledge Acquisition**, v. 5, n. 2, p. 199-229, 1993.

MCBRATNEY, Alex et al. Future directions of precision agriculture. **Precision agriculture**, v. 6, n. 1, p. 7-23, 2005.

BEWLEY, J. Exciting dairy breakthroughs: science fiction or precision dairy farming. In: **Proceedings of the precision dairy conference and expo, Rochester, Minneapolis, USA**. 2013. p. 26-27.

BAKER, Patricia G.. et al. An ontology for bioinformatics applications. **Bioinformatics**, v. 15, n. 6, p. 510-520, 1999.

BERRY, D. P.; CROWLEY, J. J. Cell biology symposium: genetics of feed efficiency in dairy and beef cattle. **Journal of Animal Science**, v. 91, n. 4, p. 1594-1613, 2013.

CROWLEY, J. J. et al. Phenotypic and genetic parameters for different measures of feed efficiency in different breeds of Irish performance-tested beef bulls. **Journal of Animal Science**, v. 88, n. 3, p. 885-894, 2010.

GOMES, R. C. et al. Ingestão de alimentos e eficiência alimentar de bovinos e ovinos de corte. **Funpec-Editora, Ribeirão Preto**, v. 77, 2012.

GRION, A. L. et al. Selection for feed efficiency traits and correlated genetic responses in feed intake and weight gain of Nellore cattle. **Journal of animal science**, v. 92, n. 3, p. 955-965, 2014.

MONTANHOLI, Yuri R. et al. Application of infrared thermography as an indicator of heat and methane production and its use in the study of skin temperature in response to physiological events in dairy cattle (Bos taurus). **Journal of Thermal Biology**, v. 33, n. 8, p. 468-475, 2008.

BONILHA, Sarah Figueiredo Martins et al. Feed efficiency, blood parameters, and ingestive behavior of young Nellore males and females. **Tropical animal health and production**, v. 47, n. 7, p. 1381-1389, 2015.

VERHOOSEL, Jack PC; VAN BEKKUM, Michael; VAN EVERT, Frits. Ontology matching for big data applications in the smart dairy farming domain. In: **OM**. 2015. p. 55-59.

GRUBER, Thomas R. et al. A translation approach to portable ontology specifications. **Knowledge acquisition**, v. 5, n. 2, p. 199-220, 1993.

GUARINO, Nicola et al. Formal ontology and information systems. In: **Proceedings of FOIS**. 1998. p. 81-97.

GUARINO, Nicola; OBERLE, Daniel; STAAB, Steffen. What is an Ontology?. In: **Handbook on ontologies**. Springer Berlin Heidelberg, 2009. p. 1-17.

BECHHOFER, Sean. OWL: Web ontology language. In: **Encyclopedia of Database Systems**. Springer US, 2009. p. 2008-2009.

HITZLER, Pascal; KROTZSCH, Markus; RUDOLPH, Sebastian. **Foundations of semantic web technologies**. CRC Press, 2009.

CAMPOS, Mariana .M., LEAO, Juliana .M., LIMA, Juliana .A.M., MACHADO, Fernanda .S. **Tecnologias de precisão na avaliação de eficiência alimentar**. Cadernos Técnicos de Veterinária e Zootecnia, 2015. n79 p. 73-85.

Integração de Ferramentas de SIG e Mineração de Dados para Utilização em Atividades de Gestão Espacialmente Diferenciada Aplicada na Agricultura de Precisão

Eduardo Antonio Speranza[1], *Ricardo Rodrigues Ciferri*[2]

[1]Embrapa Informática Agropecuária
Campinas, São Paulo, Brasil
eduardo.speranza@embrapa.br

[2]Departamento de Computação, Universidade Federal de São Carlos
São Carlos, São Paulo, Brasil
ricardo@dc.ufscar.br

RESUMO

A agricultura de precisão é uma abordagem agrícola que se utiliza de tecnologias da informação e comunicação para possibilitar uma gestão diferenciada da lavoura voltada para o aumento da produtividade de maneira sustentável, reduzindo os impactos ao meio ambiente a partir da aplicação espacialmente diferenciada de insumos agrícolas e consequentemente proporcionando o aumento do retorno econômico. Um dos conceitos utilizados por essa abordagem é o delineamento de unidades de gestão diferenciada, permitindo tratamentos localizados de acordo com as características do solo e das plantas. Devido ao crescimento da disponibilidade de ferramentas computacionais que podem auxiliar usuários finais, a criação de modelos e arquiteturas acessíveis e capazes de agregar essas ferramentas, de forma a possibilitar o seu uso de maneira integrada, tem se tornado intenso objeto de estudo. Este artigo descreve a proposta de uma arquitetura voltada para apoiar o processo de delineamento de unidades de gestão diferenciada em agricultura de precisão utilizando *software* livre e de distribuição gratuita. A arquitetura proposta é comparada com outras abordagens disponíveis na literatura, onde são identificadas as suas vantagens e desvantagens e propostas novas alternativas para trabalhos futuros. O trabalho realizado nesse artigo permitiu verificar que a solução proposta é viável para essa aplicação de AP, devido à disponibilidade das ferramentas e a utilização dos padrões OGC, amplamente disseminados nas áreas de aplicação que utilizam dados geoespaciais.

PALAVRAS-CHAVE: Agricultura de Precisão, Unidades de Gestão Diferenciada, Zonas de Manejo, Sistemas de Informações Geográficas, Mineração de Dados, Dados Espaciais.

ABSTRACT

Precision agriculture is an agricultural approach which uses information and communication technologies to enable differentiated crop management to increase yield in a sustainable way, reducing impacts to the environment from spatially differentiated application of agricultural inputs and consequently providing the increase of profits. One of the concepts used by this approach is the delineation of differentiated management units, allowing site-specific treatments according to soil and plant features. Due to the increasing availability of computational tools which can help end users to apply this concept, the proposal of models and achitetures accessible and able to aggregate these tools, so they can be used in an integrated way. This paper describes the proposal of an architeture aimed to support the delineation of differentiated management units in precision agriculture, using free available software. The proposed architeture is compared with other approaches available in the literature, where its advantages and drawbacks are identified and new alternatives are proposed for future work. The work performed in this article allowed to verify that the proposed solution is feasible for this AP application, due to the availability of tools and the use of OGC standards, widely disseminated in applied areas using geospatial data.

KEYWORDS: Precision Agriculture, Differentited Management Units, Management Zones, Geographic Information Systems, Data Mining, Spatial Data.

INTRODUÇÃO

A Agricultura de Precisão (AP) é definida como uma abordagem de gerenciamento agrícola baseada na variabilidade espacial e temporal da lavoura, com o intuito de aumentar a produtividade de maneira sustentável, reduzindo os impactos ao meio ambiente com a aplicação espacialmente diferenciada de insumos agrícolas e, consequentemente, aumentando o retorno econômico (BERNARDI et al., 2014; MOLIN; AMARAL; COLAÇO, 2015). Para atingir esse objetivos, essa abordagem comumente utiliza tecnologias de informação e comunicação (TICs) associadas a processos que visam a otimização das culturas (MOLIN, 2003; MASSRUHÁ et al., 2014). Assim, os processos de preparação do solo, plantio, acompanhamento da lavoura e colheita passam a ser auxiliados por atividades de coleta e interpretação de dados, visando a realização de intervenções sítio específicas (COMPARETTI, 2011). Um dos principais fatores contribuintes para impulsionar a adoção da AP no mundo foi a operacionalização do sinal de GPS[1], em meados da década de 90. Desse modo, os Sistemas de Informações Geográficas (SIGs) (BURROUGH, 1986; CIFERRI, 1995), em conjunto com bancos de dados espaciais (CASANOVA et al., 2005; CIFERRI, 1995), se tornaram ferramentas computacionais essenciais para dar suporte às atividades de armazenamento, manipulação, análise e recuperação de dados georreferenciados que podem ser executadas em um ciclo produtivo utilizando AP.

Nos últimos anos, com o crescimento das TICs e, consequentemente, das formas de se

[1]GPS: do inglês *Global Positioning System* - Sistema de Posicionamento Global.

obter dados georreferenciados em campo, seja por meio de sensores específicos, imagens aéreas e de satélite com altíssima resolução espacial, ou a partir de ferramentas de inspeção instaladas em *tablets* ou *smartphones*, novos conceitos relacionados ao armazenamento, manipulação e recuperação de dados, como *big data* e computação em nuvem, passaram a fazer parte do contexto das pesquisas em AP (LI; CHUNG, 2015). Desse modo, as atividades de análise, que estão fortemente relacionadas ao suporte à decisão para o usuário final, devem ser realizadas com o auxílio de processos de descoberta de conhecimento em bancos de dados (KDD[2]) (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996; WEISS; INDURKHYA, 1998). Uma das atividades de AP que pode se beneficiar desse processo é o delineamento de unidades de gestão diferenciada (UGDs), também conhecidas como zonas de manejo. As UGDs são regiões geográficas delimitadas na lavoura com variabilidade interna desprezível, a ponto de poderem ser consideradas intervenções homogêneas dentro dos seus limites (RUSS, 2012; MOLIN; AMARAL; COLAÇO, 2015). As UGDs obtidas devem ser consistentes ao longo do tempo e caracterizarem o potencial de resposta de uma área de produção agrícola, podendo ser obtidas a partir de tarefas de mineração de dados. Desse modo, as UGDs podem ser utilizadas para diversos fins gerenciais, que vão desde a determinação da densidade apropriada para coleta de amostras em solo, até o direcionamento de investimentos em insumos agrícolas.

Este artigo descreve a proposta de uma arquitetura computacional para suporte à tarefa de delineamento de UGDs em AP, utilizando a integração de SIGs e bancos de dados espaciais com algoritmos capazes de realizar as etapas de KDD necessárias para a execução dessa tarefa, tais como interpolação espacial, agrupamento de dados espaciais e validação. O restante do artigo é organizado da seguinte forma: a seção Material e Métodos descreve as TICs e conceitos utilizados para a construção da arquitetura; a seção Resultados e Discussão descreve a arquitetura em si, bem como resultados que podem ser obtidos com a sua utilização, discussões e comparações com outras soluções disponíveis na literatura; e a seção Conclusões resume os resultados obtidos e indica trabalhos futuros que podem ser realizados com o decorrer desta pesquisa.

MATERIAL E MÉTODOS

Nesta seção são apresentados os conceitos sobre TICs que dão suporte ao desenvolvimento da arquitetura proposta neste artigo, partindo desde o armazenamento dos dados até a interface com o usuário final. Também são descritos os dados reais utilizados em experimentos realizados com uma implementação da arquitetura proposta.

Sistemas Gerenciadores de Bancos de Dados Espaciais (SGBDEs)

Os Sistemas Gerenciadores de Bancos de Dados (SGBDs) são programas de computador que auxiliam nos processos de definição, construção, carga, manipulação e compartilhamento de

[2]KDD: do inglês *Knowledge Discovery in Databases*.

bancos de dados entre diversos usuários e aplicações (ELMASRI; NAVATHE, 2011). Dentre os SGBDs mais conhecidos e utilizados atualmente, destacam-se o PostgreSQL (POSTGRESQL, 2017) e o Oracle (ORACLE, 2017), que, dentre outras características, possuem mecanismos de extensibilidade que possibilitam o desenvolvimento de extensões geográficas capazes de tratar dados espaciais, que também são conhecidos como dados georreferenciados, dados geoespaciais ou dados geográficos (CASANOVA et al., 2005; CIFERRI, 1995), tornando-os SGBDs Espaciais (SGBDEs). Para as atividades de AP, pode ser utilizado qualquer SGBDE que possua a característica de armazenar dados espaciais, utilizando objetos bidimensionais para representar a disposição espacial; e dados convencionais associados a esse objetos representando fenômenos naturais que ocorrem na lavoura.

Padrões Open Geospatial Consortium

A OGC (*Open Geospatial Consortium*) é uma organização internacional sem fins lucrativos criada para o desenvolvimento de padrões de consenso para dados geoespaciais. Os padrões OGC são baseados na gramática de modelagem GML (*Geography Markup Language*), que atua no transporte, armazenamento e intercâmbio de dados geoespaciais na *Web* (PERCIVALL et al., 2003). Dentre esses padrões, destacam-se três importantes serviços: WMS (*Web Map Service*), utilizado para a recuperação e visualização de mapas como imagens; WFS (*Web Feature Service*), utilizado para acesso e edição de dados geoespaciais; e o WPS (*Web Processing Service*), utilizado para a execução de algoritmos de geoprocessamento. Diversos módulos da arquitetura proposta neste artigo são integrados utilizando os padrões OGC, que vêm sendo amplamente disseminados na comunidade produtora de dados geoespaciais.

Servidores de Mapas

A disponibilização dos SIGs na *Web* impulsionou o desenvolvimento de soluções para disseminar informações espaciais em forma de mapa, permitindo a criação de aplicações interativas com diversas funcionalidades disponíveis para o usuário final (MITCHELL, 2005). Desse modo, surgiram os servidores de mapas, como o GeoServer (GEOSERVER, 2017), capazes de acessar, manipular e entregar dados geoespaciais seguindo os padrões OGC. *Software* livre e de distribuição gratuita, o GeoServer possui suporte a diversos serviços OGC, e compatibilidade para manipular dados especiais armazenados por diversos SGBDEs, podendo ser facilmente integrado aos SIGs.

Sistemas de Informações Geográficas

Os sistemas de informações geográficas (SIGs) são utilizados para manipulação de dados que representam objetos e fenômenos onde a localização geográfica é uma característica intrínseca indispensável à análise da informação (ARONOFF, 1989). Os SIGs são compostos por vários subsistemas integrados, e devem prover suporte a diferentes tipos de dados e aplicações, com

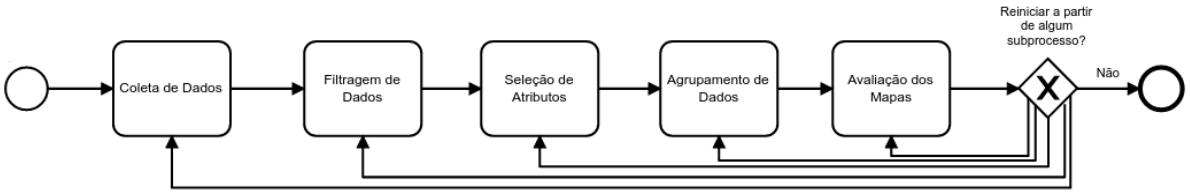
o intuito de centralizar os dados coletados de fontes heterogêneas de maneira transparente ao usuário final, por meio de funcionalidades de entrada, integração, processamento, visualização, plotagem, armazenamento e recuperação de dados (CIFERRI, 1995; CÂMARA et al., 1996). Atualmente, os SIGs trabalham de forma integrada com os SGBDEs para o armazenamento e recuperação de dados espaciais.

Dentre as ferramentas de SIG disponíveis no mercado, deve-se destacar a plataforma proprietária ArcGIS (ARCGIS, 2017), criada em 2001 pela empresa ESRI e amplamente disseminada, não só na comunidade de geoprocessamento, mas também na maioria das áreas de aplicação. Quando consideramos apenas o mercado de *software* livre e de distribuição gratuita, o QuantumGIS (ou QGIS) (QGIS, 2017) é o SIG mais utilizado. Além dos diversos complementos disponíveis para atividades de geoprocessamento, a possibilidade de implementação de algoritmos pelos próprios usuários e integração com outros SIGs e plataformas que possuem diversas bibliotecas relacionadas à geoestatística, geoprocessamento e mineração de dados, como o ambiente R (R, 2017), fazem com que o QGIS seja bastante disseminado entre os usuários e produtores de dados geoespaciais, principalmente quando o custo relacionado à aquisição de *software* é um fator relevante para o usuário.

Mineração de Dados

A transformação dos dados em conhecimento, com o intuito de fornecer suporte ao usuário final em atividades que exigem a tomada de decisão, é um dos principais motivadores para a utilização de conceitos de mineração de dados em AP, onde os valores dos atributos, que estão normalmente associados a coordenadas geográficas, podem ser influenciados por seus vizinhos espaciais (ESTER et al., 2000). A etapa de mineração de dados é considerada, em conjunto com o pré-processamento e o pós-processamento, como essencial em um processo de KDD (REZENDE, 2003). No contexto da AP, Santos, Molin e Saraiva (2013), Santos (2014) propuseram um modelo de referência para o processo de delineamento de UGDs (Figura 1). Esse modelo contempla todas as etapas do processo de KDD no âmbito da aplicação e foi utilizado como guia para a construção da arquitetura proposta neste artigo.

Figura 1: Modelo de referência para o processo de delineamento de UGDs.



Fonte: Extraído de Santos (2014)

Para a execução dos algoritmos que envolvem processos de KDD, o ambiente de desenvolvimento R (R, 2017) tem sido amplamente utilizado. O ambiente R permite o acesso aos seus

algoritmos que envolvem geoprocessamento por meio da interface do QGIS, com a utilização do complemento *Processing*. Esse complemento permite a criação de interfaces para usuários finais, e pode retornar como resultado tanto dados geoespaciais vetoriais como gráficos gerados pelos próprios algoritmos do ambiente R. Além disso, permite a criação de modelos para execução sequencial das interfaces de acesso criadas que definem *workflows* para um conjunto de aplicações.

Publicação de Dados Espaciais

Uma vez que o processo de delineamento de UGDs é finalizado, os mapas escolhidos pelo usuário final para serem utilizados em determinadas atividades devem ser armazenados para futuras consultas e até mesmo para disseminação via rede. Além do armazenamento direto dos dados espaciais em sua forma bruta no SGBDE, os mapas de UGDs também podem ser publicados no GeoServer diretamente pela interface do QGIS. Uma vez publicados, esses mapas podem ser acessados via *Web* por meio de sítios com funcionalidades de SIG compatíveis com os padrões OGC, conhecidos como *WebGIS* (MITCHELL, 2005). Um exemplo de sítios desse tipo é o GeoNode (GEONODE, 2017), uma plataforma de código aberto para compartilhamento de dados geoespaciais que permite ao usuário final explorar mapas disponibilizados por servidores como o GeoServer. Os mapas podem ser publicados diretamente no GeoNode por meio da interface gráfica do QGIS, utilizando o complemento OpenGeo.

Dados Geoespaciais

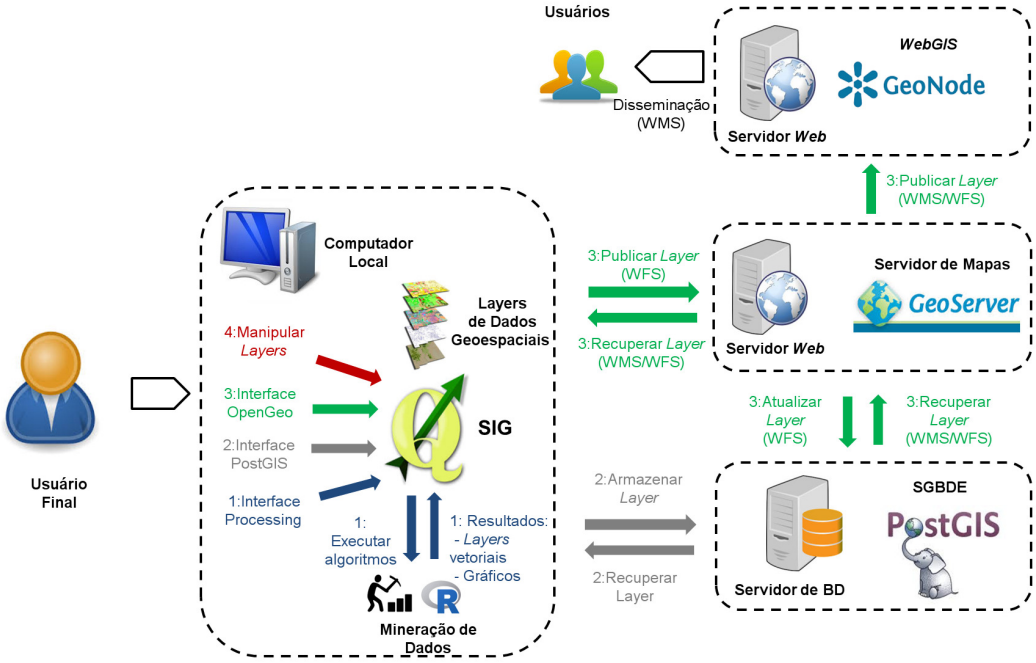
O conjunto de dados geoespaciais reais usado nos experimentos deste artigo foi obtido em um talhão de cultura de cana-de-açúcar de aproximadamente 17 hectares da Fazenda Aparecida, localizada em Mogi-Mirim, SP, com 20 atributos relacionados ao solo e a cultura, tais como pedologia, textura e condutividade elétrica do solo, índices de biomassa e produtividade. Para o delineamento de UGDs, a arquitetura proposta neste artigo permite ao usuário final: escolher os atributos que serão utilizados; submeter os dados a processos de limpeza para exclusão de valores extremos; realizar processos de interpolação espacial para acomodar os atributos em uma grade espacial única; delinear as UGDs utilizando algoritmos de agrupamento espacial; analisar os resultados obtidos por meio de critérios de validação; e publicar mapas finais de UGDs.

Na próxima seção, a arquitetura proposta é apresentada e comparada com outras soluções disponíveis na literatura, com o intuito de analisar as suas vantagens e desvantagens e também indicar melhorias que podem ser propostas para uma adequação maior ao processo à qual está sendo submetida.

RESULTADOS E DISCUSSÃO

Considerando as TICs descritas na seção Material e Métodos, foi proposto um modelo de arquitetura para permitir a realização da tarefa de delineamento de UGDs em AP, com a utilização de *software* livre e de distribuição gratuita (Figura 2).

Figura 2: Modelo de arquitetura para a tarefa de delineamento de UGDs em AP utilizando *software* livre e padrões OGC.

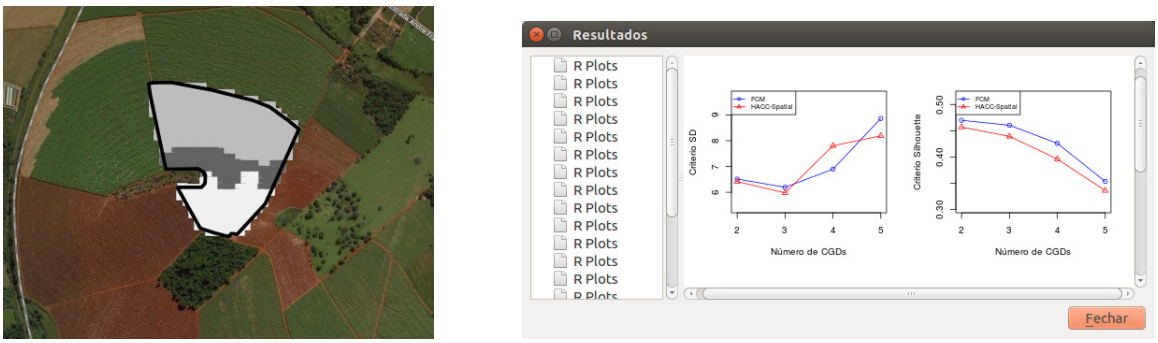


Nessa arquitetura, são observados quatro macroprocessos distintos que podem ser executados pelo usuário final. O primeiro, identificado pelo número 1, permite a integração da interface principal do QGIS com os algoritmos de mineração de dados disponibilizados pelo ambiente R, por meio da criação de *scripts* anotados utilizando o complemento *Processing*. Esses *scripts* geram janelas gráficas para que o usuário final informe os parâmetros e dados espaciais de entrada que serão utilizados pelos algoritmos de mineração de dados. Para os dados de saída, são permitidos tanto camadas vetoriais, que podem ser posteriormente manipuladas utilizando as funcionalidades específicas do QGIS, quanto figuras representando gráficos ou mapas gerados pelos próprios algoritmos de mineração de dados do ambiente R, visualizadas por meio do complemento *Processing* (Figura 3).

Além das diversas alternativas para entrada e saída de dados que a arquitetura proporciona, também é possível executar diversos algoritmos de mineração de dados de maneira sequencial, por meio da ferramenta de criação de modelos do complemento *Processing*. A Figura 4 mostra um exemplo de criação de um modelo para delineamento de UGDs, onde são executados sequencialmente algoritmos para limpeza de dados, criação de grade espacial única, interpolação espacial e agrupamento, até que o mapa vetorial de UGDs seja obtido.

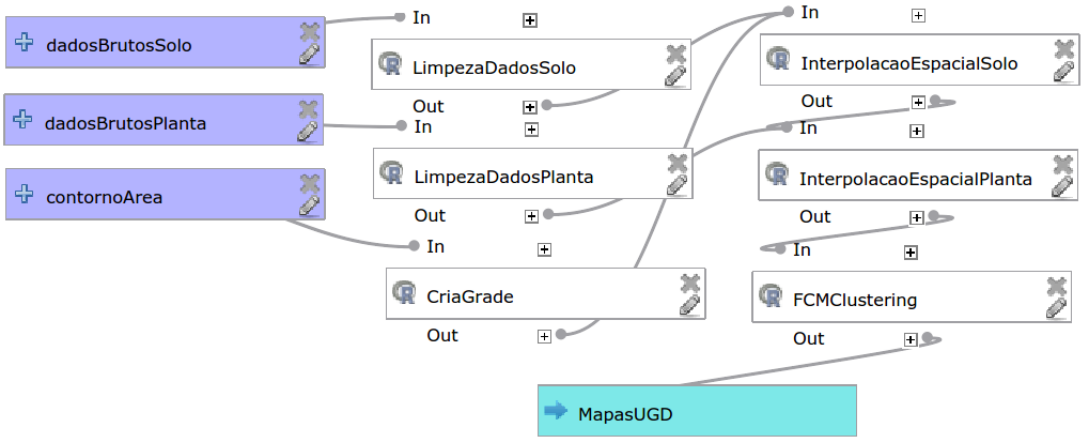
O segundo macroprocesso definido na arquitetura da Figura 2, identificado pelo número 2,

Figura 3: Exemplos de resultados retornados por algoritmos de mineração de dados executados pelo ambiente R e integrados a interface do QGIS: (a) mapa de UGDs no formato vetorial com imagem de sensoriamento remoto ao fundo; (b) gráfico com análises considerando critérios de validação.



(a) (b)

Figura 4: Exemplo de modelo para delineamento de UGDs com execução sequencial de diversos algoritmos implementados no ambiente R.



permite ao usuário final, a partir de funcionalidades específicas do QGIS, armazenar e recuperar os mapas de UGDs obtidos a partir do macroprocesso 1 diretamente no SGBDE PostGIS. Com os dados armazenados, o usuário final pode realizar o terceiro macroprocesso, identificado na arquitetura da Figura 2 pelo número 3, para disseminar pela rede os mapas de UGDs obtidos. As funcionalidades descritas para esse macroprocesso devem ser acessadas a partir da interface do complemento OpenGeo, disponível no QGIS. Por meio dessa interface, é possível publicar e recuperar mapas de UGDs já armazenados no SGBDE a partir do GeoServer, utilizando para isso os serviços OGC WFS e WMS. Utilizando a mesma interface do complemento OpenGeo, os mapas de UGDs disponibilizados no servidor de mapas GeoServer podem ser publicados no GeoNode, permitindo a sua disseminação pela rede por meio de uma interface *WebGIS* disponível para qualquer usuário que desejar utilizá-los.

Finalmente, um quarto macroprocesso, relacionado à manipulação de *layers*[3] localmente

[3] *Layer:* Termo comumente utilizado pela comunidade de SIG para se referir a conjuntos de dados geoespaciais temáticos que podem ser visualizados ou manipulados.

pelo usuário final utilizando o SIG QGIS também foi definido na Figura 2, identificado pelo número 4. A manipulação local de *layers*, sejam elas representando dados espaciais brutos disponibilizados pelo GeoServer, ou mapas de UGDs delineadas a partir do macroprocesso 1, permite ao usuário final editá-las de maneira *off-line* utilizando as funcionalidades de geoprocessamento disponibilizadas pelo QGIS. Isso permite que essas *layers* sejam publicadas em um servidor de mapas apenas quando estiverem em sua versão final.

As ferramentas e complementos disponíveis no QGIS para a sua integração com o ambiente R e posterior publicação de mapas em *WebGIS* podem ajudar muito no processo de delineamento de UGDs em AP. Entretanto, algumas dificuldades foram encontradas durante a implementação e devem ser registradas para que melhorias possam ser propostas. Com relação à anotação de *scripts* para geração de interfaces, a limitação de tipos de campos disponíveis restringe a utilização de soluções alternativas em alguns casos, por exemplo quando existe a necessidade de informar mais de um atributo da camada de dados vetorial para ser utilizado pelo algoritmo a ser executado. Com relação aos gráficos gerados pelo ambiente R e exibidos na interface do QGIS, a janela para exibição de resultados (Figura 3 (b)) utiliza imagens de tamanho fixo que foram obtidas como retorno do algoritmo executado, limitando a sua visualização e manipulação pelo usuário final. Outras dificuldades também são encontradas com relação às mensagens de retorno de execução dos algoritmos, realizada por meio de uma janela textual que dificulta a identificação de possíveis erros que possam vir a ocorrer.

Com relação às ferramentas disponíveis na literatura, algumas questões podem ser discutidas. A primeira ferramenta computacional que surgiu para auxiliar o usuário final no delineamento de UGDs em AP foi o *software Management Zones Analyst* (MZA) (FRIDGEN et al., 2004), o qual é ainda muito utilizado em diversas abordagens da literatura. O MZA possui funcionalidades importantes que fornecem estatísticas descritivas para análise e escolha do melhor mapa de UGDs a ser utilizado. Entretanto, sua disponibilidade para instalação apenas sob o sistema operacional Windows, exigência de dados de entrada já em uma grade espacial única, e indisponibilidade de acesso a bases de dados e serviços *Web*, limitam o seu uso se considerarmos os conceitos atuais de TICs e o processo de KDD que envolvem o delineamento de UGDs em AP. A arquitetura proposta neste artigo visa resolver algumas dessas questões, possibilitando ao usuário final obter dados de diversas fontes, e que podem passar por todas as etapas do KDD até a extração final do conhecimento em mapas de UGDs.

Com o objetivo de prover uma solução acessível pela *Web* aos usuários finais, disponibilizando imagens de sensoriamento remoto e uma interface amigável e simples de ser utilizada, Zhang et al. (2010) desenvolveram o aplicativo ZoneMap. Segundo experimentos realizados pelos autores, o ZoneMap permite o delineamento de mapas de UGDs bastante compatíveis com mapas delineados manualmente pelos usuários finais. Entretanto, uma dificuldade que pode ser observada é a necessidade dos usuários finais informarem os seus dados vetoriais a um sistema externo, mesmo que este esteja instalado em servidores de universidades ou órgãos governamentais. Com relação à essa questão, a arquitetura proposta neste artigo permite que

o usuário final trabalhe os seus dados em seu ambiente local, utilizando um SIG, e publique, por exemplo, apenas mapas de UGDs finais em um servidor que pode ser externo à sua rede de trabalho.

A possibilidade de utilização dos padrões OGC para aplicações em AP já vem sendo estudada a algum tempo. Murakami et al. (2007) propuseram uma estrutura completa para desenvolvimento de sistemas em AP utilizando esses padrões, que posteriormente foi atualizada e incluída em um portal *Web* para serviços agrícolas (RIBEIRO-JÚNIOR, 2007). O trabalho de Nash, Korduan e Bill (2009) também mostrou a eficiência do uso de padrões OGC em atividades de AP que exigem processamento paralelo, possibilitando a execução dessas atividades utilizando diversos servidores distintos. Nesse sentido, a arquitetura proposta neste artigo também teve como objetivo estabelecer um modelo baseado nos padrões da OGC, porém já estabelecendo opções de ferramentas livres e de distribuição gratuita que podem ser utilizadas sem custos pelo usuário final. A Tabela 1 resume as principais características das soluções encontradas na literatura no que diz respeito ao acesso, entrada e saída de dados, em comparação com a arquitetura proposta.

Tabela 1: Resumo das características das soluções da literatura com relação à proposta de arquitetura apresentada.

Ferramenta	**Tipo Aplicação**	**Fontes de Dados**	**Saídas**
Fridgen et al. (2004)	Desktop	Arquivo ASCII	Arquivo ASCII Arquivo Estatísticas Arquivo Indicadores Gráficos
Murakami et al. (2007)	Web	OGC: WMS,WFS,WCS	OGC: WCS, WMS,WFS
Nash, Korduan e Bill (2009)	Web	OGC: WMS,WFS,WCS	OGC: WMS,WFS,WCS
Zhang et al. (2010)	Web	Arquivo ASCII Arquivo Raster BD Imagens	Arquivo ASCII Arquivo Raster Arquivo Shape
Arquitetura Proposta	Desktop / Web	OGC: WMS,WFS	OGC: WMS,WFS Gráficos

A partir da Tabela 1, verifica-se que, diferentemente das soluções disponíveis na literatura, a arquitetura proposta possibilita ao usuário final manipular os dados espaciais por meio de um SIG *Desktop*, contendo todas as funcionalidades de geoprocessamento úteis para a aplicação e que, por muitas vezes, possuem custo computacional elevado de execução. Entretanto, as fontes de dados devem ser remotas, e serem acessadas a partir de requisições a provedores de serviços da OGC. O acesso às fontes de dados é realizado de maneira semelhante pelas outras soluções, exceto para a solução desenvolvida por Fridgen et al. (2004), onde é permitida apenas a utilização de um formato específico de arquivo, a ser informado pelo usuário final. Com relação à saída de resultados, a arquitetura proposta proporciona, além das camadas de dados

geoespaciais nos formatos OGC, gráficos de avaliações qualitativas a respeito dos resultados obtidos. Esse tipo de indicador também é fornecido pela abordagem desenvolvida por Fridgen et al. (2004).

CONCLUSÕES

A utilização de TICs livres e de distribuição gratuita integradas em uma arquitetura única possibilitou a análise de que é possível executar processos importantes como o delineamento de UGDs em AP utilizando os conceitos de descoberta de conhecimento em bancos de dados. Outras soluções disponíveis na literatura mostram que, ao longo do tempo, está se estabelecendo uma tendência de utilização de novas tecnologias para processos agrícolas que envolvem TICs, capazes de aumentar a interoperabilidade entre sistemas e auxiliar a tomada de decisão do usuário final.

Com relação à arquitetura proposta neste artigo, deve-se destacar o uso de ferramentas livres e de padrões OGC, realizado de maneira similar em outras soluções disponíveis na literatura. Por outro lado, a arquitetura proposta possibilita ao usuário final a manipulação dos dados espaciais de maneira local, permitindo a utilização de funcionalidades de geoprocessamento computacionalmente custosas. Essa característica pode ser interessante ao usuário final, à medida que é desejável que apenas mapas finalizados devem ser publicados e visualizados por outros usuários.

A arquitetura proposta ainda deve levar em consideração o estabelecimento de melhorias que podem tornar os seus módulos mais independentes. Uma dessas melhorias está relacionada ao acesso aos algoritmos de geoprocessamento e mineração de dados executados utilizando o ambiente de desenvolvimento R, que na versão atual da arquitetura devem estar instalados localmente em conjunto com o SIG no computador local do usuário final. Para tornar esse módulo mais independente, uma nova proposta da arquitetura, considerando a utilização de uma versão servidor do ambiente R, o RServe (RSERVE, 2017), está sendo elaborada. Nessa nova proposta, os algoritmos serão executados diretamente no servidor, por meio de requisições do serviço WPS diretamente do SIG para o RServe, utilizando a ferramenta WPS4R (HINZ et al., 2013). Assim, algoritmos específicos poderão ser disponibilizados apenas no servidor, evitando a necessidade de instalação e atualização em computadores locais.

AGRADECIMENTOS

Agradecemos à Fazenda Aparecida e as pesquisadoras da Embrapa Célia Regina Grego e Cristina Aparecida Rodrigues, pela disponibilidade dos dados coletados para uso experimental neste artigo. Agradecemos também as agências de fomento em pesquisa CNPq, CAPES, FAPESP e FINEP. O primeiro autor é apoiado pelo programa de pós-graduação da Embrapa, e o segundo autor, pela bolsa de produtividade em pesquisa do CNPq número 311868/2015-0.

REFERÊNCIAS

ARCGIS: Esri arcgis. 2017. Disponível em: <http://www.arcgis.com/>. Acesso em: 02 mai. 2017.

ARONOFF, S. Geographic information systems: A management perspective. *Geocarto International*, v. 4, n. 4, p. 58, 1989. Disponível em: <http://dx.doi.org/10.1080/10106048909354237>.

BERNARDI, A. C. d. C. et al. *Agricultura de Precisão - Resultados de um Novo Olhar*. 1. ed. Brasília: Empresa Brasileira de Pesquisa Agropecuaria, 2014. 596 p.

BURROUGH, P. A. Principles of GIS for land resources assessment. *Monographs on soil and resources survey. Clarendon, Oxford*, 1986.

CÂMARA, G. et al. *Anatomia dos Sistemas de Informações Geográficas*. [S.l.]: Campinas-SP: Instituto de Computação, UNICAMP, 1996.

CASANOVA, M. A. et al. *Bancos de Dados Geográficos*. Curitiba: MundoGeo, 2005. 506 p.

CIFERRI, R. R. *Um benchmark voltado a análise de desempenho de sistemas de informações geográficas*. Dissertação (Mestrado em Ciência da Computação) — Universidade Estadual de Campinas, 1995.

COMPARETTI, A. Precision Agriculture: Past, Present and Future. In: *Agricultural Engineering and Environment*. Akademija, Lithuania: [s.n.], 2011. p. 17.

ELMASRI, R.; NAVATHE, S. B. *Fundamentals of Database Systems*. 6th. ed. Boston: Addison-Wesley, 2011. 1172 p. ISSN 14337851. ISBN 9780136086208.

ESTER, M. et al. Spatial Data Mining: Database Primitives, Algorithms and Efficient DBMS Support. *Data Min. Knowl. Discov.*, Kluwer Academic Publishers, Hingham, MA, USA, v. 4, n. 2-3, p. 193–216, 2000. ISSN 1384-5810. Disponível em: <http://dx.doi.org/10.1023/A:1009843930701>.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Commun. ACM*, ACM, New York, NY, USA, v. 39, n. 11, p. 27–34, 1996. ISSN 0001-0782. Disponível em: <http://dx.doi.org/10.1145/240455.240464>.

FRIDGEN, J. J. et al. Management Zone Analyst (MZA): Software for Subfield Management Zone Delineation. *Agronomy Journal*, v. 96, p. 100–108, 2004.

GEONODE: Open source geospatial content management system. 2017. Disponível em: <http://geonode.org/>. Acesso em: 04 mai. 2017.

GEOSERVER. *Open Source Server for Sharing Geospatial Data*. 2017. Disponível em: <http://geoserver.org>. Acesso em: 08 mai. 2017.

HINZ, M. et al. Spatial Statistics on the Geospatial Web. In: *The 16th AGILE International Conference on Geographic Information Science, Short Papers*. [S.l.: s.n.], 2013.

LI, M.; CHUNG, S.-O. Special issue on precision agriculture. *Computers and Electronics in Agriculture*, Elsevier B.V., v. 112, p. 1, 2015. ISSN 01681699. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0168169915000897>.

MASSRUHÁ, S. M. F. S. et al. *Tecnologias da informação e comunicação e suas relações com a agricultura.* 1. ed. Brasília, DF: Embrapa Informática Agropecuária, 2014. 411 p. ISBN 978-85-7035-414-3.

MITCHELL, T. *Web mapping illustrated: using open source GIS toolkits*. [S.l.]: O'Reilly Media, Inc., 2005.

MOLIN, J. P. Agricultura de Precisão: Situação atual e perspectivas. In: FANCELLI, A. L.; NETO, D. D. (Ed.). *Milho: Estratégias de Manejo para Alta Produtividade*. Piracicaba: ESALQ/USP/LPV, 2003. p. 89–98.

MOLIN, J. P.; AMARAL, L. R. do; COLAÇO, A. *Agricultura de preci$\{\tilde{a}\}$o:*. Oficina de Textos, 2015. ISBN 9788579752148. Disponível em: <https://books.google.com.br/books?id=MX7jCgAAQBAJ>.

MURAKAMI, E. et al. An infrastructure for the development of distributed service-oriented information systems for precision agriculture. *Computers and Electronics in Agriculture*, v. 58, n. 1, p. 37–48, ago. 2007. ISSN 0168-1699. Disponível em: <dx.doi.org/10.1016/j.compag.2006.12.010>.

NASH, E.; KORDUAN, P.; BILL, R. Applications of open geospatial web services in precision agriculture: a review. *Precision Agriculture*, Springer US, v. 10, n. 6, p. 546–560, 2009. ISSN 1385-2256. Disponível em: <http://dx.doi.org/10.1007/s11119-009-9134-0>.

ORACLE. *Oracle Documentation*. 2017. Disponível em: <http://docs.oracle.com>. Acesso em: 02 mai. 2017.

PERCIVALL, G. et al. *OGC Reference Model*. Open Geospatial Consortium Inc., 2003. 108 p. Disponível em: <http://www.opengeospatial.org/standards/orm>.

POSTGRESQL. *PostgreSQL Documentation*. 2017. Disponível em: <http://www.postgresql.org/docs/9.5/static/docguide.html>. Acesso em: 02 mai. 2015.

QGIS: A free and open souerce geographic information system. 2017. Disponível em: <http://www.arcgis.com/>. Acesso em: 04 mai. 2017.

R. *The R Project for Satistical Computing*. 2017. Disponível em: <http://www.r-project.org/>. Acesso em: 15 mai. 2017.

REZENDE, S. O. *Sistemas inteligentes: fundamentos e aplicações*. [S.l.]: Editora Manole Ltda, 2003. ISBN 8520416837.

RIBEIRO-JÚNIOR, L. C. M. *Uma Arquitetura de Software para Sistemas Espaço-Temporais Baseados na Web para Agricultura de Precisão*. 190 p. Tese (Doutorado em Engenharia) — Universidade de São Paulo, 2007. Disponível em: <http://www.teses.usp.br/teses/disponiveis/3/3141/tde-01082007-182328>.

RSERVE: Binary r server. 2017. Disponível em: <http://rforge.net/Rserve>. Acesso em: 08 mai. 2017.

RUSS, G. *Spatial Data Mining in Precision Agriculture*. 251 p. Tese (Doktoringenieur) — Otto-von-Guericke-University of Magdeburg, 2012.

SANTOS, R. T. dos. *Um modelo de referência para o processo de definição de zonas de manejo em agricultura de precisão*. 115 p. Tese (Doutorado em Engenharia) — Universidade de São Paulo, 2014.

SANTOS, R. T. dos; MOLIN, J. P.; SARAIVA, A. M. A Reference Process for Management Zone Delineation. In: *EFITA Conference - Sustainable Agriculture through ICT Innovation*. [S.l.: s.n.], 2013. p. 8.

WEISS, S. M.; INDURKHYA, N. *Predictive Data Mining: A Practical Guide*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998. ISBN 978-1558604032.

ZHANG, X. et al. Zone mapping application for precision-farming: a decision support tool for variable rate application. *Precision Agriculture*, Springer US, v. 11, n. 2, p. 103–114, 2010. ISSN 1385-2256. Disponível em: <http://dx.doi.org/10.1007/s11119-009-9130-4>.

Distorções de valores altimétricos obtidos com modelos numéricos do terreno gerados com GNSS e RPA

Joel Zubek da Rosa[1], Selma Regina Aranha Reibeiro[1], Alaine Margarete Guimarães[1], Carlos Hugo Rocha [2]

[1] Departamento de Informática, Universidade Estadual de Ponta Grossa, Ponta Grossa, Paraná, Brasil, primeiroautor@xxx.br, terceiroautor@zzz.br

[2] Departamento de Agronomia, Universidade Estadual Paulista, Botucatu, São Paulo, Brasil, segundoautor@yyy.br

RESUMO

Os Modelos Numéricos do Terreno (MNT) são representações matemáticas da distribuição de fenômenos espaciais. Entre as aplicações de MNT está o armazenamento de dados de altimetria para elaboração de modelos de declividade que servem como apoio à análise geomorfológica e de erodibilidade. O objetivo deste trabalho é avaliar o comportamento dos valores de altimetria gerados por meio de imagens obtidas por Aeronave Remotamente Pilotada (RPA) e pontos amostrados em campo mediante receptores de sinal GNSS de dupla frequência (L1\L2) pós-processado, a fim de se obter mapa de declividade com menores distorções. Para a realização deste trabalho foi selecionada uma área agrícola experimental de aproximadamente 16 hectares, adaptada ao sistema de plantio direto. Foram coletados 140 pontos amostrais em forma de grade regular de 45 x 45 metros para a geração dos modelos por meio de interpoladores e o recobrimento aerofotogramétrico da área, realizado com RPA, gerou imagens com aproximadamente 4 cm de resolução espacial. O modelo de altimetria que apresentou melhor média foi o interpolador IDW, porém em desvio-padrão e variância o interpolador Krigagem apresentou melhores valores: 1,13 e 1,29 metros respectivamente

PALAVRAS-CHAVE: Interpolação, Altimetria, Erosão.

ABSTRACT

Numerical Terrain Models (NTM) are mathematical representations of the distribution of spatial phenomena. Among the applications of NTM is the storage of altimetry data for the elaboration of slope models that serve as support for geomorphological and erodibility analysis. The objective of this work is to evaluate the behavior of altimetry values generated by images obtained by Remotely Pilot Aircraft (RPA) and field sampled points using postprocessed double-frequency GNSS signal receivers (L1 \ L2) in order to Lower slope map is obtained.A total of 140 sampling points were collected in the form of a regular 45 x 45 meter grid for the generation of the models through interpolators and the aerial - photogrammetric covering of the area, performed with RPA, generated images with approximately 4 cm of spatial resolution. The altimetry model that presented the best mean was the IDW interpolator, but in standard deviation and variance, the Kriging interpolator had the best values: 1.13 and 1.29 meters, respectively.

KEYWORDS: Interpolation, Altimetry, Erosion.

INTRODUÇÃO

A erosão é considerada uma das principais causas de degradação dos solos agrícolas, devido a remoção da camada superior mais fértil do solo, onde a matéria orgânica e os nutrientes estão concentrados. Na maioria dos casos, as taxas de erosão que ocorrem em áreas agrícolas é maior do que a taxa de formação dos solos (GARCÍA-RUIZ et al., 2015). Mesmo em países altamente desenvolvidos a erosão contínua a exceder as taxas de formação dos solos (AMUNDSON et al., 2016).

Segundo a EMBRAPA (2016) um estudo envolvendo 600 pesquisadores de 60 países, coordenado pela Organização das Nações Unidas para a Alimentação e a Agricultura (FAO), mostrou que mais de 30% dos solos do planeta estão degradados com problemas de erosão, compactação e perda de matéria orgânica. A erosão em solos agrícolas e de pastagens intensiva varia entre cem a mil vezes a taxa de erosão natural e o custo anual de fertilizantes para substituir os nutrientes perdidos pela erosão chega a 150 bilhões de dólares.

Os processos de escoamento e erosão são tipicamente representados em modelos com resolução espacial insuficientes. Essas representações normalmente não conseguem incorporar os efeitos importantes de heterogeneidade topográfica na geração de modelos de escoamento superficial e erosão do solo (HUTTON; NICHOLAS; BRAZIER, 2014). Os modelos de declividade são informação básica de topografia utilizada nas metodologias de identificação de áreas potenciais aos processos de erosão (FUJIHARA, 2002).

A geração de modelos de declividade é realizada por meio de Modelo Numérico do Terreno (MNT). Um MNT é a representação matemática da distribuição espacial de uma determinada região da superfície terrestre, sendo que o Modelo Numérico de Elevação (MNE) é o termo utilizado para a representação altimétrica (BURROUGH, 1986; LONGLEY et al., 2005).

Os MNT são representados por coordenadas 3D (x, y, z) e podem ser obtidos por levantamentos de dados em campo mediante pontos amostrados com espaçamento irregular e regular (FELGUEIRAS, CÂMARA, 2001), Os equipamentos utilizados para a geração de MNT variam em relação ao custo beneficio de acordo com aplicação e precisão dos modelos. Os equipamentos utilizados vão desde teodolitos; receptores de sinal GNSS; estação total; varredura a laser do terreno, conhecida como LIDAR (Light Detection and Ranging); aerolevantamentos e sensores a bordo de satélites.

A aplicação de imagens de RPA para gerar MNT com alta resolução é uma técnica vantajosa, em relação ao custo beneficio, para quantificar e qualificar as alterações de superfície do solo (ELTNER, et al.; 2015). A principal limitação desta tecnologia consiste na obtenção de imagens aéreas válidas para a superfície do terreno quando a vegetação e os resíduos da colheita cobrem o solo, e ao movimento de sombras devido ao tempo transcorrido entre as linhas de recobrimento das fotográficas aéreas adjacentes (BAZZOFFI, 2015).

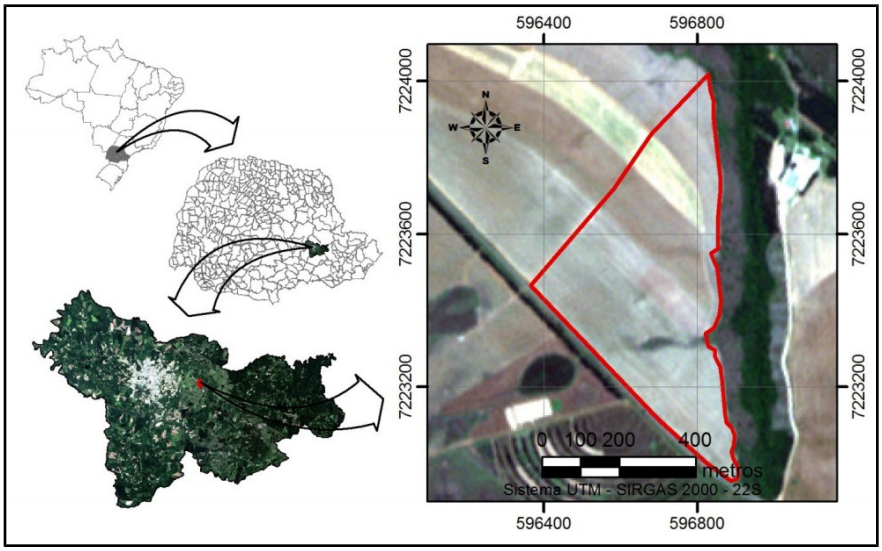
O objetivo desse trabalho é avaliar os modelos de altimetria gerados por meio de imagens obtidas por Aeronave Remotamente Pilotada (RPA) e modelos gerado por três diferentes interpoladores mediante pontos amostrados com receptores de sinal GNSS de dupla frequência (L1\L2) pós-processado.

MATERIAL E MÉTODOS

A realização deste trabalho teve como área de estudo uma parcela de aproximadamente 16 hectares, adaptada ao sistema de plantio direto, localizada na Fazenda Escola da Universidade Estadual de Ponta Grossa. A figura 1 apresenta a localização da área de estudo.

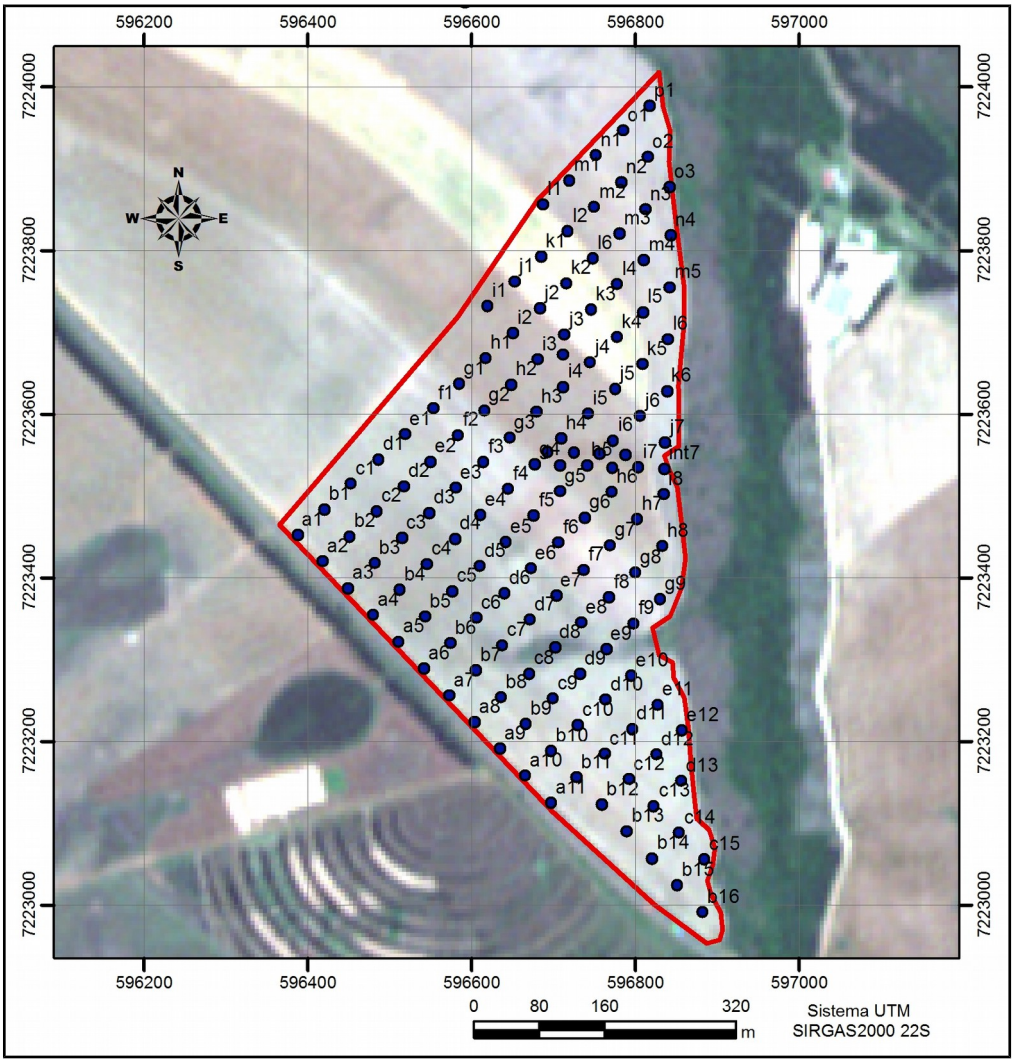
O levantamento de dados em campo foi realizado mediante a implementação de uma grade regular com pontos de 45 x 45 metros de distância. Para marcar os pontos no campo foram utilizados teodolito eletrônico para definir os ângulos retos e a distância entre cada ponto foi medida com trena. Foram marcados pontos intermediários de 22,5 metros em uma feição de erosão linear observada no campo. No total foram coletados 140 pontos para a geração da grade regular. A figura 2 representa a distribuição dos pontos marcados em campo.

Figura 1 – Localização da área de estudo.



Fonte: (Dados IBGE, 2016)

Figura 2 – Grade regular.



Fonte: (O autor, 2016)

Para o georreferenciamento de cada ponto foram utilizados receptores de sinal GNSS (Global Navegation satelitte system). O modelo dos receptores utilizados foi o Trimble R4 Geodésico. Foi utilizado o método estático rápido, permanência de 15 minutos em cada ponto, com fixação de base no campo durante o período mínimo de seis horas para a locação das coordenadas. O pós-processamento das coordenadas foi realizado no programa Trimble Business Center, com a correção de linhas de base, das estações Curitiba UFPR e Guarapuava, disponibilizadas pela Rede de Monitoramento Continuo (RBMC/IBGE).

A interpolação dos dados de altimetria obtidos com os receptores sinal GNSS, foi realizada no programa de Sistema de Informação Geográfica (SIG) ArcGIS 10.3 (ESRI). A análise espacial realizada neste programa utilizou os interpoladores IDW, Krigagem e Spline, com os pontos gerados na grade regular, para gerar modelagens numéricas do terreno.

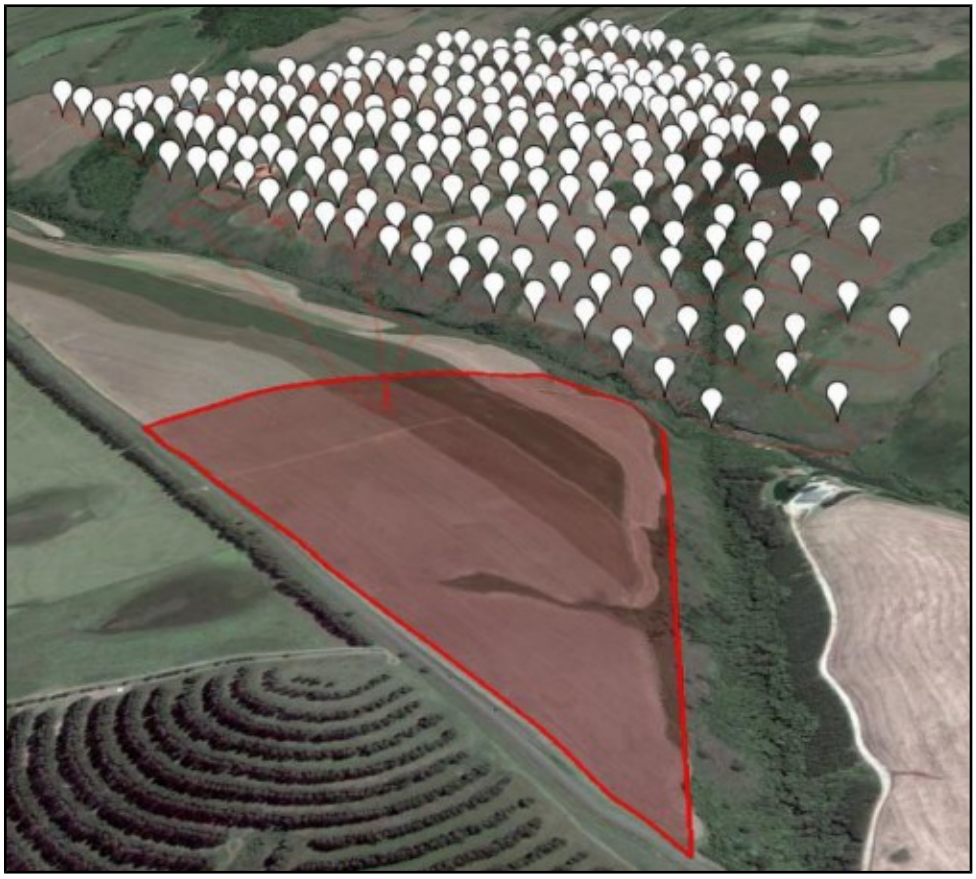
A Krigagem, entendida como um estimador que se baseia numa série de técnicas de análise de regressão, sejam essas lineares ou não, procura minimizar a variância estimada a partir de um modelo prévio levando em consideração a dependência estocástica entre os dados distribuídos no espaço.

O Interpolador IDW se baseia na dependência espacial, supõe-se que quanto mais próximo estiver um indivíduo do outro, maior será a correlação desse indivíduo com seus vizinhos. Dessa forma atribui maior peso para os indivíduos mais próximos do que para os mais distantes do ponto a ser interpolado. A Ponderação é atribuída a pontos amostrais através da utilização de um coeficiente de ponderação que controla como a influência da ponderação diminuirá à medida que a distância a partir do ponto desconhecido aumenta.

Interpolação Spline é uma técnica de aproximação que consiste em se dividir o intervalo de interesse em vários subintervalos e interpolar, da forma mais suave possível, nestes subintervalos com polinômios de grau pequeno

O levantamento de fotografias aéreas foi realizado com RPA. O equipamento utilizado foi eBee da senseFLY e o plano de vôo foi gerado no programa eMotion. A altura media do vôo foi 120m que proporcionou coletar as fotografias com resolução espacial de 3,5cm. Para o recobrimento da área de estudo foram tomadas 210 fotografias. A Figura 3 representa a posição para a tomada das fotografias. Foram distribuídos 10 pontos de controle pela área, coletados com receptores de sinal GNSS, para aumentar a precisão do aerolevantamento. O processamento das fotografias obtidas com a Aeronave Remotamente Pilotada, para gerar orto-mosaico e modelo digital de elevação (DEM), foi realizado no programa PhotoScan (Agisoft).

Figura 3 – Posição das fotografias obtidas com RPA.



Fonte: (O autor, 2016)

RESULTADOS E DISCUSSÃO

O pós-processamento das coordenadas obtidas com receptores de sinal GNSS corrigiu as linhas de base com o erro médio de 0,006 m para precisão horizontal e 0,009 m para precisão vertical. O erro médio dos levantamentos estático rápido foi de 0,011 m para a precisão horizontal e 0,013 m para a precisão vertical. A precisão dos dados gerada pelo levantamento serviu de referência para comparar com os valores altimétricos interpolados e o modelo gerado com as imagens obtidas com RPA.

O orto mosaico e o MNT foram ajustados com pontos que foram marcados com pontos de controle no campo e reconhecidos no orto mosaico. O erro médio quadrático do georreferenciamento, do orto mosaico e modelo digital de elevação foi de 12 cm. Foram coletados 14 pontos aleatoriamente, em toposequência, além dos 140 pontos que foram utilizados para gerar os modelos, para avaliar a precisão dos modelos. A tabela 1 apresenta os valores obtidos com os interpoladores e modelo gerado com as imagens obtidas com RPA, comparados com os valores altimétricos obtidos com as coordenadas obtidas com o sinal GNSS pós-processadas (referencia mais próxima do real).

Tabela 1 – Comparação entre os valores altimétricos obtidos com os modelos.

Referência (GNSS)		Valores interpolados			MNT gerado com as imagens de RPA
		IDW	KRIG.	Spline	
01	1030,20	1029,73	1029,72	1029,72	1027,65
02	1028,76	1028,36	1028,37	1028,18	1026,13
03	1025,46	1023,35	1023,32	1023,32	1024,07
04	1021,48	1021,07	1020,44	1020,71	1017,93
05	1017,57	1017,78	1017,35	1016,89	1019,33
06	1015,06	1013,26	1013,39	1013,52	1018,23
07	1013,58	1013,14	1009,81	1008,48	1013,22
08	1012,98	1011,96	1012,80	1013,63	1016,40
09	1010,30	1011,12	1011,30	1011,45	1013,32
10	1010,07	1009,38	1010,32	1010,60	1013,83
11	1006,92	1004,35	1003,95	1003,72	1007,40
12	1004,53	1002,68	1003,16	1003,06	1008,38
13	999,80	997,20	997,15	997,65	1000,57
14	987,98	990,85	988,90	989,17	992,39

Os dados apresentados na tabela anterior são correlacionados no gráfico 1 que representa os pontos coletados em uma topo-sequência.

O calculo da distorção, em metros de cada ponto, foi realizada subtraindo o valor de cada ponto, obtidos nos modelos, em relação aos valores das coordenadas obtidas com o receptor GNSS. A tabela 2 apresenta as distorções dos modelos em cada ponto e a média, desvio padrão e variância. Para o calculo da média, desvio padrão e variância todos os números assumiram os valores positivos.

Figura 4 – Gráfico com a correlação de valores altimétricos

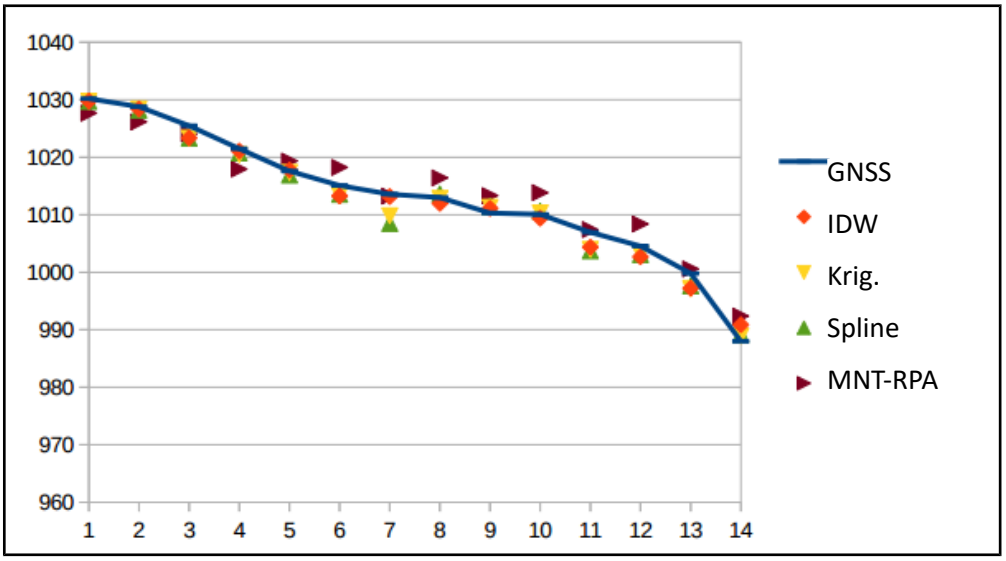


Tabela 2 – Distorções entre os valores altimétricos obtidos com os modelos.

Referência (GNSS)		Valores interpolados			MNT gerado com as imagens de RPA
		IDW	KRIG.	Spline	
01	1030,20	0,47-	0,48-	0,48-	2,55-
02	1028,76	0,40-	0,39-	0,58-	2,63-
03	1025,46	2,11-	2,14-	2,14-	1,39-
04	1021,48	0,41-	1,04-	0,77-	3,55-
05	1017,57	0,21+	0,22-	0,68-	1,76+
06	1015,06	1,80-	1,67-	1,54-	3,17+
07	1013,58	0,44-	3,77-	5,10-	0,36-
08	1012,98	1,02-	0,18-	0,65+	3,42+
09	1010,30	0,82+	1,00+	1,15+	3,02+
10	1010,07	0,69-	0,25+	0,53+	3,76+
11	1006,92	2,57-	2,97-	3,20-	0,48+
12	1004,53	1,85-	1,37-	1,47-	3,85+
13	999,80	2,60-	2,65-	2,75-	0,77+
14	987,98	2,87+	0,92+	1,19+	4,41+
Média		1,30	1,36	1,54	2,50
Desvio padrão		1,36	1,13	1,28	1,33
Variância		1,43	1,29	1,66	1,78

(O sinal de + ou – representa somente se a distorção foi acima ou abaixo da referência mais próxima do real)

O MNT gerado pelo Interpolador IDW apresentou a melhor média de erro de 1.30 m de distorção entre os valores e a menor distorção pontual 0,21 m. Mas o desvio padrão foi o que apresentou o pior comportamento.

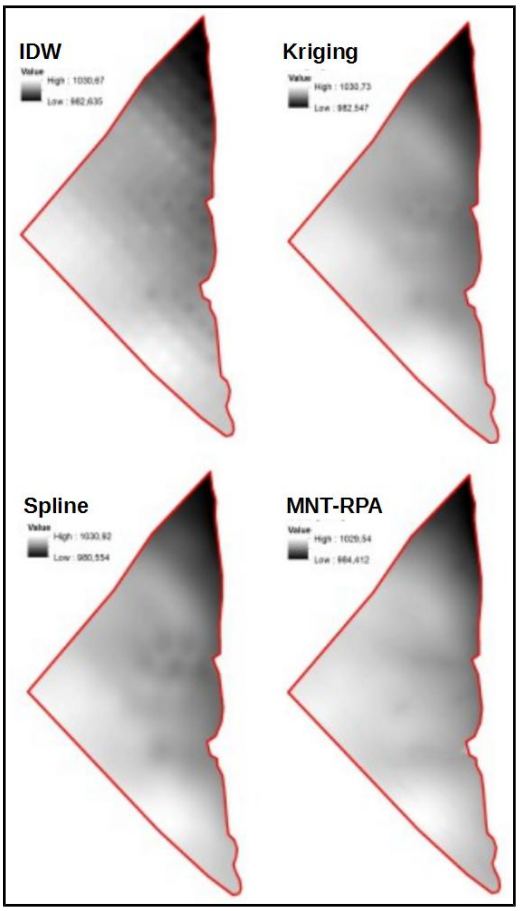
O Desvio padrão e a variância dos valores apresentaram o melhor comportamento no Interpolador krigagem, 1,13 m para o desvio padrão e 1,29 m para a variância dos dados.

O interpolador Spline teve a maior distorção, 5,10 m no ponto 7, que determinou uma média de erro alta comparada aos outros modelos obtidos pela interpolação da grade regular. O desvio padrão demonstrou comportamento razoavelmente baixo neste interpolador, isso indica que a maioria dos dados mantiveram uma tendência

O MNT gerado por meio de imagens de RPA, apresentou uma média de erro alta comparada com os modelos gerados pelos interpoladores. Mas o desvio padrão demonstrou comportamento correlacionado com os demais modelos. Outro comportamento que pode ser observado por esse modelo foi que os valores tiveram uma tendência negativa nos valores que representam maior altimetria, no topo da área de estudo, e tendência positiva nos valores que representam menor altimetria, locais que demostram maiores declividades.

A representação gráfica dos modelos numéricos do terreno podem ser observadas na figura 4. O aspecto visual do modelo gerado pelo interpolador aparenta ter as maiores discrepâncias. O efeito bolha gerado nesse interpolador possivelmente é devido à influência da proximidade de pontos que possuem pouca correlação com seus vizinhos. Os demais modelos demonstraram aspectos visuais similares.

Figura 5 – Modelos numéricos do terreno representado altimetria



Quando se trabalha com interpolação geoestatística, os casos devem ser tratados de forma independente, pois a modelagem está ligada às percepções de quem a manipula (Berveglieri et al., 2011). Neste sentido, para avaliar a precisão dos interpoladores, todos os modelos foram gerados com as configurações padrão implementadas no programa. O ajuste das configurações podem gerar resultados com menores distorções.

CONCLUSÕES

O modelo numérico do terreno gerado pelo interpolador Krigagem apresentou o melhor comportamento em relação os demais modelos.

O modelo gerado pelo RPA demonstrou valores inferiores aos demais modelos, mas apresentou boa tendência nos valores obtidos e vantagem em relação ao tempo de levantamento de dados.

Pretende-se como pesquisas futuras gerar análise espacial com os interpoladores para dados com informações de atributos do solo na mesma escala da grade regular de 45x45

metros e testar diferentes configurações para gerar modelos com fotografias de aeronave remotamente pilotada a fim de diminuir distorções.

AGRADECIMENTOS

Os autores agradecem ao Programa de Pós-Graduação em Computação aplicada/UEPG, onde este trabalho se desenvolveu e a parceria com os laboratórios de Geotecnologias, InfoAgro, Física dos solos e o Laboratório de mecanização agrícola (Lama) pela disponibilização de equipamentos.

REFERÊNCIAS

AMUNDSON, R.; BERHE, A. A.; HOPMANS,J. W.; OLSON, C.; SZTEIN, A. E.; SPARKS, D. L. Soil and human security in the 21st century. SCIENCE sciencemag.org, v. 348 ISSUE 6235, 2015.

BAZZOFFI, P. Measurement of rill erosion through a new UAV-GIS methodology. Italian Journal of Agronomy, v. 10(s1):708, 2015.

BURROUGH, P. A. Principles of Geographic Information Systems for Land Resource Assessment. Monographs on Soil and Resources Survey No. 12, Oxford Science Publications, New York, 1986.

ELTNER, A.; BAUMGART, P.; MAAS, H. G.; FAUST, D. Multi-temporal UAV data for automaticmeasurement of rill and interrill erosion onloess soil. Earth Surf. Process. Landforms, v. 40, 741–755, 2015.

EMBRAPA. Estudo revela que 30% dos solos do mundo estão degradados. Portal da Embrapa, Noticias, Recursos Naturais. 12/07/2016. Disponível em: Acesso em : 19/10/2016

FELGUEIRAS, C. A.; CÂMARA, G. MODELAGEM NUMÉRICA DE TERRENO. In: CÂMARA, G. et. al. Introdução à Ciência da Geoinformação. Cap 7. 36p, 2001. Disponível em: http://www.dpi.inpe.br/gilberto/livro/introd/.

FUJIHARA, A. K. Predição de erosão e capacidade de uso do solo numa microbacia do Oeste Paulista com suporte em geoprocessamento. (ESALQ), 2002.

GARCÍA-RUIZ, J. M.; BEGUERÍA S.; NADAL-ROMERO, E.; JOSÉ C. GONZÁLEZHIDALGO, J. C.; LANA-RENAULT, N.; SANJUÁN, Y. A meta-analysis of soil erosion rates across the world.GEOMORPHOLOGY, V. 239, P. 160–173. 2015

HUTTON, C.; NICHOLAS, A.; BRAZIER, R. Sub-grid scale parameterization of hillslope runoff and erosion processes for catchment-scale models ofsemi-arid landscapes HYDROLOGICAL PROCESSES, v. 24, 1713–1721p, 2014.

LONGLEY, P.A. et al. Geographic information systems and science. 2.ed. Chichester: Wiley, 2005. 497p

Classificação do tipo de ocupação do solo por meio de valores espectrais de imagens de satélite

Camila de Matos Alonso[1], *Ana Paula Lüdtke Ferreira*[1],
Jose Pedro Pereira Trindade[2], *Alexandro Gularte Schäfer*[1]

[1]Programa de Pós-Graduação em Computação Aplicada, Universidade Federal do Pampa
Bagé, Rio Grande do Sul, Brasil
camilamalonso@gmail.com, anaferreira@unipampa.edu.br, alexandro.schafer@unipampa.edu.br

[2]Embrapa Pecuária Sul
Bagé, Rio Grande do Sul, Brasil
jose.pereira-trindade@embrapa.br

RESUMO

O conhecimento dos dados referentes à ocupação do solo permite estabelecer elementos sobre a evolução da dinâmica da paisagem em determinada região, provendo subsídios para a tomada de decisão em relação a políticas públicas, ações de manejo e conservação e alteração em processos produtivos agropecuários. A análise automatizada de imagens de satélite pode contribuir para que os dados de ocupação de uma região possam ser estudados formalmente. O processo de classificação de dados consiste em definir uma função que permita rotular corretamente cada registro de um conjunto de informações com um identificador de um conjunto de classes. Este trabalho apresenta um estudo sobre a classificação de dados de imagens de áreas urbanas, utilizando redes Bayesianas e árvores de decisão, com vistas e discriminar solo, grama e árvores dos demais objetos presentes nas cenas. Os resultados encontrados foram analisados, comparando as técnicas de classificação utilizadas e os valores de acurácia obtidos. A técnica de Redes Bayesianas apresentou uma acurácia de 62.52, enquanto que, a técnica de Árvore de Decisão apresentou uma acurácia de 84.41, sendo a técnica mais adequada para este problema de classificação.

PALAVRAS-CHAVE: Classificação, Árvores de Decisão, Redes Bayesianas

ABSTRACT

Temporal knowledge regarding occupation areas is important in the sense they can provide elements for public policies, conservation issues and production systems decision making. Satellite images analysis can contribute towards this goal. Data classification process consists in

defining a function to correctly attribute labels to each element of a data collection representing classes of interest. This paper presents a study on the classification capabilities of two AI techniques: decision trees and bayesian networks regarding urban areas image classification on four distinct classes: soil, grass, tress and other objects on the scene. The results were analyzed by comparing the classification techniques and the accuracy values obtained. The technique of Bayesian Networks presents an accuracy of 62.52, while the Decision Tree technique has an accuracy of 84.41, being the most adequate technique for this classification problem.
KEYWORDS: *Classification, Decision Tree, Bayesian Networks, Data Mining.*

INTRODUÇÃO

O conhecimento dos dados referentes à ocupação do solo permite a análise e o entendimento sobre os processos estabelecidos da dinâmica dos assentamentos populacionais, nas alterações dos elementos de paisagem, nos processos e produtos de produção agropecuária e na evolução da infraestrura de uma determinada região. Esse entendimento é estratégico para diferentes objetivos, podendo impactar na definição de políticas públicas, em ações duráveis de manejo para a conservação da pecuária de campo e/ou intensificação sustentável da produção animal e agrícola (DIAS et al., 1999)(TRINDADE et al., 2016). Este objetivo pode ser melhor atingido por meio de análise de imagens de satélite, buscando uma classificação das áreas mapeadas em termos dos diversos tipos de ocupação de interesse.

A classificação automática de imagens (PEDRINI; SCHWARTZ, 2008) de satélite vem evoluindo ao longo dos anos com a construção de novos sensores que permitem a obtenção de imagens com alta resolução espacial e de dados hiperespectrais, mas ainda apresenta desafios oriundos da quantidade de detalhes existente em cenas heterogêneas e com possível grande variação de classes de interesse e que possuem respostas espectrais similares em comprimentos de onda no espectro visível (NOVO, 2011). Dessa forma, a investigação de técnicas capazes de classificar imagens de forma (semi)automatizada é um problema de pesquisa relevante.

A tarefa de classificação consiste em buscar uma função que permita integrar corretamente cada registro de dado de um conjunto de informações a um único rótulo de um conjunto de classes (SCOSS, 2006). Métodos de máxima verossimilhança (CERQUEIRA; ALVES, 2010) são usados com frequência, mas as técnicas atuais buscam classificadores inteligentes e probabilísticos com maior poder de discriminar entre os diferentes dados (espaciais ou espectrais) das imagens em análise. Técnicas já clássicas de inteligência artificial, tais como redes neurais, lógica fuzzy ou mineração de dados aparecem em diversos trabalhos da literatura para implementar processos de aprendizado para fins de classificação de imagens dos mais diferentes tipos.

Neste trabalho são desenvolvidos dois modelos para classificação de dados de imagens segmentadas por valores espectrais já classificados de acordo com o tipo de cobertura de solo de interesse. O primeiro modelo é construído pela indução de uma *árvore de decisão* (CARVALHO, 2005) a partir do conjunto de dados. Uma árvore de decisão é uma árvore binária onde

cada nodo contém um teste e, dependendo do resultado da avaliação do teste ser verdadeiro ou falso, a busca ocorre para o ramo esquerdo ou para o ramo direito. As folhas da árvore são os nodos rotulados com as categorias de classificação. O algoritmo de indução é baseado na Teoria da Informação de Shannon, que busca particionar o conjunto de dados de acordo com a máxima entropia de informação associada a cada variável em cada ponto de decisão. O segundo modelo é um método que proporciona a implementação de raciocínio probabilístico, tanto indutivo como abdutivo, denominado *rede Bayesiana*. Uma rede Bayesiana (RULFF; MADSEN, 2008; MARQUES; DUTRA, 2002) consiste em um grafo dirigido acíclico (DAG) em que os vértices representam as variáveis do domínio do problema e os arcos representam dependências causais entre elas. A rede representa ainda a distribuição de probabilidade conjunta, fatorizada por um conjunto de probabilidades condicionais determinadas pela estrutura do grafo. Formalmente, uma rede Bayesiana é um tupla $B = (G, P)$ onde $G = (V, E \subseteq V \times V)$ é um grafo dirigido acíclico e P_V é um coleção de distribuições de probabilidades condicionais $p(X_v|X_{pa(v)})$, indexadas por V, onde X_v é a variável do domínio do problema representada pelo nodo $v \in V$ e $pa(v) = \{v' \in V \mid (v', v) \in E\}$, ou seja, o conjunto dos nodos conectados ao nodo v em G. Como os arcos da rede expressam dependências entre variáveis – e sua ausência expressa independência – a distribuição de probabilidade conjunta $p(X_V)$ pode ser expressa como $p(X_V) = \prod_{v \in V} p(X_v|X_{pa(v)})$. Os classificadores Bayesianos generativos mais simples, conhecidos como *naïve Bayes*, partem da hipótese que todos os atributos são independentes dado a classe (CIRELO; COZMAN, 2005).

O objetivo deste trabalho é construir uma *árvore de decisão* e uma *rede Bayesiana* para classificar imagens. As classes nas quais as imagens devem ser particionadas são: grama, árvores, solo exposto e outros (ou seja, qualquer elemento diferente que possa aparecer na imagem). Ambos os treinamentos são supervisionados, ou seja, a classe à qual área da imagem pertence é conhecida pelo algoritmo. A descrição dos materiais e métodos utilizados e a comparação entre os resultados obtidos com os dois métodos implementados é feita nas seções a seguir.

MATERIAIS E MÉTODOS

Este trabalho visa estudar a pertinência de técnicas de inteligência artificial e raciocínio probabilístico – nomeadamente *árvores de decisão* e *redes Bayesianas* – na classificação de imagens por via da informação espectral disponível. Trata-se, portanto, de uma metodologia experimental de aplicação prática (CERQUEIRA; ALVES, 2010).

A base de dados escolhida para realizar a classificação foi retirada do site *Machine Learning Repository* [1]. Esse repositório digital provê gratuitamente bases de dados sobre diversos domínios do conhecimento, que podem ser usadas para o teste de técnicas de análise. Os dados foram separados no repositório digital em arquivos para treinamento (168 entradas) e teste

[1] https://archive.ics.uci.edu/ml/

(507 entradas). A separação dos dados é necessária para que possa ser verificado o nível de extrapolação do método para dados não previamente usados para o processo de aprendizado – de outra forma, os dados tendem a ser "decorados" pela árvore/rede. A base escolhida contém dados para classificação de áreas de uma imagem de alta resolução em tipos de cobertura de terra urbana. As informações de espectro brilho, média de verde, média de vermelho, proximidade de infravermelho e NDVI (*Normalized Difference Vegetation Index*), foram os parâmetros escolhidos para classificação. Estes dados foram escolhidos, pois inicialmente foi feita uma classificação com todas as variáveis e verificou-se que as mais relevantes para classificação eram as espectrais. Segundo Anderson et al. (2003), o NDVI é uma relação entre medidas espectrais de duas bandas, que são: proximidade do infravermelho (800-1100 nm); e média de vermelho (600-700 nm), que visa eliminar diferenças sazonais do ângulo do Sol e minimizar os efeitos de atenuação atmosférica, observados para dados multitemporais.

O trabalho foi dividido em duas etapas: (1) classificação utilizando o método de redes Bayesianas e (2) classificação utilizando o método de árvore de decisão. A seguir é apresentada uma descrição de cada uma das etapas. Na seção de discussão dos resultados será realizada uma comparação dos dois métodos utilizados.

Classificação por redes Bayesianas

Segundo Faria (2014), a construção de uma Rede Bayesiana pode ser dividida em três etapas: (i) definição das variáveis de interesse para a aplicação, (ii) definição da estrutura da rede, o que inclui o estabelecimento das relações de influência de uma variável sobre outras, conectando as variáveis selecionadas de forma que seja montado um grafo acíclico direcionado por meio das dependências estabelecidas e (iii) obtenção dos parâmetros da rede, definindo uma tabela das probabilidade condicionais para cada variável de interesse em relação com os seus nodos antecessores na estrutura da rede.

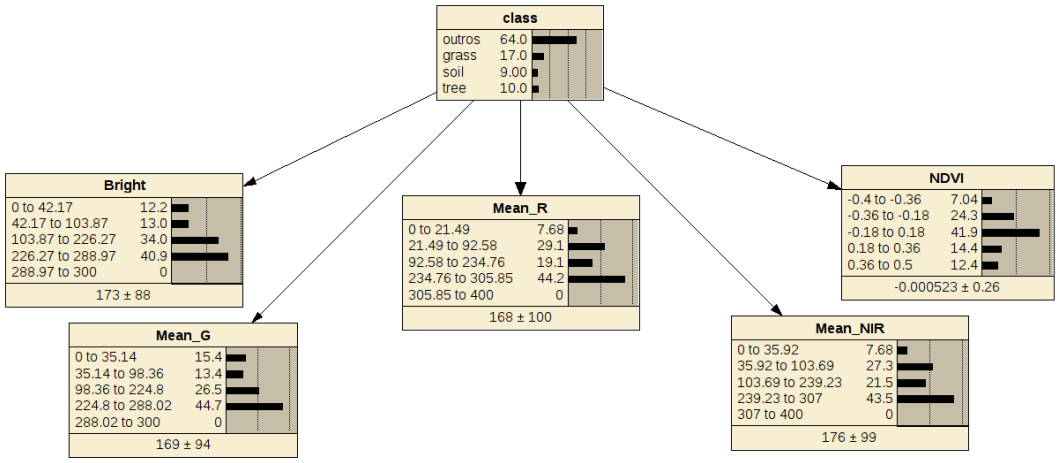
Para a etapa de seleção das variáveis foram consideradas as informações espectrais definidas como de interesse na base de dados para classificação das imagens. As variáveis de interesse são: *brilho*, *verde*, *vermelho*, *proximidade do infravermelho* e *NDVI*. A classe de cobertura é a variável de classificação alvo, que pode ter quatro valores possíveis: *grama*, *árvore*, *solo* e *outros*.

Tendo definidas as variáveis de interesse, a construção da estrutura da rede considerou as relações de causa-efeito entre as variáveis selecionadas. As variáveis espectrais foram consideradas independentes entre si, somente sendo influenciadas pelo tipo de solo representado. Dessa forma, a rede resultante possui seis nodos, sendo que o nodo que representa os possíveis valores de classificação é pai de todos os outros. A estrutura da rede assim descrita foi construída utilizado o programa Netica[2], que pode ser visto na Figura 1.

Após a criação da estrutura da rede, os intervalos de classificação foram criados pressu-

[2]**https://www.norsys.com/netica.html**

Figura 1: Estrutura da Rede Bayesiana



Fonte (Próprio Autor, 2017)

pondo uma distribuição de probabilidades Normal, com média μ, variância σ^2 e desvio padrão $\sigma = \sqrt{\sigma^2}$, conforme apresentado na Tabela 1.

Tabela 1: Discretização dos valores das variáveis de interesse

Variável de Interesse	Média μ	Desvio Padrão σ	Equação
Brilho	165.57	61.70	I1 = ($< \mu - 2\sigma$)
Verde	161.58	63.22	I2 = ($\mu - 2\sigma$) a ($\mu - \sigma$)
Vermelho	163.67	71.09	I3 = ($\mu - \sigma$) a ($\mu + \sigma$)
Proximidade do Infravermelho	171.46	67.77	I4 = ($\mu + \sigma$) a ($\mu + 2\sigma$)
NDVI	0.00	0.18	I5 = ($> \mu + 2\sigma$)

Os valores da média e do desvio padrão de cada variável de interesse, foram aplicados ãs cinco equações apresentadas na Tabela 1. Tendo calculado todos os intervalos, os resultados foram adicionados à rede. Após o cálculo dos intervalos, o arquivo de treinamento foi usado para a geração das probabilidades marginais e condicionas a partir da frequência dos dados e classes na base de treinamento. Como pode ser visto na Figura 1, a probabilidade (marginal ou *a priori*) das classes ficou definida como se segue: outros – 64%, grama – 17%, solo – 9% e – árvore 10%. Também foi definida a probabilidade condicional para cada uma das variáveis de interesse, calculada a partir do programa Netica, pela definição de probabilidade conjunta de duas variáveis aleatórias A e B: $p(A \wedge B) = p(A)p(B|A)$ ou, reescrevendo, $p(B|A) = p(A \wedge B)/p(A)$.

Após o cálculo das probabilidades condicionais e probabilidades *a priori* do conjunto de dados de treinamento, o arquivo de testes foi utilizado para testar a capacidade de aprendizado da rede. O arquivo contém 507 amostras para as quatro classes. Para cada dado de entrada, a rede calcula qual é a probabilidade *a posteriori* da amostra se enquadrar em cada uma das classes de interesse. A classe de maior probabilidade é escolhida, gerando a *matriz de confusão* do processo de treinamento/aprendizado. A matriz de confusão apresenta o número de acertos

e erros de classificação para cada uma das entradas do arquivo de testes. A Tabela 2, apresenta os resultados obtidos.

Tabela 2: Matriz de Confusão para Redes Bayesianas.

Classes	Outros	Grama	Solo	Árvore
Outros	214	72	21	8
Grama	1	81	0	1
Solo	2	16	2	0
Árvore	1	68	0	20

A Tabela 2 apresenta o número de acertos e erros gerados para a classificação de cada uma das classes. As classes de *outros* e *grama*, embora tenham apresentado erros na classificação, o número de acertos foi maior; as classes *solo* e *árvore* apresentaram um valor alto de classificações incorretas. Ocorrendo uma confusão na hora de classificar solo, onde foi classificado em muitos casos como grama e, na hora de classificar árvore, onde 68 vezes foi confundida com grama.

Classificação utilizando árvore de decisão

A árvore de decisão foi induzida, a partir dos dados de treinamento, por meio do algoritmo *J48*, implementado dentro do pacote estatístico *R* (BIVAND; PEBESMA; GóMEZ-RUBIO, 2008), que faz uso de uma metodologia gulosa (*greedy*) (CORMEN; RIVERT; LEISERSON, 2001) para indução de árvores de decisão. Segundo Martins, Marques e Costa (2009), o modelo de árvore de decisão é construído pela análise dos dados de treino e, a partir disto é gerada a árvore, onde cada nó avalia a existência ou significância de cada atributo individual. O primeiro passo foi treinar o modelo para, a partir dele, poder gerar a árvore de decisão. Tendo o arquivo de testes no pacote *R*, basicamente são necessários três comandos, apresentados na Figura 2, para gerar a árvore.

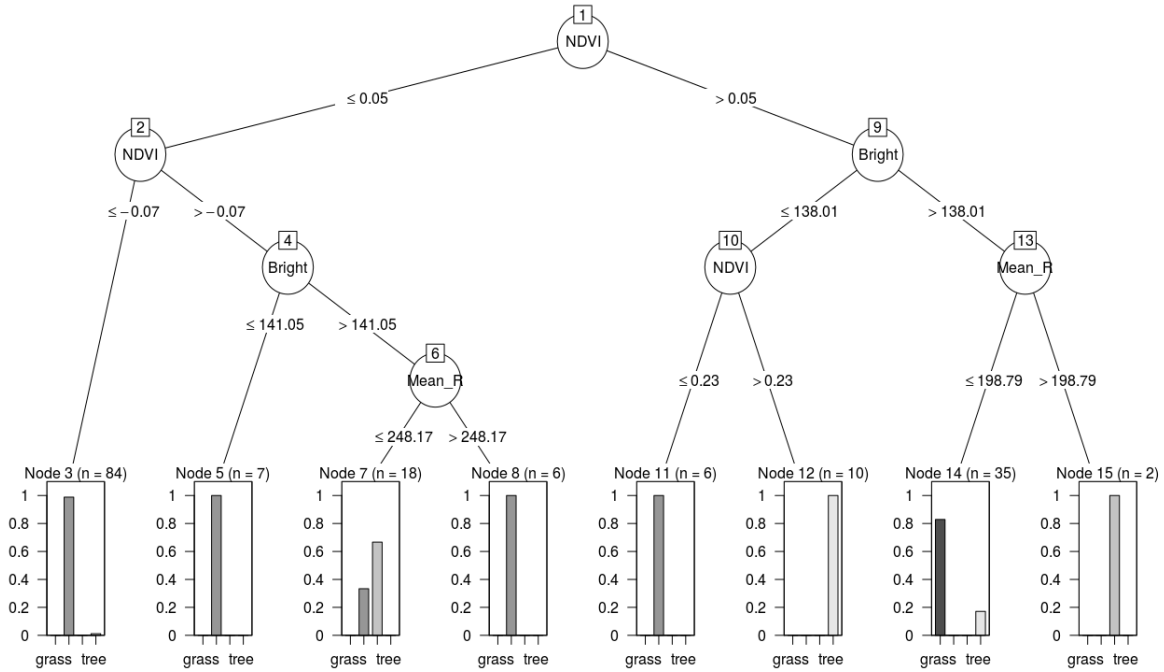
Figura 2: Comandos R para indução da árvore de decisão

```
> install.packages("RWeka")
> library(RWeka)
> arvorej48 <- J48(class ~., data=dados)
```

O primeiro comando serve para instalar o pacote *Rweka*, que contém um conjunto de algoritmos de aprendizado de máquina para tarefas de mineração de dados, contendo ferramentas para pré-processamento, classificação, regressão, agrupamento, regra de associação e de visualização de dados. O segundo comando é utilizado para ativar o pacote instalado. Por fim, o terceiro comando é responsável pela criação da árvore de decisão. O algoritmo *J48* tem dois argumentos principais: (i) a *fórmula de entrada*, que indica o nome da variável de decisão seguida do símbolo "~" e uma lista das variáveis que podem ser usadas no modelo para obter a decisão; neste caso, foi utilizado o sinal "." para indicar que todas as variáveis existentes no arquivo, podem ser usadas no modelo e (ii) o arquivo com os dados que devem ser utilizados para indução da árvore.

Após gerar a árvore de decisão, o comando plot() foi utilizado, para criar uma representação gráfica da árvore. A Figura 3 apresenta a árvore de decisão induzida.

Figura 3: Árvore de Decisão



Fonte (Próprio Autor, 2017)

A partir da árvore de decisão gerada é possível perceber que as variáveis que têm mais influências na classificação são NDVI e brilho. Note-se que no nodo 7 existe uma confusão na classificação entre *solo* e *outros*. Isto pode ser explicado pelo fato que na classe *outros* são agrupadas informações referentes a asfalto, edifícios, concreto e outros elementos urbanos que podem ser facilmente confundidos durante a classificação. Outro nodo que apresentou confusão no processo de classificação foi o 14, onde existem duvidas em relação às classes *grama* e *árvore*. Tendo analisado a árvore, o próximo passo foi gerar a matriz de confusão a partir do arquivo de testes, onde é possível ver o número de acertos e erros de classificação para cada uma das classes. A Tabela 3, apresenta os resultados obtidos.

Tabela 3: Matriz de Confusão para Árvore de Decisão.

Classes	Outros	Grama	Solo	Árvore
Outros	292	4	0	3
Grama	1	70	0	40
Solo	17	5	20	0
Árvore	5	4	0	46

Analisando a Tabela 3, podemos conferir que as classes com o número maior de classificação incorretas é as que apareceram nos nodo 7 e 14. Na classificação de grama ocorreram 70

classificações corretas e 40 classificações incorretas, onde foram classificadas como árvore. Na classificação do solo também houve um número significativo de classificações incorretas, onde 17 vezes foi classificado como outros.

RESULTADOS E DISCUSSÃO

Nesta seção será realizada uma comparação entre os dois métodos de classificação utilizados, a fim de identificar qual método foi mais eficiente na classificação dos tipos de cobertura de terra urbana e quais são as limitações descobertas no processo. Para isto, a Tabela 4 apresenta o número total de acertos, erros e o valor de acurácia para os dois métodos de classificação.

Tabela 4: Comparação entre os dois métodos de classificação.

Método	Acertos	Erros	Acurácia
Redes Bayesianas	317	190	62.52
Árvore de Decisão	428	79	84.41

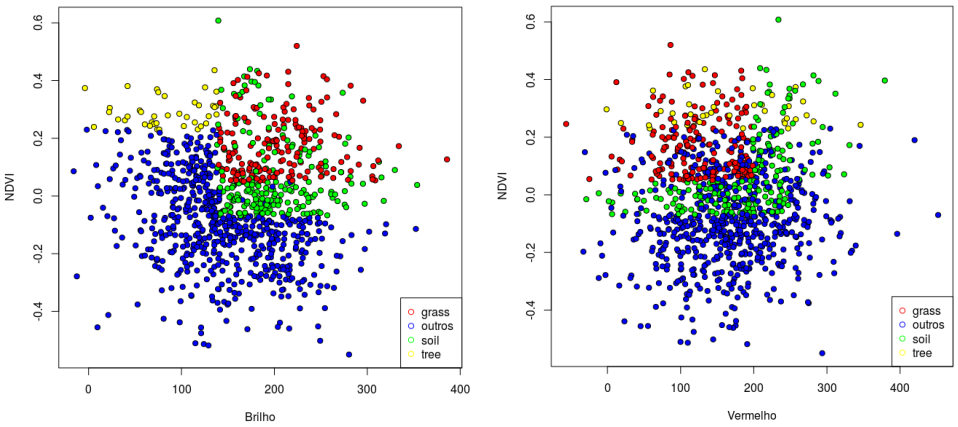
Ao analisar a Tabela 4 é possível perceber que a árvore de decisão obteve um resultado melhor em relação à rede Bayesiana, tendo uma acurácia de 84,41. Entretanto, é possível perceber que para os dois métodos houve uma grande confusão na hora de classificar *árvore* e *grama*, ainda que o valor do brilho seja discriminante no resultado dessa classificação. Aqui pode-se ver como o espectro visível gera incertezas na classificação, mesmo quando outras variáveis discriminam o resultado. A similaridade de cor - variável verde - provavelmente foi a causa do erro. Esse dado permite apontar que existe necessidade de investigação sobre a capacidade da discriminação de outras variáveis espectrais e como organizá-las em rede de forma a gerar um maior poder de discriminação dos dados.

Para visualizar como as classes estão relacionadas com as variáveis foram criados dois gráficos de dispersão, utilizando a variável NDVI, que foi a mais relevante na classificação para a árvore de decisão, em relação às classes *brilho* e *vermelho*, que foram as duas outras classes mais frequentes. A Figura 4 (a) e (b) apresentam estes gráficos.

A Figura 4(a) apresenta o gráfico de dispersão das variáveis *brilho* e *NDVI*. A classe *outros*, como previsto na rede Bayesiana, é dominante, aparecendo em torno de 64% das vezes. As classes *grama* e *árvore* só possuem valores positivos de NDVI. Enquanto isto, uma pequena parte dos solos possui valor positivo e, para a classe de *outros*, valores de brilho até 130 ou 140 possuem valor positivo de *NDVI*. Em relação ao *brilho*, excetuando a classe *outros*, existe uma divisão nos valores de 130 ou 140: abaixo desses valores temos a classe *árvore* e, acima destes valores temos as classes *grama* e *solo*.

Na Figura 4(b) é possível analisar que a classe *grama* tem valor até cerca de 200 em relação ao *vermelho* e as demais classes superam o valor 300. Em relação à variável *NDVI*, é possível perceber que os dados estão de acordo com o da Figura 3 (a), onde ficou claro que a classe *árvore* tem valores acima de 0.2 e a classe *outros* abaixo de 0.2.

Figura 4: Gráfico de Dispersão.



(a) Brilho × NDVI. (b) Vermelho × NDVI.

CONCLUSÕES E TRABALHOS FUTUROS

Este trabalho apresentou a classificação de um conjunto de dados sobre áreas urbanas, provenientes de imagens de satélite, utilizando dois métodos de classificação. Para a implementação de redes Bayesianas foi utilizado o programa *Netica* para gerar a estrutura da rede e as probabilidades condicionais e marginais do conjunto de dados de treinamento. Este método apresentou um número grande de classificações incorretas, principalmente para a classe *outros*, onde foi confundido pelo menos uma vez com todas as outras classes. Enquanto isto, para o método de árvore de decisão, foi utilizado o programa R para gerar a árvore e a matriz de confusão. Este método, embora também com erros de classificação, apresentou resultados melhores em relação às redes Bayesianas, obtendo uma acurácia de 84.41. Desta forma, o método de árvore de decisão se apresentou como uma alternativa mais eficiente para classificar a cobertura de terra urbana, a partir dos dados usados para treinamento e teste no âmbito deste trabalho.

Como trabalhos futuros, pretende-se usar árvores de decisão para investigação de variáveis discriminantes dos dados de interesse do problema. A partir deste trabalho, poderá ser definido um conjunto de variáveis que atenda melhor os critérios necessários para classificação das imagens. Poder-se-á, então verificar se os resultados podem ser qualificados com as técnicas já descritas neste trabalho. Os dados de entrada também deverão ser gerados a partir das imagens de sensoriamento remoto disponíveis para a área da região da Campanha do RS e bacia do Rio Negro, áreas de interesse para determinação da evolução dos campos nativos e interferência antrópica na paisagem.

REFERÊNCIAS

ANDERSON, L. O. et al. Sensor modis: uma abordagem geral. *São José dos Campos: INPE*, v. 58, 2003.

BIVAND, R. S.; PEBESMA, E. J.; GóMEZ-RUBIO, V. *Applied Spatial Data Analysis with R*. [S.l.]: Springer, 2008. (Use R!). ISBN 978-0-387-78170-9 (Print) 978-0-387-78171-6 (Online).

CARVALHO, D. R. Árvore de decisão/algoritmo genético para tratar o problema de pequenos disjuntos em classificação de dados. ***Universidade Federal do Rio de Janeiro**, Rio de Janeiro, Brazil. Doctor Thesis. 162pp*, 2005.

CERQUEIRA, J. A. C.; ALVES, A. D. O. ClassificaÇÃo de imagens de alta resoluÇÃo espacial para o mapeamento do tipo de pavimento urbano. In: *III Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação*. Recife: [s.n.], 2010.

CIRELO, M. C.; COZMAN, F. G. ***Aprendizado semisupervisionado de classsificadores bayesianos utlizando testes de independência***. [S.l.]: EPUSP, 2005.

CORMEN; RIVERT; LEISERSON. *Introduction to Algorithms*. [S.l.]: MIT Press, 2001.

DIAS, M. C. et al. *Manual de impactos ambientais: orientações básicas sobre aspectos ambientais de atividades produtivas*. Fortaleza: Banco do Nordeste, 1999.

FARIA, R. C. ***Redes probabilísticas: aprendendo estruturas e atualizando probabilidades***. Tese (Doutorado) — Universidade de São Paulo, 2014.

MARQUES, R. L.; DUTRA, I. Redes bayesianas: o que são, para que servem, algoritmos e exemplos de aplicações. ***Coppe Sistemas–Universidade Federal do Rio de Janeiro**, Rio de Janeiro, Brasil*, 2002.

MARTINS, A. C.; MARQUES, J. M.; COSTA, P. D. **Estudo comparativo de três algoritmos de machine learning na classificação de dados electrocardiográficos**. *Trabalho (Mestrado em Informática Médica)–Universidade do Porto, Porto*, 2009.

NOVO, E. M. L. de M. *Sensoriamento Remoto – Princípios e Aplicações*. 4. ed. [S.l.]: Blucher, 2011.

PEDRINI, H.; SCHWARTZ, W. R. (Ed.). *Análise de Imagens Digitais: Principios, Algoritmos e Aplicações*. São Paulo: Thomson, 2008. 501 p.

RULFF, U. B. K.; MADSEN, A. L. *Bayesian Networks and Influence Diagrams – A Guide to Construction and Analysis*. [S.l.]: Springer, 2008. (Information Science and Statistics). ISBN 978-0-387-74100-0 (Print) 978-0-387-74101-7 (Online).

SCOSS, A. M. A clusterização e classificação no processo de data mining para análise do desempenho docente no ensino de graduação. ***Universidade do Extremo Sul Catarinense**, Criciúma, Brasil*, 2006.

TRINDADE, J. P. P. et al. Dinâmica dos campos da campanha, fronteira oeste e missões do rio grande do sul. In: *XXVI Congresso Brasileiro de Zootecnia*. [S.l.: s.n.], 2016.

Direcionadores tecnológicos da Pesquisa Agropecuária Intensiva em Dados: mapeamento de competências, ferramentas e infraestrutura

Martha Delphino Bambini[1], *Roberto Hiroshi Higa* [2], *Maria Beatriz Machado Bonacelli*[3]

[1] Doutoranda no PPG-PCT/IG/Unicamp, Analista de Transferência de Tecnologia na Embrapa Informática Agropecuária, Campinas, SP, martha.bambini@embrapa.br

[2] Pesquisador na Embrapa Informática Agropecuária, Campinas,SP, roberto.higa@embrapa.br

[3] Professora no Departamento de Política, Científica e Tecnológica, do Instituto de Geociências, da Universidade Estadual de Campinas, Campinas, SP, bia@ige.unicamp.br

RESUMO

Este artigo apresenta um estudo de monitoramento envolvendo as tendências tecnológicas associadas à aplicação de Pesquisa Computacional Intensiva em Dados em problemas do setor agropecuário. Este estudo foi conduzido a fim de oferecer informações qualificadas para apoio à gestão de uma infraestrutura computacional corporativa com foco em armazenamento e processamento de grandes volumes de dados da Empresa Brasileira de Pesquisa Agropecuária (Embrapa). A infraestrutura visa apoiar a geração de conhecimento de impacto para o domínio agropecuário. O principal resultado desta pesquisa é a descrição dos direcionadores tecnológicos relacionadas à aplicação da Ciência de Dados (*Data Science*) ao setor agropecuário, com foco nas seguintes dimensões: aspectos conceituais, competências e drivers tecnológicos envolvendo o paradigma (ferramentas de software; organização de bases de dados; plataformas especializadas; serviços computacionais). O estudo oferece informações tecnológicas qualificadas para apoiar a gestão da infraestrutura computacional corporativa da Embrapa.

PALAVRAS-CHAVE: Pesquisa Computacional; Agropecuária; Ciência de Dados; Big Data.

ABSTRACT

This article presents a monitoring study involving the technological trends associated to the application of Data Intensive Computational Research to agricultural sector problems. This study was conducted in order to provide qualified information to manage a corporate computing infrastructure focused on the storage and processing of large volumes of data of the Brazilian Agricultural Research Corporation (Embrapa). It would support the generation

of impacting knowledge related to Agriculture. The main result of this research is the description of the technological drivers related to the application of Data Science to the agricultural sector, focusing on the following dimensions: conceptual aspects, competences and technological drivers involving the paradigm (software tools; Databases, specialized platforms, computer services). The study offers qualified technological information to support the management of Embrapa´s corporate computing infrastructure.

KEYWORDS: Computing Research; Agriculture; Data Science; Big Data.

INTRODUÇÃO

Vive-se hoje em um momento marcado pela 3ª fase da Revolução das Tecnologias de Informação e Comunicação (TIC), descrita por Cowhey, Aronson e Abelson (2009). Os autores apontam que esta fase é marcada por um ponto de inflexão, no qual as novas arquiteturas de TIC – chamadas de Nuvem ou Grid – estruturam-se em um conjunto bastante distintos de capacidades e organização de mercado do que nas fases anteriores. Estas arquiteturas consideram três tendências relacionadas: (i) a existência de redes de Internet de banda larga disponíveis, a custo acessível; (ii) a integração de um armazenamento massivo e barato de informação, associada a uma nova arquitetura de redes e serviços; e (iii) a emergência de sistemas de computação virtual que controlam e se utilizam de muitos computadores, incluindo supercomputadores, de forma a atender demandas dos usuários e aumentar eficiência e produtividade.

Avent (2014) destaca que esta nova onda tecnológica é formada por um conjunto de novas tecnologias – como inteligência de máquina, Internet Ubíqua (onipresente) e robótica avançada – capazes de possibilitar o desenvolvimento de várias invenções, entre ela: veículos não tripulados, drones, sistemas automáticos de tradução de idiomas, tecnologias móveis que possibilitam diagnóstico e tratamento médico à distância e também aprendizado virtual. No contexto desta nova fase da revolução das TIC, o volume de geração de dados de pesquisas em áreas de interesse agropecuário vem crescendo rapidamente e as necessidades de armazenamento de processamento de grandes volumes de dados se tornam prementes.

Bell (2009) descreve as 3 atividades básicas da computação intensiva em dados: captura, curadoria e análise. Os dados são coletados em escalas e formatos variados, podendo envolver resultados de experimentos e de observações individuais, laboratórios únicos, grupos de laboratórios parceiros, redes de laboratórios internacionais em larga escala bem como conjuntos de dados pessoais de indivíduos pesquisados. No Século XXI, a grande maioria do

grande volume de dados científicos é capturada por instrumentos, de forma contínua, ou informação geradas a partir de modelos computacionais (HEY, TANSLEY e TOLLE, 2009), processados por softwares específicos, sendo as informações resultantes armazenadas em bases de dados. Estes dados podem ser armazenados por longo prazo e disponibilizados de forma pública, curados e gerenciados de forma a garantir sua análise continuada, embasando a geração de novas teorias científicas derivadas. Neste contexto, a velocidade de geração de dados pode ser muito grande, dificultando o processo de captura e armazenamento. Assim, há que se criar um conjunto genérico de ferramentas que cubram o conjunto completo destas atividades científicas: captura, validação de dados, curadoria, análise e armazenamento.

O objetivo deste trabalho é identificar e analisar os principais direcionadores tecnológicos relacionados ao conceito de Pesquisa Computacional Intensiva em Dados - chamada no documento de *Data Science* - aplicada ao setor agropecuário. Este estudo se insere no âmbito das atividades do projeto GDAE, intitulado "Estruturação de processos e redes de relacionamento em apoio à gestão do arranjo Armazenamento e Processamento de Dados Experimentais da Embrapa - DATAExp". Este projeto tem entre seus objetivos delinear os processos de estruturação e gestão de uma nova infraestrutura computacional institucional para o uso por projetos de pesquisa da Empresa Brasileira de Pesquisa Agropecuária (Embrapa) que demandem armazenamento e processamento de grandes volumes de dados para gerar conhecimento de impacto para o domínio agropecuário. Algumas temáticas de pesquisa consideradas relacionadas ao uso desta plataforma são: genética e genômica, agricultura de precisão e modelagem climática. A adoção de uma infraestrutura computacional institucional visa ao uso racional dos recursos computacionais da empresa e à otimização dos investimentos em novos equipamentos e infraestruturas de processamento de alto desempenho.

MATERIAL E MÉTODOS

O trabalho aqui apresentado consiste em um estudo de monitoramento tecnológico, conforme descrito em Porter et al. (1991), envolvendo a aplicação de Pesquisa Computacional Intensiva em Dados ao domínio agropecuário. O estudo foi conduzido a partir de análise de literatura de forma a mapear as principais tendências tecnológicas relacionadas ao tema. A busca foi conduzida na base da *Web of Science* a partir das palavras-chave "*Data Science*" e "Agriculture", sendo complementada com consulta no *Google Scholar* e na plataforma LinkedIn. Foram consideradas as seguintes dimensões: aspectos conceituais, competências da

equipe, e drivers tecnológicos envolvendo o paradigma (ferramentas de software; organização de bases de dados; plataformas especializadas; serviços computacionais).

RESULTADOS E DISCUSSÃO

Esta seção apresenta os conceitos de base relacionados à pesquisa conduzida e o levantamento de competências relacionadas à Ciência de Dados; e um mapeamento dos Drivers Tecnológicos envolvendo infraestrutura, produtos e serviços relacionados à aplicação deste paradigma ao setor agropecuário.

Análise da tecnologia e seu contexto: aspectos conceituais

Jim Gray (HEY, TANSLEY e TOLLE, 2009) descreve o 4º paradigma da ciência marcado por uma dinâmica de pesquisa colaborativa, em rede e direcionada a dados, chamada pelo autor de *e-Science.* O termo descreve a síntese de Tecnologia de Informação e Ciência, levando a grandes mudanças e transformações nas atividades de pesquisa científica.

A ciência intensiva em dados, ou *Data Science*, se refere a um campo de estudo interdisciplinar que se utiliza de processos científicos e computacionais para extrair conhecimento, padrões, tendências e insights a partir de conjuntos de dados de vários formatos, estruturados ou não-estruturados. Bell, Hey e Szalay (2009) descrevem o "dilúvio de dados" que vem ocorrendo nas últimas décadas, quando dados oriundos de sensores, de simulações ou outras ferramentas computacionais de análise necessitam ser analisados em um processo científico intensivo em dados. Os autores destacam que, mesmo com um enorme potencial identificado, a ciência intensiva em dados tem se desenvolvido ainda de forma lenta, talvez pela falta de entendimento da comunidade científica em relação a este tema e às ferramentas práticas para criar, gerenciar e utilizar estas bases de dados. Ressaltava-se, naquele momento, a urgência por novas ferramentas e tecnologias para tornar possível os trabalhos de *Data Science*. Alguns campos científicos de aplicação citados pelos autores são: a Astronomia, Física de alta energia e a Genômica.

Fenômeno Big Data

Outro conceito que também se relaciona com a *Data Science*, que se transformou em uma *buzzword* na área de negócios, é *Big Data*. Conforme Diebold (2012), o termo *Big Data* - relacionado à Ciência da Computação, à Estatística e à Econometria - se originou de trabalhos não acadêmicos da empresa Silicon Graphics Inc. (SGI) em meados da década de 1990. O autor destaca que não é possível dizer que seja um novo campo de estudos tendo em vista a existência de trabalhos prévios já desenvolvidos em Ciência da Computação e Estatística já há

muitos anos. Mas, as técnicas e ferramentas utilizadas para lidar com a questão de *Big Data* são realmente inovadoras, como C*loud computing*, associação de algoritmos "massivamente paralelos", métodos para controlar descobertas falsas ao testar milhões de hipóteses, entre outras. Pospiech e Felden (2012) indicam que o termo *Big Data* vem sendo cada vez mais estudado em publicações científicas, em várias disciplinas e sob várias definições. No que se refere ao mercado, Few (2012) destaca que este termo vem sendo veiculado por várias publicações na área de *business intelligence* (BI), configurando-se pura e simplesmente como uma campanha de marketing de empresas relacionadas ao setor.

O fenômeno "*Big Data*" é geralmente caracterizado na literatura por 3 Vs: **volume** de dados (quantidade); **velocidade** de aquisição (criação, acumulação, inserção e processamento dos dados) e grande **variedade** de tipos de dados, tornando o processo de gestão e análise de dados mais complexo e pesado. Dados mais tradicionais são, em geral, estruturados, como caracteres, números, etc. Nas últimas décadas, as aplicações e sistemas passaram a gerar um grande aumento de dados não-estruturados como textos, áudio, vídeo, imagens, dados geoespaciais e de internet (como logs e registros de cliques). As implicações desta grande variedade de tipos de dados são: maior complexidade na construção e inserção destes dados em uma base computacional e maior dificuldades na transformação de dados para processamento analítico e computacional. Na próxima subseção, serão apresentados as competências e o perfil do Cientista de Dados.

Novas competências e um novo perfil profissional para lidar com Data science

Minelli, Chambers e Dhiraj (2013) destacam que os grandes conjuntos de dados não representam o maior desafio no contexto estudado, mas sim, a identificação e o desenvolvimento de métodos menos custosos e confiáveis para extrair valor dos conjuntos de *terabytes* e *petabytes* de dados. Além de empregar tecnologias e novos métodos de análise, com maior velocidade e escala para lidar com alta complexidade, torna-se necessário também **contratar e/ou capacitar profissionais e criar novos processos de trabalho**.

Davenport e Patil (2012) descrevem um **novo perfil profissional,** o cientista de dados ou *data scientist,* termo cunhado em 2008: Este profissional deve ter o treinamento e as habilidades necessárias para fazer descobertas a partir de um grande conjunto de dados. A demanda por estes profissionais é crescente, mas a oferta nem tanto, tendo em vista que não existem programas de graduação e/ou pós-graduação com este grau de especificidade.

Existem hoje alguns programas específicos de formação e iniciativas de capacitação *on the job* por parte de algumas empresas.

Uma consulta efetuada em 27/01/2017 na seção de empregos (*Jobs*) da plataforma LinkedIn (2017) utilizando o termo "*data scientist*" retornou 58 empregos no Brasil com este termo. Já nos Estados Unidos, são 11.776 vagas anunciadas para "data scientists". Na Inglaterra são 1.354 vagas para este cargo; na Alemanha, 658 vagas; na França são 449; na Holanda 229; na Espanha são 159. Em países da América Latina temos 49 vagas no México; 11 vagas no Chile e 10 na Argentina. Ainda no LinkedIn, o termo "data scientist" retornou 63.203 pessoas com este termo em seu perfil e 110 grupos relacionados ao tema.

Existe ainda pouco consenso sobre o papel deste tipo de profissional em uma organização, sua alocação funcional; como podem agregar valor e como sua performance deve ser mensurada (DAVENPORT e PATIL, 2012). Em termos de competências técnico-científicas, espera-se que cientistas de dados possuam fundamentos sólidos em matemática, estatística, probabilidade e computação, sendo a formação científica desejável independente do campo de formação. Habilidades específicas podem ser construídas em treinamentos *on-the-job* e eventuais cursos sobre ferramentas tecnológicas. Os profissionais devem saber codificar software, ainda que não seja sua principal habilidade, e também conseguir comunicar os resultados de seu trabalho de forma a que os *stakeholders* consigam compreendê-los e utilizá-los. É muito importante que cientistas de dados se conectem com comunidades de prática, seja na organização em que trabalha ou externamente com outros profissionais da área. Outras habilidades indicadas pelos autores são: saber identificação e integração de bases de para completar dados faltantes e gerar melhores resultados; extrair *insights* e informações relevantes; comunicar estes *insights* e informações às equipes; apresentação e visualização criativa dos dados, criando padrões de interesse; sugerir implicações para o direcionamento dos negócios, sobre processos executados e produtos comercializados.

Análise de Drivers Tecnológicos envolvendo Data Science

Segundo Minelli, Chambers e Dhiraj (2013), as crescentes demandas para tratamento de dados com **volume, variedade e velocidade** aumentam as demandas por plataformas de computação e tecnologias de software capazes de lidar com escala, complexidade e velocidade a fim de garantir a competitividade de organizações, sejam empresas ou institutos de pesquisa. Os autores indicam que diferentes sistemas como software para gestão e análise

de dados, sistemas livres e proprietários bem como configurações de hardware, estão sendo combinadas para criar novas soluções em Tecnologias de Informação para lidar com os desafios de *Big Data Analytics.*

Davenport e Patil (2012) destacam também que é comum encontrar casos em que os cientistas de dados desenvolvem suas próprias ferramentas, conduzindo pesquisas em estilo acadêmico, muito embora Minelli, Chambers e Dhiraj (2013) apontem a existência também soluções proprietárias.

Pospiech e Felden (2012) indicam que os pesquisadores divergem quanto ao uso debasesescalabilidade e performance exigida por este tipo de aplicação. Outras abordagens mencionadas pelos autores são: bases de dados funcionando em paralelo; NoSQL; e bases de dados distribuídas para analisar conjuntos de dados muito grandes.

As arquiteturas computacionais para processamento intensivo em dados precisam ser escaláveis, ou seja, devem ser projetadas para que sua capacidade de processamento e armazenamento acompanhe o crescimento da complexidade dos problemas tratados e o correspondente volume de processamento de dados requerido. Usualmente, essas arquiteturas são classificadas em duas categorias, do ponto de vista de escalabilidade (CINTRA, NAKAI e CORREA, 2014): **escalabilidade vertical** em que os computadores podem ter até milhares de processadores, dezenas de *terabytes* de memória RAM e sistemas de armazenamento de dados especializados que podem atingir *petabytes* de capacidade; e **escalabilidade horizontal**, representadas por *cluster* e *grids* de computadores, cuja capacidade é expandida pela adição de novos nós, ou seja, computadores ao sistema. Enquanto este último possui a vantagem de ter um custo por processador ou memória mais acessível, é preciso ter em mente que existem problemas para os quais uma solução só é viável em sistemas verticalmente escaláveis. É por esse motivo que centros de computação científica, dedicados a apoiar projetos de pesquisa intensivo em dados (ex: *Commonwealth Scientific and Industrial Research Organisation* (CSIRO, 2017)) em geral possuem os dois tipos de soluções.

O projeto Apache Hadoop (APACHE, 2017), formado por uma biblioteca de software livre, detecta e manipula falhas na camada de aplicação, provendo um serviço de alta disponibilidade sobre o cluster de computadores, cada um dos quais pode estar sujeito a falhas. Ele é composto por diversos módulos, dentre os quais destaca-se o MapReduce (DEAN e GHEMAWAT, 2008), um modelo de programação simples, originalmente desenvolvido pelo Google, e que permite o processamento e geração de grandes volumes de

dados. Ele baseia-se no princípio de solução de problemas conhecido como "dividir para conquistar", consistindo em dividir o problema original em diversos problemas menores que podem ser distribuídos para serem processados, de forma independente e em paralelo, em nós de um cluster de computadores; e então, posteriormente, consolidados para obter a solução do problema original. Além dos módulos que o compõem, o Hadoop também possui um rico ecossistema de projetos associados, com funcionalidades para processamento de grandes volumes de dados, que podem ser consultados em Apache (2017).

Um outro direcionador-chave do fenômeno *Big Data* e das pesquisas em *Data Science* é o *Cloud computing*, ou computação em nuvem, que pode ser vista com uma forma de consumo de recursos computacionais (máquina virtual, *storage* ou aplicação) como um serviço, sem a necessidade de constituir e manter uma infraestrutura computacional própria. Esse tipo de serviço possui vários atrativos para o usuário, tais como o auto provisionamento, a elasticidade e o pagamento apenas pelo uso dos recursos; pode ser oferecido por provedores comerciais (nuvem pública), pela própria instituição (nuvem privada) ou em um modelo híbrido, que contempla os dois casos anteriores; sendo o serviço categorizado de 3 diferentes formas: infraestrutura como serviço (IaaS), plataforma como serviço (PaaS) e software como serviço (SaaS). Esses modelos subsidiam novos modelos de negócios envolvendo a prestação de serviços associados ao armazenamento, organização, validação, processamento e análise de grandes volumes de dados.

Embora, seu uso esteja bastante difundido em aplicações comerciais, a comunidade de computação de alta performance ainda tem hesitado em entrar nesse mercado (SCIENTIFIC COMPUTING WORLD, 2017). Isso, contudo, parece estar mudando à medida que caem os custos e tornam-se disponíveis novos modelos de programação que permitem a execução mais eficiente na nuvem de processamentos de tarefas que exigem alta performance. Provedores de nuvens públicas estão instalando configurações de hardware próprios para HPC (*High Perfomance Computing*) ao mesmo tempo em que nuvens privadas estão dando aos usuários de computação científica a experiência sobre como executar tarefas em um ambiente em nuvem. Entretanto, ainda na mesma referência, o autor chama atenção para o fato de que nem toda aplicação de HPC é apropriada para execução em um ambiente baseado em nuvem. Tarefas fracamente acopladas sem muitas operações de entrada e saída (I/O) são bons candidatos para execução em ambiente de nuvem; enquanto aplicações com requerimentos de

memória muito elevados ou tarefas altamente paralelizadas em dezenas de milhares de núcleos não.

Ainda, outras questões relacionadas ao uso dessa tecnologia são a **segurança da informação e o armazenamento de dados sensíveis** (CINTRA, NAKAI e CORREA, 2014), quando se está utilizando o serviço oferecido por provedores públicos. Uma alternativa para essa situação é a estruturação de uma nuvem híbrida e a utilização de uma estratégia que preserve os dados mais sensíveis na parte privada da nuvem. *Scientific Computing World* (2017) aponta que a tendência atual é que o próximo passo no desenvolvimento da tecnologia de nuvem seja a interligação de várias nuvens privadas, formando uma federação de nuvens. Dessa forma, uma empresa multinacional (ou um conjunto de instituições) com centros de dados em múltiplas localizações poderá balancear a carga de processamento entre todos os centros de dados. Para instituições de pesquisa, essa é uma tendência que vai ao encontro com o que se espera para o futuro da ciência intensiva em dados e e-Science (HEY, TANSLEY e TOLLE,2009).

CONCLUSÕES

Esta pesquisa buscou contribuir para o projeto GDAE da Embrapa gerando informações qualificadas para apoiar a estruturação dos processos e as redes de relacionamento para apoiar a gestão desta nova infraestrutura computacional. Entende-se que, em consonância com as tendências encontradas na literatura e com o trabalho de ARS (2013), as instituições de pesquisa que, como a Embrapa, busquem uma posição de liderança em computação científica aplicada a agropecuária, necessitam efetuar investimentos em infraestrutura computacional, em novas práticas e no desenvolvimento de competências. Os investimentos devem prever: aumento da eficiência computacional na instituição com a redução de redundância nas compras de hardware; desenvolver a capacidade de compartilhar dados entre laboratórios, instituições e disciplinas; aprimorar competências da força de trabalho; criação de novos métodos computacionais e estabelecimento de padrões para coleta de dados, estruturação, bem como sua documentação e disseminação, entre outros.

REFERÊNCIAS

APACHE - *Apache Software Foundation*. Disponível em: <https://www.apache.org/>. Acesso em: 06 jan.2017.

AVENT, R. The third great wave. *The Economist*, October, v. 4, 2014.

ARS - Agricultural Research Service. United States Department of Agriculture – USDA. *Big Data and Computing Building a Vision for ARS Information Management. Workshop Summary Feb. 5-6, 2013*. Disponível em: <https://www.ars.usda.gov/> Acesso em: 13.fev.2017.

BELL, G. Foreword. In: HEY, T., TANSLEY, S. , TOLLE, K. M. *The fourth paradigm: data-intensive scientific discovery*. Redmond, WA: Microsoft Research, 2009. pp. xi-xiv.

BELL, G.; HEY, T.; SZALAY, A. Beyond the data deluge. *Science* v. 323, n. 5919, p. 1297-1298, 2009.

CINTRA, L.C; NAKAI, A.; CORREA; J. L. Métodos, procedimentos e técnicas utilizadas na construção de AgroTIC. In Tecnologias da Informação e Comunicação e suas relações com a agricultura. Massruhá, S. M. F. S.; Leite, M. A. A.; Luchiari Jr., A.; Romani, L. A. S. (eds). Embrapa. Brasília, DF. 2014.

COWHEY, P.F., ARONSON, J.D., ABELSON, D. *Transforming Global Information and Communication Markets: The Political Economy of Innovation*. Cambridge, USA: Massachusetts Institute of Technology, 2009. 341 p.

CSIRO - Commonwealth Scientific and Industrial Research Organisation. Disponivel em: <https://confluence.csiro.au/display/SC/Facilities> Acesso em: 06 jan.2017.

DAVENPORT, T. H.; PATIL, D. J. Data scientist. *Harvard business review*, v. 90, n. 5, p. 70-76, 2012.

DEAN, J.; GHEMAWAT, S. MapReduce: simplified data processing on large clusters. Communications of the ACM, v. 51, n. 1, p. 107-113. 2008.

DIEBOLD, F. X. "On the Origin(s) and Development of the Term "Big Data". *PIER Working Paper* 12-037. Philadelphia, PA: Penn Institute for Economic Research, 2012.

FEW, S. Big data, big ruse. Visual Business Intelligence Newsletter, 2012.

HEY, T., TANSLEY, S. , TOLLE, K. M. *The fourth paradigm: data-intensive scientific discovery*. Redmond, WA: Microsoft Research, 2009. 252p.

PORTER, A.; BANKS, J; ROPER, A. T.; MASON,T.; ROSSINI, F.; WIEDERHOLT, B. *Forecasting and management of technology*. New York, NY: John Wiley and Sons, 1991. pp. 114-132.

POSPIECH, M.; FELDEN, C. Big data–a state-of-the-art. (July 29, 2012). *AMCIS 2012 Proceedings*. Paper 22. Disponível em: <http://aisel.aisnet.org/> Acesso em: 24 Jan 2017.

LINKEDIN. Disponível em: < https://pt.linkedin.com/> Acesso em: 06 jan.2017

MINELLI, M.; CHAMBERS, M.; DHIRAJ, A. What Is Big Data and Why Is It Important?. In: *Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses*, 2013. pp. 1-18.

SCIENTIFIC COMPUTING WORLD. *HPC finally climbs into the cloud*. 16 de fevereiro de 2016. Disponível em: <https://www.scientific-computing.com/feature/hpc-finally-climbs-cloud/> Acesso em: 6 Jan.2017.

SZALAY,A. S. , BLAKELEY,J. A. Gray's laws: database-centric computing in science. In: HEY, T., TANSLEY, S. , TOLLE, K. M. *The fourth paradigm: data-intensive scientific discovery*. Redmond, WA: Microsoft Research, 2009. pp. 5-11.



Farming 4.0: Estado da Arte e Potenciais Futuros

Henike Guilherme Jordan Voss[1], *Sergio Luiz Stevan Jr*[1,2]

[1]Programa de Pós-Graduação em Computação Aplicada (PPGCA)
Universidade Estadual de Ponta Grossa (UEPG)
Ponta Grossa, Paraná, Brasil
henike_jordan@hotmail.com

[2]Departamento de Engenharia Eletrônica
Universidade Tecnológica Federal do Paraná (UTFPR)
Ponta Grossa, Paraná, Brasil
sstevanjr@utfpr.edu.br

RESUMO

A evolução das tecnologias computacionais e de conectividade nos últimos anos tem avançado a passos largos em diferentes setores da economia, da indústria à agricultura, alterando a forma como pensamos, agimos e produzimos. Desde a inserção da mecanização no campo até a utilização abundante de sensoriamento, a agricultura tem evoluído proporcionando o aumento contínuo de produção e da qualidade dos cultivos, inserindo o conceito de Agricultura Inteligente (*Smart Farming*). Entretanto, com o advento da conectividade e da circuitaria de comunicação de dados, sistemas de Internet da Coisas (IoT) tem extrapolado os limites das cidades e também alcançado o campo. Assim, os sinais destes sensores estão sendo coletados e disponibilizados em nuvem, fomentando o conceito ainda em formação denominado de *Farming 4.0*, que é a evolução da Agricultura Inteligente utilizando a IoT, proporcionando assim um controle cada vez mais otimizado do agronegócio. Neste cenário, este trabalho apresenta uma revisão bibliográfica de Agricultura Inteligente, identificando os trabalhos correlatos que embasam a definição de *Farming 4.0* na evolução do agronegócio.

PALAVRAS-CHAVE: *Smart Farming*, Internet das Coisas, IoT, Sensoriamento, *Big Data*.

ABSTRACT

The evolution of computing and connectivity technologies in recent years has been making strides in different sectors of the economy, from industry to agriculture, changing the way we think, act and produce. Since the inception of mechanization in the field to the abundant use of sensing, agriculture has evolved providing continuous increase of production and quality of

crops by inserting the concept of Intelligent Agriculture (Smart Farming). However, with the advent of connectivity and the circuitry of data communication, Internet of Things (IoT) systems has extrapolated the boundaries of cities and also reached the field. Thus, the signals from these sensors are being collected and made available in the cloud, fomenting the concept still in formation denominated of Farming 4.0, that is the evolution of the Intelligent Agriculture using the IoT, thus providing an increasingly optimized control of agribusiness. In this scenario, this work presents a bibliographic review of Intelligent Agriculture, identifying the related works that supports the definition of Farming 4.0 in the evolution of agribusiness.
KEYWORDS: Smart Farming, Internet of Things, IoT, Sensing, Big Data.

INTRODUÇÃO

A agricultura tem sido uma das mais importantes atividades humanas, uma vez que é responsável pela produção de recursos indispensáveis, como alimentos, medicamentos, energia, entre outros. Após a revolução indústrial, a maioria dos países têm enfatizado os papéis essenciais da indústria agrícola e tecnologias relacionadas que afetam a produção agrícola. As práticas agrícolas tenderam a apoiar um maior consumo de energia através dos grandes maquinários e o aumento do uso de fertilizantes e produtos químicos. Por exemplo: uso de equipamentos pesados e aviões agrícolas implementados foram capazes de aumentar a eficiência do trabalho na agricultura, economizando recursos humanos e aumentando drasticamente a produtividade agrícola. Além dessas práticas, outras têm implicações sociais e ambientais negativas, tais como: a erosão do solo, a salinização, a fertilidade do solo, a compactação dos subsolos, a poluição do solo e da água. Essas práticas geralmente têm se apoiado nas necessidades de alimentos de uma população humana em rápido crescimento. É necessária uma mudança de paradigma para um novo método de produção que garanta uma agricultura segura e sustentável (LIAGHAT; BALASUNDRAM et al., 2010).

A chave para o futuro agronômico reside no uso de análises personalizadas de dados por produtores individuais, trabalhadores agrícolas e produtores de alimentos. Para que as análises se tornem instrumentos agrícolas deste século, os novos desenvolvimentos da Tecnologia da Informação (TI) que aceleram o comércio eletrônico (por exemplo, computação em nuvem, aprendizagem de máquinas, sistemas cliente-servidor móveis) devem ser baratos e acessíveis, abordar uma série de problemas distintos, incluindo aumentar os rendimentos, conservar a água e garantir a saúde dos solos e das plantas, e facilitar novas práticas agrícolas sustentáveis (KRINTZ et al., 2016).

Com tudo isso exposto, este trabalho tem o objetivo de demonstrar o estado da arte e explicar o *Farming 4.0*, que envolve todos esses conceitos recentes de TI, aplicados à agricultura, sendo uma das tendências para o agronegócio e tema de discussão em eventos por todo o mundo. A organização do trabalho se pauta em fazer uma fundamentação teórica dos assuntos: Internet das Coisas e *Smart Farming*; discutir a aplicação de sensores na agricultura; trazer o Estado da

Arte de sistemas inteligentes implantados nas fazendas nesses últimos anos; mostrar a evolução da agricultura com a utilização da mecanização com o sensoriamento e monitoramento; buscar conceitos que possam esclarecer o *Farming 4.0*; apresentar ideias de propostas e projetos que não estejam incluídas na literatura que podem ser produzidas como trabalhos futuros; por fim, discutir a importância de obter todos esses dados conjuntos da fazenda e disponibilizar na nuvem para o acesso através de dispositivos eletrônicos.

ESTADO DA ARTE

Na década de 50, foi verificado que pela primeira vez o número de tratores excedeu o número de animais usados na agricultura para tração, sendo que nessa época já havia o uso intensivo de fertilizantes e pesticidas. Na década seguinte, pela primeira vez, foi implantado dispositivos eletrônicos de monitoração que permitiu aos agricultores plantarem de forma mais eficiente, de forma que são ligados a plantadores mecânicos e semeadores de ar que monitoram o número e o espaçamento das sementes sendo plantadas (NORMAN, 2002).

Nos anos 80 houve o crescimento de métodos de plantio direto para limitar a erosão, um crescimento acentuado de equipamentos agrícolas e mais agricultores começavam a usar técnicas de agricultura sustentável de baixo insumo para reduzir as aplicações químicas. Na década de 90, a TI e as técnicas de precisão são utilizadas cada vez mais na agricultura. Os agricultores começavam a usar tecnologia de satélite (*Global Positioning System* - GPS) para rastrear e planejar suas práticas agrícolas. Usando a nova Agricultura de Precisão (AP), os agricultores usavam receptores do GPS para registrar locais precisos em suas fazendas para determinar quais áreas precisam de quantidades particulares de água, fertilizantes e pesticidas. As informações puderam ser armazenadas em um cartão e transferidas para um computador doméstico. Os agricultores conseguiam combinar esses dados com informações de rendimento, previsões meteorológicas e análise de solo para criar planilhas. Estas ferramentas permitiram ainda maior eficiência na produção de alimentos. Nesta mesma década também foi concedido pela primeira vez a aprovação para um alimento ser produzido através da biotecnologia (NORMAN, 2002).

Na virada do século houve a expansão da Internet, a disponibilidade das informações na *Web 2.0*, a inserção das informações na nuvem em 2009 e a criação do conceito de *Big Data* por meados de 2011. Até que em 2012 surgiu os primeiros conceitos relacionados a Internet das Coisas (*Internet of Things* - IoT), juntamente com os princípios do *Farming 4.0* (CLASEN, 2016).

FUNDAMENTAÇÃO TEÓRICA

A IoT é uma rede de dispositivos físicos circundantes, veículos, dispositivos relacionados à eletrônica, software embutido e hardware. As redes conectadas os ajudam a coletar e trocar dados entre si. Pode-se comparar os dados obtidos com a localização da área especificada e

análise de dados anteriores ou qualquer padrão pré-definido. IoT ajuda ou permite detectar os dados do monitoramento, bem como controlar os componentes de qualquer lugar do mundo, onde existe disponibilidade de conexão com a Internet. Portanto, este tipo de sistema resulta na melhoria da eficiência e precisão (SUKHADEVE; ROY, 2016). A Figura 1 ilustra um esquema de tendência futuras abordando a IoT.

Figura 1: Tendências futuras em IoT



Adaptado de: (SUKHADEVE; ROY, 2016)

Portanto, a IoT torna-se ideal para o rastreamento (localização e seguimento) de alimentos. Para este fim, uma aplicação Web, pode reunir os dados para um produto das bases de dados descentralizadas e mostrar claramente através de uma página ou de um aplicativo. Com base nisso, a infraestrutura genérica da Internet das Coisas oferece muitos usos, e este termo não deve mais ser usado para aplicações individuais (CLASEN, 2016).

O uso de tecnologias de sensores é cada vez mais aplicável na agricultura hoje em dia. No campo da agricultura de precisão (AP), é um auxílio inevitável para a geração de informações espaciais e temporais específicas do local para apoiar estratégias de manejo de culturas. Na última década, várias empresas de maquinária agrícola e de construção de sensores estabeleceram uma multiplicidade de sistemas de sensores para detectar parâmetros relacionados ao solo e às plantas, bem como para detectar fatores de impacto ambiental, influenciando o desenvolvimento das plantas cultivadas (GEIPEL et al., 2015). O desenvolvimento recente adicionou a possibilidade de utilização móvel em plataformas de sensores aéreos, por exemplo, sistemas de câmeras montadas em Aeronaves Remotamente Pilotadas (*Remotely Piloted Aircraft* - RPA).

O *Smart Farming*, incorporado com sistemas IoT, poderia ser chamado de uma fazenda conectada, que pode suportar uma ampla gama de dispositivos de diversos fabricantes de dispositivos agrícolas com diversas tecnologias envolvidas. Além do mais, fazendas conectadas poderiam fornecer serviços agrícolas mais inteligentes com base no conhecimento de especialistas compartilhados (RYU et al., 2015). Por exemplo, as pessoas com pouca experiência na agricultura podem cultivar plantas ou culturas a fim de obter lucro. A prevenção de doenças

infecciosas também pode ser outro benefício do desenvolvimento das fazendas conectadas na qual pode-se, através da detecção do vírus da gripe em porcos, realizar o isolamento daqueles saudáveis. O *Smart farming* é baseado na evolução das seguintes tecnologias: sensores e atuadores mais baratos e aperfeiçoados, sistemas de software e hardware de baixo custo, comunicação de aparelhos móveis com alta largura de banda, aplicativos para *Smartphones* e *Tablets*, tecnologias de posicionamento terrestre e análise de grande quantidade de dados, como mostra a Figura 2.

Figura 2: Visão geral sobre *Smart Farming*



Adaptado de: (RETHINK, 2015)

O contínuo aumento no volume e detalhe dos dados capturados pela IoT produziu um fluxo esmagador de dados em formato estruturado ou não estruturado e emergiu como uma tendência amplamente conhecida como *Big Data* (Grande volume de dados). O *Big Data* inicia com grande volume heterogêneo, fontes autônomas com controle distribuído e descentralizado, e procura explorar relações complexas e em evolução entre os dados (WU et al., 2014). Com isso, torna-se um desafio complexo em extrair conhecimentos e informações a partir do *Big Data*, considerando que cada fonte de informação possui um protocolo (linguagem) distinto e pode haver até políticas de privacidade na troca dessas informações. Assim, algumas áreas como o *Data Mining* (Mineração de dados) procura encontrar padrões nesses dados, a fim de facilitar a extração de informações.

A computação em nuvem é uma das mudanças mais significativas nas modernas tecnologias de informação e tornou-se uma arquitetura poderosa para executar computação em grande escala e complexa. As vantagens da computação em nuvem incluem recursos virtualizados, processamento paralelo, segurança e integração de serviços de dados com armazenamento de dados escalável (HASHEM et al., 2015). A computação em nuvem não só pode minimizar o custo e a restrição para automação e informatização para a agricultura, mas também pode proporcionar custos reduzidos de manutenção de infra-estrutura, gerenciamento eficiente e acesso de

usuários, tornando-se indispensável a sua utilização na cadeia agrícola.

Recentemente, a fazenda inteligente (*Smart Farming*) evoluiu para o *Farming 4.0* que seria a possível identificação (endereçamento) de todos os dispositivos na fazenda, para que eles possam ser acessados (através da Internet) por qualquer pessoa em qualquer lugar do mundo. A Figura 3 resume a definição de *Farming 4.0*, que engloba todos os conceitos explicados.

Figura 3: Visão geral sobre *Farming 4.0*



Fonte: Elaborado pelos autores

ESTUDO DE CASOS, DISCUSSÃO E EXEMPLOS

Vários trabalhos relacionados ao *Smart Farming* baseiam-se em desenvolver um sistema integrado de gerenciamento da fazenda, desde captar dados através de sensores e disponibilizá-los em tempo real na rede para o agricultor, que com base nessas informações geradas pelo sistema, tomar decisões que podem ser captadas pelos atuadores na fazenda.

Roham et al. (2015) desenvolveram um sistema para monitorar condições climáticas, como gás carbônico, temperatura e umidade de diferentes estufas utilizando redes de sensores sem fio, com o dispositivos *ZigBee* para comunicação, uma aplicação Web, de *Smartphone* e um controlador *BeagleBone*. Jagannathan, Priyatharshini et al. (2015) propuseram uma nova metodologia para o *Smart Farming* ligando um sistema de sensoriamento inteligente e um sistema irrigação inteligente através da tecnologia de comunicação sem fio. Este sistema foca em mensurar parâmetros físicos como: umidade, quantidade de nutrientes e pH do solo que são de vital importância nas atividades agrícolas. Baseado em parâmetros químicos e físicos essenciais do solo avaliado, a quantidade requerida de água é irrigada nas culturas usando um irrigador inteligente, que é montado em um sistema móvel no topo do trator. Ryu et al. (2015) apresentaram uma fazenda conectada baseada em sistemas de Internet das Coisas, que permite fornecer sistemas de *Smart farming* para usuários finais. Um projeto detalhado e implementação para fazendas conectadas são ilustradas, e suas vantagens são explicadas com cenários de serviços em comparação com sistemas de *Smart Farming* desenvolvidos anteriormente. Assim como alguns dos trabalhos relacionados, o monitoramento da fazenda é realizado através de *Smartphone*.

Já Lavanya (2016) projetou um robô agrícola de alta potência e alta capacidade de limpeza do solo para arar, semear e regar colheitas com uma mão-de-obra reduzida, tornando esse

veículo eficiente. Este robô opera o cultivo da fazenda considerando linhas e colunas específicas a uma distância fixa dependendo da cultura. Este robô pode ser controlado via *BlueTooth* usando um *Smartphone* e o processo de cálculo, processamento e monitoramento são projetados com motores e uma interface com o microcontrolador. Patil et al. (2016) introduziram um sistema automatico de irrigação baseado na plataforma Arduino e uma aplicação *Android* que irá auxiliar no controle do Arduino via Internet. Além disso, essa aplicação *Android* ajuda os agricultores com informações agrícolas como o custo de grãos, nível de umidade do solo requerido, tipo de solo necessário, previsão do tempo, fertilizantes e pesticidas para serem utilizados.

Viani (2016) mostrou as potencialidades de economia de água de um sistema de irrigação inteligente escalável aplicado à agricultura de precisão. Uma estratégia de lógica *Fuzzy* é integrada em um sistema de monitoramento distribuído baseado na tecnologia de rede de sensores sem fio. O cronograma ideal de irrigação é estimado adaptativamente de acordo com a medição em tempo real do estado do solo e das condições climáticas. O objetivo do sistema é otimizar o conteúdo de água no solo para maximizar o estado de saúde das plantas, melhorar a qualidade dos produtos e reduzir o desperdício de água.

Jayaraman et al. (2016) projetaram o *SmartFarmNet*, que é uma plataforma baseada na Internet das Coisas que pode automatizar a coleta dos dados ambientais, de solo, de fertilização e irrigação; correlacionar automaticamente tais dados e filtrar os dados inválidos da perspectiva de avaliar o desempenho da cultura; e calcular as previsões de safras e recomendações de culturas personalizadas para qualquer fazenda em particular. Este sistema pode integrar virtualmente qualquer dispositivo IoT, incluindo sensores disponíveis comercialmente, câmeras, estações meteorológicas, entre outros, e salvar seus dados na nuvem para análise de desempenho e recomendações. O *SmartFarmNet* fornece análises e recomendações de desempenho de culturas.

Kumar (2017) desenvolveu um sistema de irrigação de culturas totalmente automatizado. A principal motivação por trás deste sistema é diminuir o desperdício de água e gerir de forma eficaz a quantidade de água regada nas plantas. São usados painéis solares para fornecer energia para o sistema durante o dia. Quando o teor de umidade do solo vai abaixo de um certo limite para uma cultura, o sistema de bomba é acionado e a cultura é irrigada. As plantas são irrigadas até que o valor desejado seja alcançado e então a bomba é desligada automaticamente. Por fim Inamdar e Roy (2017) mostram uma visão geral da arquitetura básica da Internet das Coisas, seus requerimentos de segurança, e também diferentes domínios de aplicações como: cidades inteligentes e casas inteligentes que, como mostra o artigo, provavelmente serão a condução das pesquisas futuras próximas em relação à IoT.

A partir do que foi exposto, pode-se verificar que o *Smart Farming* está atualmente presente nas fazendas nas formas de monitorar alguns setores de interesse. Mas o conceito de *Farming 4.0* vai mais a fundo e se baseia em manter todos os setores da fazenda conectada de modo que o produtor tenha o acesso em tempo real de tudo que está acontecendo na fazenda, e não apenas de uma parte dela.

Farming 4.0 significa que as máquinas agrícolas podem agir de forma autônoma, tornando-se responsáveis pela produção agrícola. Tratores inteligentes, colheitadeiras e robô de ordenha já são uma realidade e estão perto de produção em série (CLASEN, 2016). Com isso metas de produção são definidas no futuro para que sejam cumpridas. Mas como isso deve ser feito em detalhes, permanece sob a área das máquinas. Para o sistema de produção de culturas o agricultor pode, por exemplo, ter maior produtividade. Os pulverizadores e distribuidores de adubo, então, podem decidir por si só quando e onde irão implantar os químicos. O agricultor, no futuro, provavelmente terá que intervir cirurgicamente apenas em caso de falha, mantendo a sua atividade limitada a fixação de metas.

A partir disso as chamadas tecnologias inteligentes também são incrementalmente montadas em tratores, colheitadeiras, e em outros equipamentos como: dispositivos de controle inteligente, sensores para a operação da máquina e processos agronômicos, capacidade de automação avançada (orientação, colocação de sementes, pulverização) e tecnologias de comunicação embutidas no veículo.

Além disso, algumas aplicações do *Farming 4.0* são também introduzidas como: produtos físicos são melhorados com serviços digitais adicionais com novos algorítmos sendo desenvolvidos para transformar dados em informações para otimizar produtos e processos na agricultura, reduzir riscos e limitar a vulnerabilidade de influências externas como falhas do maquinário, temperatura e doenças. Essas aplicações são introduzidas entre diferentes setores na cadeia agrícola, ou seja, a informação na forma digital existe para todos os setores e processos da fazenda (CLASEN, 2016).

Como resultado, o equipamento agrícola tem se tornado um dos vários elementos no sistema de produção, embora seja muito importante, não apenas como o maior gerador de dados, mas também o executor de mapas e planos gerados por plataformas de dados e sistemas agrícolas.

Com base nas definições e na literatura, pode-se expor as seguintes possibilidades de trabalhos futuros para o *Farming 4.0* na área de Agricultura de Precisão: acompanhar o crescimento das culturas através de RPAs e fazer uma mapeamento da área para possíveis correções de forma não linear. Outra possibilidade seria fazer o mapeamento do solo, enquanto ara, e assim fazer uma correção pontual e personalizada. Também há uma lacuna com a possibilidade de o agricultor fazer a gestão de manutenção do maquinário, mantendo um Banco de Dados na nuvem em contato com os fabricantes e fornecedores desses equipamentos; gestão de consumo de insumos como: sementes, adubos, defensivos agrícolas e fertilizantes, também em contato bilateral com os fornecedores desses produtos; gestão da qualidade de grãos; consulta e monitoramento de uso de defensivos controlados;. Para os produtores rurais que possuem geração própria de energia, seria interessante eles obterem um histórico e informações instantâneas de consumo de energia, principalmente nas granjas e nos lacticínios.

Alguns dos trabalhos correlatos citados nesse artigo, basearam-se em desenvolver sistemas de controle automatizado de irrigação, mas existe ainda a possibilidade em fazer um sis-

tema de gestão de bombas e válvulas controlados à distância, permitindo uma maior liberdade para o agricultor.

Por fim, há a possibilidade de fazer a gestão de deslocamento do gado, usando tecnologias de identificação individual por radiofrequência (RFID - *Radio-Frequency IDentification*), monitoramento distribuído das condições climáticas na propriedade rural e parâmetros de existentes nas granjas, gestão da aplicação de vacinas, rastreio de matrizes, acompanhamento de mercado financeiro e de cooperativas.

Com essas possibilidades expostas, a fim de criar um sistema *Farming 4.0* é necessário manter todos os dados na nuvem, a fim de integrar tudo isso e possibilitar ao agricultor o acesso e o controle total da fazenda.

Com base no que foi exposto, é visível que o sensoriamento cada vez mais está integrado nas fazendas e que auxilia o agricultor para o obtenção de dados a respeito da fazenda. Com o acesso cada vez mais facilitado e a redução dos preços de sensores e atuadores, os agricultores têm a oportunidade de aumentar a produtividade e reduzir os custos nos diversos setores envolvidos na fazenda. Assim é necessário disponibilizar esse apanhado de dados de forma organizada na nuvem, a fim de o agricultor obter toda a informação organizada a respeito da fazenda através de aplicações a fim de facilitar o controle agrícola.

DISCUSSÕES FINAIS

Pode ser verificado a rápida evolução nos últimos dez anos em TI, com aplicação no campo agrícola, como alguns dos principais conceitos abordados nesse trabalho como: *Smart Farming*, Internet das Coisas, *Big Data* e Computação em Nuvem, que fundamentam o *Farming 4.0* e que juntos, apresentam uma grande importância para a produção atual do agronegócio. Vários benefícios podem ser verificados, como: controle de todos os setores e processos da fazenda, obtenção de informações precisas em tempo real, aumento da produtividade e redução de custos. Cada vez mais, a agricultura está inserida no desafio do aumento da quantidade e qualidade de produção agrícola, uma vez que a demanda de alimentos está em contínuo crescimento, contrastando com a redução de produtores rurais. Não há dúvidas que o *Farming 4.0* é a tendência futura mais próxima para a criação de soluções agrícolas e que poderá tornar-se uma revolução na maneira como a gestão da fazenda é realizada.

REFERÊNCIAS

CLASEN, M. Farming 4.0 und andere anwendungen des internet der dinge. In: *GIL Jahrestagung*. [S.l.: s.n.], 2016. p. 33–36.

GEIPEL, J. et al. A sensor web-enabled infrastructure for precision farming. *ISPRS International Journal of Geo-Information*, Multidisciplinary Digital Publishing Institute, v. 4, n. 1, p. 385–399, 2015.

HASHEM, I. A. T. et al. The rise of "big data" on cloud computing: Review and open research issues. *Information Systems*, Elsevier, v. 47, p. 98–115, 2015.

INAMDAR, S.; ROY, S. Internet of things: Architecture, security and applications. *Transactions on Engineering and Applied Sciences*, v. 1, n. 1, p. 9–19, 2017.

JAGANNATHAN, S.; PRIYATHARSHINI, R. et al. Smart farming system using sensors for agricultural task automation. In: IEEE. *Technological Innovation in ICT for Agriculture and Rural Development (TIAR), 2015 IEEE*. [S.l.], 2015. p. 49–53.

JAYARAMAN, P. P. et al. Internet of things platform for smart farming: Experiences and lessons learnt. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 16, n. 11, p. 1884, 2016.

KRINTZ, C. et al. Smartfarm: Improving agriculture sustainability using modern information technology. In: *KDD Workshop on Data Science for Food, Energy, and Water*. [S.l.: s.n.], 2016.

KUMAR, N. Smart farming system for indian farmers using arduino based technology. 2017.

LAVANYA, M. *WSN based advanced agricultural vehicle operated using smart phone - agribot*. Tese (Doutorado) — Visvesvaraya Technological University, 2016.

LIAGHAT, S.; BALASUNDRAM, S. K. et al. A review: The role of remote sensing in precision agriculture. *American journal of agricultural and biological sciences*, Science Publications, v. 5, n. 1, p. 50–55, 2010.

NORMAN, D. The farming systems approach: A historical perspective. In: *Presentation held at the 17th Symposium of the International Farming Systems Association in Lake Buena Vista, Florida, USA*. [S.l.: s.n.], 2002. p. 17–20.

PATIL, A. et al. Smart farming using arduino and data mining. In: IEEE. *Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on*. [S.l.], 2016. p. 1913–1917.

ROHAM, V. S. et al. Smart farm using wireless sensor network. Citeseer, 2015.

RYU, M. et al. Design and implementation of a connected farm for smart farming system. In: IEEE. *SENSORS, 2015 IEEE*. [S.l.], 2015. p. 1–4.

SUKHADEVE, V.; ROY, S. Advance agro farm design with smart farming, irrigation and rain water harvesting using internet of things. *International Journal of Advanced Engineering and Management*, v. 1, n. 1, p. 33–45, 2016.

VIANI, F. Experimental validation of a wireless system for the irrigation management in smart farming applications. *Microwave and Optical Technology Letters*, Wiley Online Library, v. 58, n. 9, p. 2186–2189, 2016.

WU, X. et al. Data mining with big data. *ieee transactions on knowledge and data engineering*, IEEE, v. 26, n. 1, p. 97–107, 2014.

Agência Embrapa de Informação Tecnológica (Ageitec): avaliação de seus impactos econômicos

Junia Rodrigues de Alencar[1]; Ana Paula da Silva Dias[2]; Fernando Attique Maximo[3]; Miguel Sardenha Francisco[4]

[1]Economista, Pesquisadora da Embrapa Informática Agropecuária, Campinas – SP, junia.alencar@embrapa.br
[2]Gestora da Informação e do Conhecimento, Analista da Embrapa Informação Tecnológica, Brasília – DF, anapaula.dias@embrapa.br
[3]Matemático, Pesquisador da Embrapa Informática Agropecuária, Campinas – SP, fernando.maximo@embrapa.br
[4]Graduando em Gestão do Agronegócio, Bolsista da Embrapa Informática Agropecuária, Campinas - SP miguel_sf_@hotmail.com

RESUMO

A agricultura é um setor da economia que abrange um grande escopo de atividades que têm como base o cultivo da terra, da criação de animais, do beneficiamento, da distribuição e comercialização, gerando assim diversidade de informações. A Agência Embrapa de Informação Tecnológica (Ageitec) foi criada com o objetivo de reunir e disponibilizar esse conjunto de informações detalhadas sobre o processo produtivo agropecuário, em endereço eletrônico, por meio de árvores do conhecimento (AC), para um público diversificado, com grande impacto na vida de milhares de pessoas. No entanto, o desenvolvimento e a manutenção da Ageitec envolveram recursos que precisam ser computados e analisados para mostrar à sociedade a eficiência do uso dos recursos públicos oriundos de seus impostos. Sendo assim, objetivando conhecer os impactos econômicos, foram analisados os benefícios gerados pela redução de custos, do incremento de produtividade na agricultura e na mão de obra pela adoção das informações agropecuárias inseridas na Ageitec. A metodologia de avaliação na dimensão econômica é baseada no método do excedente econômico, que busca analisar e quantificar os impactos gerados pela tecnologia para a sociedade brasileira. O período de análise foi de 2000 a 2016, que resultou, economicamente, uma Taxa Interna de Retorno (TIR) de 17,8%, Relação Beneficio/Custo (B/C) de 1,98 e Valor Presente Líquido (VPL) de R$ 6.743.382,10, considerando a taxa de juros de 6% ao ano.

PALAVRAS-CHAVE: Informação tecnológica, Árvore do conhecimento da agricultura, Organização da informação, Repositório, Transferência de tecnologia.

ABSTRACT

Agriculture is a sector of the economy that encompasses a wide range of activities based on land cultivation, animal breeding, processing, distribution and marketing, thus generating a great diversity of information. The Embrapa Agency for Technological Information (Ageitec) was created to gather and make available this set of detailed information on the agricultural production process, in electronic address, through knowledge trees (AC), to a diverse public, with large Impact on the lives of thousands of people. However, the development and mainte-

nance of Ageitec involved resources that need to be computed and analyzed to show to the society the efficient use of public resources. Thus, in order to know the economic impacts, we analyzed the benefits generated by the reduction of costs, the increase of productivity in agriculture and labor by the adoption of agricultural information inserted in Ageitec. The methodology of evaluation in the economic dimension is based on the method of economic surplus, which seeks to analyze and quantify the impacts generated by technology for Brazilian society. The period of analysis was from 2000 to 2016, which resulted in an internal rate of return (IRR) of 17.8%, a profit / cost ratio of 1.98 and a net present value (NPV) of R $ 6,743,382.10, considering the interest rate of 6% per year.

KEYWORDS: Technological information, Knowledge tree of agriculture, Information organization, Repository, Technology transfer.

INTRODUÇÃO

Ao longo da sua existência, a Empresa Brasileira de Pesquisa Agropecuária (Embrapa) tem gerado inúmeros conteúdos e tecnologias em seus diversos Centros de Pesquisa (CP). Porém, todo esse estoque de informações permanecia pulverizado, tornando difícil ao público obter as informações com eficácia, de forma sistematizada.

O acesso da sociedade às novas Tecnologias da Informação e Comunicação (TIC) e a exigência do público pelo acesso à informação em tempo real motivaram a Empresa a buscar novas condutas e procedimentos para a organização da informação, que resultem em efetiva transferência de tecnologia e conhecimento.

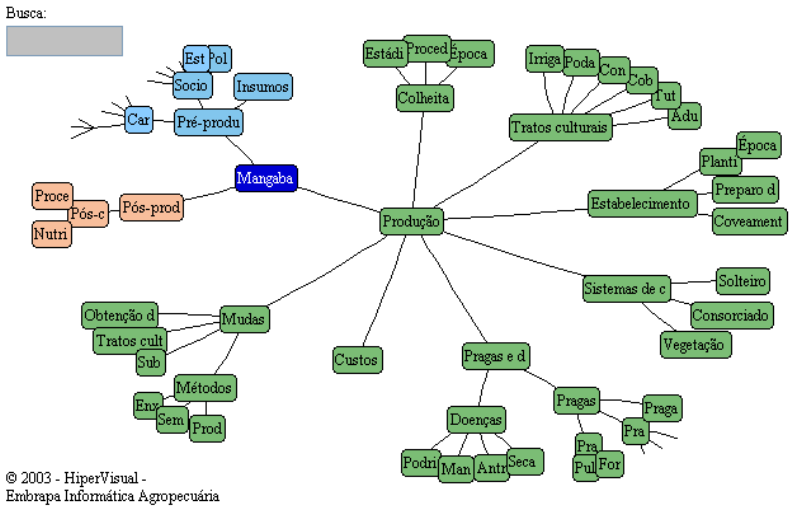
Nesse intuito, a Embrapa criou a Agência Embrapa de Informação Tecnológica (Ageitec), com o objetivo de armazenar, tratar e divulgar informações tecnológicas e conhecimentos gerados pela Empresa e outras instituições parceiras, organizadas por cultivo, criação, área temática e territorial. As informações estão organizadas em uma estrutura ramificada em forma de árvore hiperbólica, denominada Árvore do Conhecimento (AC), de acordo com a Figura 1, na qual o conhecimento é organizado de forma hierárquica, considerando todas as etapas da cadeia produtiva do negócio agrícola (ALENCAR et al., 2015; HANASHIRO; BAMBINI, 2008; SOUZA et al., 2006).

Esta ferramenta teve seu desenvolvimento iniciado no ano 2000, por meio de parceria entre a Embrapa Informação Tecnológica, a Embrapa Informática Agropecuária e a Embrapa Gado de Corte. Em 2002, foi feito um treinamento com todos os Centros de Pesquisa da Embrapa para uso da ferramenta e, em 2004, foi aprovado projeto com vistas ao desenvolvimento de AC pelos mesmos.

As AC permitem agregar vários elementos da informação de forma integrada, em linguagem acessível, como textos, vídeos, áudios, imagens, mapas e base de dados. Recursos esses que potencializam a eficácia na transmissão da informação. Com vista a assegurar a integração e troca de dados com sistemas similares e a interoperabilidade entre aplicações,

bem como a recuperação e o acesso ágil à informação desejada, a Ageitec utiliza o padrão internacional de classificação de informação por metadados *Dublin Core* (ALVES; SOUZA; SANTOS, 2008).

Figura 1 – Árvore do Conhecimento.



O lançamento das primeiras AC foi realizado em 2005, com a publicação na internet das Árvores do feijão, do agronegócio do leite e das espécies arbóreas da Amazônia. Até 2016, a Ageitec dispunha de 50 Árvores publicadas. As informações disponíveis na Ageitec atendem aos interesses de diversos públicos, quais sejam: docentes, pesquisadores, estudantes, extensionistas, produtores, processadores, distribuidores etc. A expectativa é de que todos eles empreguem a Ageitec como fonte de informações confiáveis, atualizadas, conforme as suas necessidades. A consulta é gratuita e pode ser feita pelo site www.embrapa.br/ageitec.

No entanto, para o desenvolvimento e a manutenção desta Ageitec, envolveram-se recursos humanos, financeiros, materiais, que precisam ser computados e analisados para mostrar à sociedade a eficiência do uso dos recursos públicos oriundos de seus impostos (ALENCAR et al., 2004; 2015). Portanto, o objetivo deste trabalho consiste em avaliar os impactos econômicos derivados dos gastos e benefícios com pesquisa, desenvolvimento e a transferência, gerados pela utilização das informações agropecuárias disponíveis na Ageitec, referente ao período de 2000 a 2016, considerando desde o início dos custos com o desenvolvimento do sistema em 2000 ao início dos benefícios em 2005.

MATERIAL E MÉTODOS

Fontes dos dados

Inicialmente, foi necessário entender a importância das informações da Ageitec para o público do site, seus acessos, a evolução das entradas de novas AC no Sistema Web. Em 2005 foram publicadas as três primeiras AC, com acesso pela web das informações contidas nestas. Esse

ano foi considerado como o início da adoção da tecnologia. Em 2007, passou para oito AC, 10 em 2008, 14 em 2010, 23 em 2011, 41 em 2012, 48 em 2013 e, em 2016, o Sistema Ageitec contava com 50 AC publicadas.

No intuito de melhorar a contagem de acessos destas informações pelo público, e entender qual área do sistema era mais buscada, foi utilizada a ferramenta denominada de Logômetro[1], pela equipe de desenvolvimento da Ageitec, a partir de 2013. A ferramenta disponibiliza informações sobre a localização dos usuários que acessam as AC, bem como o total geral de acessos, tanto no Brasil como no exterior. Estes acessos são considerados o principal indicador de medida para os cálculos dos impactos econômicos.

Como exemplo deste indicador, a Tabela 1 mostra o número de acessos para todas as AC, referente ao ano de 2016.

Tabela 1 - Total de acessos por Árvores do Conhecimento, no período de 01/01/2016 a 31/12/2016.*

CULTIVOS (33)		
Açaí (17.164)	Centeio (5.777)	Morango (7.046)
Ameixa (2.052)	Cevada (3.122)	Pera (2.441)
Amora (1.904)	Coco (38.780)	Pêssego (4.279)
Arroz (104.569)	Eucalipto (26.208)	Pimenta (50.177)
Banana (29.234)	Feijão (61.683)	Sisal (5.365)
Batata (17.940)	Feijão-caupi (28.475)	Soja (112.220)
Caju (27.111)	Gergelim (9.451)	Tomate (38.964)
Cana-de-açúcar (370.431)	Mamona (9.423)	Trigo (11.788)
Castanha-do-brasil (5.673)	Manga (55.808)	Triticale (3.297)
Cebola (43.709)	Mangaba (5.938)	Uva de Mesa (11.458)
Cenoura (19.357)	Milho (167.185)	Uva para Processamento (17.730)
CRIAÇÕES (4)		
Agronegócio do Leite (127.204)	Frango de Corte (102.054)	Ovinos de Corte (30.603)
Suínos (6.352)	-	-
TEMÁTICAS (11)		
Agricultura e Meio Ambiente (62.628)	Espécies Arbóreas Brasileiras (21.748)	Sistema Plantio Direto (26.944)
Agroenergia (84.242)	Espécies Arbóreas da Amazônia (4.241)	Solos Tropicais (139.149)
Bioma Caatinga (55.259)	Manejo Florestal (5.188)	Tecnologia de Alimentos (151.646)
Bioma Cerrado (87.799)	Reprodução Animal (2.669)	-
TERRITÓRIOS (2)		
Território Mata Sul Pernambucana (67.809)	Território Sisal (5.365)	-

*50 AC, sendo 33 sobre culturas, quatro sobre criação, 11 temáticas e duas territoriais. Em 2016 foram 2.296.659 acessos (visitas) via web.

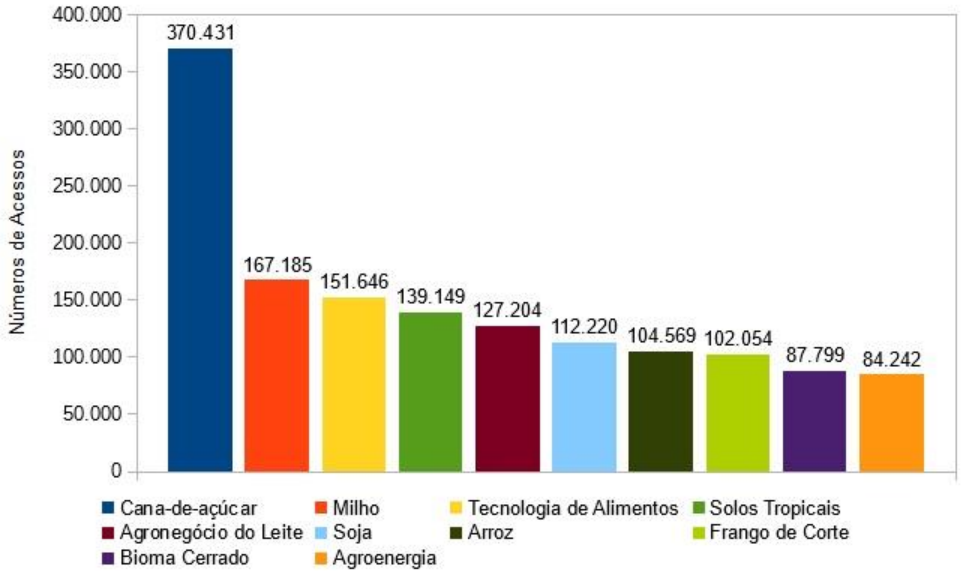
1 O Logômetro é um sistema de análise de visitas a websites que provê informações qualificadas sobre o tráfego e o acesso aos conteúdos e páginas disponibilizadas. Números de acessos, localização georreferenciada das visitas, palavras-chave utilizadas, páginas mais acessadas, tempo médio de visita entre outras informações específicas podem ser facilmente recuperadas (USER guides, 2017). Integrado ao website da Agência, permite descobrir e minerar dados das visitas, de maneira geral ou de uma AC específica.

Embora a Tabela 1 mostre apenas os acessos de 2016, para os cálculos dos impactos econômicos foi considerado este indicador de sucesso desde 2005.

Em 2012 existiam 41 AC; no ano de 2013 este número subiu para 48; já em 2014 passou para 50. Nos anos de 2015 e 2016 esse número manteve-se inalterado, dos quais, 33 são de culturas, quatro de criações, 11 temáticas e dois de territórios. Por outro lado, o número de acessos vem aumentando ao longo do tempo, considerando que em 2013 foram 1.155,290 acessos, 2014 passou para 1.378.200, em 2015 de 1.572.220 e, em 2016, o número de acessos foi de 2.296.659, ou seja, um aumento de 46,08% em relação ao ano anterior.

Para corroborar com estes dados, a Figura 2 mostra as 10 AC mais acessadas em 2016, o que equivale a 62,98% do total dos acessos, ou seja, 1.446.499. A título de informação, comparando aos anos anteriores, a AC com o maior número de acessos foi a da Cana-de-açúcar, com 280.274 em 2014, 300.484 em 2015 e, em 370.431 em 2016. Nota-se um aumento de 69.947 acessos, ou seja, 23,28% entre 2015 e 2016. A segunda mais acessada, em 2016, foi a AC do Milho, com 167.185 acessos. A terceira mais acessada foi a Tecnologia de Alimentos (151.646 acessos); a quarta a de Solos Tropicais (139.149 acessos); a quinta a AC Agronegócio do Leite (127.204 acessos); a sexta a de Soja (112.220 acessos), a sétima Arroz (104.569 acessos), a oitava Frango de Corte (102.054 acessos), a nona Bioma Cerrado (87.799 acessos) e a 10ª a AC Agroenergia com 84.242 acessos.

Figura 2 – Dez AC da Ageitec mais acessadas em 2016 (1.446.499 acessos – 62,98% do total).



Metodologia de avaliação dos impactos econômicos

A metodologia de avaliação na dimensão econômica é baseada no método do excedente econômico, incluindo as melhorias operacionais em função da experiência acumulada pela

Embrapa no processo de avaliação de tecnologias (AVILA et al., 2008). Essa avaliação econômica busca analisar e quantificar os impactos gerados pela tecnologia para a sociedade brasileira, sendo estimados os benefícios oriundos do incremento de produtividade, da redução de custos, expansão da produção em novas áreas e agregação de valor pelo uso e adoção das tecnologias geradas pela Empresa no negócio agrícola. No entanto, no caso da Ageitec foram utilizados os benefícios do incremento de produtividade e da redução de custos no empreendimento agrícola pelo uso das informações contidas nas AC.

Estimativa dos custos e análises de rentabilidade da Ageitec

Além dos benefícios, toda tecnologia para ser gerada e adotada acarreta custos. Os custos da Ageitec foram calculados usando a metodologia definida no mesmo manual da avaliação de impactos (AVILA et al., 2008) considerando desde o início do desenvolvimento da ferramenta, ou seja, o ano de 2000, que difere do ano de adoção. Para os custos com pessoal, ou seja, da mão de obra envolvida no desenvolvimento, manutenção e transferência da tecnologia Ageitec, foram levantados os dados referentes aos salários de empregados nos agrupamentos de pesquisadores, analistas e assistentes a partir do Sistema Integrado de Recursos Humanos (SIRH) da Empresa, considerando o tempo de dedicação (%) ao projeto.

O custeio de pesquisa foi calculado a partir de um amplo levantamento dos projetos de pesquisa voltados ao desenvolvimento da tecnologia. Toda a base inclui informações de projetos do Sistema Embrapa de Gestão (SEG) resgatando-se informações de orçamentos de planos de ação liderados por pesquisadores da Embrapa Informática Agropecuária.

Para o cálculo da depreciação de capital, foram considerados os custos de pessoal, pesquisa, administração e transferência de tecnologia, tomando por base a vida útil da tecnologia.

Gastos com infraestrutura foram atribuídos calculando um rateio dos custos de administração do centro de pesquisa. Por outro lado, os custos com transferência tecnológica foram aferidos por meio de gastos com viagens das equipes de pesquisa e transferência de tecnologia, com a elaboração de material de comunicação e divulgação, a exemplo de fôlderes, gastos com palestras e cursos, com indicação de horas de treinamento dos instrutores envolvidos na transferência da tecnologia. O período considerado para o cálculo dos custos da Ageitec foi desde o início do projeto, ou seja, de 2000 até 2016.

Estimativa dos benefícios e análises de rentabilidade da Ageitec

Para aferir os benefícios oriundos do uso e adoção das informações contidas em todas as AC da Ageitec, foram considerados os impactos no incremento de produtividade e na redução de

custos para o empreendimento agropecuário. Os cálculos dos benefícios econômicos foram realizados a partir de 2005, quando a tecnologia passou a ser adotada e acessada por via web. Duas metodologias de cálculo foram utilizadas, as quais, somadas, resultam no valor total anual dos benefícios em virtude do incremento de produtividade gerado pela Ageitec. Estima-se que o valor médio de cada AC seja de US$ 25.000,00, sendo que, para cada ano, multiplica-se esse valor à taxa do dólar ao ano referente e, por fim, se multiplica esse valor em moeda nacional pelo número de árvores disponíveis. O valor encontrado é referente ao item "economizado" pelo uso e adoção das informações contidas na Ageitec.

Ainda para o incremento de produtividade, foi levantado o valor médio por propriedade agrícola (a base deste cálculo é o mesmo feito para o Lactus – ferramenta de gestão da produção de gado leiteiro – 2008, disponibilizado pela Rede AgroLivre[2], acompanhado do valor de visitas médias para as árvores de maior relevância comercial, utilizando-se de estimativas conservadoras de 5% de proprietários individualizados e de 1,065% de redução de perdas verificadas com o auxílio das informações disponibilizadas para a tomada de decisões.

Com relação à redução de custos que o empreendedor agrícola obtém pelo uso das informações para diminuir os riscos na tomada de decisão, foi levado em conta o valor calculado para o item "economizado" somado ao número de acessos ao sistema, multiplicado o percentual estimado de acessos realizados por produtores que se utilizam do sistema para a tomada de decisão. Em seguida, multiplicou-se pelo valor de uma publicação importante, somado ao percentual estimado de professores que buscam material didático, economizando na aquisição de publicações, multiplicado pelo valor médio de uma publicação desse tipo. Com a soma desses valores encontrados se obtém o item redução de custo.

Após realizar os cálculos dos custos da geração da tecnologia e os benefícios resultantes de seu uso e adoção, calculou-se o VPL, a TIR e a relação B/C, usando as metodologias de Administração Financeira.

[2] Um dos softwares que auxiliam no processo de gestão da atividade leiteira é o Lactus, disponibilizado pelo Repositório da Rede AgroLivre. A prática da gerência racional da atividade leiteira traz mudanças significativas nos sistemas de produção de leite, pois poderá ser feito um controle mais efetivo da atividade, avaliando os índices de custo-benefício de todas as atividades desenvolvidas, no sistema de produção de leite. Estas facilidades possibilitam que se tenha uma avaliação econômica da atividade, baseada em dados reais, permitindo que as decisões produzam melhores resultados econômicos para o produtor. Durante o período compreendido entre 1/3/2004 e 7/2/2008, foram constatados 13.001 downloads deste software, ou seja, quase 3.250 downloads/ano. Em fevereiro de 2012, o número de downloads estava em 34.432 unidades, ou seja, crescente. Assim, para estimarmos o impacto devido a este aplicativo, tomaremos como base dados dos Sistemas de Produção da Embrapa (2008), referentes a 2001, onde o Valor Bruto da Produção Agropecuária foi de 91 bilhões de reais, sendo que o leite ocupou o valor de 6,7 bilhões de reais. E, quanto ao número de propriedades existentes, estima-se que haja em torno de um milhão e cem mil propriedades que exploram leite, ocupando diretamente 3,6 milhões de pessoas. Desses números, podemos chegar ao valor médio de R$ 6.090,91/propriedade, que também foi mantido como referência para os demais anos, dada a disponibilidade do valor bruto da produção da pecuária de leite para 2001 (ALENCAR et al., 2015).

RESULTADOS E DISCUSSÃO

De acordo com as informações apresentadas no item anterior, foram feitos todos os cálculos dos custos da geração, manutenção e transferência da Ageitec, desde a concepção do projeto em 2000. Em seguida, foram calculados os benefícios econômicos pelo uso e adoção das informações contidas nas 50 AC considerando tanto o aumento de produtividade como da redução de gastos que os usuários teriam por este uso, a partir de 2005, tendo como seu principal indicador os acessos à web. Finalmente, foi calculada a rentabilidade dos investimentos por meio da TIR, o VPL e B/C ao longo do tempo.

Custos da geração, manutenção e transferência da tecnologia

A Tabela 2 indica os custos de pessoal, pesquisa/desenvolvimento, depreciação do capital, administrativos, transferência da tecnologia da Ageitec, no período de 2000 a 2016.

Tabela 2 - Estimativas de custos de geração e manutenção da tecnologia Ageitec.

Ano	Custo de pessoal (R$ 1.000,00)	Custeio de Pesquisa (R$ 1.000,00)	Depreciação de Capital (R$ 1.000,00)	Custos de Administração (R$ 1.000,00)	Custos de Transferência Tecnológica (R$ 1.000,00)	Total (R$ 1.000,00)
2000	325, 04	399,46	41,79	122,11	0,00	888,40
2001	473,31	105,31	57,65	167,04	0,00	803,34
2002	536,81	58,51	52,91	156,26	0,17	804,66
2003	367,68	0,00	33,24	75,86	0,17	476,94
2004	418,59	89,33	30,10	85,47	0,47	623,97
2005	324,24	120,10	64,29	69,28	6,56	584,47
2006	257,02	95,01	24,05	74,67	4,81	455,56
2007	403,36	68,09	39,33	122,34	3,25	636,38
2008	391,07	31,33	33,06	113,02	18,30	586,78
2009	370,68	22,22	30,06	113,67	5,91	542,53
2010	390,18	21,04	28,42	108,38	3,55	551,56
2011	336,89	25,71	36,82	44,62	5,63	449,67
2012	340,75	0,00	67,66	37,20	2,40	448,00
2013	439,43	27,10	71,30	39,20	8,93	585,96
2014	534,96	28,12	74,00	40,69	12,27	690,04
2015	820,57	0,00	74,89	41,18	38,88	975,53
2016	946,64	0,00	74.91	39,54	5,55	1.063,65

Como citado anteriormente, os custos tiveram seu início em 2000, momento da aprovação do projeto para o desenvolvimento da plataforma web Ageitec, sendo o custo com a pesquisa o mais elevado, não diferente de outras atividades semelhantes. É possível destacar que o maior custo gerado pela Ageitec se refere ao custo de pessoal, crescendo anualmente, chegando, em 2016, no valor de R$ 946.642,73, ou seja, 88,99% do total. Isso se justifica porque a geração de conhecimento necessita de recursos humanos altamente qualificados para gerar, manter e inovar tecnologias, produtos e serviços. Por outro lado, os gastos de pesquisa

têm diminuído, visto que, nos anos de 2015 e 2016, não houve dotação orçamentária para a manutenção e novos aprimoramentos no Sistema Web, inclusive prejudicando a entrada de novas AC.

Considerando ainda 2016, os gastos com a depreciação do capital foram de 7,04%, uma vez que os materiais de informática necessários para a manutenção do site (servidores) têm uma desvalorização a cada ano. Os custos de administração, que são decorrentes de gestão do sistema e manutenção do endereço web, vêm se mantendo quase inalterados. Por outro lado, ocorreu uma queda considerável nos gastos com transferência da tecnologia, prejudicando o acompanhamento junto aos usuários e novas capacitações. Porém, o aumento da divulgação da Ageitec em feiras agropecuárias, apresentação de palestras, divulgação na mídia falada e escrita fez com que os acessos, citados na fonte de dados, aumentassem entre 2015 e 2016 em 46,08%, impactando o retorno do investimento.

Benefícios gerados: incremento de produtividade e redução de custos

A Tabela 3 mostra os benefícios gerados pela Ageitec, a partir de 2005, quando houve a implantação de três AC e se iniciou a mensuração de seus acessos por meio da web. O período de análise foi de 2015 a 2016, levando em consideração a taxa do dólar, o número de árvores, o quanto foi economizado polo uso das informações, considerando como parâmetro a AC Agronegócio do Leite e suas relações com o incremento de produtividade, bem como a redução de custos pelo uso e adoção das informações.

Tabela 3 - Benefícios econômicos da Ageitec: redução de custos e incremento de produtividade.

Ano	Taxa do Dólar (R$)	Nº de Árvores	Economizado (R$ 1.000,00)	Redução de Custos (R$ 1.000,00)	Incremento de Produtividade (R$ 1.000,00)	Total (R$ 1.000,00)
2005	2,44	3	182,64	510,10	90,86	600,91
2006	2,18	3	163,21	623,91	113,52	737,43
2007	1,95	8	389,58	846,08	136,22	982,30
2008	1,84	10	458,75	952,98	155,46	1.108,44
2009	2,00	10	499,40	1.093,77	196,13	1.289,87
2010	1,76	14	616,00	1.476,16	217,10	1.693,26
2011	1,67	23	960,25	1.911,05	209,13	2.120,18
2012	1,91	41	1.957,75	2.680,16	217,25	2.897,41
2013	2,16	48	2.589,60	3.210,35	197,26	3.407,62
2014	2,43	50	3.036,25	3.792,20	194,13	3.986,33
2015	3,33	50	4.164,50	4.815,23	205,44	5.020,66
2016	3,49	50	4.362,50	5.831,39	288,80	6.120,18

Na Tabela 3, observa-se que o indicador de maior benefício econômico foi o de redução de custo, que variou, no período, de R$ 510.098,63 para R$ 5.531.378,77, uma vez que os

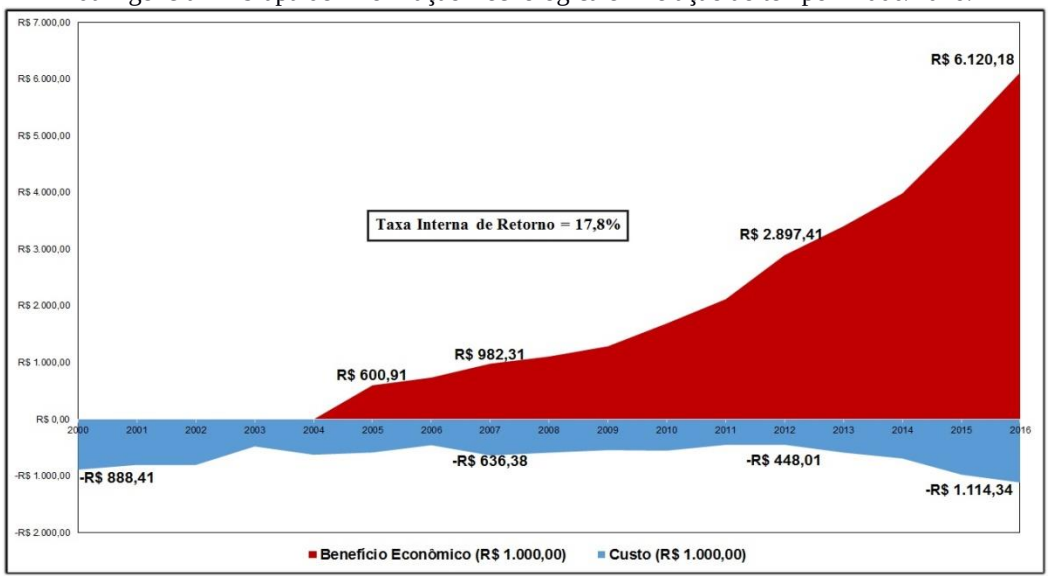
tomadores de decisão munidos de informações confiáveis sobre sistemas de produção tanto para cultivo, como para a criação, áreas temáticas, como biomas cerrado e caatinga, solos, meio ambiente, manejo florestal, tecnologia de alimentos e outros, contidas nas AC, proporcionam menores riscos e maior eficiência em decidir o quê, onde, porquê e como produzir. Em seguida, vem o indicador "economizado", que parte do pressuposto de que o processo de disponibilizar, e mesmo buscar informações sobre a pesquisa agrícola, por parte dos agentes econômicos, é facilitado. Por outro lado, o incremento de produtividade se explica pelo fato de ter todas essas informações contidas em um único endereço eletrônico, facilitando e agilizando a busca desses conhecimentos, fazendo com que o tempo de trabalho seja reduzido, disponibilizando horário livre para pesquisadores, estudantes, produtores rurais, extensionistas e outros desempenharem outras atividades. O benefício econômico, somente em 2016, foi de R$ 6.120.181,00.

Análise de rentabilidade dos investimentos com a Ageitec

Foram considerados os custos e benefícios contidos nas Tabelas 2 e 3 entre o período de 2000 a 2016. Desta forma, calculou-se a TIR, que foi de 17,8%, superior as praticadas no mercado. A relação B/C, foi de 1,98, significando que para cada R$1,00 investido na tecnologia foi obtido um retorno de R$ 1,98. Por fim, o VPL resultou em R$ 6.743.382,10, mostrando a viabilidade do investimento. Todos os cálculos foram realizados tendo como base a taxa de juros de 6%, quase semelhante à taxa de inflação de 2016, que, segundo o índice IPCA-15/IBGE (IBGE, 2017), foi de 6,29%.

A Figura 3 mostra, com mais clareza, a relação de benefícios e custos do resultado do uso e adoção do Serviço Web Ageitec, onde se pode observar ao longo do tempo a evolução dos benefícios superiores aos custos.

Figura 3 - Benefícios líquidos da pesquisa, desenvolvimento, manutenção e transferência da Agência Embrapa de Informação Tecnológica em relação ao tempo - 2000/2016.



CONCLUSÕES

Com estes resultados econômicos altamente favoráveis, prova-se, mais uma vez, a importância de se investir em pesquisa. Deve-se ressaltar que há necessidade de se fazer melhoria de processo na ferramenta e nos conteúdos da Ageitec, os quais aumentariam seus benefícios. Além disso, espera-se que, ao longo dos anos, com a expansão crescente das tecnologias da informação e comunicação, no contexto da agricultura digital, aumentem os acessos na Ageitec, de modo que auxilie, mais e mais, a tomada de decisão no campo e contribua para as políticas públicas de acesso à informação no setor agrário.

REFERÊNCIAS

ALENCAR, J. R. de; DIAS, A. P. da S.; MAXIMO, F. A.; FRANCISCO, M. S.; **Relatório de Avaliação dos Impactos das Tecnologias Geradas pela Embrapa – Agência Embrapa de Informação Tecnológica - Ageitec ,** 2016. 68 p. Disponível <http://bs.sede.embrapa.br/2015/relatorios/informaticaagropecuaria_2015_agencia.pdf>. Acesso em: 3 jan. 2017

ALENCAR, J. R. de; HERRUZO MARTINEZ, A. C.; HOEFLICH, V. A.; OLIVEIRA, E. B. de; Impactos econômicos da pesquisa e desenvolvimento de um sistema informatizado para manejo florestal de Pinus spp. **Cadernos de Ciência & Tecnologia,** Brasília, DF, v. 21, n. 3, p. 425-446, set./dez. 2004.

ALVES, M. das D. R.; SOUZA, M. I. F.; SANTOS, A. D. dos. **Manual de representação descritiva de recursos eletrônicos no Sistema Agência**. Campinas: Embrapa Informática Agropecuária, 2008. 139 p. (Embrapa Informática Agropecuária. Documentos, 81).

AVILA, A. F. D., RODRIGUES, G. S., VEDOVOTO, G. L. **Avaliação dos impactos de tecnologias geradas pela Embrapa**: metodologia de referência. Brasília, DF: Embrapa Informação Tecnológica, 2008. 189 p.

HANASHIRO, M. M.; BAMBINI, M. D. **Relatório de avaliação dos impactos das tecnologias geradas pela Embrapa – Agência de Informação Embrapa (ano-base 2007)**. Campinas: Embrapa Informática Agropecuária. 2008. Não publicado.

IBGE. **IPCA**: séries históricas. Disponível em: http://www.ibge.gov.br/home/estatistica/indicadores/precos/inpc_ipca/defaultseriesHist.shtm. Acesso em: 3 jan. 2017.

SOUZA, M. I. F.; SANTOS, A. D. dos; MOURA, M. F.; ALVES, M. das D. R. Agência de informação Embrapa: uma aplicação para a organização da informação e gestão do conhecimento. In: SIMPÓSIO BRASILEIRO DE ENGENHARIA DE SOFTWARE, 20.; SIMPÓSIO BRASILEIRO DE BANCO DE DADOS, 21.; WORKSHOP DE BIBLIOTECAS DIGITAIS, 2., 2006, Florianópolis. **Anais**... Florianópolis: SBC, 2006. p. 51-56.

USER guides: learn how to install, maintain and configure Piwik and how to effectively analyze your data. 2017**.** Disponível em: <http://piwik.org/docs/user-guide/>. Acesso em: 5 maio 2017.

Estimativa do volume de ameixa ‘Reubennel’ através da técnica da silhueta

Daniele Candido da Costa[1], Juliana Sanches[2], Antônio C. L. L[3], Edson D`Ávila[4], Patrícia Cia[3]

[1] Instituto Agronômico (IAC), Campinas, São Paulo, Brasil, danielecandido90@gmail.com

[2] Centro de Ecofisiologia e Biofísica, Instituto Agronômico, Campinas, São Paulo, Brasil, jsanches@iac.sp.gov.br

[3] Centro de Engenharia e Automação, Jundiaí, São Paulo, Brasil, lino@iac.sp.gov.br; pcia@iac.sp.gov.br

[4] Faculdade de Engenharia Agrícola, UNICAMP, Campinas, São Paulo, Brasil, davila.dre@gmail.com

RESUMO

Este trabalho teve como objetivo estimar o volume de ameixa ‘Reubennel’, a partir da filmagem de sua silhueta, com o auxílio de rotinas para o *software* ImageJ. As imagens foram obtidas através de uma câmera tipo webcam, iluminação de fundo para gerar a silhueta e um sistema para rotacionar a fruta. Para a calibração do sistema foram utilizadas esferas de diâmetros conhecidos. Desenvolveram-se rotinas para obtenção das imagens quadro a quadro do filme obtido, transformação de cada quadro em imagens binárias (sendo preto referente à silhueta e branco o fundo) e geração de uma matriz de coordenadas do contorno da fruta para cada uma das imagens do filme. Uma vez calibrado o sistema, a rotina estimou o volume de 50 ameixas e comparou-se com o volume determinado. A metodologia proposta mostrou-se capaz de estimar o volume de pêssegos com alta precisão e exatidão.

PALAVRAS-CHAVE: Processamento de imagens digitais, ImageJ, imagens tridimensionais

ABSTRACT

The present work had as an objective to estimate the volume of a ‘Reubennel’ plum using the captured silhouette. The images were obtained through a webcam camera, backlight illumina-

tion in order to obtain a silhouette and a mechanism to turn the fruit. Spheres with a known diameter were used in order to calibrate the system. Routines were developed to generate frame-by-frame images from the captured video, transformation of every frame in binary images (black for silhouette and white for background) and the generation of a coordinated matrix of the fruit silhouette for every frame. Once the system has been calibrated, the routine has estimated the volume of 50 plums and compared those values with the volume obtained through the calculated volume. The present methodology has shown capable of estimating the volume of plums with high precision and exactness.
KEYWORDS: Digital image processing, ImageJ, tridimensional images

INTRODUÇÃO

O uso de imagem digital permite maior precisão e eficiência na classificação, importante fator na determinação do preço dos frutos, sendo um processo essencial para as indústrias de processamento de alimentos (LEE et al., 2003). É uma alternativa tanto para o setor produtivo quanto para pesquisa. O sistema pode funcionar com *softwares* livres, o que garante que pode ser usado, copiado, estudado, modificado e redistribuído, como o ImageJ.

Nesse processo, as imagens bidimensionais (2D) são frequentemente utilizadas, porém, há uma limitação nas imagens em 2D pela falta de informações sobre as dimensões de altura e profundidade. Sendo assim, os modelos tridimensionais têm um papel importante na estimativa precisa do volume e da area superfícial (MOREDA et al., 2009). Os sistemas de inspeção tridimensional permitem o cálculo do volume dos frutos por meio da técnica de silhueta, como o proposto por LEE et. al (2006), que obtém contornos topográfico de objetos com formatos irregulares, como os frutos.

Dentro deste contexto, este trabalho teve por objetivo estimar o volume de ameixa ‘Reubennel’ utilizando imagens tridimensionais através da técnica da silhueta.

MATERIAL E MÉTODOS

O experimento foi conduzido no Laboratório de Imagens Aplicadas à Agricultura do Centro de Engenharia e Automação / IAC, em Jundiaí, São Paulo.

Volume dos frutos

Foram utilizados 50 frutos de ameixa ‘Reubennel’, obtidos com produtores da região de Jundiai, São Paulo. O arranjo experimental foi constituído de uma câmera tipo webcam, acoplada a um computador, onde cada fruta foi colocada sobre uma plataforma

acionada por um motor elétrico a velocidade constante. Para obter as silhuetas, os frutos foram colocados em frente a um negatoscópio com iluminação de fundo (conforme Figura 1), rotacionadas em meia volta obtendo-se um vídeo, de resolução de 640x480 pixels com as silhuetas obtidas a cada 4,8° de rotação. O negatoscópio é constituído por uma caixa com lâmpadas e uma tampa feita de um material semitransparente.

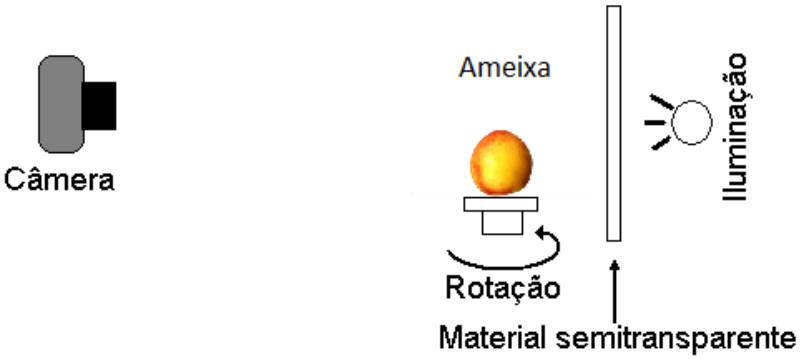


Figura 1. Arranjo experimental para captura das imagens pela técnica da silhueta.

Após as filmagens, os vídeos foram divididos em 50 quadros, tratados e os volumes calculados no *software* livre, linguagem Java, ImageJ. Para tal, foi desenvolvida uma rotina computacional que abre cada imagem do vídeo (Figura 2A), e a transforma em uma imagem binária (Figura 2B) o que permite o reconhecimento do contorno (silhueta) e determinação das coordenadas do fruto.



Figura 2. Imagem de um dos quadros obtidos a partir da divisão dos vídeos (A); Imagem binaria (B). Imagens de ameixa 'Reubennel' obtidas a partir da rotina computacional desenvolvida no *software* ImageJ.

A rotina desenvolvida a fim de determinar o volume pelo *software* ImageJ calcula a área média de todas as silhuetas da fruta. A seguir calcula as dimensões de um quadrado com a mesma área (Equações 1 e 2). O volume da fruta é igual ao volume de um cilindro de revolução (Figura 3).

$$A = L^2 \quad (1);$$

Onde *A* é a área média das silhuetas e *L* são os lados do quadrado de mesma área (cm^2).

$$V = \pi * r^2 * h \quad (2);$$

Onde V é o volume do cilindro de revolução, e, portanto, da fruta em estudo, *r* é o raio do cilindro e é igual a *L/2*. E *h* é a altura do quadrado, portanto igual a *L* (cm^3).

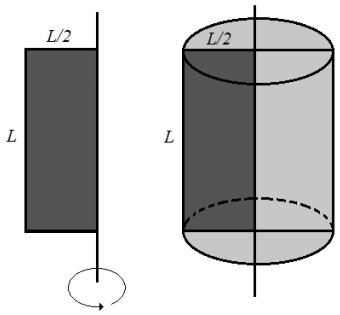


Figura 3. Cilindro de revolução utilizado como base para os cálculos de volume dos frutos de ameixa 'Reubennel' pelo *software* ImageJ.

O volume do fruto também foi calculado a partir dos dados estimados da densidade da ameixa 'Reubennel'. Dez frutos foram pesados em balança semianalítica e determinada a massa em gramas, submersos em água e o volume extravasado do recipiente foi considerado como o volume final do fruto. A densidade foi calculada pela Equação 4 e expressa em g/cm^3.

$$d = \frac{m}{v} \quad (3);$$

Onde d é a densidade do fruto, V é o Volume, m é a massa

Calibração do sistema

Com o objetivo de testar a metodologia do volume gerado pelas silhuetas foram utilizadas cinco esferas simétricas de borracha, de diâmetros conhecidos (4,9 cm, 5,6 cm, 6,3 cm, 6,9 cm e 7,5 cm, respectivamente), cujos volumes foram calculados pela equação de volume (Equação 4), pelo método de deslocamento de água e pelo *software* ImageJ.

$$V = \frac{4}{3}\ \pi\ r^3$$

(4);

Onde V é volume e r é o raio da esfera (cm^3)

O deslocamento de água consistiu em pesar as esferas em balança semianalítica para determinação de sua massa, em gramas, submergi-las em água e considerar o volume extravasado como o volume final.

Os volumes obtidos pela técnica da silhueta realizado através do *software* imageJ e pela equação de volume pelo deslocamento de água, tanto dos frutos como das esferas, foram comparados através do coeficiente de determinação R^2.

RESULTADOS E DISCUSSÃO

A Tabela 1 apresenta os resultados de volume das esferas calculado pela equação 4, os volumes determinados pelas silhuetas das esferas por meio do *software* ImageJ e o volume calculado das esferas a partir do deslocamento de água. Os três métodos de cálculo do volume das esferas apresentaram resultados semelhantes, sendo que a esfera de 4,9 cm foi a que obteve a maior diferença (15,9 cm^3 entre os dados de volume pelo ImageJ e água). O cálculo do volume por meio do deslocamento de água se mostrou interessante para estabelecer e comparar os resultados para objetos assimétricos e de densidade desconhecida.

Tabela 1. Volumes obtidos através da equação de volume, pelo *software* ImageJ e pelo método do deslocamento de água.

Esferas Diâmetro (cm)	Volume equação (cm^3)	Volume ImageJ (cm^3)	Volume água (cm^3)
4,9	62,3	56,1	72,0
5,6	93,7	86,8	88,6
6,3	129,9	123,9	125,4
6,9	176,3	170,3	174,3
7,5	219,9	222,2	221,2

Foram calculados também os coeficientes de determinação (R^2) e correlação de Pearson (r) dos volumes comparados para garantir a similaridade dos métodos (Figura 4). Os resultados apresentados indicaram correlações muito altas (R^2=0,99 e r=0,99), o que evidencia a possibilidade de utilizar os métodos propostos para calcular volumes. É importante ressaltar que a técnica da silhueta desenvolvida através de ImageJ permitiu determinar com grande acurácia o volume das esferas, que são simétricas.

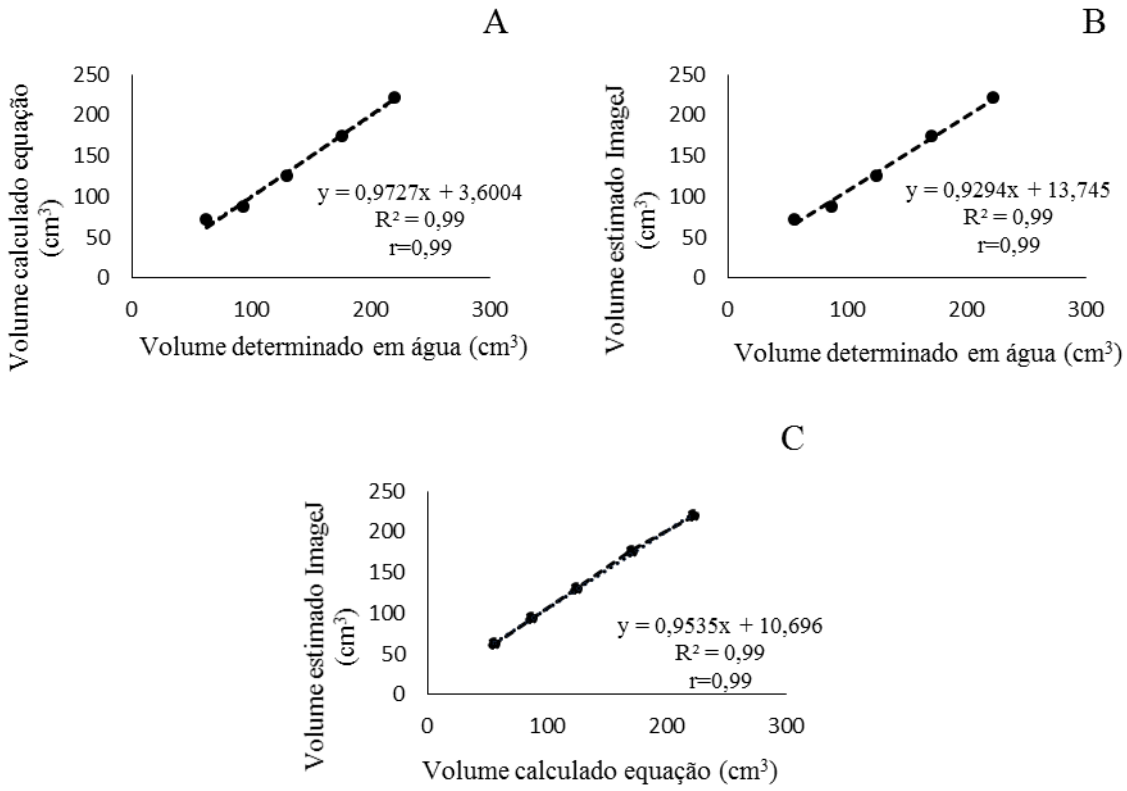


Figura 4. Correlação entre o volume calculado pela equação 4* e o volume determinado pelo deslocamento da água (A); Correlação entre o volume estimado pela técnica da silhueta através do *software* ImageJ e determinado pelo deslocamento da água (B); Correlação entre o volume estimado pela técnica da silhueta através do *software* ImageJ e pela equação 4 (C). *Equação 3: $V = \frac{4}{3}\ \pi\ r^3$, Onde V é volume e r é o raio da esfera.

.

Supondo-se que a técnica empregada para se calcular os volumes das esferas possa ser utilizado para objetos assimétricos, foram calculados os volumes da ameixa 'Reubennel'. O volume das frutas calculado pela equação 2 e pelo *software* ImageJ apresentaram alta correlação (R^2=0,96 e r=0,98) (Figura 5).

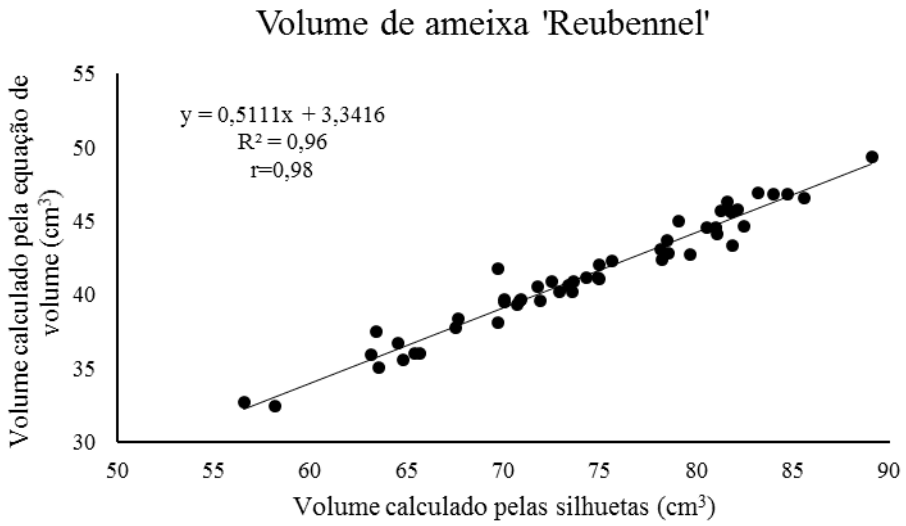


Figura 5. Correlação entre o volume de ameixas ‘Reubennel’ calculado pela equação de volume (V= m/d) e pela técnica da silhueta através do *software* ImageJ.

CONCLUSÕES

A técnica da silhueta através do software imageJ, mostrou-se capaz de estimar o volume ameixas ‘Reubennel’ com alta precisão e exatidão.

AGRADECIMENTOS

À FAPESP (Proc. n. 2013/25728-7), ao CNPQ e a CAPES pelo auxílio financeiro.

REFERÊNCIAS

LEE, D. et. al. Fast surface approximation for volume and surface area measurements using distance transform. **Optical engineering**, v. 42, n. 10, p. 2947-2955, 2003.

LEE, D. et. al. Area and volume measurements of objects with irregular shapes using multiple silhouettes. **Optical Engineering**, v. 45, n. 2, p. 027202-027202-11, 2006.

MOREDA, G.P. et al. Non-destructive technologies for fruit and vegetable size determination–a review. **Journal of Food Engineering**, v. 92, n. 2, p. 119-136, 2009.



A prospective study on the application of Data Science in agriculture

Kleber Xavier Sampaio de Souza, Sônia Ternes, Stanley Robson de Medeiros Oliveira, Maria Fernanda Moura, Luís Gustavo Barioni, Roberto Hiroshi Higa, Maria do Carmo Ramos Fasiaben

Embrapa Informática Agropecuária, Campinas, SP, Brasil,
kleber.sampaio@embrapa.br, sonia.ternes@embrapa.br,
stanley.oliveira@embrapa.br, maria-fernanda.moura@embrapa.br,
luis.barioni@embrapa.br, roberto.higa@embrapa.br, maria.fasiaben@embrapa.br

RESUMO

A quantidade e diversidade de dados disponíveis têm o potencial de causar profundas transformações na maneira que se realiza pesquisa e se propõe inovações na agricultura. Na chamada era do Petabyte, caracterizada pela ubiquidade de sensores e computadores, armazenamento quase infinito, computação em nuvem, robótica e IoT, a demanda e as oportunidades para aplicação da computação científica são extraordinárias, tanto na extração do conhecimento quanto na compreensão dos mecanismos associados a sistemas complexos. Este artigo apresenta um estudo prospectivo com base no estado da arte e enumera algumas áreas nas quais a aplicação da Ciência de Dados resultaria em grande benefício para pesquisadores, agricultores e agentes públicos.

PALAVRAS-CHAVE: Computação científica, Aprendizado de máquina, Modelagem e simulação, Agricultura, Redes de sensores.

ABSTRACT

The amount and diversity of data available have the potential to cause profound transformations in the way how research is conducted and innovations in agriculture are proposed. In the so-called Petabyte era, characterized by the ubiquity of sensors and computers, almost infinite storage, cloud computing, robotics and IoT, the demand and opportunities for scientific computing applications are extraordinary, both in extracting knowledge and in understanding mechanisms associated with Complex systems. This paper

conducts state-of-the-art research and enumerates some areas in which the application of data science would be of great benefit to researchers, farmers, and public agents.

KEYWORDS: Scientific computing, Machine learning, Simulation and modeling, Agriculture, Sensor networks.

INTRODUCTION

The enormous amount of data generated every day by humanity, which in accordance to IBM[1] amounts 2.5 exabytes (2.5 quintillion bytes) of data has caused a profound impact in research, development and innovation in agriculture. This occurs especially because agriculture deals with data ranging from nanoscale (DNA, RNA, proteins) to macroscale (remote sensing data from satellites, climate data, , agricultural census and global production databases), and have to combine them with data collected by sensor networks, harvesters, actuators etc.

In terms of complexity, the amount of data to be analyzed (terabytes or petabytes), its dimensionality, heterogeneity and dispersion in cloud storages characterizes what is called Big Data. This study performs literature review and proposes a framework to structure the actions of research, development and innovation of Data Science in agriculture.

STATE OF THE ART SURVEY

Data Science

The abundance of data is causing an enormous impact in the way how science is performed, not only in agriculture, but in general. In 2008, Wired Magazine editor, Chris Anderson (ANDERSON, 2008) published a paper on how the abundance of data renders the scientific method obsolete. The rationale behind this affirmation is that theories are inherently incomplete. The Newtonian model of the universe did not explain all questions and was supplanted by Einsteinian model, which in turn does not answer questions of the subatomic world.

For example, Google translator does not understand language semantics, nor Watson – IBM's computing system capable of answering questions in natural language – understands the meaning of the questions of the television show Jeopardy. Everything is mathematically calculated by machine learning algorithms that do not understand the nature of the data they are dealing with. Despite this, translations only get better and Watson defeated human players in Jeopardy.

[1] https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

Therefore, Anderson claims that instead of constructing (incomplete) theories supported by mental exercise over previously established theories, using data simply to validate them, one can analyze mathematically and computationally the data and contextualize what was learned a posteriori. This data intensive way of doing science is what has been called Data Science (DHAR, 2013; HEY et al., 2009; BELL et al., 2009). There are, of course, conflicting opinions: Marzzocchi (2015) analyzes the proposition from epistemological point of view and argues that "even the computational approach involves the testing of certain assumptions", but it is undeniable that computational analysis of massive data plays a complementary role in scientific discovery.

A report to the president of the United States (PITAC, 2005) advocates that Computational Science "has become the third pillar" of scientific research, along with theory and experimentation. This report highlights examples linked to agriculture, such as the use of biomolecular simulations to explore research areas considered impossible via experimentation. Architectural modeling of plants and bioinformatics will allow, when sufficiently developed, the reduction of field experiments.

Synthetic biology, the design and construction of new parts, devices and systems not existing in real world as well as the redesign of existing biological systems to perform desired functions, is another area benefited from the combination of information technology (IT) with nanotechnology. The report Nanofrontiers: visions for the future of nanotechnology (SCHMIDT, 2007), organized by National Science Foundation and National Institutes of Health of the USA, pointed out the importance of IT for the organization of a library of nanometric world (Nano Library), and for treatment of information generated by nanosensors.

Another report by a Canadian initiative called ETC Group (2007) predicts that biological engineers of the future will start their work on laptops instead of their laboratories. In accordance to this group, there is no barrier to the synthesis of plants and animals, and this will occur as soon as someone decides to finance it. Synthetic biology also will change intellectual property protection, because it will involve not only the organisms and biological processes designed, but also the software used in the process of creation.

In 2008, Jay Keasling, from UC-Berkeley, succeeded in inserting genes of *Artemisia annua* into a yeast cell and then reprogrammed some genes to transform sugar in artemisin, a powerful anti-malaria drug (INTERLANDI, 2008), reducing substantially its production costs. Besides the influence on biological research, Data Science has the potential to cause impacts on farming, as discussed in the next section.

Smart Farming and Internet of Things

In the macro scale, the Food and Agriculture Organization of the United Nations (FAO) predicts that world population will be around 9,6 billion people by 2050, increasing demand on agricultural production by 70% until then (ALEXANDRATOS; BRUINSMA, 2012). The need for production increase has to be combined with restricting factors such as deceleration of productivity growth, pressure to reduce deforestation, limited availability of arable land and water and climate change. From a global perspective, FAO estimates that water withdrawal for use in agriculture is around 70%. Agricultural processes will have to increase productivity while reducing the use of natural resources. These constraints require optimized use of resources.

Smart farming is applied in many sectors, such as tracking of farm vehicles, arable farming, livestock monitoring, greenhouses and stables, fish farming, forestry and storage monitoring of tanks (TOWARDS..., 2015). It also depends on sensing technologies, telematics and geographic positioning and communication systems. Internet of Things (IoT) has gained especial attention recently. IoT comprises a global network of machines and devices capable of interacting with each other (LEE; LEE, 2015). There are many applications of IoT in agriculture. For instance, the use of RFIDs[2] and temperature tags can capture temperature, humidity, animal feed rate and drinking water flow to evaluate animal health and comfort (ENABLING..., 2017).

Sensor networks data analysis in smart farming applications often require Data Science expertise to make sense of the huge volume of data, filter out noise and calculate appropriate interventions. A sensor network linked to an irrigation system, for instance, collect data on temperature and humidity from local weather stations and on moisture from soil sensors; filter out noise from faulty sensors; mix information from weather forecast service for that location; and calculates the amount of irrigation need. The decisions on field actuators could either be a fixed set of rules for every possible situation, or dynamic, through the use of machine learning algorithms and cognitive computation (KELLY III, 2015). A cognitive irrigation system proposes the amount of irrigation needed based on calculations and probabilities, but learns from the human behavior if the amount of irrigation actually applied is different.

Machine learning can also be used for identification of crop diseases (BARBEDO, 2016) and in robotics. Astrand and Baerveldt (2002) presented an autonomous agricultural robot that recognized crop rows and identified a single crop among weed plants. Once recognized, a weed tool removed weeds from the row of crops. The University of Sydney has

[2] RFID – radio frequency identifiers

also developed a prototype called RIPPA (Robot for Intelligent Perception and Precision Application). Instead of physically removing weeds, this robot shoots them with a micro-dose of pesticide (KING, 2017), thereby using only 0.1% of the volume usually consumed in whole field spraying.

Harper Adams University is working on machine vision to recognize weeds and then using laser to kill the weed by heating the growing point (BLACKMORE, 2014). Machine vision is also at the heart of drones in agriculture (MARINELLO et al., 2016). Drones have been used, for instance, in the context of integrated pest management, to evaluate spectral variations in plants to identify spots where problems are likely occurring, such as insect or pathogen infestations (BAYLIS, 2014). Moreover, the surgical elimination of weeds satisfies the increasing consumer's demand for food with less agrochemicals. Consumer awareness and water supply has a direct impact on public policies, as discussed in next section.

Public policies

In the context of social awareness, democratic governments are compelled to discuss and substantiate their decisions based on scientific evidences (DAVIES; NUTLEY, 2001). However, improving data-driven decision support requires the ability to integrate data from different sources and formats and the production of evidence based on robust data analysis techniques (WILSDON; DOUBLEDAY, 2015). Thus, the production of evidence to support the formulation of public policies may include several techniques of scientific computing to assess quantitatively the available options in terms of costs and goals.

In Brazil, one of such policies is the Climatic Risk Agricultural Zoning (ZARC), an instrument of agricultural policy and risk management in agriculture (ARIAS et al., 2015). Developed by the Ministry of Agriculture, Livestock and Supply (MAPA), Embrapa and partner research institutions, ZARC uses a series of simulation and analysis techniques. Assad et al. (2013) assess how climate change is expected to affect ZARC for cotton in Brazil until 2040.

Applications of Data Science in modeling and analysis, digital agriculture and public policies discussed until now form the main axis of the prospective areas discussed in the sequel.

PROSPECTIVE AREAS FOR DATA SCIENCE IN AGRICULTURE

As discussed in previous sections, Data Science has been applied in many different areas and scales, ranging from the nanoscale of synthetic biology to the macroscale of agricultural zon-

ing. The remainder of this paper will enumerate some prospective areas which would benefit from the application of Data Science in research, development and innovation in agriculture.

Integrated data modeling and analysis

- Integration of agricultural databases, including zootechnical, climatic, meteorological, crop production, animal production etc.;
- Systemic analysis of data from advanced biology, considering interactions between genes, proteins, metabolic pathways and regulatory networks. This analysis includes the mathematical simulation of biological processes (PEDERSEN et al., 2014), the application of machine learning processes to search for patterns and the cognitive assistants anchored in domain ontologies;
- Application of statistical and machine learning algorithms in genome-wide association studies (GWAS) to search for genomic markers corresponding to certain characteristics, such as resistance to diseases or tenderness of meat in animals. This search occurs in molecular markers of the type single nucleotide polymorphisms (SNPs);
- Analytical models for genetic evaluation, traditional and genomic, for large animal populations (millions of animals): genetic merit evaluations of breeders, screening characteristics of market interest, depend on the adjustment of complex statistical models to data collected from the population under evaluation.
- Land use change dynamics modeling, involving the use of different temporal and spatial scales in the same model, which can even be treated by different approaches and usually demand high computational capacity (SOARES-FILHO et al., 2010).

Applications for digital agriculture

- Integrated analysis of data from sensor networks of different nature with the objective of identifying patterns and computing the best intervention in human, environmental and economic terms. Interventions could be assisted by cognitive agents;
- Applications developed in the logic of games (serious games) with the purpose of training the user in a particular product, service or procedure (MICHAEL; CHEN, 2006);

- Development of simulators of agricultural systems or their components using "digital twin" technology, which integrates sensors, simulation and actuators in an interactive way to reproduce the behavior of the real system;
- Augmented reality systems incorporated into smartphones so that producers can view on-screen additional information about the components of their farms;
- Pest and disease alert systems based on climate conditions, computed from dispersion and susceptibility models using private and public sensor networks data;
- Recommendation systems that analyze users' interaction with systems and can suggest new content based on previous similar behaviors;
- Use blockchain technology in the development of applications for agri-food traceability (TIAN, 2016). The technology has many applications and has also been proposed to monitor contract length and transparency on Carbon Dioxide Removal (CDR) contracts (COFFMANN; LOCKLEY. 2017).6

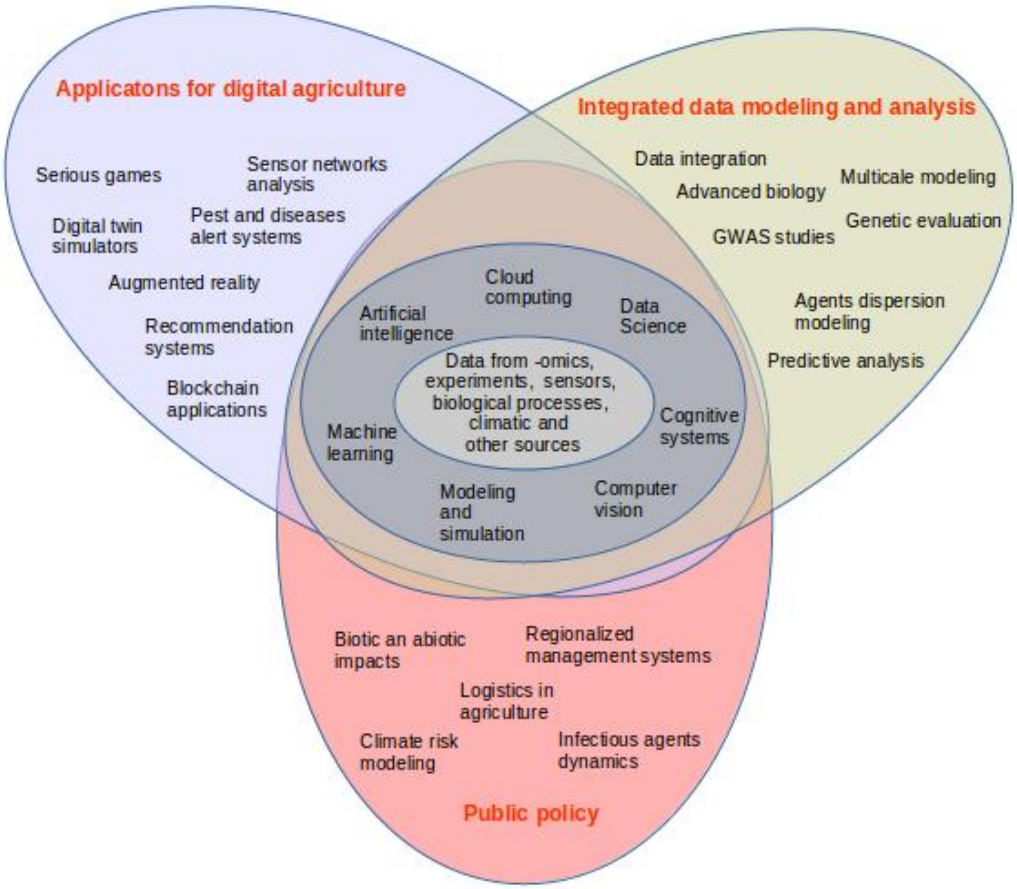
Public policy

- Modeling, simulation and optimization of agricultural production systems in response to risks and climate trends and evaluation of their environmental, social and economic sustainability;
- Predictive analysis of impacts on productivity caused by biotic or abiotic agents. The analysis could include both structured and unstructured data such as agricultural news;
- Data-driven regionalized management systems, considering aspects such as zoning, insurance and sustainable land management;
- Issuance of alerts and establishment of blocking areas to contain the spread of infectious agents in the national territory, based on predictive models (CUNNIFFE et al., 2015);
- Analysis of opportunities in logistics solutions and their impacts on the sustainability and competitiveness of agricultural production. From this analysis, the economic and environmental impacts of investments in infrastructure and logistics can be estimated.

Figure 1 shows schematically the interaction among the prospective areas. The inner oval represents data from sensors, experiments, biological processes, climatic or any other sources. In the next level, one can find the mathematical and computational disciplines which will be jointly applied to build applications. The other ovals represent each of the prospective

areas with their corresponding specific applications.

Figure 1 – Interaction among prospective areas for data science in agriculture.



Source: authors

CONCLUSION

In the era of Big Data, Data Science becomes the protagonist of digital transformation in agriculture. Several opportunities and directions for research have emerged as a consequence of a myriad of devices, standards, technologies, systems, models, algorithms, concepts and equipments. In particular, Data Science methods can be applied in a number of prospective areas, including research, digital farming public policies, marketing and commercialization. More specifically, such methods can: i) benefit researchers and practitioners from a better and more comprehensive data modeling and analysis tools applied to biological processes; ii) assist farmers to have more precision in their interventions with sensor networks – they would spend less and would have access to restricted markets that require stringent use of agrochemicals; and, iii) uphold governments to increase precision in global forecasts.

REFERENCES

ALEXANDRATOS, N.; BRUINSMA, J. World agriculture towards 2030/2050: the 2012 revision. ESA Working paper No. 12-03. Rome, FAO. 154p. 2012.

ANDERSON, C. The end of theory: The data deluge makes the scientific method obsolete. **Wired**, v. 16, n. 7, 2008.

ARIAS, D.; MENDES, P.; ABEL, P. (Coord.). **Revisão rápida e integrada da gestão de riscos agropecuários no Brasil**: caminhos para uma visão integrada. Brasília, DF: Banco Mundial, 2015. 76 p. il. Disponível em: <http://documents.worldbank.org/curated/en/717561467986362017/pdf/AUS12876-PORTUGUESE-REVISED-PUB-OUO-9-Riscos-Agropecu%C3%A1rios-no-Brasil-World-Bank-Group-paginas-compressed.pdf>. Acesso em: 9 ago. 2017.

ASSAD, E. D.; MARTINS, S. C.; BELTRAO, N. E. M.; PINTO, H. S.. Impacts of climate change on the agricultural zoning of climate risk for cotton cultivation in Brazil. **Pesq. agropec. bras.** [online]. 2013, vol.48, n.1, pp.1-8. ISSN 0100-204X. http://dx.doi.org/10.1590/S0100-204X2013000100001.

ÅSTRAND, B.; BAERVELDT, A. J. Autonomous Robots (2002) 13: 21. doi:10.1023/A:1015674004201

BARBEDO, J. G. A. Expert Systems Applied to Plant Disease Diagnosis: Survey and Critical View. **IEEE Latin America Transactions**, v. 14, n. 4, p. 1910-1922, 2016.

BAYLIS, A. Is precision agriculture ready to tackle pest management?. **Outlooks on Pest Management**, v. 25, n. 6, p. 350, 2014.

BELL , G.; HEY, T.; SZALAY, A. Beyond the Data Deluge, **Science**, vol. 323, no. 5919, pp. 1297–1298, 2009, doi: 10.1126/science.1170411.

BLACKMORE, S. Farming with robots 2050. In: Presentation delivered at Oxford Food Security Conference, 2014.

COFFMANN, D.; LOCKLEY, A. Carbon dioxide removal and the futures market. **Environmental Research Letters**, v. 12, n. 1, p. 015003, 2017.

DAVIES, H.; NUTLEY, S.;Evidence-based policy and practice: moving from rhetoric to reality. Third International, Inter-disciplinary Evidence-based Policies and Indicator Systems Conference, University of Durham, July 2001.

DHAR, V. Data Science and Prediction. **Communications of the ACM**, v. 56, n. 12, p. 64-73, 2013.

ENABLING the smart agriculture revolution: the future of farming through the IoT perspective. [S.l.]: Beecham Research, 2016. 23p. Available at: <http://www.beechamresearch.com/download.aspx?id=1051>. Accessed: 15 May 2017.

ETC Group. Extreme Genetic Engineering: an Introduction to Synthetic Biology. 2007.

HEY, T.; TANSLEY, S. & TOLLE, K., ed., *The Fourth Paradigm: Data-Intensive Scientific*

Discovery , Microsoft Research , Redmond, Washington,. 2009.

INTERLANDI, J. Power 2009: Jay Keasling, Molecular Biologist. **Newsweek** 19/12/2008. http://www.newsweek.com/power-2009-jay-keasling-molecular-biologist-83051. Acessado em 10/05/2017.

KELLY III, J. E. Computing, cognition and the future of knowing: how humans and machines are forging a new age or understanding. IBM Global Services, Sommers, NY, 2015.

KING, A.. The Future of Agriculture. **Nature**, v. 544, n. 7651, p. S21-S23, 2017.

LEE, I.; LEE, K. The Internet of Things (IoT): applications investments and challenges for enterprises. Business Horizons, v. 58, n. 4, p. 431-440, 2015.

MARINELLO, F.; PEZZUOLO, A.; CHIUMENTI, A.; SARTORI, L. Technical analysis of unmanned aerial vehicles (drones) for agricultural applications. **Engineering for Rural Development**, 15, 2016.

MAZZOCCHI, F. Could Big Data be the end of theory in science?. **EMBO reports**, p. e201541001, 2015.

MICHAEL, D.; CHEN, S. Serious Games: Games That Educate, Train and Inform. Course Technology, Mason, USA, 2006.

CUNNIFFE, N. J.; STUTT, R. O.; DESIMONE, R. E.; GOTTWALD, T. R.; GILLIGAN, C. A. Optimising and communicating options for the control of invasive plant disease when there is epidemiological uncertainty. **PLoS computational biology**, v. 11, n. 4, p. e1004211, 2015.

PEDERSEN, M.; OURY, N.; GRAVILL, C.; PHILLIPS, A. Bio Simulators: a web UI for biological simulation. **Bioinformatics**, v. 30, n. 10, p. 1491-1492, 2014.

PITAC - President's Information Technology Advisory Committee. Computational Science: Ensuring America's Competitiveness. Report to the President. 2005.

SCHMIDT, K. F. Nanofrontiers: visions for the future of nanotechnology. Project on Emerging Technologies. Woodrow Wilson International Center for Scholars. 2007.

SOARES-FILHO, B. S.; RODRIGUES, H. O.; COSTA, W. L. Modeling Environmental Dynamics with DINAMICA EGO. Belo Horizonte, 2010. 120 p.

TIAN, F. An agri-food supply chain traceability system for China based on RFID & blockchain technology. **In: Service Systems and Service Management (ICSSSM), 2016 13th International Conference on. IEEE**, 2016.

TOWARDS smart farming: agriculture embracing the IoT vision. [S.l.]:Beecham Research, 2015. 35 p.

WILSDON, J.; DOUBLEDAY , R. (Ed.). Future directions for scientific advice in europe. Cambridge: Centre for Science and Policy, 2015. 178p. Disponível em: <http://sro.sussex.ac.uk/54004/4/future-directions-for-scientific-advice-in-europe-v10.pdf>. Acesso em: 20 abr. 2017.

Sistema para recomendação de adubação

Cinthia Cabral da Costa[1]

[1] Embrapa Instrumentação, São Carlos, São Paulo, Brasil, cinthia.costa@embrapa.br.

RESUMO

As recomendações de aplicação de fertilizantes são realizadas, atualmente, usando como parâmetro tabelas elaboradas para diferentes condições de solo e cultura. Entretanto, as funções de respostas das plantas àqueles fertilizantes e os pontos de produtividade máxima econômica que o produtor pode alcançar para diferentes preços do produto e do insumo não são conhecidos. Alguns sistemas informatizados para esta recomendação já foram elaborados, mas apenas reproduzindo as tabelas existentes, sem a flexibilização de acordo com os preços de mercado. Uma vez que o objetivo do produtor agropecuário é obter máximo lucro, o qual depende dos preços do produto final e dos insumos utilizados, um sistema de recomendação de adubação informatizado, que permita que o produtor indique os preços da época e região faz-se necessário. O objetivo deste trabalho foi modelar e simular os resultados de um sistema deste tipo para demonstrar sua importância econômica. Verificou-se que o produtor rural pode obter ganhos financeiros considerando os preços de mercado em relação à recomendação dos manuais atualmente utilizados.

PALAVRAS-CHAVE: Produtividade máxima econômica, Fertilizante, Impacto econômico.

ABSTRACT

Currently, the fertilizer application recommendations are performed using tables building for different soil type and crop conditions. However, the plant responses to these fertilizers and the points of maximum economic productivity that producer can achieve for different prices of the product and the input are not known. Some computerized systems for this recommendation have already been elaborated, but only reproducing the existing tables, that is, no make these tables flexible according to the market prices. Since the objective of the agricultural producer is to obtain maximum profit, which depends on the prices of the final product and the inputs used, a computerized system of fertilizer recommendation where the producer could indicate the current prices is fundamental. The objective of this work was to

model and simulate the results of this system to demonstrate its economic importance. It was verified that the producer could obtain financial gains considering the market prices in relation to the recommendation of the manuals currently used.
KEYWORD: Maximum economic productivity, Fertilizer, Economic impact.

INTRODUÇÃO

Este artigo teve como objetivo identificar o impacto sobre o lucro para o produtor rural considerando uma recomendações de adubação de modo que o produtor atinja o ponto de máximo econômico ao invés das recomendações feitas com base nas tabelas atualmente utilizadas neste processo[1]. Foram considerados os produtos agrícolas milho e soja, e os principais nutrientes utilizados para fertilização química nestas culturas.

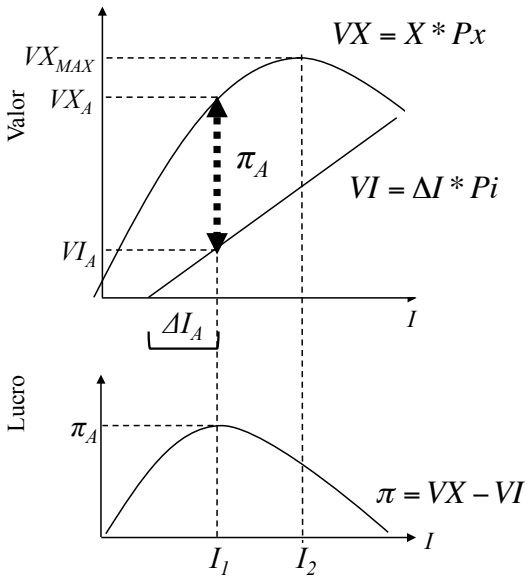
O aumento na produtividade é benéfico para o produtor rural pois, assim, ele aumenta sua receita por área com a venda do produto. Entretanto, o aumento no uso de insumos para que esta maior produtividade seja alcançada gera também custos adicionais. Assim, o desafio para o produtor é estabelecer a quantidade de insumos que irá originar seu máximo retorno econômico. Este ponto de máximo econômico é ilustrado na Figura 1, utilizando uma função representativa da receita do produtor (pela multiplicação da quantidade produzida, X, pelo preço do produto, Px) e do seu custo (multiplicando a quantidade do insumo fertilizante, I, pelo seu preço, Pi). Nesta figura observa-se que, a produção máxima (X_{MAX}) ocorre com a quantidade de insumo I_2. Neste ponto, a receita também é a máxima (VX_{MAX}). Entretanto, este ponto não corresponde ao nível de produção que origina o máximo de retorno econômico para o produtor, que é o seu lucro (π). O máximo lucro para o produtor, descrito como π_A nesta figura, corresponde à quantidade de insumo I_1, que origina um nível de produção X_A e receita VX_A, menor do que àquela descrita para o ponto de máxima produção (VX_{MAX}).

Portanto, podemos conceituar dois termos distintos: o de produtividade máxima (PM) e o de produtividade máxima econômica (PME). Entretanto, com o objetivo de simplificar o processo de recomendação de adubação, algumas publicações produzidas pelos principais órgãos de pesquisa agropecuários do país são adotadas como padrão. Contudo, apesar de procurar considerar as respostas das plantas aos nutrientes e, provavelmente, o preço do

[1] Muitas tabelas de recomendação de adubação descrevem níveis diferentes de insumo para diferentes níveis de produtividade esperada da cultura agrícola. Esta informação pode, indiretamente, ser utilizada para, por exemplo, em situações de baixo nível do preço do produto e, ou, alto nível do preço do insumo, uma recomendação considerando uma quantidade menor de fertilizante possa ser utilizada. Entretanto, não se tem uma relação direta entre esta relação de preços e as recomendações. Tal relação seria, inclusive, muito difícil de ser colocada no formato destas tabelas uma vez que podemos ter infinitas destas relações. A Figura 1 descreve o comportamento de algumas destas relações de preço apenas para os últimos 15 anos, e utilizando dados do Estado de São Paulo.

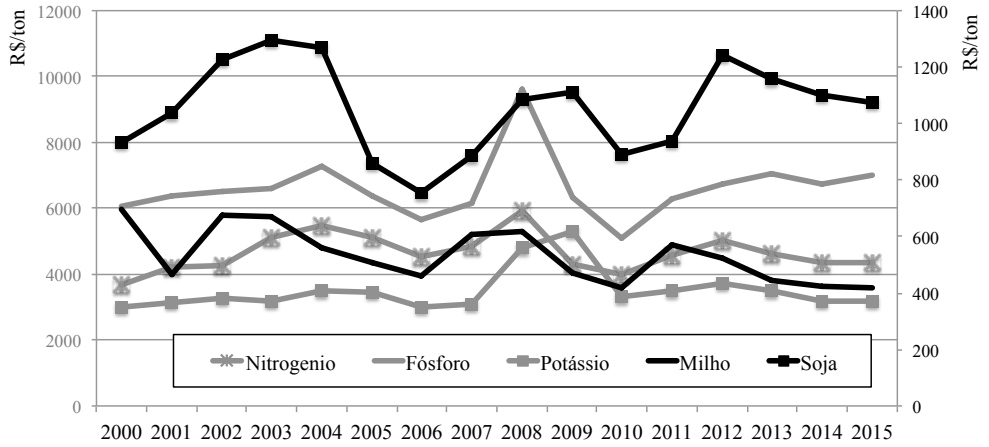
produto e insumo, estes preços se alteram com muita frequência (como ilustra a Figura 2) podendo, portanto, modificar o nível do nutriente para atingir o máximo econômico. Uma vez que a magnitude dos preços de produtos e nutrientes são muito diferentes, na Figura 2 o preço dos nutrientes é descrito no eixo da esquerda e dos produtos (milho e soja) no eixo da direita.

Figura 1 – Representação da receita, custo e do lucro do produtor elaborados a partir da função de produção agrícola



Fonte: Elaborado pelos autores.

Figura 2 – Média anual dos preços, deflacionados para 2015, dos produtos milho e soja, no eixo da direita, e dos macronutrientes nitrogênio, fósforo e potássio, no eixo da esquerda, no Brasil. Período de 2000 a 2015.



Fonte: IEA, 2015.

Por causa da grande variação à que os preços agrícolas estão sujeitos, e da importância dos mesmos para o lucro do produtor, como este artigo demonstrou, um sistema dinâmico, como um software, onde estes preços possam ser alterados de acordo com a realidade do produtor é preconizado pelos autores. A próxima sessão mostra o método e dados utilizados para analisar os impactos sobre a lucratividade do produtor rural ao utilizar o método

economicamente mais eficiente, que pode ser implementado por um aplicativo, em comparação ao método atual de tabelas. A seguir são apresentados os resultados da pesquisa e as conclusões.

MATERIAL E MÉTODOS

Para obter as recomendações relacionadas aos pontos de produtividade máxima econômica (PME) são necessárias as seguintes informações: função de resposta do milho e da soja a cada um dos macronutrientes; preço destes produtos agrícolas e de cada um dos nutrientes. As funções de produção utilizadas neste estudo foram obtidas da literatura e estão descritas na Tabela 1, assim como a fonte do estudo que originou tal função. Nesta tabela são também apresentados as quantidades de cada nutriente que origina o ponto de máxima produtividade física (PM)[2]. O nitrogênio não foi considerado como nutriente na produção da soja uma vez que ele é disponibilizado para esta cultura pela fixação biológica através da inoculação de bactéria nas sementes, e não como adubação química.

Tabela 1 – Funções de produção utilizadas para produção de milho e soja, com respectivas fontes e níveis de nutrientes que originam o ponto de máxima produtividade (PM)

Produto agrícola (*X*)	**Nutriente Avaliado (*I*)**	**Função de produção**	**Fonte**	**Quantidade de nutriente no PM (Kg/ha)**
Milho	Nitrogênio (N)	$X=651+28,9*I-0,13*I^2$	Lucena et al. (2000)	111
	Fósforo (P_2O_5)	$X=814,28+14,62*I-0,037*I^2$	Lucena et al. (2000)	198
	Potássio (K_2O)	$X=-2981+115,5*I-0,318*I^2$	Silva & Ritchey (1982)*	182
Soja	Fósforo (P_2O_5)	$X=1.523+31,84*I-0,199*I^2$	Souza et al. (2014)	80
	Potássio (K_2O)	$X=2.726+13,415*I-0,068*I^2$	Leal et al. (2015)	98

Nota: *Equação estimada a partir do gráfico da função de produção apresentada naquele estudo.

A influencia dos preços de milho, soja, nitrogênio (N), fósforo (P_2O_5) e potássio (K_2O) foi analisada para a média anual dos preços reais observados no país nos últimos quinze anos (2000 a 2015). Todos os preços foram deflacionados para 2015 de maneira que pudessem ser comparados os diferentes anos. Para isto foi utilizado o Índice Geral de Preços – Disponibilidade Interna (IGP-DI) disponibilizado por FGV (2016).

Em relação aos preços dos macronutrientes, uma vez que há diferentes fertilizantes que os contém, os preços dos mesmos para compor a análise realizada neste estudo foram os seguintes: uréia, que contem 45% de N; fosfato simples, que computa 18% de P_2O_5 e; cloreto

[2] A formula para obter este valor pode ser visualizada em Costa & Burnquist (2016).

de potássio, que conta com 60% de K_20. Com as funções de produção e os preços foram estimadas as quantidades ótimas (que origina a PME) de cada fertilizante, em cada cultura, pela equação (1) descrita na seção anterior. Já o lucro do produtor, por unidade de área (R$/ha) foi estimado conforme descrito na equação (2).

$$I = (P_i - b * P_x) / (2 * c * P_x) \quad (1)$$

$$\pi = VX - VI = X * P_x - I * P_i \quad (2)$$

Entretanto, a quantidade de fertilizante aplicado não é necessariamente igual ao nível que acarreta a PME, uma vez que a análise de solo pode indicar já algum nível do nutriente no solo. Isto ocorre para fósforo e potássio. O nitrogênio não é analisado uma vez que é muito móvel no solo e, portanto, são aplicadas suas quantidades totais indicadas para o nível de PME. Para as análises realizadas neste estudo foram considerados três diferentes níveis de fertilidade do solo, para cada um dos nutrientes analisados: solos com baixo, médio e alto níveis de nutriente, descritos por Ribeiro et al. (1999); Sousa & Lobato (2004); Manual (2003); Oliveira (2003); Raij et al. (1996) e Malavolta (1989). Para o fósforo este níveis foram, respectivamente, de 9,8; 21,3 e 47,7 kg/ha. Para o potássio, os níveis foram de 41,5; 77,5 e 158 kg/ha, respectivamente para solos de baixo, médio e alto níveis do nutriente[3]. Uma vez que a quantidade de nutriente que origina o máximo retorno econômico foi comparado aos manuais atualmente utilizados com este propósito, a Tabela 2 descreve as indicações de adubação para as culturas agrícolas analisadas considerando algumas destas principais tabelas de recomendação utilizadas no Brasil.

Tabela 2 – Recomendações de adubação segundo alguns dos principais manuais usados no país, em kg/ha

			Sousa & Lobato (2004)	Raij et al. (1996)	Ribeiro et al. (1999)	Manual (2004)	Oliveira (2003)
Milho	N	solo pobre		170	160	**90**	180
		solo médio	210	170	160	**90**	180
		solo rico	210	170	160	**90**	180
	P_2O_5	solo pobre		100	120	125	**90**
		solo médio	120	70	100	85	**70**
		solo rico	60	50	70	75	**50**
	K_2O	solo pobre		160	90	110	**70**
		solo médio	150	130	80	70	**60**
		solo rico	150	100	60	60	**50**
Soja	P_2O_5	solo pobre		**80**	120	110	
		solo médio	100	**50**	80	70	
		solo rico	50	**40**	40	60	
	K_2O	solo pobre		**80**	120	125	
		solo médio	100	**60**	80	85	
		solo rico	70	**40**	40	75	

[3] Os níveis destes nutrientes no solo são usualmente classificados na unidade de mg/dm^3 de solo. Entretanto, os valores descritos na Tabela 1 foram transformados para kg/ha para ficar mais claro a concentração dos nutrientes que já existiam e da quantidade que será aplicada.

Fonte: elaborado pelos autores, a partir de: Ribeiro et al., 1999; Sousa & Lobato, 2004; Manual, 2003; Oliveira, 2003; Raij et al., 1996 e Malavolta, 1989.

A construção da Tabela 2 foi feita de maneira aproximada uma vez que as recomendações são, para algumas das fontes citadas, feitas com base em algumas diferentes expectativas de rendimento da cultura. Assim, foi adotado como critério o uso da quantidade de insumo que origina o maior nível de rendimento, quando esta variável foi considerada. Para não superestimar a simulação do impacto para o uso do nutriente segundo estas tabelas de recomendação, o nível de nutriente aplicado neste caso foi considerado como a menor quantidade descrita na Tabela 2 entre as diferentes fontes descritas. Este níveis estão em negrito nesta tabela e corresponde ao descrito no: Manual (2004) para os níveis de nitrogênio (N) no milho; Oliveira (2003) para fósforo (P) e potássio (K) no milho e; Raij et al. (1996) para os níveis de fósforo (P) e potássio (K) na cultura da soja.

A seguir, foram simuladas as variações no lucro para o produtor rural calculando a diferença de lucratividade ao considerar o ponto de máximo lucro com o lucro estimado simulado caso o produtor considere as informações da Tabela 2 para as recomendações de adubação. A variação no lucro ($\Delta\pi$) é estimada da acordo com o descrito na equação (3):

$$\Delta\pi = [Xa*PX - IA*PI] - [Xb*PX - IB*PI] \quad (3)$$

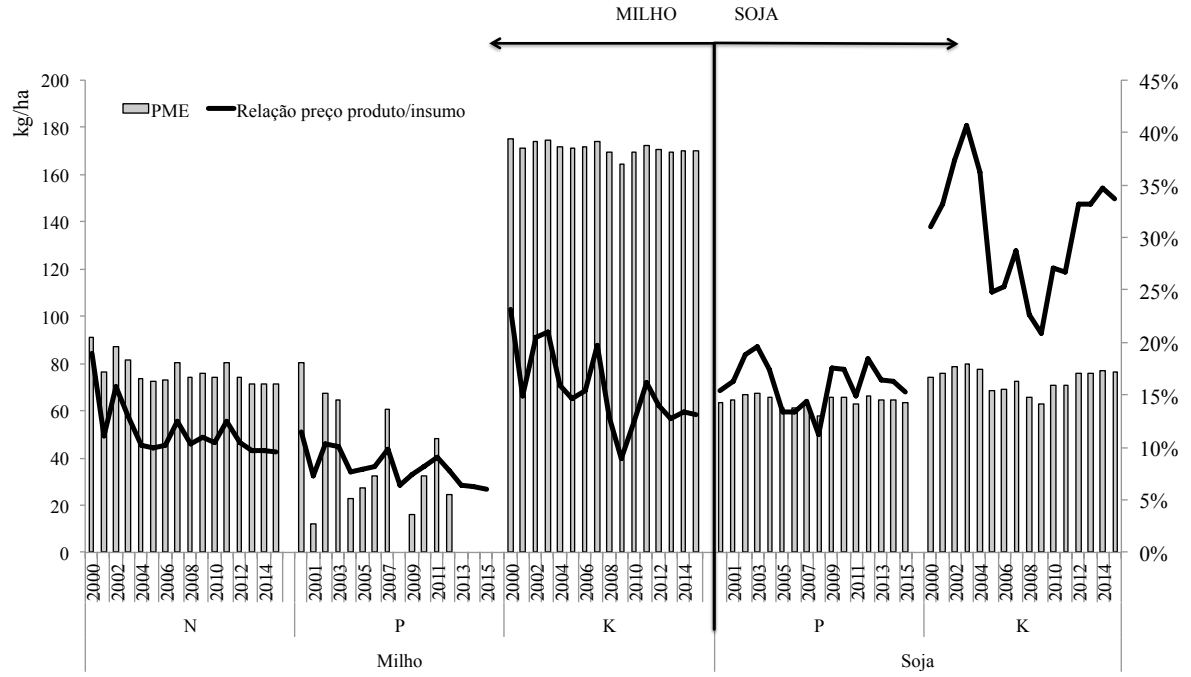
onde X_a corresponde ao nível de produtividade ao utilizar a quantidade de insumo I_a, o qual foi calculado para o ponto de máxima produtividade econômica considerando as funções de produção e preços de produto (P_X) e insumo (P_I) como apresentado na equação (1). Como já descrito previamente, a quantidade de fertilizante aplicado (I^A) não é necessariamente igual ao nível I_a, uma vez que a análise de solo pode indicar já algum nível do nutriente no solo. Assim, $I_a = I^A$ + nutriente existente no solo. E X_b corresponde ao nível de produtividade correspondente à quantidade I_b de nutriente disponível para a planta. Esta quantidade refere-se à quantidade de nutriente recomendada pelas tabelas de referencia atualmente encontradas na literatura (I^B) adicionado ao nível do nutriente no solo já existente no solo. Assim, $I_b = I^B$ + nutriente existente no solo.

RESULTADOS E DISCUSSÃO

A Figura 3 mostra a variação nos níveis de adubação economicamente mais eficientes para o produtor, no período de 2000 a 2015. A mudança no nível ótimo econômico dos nutrientes, em kg/ha, causados pela variação nos preços neste período, mostrada nas colunas da Figura 3, já indica a importância de não haver uma recomendação de adubação padrão, como ocorre atualmente. Esta mudança nos níveis de nutrientes para atingir o PME decorre não apenas da

variação no preço de produtos e insumos, mas da variação na relação entre estes preços (indicada nas linhas da Figura 3).

Figura 3 – Relação (%) entre os preços do produto e do insumo, nas linhas, e níveis dos nutrientes nitrogênio, fósforo e potássio para atingir o Ponto de Máximo Econômico (PME) - I_a, em kg/ha, nas colunas, em cada ano do período analisado



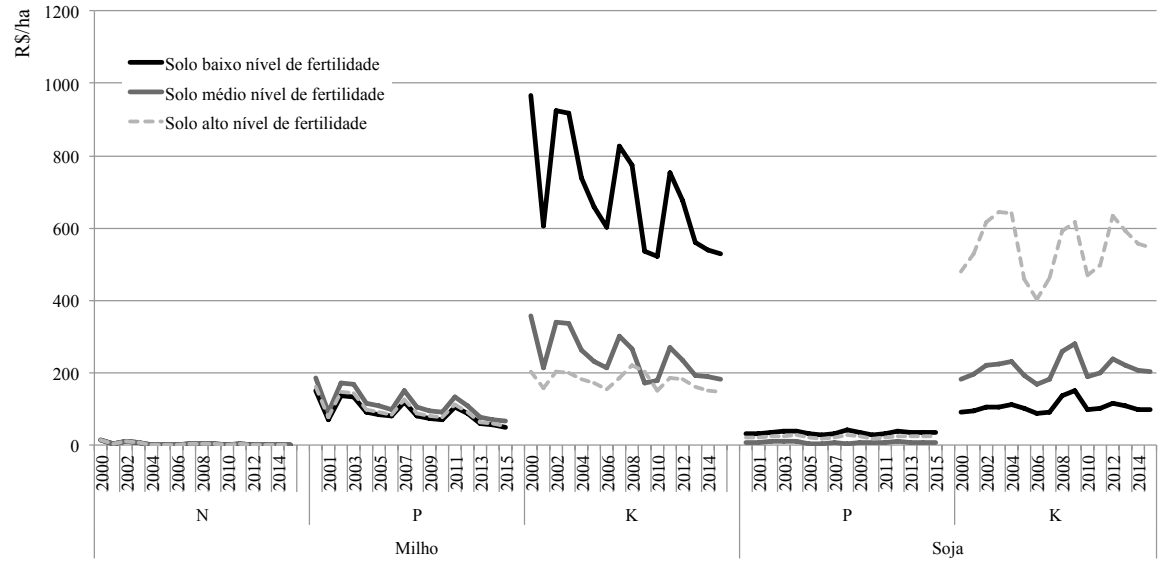
Fonte: Resultado da pesquisa.

O nível aplicado de nitrogênio para a produção de milho, por exemplo, considerando a quantidade para atingir o PME, variou de 71 a 91 kg/ha no período. Já para a aplicação do fósforo, observa-se uma variação maior, desde o nível zero, ou seja, não aplicar este nutriente, o que ocorreu em vários anos, até o nível de 80 kg/ha em 2000 para alcançar o PME. Para a aplicação do potássio observou-se uma maior estabilidade, ficando entre 164 e 175 kg/ha do nutriente. Já na produção de soja, as quantidades de fósforo e potássio necessárias para atingir o PME ficaram entre 58 e 67 kg/ha e entre 63 e 80 kg/ha, respectivamente.

Aplicando os valores apresentados em cada ano na fórmula descrita na equação (3) tem-se a variação no lucro do produtor rural resultante da mudança de recomendação. Ou seja, ao invés da aplicação dos nutrientes seguindo as tabelas de recomendação, considerar a função de produção do produto e os preços de mercado vigentes. Esta variação no lucro, por hectare cultivado de milho e soja, é descrita na Figura 4 para as condições de fertilidade baixa, média e alta dos solos. O resultado mais significativo observado nesta figura é que, aplicar a dose da PME em oposição às doses recomendadas nos manuais, gerou sempre um ganho econômico para o produtor. Em geral, podem ser observados o aumento no lucro do produtor acima de R$100,00 por hectare. Os maiores ganhos foram obtidos quando a mudança ocorreu

considerando a aplicação do potássio, na produção de ambas as culturas. Na produção de milho os maiores ganhos foram em solo com baixo nível de fertilidade e na produção de soja, este ganho de lucro do produtor foi maior em solo com maior fertilidade[4].

Figura 4 - Variação no lucro do produtor se aplicar nutrientes para a condição economicamente ótima ao invés daqueles indicados nas tabelas de recomendação de adubação, preços médios anuais de mercado entre 2000-2015



Fonte: Resultado da pesquisa.

Considerando uma área de 1.000 hectares, o ganho de lucro para o produtor alterando a aplicação de nitrogênio poderia ficar em cerca de R$20,00 mil na produção de milho para os preços que vigoraram nos anos de 2004, 2005, 2008, 2013 a 2015, e em menos de R$1,00 mil com os preços de 2000 e 2002, por exemplo. Já a alteração na quantidade aplicada de fósforo produziria um ganho no lucro do produtor entre R$3,00 mil (para preços observados em 2000 e solos com fertilidade média) e mais de R$200,00 mil (em solos com fertilidade baixa). Considerando o potássio, o ganho de lucro ao mudar o tipo de recomendação (da tabela para o calculado objetivando alcançar a PME) na produção de milho, numa lavoura desta mesma dimensão, ficaria entre R$240,00 mil e R$890,00 mil, dependendo do nível de fertilidade do solo, com as condições de mercado de 2000.

Para a produção de soja, os ganhos no lucro do produto ao mudar o nível aplicado de fósforo e potássio foram menores do que os observados para o milho. Observou-se ganhos no lucro entre R$5,00 por hectare (solo de fertilidade média) a cerca de R$150,00/ha (em solos menos férteis) no caso da aplicação de fósforo e entre R$140,00 (solo de baixa fertilidade) e cerca de R$700,00 (solo de fertilidade alta) por hectare, no caso de aplicação de potássio.

[4] Dado que a variação na lucratividade nestes casos foi muito maior que nos demais, pode-se suspeitar também de que as funções de produção utilizadas para mensuração da resposta destas culturas ao potássio não tenha sido bem estimada. Entretanto, não foram identificadas outras funções na literatura que melhor se ajustassem.

Portanto, observa-se como o aumento no lucro pode ser alterado quando são alterados os cenários de preços. A magnitude da variação no lucro, para os preços médios anuais dos últimos 15 anos, são resultados que demonstram a importância de um sistema flexível, que leve em conta os preços de mercado como informação para tomada de decisão pelo produtor.

CONCLUSÕES

Entre as questões colocadas para melhoria das condições da agricultura brasileira estão o aumento da produtividade e a redução de custos. Este estudo identificou um ponto onde atuar para atingir ambos propósitos: a adubação. A adubação da produção agrícola é, até os dias de hoje, essencialmente guiada por tabelas padrões. Este estudo mostrou que há possibilidade de ganhos no lucro mudando este instrumento de recomendação de adubação. Ao invés das tabelas usuais, partiu-se da teoria econômica de maximização de lucro e se utilizou este conceito para proceder às recomendações. Um sistema computacional pode ser um instrumento facilmente utilizado pelo produtor para obter este resultado.

De modo geral, as recomendações de fertilizantes ficaram abaixo dos níveis recomendados nas tabelas, mesmo utilizando os níveis mais baixos de recomendação entre as tabelas analisadas. Isto pode também ser uma indicação da superestimação existente nas tabelas de recomendação atualmente em uso, ou de que as funções de produção encontradas na literatura que deram suporte para este estudo não são condizentes com a realidade agronômica. Entretanto, registra-se aqui a grande dificuldade de se encontrar funções de produção na literatura, as quais são essenciais para se definirem as recomendações de adubação nas culturas agrícolas.

Uma limitação neste estudo é a hipótese que os impactos da mudança nos níveis de recomendação dos três macronutrientes (N, P e K) e, consequentemente, os ganhos derivados desta mudança são adicionais um ao outro. Na prática, tem-se que o efeito da interação entre os nutrientes analisados pode gerar impactos diferentes dos descritos para os efeitos individualizados nas funções de produção. Entretanto, esta limitação também ocorre nas tabelas de recomendação.

REFERÊNCIAS

FUNDAÇÃO GETÚLIO VARGAS-FGV. **Pesquisa. Instituto Brasileiro de Economia. Indicadores de Preços**. Disponível em: <http://portalibre.fgv.br/main.jsp?lumChannelId=402880811D8E34B9011D92B6B6420E96> . Acesso em: 20 de jan. 2016.

INSTITUTO DE ECONOMIA AGRÍCOLA – IEA. **Banco de Dados. Preços Agrícolas**. Disponível em: < http://www.iea.sp.gov.br/out/bancodedados.html>. Acesso em: 20 de jan. 2016.

INTERNATIONAL FERTILIZER INDUSTRY ASSOCIATION – IFIA. **Statistics. Statistics Database. IFA**. Disponível em: < http://ifadata.fertilizer.org/ucSearch.aspx>. Acesso em: 02 de fev. 2016.

LEAL, A.J.F.; VALDERRAMA, M.; KANEKO, F.H.; LEAL, U.A.S.; PERIN, A.; LUCHESE, K.U.O. Produtividade da soja de acordo com diferentes doses de cloreto de potássio revestido ou não com polímeros. **Global Science and Technology**, v.8, n.1, p.19-30. 2015.

LUCENA, L.F.C.; OLIVEIRA, F.A.; SILVA, I.F.; ANDRADE, A.P. Resposta do milho a diferentes dosagens de nitrogênio e fósforo aplicados ao solo. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.4, n.3, p.334-337. 2000.

MALAVOLTA, E. **ABC da adubação**. 5. ed. rev. at. São Paulo: Agronômica Ceres, 1989. 292 p.

MANUAL de adubação e de calagem para os Estados do Rio Grande do Sul e Santa Catarina. 10 ed. Porto Alegre: Sociedade Brasileira de Ciência do Solo - Núcleo Regional Sul - Comissão de Química e Fertilidade do Solo, 2004. 400 p.

OLIVEIRA, E. L. de. (Coord.). **Sugestão de adubação e calagem para culturas de interesse econômico no Estado do Paraná**. Londrina, PR: IAPAR, 2003. 30 p. (IAPAR. Circular Técnica, 128).

RAIJ, B. van; CANTARELLA, H.; QUAGGIO, J. A.; FURLANI, A. M. C (Ed.). **Recomendações de adubação e calagem para o estado de São Paulo**. 2. ed. Campinas: Instituto Agronômico, 1996. 285p. (IAC. Boletim Técnico, 100).

RIBEIRO, A. C.; GUIMARAES, P. T. G.; ALVAREZ V., V. H. (Ed.). **Recomendação para o uso de corretivos e fertilizantes em Minas Gerais:** 5a. aproximação. Viçosa, MG: Comissão de Fertilidade do Solo do Estado de Minas Gerais, 1999. 359 p.

SILVA, J.E.; RITCHEY, K.D. Adubação potássica em solos do cerrado. In: Anais Simpósio sobre potássio na agricultura brasileira. Londrina. 1982.

SOUSA, D. M. G. de; LOBATO, E. (Ed.). **Cerrado:** correção do solo e adubação. 2. ed. Brasília, DF: Embrapa Informação Tecnológica; Planaltina, DF: Embrapa Cerrados, 2004. 416 p. il.

SOUZA, J.R.; RIBEIRO, B.N.; RAPOSO, T.P.; FIORIN, J.E.; CASTRO, G.S.A.; MAGALHÃES, R.S. Eficiência do fósforo revestido com polímeros na cultura da soja. **Acta Iguazu**, v.3, n.4, p.1-9. 2014.

Revisão de literatura de técnicas de caracterização de leite e derivados com uso de ultrassom

Laura Beatriz Alves[1], *Francisco José Arnold*[1], *Elaine Cristina Catapani Poletti*[1], *Eduardo Tavares Costa* [2]

[1] Faculdade de Tecnologia, Universidade Estadual de Campinas, Limeira, São Paulo, Brasil, laurab.alves@gmail.com, arnold@ft.unicamp.br, elainec@ft.unicamp.br

[2] Centro de Engenharia Biomédica, Universidade Estadual de Campinas, Campinas, São Paulo, Brasil, educosta@ceb.unicamp.br

RESUMO

O controle da qualidade do leite é um assunto que desperta interesse sob diferentes aspectos, como saúde, economia, produção, entre outros. Há muitas técnicas que são utilizadas para avaliar possíveis adulterações do leite, todavia, muitas delas são de custo elevado e necessitam de laboratórios especializados. Técnicas baseadas em ultrassom apresentam-se como uma possibilidade prática e de menor custo. O presente artigo tem o objetivo de reunir informações sobre métodos de análise de leite e derivados com uso de ultrassom. As técnicas ultrassônicas, bem como a sua amplitude, possíveis falhas, relacionando-as com a caracterização de amostras e o desempenho das técnicas são apresentadas. Constatou-se a predominância do uso das variáveis acústicas, velocidade de propagação e o coeficiente de atenuação das ondas, nas avaliações experimentais. Os estudos mostraram que o ultrassom é uma alternativa para análise a qual fornece informações rápidas e que pode caracterizar os meios analisados.

PALAVRAS-CHAVE: Ultrassom, Leite, Adulteração.

ABSTRACT

The quality control of milk is a subject that arouses interest in different aspects, such as health, economy and production. There are many techniques that are used to evaluate possible milk adulterations, however, many of them are expensive and need specialized laboratories. Ultrasound-based techniques present themselves as a practical and lower-cost possibility. The

aim of this article is to gather information on methods of analysis of milk and milk products using ultrasound. Ultrasonic techniques, as well as their amplitude, possible failures, relating them to the characterization of samples and the performance of the techniques are presented. It was recorded the predominance of the use of acoustic variables, propagation velocity and wave attenuation coefficient, in the experimental evaluations. Studies have shown that ultrasound is an alternative for analysis which provides rapid information and can characterize the means analyzed.

KEYWORDS: Ultrasound, Milk, Adulteration.

INTRODUÇÃO

O leite é um alimento rico em nutrientes e está presente diariamente na alimentação da população mundial, seja de maneira integral ou em forma de derivados, e movimenta bilhões de reais no Brasil. Possui alto valor nutritivo, composto por água, carboidratos, proteínas, gorduras, vitaminas e minerais em diferentes estados de dispersão (BRASIL, 1997).

Algumas análises físico-químicas são obrigatórias para controle e garantia de qualidade no país e, geralmente, são realizadas em laboratório próprio ou terceirizado. Tais análises demandam tempo e capital investidos. São feitas de maneira amostral e com periodicidade variada. As análises exigidas são: gordura, acidez titulável, densidade, índice crioscópico, sólidos não gordurosos, alizarol, contagem padrão em placas, contagem de células somáticas e pesquisa de resíduos de antibióticos (BRASIL, 2011). No mercado há opções de aparelhos que usam diferentes técnicas: os ultrassônicos (Lactoscan, BulltechEkomilk), os com base em absorção de raios infravermelhos (Laktron LCF-980, MirisDairyMilkAnalyser) e os que avaliam o índice crioscópico das amostras de leite (Laktron LK-7000), entre outros. Todos avaliam a composição do leite, e trabalham com análises simultâneas de amostras e possuem alto custo.

Os métodos de análise atuais mais utilizados não apresentam uma resposta rápida quanto à caracterização das amostras, seja para estabelecer sua composição no pagamento por conteúdo ou para identificar irregularidades. Com isso, a necessidade de uma técnica alternativa e com resposta rápida se faz presente dentro da cadeia produtiva. Neste cenário, o ultrassom se apresenta como uma possibilidade pouco explorada, menos onerosa e que pode fornecer informações relevantes sobre o meio analisado (CHRISTENSEN, 1988; MILES, SHORE E LANGLEY, 1990).

A partir de então, este artigo reúne informações para sobre o uso de ultrassom para caracterizar amostras de leite e derivados e verificar, diante dos estudos já desenvolvidos, a viabilidade e a amplitude da técnica. Com o levantamento das deficiências e consistências de cada estudo, averígua-se a aplicabilidade do ultrassom como um método de caracterização, além de encontrar lacunas para estudos posteriores a serem desenvolvidos.

MATERIAL E MÉTODOS

Há diversos métodos de análise de leite no mercado, dentre eles os testes de pH - que identifica a acidez do leite - e ponto de congelamento do leite cru, que pode identificar a adição de água (GLAESSER, 2003). Para o teste de pH são necessários aparelhos potenciométricos, enquanto que o ponto de congelamento envolve crioscópios eletrônicos (IDF, 2002). Estas metodologias limitam o atendimento da demanda por elevado número de análises, uma vez que os aparelhos possuem reduzida capacidade analítica (OLIVEIRA et al., 2006). Por isso, os laboratórios têm buscado novas metodologias por meio de equipamentos mais rápidos e precisos. Uma delas é a espectroscopia de infravermelho.

Os equipamentos com infravermelho são baseados na emissão de energia eletromagnética, e a interação sobre a matriz que absorve esta energia gera uma absorção - que é característico de cada tipo de material. Portanto, ao adicionar algum produto ele pode ser identificado (COITINHO et al., 2017). Por ter maior capacidade de análises simultânea e envolverem tecnologias mais recentes, torna-se um equipamento acessível apenas a grandes laboratórios especializados. Outra tecnologia muito utilizada nas análises de leite é o ultrassom. Há equipamentos que operam em larga escala, assim como os infravermelhos, e equipamentos que fazem caracterizações em menor volume.

O ultrassom pode ser entendido como uma onda mecânica, com frequência maior que 20 kHz e que induz à vibração as partículas que compõem o material analisado por meio de ondas acústicas (FERNANDES, 2010). Utilizado em processos não destrutivos, identifica estruturas, propriedades, auxilia em diagnósticos, etc.

Ao se propagar em um meio de análise, um sinal ultrassônico perde sua intensidade devido a alguns fatores, como: o espalhamento do sinal pelas partículas que compõem o meio, o efeito de divergência em relação ao eixo de propagação e absorção da energia. Considerando a propagação de ondas em amostras de leite, o sinal de ultrassom sofre atenuações pelo espalhamento do sinal ao encontrar partículas de gordura e pela absorção do

meio. Com isso, a atenuação aumenta de acordo com o grau de homogeneização das amostras (CHRISTENSEN, 1988; MILES, SHORE E LANGLEY, 1990; ALVES et al., 2015).

Os métodos de experimentação com ultrassom para caracterização de líquidos mais utilizados são o de pulso eco e o de transmissão-recepção (BACANELI, 1999; POVEY, 1999; MILES, SHORE E LANGLEY, 1990; NAZARIO, 2006; DORABIATO, 2006; ELVIRA, 2007). O primeiro método necessita de um transdutor com a função de emissão do sinal acústico enquanto que um segundo transdutor possua a função de recepção do sinal. Ambos devem ser dispostos de maneira paralela. O segundo método utiliza apenas um transdutor que alterna sua função de emissão e recepção do sinal. Com o uso de um destes métodos, é possível obter dados acústicos como a velocidade de propagação e o coeficiente de atenuação das ondas nos meios analisados (CHRISTENSEN, 1988; ASSEF et al., 2009).

A velocidade de propagação de um pulso de ultrassom que se propaga por um meio qualquer e pode ser entendida pelo conceito básico de velocidade: variação de espaço pelo tempo percorrido do emissor ao receptor do sinal (ADAMOWSKI et al., 2004).

O coeficiente de atenuação de ondas é uma medida que corresponde ao decréscimo da intensidade do sinal em função da distância da fonte das ondas em um meio. Este efeito pode ocorrer por fatores geométricos e absorção, característicos de cada meio (ASSEF, 2009; KINSLER, 2000). Também pode sofrer interferências pela divergência do feixe em relação ao eixo central, espalhamento pela não homogeneidade do meio, conversão em outros modos de vibração, relaxação, movimento relativo e bolhas de gás (BACANELI, 1999). Há diversas maneiras de calcular o coeficiente de atenuação de ondas (ADAMOWSKI et al., 2004).

RESULTADOS E DISCUSSÃO

No levantamento bibliográfico foram analisados 15 artigos científicos que utilizam ultrassom para análise e/ou caracterização de amostras de leite e derivados. A Tabela 1 apresenta a relação entre os estudos selecionados e suas características, como o material analisado, qual variável acústica foi isolada e qual o objeto de estudo.

Tabela 1- Quadro com identificação dos estudos e suas particularidades

Autores (ano)	Variáveis	Material	Objeto de Estudo
Alves et al. (2015)	Velocidade e atenuação	Leite cru	Adulteração com água
Alouche et al (2015)	Atenuação	Iogurte	Monitoramento da fermentação
Izbaim et al. (2012)	Velocidade e atenuação	Iogurte	Monitoramento da fermentação

Bosiljkov et al. (2011)	Absorção e viscosidade	Leite	Alteração de propriedades
Elvira et al.(2009)	Velocidade e densidade	Leite	Adulteração com água e melamina
Nazario et al. (2006 e 2009)	Velocidade, atenuação e densidade	Leite	Identifica adição de água e relaciona com gordura
Savaroglu e Aral (2007)	Velocidade e densidade	Leite	Adulteração com água e bicarbonato de sódio ou peróxido de hidrogênio
Dorabiato (2007)	Tempo de propagação e amplitude	Leite UHT	Adulteração com água
Elvira (2006)	Velocidade e atenuação	Leite	Identifica microorganismos em embalagens TetraPack
Dukhin et al. (2005)	Velocidade e atenuação	Leite e manteiga	Identifica partículas de gordura
Mohanan et al (2002)	Velocidade e densidade	Leite	Adulteração com carbonato de sódio, bicarbonato de sódio e formol
Bacaneli (1999)	Velocidade e atenuação	Leite	Identifica partículas de gordura
Mc Clements (1997)	Velocidade e atenuação	Leite e derivados	Caracterização de amostras
Miles et al. (1990)	Atenuação	Leite, creme de leite e proteína	Caracterização de amostras

Primeiramente, com o leite bovino cru há um estudo com técnicas de ultrassom, lógica fuzzy e adulterações por água (ALVES et al., 2015). Neste artigo, amostras de leite são adulteradas com água de maneira assistida, e as variáveis coeficiente de atenuação em banda larga e velocidade de propagação são medidas de acordo com o percentual de adulteração. Foi constatado que as variáveis carregam informações sobre o meio analisado e modificam seu comportamento, sendo que a velocidade diminui e o coeficiente de atenuação aumenta com o acréscimo de água adicionada. Um sistema de auxílio à decisão com base de regras fuzzy recebia as variáveis de entrada e classificava as amostras.

Em Alouche et al (2015), o coeficiente de atenuação foi definido levando em consideração o coeficiente de reflexão e transmissão, e correlacionado ao processo de fermentação no monitoramento dos estágios de fermentação do iogurte. As alterações no coeficiente, originadas por um sinal e obtidas por um osciloscópio, foram relacionadas ao pH (medido com outros equipamentos) com uso do hardware e software IEEE-GPIB.

Izbaim et al. (2012) também caracterizaram derivados com velocidade de propagação e coeficiente de atenuação utilizando o módulo elástico, a densidade e a razão de Poisson do meio para o cálculo da velocidade, e a diferença entre a amplitude inicial e final da onda no

espectro da frequência para o cálculo da atenuação. Neste caso, foi acompanhado o processo de fermentação do iogurte. O crescimento das bactérias associado à mudança na consistência do meio de análise facilita a caracterização.

Em Bosiljkov et al. (2011), é possível identificar que a emulsificação ultrassônica atraiu muito interesse na homogeneização do leite de ovelha. As aplicações de diferentes diâmetros de onda em diferentes intervalos de tempo têm influência significativa sobre todas as propriedades físicas observadas, como temperatura, absorção e viscosidade. Apesar de não tratar da análise de características das amostras, este estudo usou o ultrassom como um modificador de propriedades físicas.

O ultrassom pode ser utilizado em estudos para identificação de adulteração do leite desnatado por determinados produtos como a melamina, como ocorreu na China em 2008, ao relacionar a velocidade de propagação, que usou a diferença entre os picos de emissão e recepção, e a densidade das amostras (ELVIRA et al., 2009). Os valores de densidade e percentual de melanina adicionada, bem como o de velocidade foram relacionados.

Outro estudo Nazario et al. (2009) e Nazario et al. (2006) desenvolveu a própria célula de medição, analisou leite UHT quanto à velocidade de propagação e o coeficiente de atenuação e relacionou-os a densidade e o teor de gordura. A atenuação foi medida levando em consideração os ecos do sinal e a reflexão na interface entre a amostra e o refletor. Na densidade, levou em consideração a linha de retardo do vidro e a velocidade de propagação. Usou redes neurais para relacionar os dados coletados, porém faltaram mais experimentos para melhorar a resolução, pois classificaram o leite em apenas cinco concentrações.

Em Savaroglu e Aral (2007), a velocidade de propagação da onda e a densidade do leite foram associadas, e deram origem ao cálculo da impedância acústica, a compressibilidade isentrópica e a função de som molar de Rao (VIGAUREUX, 1952). Os experimentos foram feitos com amostras de leite adulteradas por bicarbonato de sódio e água oxigenada (peróxido de hidrogênio). Neste caso, foi necessário um equipamento auxiliar para o cálculo da densidade do meio. Relacionaram-se todas as variáveis de acordo com o grau de adulteração e descobriu-se a dependência entre a concentração de adulterantes, a temperatura e as variáveis escolhidas para análise. O ponto frágil da pesquisa é o envolvimento de um equipamento a mais na experimentação, que pode ser evitado usando-se outras variáveis para análise, além da baixa variação na densidade que pode ser muito baixa levando em consideração a variabilidade da composição do leite.

Em Dorabiato (2007), foi desenvolvida uma placa acoplada a um sistema ultrassônico para qualificar amostras de leite de acordo com o índice de gordura, e comparou com analisadores comerciais. Com uso da velocidade das amostras é possível calcular a concentração de gorduras totais e extrato seco desengordurado através de um sistema de equações lineares. Desenvolveu sua própria célula de medição e avaliou leite UHT de acordo com índice de gordura relacionado à velocidade e atenuação, sendo que o transdutor se mostrou sensível a 0,1% de gordura. A vantagem é a exatidão da informação oferecida, porém não identifica padrões de adulterações assistidas. Foi utilizado um equipamento comercial para comparar as medições realizadas com ultrassom e as propriedades físico-química dos materiais analisados. Também houve a adição de água gradualmente nas amostras.

Outra possibilidade do ultrassom é acompanhar o crescimento de micro-organismos no leite UHT dentro da embalagem original e sem precisar rompê-la, com a análise da velocidade de propagação e coeficiente de atenuação (ELVIRA et al., 2006). A limitação deste estudo é que apenas pode ser feito com leite em embalagens Tetra Pack. O cálculo da velocidade foi feito com uso da mesma técnica de Elvira et al. (2009), e para a atenuação foi comparado a amplitude do sinal. Ambas as variáveis foram relacionadas com os níveis de evolução de bactérias nas amostras.

Já Dukhin, Goetz e Travers (2005) usou um sistema de ultrassom com transmissão-recepção, e encontrou padrões de reconhecimento para a quantidade e tamanho de partículas de gordura no leite e manteiga, com dependência da frequência, temperatura, velocidade de propagação – definida pela razão entre tempo e distância percorrida - e coeficiente de atenuação, que foi calculado pela diferença de amplitude do sinal inicial e final. As variáveis foram relacionadas de acordo com o tipo de amostra e a quantidade de gordura. Não houve experimentos com adulterações.

Mohanan et al (2002) analisou leite adulterado com carbonato de sódio, bicarbonato de sódio e formol. Neste trabalho foram isoladas a densidade e velocidade. Foram utilizadas as variáveis: velocidade acústica específica de Rao, impedância acústica, compressibilidade adiabática e densidade. Foram isoladas cada variável e comparada de acordo com a temperatura e o grau de adulteração.

Já Bacaneli (1999) também usou velocidade e atenuação com a própria célula de medição, e também verificou a dependência de ambas com a variação na distribuição e tamanho de partículas. A atenuação foi definida pelo coeficiente de reflexão na interface

amostra- refletor, enquanto que a velocidade levou em consideração a medição do intervalo de tempo entre dois ecos consecutivos. Com a sua célula, apenas o coeficiente de atenuação se mostrou sensível às alterações na quantidade de gordura das amostras. Não houve teste com nenhum tipo de adulterante.

Mc Clements (1997) mostrou as formais mais comuns de análise de alimentos por ultrassom geralmente, que utilizam o coeficiente de atenuação e a velocidade de propagação para caracterização dos meios. Também cita a importância de um sistema de tomada de decisão acoplado, baseado em inteligência artificial (como a lógica fuzzy e redes neurais) para processamento dos dados, baseado em uma gama de experimentos já realizados anteriormente voltados ao setor alimentício.

O estudo referente ao ultrassom na caracterização de amostras foi feito por Miles, Shore e Langley (1990), usou transdutores de 1 a 7 MHz para análise do leite integral, desnatado e semidesnatado, sendo UHT ou pasteurizado e homogeneizado. Este é o material primário sobre o assunto e mais citado dentre as pesquisas posteriores. Para a velocidade, usou-se a frequência angular e o volume fracionado de partículas, enquanto que para a atenuação levou-se em consideração compressibilidade adiabática, viscosidade e atenuação termal. Concluiu-se que a atenuação sofreu interferências da quantidade de gordura, da temperatura da amostra e do grau de homogeneização. Também não houve o estudo de adulterantes.

Foram encontrados também algumas revisões sobre os diversos métodos identificação de adulteração no leite, como em Poonia et al. (2016) e Das et al. (2016). Os trabalhos discutidos usam diversas técnicas, sendo que Das et al. (2012) apresentaram um equipamento com sensibilidade à impedância acústica que foi o único que mais se aproximou dos estudos aqui mostrados. Usou a ângulo da constante de fase para caracterização das amostras.

Apenas alguns dos estudos encontrados trabalharam com a caracterização de amostras adulteradas de forma assistida. Além disso, a maior parte dos estudos usou a velocidade de propagação e o coeficiente de atenuação de ondas como as variáveis acústicas selecionadas. Ambas são suficientes para a caracterização do leite e seus derivados, desde o acompanhamento da evolução de micro-organismos até a identificação de material não desejado no meio. Foi possível identificar que há poucos estudos sobre adulteração, um problema real em diversos países e que afeta a população de forma direta; o que indica uma área a ser explorada em estudos posteriores.

CONCLUSÕES

O ultrassom é um método frequentemente usado na avaliação de alimentos, inclusive para análise de leite e derivados. Por ser um meio não invasivo e de resposta rápida, é uma alternativa às técnicas existentes. A avaliação do leite é feita a partir de medições que levam a velocidade de propagação e a atenuação das ondas. Em seguida, por meio de processamentos apropriados são estabelecidas correlações entre estas variáveis e a adulteração presente. O equipamento utilizado baseia-se na técnica de pulso-eco. Apesar de ser um problema cotidiano, a investigação da adulteração de leite com ultrassom ainda é pouco explorada e, com isso, abre boas possibilidade para futuros estudos.

REFERÊNCIAS

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Regulamento da Inspeção Industrial e Sanitária dos Produtos de Origem Animal-RIISPOA. Brasília – DF, 1997.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. IN 62-2011, Brasília – DF, 2011.

CHRISTENSEN, D. A. Ultrasonic Bioinstrumentation. Harper & Row, 1988. p. 256.

MILES, C. A, Shore, D. e Langley, K. R. Attenuation of ultrasound in milks and creams. Ultrasonics, v. 28, n. 6, p. 394 – 400, 1990.

ALVES, L. B., Poletti, E. C. C., Arnold, F. J. Método de identificação de leite cru adulterado via ultrassom e regras fuzzy. Biomatemática (UNICAMP), v. 25, p. 31-44, 2015.

GLAESER, H. Control of the water content of dairy products – definition of limits, consideration of process variation, official use autocontrol data. Food Chemistry, v.82, p.121-124, 2003.

INTERNATIONAL DAIRY FEDERATION - IDF. Milk. Determination of freezing point. Thermistor cryoscope method (reference method). 108:2002/ISO 5764. Brussels: IDF International, 2002, p. 12.

OLIVEIRA, W.P.S.; OLIVEIRA, A.N.; MESQUITA, A.J. et al. Avaliação da metodologia de infravermelho para determinação índice crioscópico do leite cru. In: Cong Bras de Qual do Leite, 2., 2006.

COITINHO, T., Cassoli, L. D., Cerqueira, P. H. R. C., Silva, H. K., Coitinho, J. B., Machado, P. F. Adulteration identification in raw milk using Fourier transform infrared spectroscopy. J. Food Sci and Tech. p. 1-9, 2017.

FERNANDES, J. C. Acústica, Ruídos e Perda de Audição. In: Proceedings of the 9th Brazilian Conference on Dynamics Control and their Applications, Anais...Serra Negra: DINCON'10, 2010.

BACANELI, F. Determinação do teor de gordura no leite através de parâmetros acústicos. 4º Simpósio Brasileiro de Automação Inteligente, UFRGS: São Paulo, 1999, p. 609–614.

POVEY, M. J. W. Ultrasonic determination of the properties of food dispersions. In: Seminars in Food Analysis, 4. 1999, Londres: Chapman and Hall, 1999, p. 95–111.

NAZARIO, S. L. S., Kitano, C., Isepon, J. S. e Higuti, R. T. Obtenção do teor de gordura do leite através de técnicas de ultra-som e redes neurais. Brazilian Conf on Dynamics Control, São Paulo UNESP, 2006, p. 1-6.

DORABIATO, L. F., Maia, J. M. e Gamba, H. R. Caracterização de leite bovino por ultra-som. In: XX Congresso de Engenharia Biomédica, São Pedro: USP, 2006, p. 788- 791.

ELVIRA, L., Durán, C., Sierra, C., Resa, P., Montero de Espinosa, F. Ultrasonic measurement device for the characterization of microbiological and biochemical processes in liquid media. Meas Sci Technology, v. 18, p. 2189–2196, 2007.

ASSEF, A. A., Maia, J. M., Gewehr, P. M., Gamba, H. R., Costa, E. T. e Button, V. L. S. N. Sistema para geração, aquisição e processamento de sinais de ultra-som. Sba: Controle e Automação - Sociedade Brasileira de Automação, v. 20, p. 145 – 155, 2009.

KINSLER, L. E., Frey, A. R., Coppens, A. B., Sanders, J. V. Fundamentals of Acoustic. John Wiley & Sons, Inc., 2000.

ALOUACHE, B., Touat, A., Boutkedjirt, T., Bennamane, A. Monitoring of Lactic Fermentation Process by Ultrasonic Technique. Physics Procedia, v. 7, p. 1057 – 1060, 2015.

IZBAIM, D., Faiz, B., Moudden, A., Malainine, M., Aboudaoud, I. Ultrasonic characterization of yogurt fermentation process. Societè Francaise d'Acoustique. Acoustics, Nantes, France, p. 455-459, 2012.

BOSILJKOV, T., Tripalo , B., Brnčić ,M., Ježek D., Karlović , Jagušt S. I. Influence of high intensity ultrasound with different probe diameter on the degree of homogenization (variance) and physical properties of cow milk. African Journal of Biotechnology, v. 10, n. 1, p. 34–41, 2011.

ELVIRA, L., Lynnworth, J. Rodriguez. L. C. Sound speed and density characterization of milk adulterated with melamine. The Journal of the Acoustical Society of America, v. 125, n.5, p. 177-182, 2009.

NAZARIO, S. L. S., Isepon, J. S., Buiochi, F., Adamowski, J. C., Kitano, C., and Higuti, R. T. Caracterização de leite bovino utilizando ultra-som e redes neurais artificiais. Sba Controle e Automação, v. 20, p. 625-634, 2009.

VIGAUREUX, P. Ultrasonics. Londres: Chapman & Hall, 1952, Cap. 4.

SAVAROGLU, G. e Aral, E. Acoustic parameters of cow's milk added hydrogen peroxide and sodium bicarbonate at different temperatures. J Food Eng, v. 79, p. 287–292, 2007.

ELVIRA, L., Sierra, C., Resa, P., Montero de Espinosa, F., Matesanz, J., Gómez-Ullate, Y., Resa, P., Echevarría, F., Iglesias, J. Eight-channel ultrasonic device for non-invasive quality evaluation in packed milk, Ultrasonics, v. 45, p. 92–99, 2006.

DUKHIN, A. S., Goetz P. J. e Travers B. Use of Ultrasound for Characterizing Dairy Products. Journal of Dairy Science, v. 88, p. 1320–1334, 2005.

ADAMOWSKI, J. C., Furukawa, C. M., Higuti, R. T. e Buiochi, F. Sensores: Tecnologias e Aplicações In Caracterização de Líquidos por Ultra-Som. 1 ed. São Paulo: Nobel, 2004, Cap. 4, p. 98-148.

MOHANAN, S., Panicker P. G., Iype, L., Laila, M., Domini, I. e Bindu, R. G.A new ultrasonic method to detect chemical additives in branded milk. Pramana Journal of Physics, v. 59, n. 3, p. 525-529, 2002.

MC CLEMENTS, D. J. Ultrasonic characterization of food and drinks: Principles, methods and applications. Critical Reviews in Food Science and Nutrition, v. 37, p.1– 46, 1997.

MILES, C. A, Shore, D. e Langley, K. R. Attenuation of ultrasound in milks and creams. Ultrasonics, v. 28, n. 6, p. 394 – 400, 1990.

POONIA, A., Jha, A., Rai, Ak., Sharma, N. Detection of adulteration in milk: A review. International Journal of Dairy Technology, v. 69, p. 1-20, 2016.

DAS S., Sivaramakrishna M., Biswas K. and Goswami B. Performance of a constant phase angle based impedance sensor to detect milk adulteration. Sensors and Actuators A: Physical, v.167, p. 273–278, 2012.

DAS, S., Goswami, B., Biswas, K. Milk Adulteration and Detection: A Review. Sensor Letters. v. 14, n.1, p. 4-18, 2016.

Risco Potencial de Lixiviação do 2,4-Diclorofenoxiacético no Cultivo do Café Conilon (*Coffea canephora*)

Samira Luns Hatum de Almeida[1], *Alexandre Rosa dos Santos*[2], *Gleissy Mary Amaral Dino Alves dos Santos*[2], *Samuel Ferreira da Silva*[2], *José Romário Carvalho*[2]

[1] Departamento de Engenharia Agrícola, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brasil, samiraluns@hotmail.com

[2] Departamento de Engenharia Rural, Universidade Federal do Espírito Santo, Alegre, Espírito Santo, Brasil, mundogeomatica@yahoo.com.br, gleissym@yahoo.com.br, samuelfd.silva@yahoo.com.br, jromario_carvalho@hotmail.com

RESUMO

O Estado do Espírito Santo é o maior produtor nacional de café conilon, representando aproximadamente 35% do Produto Interno Bruto (PIB) do Estado. Contudo, está produtividade pode ser prejudicada de forma considerável pela presença de plantas daninhas nas áreas cultivadas. Buscando mitigar tais danos, utilizam-se herbicidas com o intuito de controlar possíveis perdas econômicas. Neste cenário, esses agrotóxicos representam um alto potencial de contaminação do solo e dos sistemas aquáticos, por meio, principalmente, da lixiviação dos princípios ativos contidos em sua composição. Neste contexto, objetivou-se avaliar o potencial de risco de lixiviação de agrotóxico em áreas de aptidão edafoclimática para a cultura do café conilon (*Coffea canephora* Pierre ex Froehner) no estado do Espírito Santo, utilizando modelo Fator de atenuação/Fator de Retardamento (Attenuation Factor/ Retardation Factor - AF/RF). Para alcançar tais objetivos, foram definidas as áreas de aptidão edafoclimática para a cultura do café conilon no estado e, posteriormente, estimou-se o fator do potencial de adsorção, através do fator de retardamento, e em sequência, por meio do fator de atenuação, permitindo estimar o risco de lixiviação do princípio ativo 2,4-diclorofenoxiacético (2,4D). Com base nos resultados obtidos foi possível concluir que o composto estudado apresentou acima de 96% do território do Estado risco Muito Baixo de lixiviação, variando de acordo com a profundidade fixada e as lâminas de irrigação utilizadas.

PALAVRAS-CHAVE: Modelagem, Fator de Atenuação e Retardamento, Sistemas de Informações Geográficas, Agrotóxicos.

ABSTRACT

The state of Espírito Santo is the largest producer of conilon coffee in Brazil, being responsible for almost 35% of the State Gross Domestic Income. However, this productivity can be significantly damaged by the presence of weeds in the cultivated areas. In order to mitigate such damages, herbicides are used in order to control possible economic losses. In this scenario, these pesticides represent a high potential for soil and aquatic systems contamination, mainly through the leaching of the active principles contained in the pesticides composition. In this context, the objective of this study was to evaluate the potential of agrochemical leaching in areas of edaphoclimatic aptitude for conilon coffee (*Coffea canephora* Pierre ex Froehner) in the state of Espírito Santo, using the Attenuation Factor / Retardation Factor - (AF/RF). In order to reach these objectives, initially, areas of edaphoclimatic aptitude were defined based on the analyzed crop, and later, the adsorption potential factor was estimated by the retardation factor, and in sequence, by the attenuation factor, allowing the estimation of the risk of leaching of the active principle 2,4-dichlorophenoxyacetic (2,4D). Based on the results obtained, it was possible to conclude that the compound studied showed very low leaching risk in above 96% of the State territory, varying according to the determined depth and the irrigation plates used.

KEYWORDS: Modeling, Attenuation and Retardation Factor, Geographic Information Systems, Pesticides.

INTRODUÇÃO

No cenário atual, o Brasil destaca-se como o maior produtor mundial de café, sendo esta, uma atividade considerada de grande importância no âmbito econômico e social. Principal produtor de café conilon (*Coffea canephora* Pierre ex Froehner), o Estado do Espírito Santo é responsável por 76% da produção nacional desta espécie (INCAPER, 2009). A atividade cafeeira representa 35% do PIB agrícola do Estado (INCAPER).

Um fator determinante no desenvolvimento do cafeeiro é a presença de plantas daninhas na área. Sua influência negativa se dá principalmente pela competição pelos recursos do meio, como água, luz e nutrientes. Além disso, as plantas infestantes podem liberar substâncias alelopáticas, além de atuar como hospedeiras de pragas e doenças comuns à cultura (KUVA et al., 2003). Visando o controle dessas plantas daninhas, a utilização de produtos químicos tem se intensificado nos últimos anos (PAZ; RUBIO, 2006).

A lixiviação de pesticidas em solo está diretamente relacionada às características físico-químicas do solo e do próprio agrotóxico (KELEMEN; BEHN, 2016). Outros fatores que podem levar ao aumento da contaminação de águas subterrâneas por lixiviação são o manejo do solo e métodos de irrigação das lavouras (BOZZO et al., 2001).

Nesse contexto, buscou-se com este trabalho, avaliar o potencial de risco de lixiviação de agrotóxicos em áreas em que o clima e o solo são favoráveis ao desenvolvimento do cafeeiro, ou seja, áreas de aptidão edafoclimática para a cultura do café conilon no estado do Espírito Santo, por meio do emprego do modelo AR/RF (RAO et. al., 1985), utilizando técnicas de geotecnologia.

O trabalho em questão foi estruturado em mais cinco seções, sendo elas: Material e Métodos, em que foi abordado de que forma foi realizado o trabalho, definindo a área de estudo, assim como as etapas subsequentes, zoneamento edafoclimático da região, estimação do fator de retardamento e fator de atenuação. Em seguida, os resultados de cada etapa foram analisados e discutidos, gerando uma conclusão. O trabalho contém ainda agradecimentos e as referências utilizadas para a realização e escrita do mesmo.

MATERIAL E MÉTODOS

Área de estudo

A área de estudo é representada pelo Estado do Espírito Santo, com uma área territorial de 46.053,19 km². Está localizado entre os paralelos de 17°53'29" a 21°18'03" de latitude Sul e os meridianos 39°41'18" a 41°52'45" longitude Oeste de Greenwich.

Zoneamento edafoclimático para o café conilon (Coffea canephora Pierre ex Froehner)

Para a realização deste estudo, foi utilizada uma série de dados meteorológicos de 30 anos, representativa de 109 estações localizadas no Estado do Espírito Santo e áreas limítrofes. A base de dados meteorológica foi importada para o aplicativo computacional *Microsof Office Excel*®, versão 2010, sendo processado o balanço hídrico agroclimatológico proposto por Thornthwaite e Matter (1955) com o objetivo de gerar uma planilha eletrônica. Planilhas essas posteriormente importadas para o aplicativo computacional ArcGIS®, versão 10.3.

Em seguida, realizou-se uma regressão linear múltipla (RIBEIRO-JUNIOR, 2011), utilizando a altitude e as coordenadas UTM como variáveis independentes, além da temperatura como variável dependente. De posse dos campos das coordenadas *UTM X* e *Y* representativas das estações meteorológicas, foi aplicada a interpolação espacial pelas tendências com o objetivo de gerar as imagens matriciais das coordenadas *UTM X* e *Y*.

Aplicou-se álgebra de mapas, tendo como entrada as equações de regressões lineares múltiplas com suas respectivas variáveis independentes representadas pelas imagens matriciais do Modelo Digital de Elevação (MDE) /*Shuttle Radar Topography Mission* (SRTM), coordenadas *UTM X* e *Y*, obtendo como saída, as imagens matriciais de temperaturas médias mensais e, posteriormente, a temperatura média anual.

Para obter imagens matriciais de déficit hídrico anual para as culturas do café conilon, aplicou-se a interpolação espacial por krigagem (MAZZINI; SCHETTINI, 2009), que considera a dependência espacial entre amostras vizinhas, para estimar valores dentro do espaço analisado (MOTOMIYA et al., 2006). Sobre as imagens matriciais de temperatura média anual e déficit hídrico anual, foi aplicada a função de "reclassificação espacial" com o objetivo de representar as classes de aptidão, restrição e inaptidão para o café conilon (OMENA, 2014).

Tabela 1 - Classes de aptidão climática para culturas agrícolas

Classe	Característica
Apta	Quando as condições térmicas e hídricas da área apresentam-se favoráveis para o bom desenvolvimento e produção da cultura em escala comercial.
Restrita	Quando a área apresenta restrições hídricas ou térmicas, ou ambas, que podem eventualmente prejudicar as fases de desenvolvimento da cultura, repercutindo negativamente em sua produção.
Inapta	Quando as características normais do clima não se apresentam adequadas à exploração comercial da cultura, por apresentarem limitações severas quanto aos fatores hídricos ou térmicos, ou ambos, com marcante repercussão em sua produção, exigindo, para que sejam corrigidas, práticas agrícolas dispendiosas.

Fonte: Adaptado de Zolnier (1994); Santos (2017)

Avaliação do potencial de lixiviação empregando o método RF/AF

Este método, proposto por Rao et al. (1985), foi empregado para estimar o retardo da lixiviação do composto orgânico relativo à percolação da água (RF) e a atenuação estimada por meio do aporte relativo de massa de um composto orgânico na água subterrânea (AF).

- *Distribuição espacial da taxa de recarga das águas subterrâneas*

O banco de dados utilizado para o zoneamento edafoclimático foi novamente utilizado para obtenção de dados de precipitação e a evapotranspiração da cultura. Em seguida foi aplicada interpolação espacial por krigagem (MAZZINI; SCHETTINI, 2009), com o objetivo de gerar as imagens matriciais dos dados em questão. Com a utilização da calculadora Raster aplicou-se a equação 1 para a obtenção da taxa de recarga (q), que leva em consideração a precipitação

(P), evapotranspiração (ETC), e irrigação (I). Para a irrigação, foram assumidas duas lâminas de irrigação anual, sendo elas 600 e 1200 mm.

$$q = P + I - ETC \quad \text{(Equação 1)}$$

- *Distribuição espacial do fator de retardamento*

O fator de retardamento é um modelo utilizado para estimar o potencial de adsorção, ou seja, avaliar o potencial do composto químico de ficar retido nas partículas do solo. Para sua estimação, utilizou-se o mapa de Distribuição e Classificação de Solos do Brasil (Embrapa, 2013). Mapa este que foi editado de modo a adicionar informações oriundas de pesquisa bibliográfica, relacionadas às características físicas e químicas do solo (ρ: densidade do solo em gcm^{-3}; OC: teor de carbono orgânico do solo em gg^{-1}; δ: porosidade do solo na capacidade de campo em vv^{-1}; Fc: capacidade de campo do solo em vv^{-1}) e também propriedades dos agrotóxicos (Koc: coeficiente de adsorção ao carbono orgânico em cm^3g^{-1}; K_H: coeficiente de partição ar-água do agrotóxico em cm^3g^{-1}).

Com base na equação 2, e através da utilização da calculadora raster, estimou-se o valor de fator de Retardamento (RF). Em seguida, sobre o RF calculado aplicou-se função de "reclassificação espacial", gerando cinco classes (muito baixo, baixo, médio, alto e muito alto), conforme descrito por Mattos e Silva (1999).

$$RF = 1 + \left(\frac{\rho \cdot OC \cdot K_{oc}}{Fc}\right) + \left(\frac{\delta \cdot K_H}{Fc}\right) \quad \text{(Equação 2)}$$

- *Distribuição espacial do fator de atenuação*

Para o cálculo de fator de atenuação foram consideradas duas profundidades distintas, sendo elas a profundidade de água subterrânea a partir da superfície, e a profundidade efetiva do sistema radicular do café conilon, assumida como 0,30 m. Com base na capacidade de campo (Fc - em vv^{-1}) na recarga anual (q - em $mdia^{-1}$), profundidade definida (L - em m), meia vida do pesticida ($t_{½}$ - em dia) e o fator de retardamento estimado (RF - adimensional), calculou-se a estimativa do fator de atenuação, através da calculadora raster, de acordo com a equação 3. Em seguida, sobre o AF calculado aplicou-se função de "reclassificação espacial", gerando seis classes, sendo elas: nulo, em que AF é igual a 0; muito baixo, AF varia de 0,0 a 0,0001; baixo, de 0,0001 a 0,01; médio, de 0,01 a 0,1; alto, de 0,1 a 0,25; e muito alto, em que AF varia de 0,25 a 1, conforme descrito por Mattos e Silva (1999).

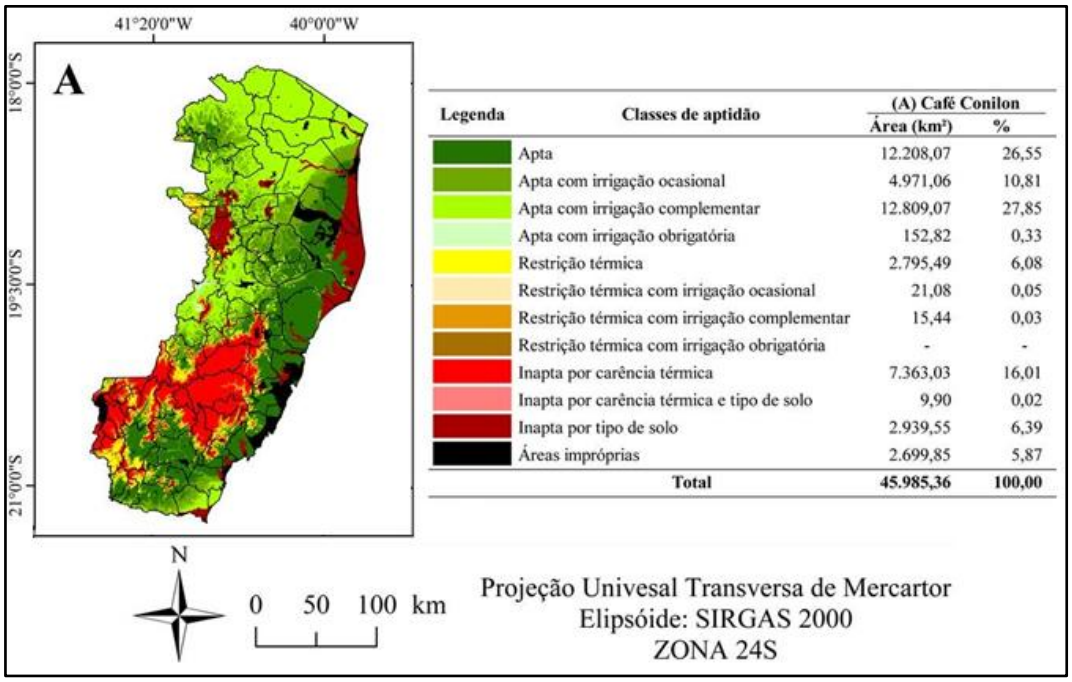
$$AF = \exp\left(\frac{-0{,}693 \cdot L \cdot RF \cdot Fc}{q \cdot t_{½}}\right) \quad \text{(Equação 3)}$$

RESULTADOS E DISCUSSÃO

Por meio do zoneamento edafoclimático para a cultura do café conilon no Estado do Espírito Santo, foi possível identificar áreas consideradas aptas, restritas e inaptas para o cultivo desta cultura (Figura 1). Dentro do zoneamento verificou-se uma área de aptidão de 26,55%. Considerado como inapta observou-se uma área equivalente a 22,42% do Estado. Grande parte destas áreas inaptas encontram-se na região Sul e Serrana do estado, sendo a inaptidão nestas regiões decorrente, principalmente, de baixas temperaturas, que não atendem as exigências térmicas da cultura. Dos 45,16% restantes, verificou-se a ocorrência de áreas com aptidão e restrição térmica com carência hídrica, ou seja, necessidade de irrigação.

Considerando as exigências térmicas da cultura do café conilon, 65,54% da área de estudo possui aptidão térmica para o cultivo. Porém, grande parte dessas áreas não atende as necessidades hídricas da cultura. Dessa forma, para serem consideradas áreas aptas atendendo as exigências térmicas e hídricas, devem ser manejadas com irrigação. As áreas em que se encontra maior déficit hídrico durante o ano, deve-se utilizar a irrigação obrigatória, e isso ocorre em 0,33% do Estado. À medida que se reduz esse déficit podem ser utilizadas irrigação complementar, e a irrigação ocasional.

Figura 1. Zoneamento Edafoclimático da cultura do café conilon.



Legenda	Classes de aptidão	(A) Café Conilon Área (km²)	%
	Apta	12.208,07	26,55
	Apta com irrigação ocasional	4.971,06	10,81
	Apta com irrigação complementar	12.809,07	27,85
	Apta com irrigação obrigatória	152,82	0,33
	Restrição térmica	2.795,49	6,08
	Restrição térmica com irrigação ocasional	21,08	0,05
	Restrição térmica com irrigação complementar	15,44	0,03
	Restrição térmica com irrigação obrigatória	-	-
	Inapta por carência térmica	7.363,03	16,01
	Inapta por carência térmica e tipo de solo	9,90	0,02
	Inapta por tipo de solo	2.939,55	6,39
	Áreas impróprias	2.699,85	5,87
	Total	**45.985,36**	**100,00**

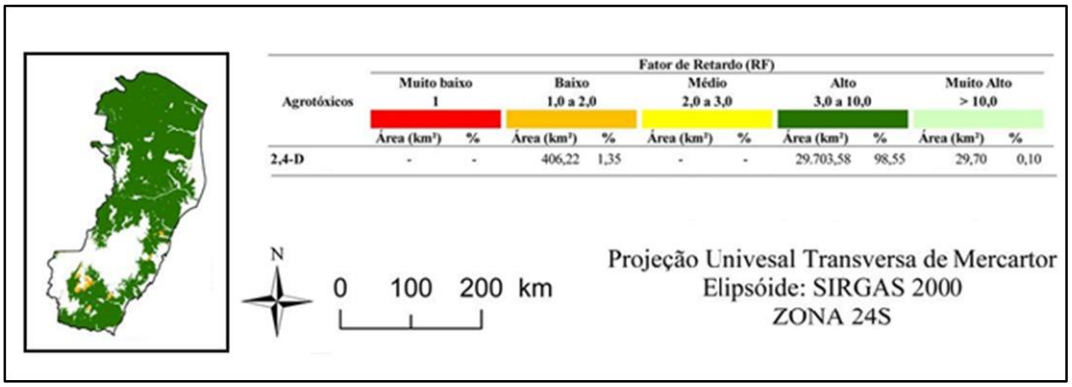
Fonte: Elaborado pelos autores.

Fator de Retardamento

Ao calcular o fator de retardamento (Figura 2), são obtidos valores que vão do intervalo de um a infinito. Quanto maior as estimativas de RF supõe-se que mais partículas do produto químico irão ficar retidas nas partículas do solo. Logo, reduzem-se as chances deste produto ser perdido através do perfil do solo, por meio da lixiviação.

Para o composto químico 2,4D 98,55% da área do Estado apresentou potencial de adsorção variando entre 3 e 10, sendo classificada como Alto. 0,10% do território apresentou RF muito alto, acima de 10. Outros 1,35% do Estado foram classificados como Baixo. Essa variação de classes de potencial de adsorção do herbicida 2,4D foi em função do tipo de solo. As regiões do Estado que apresentam Neossolo Litólico, foram os que apresentaram valores mais baixos de fator de retardamento, ou seja, nesse solo as partículas ficam menos retidas, sendo mais facilmente lixiviadas. Esse fato ocorreu devido às características físicas no Neossolo, que apresenta em sua composição, uma parcela mais significativa de areia (EMBRAPA, 2013).

Figura 2. Fator de Retardamento para o Estado do Espírito Santo.



Fonte: Elaborado pelos autores.

Fator de Atenuação

Com base na equação de Fator de Atenuação (AF) obtiveram-se mapas com estimativas de AF para quatro situações distintas (Figura 3). Dentre elas, duas levam em consideração a profundidade do lençol freático, com lâminas de irrigação anual de 600 e 1200 mm nas regiões aptas para o cultivo do café conilon. Outras duas, com as mesmas lâminas de irrigação, porém considerando a profundidade de raiz efetiva do cafeeiro,0,30 metros, região onde se encontra grande parte de seu sistema radicular (SILVA et al., 2009).

Com a lâmina de irrigação de 600 mm, na profundidade de 0,30m, o 2,4D apresentou um risco de lixiviação muito baixo em 96,32% do Estado, e um risco baixo em 0,03%. Com o

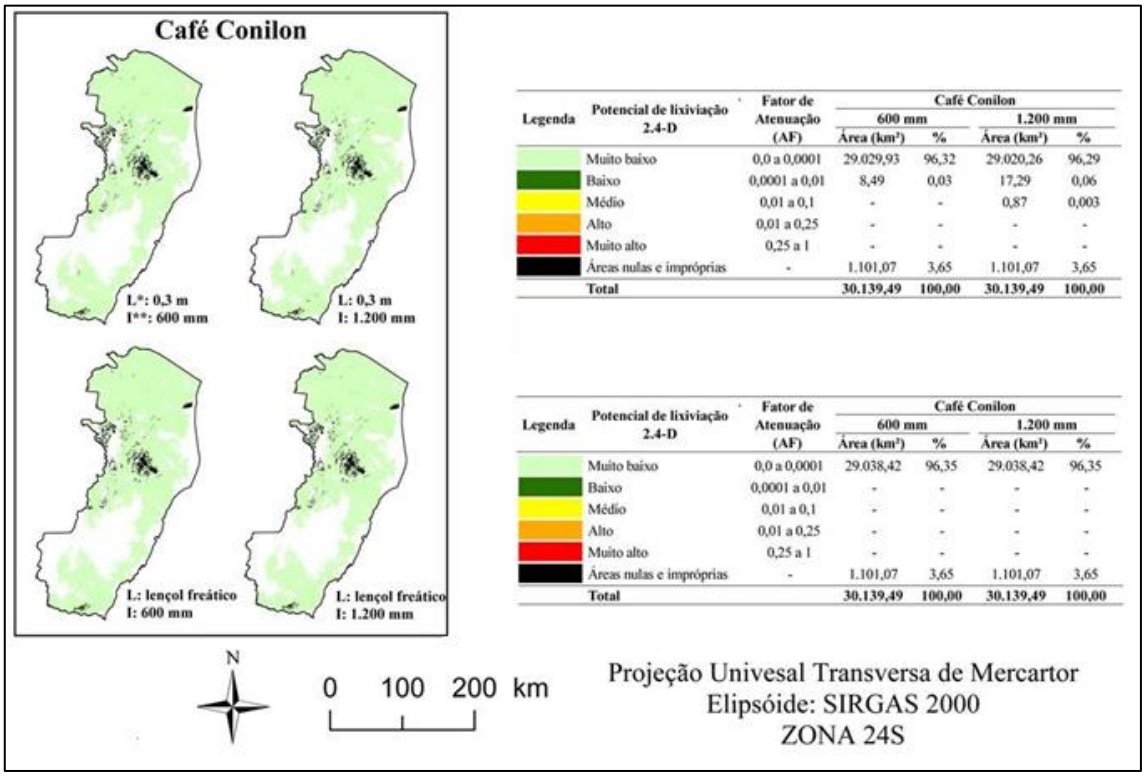
aumento da lâmina 1200 mm houve a presença de mais uma classe de risco, a classe Média, presente em 0,003% da área de estudo. Ainda assim, para grande parte do Estado, 96,29% o composto analisado apresentou uma estimativa de risco de lixiviação muito baixo.

Analisando a profundidade do lençol freático, com o objetivo de estimar o risco do herbicida de atingi-lo, o 2,4 D apresentou um risco muito baixo de lixiviação. Esta classe de risco encontrada refere-se às duas lâminas de irrigação estipuladas, ou seja, mesmo com o aumento da lâmina de irrigação, o 2,4D não apresentou variação de risco ao longo do Estado do Espírito Santo, estando toda a totalidade do território, exceto as áreas impróprias ao cultivo, apresentando uma classe muito baixa de risco de lixiviação.

Lourencetti et al. (2005) por meio de estudo para comparar os modelos de lixiviação, encontrou para o composto 2,4D, resultados compatíveis com o encontrado no presente trabalho. Em um dos modelos que usou, o modelo RF/AF, encontrou valores que classificam o composto com um risco de lixiviação Muito Baixo.

É importante considerar que a ocorrência de determinado pesticida em água subterrânea depende além do seu potencial de lixiviação, da vulnerabilidade natural do aquífero e da quantidade do pesticida aplicado (LOURENCETTI et al., 2005).

Figura 3. Fator de Atenuação do 2,4D para o Estado do Espírito Santo.



Legenda	Potencial de lixiviação 2.4-D	Fator de Atenuação (AF)	Café Conilon			
			600 mm		1.200 mm	
			Área (km²)	%	Área (km²)	%
	Muito baixo	0,0 a 0,0001	29.029,93	96,32	29.020,26	96,29
	Baixo	0,0001 a 0,01	8,49	0,03	17,29	0,06
	Médio	0,01 a 0,1	-	-	0,87	0,003
	Alto	0,01 a 0,25	-	-	-	-
	Muito alto	0,25 a 1	-	-	-	-
	Áreas nulas e impróprias	-	1.101,07	3,65	1.101,07	3,65
	Total		**30.139,49**	**100,00**	**30.139,49**	**100,00**

Legenda	Potencial de lixiviação 2.4-D	Fator de Atenuação (AF)	Café Conilon			
			600 mm		1.200 mm	
			Área (km²)	%	Área (km²)	%
	Muito baixo	0,0 a 0,0001	29.038,42	96,35	29.038,42	96,35
	Baixo	0,0001 a 0,01	-	-	-	-
	Médio	0,01 a 0,1	-	-	-	-
	Alto	0,01 a 0,25	-	-	-	-
	Muito alto	0,25 a 1	-	-	-	-
	Áreas nulas e impróprias	-	1.101,07	3,65	1.101,07	3,65
	Total		**30.139,49**	**100,00**	**30.139,49**	**100,00**

Fonte: Elaborado pelos autores.

CONCLUSÕES

Diante dos resultados obtidos é possível concluir que o princípio ativo estudado apresentou para grande parte do território do Estado do Espírito Santo (acima de 96%), risco Muito Baixo de lixiviação, variando de acordo com a profundidade fixada e as lâminas de irrigação aplicadas. Esta metodologia apresenta potencial para ser aplicada de forma eficiente em outras áreas agrícolas ou biomas do mundo.

AGRADECIMENTOS

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes). Ao grupo de pesquisa *Geotechnology Applied to Global Environment* (GAGEN).

REFERÊNCIAS

BOZZO, S., AZIMONTI, G., VILLA, S., DI GUARDO, A., FINIZIO, A.. Spatial and temporal trend of groundwater contamination from terbuthylazine and desethyl-terbuthylazine in the Lombardy Region (Italy). Environmental Science. Processes Impacts. v. 15, p. 366-372, 2013.

EMBRAPA. Empresa Brasileira de Pesquisa Agropecuária. Embrapa Solos. Ministério da Agricultura, Pecuária e Abastecimento. Sistema Brasileiro de Classificação de Solos. 3ª edição revisada e ampliada. Brasília, DF: Embrapa, 2013, 353 p.

INCAPER - Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural. Governo estimula renovação das lavouras de café arábica em Dores do Rio Preto. 2009.

INCAPER - Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural. Cafeicultura. Disponível em: < https://incaper.es.gov.br/cafeicultura >.

KELEMEN, P. B., BEHN, M. D. Formation of lower continental crust by relamination of buoyant arc lavas and plutons. Nature geoscience. v. 9, p. 197-205, 2016.

KUVA, M. A.; GRAVENA, R.; PITELLI, R. A.; CHRISTOFFOLETI, P. J.; ALVES, P. L. C. A. Períodos De Interferência Das Plantas Daninhas Na Cultura Da Cana-De-Açúcar. III – Capim-Braquiária (Brachiaria decumbens) e Capim-Colonião (Panicum maximum). Planta Daninha, Viçosa-MG, v. 21, n. 1, p. 37-44, 2003.

LOURENCETTI, C.; SPADOTTO, C. A.; SILVA, M. S.; RIBEIRO, M. L. Avaliação do potencial de contaminação de águas subterrâneas por pesticidas: comparação entre métodos de previsão de lixiviação. Pesticidas: Revista Ecotoxicologia e Meio Ambiente, v.15, p.1-14, 2005.

MATTOS, L. M.; SILVA, E. F. Influência das propriedades de solo e de pesticidas no potencial de contaminação desolos e águas subterrâneas. Pesticidas: Revista de ecotoxicologia e meio ambiente, Curitiba, v. 9, p. 103-124,1999.

MAZZINI, P. L. F.; SCHETTINI, C. A. F. Avaliação de metodologias de interpolação espacial aplicadas a dados hidrográficos costeiros quasesinóticos. Brazilian Journal of Aquatic Science and Technology, v.13, n. 1, p. 53-64, 2009.

MOTOMIYA, A. V. A.; CORÁ, J. E.; PEREIRA, G. T. Uso da krigagem indicatriz na avaliação de indicadores de fertilidade do solo. Revista Brasileira de Ciência do Solo, vol. 30, núm. 3, 2006, pp. 485-496.

OMENA, M. S. Conjunto de ferramentas computacionais para análises agroclimáticas. Dissertação (Mestrado - Produção Vegetal) – Universidade Estadual do Norte Fluminense Darcy Ribeiro, Centro de Ciências e Tecnologias Agropecuárias, Campos dos Goytacazes, RJ. 2014, 106 p.

PAZ, J. M., RUBIO, J. L. Application of a GIS-AF/AR model to assess the risk of herbicide leaching in a citrus-growing area of the Valencia Community, Spain. Science Of The Total Environment. v. 371, p. 44–54, 2006.

RAO P. S. C., HORNSBY A. G., JESSUP R. E. Indices for ranking the potential for groundwater contamination. Proc - SoilCrop Sci Soc Fla, 44: p. 1–8, 1985

RIBEIRO JR, J. I. Análises estatísticas no Excel: guia prático. Viçosa: UFV, 258p, 2011.

THORNTHWAITE, C. W.; MATHER, J. R. The water balance. Publications in Climatology. New Jersey: Drexel Institute of Technology. 1955. 104p.

SANTOS, G. M. A. D. A. Espacialização do risco de lixiviação de agrotóxicos em áreas de cafeicultura no Estado do Espírito Santo. Tese (Doutorado – Agroquímica) – Universidade Federal de Viçosa, Viçosa. MG. 2017, 32p.

SILVA, A. L.; BRUNO, I. P; REICHARDT, K.; BACCHI, O. O. S.; DOURADO-NETO, D.; FAVARIN, J. L.; COSTA, F. M. P; TIMM, L. C. Soil water extraction by roots and Kc for the coffee crop. Revista brasileira de engenharia agrícola e ambiental. v. 37, n. 3. Campina Grande May/June 2009.

ZOLNIER, S. Zoneamento climático. (Caderno Didático: Série 20). Viçosa, MG.: UFV, 1994.

Modos de troca cognitiva nas redes sociais do agrossistema digital

Alessandro Mancio de Camargo[1]

[1] Pesquisador em nível de Doutorado do Programa de Estudos Pós-Graduados em Tecnologias da Inteligência e Design Digital (TIDD) da Faculdade de Ciências Exatas e Tecnologia da PUC-SP, membro do grupo de pesquisa TransObjeto (vinculado ao CNPq), almancio@uol.com.br

RESUMO

A atividade agrícola tira proveito da emergência de uma série de novas tecnologias, cujos avanços estruturais e técnicos redundaram, desde o século passado, no código digital, na multimídia e na mobilidade interativa. Particularidades desse processo são citadas neste artigo por meio de registros da micro-história, tais como o uso por agricultores do WhatsApp na extensão rural ou hacktivismo em fóruns on-line contra firmware agrícola. No sentido proposto por pesquisadores da Universidade Wageningen, adota-se como estratégia metodológica uma postura tecnográfica, que leva em conta a integração das dimensões técnicas, biofísicas, culturais e econômicas das redes sociais. O problema levantado é: De que maneira sistemas de naturezas diversas (clima, solo, máquinas, pessoas) têm suas trocas cognitivas afetadas pela ubiquidade conectiva? O objetivo é corroborar para que haja uma melhor compreensão das contingências encontradas nas redes sociais do agronegócio.

PALAVRAS-CHAVE: Hacktivismo, *Firmware*, IoT, WhatsApp, Extensão Rural.

ABSTRACT

The paper is an essay on the emergence of digital agriculture and its role in media ecology, particularly in social networks, that presents particularities of this process through microhistory registers, such as the use by farmers of WhatsApp in rural extension or hacktivism in online forums against agricultural firmware. As a methodological strategy, a technography posture is adopted, in the sense proposed by researchers from Wageningen University. Technography is used to study rural social networks because it takes into account the technical, biophysical and local economic and culture dimensions of the agro system. The problem raised is: In what way different systems (climate, soil, machines, people, digital codes) have their relations and cognitive exchanges affected by connective ubiquity? The answer to this problem is important to corroborate with a better understanding of the contingencies seen in the connection to social digital networks of the agribusiness.

KEYWORDS: Hacktivism, Firmware, IoT, WhatsApp, Rural Extension.

INTRODUÇÃO

A interação da espécie humana com o ecossistema ocorre tanto no nível da biosfera quanto no nível da noosfera – mais conhecida como sociedade da informação e baseada nas linguagens

criadas pelo homem –, que orienta o crescente papel do poder cognitivo e tecnológico para conformar o futuro (CRUTZEN; STOEMER, 2000; SANTAELLA, 2015). A agricultura faz parte dessas linguagens, sendo eficaz para multiplicar alimentos por meio da cultura de plantas e criação de rebanhos que realizam trocas com os fluxos de energia, matéria e informação do sistema terrestre a fim de se desenvolver em sintonia com os processos biogeoquímicos, como estações do ano.

De forma cooperativa e competitiva entre si, a relação da agricultura com o ecossistema – que abrange os infindáveis vínculos, mediações e trocas entre a diversidade de "seres vivos" (aves, bactérias, florestas) e "o conjunto das características físicas e químicas" de um habitat, tais como as condições climáticas (LÉVÊQUE, 2001, p. 21) –, ganhou relevância com o assentamento do homem no campo há cerca de 10 mil a 8 mil anos, conforme revelam tecnofósseis (ZALASIEWICZ et al., 2014; MATTHEWS, 2016). A partir daí a atividade agrícola permitiu que pequenos grupos e tribos se acomodassem em espaços geográficos delimitados, de forma a favorecer em cada comunidade a comunicação e o comércio entre seus integrantes, constituídos como sistemas sígnicos (VAN DIJK, 2009).

No decorrer dos séculos, a integração espacial e temporal entre esses primeiros assentamentos, por meio do estabelecimento de redes sociais, adquiriu escala e diversidade. Isso permitiu o surgimento das primeiras vilas e cidades, além de gerar recursos necessários para a emergência da revolução científica (HARARI, 2015). Num processo de grande aceleração e globalização, evidenciado pelo Antropoceno a partir de 1950 (CRUTZEN; STOEMER, 2000; STEFFEN et al., 2015), esse fenômeno já afeta a comunicação, a cognição, a economia, as máquinas, a sociedade e o planeta, em desdobramentos descritos na obra de autores como Santaella (2015), Garnet e Parikka (2016), Zalasiewicz et al. (2016).

Especialmente a partir do século XX, com o desenvolvimento das Tecnologias de Comunicação e Informação (TIC), a integração entre diversos sistemas sígnicos e objetos tecnológicos ocorre de maneira mais constante por meio de infinitas relações de parte a parte possibilitadas pelo código digital e pela mobilidade conectiva, formando um imenso conjunto de nós interconectados, segundo Castells (2006). Essa rede sociotécnica comum assumiu um caráter rizomático de organização. Ou seja: opera sob um sistema dinâmico que remete a imagem de linhas ou filamentos entrecortados, capazes de se movimentar em todas as direções, ao mesmo tempo que estabelecem infinitas relações e agenciamentos entre si, conforme antecipado pela obra de Gilles Deleuze e de Félix Guattari (2000).

Isso constitui o ecossistema midiático que extrapola no século XXI a relação humana com o conjunto formado por animais, plantas, bactérias, água, solo, vento e abrange, também,

o nível cultural e a vinculação de tudo isso por meio de interações sociotécnicas como as criadas pelo advento da agricultura de precisão, da Internet das coisas (IoT); ou seja, de diversos tipos de redes sociais digitais (CAMARGO, 2016a; CAMARGO, 2016b). Em outras palavras, esse ecossistema constitui a sociedade em rede que é um sistema aberto, complexo, influenciado por toda sorte de mudanças que afetam o seu equilíbrio.

Pari passu com essa evolução a tecnologia agrícola absorve tanto conhecimentos tácitos e populares quanto religiosos e científicos (LOWENTHAL, 2016). Nesse quadro, o poder de persuasão da tradição oral ainda ocorre nas relações agrícolas (CAMARGO, 2016b). Tal fato já é analisado desde o estudo clássico na área de extensão rural sobre os fatores que influenciam a adoção de novas tecnologias na agricultura de Ryan e Gross (1943). Mas no século XXI trata-se de uma ampliação da tradição oral com base nas oportunidades criadas para o "agrossistema" (CONWAY, 2003, p. 197-198) com o advento das TIC (AgroTIC).

Isso tem apontado para uma mudança de paradigma no campo, no qual, acredita-se, serão os "dados coletados por tratores e outras máquinas e processados por computadores que *dirão* [grifo meu] aos agricultores como aumentar a produção de culturas como milho e soja" (BUNGE, 2014, p. B13). Mas persistem dúvidas quanto à adoção das AgroTIC. Os agricultores, ciosos do relacionamento milenar com a terra, concluem que há muito valor nas informações que elas coletam; e não está claro quem deterá o monopólio disso após o processamento digital: se eles ou as empresas do agronegócio. Outro embaraço é como cuidar da "transferência" dessa nova tecnologia para o campo (BERNARDI et. al, 2014, p. 548).

Especificamente sobre as redes sociais do agrossistema, o problema que se coloca é:

– De que maneira sistemas de natureza diversa (clima, solo, máquinas, pessoas, códigos digitais) têm suas relações e trocas cognitivas afetadas pela ubiquidade conectiva?

Não há resposta pronta para o problema apresentado. A revisão bibliográfica ressalta que as possibilidades da tecnologia ubíqua (ou pervasiva) ainda não foram avaliadas na totalidade. Exemplarmente no caso da IoT, um campo no qual os efeitos da tecnologia ubíqua são notáveis, a conectividade "é um conceito tão abrangente que há um desafio até mesmo para imaginar todas as maneiras possíveis em que a IoT irá afetar as empresas, as economias e a sociedade" (MANYIKA et. al, 2013, p. 59).

Ao lançar luz sobre esse problema, o artigo tem como objetivo corroborar para que haja uma melhor compreensão quanto às contingências encontradas nos modos de conexão e relacionamento das redes sociais do agronegócio.

MATERIAL E MÉTODOS

Este artigo visa retratar de forma comparativa os desdobramentos da tecnologia ubíqua no agrossistema, principalmente a extensão disso nas redes sociais rurais, nas quais as relações e trocas cognitivas passam a se estabelecer cada vez mais entre objetos/sistemas sígnicos de natureza técnica diversa (pessoas, softwares, máquinas, meio ambiente), tal como já ocorre em sistemas agrícolas digitais e em comunidades rurais com acesso às TIC.

Como estratégia metodológica adota uma postura tecnográfica. A tecnografia pode ser resumida como uma etnografia da tecnologia, segundo pesquisadores do departamento de desenvolvimento agrário e tecnológico da Universidade Wageningen, da Holanda (JANSEN; VELLEMA, 2011). A tecnografia se assemelha, porém, não é idêntica, a teoria ator-rede, cujo estado da arte está descrito no livro Ciência em Ação do antropólogo, sociólogo e filósofo da ciência Bruno Latour (2000).

A interdisciplinaridade envolvida nas relações do agrossistema resulta do emprego em suas atividades tanto de conhecimentos tácitos e populares quanto religiosos e científicos. Para estudar a natureza dessa linguagem complexa a tecnografia leva em conta que:

> A agricultura integra a dimensão técnica (como o tipo de práticas de lavoura do solo que reduzem a erosão do solo) com a dimensão social e econômica (a maneira como os agricultores mobilizam e mantêm uma força de trabalho ou selecionam as culturas de acordo com as condições do mercado). Embora saibamos pesquisar cada uma dessas áreas separadamente, há pouco consenso sobre como projetar pesquisas que integrem os processos biofísicos e a determinação social das práticas tecnológicas envolvidas (JANSEN; VELLEMA, 2011, p. 169).

Nesse sentido, “a tecnografia pode ser considerada como uma ciência social descritiva da tecnologia que examina a interação homem / máquina ou ferramenta” com o ecossistema (JANSEN; VELLEMA, 2011, p. 169). Assim como esses autores, emprega-se aqui o termo tecnologia no sentido de objetos técnicos, artefatos ou linguagens que resultam da aplicação prática de descobertas humanas e científicas. No caso da tecnografia, o interesse primário recai na explicação do impacto de uma tecnologia em seu momento de uso, principalmente. E pressupõe as trocas sistêmicas e semióticas envolvidas nisso (SANTAELLA; VIEIRA, 2008).

“O foco principal da pesquisa recai nos processos, conhecimentos e habilidades envolvidos na tecnologia entendida como a capacidade humana de fazer” (JANSEN; VELLEMA, 2011, p. 169). A estratégia metodológica também compreende a observação não-participante de material intersemiótico sobre o uso de redes sociais digitais no ambiente rural. Isso é visto, por exemplo, no caso do hacktivismo mediado por fóruns on-line para burlar contrato de licença de usuário final de equipamentos agrícolas de precisão, estimulado por

uma cultura estreitamente relacionada à economia política de uso das TIC (GARNET; PARIKKA, 2016).

WhatsApp, fóruns on-line e redes sociotécnicas sustentáveis

O que se concebe como redes sociais é, principalmente, fruto de um processo evolutivo iniciado no fim do século XIX e ampliado com o surgimento das três últimas eras culturais – de massa, das mídias e digital – descritas por Santaella (2003). Entre os séculos XX e XXI, esse processo acelerou-se por meio de uma infraestrutura de redes social e de mídia que permite um modo especial de organização, cada vez mais abrangente, baseado na:

> Incessante evolução dos dispositivos tecnológicos [que] possibilita que paulatinamente os computadores, como estamos acostumados a conhecê-los, desapareçam de nossas vistas e cedam espaço a tecnologias pervasivas intercomunicantes, instaurando uma nova ecologia comunicativa em que objetos deixam sua posição de suporte para a ação do homem e se fazem notar como seres [objetos] sencientes capazes de estabelecer diálogos com o humano e entre si. […]. Do ponto de vista estrutural, a comunicação entre máquinas possibilitou a expansão do conhecimento técnico, barateando a criação de tecnologias que hoje são utilizadas para impulsionar o desenvolvimento da Internet das Coisas [IoT]. (SANTAELLA et. al, 2013, p.28)

Sob o signo da ubiquidade (SANTAELLA, 2013), tem sido afirmado que a evolução dessas tecnologias produzirá uma "verdadeira inteligência das máquinas e a conexão dela a todos os seres humanos via uma rede digital comum, transformando a economia do planeta" (BRYNJOLFSSON; McAFEE, 2014, p. 250).

As características particulares desse rizoma, no sentido de Deleuze e Guattari (2000), já afetam tanto as relações entre diferentes objetos/sistemas sígnicos nas cidades (*smart cities*) quanto no campo, por meio da agricultura digital (FRAGALLE e SILVA, 2014). No meio rural verifica-se importantes mudanças de comportamento devido à adoção de inovações no processo de pré-produção, produção e pós-produção de alimentos (MASSRUHÁ et. al, 2014).

Registros da micro-história revelam como comunidades agrícolas têm se organizado de forma cooperativa e competitiva por meio das redes sociais digitais. Exemplo disso é o uso do aplicativo WhatsApp, conforme registra o The Global Forum on Agricultural Research:

> Existem cerca de 250 agricultores brilhantes, comprometidos, habilidosos e profissionais, que fazem parte de uma plataforma que é apoiada pela Rede de Engajamento de Jovens em Agricultura (YEAN) (https://yeanrwanda.org/), um parceiro do GFAR. Esta organização partilha informação agrícola e permite a prestação de serviços de extensão agrícola a estes jovens agricultores […]. Com a plataforma YEAN, as informações são compartilhadas digitalmente sobre como cultivar a cultura, como comprar ou vender produtos e até como importar ou exportar produtos agrícolas. É uma maneira fácil e inteligente de conectar os

> agricultores [africanos] a outros agricultores e aos compradores e formuladores de políticas. A rede social digital está influenciando fortemente esta nova geração de agricultores profissionais. Um amigo meu realmente comprou sementes de mamão do Brasil graças ao apoio do grupo WhatsApp e outra jovem também conseguiu iniciar a criação de pato desta forma (GFAR, 2016).

Via redes sociais digitais, agricultores também têm se posicionado contra as restrições criadas pela agricultura de precisão à tomada de decisão deles. Nesse sentido, nos Estados Unidos e na Ucrânia ganha espaço o hacktivismo contra os sistemas informacionais implantados em tratores de última geração. Os fabricantes embarcam nessas máquinas softwares para controlar funcionalidades essenciais delas. Entre outras desvantagens isso impede, na visão dos hacktivistas, a realização de reparos "não autorizados" em equipamentos agrícolas, o "que os agricultores veem como um ataque à sua soberania e muito possivelmente uma ameaça existencial ao seu sustento quando seu trator para de funcionar em situações inoportunas para solicitar o reparo por parte da assistência autorizada" (KOEBLER, 2017).

Para contornar o problema, agricultores do estado norte-americano do Nebraska burlam o *firmware* das máquinas por meio de informações compartilhadas em fóruns on-line apenas para convidados. Via redes sociais digitais do agrossistema esse movimento representa uma extensão da cultura do Faça-Você-Mesmo (DIY, na sigla em inglês), do desvio de circuito, modificação de hardware e outros exercícios hacktivistas (GARNET; PARIKKA, 2016), os quais estão intimamente relacionados ao uso social e político das TIC.

Levando em conta os tipos de prestação de serviço e extensão rural descritos por meio dessas duas tecnografias, apresentadas em torno do uso por parte dos agricultores tanto do WhatsApp quanto dos fóruns on-line, observa-se como os participantes das redes sociais do agronegócio valorizam o livre acesso e a troca de conhecimentos tecnológicos entre eles, em oposição a sistemas organizados menos transparentes, do tipo caixa preta (LATOUR, 2000).

Nesse cenário, os desenvolvedores das emergentes redes sociotécnicas do tipo IoT podem obter mais sucesso na relação tanto com os agricultores quanto com o ecossistema ao posicionar a sustentabilidade como principal indicador de desempenho das soluções criadas por eles. Sintonizados ao conceito de desenvolvimento sustentável do relatório Brundtland (WCED, 1987), indicadores sustentáveis são aqueles que abrangem a tríade do crescimento econômico, social e ambiental no desenvolvimento de novas soluções, e levam em conta as expectativas de seus públicos de interesse atuais e futuros, passando a englobar em suas diretrizes uma maior consciência sobre questões culturais e ecossistêmicas.

É fato que só o desenvolvimento sustentável não reverte problemas como os assinalados pelo Antropoceno – até porque o próprio ecossistema, isoladamente, sempre

estará sujeito a mudanças bruscas de um estado homeostático da Terra para outro, como ocorreu entre 3,8 bilhões e 2 bilhões de anos com impactos duradouros na atmosfera devido à elevação do oxigênio na Terra (VEIGA, 2012). Mas diante da iminência das mudanças geográficas e climáticas (STEFFEN et al., 2015), o risco de falência dos recursos naturais pode ser mais bem manejado por meio da adoção de indicadores de desempenho sustentáveis. Assim, eles devem exercer para criação e manutenção das emergentes redes sociotécnicas rurais do tipo IoT incentivo similar ao que representou o fantasma malthusiano da fome, no século XX, para o sucesso da implantação das tecnologias em torno da Revolução Verde.

RESULTADOS E DISCUSSÃO

Em plena evolução no século XXI, a linguagem digital aplicada aos agrossistemas em geral e às redes sociais em particular envolve uma série de trocas cognitivas entre as diversas partes humanas, técnicas, biofísicas; e impõe fluxos informativos cada vez mais velozes, vastos e variados ao setor rural (CAMARGO, 2015/2017). Ao mesmo tempo, interesses populares, científicos, religiosos continuam presentes no processo de aceitação das novas tecnologias agrícolas. Observe-se novamente o caso da emergência da Revolução Verde. Extremamente controversa, ela teve a seu favor para se estabelecer a equação malthusiana da fome. Por isso, tornou-se uma alternativa viável para alimentar a população mundial de 2 bilhões de pessoas e em franco crescimento em 1950. Mesmo no século XXI, embora rivalizem com a consciência agroecológica em voga, as técnicas da Revolução Verde continuam ativas. Por sua vez, ao ter de fato se configurado neste início de século um fantasma malthusiano, o risco da falta de alimentos (mais associado hoje aos problemas de carência de renda e à logística) não foi uma vantagem competitiva para introdução das sementes transgênicas no mercado (VEIGA, 2007).

Essas diferentes situações podem servir como baliza para o novo sistema sígnico a ser implantado pelo setor agrícola em torno da universalização da cultura digital no campo, especialmente por meio das redes sociais. No agrossistema, as trocas entre as partes (solo, clima, consumidor, fornecedor) ultrapassam os limites dos próprios ecossistemas envolvidos. Assim, não basta inferir relações causais só por meio de configurações internas de conhecimento, habilidades e técnicas. No agrossistema, fatores biofísicos e locais fora de controle têm papel ativo nas controvérsias, especialmente as criadas em torno das inovações. Comparativamente, esse fenômeno foi observado no caso da tentativa de lançamento da semente transgênica Terminator, contrária a continuidade da própria vida vegetal, e que foi rejeitada tanto pelo mercado consumidor quanto pelos agricultores e comunidade científica (CONWAY, 203; VEIGA, 2007). “Conclui-se [daí] que a tecnografia, ao integrar uma análise

social de processos políticos, econômicos e culturais concretos [com as implicações biofísicas que levaram ao Antropoceno, por exemplo], permite avançar [mais] nas explicações causais da dinâmica da mudança tecnológica" agrícola (JANSEN; VELLEMA, 2011, p. 169-170).

CONCLUSÕES

Como parte do ecossistema, a agricultura digital continua, pelo menos até 2017, dependente dos processos de fotossíntese, do nitrogênio atmosférico, dos consumidores de alimentos e dos metais extraídos das terras raras que mantém as baterias de força dos equipamentos eletrônicos operantes. O resultado é que todos esses elementos (vivos e inanimados, geológicos e digitais) estão impregnados uns dos outros (CAMARGO, 2016b).

Isso significa que antes de as tecnologias agrícolas emergentes (agricultura de precisão, redes sociotécnicas do tipo IoT) darem forma racional à fertilidade, à lotação, à adubação, à produtividade do solo na produção das culturas vegetais e animais, elas mesmas deveriam incluir nos seus algoritmos um viés de confirmação (equivalente à verificação da coerência e da relevância daquilo que se promove) que contemple a natureza humana, biofísica, política (tradição oral, ecologia, hacktivismo) que seguem no comando das redes sociais digitais.

Assim, até a simples promoção das AgroTIC nas redes sociais poderia se tornar mais previsível se começasse a pautar os seus processos de extensão rural por meio de mecanismos mais conectivos e ecossistêmicos. De forma mais específica, e tendo como base os exemplos de protagonismo da micro-história agrícola já descritos, as AgroTIC deveriam guiar seus modos de troca cognitiva nas redes sociais do agrossistema digital em torno da tríade do crescimento econômico sustentável, conforme apresentado no desenvolvimento deste artigo.

Isso requer propor e criar soluções digitais que sejam até mais colaborativas do que precisas ou exatas. Fato que força o agronegócio a rever entre os indicadores de desempenho o peso dado à produtividade, sempre tão valorizado, porém insuficiente para garantir a continuidade de seus interesses nas novas redes sociotécnicas do agrossistema, movidas por trocas cognitivas singulares e/ou imprevisíveis entre as diversas partes envolvidas.

REFERÊNCIAS

BRYNJOLFSSON, Erik; McAFEE, Andrew. *The Second Machine Age*: Work, Progress, and Prosperity in a Time of Brilliant Technologies. New York, NY: W. W. Norton, 2014.

BUNGE, Jacob. Tecnologia do 'big data' chega à lavoura e semeia desconfiança. *Valor Econômico*, Opinião, São Paulo, 27 fev. 2014.

CAMARGO, Alessandro Mancio de. *Sociedade em rede*: comunicação científica na nova mídia. Curitiba: Appris, 2016a.

_____. A agricultura inteligente na ecologia midiática. *Teccogs*: Revista Digital de Tecnologias Cognitivas, TIDD - PUC-SP, São Paulo, n. 13, p. 78-92, jan./jun. 2016b.

_____ (ed.). Tag Cloud "CT&I". *Alimento Seguro*, Eiracom, 2015/2017. Disponível em: <http://goo.gl/XRVocl>. Acesso em: 14 fev. 2017.

CASTELLS, Manuel. *A sociedade em rede* – A era da informação: economia, sociedade e cultura. v. 1. 9. ed. São Paulo: Paz e Terra, 2006

CONWAY, Gordon. *Produção de alimentos no século XXI*: biotecnologia e meio ambiente. São Paulo: Estação Liberdade, 2003.

CRUTZEN, Paul J. STOERMER, Eugene F. The "Anthropocene". *IGBP Newsletter*, n. 41, p. 12, 2000. Disponível em: <https://goo.gl/Ksl3t9>. Acesso em: 12 fev. 2017.

DELEUZE, Gilles; GUATTARI, Félix. *Mil Platôs*. São Paulo: Editora 34, 2000.

FRAGALE, Edilson; SILVA, Joana. Agricultura inteligente inspira os 30 anos da Embrapa Instrumentação. *Embrapa*, Notícias, São Carlos, 2014. Disponível em: <https://goo.gl/jHN0JI>. Acesso em: 18 nov. 2015.

GARNET, Hertz; PARIKKA, Jussi. Mídia zumbi: desvio de circuito da arqueologia da mídia para um método de arte. Tradução de Alessandro Mancio de Camargo. *Teccogs*: Revista Digital de Tecnologias Cognitivas, TIDD, PUC-SP, São Paulo, n. 14, jul-dez. 2016 (no prelo).

GFAR. WhatsApp for Agriculture: Digital Farming highlights the need for Digital Agriculture Extension. *The Global Forum on Agricultural Research (GFAR) Blog*, Roma, 8 dez. 2016. Disponível em: <https://goo.gl/qenOXA>. Acesso em: 7 maio 2017.

HARARI, Yuval Noah. *Sapiens* – Uma breve história da humanidade. Tradução de Janaína Marcoantonio. 4. ed. Porto Alegre: LP&M, 2015.

JANSEN, K.; VELLUMA, S. What is technografphy? *NJAS* - Wageningen Journal of Life Sciences, Wageningen (The Netherlands), n. 57 , p. 169–177, 2011.

KOEBLER, Jason. Why American Farmers Are Hacking Their Tractors With Ukrainian Firmware. *Vice*, Motherboard, 21 mar. 2017. Disponível em: <https://goo.gl/3f4wXn>. Acesso em: 7 maio 2017.

LATOUR, Bruno. *Ciência em ação*: como seguir cientistas e engenheiros sociedade afora. São Paulo: Unesp, 2000.

LÉVÊQUE, Christian. *Ecologia*: do ecossistema à biosfera. Lisboa: Instituto Piaget, 2001.

LOWENTHAL, David. Origins of Anthropocene awareness. *The Anthropocene Review*, v. 3, n. 1, p. 52-63, 2016.

MANYIKA, James et. al. Disruptive technologies: Advances that will transform life, business, and the global economy. *McKinsey Global Institute*, p. 1-162, maio, 2013.

MASSRUHÁ, Silvia Maria Fonseca Silveira et. al. *Tecnologia da informação e comunicação e suas relações com a agricultura*. Brasília: Embrapa, 2014.

MATTHEWS, Wendy. Humans and fire: Changing relations in early agricultural and built environments in the Zagros, Iran, Iraq. *The Anthropocene Review*, v. 3, n. 2, p. 107-139, 2016.

RYAN, R; GROSS, N. The diffusion of hybrid seed corn in two Iowa communities. *Rural Sociology*, n. 8, p. 15-24, 1943.

SANTAELLA, Lucia. *Culturas e artes do pós-humano*: da cultura das mídias à cibercultura. São Paulo: Paulus, 2003.

_____; VIEIRA, Jorge de Albuquerque. *Metaciência como guia de pesquisa*: uma proposta semiótica e sistêmica. São Paulo: Editora Mérito, 2008.

_____. *Comunicação ubíqua*: Repercussões na cultura e na educação. São Paulo: Paulus, 2013. Kindle Edition.

_____ et. al. Desvelando a Internet das Coisas. *Geminis*, Universidade Federal de São Carlos - UFSCar, Brasil, v. 1, n. 2, Ano 4, p. 19-32, 2013.

_____. A grande aceleração & o campo comunicacional. *Intexto*, Porto Alegre, UFRGS, n. 34, set./dez. 2015, p. 46-59.

STEFFEN, Will et al. The trajectory of the Anthropocene: The Great Acceleration. *The Anthropocene Review*, 2015, Vol. 2, n. 1, p. 81–98, 2015.

VAN DIJK, Jan A. G. M. *The Network Society*: Social Aspects of New Media. 3 ed. Thousand Oaks, CA: Sage, 2009.

VEIGA, José Eli da (org.). *Transgênicos*: sementes da discórdia. São Paulo: Senac, 2007.

VEIGA, José Eli da (org.). *Gaia*: do mito a ciência. São Paulo: Senac, 2012.

WORLD COMMISSION ON ENVIRONMENT AND DEVELOPMENT (WCED). *Our common future*. Oxford: Oxford University Press, 1987.

ZALASIEWICZ, Jan et al. The technofossil record of human. *The Anthropocene Review*, v. 1, n. 1, p. 34-43, 2014.

ZALASIEWICZ, Jan et al. Anthropocene Working Group (AWG). *University of Leicester*, United Kingdom, Press Office, media note, 29 ago. 2016. Disponível em: <https://goo.gl/i9ahZR>. Acesso em: 4 abr. 2017.

RepiLeite: uso de rede social temática para compartilhar conhecimento

Leonardo Mariano Gravina Fonseca[1], Denis Teixeira da Rocha[1], Carolina Rodrigues Pereira[1], Vanessa Maia Aguiar de Magalhaes[1], Luiz Ricardo da Costa[1], Rita de Cássia Bastos Souza[1], Rosangela Zoccal[1]

[1] Embrapa Gado de Leite, Juiz de Fora, Minas Gerais, Brasil, leonardo.gravina@embrapa.br, denis.rocha@embrapa.br, carolina.pereira@embrapa.br, vanessa.magalhaes@embrapa.br, ricardo.da-costa@embrapa.br, rita.bastos@embrapa.br, rosangela.zoccal@embrapa.br

RESUMO

A Rede de Pesquisa e Inovação em Leite (RepiLeite)[1] é um projeto da Embrapa Gado de Leite lançado em 2011, que consiste em uma rede social temática cujo objetivo é o compartilhamento e o debate de conhecimentos relacionados à cadeia produtiva do leite. Para atingir seus propósitos, a Rede tem como principais ferramentas: fóruns, palestras ao vivo, vídeos e *downloads* de publicações. O objetivo deste trabalho é apresentar a principal postagem dessas ferramentas ao longo dos seis anos do projeto e seus resultados em termos de visualizações, participações e alcance. A RepiLeite foi desenvolvida com a ferramenta Ning e tem seus acessos monitorados pelo *Google Analytics*. As postagens apresentadas demonstram que o projeto tem aproximado diversos representantes da cadeia produtiva do leite, proporcionando aos participantes debates técnicos e o compartilhamento de conhecimentos e tecnologias.

PALAVRAS-CHAVE: *Web*, Debate, Cadeia produtiva leite.

ABSTRACT

The Dairy Research and Innovation Network (RepiLeite)[1] is a Embrapa Dairy Cattle project that started in 2011, which consists of a thematic social network whose objective is the sharing and the debate of knowledge related to the dairy production value chain. To achieve its purposes, the Network has their main tools: forums, live lectures, videos and downloads of publications. The purpose of this paper is to present the main posting of these tools over the

[1] http://www.repileite.com.br

six years of the project as well as their results in terms of visualizations, participation and geographical assessment. The RepiLeite was developed using the Ning tool and has its access monitored by Google Analytics. The posts presented demonstrate that the project has approached various representatives of the dairy production value chain, providing to participants technical discussions and the sharing of knowledge and technologies.
KEYWORDS: Web, Debate, Milk production value chain.

INTRODUÇÃO

A Embrapa Gado de Leite é uma unidade da Empresa Brasileira de Pesquisa Agropecuária (Embrapa), vinculada ao Ministério da Agricultura, Pecuária e Abastecimento (MAPA). A missão da empresa é viabilizar soluções de pesquisa, desenvolvimento e inovação para a sustentabilidade da agricultura, em benefício da sociedade brasileira (EMBRAPA, 2017).

De acordo com o documento Visão 2014-2034 (EMBRAPA, 2014), a pesquisa gera conhecimentos que precisam ser adequadamente encaminhados para produtores rurais, técnicos e sociedade em geral, de modo a possibilitar que as recomendações científicas sejam efetivamente adotadas. Afirma ainda que as redes sociais vão permitir cada vez mais que todos participem e influenciem diretamente o debate público sobre temas como agropecuária, alimentos, biotecnologia e outros, na velocidade da *web*.

Diante deste contexto, a Embrapa Gado de Leite, por meio da Área de Transferência de Tecnologia, vislumbrou a oportunidade da utilização de redes sociais temáticas para o compartilhamento e o debate de conhecimentos sobre a cadeia produtiva do leite (FONSECA et al., 2011). Essas redes são comunidades *online* destinadas à discussão de temas específicos e são formadas por elementos (pessoas, discussões, tempo) que, combinados na *web*, possibilitam a formação de redes de relações sociais (RECUERO, 2009).

Dessa forma, foi lançada em maio de 2011 a Rede de Pesquisa e Inovação em Leite (RepiLeite), uma rede social temática cujo objetivo é o compartilhamento e o debate de conhecimentos relacionados à cadeia produtiva do leite. Ao longo desse tempo foram realizadas melhorias visando dinamizar o fluxo de mensagens e ofertar conteúdos em formatos mais atraentes de modo a ampliar o número de usuários e potencializar o compartilhamento de conhecimentos na Rede. A Figura 1 mostra parte da página inicial do projeto.

Figura 1 – Página inicial da RepiLeite.



Fonte: RepiLeite (2017a).

As ferramentas da RepiLeite que tiveram maior destaque para compartilhamento e debate de conhecimentos, desde sua criação, foram os fóruns, as palestras ao vivo, os vídeos e os *downloads* de publicações. Dessa forma, o objetivo deste trabalho é apresentar a principal postagem dessas ferramentas ao longo dos seis anos do projeto e seus resultados em termos de visualizações, participações e alcance.

MATERIAL E MÉTODOS

Para o desenvolvimento da RepiLeite, optou-se por utilizar a plataforma Ning, que é a maior plataforma de *Software* como serviço, *Software as a service* (SaaS), do mundo para criar sites de redes sociais (NING, 2017). No modelo de distribuição e comercialização SaaS, o fornecedor se responsabiliza por toda a estrutura necessária à disponibilização do sistema

(servidores, conectividade, segurança da informação) e o cliente utiliza o *software* via internet, pagando um valor pelo serviço oferecido.

A Ning disponibiliza uma série de recursos prontos para construção de redes sociais, permitindo personalizações de *layout* (temas, cores, fontes) e ferramentas (fóruns, *blogs*, vídeos, fotos etc.). Essas personalizações podem ser feitas por meio de interface gráfica ou pela linguagem de folha de estilo *Cascading Style Sheets* (CSS).

Para monitorar os acessos à Rede, é utilizado o *Google Analytics*, serviço gratuito oferecido pela *Google*. Por meio desse serviço, é possível obter informações como: acessos por página, tempo de permanência nas páginas, disposição geográfica dos acessos, páginas mais visualizadas e informações dos usuários (navegadores, sistemas operacionais, resoluções de tela, etc.).

Os dados quantitativos apresentados neste trabalho são baseados no *Google Analytics* e nos dados de acesso que a própria Ning disponibiliza[2].

RESULTADOS E DISCUSSÃO

A RepiLeite possui mais de 6.700 membros, distribuídos por todos os estados brasileiros e alguns até em outros países. Apresenta ainda uma média de 8.000 acessos mensais vindos de cerca de 700 cidades.

Qualquer membro pode propor uma postagem, desde que esta seja relacionada ao assunto da Rede. A seguir são apresentadas a principal postagem das quatro ferramentas de maior destaque da Rede (fórum, palestra ao vivo, vídeo e *download* de publicação) ao longo dos seis anos de existência do projeto.

Fórum

A RepiLeite possui 344 fóruns, que abrangem diversos assuntos relacionados à cadeia produtiva do leite. Além da possibilidade de iniciar uma discussão para propor um debate ou colocar alguma dúvida sobre determinado tema, esta ferramenta tem se mostrado eficaz para apresentar conteúdo adicional, como publicações para *download*, apresentações de *slides*, vídeos, *links* de notícias e fotos, complementando sobremaneira as discussões.

O fórum de maior repercussão da Rede é o de um membro que disponibiliza gratuitamente uma planilha elaborada por ele para o gerenciamento de uma propriedade que trabalha com gado leiteiro. Nesta postagem, ele solicita ainda que sejam enviadas sugestões para melhoria da planilha, que assim poderia ser aperfeiçoada em novas versões (Figura 2).

[2] Todos os dados quantitativos da RepiLeite citados nesse trabalho foram coletados no dia 24 de julho de 2017.

Figura 2 – Fórum de discussão na RepiLeite.



Fonte: RepiLeite (2014).

Esse fórum foi postado em 22 de março de 2014 e obteve 45.034 acessos e 1.474 comentários. Analisando-se os comentários feitos pelos membros que se identificaram, percebe-se a participação de diversos tipos de profissionais (agrônomos, estudantes, extensionistas, médicos veterinários, pesquisadores, produtores, professores, zootecnistas), de diversas instituições (Agraer, Cooperalfa, Emater, Embrapa, Empaer, Iagro/MS, IFPA/Campus Castanhal, Instituto Federal de Minas Gerais, Instituto Federal do Triângulo Mineiro, Instituto Federal Farroupilha, Instituto Federal Rio Grande do Sul, Sebrae, Senar/MG, Sindileite, UFLA, Unemat, Uniguaçu, Unimontes, Universidade Estadual do Sudoeste da Bahia, Unochapeco) e de variados estados (AL, AM, CE, DF, GO, MG, MS, MT, PA, PE, PR, RJ, RN, RS, SC, SE, SP).

Essa postagem evidencia o poder de compartilhamento de informações e de construção coletiva do conhecimento propiciado pela ferramenta. Ao mesmo tempo que disponibiliza uma planilha útil a muitos envolvidos na área, o fórum permite aos participantes contribuir para seu aperfeiçoamento com base nas suas experiências de uso no dia a dia.

Palestra ao vivo

Já foram realizadas pela RepiLeite transmissões ao vivo de 61 palestras. Iniciada em 2012, essa atividade amplia a oportunidade para o público estabelecer uma troca multilateral de conhecimentos e experiências com diferentes profissionais. As palestras podem ser de dois tipos: exclusiva para a Rede ou pertencente a algum evento. Em ambos os casos, além de acompanhar o vídeo da transmissão da palestra, os membros podem enviar perguntas e comentários por meio de um *chat* e visualizar a apresentação do palestrante à parte (Figura 3). Ao final, ainda são oferecidos materiais para o aprofundamento no tema apresentado.

Figura 3 – Ambiente para transmissão de palestras ao vivo na RepiLeite.



Fonte: RepiLeite (2017b).

Essa estratégia de transmissões ao vivo tem-se mostrado útil para potenciar o compartilhamento de informações de especialistas a um público maior, não restrito ao espaço físico onde o evento está ocorrendo. Dessa forma, sem gastos de tempo e recursos financeiros com deslocamentos, é possível a interação com palestrantes especialistas em diversos temas.

A transmissão ao vivo de maior repercussão ocorreu em 13 dezembro de 2016. Nesta oportunidade, foi transmitida a final do evento *Ideas for Milk*[3], um desafio de *startups* com foco em soluções para aumentar a eficiência da cadeia produtiva do leite. Entre 9h e 17h (duração do evento), a página da transmissão recebeu 550 acessos únicos, vindos de 16

[3] www.ideasformilk.com.br

estados (CE, DF, ES, GO, MA, MG, PA, PR, PI, RN, RS, RO, RR, SC, SP, SE). Nesta data, a RepiLeite registrou ainda o maior número de acessos em um único dia: 1.023. Também contribuiu para isso o fato de membros da Rede continuarem entrando na página, mesmo após a transmissão, para solicitar o *link* da gravação do evento.

Vídeo

A Rede possui 220 vídeos, que são um dos recursos mais estratégicos de compartilhamento de conhecimento. Eles permitem que, para diversos públicos, boas práticas, procedimentos e tecnologias em geral possam ser melhor absorvidos e compreendidos do que se fossem apresentados por texto, por exemplo. Além disso, esse tipo de mídia digital tem se popularizado ao longo dos últimos anos, estando cada vez mais acessível sua localização, produção e distribuição.

O vídeo "Produção de energia elétrica a partir de biogás gerado por dejetos da pecuária leiteira" (Figura 4), lançado em oito de julho de 2011, é o que obteve mais acessos: 2.549. Sua duração é de 5 minutos e 48 segundos e o tempo médio do usuário na página é de 4 minutos e quatro segundos, o que qualifica o volume de acessos, pois mostra que este tem sido assistido, em média, na sua quase integralidade. Os dados de visualização mostram ainda que o conteúdo apresentado tem chegado a diversas partes do país: 25 estados no total (AC, AL, AM, BA, CE, DF, ES, GO, MA, MG, MS, MT, PA, PB, PE, PI, PR, RJ, RN, RO, RS, SC, SE, SP, TO).

Figura 4 – Vídeo com mais acessos na RepiLeite.



Fonte: RepiLeite (2011).

Download de publicações

Os membros da RepiLeite podem disponibilizar uma publicação para *download* por meio de fórum, *blog* ou comentário em alguma postagem. Essas publicações podem ser apresentadas para divulgação, como complemento de algum debate e/ou para iniciar um canal de comunicação com o autor.

Um *blog* que disponibiliza as cartilhas da Embrapa Gado de Leite com linguagem adaptada ao dia a dia no campo é a postagem que mais se destaca em termos de *download* de publicação apresentando 4.679 acessos. A Figura 5 mostra algumas das cartilhas disponíveis na postagem, que foi lançada em 29 de maio de 2015 e é atualizada toda vez que uma nova cartilha é lançada. A Tabela 1 apresenta a lista de todas as cartilhas disponíveis, com a data de lançamento e o número de *downloads* de cada uma delas.

Figura 5 – *Blog* com cartilhas disponíveis para *download*.



Fonte: RepiLeite (2015).

Tabela 1 – Título das cartilhas disponíveis, data de lançamento e número de *downloads*.

Título cartilha	Data lançamento	Número de *downloads*
Uso econômico do concentrado por meio do controle leiteiro	29/05/2015	1.162
Como obter leite de qualidade utilizando ordenhadeira mecânica	29/05/2015	1.111
Cria de Bezerras	29/05/2015	1.106
Cana com ureia na alimentação de bovinos	29/05/2015	917
Como montar e usar a fossa séptica modelo Embrapa	29/05/2015	784
Como montar e usar o clorador de pastilhas em residências rurais	29/05/2015	743
Como identificar a vaca com mastite em sua propriedade	11/04/2016	742
Montagem do Kit Embrapa de Ordenha Manual para produzir leite com qualidade	29/05/2015	730
Sete passos para uma boa ensilagem de milho	29/01/2016	702
O uso do Kit Embrapa de Ordenha Manual para produzir leite com qualidade	29/05/2015	682
Aprenda a escolher o touro de raças leiteiras para inseminação artificial do seu rebanho	08/07/2015	583
Administrando a propriedade leiteira	05/04/2017	545
Sete passos para uma boa ensilagem de sorgo	29/01/2016	463
Palma forrageira na alimentação de bovinos leiteiros	10/02/2017	364
Plantio e manejo da palma forrageira no Semiárido	10/02/2017	301

Essa disponibilização e divulgação das cartilhas para *download* tem facilitado sua distribuição pelo país, já que os dados mostram acessos de 25 estados (AL, AM, BA, CE, DF, ES, GO, MA, MG, MS, MT, PA, PB, PE, PI, PR, RJ, RN, RO, RR, RS, SC, SE, SP, TO). Além disso, tem economizado recursos financeiros com as impressões de exemplares.

CONCLUSÕES

A utilização da *web* e de mídias sociais já é uma realidade consolidada, inclusive no ambiente rural, onde parte do público da RepiLeite se localiza. Dessa forma, uma rede social temática como esta, voltada para a cadeia produtiva do leite, potencializa o compartilhamento de conhecimentos no setor. Os dados de acesso à Rede e os exemplos de postagens apresentados neste trabalho comprovam isso, permitindo que conteúdos multimídias estejam acessíveis a diversos usuários, inclusive multiplicadores, em diferentes locais do país.

Os diferentes perfis de membros cadastrados na plataforma potencializam um ambiente colaborativo, onde são possíveis interações, por exemplo, entre pesquisadores e produtores, e entre extensionistas e estudantes. Além, é claro, de interações entre membros do mesmo perfil.

Assim, o projeto tem aproximado diversos representantes da cadeia produtiva do leite, proporcionando aos participantes debates técnicos e o compartilhamento de conhecimentos e tecnologias.

Como trabalhos futuros, indicamos uma pesquisa a ser feita com os membros da Rede ainda esse ano. É fundamental para a RepiLeite estar alinhada com as necessidades e

preferências de seu público. Assim, nessa pesquisa serão questionados quais as ferramentas e temas merecem mais destaque, além de outros aspectos de comunicação, como *layout* e divulgação. A partir dos resultados, novas melhorias serão propostas. Como o ambiente *web* é muito dinâmico, uma rede social temática tem que estar em constante processo de melhoria para aumentar visualizações, participações e alcance.

REFERÊNCIAS

EMBRAPA. **Missão, visão e valores**. Brasília, DF, 2017. Disponível em: <https://www.embrapa.br/missao-visao-e-valores>. Acesso em: 27 abr. 2017.

EMBRAPA. **Visão 2014-2034:** o futuro do desenvolvimento tecnológico da agricultura brasileira. Brasília, DF, 2014. 194 p.

FONSECA, L. M. G.; LIMA, V. M. B.; NOBRE, M. M.; MORAIS, L. S. de RepiLeite: Rede de Pesquisa e Inovação em Leite. In: CONGRESSO INTERNACIONAL DO LEITE, 10.; WORKSHOP SOBRE POLÍTICAS PÚBLICAS, 10.; SIMPÓSIO DE SUSTENTABILIDADE DA ATIVIDADE LEITEIRA, 11., 2011, Maceió. **Anais...** Juiz de Fora: Embrapa Gado de Leite, 2011. 3 p. 1 CD-ROM.

NING. **Sobre o Ning.** Menlo Park, CA, 2017. Disponível em: <https://www.ning.com/pt-br/about-us>. Acesso em: 27 abr. 2017.

RECUERO, R. **Redes sociais na internet.** Porto Alegre: Sulina, 2009.

REPILEITE. **Produção de energia elétrica a partir de biogás gerado por dejetos da pecuária leiteira**. Juiz de Fora, MG, 2011. Disponível em: <http://www.repileite.com.br/video/producao-de-energia-eletrica-a>. Acesso em: 27 abr. 2017.

REPILEITE. **FREE - PLANILHA - Software Controle Gado Leiteiro / Fazenda Versao 19/10/2016**. Juiz de Fora, MG, 2014. Disponível em: <http://www.repileite.com.br/forum/topics/novo-software-controle-gado-leiteiro-fazenda>. Acesso em: 27 abr. 2017.

REPILEITE. **Download cartilhas com linguagem adaptada ao dia a dia no campo**. Juiz de Fora, MG, 2015. Disponível em: <http://www.repileite.com.br/profiles/blogs/download-cartilhas-com-linguagem-adaptada-ao-dia-a-dia-no-campo>. Acesso em: 27 abr. 2017.

REPILEITE. **Rede de Pesquisa e Inovação em Leite**. Juiz de Fora, MG, 2017a. Disponível em: <https://www.repileite.com.br>. Acesso em: 27 abr. 2017.

REPILEITE. **Palestras ao vivo**. Juiz de Fora, MG, 2017b. Disponível em: <http://www.repileite.com.br/page/transmissoes-ao-vivo>. Acesso em: 27 abr. 2017.

Lista de autores

——/ A /——

Abreu, Urbano G. P. 197, 267
Almeida, Maxwell Merçon Tezolin Barros 267
Almeida, Samira Luns Hatum 743
Alonso, Camila de Matos 663
Alves, Bruno José 129, 379
Alves, Laura Beatriz 733
Amaro, Rafaella Pironato 187
Apolinário, Daniel Rodrigo de Freitas 59
Arbex, Wagner 625
Arnold, Francisco José 733

——/ B /——

Bambini, Martha Delphino 673
Barbedo, Jayme Garcia Arnal 69
Barbieri, Elvis William de Oliveira 379
Bariani, Cassiane Jrayj de Melo Victoria . . . 389, 557, 567
Bariani, Joice Machado 59
Bariani, Nelson Mario Victoria 389, 557, 567
Barioni, Luis Gustavo 207, 217, 267, 713
Barros, Flavio M. M. 615
Bassoi, Luis Henrique 89, 537
Bazzano, Cristina Freitas 207, 217
Bazzi, Claudio Leones 419
Blanco, Daniel Rodrigues 343
Bocca, Felipe Ferreira 179, 187
Bolfe, Edson Luis 459
Bonacelli, Maria Beatriz Machado 673
Borges, Carlos Cristiano Hasenclever 625
Borges, Ligia Framcielle 419
Botelho, Tiago Gonçalves 257
Braga, Regina 625
Bruno, Karina Maria Berbert 399

——/ C /——

Caires, Eduardo Fávero 9, 19, 149, 595
Camargo, Fabiane Nunes Prates 375, 527
Camargo, Leandro da Silva 375
Camargo, Marcel da Silva 383
Camargo, Sandro da Silva 527
Campos, Fernanda 625
Campos, Leonardo Barreto 41, 237
Campos, Mariana Magalhães 625
Candido, Leandro Rosendo 577
Caresani, José Rodrigo Fernandez 389
Carvalho, José Romário 743
Casagrande, Luan 499, 507
Cavazim Neto, Danilo Luiz 343
Chaves, Cesar G. 585
Cia, Patrícia 705
Ciferri, Ricardo Rodrigues 639
Contini, Elisio 459
Correa, Jorge Luiz 59
Costa, Eduardo Tavares 733
Costa, José Leonardo Dos Santos 257
Crotti, Yuri 499, 507
Cuadra, Santiago Viana 409
Cugnasca, Carlos Eduardo 41, 237, 247
Cunha, Renan 499, 507

——/ D /——

d'Avila, Edson 327, 335
d'Avila, Kirlian Kitzinger 335
D'Ávila, Edson 705
Dalla Pria, Maristella 169
Dal Fabbro, Inácio 335
David, José Maria N. 625
da Costa, Cinthia Cabral 723
da Costa, Daniele Candido 705
da Costa, Luiz Ricardo 763
da Cruz, Ederval Pablo Ferreira 517
da Fontoura, Diego de Carvalho Neves 375
da Hora, Henrique Rego Monteiro 605
da Mota, Fernando Maia 29
da Rocha, Denis Teixeira 763
da Rocha, José Carlos F. 9
da Silva Junior, José Cláudio 317
da Silva, Diego Félix 429
da Silva, Jonathan Galdino 343
da Silva, João Paulo 119
da Silva, Marcos Valério Gebra 327
da Silva, Matheus Phelipe Miller Dias 399
da Silva, Paulo Igor Moraes 51, 139
da Silva, Roberto Fray 247
da Silva, Samuel Ferreira 743
da Silva, Silvando Carlos 439
Della Justina, Diego 187
de Abreu, Deyvison Bruno Morais 41
de Albuquerque, Fernando H. M. A. R. 197

de Alencar, Junia Rodrigues 693
de Almeida Junior, Franciscone Luiz 387
de Andrade, Aluísio Granato 449
de Camargo, Alessandro Mancio 753
de Campos Júnior, Arion 9
de Carvalho, Rogério Atem 605
de Felice, Roberto Dutra 389
de Jesus, Leandro 363
de Lima, André Gomes 605
de Lima, Eduardo Rodrigues 585
de Lima, Helano Póvoas 267
de Lima, Lisiane D. 197
de Magalhães, Vanessa Maia Aguiar 763
de Marchi, Vitor Alex Alves 343
de Melo, Gabriel Marques 387
de Mendonça, Lucas Diogo 585
de Moura, Danilo Gomes 429
de Oliveira, Aryeverton Fortes 409
de Oliveira, Carolina Tavares 537
de Oliveira, Henrique Lima Cará 277
de Oliveira, Luciene Fróes Camarano 439
de Oliveira, Octávio Costa 267
de Oliveira Júnior, Arilson José 377
de Paula Filho, Pedro Luis 343
de Pinho, Leonardo Bidese 375, 547
de Queiroz, Fernando Nakayama 419
de Souza, Eduardo Godoy 419
de Souza, Igo Romero Costa 41, 237
de Souza, Klauss Machareth 29
de Souza, Kleber Xavier Sampaio 713
de Souza, Silvia Regina Lucas 377
de Souza, Ícaro Ramires Costa 41
de Vasconcelos, Aline Pires Vieira 605
Dias, Ana Paula da Silva 693
dos Santos, Alexandre Rosa 743
dos Santos, Fábio 9
dos Santos, Gleissy Mary Amaral Dino Alves 743
dos Santos, Paulo César 257
Dourado Neto, Durval 307
Drucker, Debora Pignatari 59
Duft, Daniel G. 399

——/ **E** /——

Evangelista, Sílvio Roberto Medeiros 59, 129

——/ **F** /——

Faria, Fabio A. 381
Fasiaben, Maria do Carmo Ramos 267, 713
Fernandes, José Mauricio Cunha 99
Ferraciolli, Matheus Agostini 179
Ferreira, Ana Paula Lüdtke 297, 383, 663
Ferreira, Matheus Galvão 387
Ferreira, Pablo Henrique 149
Ferreira, Victor Mesquita 79
Ferronato, Ana Carolina Clivatti. 605
Fiss, Rebeca Einhardt 527
Fonseca, Leonardo Mariano Gravina 763
Francisco, Miguel Sardenha 693
Franquitto, Jessyca Kawana 595
França, Maria Bernadete de Morais 317

——/ **G** /——

Gabardo, Gislaine 169
Gamba, Adriano Ricardo de Abreu 353
Girardi, Gustavo Carolino 343
Gomes, Marina Silva 297
Gomes, Rodrigo da Costa 29
Gonda, Luciano 363
Gonzales, Luis Eduardo 59
Guimarães, Alaine Margarete . . 9, 19, 149, 169, 595, 653
Guimarães, Elcio Perpétuo 439

——/ **H** /——

Haliski, Adriano 19
Higa, Roberto Hiroshi 429, 673, 713
Hölbig, Carlos Amaral 287

——/ **I** /——

Iliuk, Itamar 343
Ishii, Renato Porfirio 29
Itako, Adriana Terumi 159

——/ **J** /——

Jorge, Lúcio André de Castro 577
João, Rafael Stoffalette 489

——/ **K** /——

Kawano, Bruno Rogora 247

——/ **L** /——

Lampert, Vinícius do Nascimento 297
Leonardo, Matheus M. 381
Lino, Antonio Carlos Loureiro 327, 335, 705
Lomba, Luiz Fernando Delboni 363
Luciano, Ana Cláudia dos Santos 399
Luppe, Maximiliam 577

——/ **M** /——

Machado, Luiz Rodolfo 149
Maciel, Denise do Rocio 479

Maciel, Rodrigo 499, 507
Magaldi, Heitor 625
Maia, Aline de Holanda Nunes 307
Mancini, Adauto Luiz 207, 217
Mantuani, Silvia Ribeiro 469, 479
Marcelino, Roderval 499, 507
Mariath, Fabiano 459
Martins, Roberto Luvisutto 89
Marér, Renata Maria 247
Matos, Juliana Martins de Mesquista 79
Mattos, Amanda do Prado 159
Maximo, Fernando Attique 693
Mazzonetto, Angela 287
Mengarda, Diego 375
Monteiro, José Eduardo Boffino de Almeida 277
Moura, Maria Fernanda 537, 713

——/ N /——

Nacer, João Felipe Resende 227
Narciso, Marcelo Gonçalves 307, 439
Nicolau, Márcio 99, 207, 217
Noleto, Fábio Fernandes 439

——/ O /——

Oldoni, Henrique 89
Oliveira, Monique Pire Gravina 109
Oliveira, Márcia Gonzaga de Castro 439
Oliveira, Rodolfo 517
Oliveira, Stanley R. de M. 267, 277, 537, 615, 713

——/ P /——

Pavan, Willingthon 99, 287
Pena Jr., Marcos A. G. 459
Pereira, Bruno Henrique 207, 217
Pereira, Carolina Rodrigues 763
Perez, Naylor Bastiani 383
Pichorim, Sérgio Francisco 353
Pimentel, Marcia Barrocas Moreira 99
Pires, Pedro Paulo 363
Poletti, Elaine Cristina Catapani 733
Portocarrero, Hugo 449

——/ R /——

Rebonatto, Marcelo Trindade 287
Ribeiro, Carlos A. B. 381
Ribeiro, Marcela Xavier 489
Ribeiro, Selma Regina Aranha 595, 653
Richa, César Huegel 547
Rocha, Carlos Hugo 653
Rodrigues, Giancarlo 9
Rodrigues, Luiz Henrique Antunes 109, 179, 187
Rodrigues, Rulian 449
Romani, Luciana Alvim Santos 59, 119, 129, 489
Rosa, Joel Zubek 653
Rosa, Renan de Paula 385
Rubinsztejn, Hana Karina Salles 363

——/ S /——

Salame, Marcos Filipe Alves 51, 139
Salles, Allan Silva 449
Sanches, Guilherme Martineli 399
Sanches, Juliana 705
Santos, Bruno Queiroz 387
Santos, Thiago Teixeira 89
Saqui, Diego 257
Schafer, Alexandro Gularte 663
Schenatto, Kelyn 419
Scrivani, Rachel 129
Silva, Denival da Costa 327
Silva, Giani 459
Silva, Gisela Maiara 399
Silva, Ivan Lucas Cabral 537
Sotoma, Irineu 227
Sousa, Allex de Lima 51
Souza, Juan D. F. 197
Souza, Rita de Cássia Bastos 763
Speranza, Eduardo Antonio 639
Stachak, Alessandro 9, 169, 595
Stevan Jr., Sergio Luiz 683

——/ T /——

Takemura, Celina Maki 537
Teixeira, Guilherme da Silva 387
Teixeira, Gustavo Ribeiro 79
Ternes, Sônia 713
Tibola, Casiane Salete 99
Tolentino Júnior, João Batista 159
Trindade, José Pedro Pereira 527, 547, 663
Tápias, Luisa Miyashiro 537

——/ V /——

Vacari, Isaque 59
Vaz, Glauber José 59
Vaz, Maria Salete Marcon Gomes 469, 479
Vaz, Monica Cristine Scherer 469, 479
Vicente, Luiz Eduardo 585
Victoria, Daniel de Castro 409
Vidal, Alda Mieko Rocha Kimura 79

Vidal, Flavio de Barros 79
Vieira, Ana Paula dos Santos Braatz 489
Vieira, Fábio Danilo 429
Viniski, Antônio David 19, 149, 169, 595
Volk, Leandro Bochi da Silva 527, 547
Voss, Henike Guilherme Jordan 683

—— / Z / ——

Zampier, David Patrick 169
Zerlotini, Adhemar 429
Zoccal, Rosangela 763
Zullo Jr., Jurandir 119

Realização:

Promoção:

Patrocínio:

Patrocinador diamante

Patrocinador ouro

Patrocinadores prata

Patrocinadores bronze

Apoio: