



5 Computação científica na agricultura

Sônia Ternes
Maria Fernanda Moura
Kleber Xavier Sampaio de Souza
Glauber José Vaz
Stanley Robson de Medeiros Oliveira
Roberto Hiroshi Higa
Helano Póvoas de Lima
Celina Maki Takemura
Enilda Coelho
Francisco Ferraz Laranjeira Barbosa
Marcos Cezar Visoli
Gilberto Romeiro de Oliveira Menezes
Luiz Otávio Campos da Silva
Sandra Aparecida Santos
Sílvia Maria Fonseca Silveira Massruhá
Urbano Gomes Pinto de Abreu
Balbina Maria Araujo Soriano
Suzana Maria Salis
Márcia Divina de Oliveira
Walfrido Moraes Tomas

1 Introdução

O desenvolvimento recente de tecnologias digitais tem estimulado um crescimento vertiginoso da capacidade de aquisição de grandes volumes de diferentes tipos de dados, provenientes das mais variadas fontes. Na agricultura, ao longo de sua cadeia de geração de valor, esses dados podem compreender: (i) dados ômicos (genômica, proteômica, transcriptômica e metabolômica); (ii) atributos físico-químicos com localização espaço-temporal adquiridos

por meio de sensores (iii) imagens aéreas e de satélites com localização espaço-temporal; (iv) dados socioeconômicos; dentre outros.

De forma similar ao que ocorre com dados oriundos de fontes mais tradicionais, a utilização desse grande volume de dados, seja no desenvolvimento de um novo ativo biotecnológico, no monitoramento de uso da terra ou no controle de um processo de produção por um controlador, pressupõe sua análise por meio de modelos e algoritmos capazes de extrair informação útil ao processo de tomada de decisão. Assim, a computação científica, entendida como uma coleção de técnicas, ferramentas e teorias que englobam matemática, estatística, física e computação e abrangem conhecimentos específicos de subáreas, tais como estatística aplicada, econometria, matemática aplicada, inteligência computacional, visualização científica e biometria, continuará a ser central no desenvolvimento de novas tecnologias agrícolas, agora no contexto da emergente Agricultura Digital. Nas últimas décadas, inclusive, a computação científica tem sido apontada como o terceiro pilar da pesquisa científica, junto com a experimentação e a teoria (Souza et al., 2017).

Nas seções seguintes apresentamos exemplos de aplicações que utilizam algoritmos e técnicas de computação científica para contribuir com a solução de problemas do setor agropecuário. Na Seção 2 são apresentadas duas aplicações que, de modo geral, partem da observação de um grande conjunto de dados brutos visando reconhecer padrões embutidos nos mesmos e derivar conhecimento e ações a partir dos padrões observados que podem, inclusive, ser utilizados por um Sistema Especialista. Na Seção 3 são apresentadas três aplicações baseadas na construção de modelos matemáticos e estatísticos, a partir dos quais podem ser realizadas previsões e análises baseadas em cenários de simulação hipotéticos, para auxiliar a tomada de decisão do poder público. Por meio dessas diferentes aplicações é possível observar que a computação científica é uma área de pesquisa eminentemente transversal a outras.

2 Inteligência artificial

Inteligência Artificial é uma área ampla, que teve seu início na segunda metade da década de 1940, quando foi concebida uma rede neural artificial descrevendo como os neurônios humanos deveriam aprender a executar cálculos. A área passou por muitas modificações e intersecções com outras áreas, especialmente a modelagem estatística e vários métodos de reconhecimento de padrões. Essas intersecções formam um grupo de técnicas conhecidas como sistemas inteligentes, que são baseados em aprendizado de máquina.

Um modelo de aprendizado de máquina é apoiado em dados observados previamente, provenientes de bases de dados, experimentos, imagens ou textos. Os dados possuem atributos, que precisam ser descritos para cada

observação. Por exemplo, se coletamos dados de uma pastagem em diferentes locais da propriedade, teremos os mesmos atributos para cada coleta de dados, tais como: local de coleta, tipo de capim, data da coleta, situação do pasto (degradado, não degradado, em degradação), posição geográfica da coleta, percentual de cobertura do solo, tipo de solo, altura do pasto, etc. Com esses atributos e os dados coletados, pode-se construir um modelo de classificação da situação da pastagem. Se os dados observados fossem textos, os atributos poderiam ser os vocábulos presentes nos textos; se fossem imagens, os atributos poderiam ser essas imagens divididas em pedaços bem pequenos, chamados *pixels*, considerando-se, por exemplo, a cor de cada pixel.

Nos itens 2.1 e 2.2 são apresentados, respectivamente, exemplos de classificação automática de solos e de exploração de informações em textos, fazendo uso de diferentes técnicas de Inteligência Artificial.

2.1 Classificação automática de solos

Para classificar um perfil de solo, o Sistema Brasileiro de Classificação de Solos (SiBCS) considera uma ampla gama de atributos morfológicos, físicos, químicos e mineralógicos, além de aspectos ambientais como clima, vegetação, relevo, material originário, condições hídricas, características externas ao solo e relações solo e paisagem (Santos et al., 2013).

Visando auxiliar nesse desafio, que é um processo laborioso, a Embrapa Informática Agropecuária e a Embrapa Solos conceberam duas ferramentas inteligentes para classificação automática de solos. A primeira refere-se ao desenvolvimento de um Sistema Especialista que contém as regras do SiBCS de forma automática para classificação de solos. A segunda iniciativa diz respeito ao sistema Web (SoloClass) para classificação de perfis de solos, por meio de um comitê de soluções inteligentes baseadas em algoritmos de aprendizado de máquina. Essas ferramentas inteligentes foram desenvolvidas dentro do escopo do projeto “Uso de dispositivos móveis inteligentes na classificação de solos brasileiros – SmartSolos”, liderado pela Embrapa Solos. Ambas as ferramentas são apresentadas nas subseções a seguir.

2.1.1 Sistema especialista baseado no SiBCS

O sistema especialista baseado nas regras do SiBCS simula o raciocínio de um profissional especialista do domínio ao realizar a classificação de perfis de solo. Dessa forma, pode ser usado para classificar perfis de solos ainda não conhecidos e para validar perfis previamente classificados (Vaz et al., 2018).

Vaz et al. (2019a) usaram o sistema especialista para analisar dados de solos disponibilizados pelo IBGE. Mostraram que se trata de importante ferramenta para a curadoria dos dados de solos brasileiros, pois permite executá-la de forma mais eficiente e com menos erros, beneficiando a governança de solos no Brasil.

Também foram mostradas as vantagens de se disponibilizar o sistema especialista por meio de uma API e a importância dessa ferramenta para facilitar a curadoria dos dados de solos e orientar um registro mais adequado desses dados (Vaz et al., 2019b). A Figura 1 mostra que, disponibilizando-se o sistema especialista por meio da API, o usuário pode obter as classificações de perfis de solo do sistema especialista e compará-las com classificações previamente conhecidas. Assim, podem-se analisar possíveis erros nos dados de solos e corrigi-los, tornando a ferramenta poderosa para melhoria da qualidade de dados de solos do Brasil.



Figura 1. Esquema da análise de classificação de perfis de solo.

Fonte: Vaz et al. (2019b).

O grande desafio desse sistema está em codificar todas as regras do SiBCS para tratar seus quatro primeiros níveis categóricos. A taxonomia de classificação conta com mais de mil classes entre o primeiro e o quarto nível categórico. Além disso, as regras são bastante complexas, de maneira que é fundamental o trabalho conjunto e muito empenho de cientistas da computação e de solos para viabilizar o desenvolvimento de tal sistema.

Embora um aplicativo específico esteja sendo desenvolvido pela Embrapa para utilizar essa API de classificação de solos, instituições parceiras também podem utilizá-la para criar novas soluções que contem com a classificação de solos, desde que seus dados estejam codificados de acordo com os padrões estabelecidos pelo sistema.

Em relação aos padrões de dados de solos do Brasil, há diferentes iniciativas que buscam organizá-los. No entanto, nenhuma delas se consolidou como um padrão, nem atende às necessidades do sistema especialista desenvolvido, de maneira que muitas observações puderam ser realizadas ao longo deste trabalho em relação à organização desses dados no Brasil. É comum, por exemplo, observar redundância de dados em diferentes campos, ausência de campos necessários para o registro de informações importantes para a classificação de solos e representações de dados que dificultam o processamento computacional e sua recuperação. A próxima etapa deste trabalho é, portanto, consolidar uma série de recomendações para a estruturação de dados de solos brasileiros a fim de tornar mais simples a manipulação computacional desses

dados, garantir maior qualidade dos dados armazenados e facilitar a criação de novas soluções que dependam deles.

Pesquisas futuras estão na possibilidade de automatizar outros processos que normalmente tomam muito tempo do profissional ou aumentam muito a incerteza dos dados coletados no campo. Por exemplo, cor, textura, limites das camadas de solos e outros atributos são determinados de uma maneira subjetiva, de acordo com interpretações pessoais realizadas no momento do trabalho em campo. A coleta desse tipo de informação pode ser facilitada e automatizada por meio de ferramentas computacionais que extraiam características de imagens tiradas no campo.

2.1.2 Sistema inteligente de classificação de solos

Uma alternativa promissora para classificação automática de solos é a combinação de algoritmos de aprendizado de máquina (AM) com métodos de seleção de atributos. Algoritmos de AM operam construindo um modelo a partir de amostras de treinamento a fim de fazer previsões guiadas pelos dados, contendo observações de perfis de solos previamente classificados por pedólogos. Por sua vez, os métodos de seleção de atributos têm como objetivo principal encontrar um subconjunto de variáveis relevantes para uma tarefa-alvo, tornando o processo de aprendizado mais eficiente, por meio da simplificação do custo operacional dos modelos, possibilitando um melhor entendimento sobre os resultados encontrados (Guyon; Elisseeff, 2003).

O SoloClass é um sistema inteligente desenvolvido para classificação de perfis de solos. Esse sistema permite que um usuário forneça como entrada um conjunto de variáveis de um ou mais perfis de solo, e receba como resultado a classificação de cada perfil de acordo com o SiBCS, com uma probabilidade associada à classe prevista.

Cinco classes de algoritmos de AM foram utilizadas para a classificação inteligente de solos: i) simbólica: árvores de decisão; ii) baseada em instâncias: k -NN ou k vizinhos mais próximos; iii) aprendizado estatístico: SVM - *Support Vector Machines*; iv) *bootstrap aggregating: Random Forest*; e v) conexionista: Redes Neurais Profundas. Todos esses algoritmos foram treinados para os quatro níveis categóricos (ordens, subordens, grandes grupos e subgrupos) adotados pelo SiBCS.

A arquitetura do sistema SoloClass é baseada em um comitê de classificadores, conforme pode ser visto na Figura 2.

Ao receber um conjunto de perfis de solos, com números diferentes de horizontes, o usuário pode selecionar um ou mais classificadores que foram treinados a partir de uma base de dados pré-classificada por pedólogos (processo de indução). Em seguida, o sistema dispara os classificadores selecionados e armazena os resultados apresentados individualmente. Ao final do processo de dedução, o comitê de classificadores (Figura 2) atribui o resultado

da classificação ao perfil de solo por votação, isto é, a classificação associada ao perfil é aquela que obteve maior frequência ou votação majoritária.

Essa arquitetura baseada em comitê de classificadores tem algumas vantagens como: a) aumento do poder de predição do sistema em função do uso de vários classificadores ajustados aos dados e combinados para esse aumento; b) redução da variância e do viés quando comparada ao uso de apenas um método de aprendizado de máquina; c) arquitetura extensível, isto é, outros classificadores podem ser adicionados.

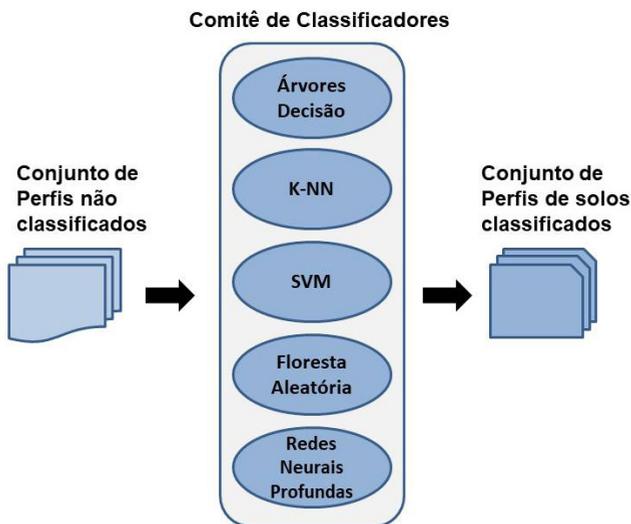


Figura 2. Arquitetura do SoloClass baseada em um comitê de classificadores.

Os principais benefícios do sistema SoloClass são: a) auxiliar projetos e programas nacionais de levantamento de solos, como o Pronasolos (Polidoro et al., 2016), atuando como uma ferramenta facilitadora nos trabalhos de classificação de solos; b) facilitar o entendimento da classificação de solos para agricultores, estudantes, professores, extensionistas e pesquisadores; e c) minimizar eventuais erros humanos durante a atividade de classificação de solos.

Por se tratar de um Sistema Web, o SoloClass¹ pode ser acessado por meio de dispositivos móveis ou computadores pessoais, sem nenhuma restrição de sistema operacional, contribuindo para ampliar o acesso e a inclusão de um maior número de usuários. O SoloClass possui interface responsiva, isto é, pode ser caracterizado pela adaptação visual de uma página ou interface a qualquer dispositivo em que seja visualizada, sem a necessidade do uso de versões específicas para cada modelo.

¹ Disponível em: www.soloclass.cnptia.embrapa.br

2.2 Mineração de textos em publicações técnico-científicas

O processo de aprendizado humano é baseado em observações e formação de padrões, hipóteses e inferências a partir dessas observações. Nos dias de hoje, sobram observações: há um volume excessivo de dados, tanto organizados em bases de dados quanto publicados no formato textual. A mineração de dados usa processos de análise estatística, cujos algoritmos são implementados em programas computacionais que conseguem lidar com um grande volume de dados, para encontrar padrões e auxiliar a formatação de hipóteses e modelos que permitam descrever esses padrões.

Pode-se afirmar que a mineração de textos (MT) é uma especialização do processo de mineração de dados. A principal diferença entre os dois processos é que, enquanto a mineração de dados convencional trabalha exclusivamente com dados estruturados (pré-organizados em bases de dados ou alguma representação, como uma planilha), a mineração de textos lida com dados inerentemente não estruturados. Logo, na MT o primeiro desafio é estruturar os dados com seus respectivos atributos, a partir dos textos, para que se possa utilizar algoritmos de mineração de dados.

A estruturação dos textos depende do problema abordado. Por exemplo, se queremos conhecer ou relacionar quais os tipos de tecnologias agrícolas estão ligados ao uso da água na agricultura brasileira, pode-se delimitar um conjunto de publicações técnico-científicas sobre o tema e dele extrair essas informações. Nesse caso, uma opção é o uso de ferramentas linguísticas que permitam identificar o vocabulário de interesse (por exemplo: irrigação, colheita, recursos hídricos, pivô, etc.) e delimitar e desambiguar localizações geográficas (tais como: Rio São Francisco, Igreja de São Francisco, Cidade São Francisco, etc.) nos textos.

Para problemas similares a esse, na metodologia proposta por Moura et al. (2017) encontra-se um passo a passo semiautomatizado, com o uso de ferramentas de software desenvolvidas especificamente para isso, contendo as seguintes etapas: 1) delimitação das publicações de interesse; 2) extração e desambiguação de topônimos com a ferramenta TopExtract (Takemura et al., 2013); 3) formatação de um dicionário dos termos de interesse, que é uma atividade manual, realizada por especialistas do domínio; 4) uso da ferramenta ExtracTrans (Extração de Transações) para: i) extração de termos dos textos, por similaridade e sinonímia; ii) criação das transações presentes nos textos (todo o vocábulo de interesse que apareceu no texto); e iii) eliminação de dados redundantes, que não contribuem para os resultados; 5) extração de padrões, utilizando algoritmos de aprendizado de máquina, como os de regras de associação, ou mesmo colocando os resultados em uma planilha Excel ou outro software semelhante. Por exemplo, em Moura et al. (2017) foram encontradas para a região Nordeste 40 regras de associação, dentre as quais:

Se tecnologiaClasse = engenharia agricola & cultura = uva & culturaClasse = frutíferas & regioao = NE ==> tecnologia = irrigacao.

Uma outra aplicação de interesse bem específico foi descrever quais métodos quantitativos e de computação científica foram citados nas publicações científicas da Embrapa, dentre as consideradas de mais alto nível, segundo o indicador Qualis CAPES (A1, A2, B1 e B2), entre os anos de 2000 e 2018. A Embrapa possui um sistema próprio de catalogação de todas as suas publicações e tecnologias (Embrapa, 2020), cujos metadados, dos quais as palavras-chave fazem parte, são auditados, o que mostra sua altíssima qualidade. Porém, dois grandes problemas apresentam-se para o estudo: i) a grande quantidade de publicações nesse intervalo, aproximadamente vinte e dois mil artigos; e ii) o fato de que as palavras-chave dos artigos cobrem termos de agricultura, e não necessariamente termos de métodos quantitativos e computação científica. Ou seja, as palavras-chave de interesse dessa análise de dados não faziam parte do repertório de palavras-chave convencionais desses artigos e, portanto, não poderiam ser localizadas apenas por resultados de busca e muito menos lendo-se cada um dos vinte e dois mil artigos, o que seria uma tarefa muito extensa. Dessa forma, a metodologia de Moura et al. (2017) foi adaptada da seguinte forma para utilização nesse processo: 1) já se encontravam selecionados os 22 mil artigos de interesse; 2) o processo de geolocalização não foi necessário e em seu lugar foi criado um processo para realizar o *download* dos artigos e convertê-los para texto plano; 3) os especialistas do domínio, de métodos quantitativos e computação científica, organizaram o dicionário de termos necessário à área; 4) foi adaptada uma ferramenta (a partir da ExtracTrans) para extrair, por similaridade, os vocábulos de interesse da coleção de textos e, na sequência, colocar os dados em uma planilha; e 5) a partir das planilhas de dados, em formato Excel, foram aplicadas as técnicas de cruzamento de tabelas dinâmicas, agregação de dados de outras fontes, agrupamentos, seleção e filtros para facilitar a exploração dos dados em diferentes visões.

Algumas aplicações exploratórias sobre um grande volume de textos fazem uso de um processo semelhante ao utilizado pelas máquinas de busca, como Google, Yahoo, etc. Indexa-se a coleção textual, na qual cada texto (dado) corresponde a uma linha de uma tabela e cada vocábulo (atributo) a uma coluna, nem sempre sendo necessário o conhecimento da língua na qual o texto esteja escrito, muito menos se há dependências entre os vocábulos. Em cada célula coloca-se a frequência do vocábulo no texto, ou alguma medida derivada desta. Então, como essa tabela tem um número exagerado de colunas e muitas células com o valor zero, procura-se diminuir o número de colunas, selecionando-se os vocábulos ou a composição de vocábulos estatisticamente mais significativos.

Existem muitas técnicas para diminuição do número de colunas da tabela, sendo todas dependentes do que se quer responder em relação à coleção de

textos. Para formatar uma coleção de textos em uma tabela como essa, tem-se a ferramenta I-PreProc (Pereira; Moura, 2015). Uma aplicação comum para essa formatação é agrupar os textos com conteúdos semelhantes de modo que eles devam corresponder a tópicos específicos, isto é, ficar subdivididos em assuntos mais correlatos, como realizado no projeto “CRITIC@ - Compilação e Recuperação de Informações Técnico-científicas e Indução ao Conhecimento”, desenvolvido pela Embrapa, que também utiliza outras ferramentas como a TopExtract, já citada.

Na Figura 3 observa-se, à esquerda, que a partir de uma expressão de busca na base de publicações, os resultados da busca são organizados hierarquicamente em grupos de documentos, para os quais são encontrados termos mais estatisticamente significativos no grupo, que são considerados “tópicos”; no meio tem-se a distribuição de frequências acumuladas no tempo para o grupo representado por “Trator, Efeito, Termo, Diferença, Aplicada, Folha, Poda” e, à sua direita, os locais no país citados nesses documentos. Esse resultado, que é fruto de exploração dos dados, nos traz pistas de: i) como esses documentos poderiam estar organizados de acordo com os grupos; ii) quais seriam os tópicos ou o conjunto de palavras-chave desse grupo, por exemplo “trator, poda, poda aplicada à folha”, ou seja, o que um especialista da área julgar mais importante no resultado apresentado; e iii) localização geográfica, isto é, onde esses grupos aparecem de modo mais significativo.

Figura 3.
Exemplo de resultado do projeto CRITIC@.



Como visto nas aplicações citadas, processos de mineração de textos, sejam quais forem as técnicas, auxiliam a exploração e a identificação de informação em um grande volume de textos.

3 Modelagem matemática e estatística

A modelagem matemática é uma área ainda mais ampla que a de Inteligência Artificial, que usa um pequeno arcabouço de soluções matemáticas, o mesmo ocorrendo com a modelagem estatística. A ideia geral do processo de modelagem é a interpretação simplificada de um fenômeno, que passa a ser descrito em linguagem matemática, permitindo posteriormente a realização

de simulações por meio do computador. Assim, o usuário do modelo posiciona-se como um experimentador do mundo real, pois a partir dos resultados de diversas simulações computacionais, pode compreender detalhes do fenômeno em situações não vivenciadas na prática. Na pesquisa agropecuária, por exemplo, modelos matemáticos e estatísticos são essenciais para complementar os experimentos biológicos, permitindo, entre outros, o estudo da dinâmica de doenças no campo a partir de simulações computacionais, ou seja, sem impacto ambiental e com economia de recursos. Como exemplo, no item 3.1 é apresentado um modelo de simulação para análise da dispersão intrapomar da doença conhecida como HLB do citros.

Para entender a diferença entre modelos matemáticos e estatísticos, consideremos um exemplo simples como a equação matemática que representa uma reta no plano cartesiano (x, y) , dada por “ $y = a + bx$ ”, onde “ a ” é a inclinação da reta e “ b ” o fator que correlaciona cada valor de x a exatamente um valor de y no plano, dada a inclinação “ a ”. Por outro lado, se observarmos os valores de peso e altura de um grupo de pessoas, sabemos *a priori* que o comportamento dos pontos observados (peso, altura) é linear, ou seja, pode ser representado por uma reta, mas não corresponde exatamente à relação de peso e altura de toda a população, ou seja, esse conjunto de pontos é apenas uma amostra dessa população. Uma boa amostra deve ser aleatorizada, isto é, sorteia-se aleatoriamente cada pessoa para se medir o peso e a altura, e ter um tamanho estatisticamente razoável. Assim, com essa amostra coletada, estima-se o comportamento da população para o problema em estudo (peso, altura), que é uma reta composta por valores estimados da inclinação da reta (média dos valores observados de peso), e do fator que correlaciona a altura ao peso, considerando os erros do modelo, no qual as estimativas dependem de distribuições de probabilidade. No item 3.2 é apresentado um modelo linear multivariado, isto é, várias variáveis dependentes (que entrariam no lugar do peso) e várias variáveis independentes (que entrariam no lugar de altura), e ainda: i) efeitos fixos, que são os dos fatores que podem ser observados, assim como a altura no nosso exemplo; e ii) efeitos aleatórios, que não são observados na coleta da amostra, mas precisam ser estimados pelo modelo.

Um outro arcabouço dentro de modelagem matemática são os modelos lógicos indutivos, por exemplo, “*se A é um haras então A possui cavalos*”. Dentre esses modelos estão os de lógica *fuzzy* (lógica nebulosa ou difusa). Por exemplo, se temos um copo com água, este pode estar cheio, meio cheio, meio vazio ou vazio, de acordo com a interpretação de cada pessoa que olha para o copo. Então, podem-se formar regras, tais como o copo está vazio se tem de 0 a 20 mL de água, está meio vazio se tem de 10 a 100 mL de água, está meio cheio se tem de 50 a 160 mL de água e está cheio se tem mais de 140 mL de água. Para resolver uma classificação de como está cada copo, desenvolve-se um sistema baseado em regras *fuzzy*. O item 3.3 mostra uma aplicação de

sistemas baseados em regras *fuzzy* para auxiliar a tomada de decisão relativa à sustentabilidade em fazendas pantaneiras, considerando ponderações entre valores ambientais, sociais e econômicos.

3.1 Modelagem da dinâmica de dispersão do “HLB do citros”

O Brasil é o maior produtor mundial de laranja, sendo a safra 2020/21 estimada em quase 288 milhões de caixas (40,8 kg) (Fundecitrus, 2020). A doença conhecida como Huanglongbing (HLB) ou Greening, identificada no país em 2004, é atualmente a de maior importância para a citricultura nacional. O HLB do citros é causado pela bactéria *Candidatus Liberibacter asiaticus* e transmitida no Brasil principalmente pelo psilídeo *Diaphorina citri*, que adquire as bactérias ao se alimentar da seiva de plantas infectadas, transmitindo-as posteriormente para plantas sadias.

Devido à sua importância para a economia nacional, desde 2012 a Embrapa vem desenvolvendo ferramentas biomatemáticas para auxiliar no monitoramento, na amostragem, na detecção e na erradicação do HLB do citros. Inicialmente foi desenvolvido, entre outros, um modelo matemático compartimental determinístico (Vilamiu et al., 2013) para avaliar o impacto na diminuição dos níveis populacionais do inseto vetor *D. citri* no Recôncavo baiano, em áreas onde são plantados citros e hospedeiros alternativos (murta-de-cheiro - *Murraya paniculata*) em diferentes proporções, objetivando colaborar com políticas públicas para o setor. Nesse estudo as populações de citros e murta foram divididas em compartimentos (plantas suscetíveis, expostas, infectadas e recuperadas) e as características gerais de cada compartimento foram expressas por meio de equações matemáticas para análise da dinâmica temporal de propagação do HLB.

Mais recentemente (Barbosa, 2015), uma nova abordagem de modelagem foi usada para avaliar, entre outros aspectos, o papel da murta como fator de repulsão (*push*) ou atração (*pull*) de insetos vetores nas áreas de cultivo, a partir de cenários de simulação com diferentes configurações espaciais de murta e citros. Para tal utilizou-se a modelagem baseada no indivíduo (MBI) (Grimm; Railsback, 2005), que considera no modelo a presença e a particularidade de cada indivíduo das populações envolvidas e observa o sistema final como sendo resultante das interações entre os indivíduos das diferentes populações. A abordagem MBI mostra-se adequada para os objetivos do estudo porque permite explorar conjuntamente os aspectos temporais e espaciais do sistema “hospedeiro-inseto vetor-HLB”, de modo mais intuitivo e flexível do que os modelos matemáticos clássicos como os usados em Vilamiu et al. (2013).

O MBI foi desenvolvido em linguagem de programação Python e considera uma paisagem agrícola padrão do Recôncavo Baiano, contendo 9 talhões com 20 x 42 plantas em cada talhão (total de 840 plantas hospedeiras

por talhão ou 7.560 plantas na paisagem), espaçamento entre linhas de 6m e espaçamento entre colunas de 4m, totalizando uma área de 120m (largura) x 168m (comprimento), pouco mais de 2ha.

Para análise da dispersão intrapomar do inseto vetor e propagação do HLB foram testadas e comparadas 3 diferentes paisagens: a) Cenário 1: somente citros; b) Cenário 2: citros e murta ao redor da área toda; c) Cenário 3: citros e murta nas bordas de cada talhão. Assim, as populações consideradas no MBI e envolvidas nas simulações computacionais são: a) planta hospedeira principal (citros); b) planta hospedeira alternativa (murta) para teste do efeito de repulsão e atração; c) inseto vetor *D. citri* no estágio de ninfa; d) inseto vetor adulto.

De modo geral, na execução das simulações do modelo, o usuário do simulador pode escolher diferentes valores para os seguintes parâmetros biológicos, obtidos a partir de estudos e experimentos biológicos realizados nos campos experimentais da Embrapa Mandioca e Fruticultura (Cruz das Almas, Bahia):

- tempo da fase de incubação da doença na planta: 180 a 540 dias;
- tempo de duração da fase de latência da doença na planta: 30 ou 60 dias;
- proporção de insetos por planta: 0,41 a 5;
- tempo de simulação: 1, 2, 5, 10 ou 20 anos;
- modo de simulação: 1 (*single*) ou 2 (*multi*);
- probabilidade de infecção primária (PIP), conforme incidência na região: 0,01 (baixa), 0,15 (média) ou 0,30 (alta);
- probabilidade de detecção da doença no campo pelo fiscal de inspeção: 0 ou 0,476.

As simulações se iniciam com todas as plantas sadias e a chegada de certa proporção de insetos infectivos, conforme os valores de PIP. As populações evoluem estocasticamente no tempo (conforme a probabilidade de ocorrência) a partir de processos do tipo nascimento e morte de ninfas e insetos adultos, infecção de plantas hospedeiras, aquisição da bactéria por ninfas e insetos adultos, reprodução e voo de insetos adultos.

Ao final das simulações computacionais são gerados dois tipos de resultados: a) para simulação do tipo “single”: a cada 10 dias de execução da simulação é gerado um arquivo em disco contendo o status das populações em cada posição da área de plantio (tipo do hospedeiro, status de infecção, número de insetos na posição); b) para simulação do tipo “multi”: ao final de 100 execuções automáticas (processo de Monte Carlo) são gerados gráficos do número de plantas suscetíveis, infectadas e sintomáticas ao longo do tempo.

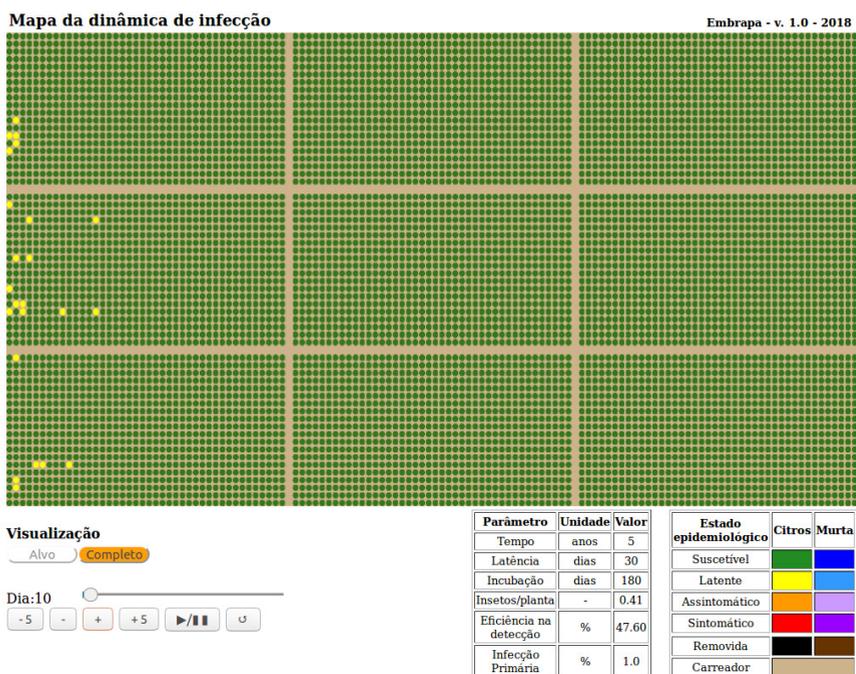
Os resultados do MBI são gravados em disco, e um software, desenvolvido em linguagem Java, permite a visualização dos resultados do modelo via Web, conforme os exemplos mostrados nas figuras 4, 5 e 6, relativos aos 3 cenários de simulação que representam diferentes paisagens agrícolas (distintas configurações e proporções de citros e murta), para as simulações do tipo “single”.

Nas figuras 4, 5 e 6 é visualizada a dinâmica do espalhamento do HLB, depois de um determinado número de dias após o início da simulação, que se dá a partir da chegada de insetos em plantas aleatórias das duas primeiras colunas dos talhões da esquerda. A quantidade de insetos infectivos que chegam à área depende da proporção de insetos por planta e do valor de PIP escolhidos pelo usuário. Por exemplo, para a proporção de 0,41 insetos por planta, tem-se 1.033 insetos no início da simulação, dos quais: a) para PIP = 0,1%: 1 inseto infectivo; b) para PIP = 1%: 10 insetos infectivos; c) para PIP = 15%: 154 insetos infectivos.

Figura 4.

Imagem da simulação após 10 dias para o Cenário 1.

HLB Biomath2 - Simulador MBI



Para análise de técnicas de repulsão e atração, os cenários 1, 2 e 3 foram testados separadamente para inúmeras combinações dos parâmetros anteriormente citados. Não foram encontradas diferenças visuais nos gráficos gerados pela execução “multi”, bem como para as dinâmicas observadas nas execuções “single”. Na sequência das análises optou-se por comparar os cenários dois a dois, e foram executadas várias simulações para cada cenário. Testes estatísticos foram realizados para comparar o tempo de chegada da doença no talhão-alvo, e em todas as comparações entre cenários, considerando diferentes probabilidades de infecção primária, verificou-se que, estatisticamente, não há diferença (comparações entre PIP igual a 1% e 15%) com relação ao tempo de chegada da doença no talhão-alvo.

Os resultados das simulações comprovam observações realizadas em experimentos de campo ao longo da execução deste estudo: a infecção primária possui muito mais peso na dinâmica de propagação da doença do que as diferentes configurações espaciais de murta e citros consideradas nos cenários de simulação.

Dessa forma, a principal conclusão obtida é que a simples presença do hospedeiro alternativo (murta) não influencia significativamente o processo epidêmico. Isso nos leva a questionar como seria a interação do sistema “HLB-inseto vetor-citros” com o uso de métodos de controle populacional do inseto vetor, como a aplicação de inseticidas, que poderia afetar significativamente a infecção primária.

Ao mesmo tempo, a busca de um valor limiar (*threshold*) de infecção primária pode nos levar a estimar o esforço do manejo regional para estabilizar o processo epidêmico. Além disso, a determinação de níveis de infectividade dos vetores pode ser um indicador, a ser usado no futuro, da efetividade de medidas de controle do manejo regional. Esse indicador pode ser obtido mais facilmente do que extensos levantamentos com plantas infectadas.

Atualmente (Barbosa, 2019) o MBI encontra-se em evolução, com a inclusão de novos hospedeiros alternativos para avaliação em configurações de repulsão e atração, bem como para testar estratégias de controle periódico de inseticidas, minimizando o efeito da infecção primária nas paisagens em estudo.

HLB Biomath2 - Simulador MBI

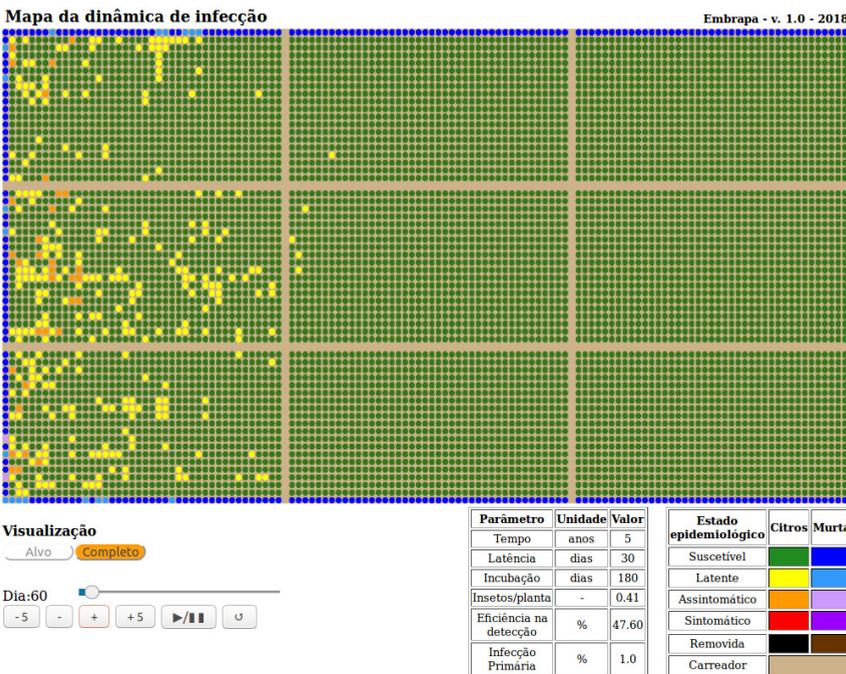


Figura 5. Imagem da tela após 60 dias de simulação para o Cenário 2.

A partir da observação da dinâmica espaço-temporal do HLB do citros representada no modelo é possível simular cenários de disseminação complexos e realizar a seleção de configurações de repulsão e atração mais promissoras para o controle da disseminação do inseto vetor, que poderão ser testadas em experimentos futuros, assim como obter indicadores de eficácia com potencial para estudos mais detalhados em outros projetos.

Figura 6.

Imagem da tela após 60 dias de simulação para o Cenário 3.

HLB Biomath2 - Simulador MBI



3.2 Avaliação genética de animais

Programas de melhoramento genético de animais têm por objetivo incrementar geneticamente a população quanto a características econômicas demandadas pelo mercado, adotando índices apropriados para o sistema de produção. Como linhas básicas, consideram a identificação e a discriminação genética dos indivíduos da população, a seleção dos superiores para reposição, quer machos ou fêmeas, e o acasalamento entre eles. Parte integrante desses programas são os processos de avaliação genética, que consistem em coletar contínua e cumulativamente dados biométricos e genealógicos da população em melhoramento e, periodicamente, utilizar um modelo genético-estatístico para prever os valores genéticos de cada animal. Os dados coletados incluem a observação da expressão dos atributos

físicos ou comportamentais de interesse do mercado, denominados fenótipos, e os dados de pedigree, ou seja, as relações que definem a genealogia da população.

Atualmente a metodologia usada em avaliações genéticas de animais é baseada na teoria dos modelos mistos (Henderson, 1963), conhecida como BLUP (Best Linear Unbiased Prediction). Consiste, basicamente, na predição dos valores genéticos, ajustando-se os dados, concomitantemente, para os efeitos fixos e um número desigual de observações por classe (Lopes, 2005). Dentre as vantagens de uma avaliação genética usando BLUP estão: inclusão da informação completa de família por meio da matriz de parentesco; comparação de indivíduos de diferentes níveis de efeitos fixos; avaliação simultânea de reprodutores, fêmeas e progênes; avaliação de indivíduos sem observações, com observações perdidas e com observações em apenas algumas características (Lopes, 2005). O BRBLUP (Higa, 2020) é um software para avaliação genética de animais desenvolvido pela Embrapa, baseado na plataforma de desenvolvimento da linguagem de programação Python e nas bibliotecas associadas de computação científica denominadas Scipy/Numpy e PyTables. Ele suporta a especificação de equações de modelos mistos, tal que diferentes modelos genético-estatísticos podem ser especificados, incluindo o modelo animal multivariado (MAM), que simultaneamente avalia os efeitos fixos e aleatórios para um conjunto de fenótipos quantitativos, levando em consideração as correlações tanto entre os fenótipos quanto entre os efeitos aleatórios, tais como os efeitos de origem genética.

Para ilustrar a utilização do BRBLUP, utilizamos um exemplo didático (exemplo 5.4 de Mrode (2014)). Nesse exemplo é considerado um modelo animal com 2 fenótipos (bivariado): (i) FAT1: rendimento de gordura no período de lactação 1; (ii) FAT2: rendimento de gordura no período de lactação 2. Associado a cada fenótipo também é considerada a presença de um efeito fixo referente ao grupo de contemporâneos ou *herd-year-season* (HYS1 e HYS2). O conjunto de dados é apresentado na Tabela 1: são 8 animais, numerados de 0 a 7, sendo que apenas aqueles que possuem fenótipos observados (os animais 0, 1 e 2) aparecem no pedigree (colunas Sire e Dam). As variâncias residuais são 65 para o fenótipo FAT1 e 70 para o fenótipo FAT2, com a covariância entre eles igual a 27; já as variâncias genéticas são 35 para o fenótipo FAT1 e 30 para o fenótipo FAT2, com a covariância entre eles igual a 28.

Para resolução do modelo, o software BRBLUP é executado por meio de linha de comando, passando como parâmetro um arquivo de configuração com a especificação do modelo, sendo o resultado armazenado em um arquivo de saída.

Tabela 1.

Conjunto de dados (colunas 0, 1, 2: pedigree – colunas 0, 3, 4, 5, 6: dados observados).

Fonte: Adaptado de Mrode (2014).

| Animal | Pai | Mãe | HYS1 | HYS2 | FAT1 | FAT2 |
|--------|-----|-----|------|------|------|------|
| 3 | 0 | 1 | 0 | 0 | 201 | 280 |
| 4 | 2 | 1 | 0 | 1 | 150 | 200 |
| 5 | 0 | 4 | 1 | 0 | 160 | 190 |
| 6 | 2 | 3 | 0 | 0 | 180 | 250 |
| 7 | 0 | 6 | 1 | 1 | 285 | 300 |

A Tabela 2 apresenta o conteúdo do arquivo de saída gerado. Ele contém 4 colunas (*Trait*: coluna no arquivo de dados correspondente a um fenótipo; *Effect*: efeito especificado no modelo; *Level*: nível do efeito no arquivo de dados; *Sol*: solução obtida). Neste exemplo, a primeira linha do arquivo significa que a solução para o nível 0 do efeito 1 (HYS1) para o fenótipo na coluna 3 é igual a 175.73126996362862. Já a sétima linha significa que o nível 1 (animal 1) para o efeito 0 (valor genético) para o fenótipo na linha 3 é igual a -2.999142788478058.

Tabela 2.

Resultado da avaliação genética.

| Característica | Efeito | Nível | Solução |
|----------------|--------|-------|---------------------|
| 3 | 1 | 0 | 175.73126996362862 |
| 3 | 1 | 1 | 219.61329398893875 |
| 4 | 2 | 0 | 243.23908674216108 |
| 4 | 2 | 1 | 240.54972646633607 |
| 3 | 0 | 0 | 8.969159144237393 |
| 4 | 0 | 0 | 8.840288629082728 |
| 3 | 0 | 1 | -2.999142788478058 |
| 4 | 0 | 1 | -2.7772802747175986 |
| 3 | 0 | 2 | -5.970016355758499 |
| 4 | 0 | 2 | -6.063008354365654 |
| 3 | 0 | 3 | 11.75424243135119 |
| 4 | 0 | 3 | 11.657587566164255 |

Continua...

| Característica | Efeito | Nível | Solução |
|----------------|--------|-------|---------------------|
| 3 | 0 | 4 | -16.252956614066754 |
| 4 | 0 | 4 | -15.823507978243187 |
| 3 | 0 | 5 | -17.31429689333114 |
| 4 | 0 | 5 | -15.719126003080525 |
| 3 | 0 | 6 | 8.690473723985185 |
| 4 | 0 | 6 | 8.137644915235219 |
| 3 | 0 | 7 | 22.702139483291525 |
| 4 | 0 | 7 | 20.930688340763133 |

Tabela 2.
Continuação.

Neste exemplo não foram apresentados os correspondentes valores de acurácia, que representam a confiabilidade da solução obtida para o valor genético e são sempre utilizados em conjunto. Finalmente, outro aspecto não abordado refere-se ao fato de que, atualmente, os programas de melhoramento genético animal estão desenvolvendo esforços para inclusão de informações genômicas no processo de avaliação genética, com implicações diretas na construção e na resolução do modelo genético-estatístico empregado.

3.3 Fazenda Pantaneira Sustentável (FPS)

Nas últimas décadas, diante da globalização da economia e da criação de mercados competitivos, intensificaram-se as pressões para aumento de produtividade das fazendas pantaneiras, comprometendo a sustentabilidade de seus sistemas de produção, dada a fragilidade dos ecossistemas. Diante desse cenário, um grupo multidisciplinar de pesquisadores da Embrapa Pantanal, utilizando experiência anterior voltada para a caracterização das fazendas pantaneiras (Santos et al., 2017), elaborou projeto em parceria com a Embrapa Informática Agropecuária para desenvolver uma ferramenta para avaliar a sustentabilidade de sistemas de produção de gado de corte em regiões complexas e dinâmicas, como é o caso do Pantanal, de modo que fosse possível verificar os pontos fracos do sistema e buscar boas práticas de manejo para obter a sustentabilidade.

O Pantanal está localizado na região Centro-Oeste do Brasil (80%), abrangendo parte da Bolívia e do Paraguai. Constitui uma extensa planície periodicamente inundável, com variabilidade temporal e espacial de diversidade controlada pelo pulso de inundação, o que torna a região um sistema complexo, dinâmico e incerto (Santos et al., 2017). Pelo fato de possuir extensas áreas de campos naturais com predominância de forrageiras, o Pantanal

apresenta vocação para a criação extensiva de gado de corte com baixa utilização de insumos externos, o que tem contribuído para a sua conservação há mais de dois séculos. Essa tem sido a atividade econômica principal nas fazendas pantaneiras, consistindo num importante setor socioeconômico em nível regional e nacional. Considerando que as fazendas compreendem cerca de 95% da planície pantaneira, o principal desafio dos tomadores de decisão é definir sistemas de produção de gado de corte que não causem grandes impactos ambientais e que tragam benefícios econômicos e sociais para a população local, garantindo a conservação e o uso sustentável dos recursos naturais.

Para compreender as fazendas pantaneiras de maneira holística foram definidos, de forma hierárquica, os aspectos e os indicadores necessários para avaliar o sistema de produção de gado de corte em nível de fazenda e região (Figura 7). Esses aspectos e indicadores foram selecionados em função da praticidade em representar e simplificar fenômenos complexos e sistêmicos. Alguns dos indicadores foram definidos a partir dos estudos científicos desenvolvidos pela equipe multidisciplinar, enquanto outros foram determinados por meio da realização de diversas oficinas participativas, envolvendo tomadores de decisão para validação dos indicadores. Alguns desses indicadores devem ser avaliados diretamente no campo, enquanto outros podem ser estudados por meio de análise de imagens e cálculos matemáticos ou definidos no próprio sistema de inferência adotado. Para guiar a avaliação em campo e a coleta de informações necessárias para os cálculos, foram elaborados e publicados vários protocolos (Soares et al., 2014; Santos et al., 2014a, 2014b, 2015; Abreu et al., 2015; Amâncio et al., 2016). Esse processo hierárquico (Figura 7) permite avaliar cada aspecto da sustentabilidade de forma isolada ou em conjunto.

Figura 7.
Estrutura
hierárquica da FPS.



Análises Opcionais:

Viabilização da Pecuária | **Aptidão Produtiva Natural para a Pecuária**

3.3.1 O software FPS

Alguns problemas surgem na avaliação da sustentabilidade, sendo preciso levar em conta o nível de abstração envolvido no conceito, bem como a existência de variabilidades naturais em alguns fenômenos implicados. A síntese dos indicadores levantados em um dado “grau de sustentabilidade” exige uma metodologia robusta para tratar incertezas, exprimir inter-relações complexas e, ao mesmo tempo, ser interpretável e transparente para garantir confiança na avaliação.

Um arcabouço matemático e computacional capaz de lidar com essas dificuldades são os originários da teoria de conjuntos *fuzzy* (CF), lógica *fuzzy* e sistemas baseados em regras *fuzzy* – SBRF. Tais sistemas têm sido aplicados em áreas como engenharia, modelagem e controle, entre outras. Historicamente, seu sucesso deve-se à habilidade de modelar o conhecimento baseando-se em linguagem natural e à boa capacidade de generalização, além da notável competência dos SBRF em “explicar” a elaboração do resultado a partir dos valores de entrada fornecidos.

O software FPS foi construído como um sistema de suporte à decisão baseado em modelos expressos em SBRF. A sustentabilidade é avaliada pela dimensão ambiental, econômica e social, em nível de fazenda e regional. Para cada avaliação (Figura 7) foram definidos modelos cujas variáveis de entrada são os próprios indicadores com suas escalas definidas em linguagem natural (como Bom, Moderado e Ruim). As relações entre os indicadores são expressas como um conjunto de regras definidas pelos especialistas de domínio. Os resultados da avaliação (índices e subíndices), por sua vez, além de fornecerem um valor numérico comparativo (1 a 10), têm uma saída qualitativa correspondente. Cada modelo (índice) mais abaixo hierarquicamente alimenta os modelos mais gerais, culminando no modelo de sustentabilidade da fazenda.

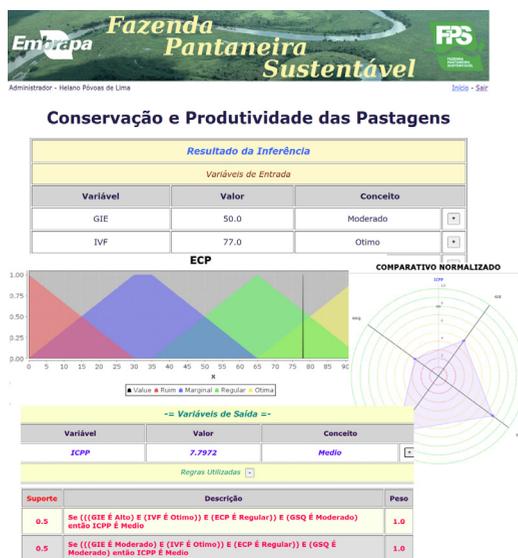
A interface da FPS para a Internet² (Figura 8) constitui um sistema interativo em que o usuário, dados os valores dos indicadores fornecidos, é capaz de inferir conceitos qualitativos e valores numéricos sobre ela, bem como comparar o quão bons são os valores em relação ao desejável. É possível também, através de gráficos, visualizar quais indicadores tiveram mais influência no resultado. Ao usuário também são enumeradas as regras que foram utilizadas para a conclusão apresentada, garantindo interpretabilidade e transparência. O sistema também permite que o usuário simule cenários para planejar quais deles levam ao nível de sustentabilidade que ele deseja. O software FPS possui ainda uma segunda interface, voltada para dispositivos móveis como *tablets* e *smartphones* (Figura 9) que utilizam sistema operacional Android (disponível

² Disponível em: <https://www.fps.cnptia.embrapa.br/>

na loja de aplicativos Google Play). Essencialmente ela fornece as mesmas funcionalidades e é baseada nos mesmos modelos matemáticos, mas, dada uma restrição regional do Pantanal e em geral, das fazendas, essa versão não precisa de conexão com a internet para funcionar, pois tem embutido no aplicativo o próprio mecanismo de inferência.

Figura 8. Elementos da interface do software FPS para a internet.

Fonte: Fazenda Pantaneira Sustentável (2020).



A ferramenta FPS é passível de adoção por diversos tomadores de decisão (pesquisadores, proprietários, técnicos, políticos, legisladores, certificadoras, entre outros). Seu principal uso está relacionado com o diagnóstico (grau de sustentabilidade) do sistema de produção de bovinos de corte na planície pantaneira, por meio da avaliação de impactos ambientais, sociais e econômicos da atividade para, assim, auxiliar na gestão eficiente mediante seleção de tecnologias e boas práticas de manejo. Porém, sua aplicação pode ser muito mais ampla por meio de subsídios aos programas de financiamento, certificação e estratégias de marketing que valorizem os produtos provenientes da região, podendo também oferecer subsídios necessários para a reformulação da legislação vigente e de políticas públicas de incentivo à produção sustentável na região. Pretende-se inserir, no futuro, o aspecto da multifuncionalidade e dos serviços ecossistêmicos, essenciais para a sustentabilidade dos sistemas de produção. A ferramenta está sendo implantada em 15 fazendas no Pantanal do Mato Grosso, com apoio da FAMATO, da ACRIMAT, do SENAR, do IMEA e de sindicatos rurais, e em seis fazendas no Pantanal do Mato Grosso do Sul, com apoio da FAMASUL, do SENAR e de sindicatos rurais. Melhorias serão incorporadas ao longo do tempo, em conjunto com técnicos, produtores e pesquisadores.

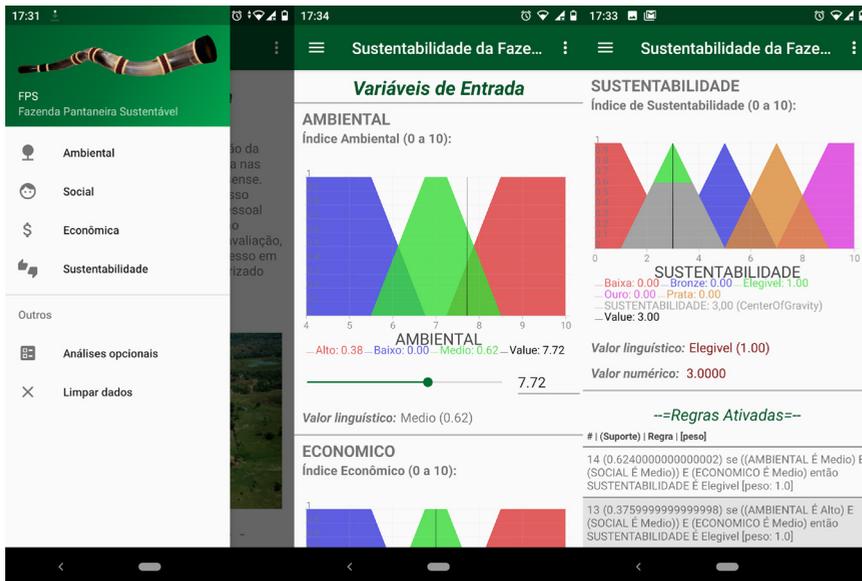


Figura 9. Interface do aplicativo Android da FPS.³

4 Considerações finais

Neste capítulo apresentaram-se várias técnicas de computação científica aplicadas na resolução de problemas do setor agropecuário. Na área de Inteligência Artificial, técnicas da lógica clássica foram aplicadas ao desenvolvimento de um sistema especialista para classificação de solos. O mesmo problema foi abordado também por uma técnica completamente diferente, com o uso de algoritmos de aprendizado de máquina, que são fundamentalmente interligados à estatística. A análise estatística é também a base das técnicas de mineração de textos usadas para agrupar documentos com conteúdos semelhantes na área agrícola.

Outra área da computação científica, a modelagem matemática, foi explorada de três formas distintas. Na primeira delas, o Modelo Baseado no Indivíduo fundamentou uma modelagem totalmente computacional, por meio de um sistema de simulação que compara três configurações de plantio envolvendo citros e murta para avaliar estratégias de controle da propagação do HLB dos citros. Na segunda aplicação, modelos de preditores lineares, compostos de equações matemáticas clássicas, foram usados para avaliar

³ A tela (em validação) estará disponível pelo endereço: <https://play.google.com/store/apps/details?id=br.embrapa.cnptia.fps>

os valores genéticos dos animais, com o objetivo de descobrir quais deles reforçam as características desejadas pelo mercado. No terceiro modelo, os cálculos matemáticos foram realizados internamente a um sistema de inferência, baseado na lógica *fuzzy*, para avaliar a sustentabilidade de fazendas pantaneiras. Explora-se, nesse caso, a vantagem da lógica *fuzzy* em aliar a linguagem natural na construção de um modelo lógico cuja resposta seja explicável ao tomador de decisão.

As técnicas de computação científica são essenciais para a análise do grande volume de dados que vêm sendo produzidos no processo de transformação digital da agricultura. Por meio dessas técnicas será possível, a partir dos dados coletados, extrair informações e conhecimentos que auxiliarão no processo de tomada de decisão em todos os elos das cadeias produtivas, tornando-se centrais no desenvolvimento de novas soluções e tecnologias agrícolas no contexto da Agricultura Digital. As aplicações apresentadas neste capítulo ilustram a variedade de problemas que podem ser abordados pelo arcabouço metodológico da computação científica, incluindo a modelagem matemática e estatística, os sistemas de lógica clássica e nebulosa, os modelos de simulação e os modelos de aprendizado de máquina.

Considerando essas aplicações, é válido ressaltar que o constante crescimento da disponibilidade de dados, o avanço tecnológico e a ampliação da dimensionalidade e da complexidade das demandas da sociedade brasileira impõem enormes desafios e oportunidades para pesquisa e desenvolvimento em computação científica aplicada à agropecuária.

5 Referências

ABREU, U. G. P.; LIMA, H. P.; SANTOS, S. A.; MASSRUHÁ, S. **Protocolo**: Índice Financeiro (IF) para a Fazenda Pantaneira Sustentável (FPS). Corumbá: Embrapa Pantanal, 2015. 12 p. (Embrapa Pantanal. Documentos, 134).

AMÂNCIO, C. O. G.; ARAÚJO, M. T. B.; SANTOS, S. A.; NARCISO, M.; OLIVEIRA, M. D. **Protocolo**: Índice de Bem-Estar Social (IBS) para a Fazenda Pantaneira Sustentável (FPS). Corumbá: Embrapa Pantanal, 2016. 16 p. (Embrapa Pantanal. Documentos 139).

BARBOSA, F. F. L. **HLB BioMath fase 2**: abordagem biomatemática como suporte a defesa fitossanitária e avaliação ex-ante de tecnologias de manejo. Cruz das Almas: Centro Nacional de Pesquisa de Mandioca e Fruticultura, 2015. 26 p. (Embrapa. Macroprograma 2 - Código SEG 02.13.03.007.00.000).

BARBOSA, F. F. L. **HLB BioMath fase 3**: biomatemática aplicada à otimização de tecnologias de interposição de barreiras, modificação microambiental e exclusão para manejo do huanglongbing dos citros. Cruz das Almas: Embrapa Mandioca e Fruticultura, 2019. 40 p. (Embrapa. Tipo II - Código SEG 20.18.03.044.00.00).

EMBRAPA. **Bases de Dados da Pesquisa Agropecuária**. Disponível em: <https://www.bdpa.cnptia.embrapa.br>. Acesso em: 19 maio 2020.

FAZENDA PANTANEIRA SUSTENTÁVEL. Disponível em: <https://www.fps.cnptia.embrapa.br/>. Acesso em: 19 maio 2020.

FUNDECITRUS. **Sumário executivo:** estimativa da safra de laranja 2020/21 do cinturão citrícola de São Paulo e Triângulo/Sudoeste Mineiro: cenário em maio de 2019. Disponível em: https://www.fundecitrus.com.br/pdf/pes_relatorios/2020_05_11_Sumario-Executivo-da-Estimativa-da-Safra-de-Laranja-2020-2021.pdf. Acesso em: 19 maio 2020.

GRIMM, V.; RAILSBACK, S. F. **Individual-based modeling and ecology**. Princeton: Princeton University Press, 2005. DOI: [10.1515/9781400850624](https://doi.org/10.1515/9781400850624).

GUYON, I.; ELISSEEFF, A. An introduction to variable and feature selection. **Journal of Machine Learning Research**, v. 3, p. 1157-1182, 2003.

HENDERSON, C. R. Selection index and expected genetic advance. **Statistical Genetics and Plant Breeding**, v. 982, p. 141-163, 1963.

HIGA, R. H. **Tutorial:** introdução ao software brblup. Campinas: Embrapa Informática Agropecuária, 2020. (Embrapa Informática Agropecuária. Documentos, 168). 2020. No prelo.

LOPES, P. S. **Teoria do melhoramento animal**. Belo Horizonte: FEPMVZ, 2005. 118 p.

MOURA, M. F.; TAKEMURA, C. M.; SILVA, I. L. C.; TÁPIAS, L. M.; OLIVEIRA, C. T. de; BASSOI, L. H.; OLIVEIRA, S. R. de M. Metodologia para a construção de portfólios tecnológicos agrícolas a partir de publicações técnico-científicas. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 11., 2017, Campinas. **Ciência de dados na era da agricultura digital:** anais. Campinas: Ed. Unicamp: Embrapa Informática Agropecuária, 2017. p. 537-546. SBIAgro 2017.

MRODE, R. A. **Linear models for the prediction of animal breeding values**. 3rd ed. CAB, 2014. DOI: [10.1079/9781780643915.0000](https://doi.org/10.1079/9781780643915.0000).

PEREIRA, R. G.; MOURA, M. F. I-Preproc: uma ferramenta para pré-processamento e indexação incremental de documentos. In: MOSTRA DE ESTAGIÁRIOS E BOLSISTAS DA EMBRAPA INFORMÁTICA AGROPECUÁRIA, 11., 2015, Campinas. **Resumos expandidos**. Brasília, DF: Embrapa, 2015. p. 17-23.

POLIDORO, J. C.; MENDONÇA-SANTOS, M. de L.; LUMBRERAS, J. F.; COELHO, M. R.; CARVALHO FILHO, A. de; MOTTA, P. E. F. da; CARVALHO JUNIOR, W. de; ARAUJO FILHO, J. C. de; CURCIO, G. R.; CORREIA, J. R.; MARTINS, E. de S.; SPERA, S. T.; OLIVEIRA, S. R. de M.; BOLFE, E. L.; MANZATTO, C. V.; TOSTO, S. G.; VENTURIERI, A.; SA, I. B.; OLIVEIRA, V. A. de; SHINZATO, E.; ANJOS, L. H. C. dos; VALLADARES, G. S.; RIBEIRO, J. L.; MEDEIROS, P. S. C. de; MOREIRA, F. M. de S.; SILVA, L. S. L.; SEQUINATTO, L.; AGLIO, M. L. D.; DART, R. de O. **Programa Nacional de Solos do Brasil (PronaSolos)**. Rio de Janeiro: Embrapa Solos, 2016. 53 p. (Embrapa Solos. Documentos, 183).

SANTOS, H. G. dos; JACOMINE, P. K. T.; ANJOS, L. H. C. dos; OLIVEIRA, V. A. de; LUMBRERAS, J. F.; COELHO, M. R.; ALMEIDA, J. A. de; CUNHA, T. J. F.; OLIVEIRA, J. B. de. **Sistema brasileiro de classificação de solos**. 3. ed. rev. ampl. Brasília, DF: Embrapa, 2013. 353 p.

SANTOS, S. A.; CARDOSO, E. L.; CRISPIM, S. M. A.; SORIANO, B. M. A.; GARCIA, J. B.; BERSELLII, C. **Protocolo:** Índice de Conservação e Produtividade das Pastagens (ICPP) para a Fazenda Pantaneira Sustentável (FPS). Corumbá: Embrapa Pantanal, 2014a. 18 p. (Embrapa Pantanal. Documentos, 130).

SANTOS, S. A.; LIMA, H. P. de; BALDIVIESO, H. P.; OLIVEIRA, L. O.; TOMÁS, W. M. GIS-fuzzy logic approach for building indices: regional feasibility and natural potential of ranching in tropical wetland. **Journal of Agricultural Informatics**, v. 5, n. 2, p. 26-33, 2014b. DOI: [10.17700/jai.2014.5.2.140](https://doi.org/10.17700/jai.2014.5.2.140).

SANTOS, S. A.; LIMA, H. P. de; MASSUHÁ, S. M. F. S.; ABREU, U. G. P. de; TOMÁS, W. M.; SALIS, S. M.; CARDOSO, E. L.; OLIVEIRA, M. D. de; SOARES, M. T. S.; SANTOS JR., A. dos; OLIVEIRA, L. O. F. de; CALHEIROS, D. F.; CRISPIM, S. M. A.; SORIANO, B. M. A.; AMÂNCIO, C. O. G.; NUNES, A. P.; PELLEGRIN, L. A. A fuzzy logic-based tool to assess beef cattle ranching sustainability in complex environmental systems. **Journal of Environmental Management**, v. 198, part 2, p. 95-106, Aug 2017. DOI: [10.1016/j.jenvman.2017.04.076](https://doi.org/10.1016/j.jenvman.2017.04.076).

SANTOS, S. A.; OLIVEIRA, L. O. F.; LIMA, H. P.; ABREU, U. G. P.; OLIVEIRA, M. D.; ARAÚJO, M. T. B. D.

Protocolo: Índice de Manejo e Bem-Estar do Rebanho (IMBA) para a Fazenda Pantaneira Sustentável (FPS). Corumbá: Embrapa Pantanal, 2015. 20 p. (Embrapa Pantanal. Documentos, 135).

SOARES, M. T. S.; OLIVEIRA, M. D.; CALHEIROS, D. F.; SANTOS, S. A.; LIMA, H. P. **Protocolo:** Índice de Conservação de Corpos de Água Naturais (ICA) para a Fazenda Pantaneira Sustentável (FPS). Corumbá: Embrapa Pantanal, 2014. 22 p. (Embrapa Pantanal. Documentos, 128).

SOUZA, K. X. S.; TERNES, S.; OLIVEIRA, S. R. M.; MOURA, M. F.; BARIONI, L. G.; HIGA, R. H.; FASIABEN, M. C. R. A perspective study on the application of Data Science in agriculture. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 11., 2017, Campinas. **Ciência de dados na era da agricultura digital:** anais. Campinas: Ed. Unicamp: Embrapa Informática Agropecuária, 2017. p. 537-546. SBIAgro 2017.

TAKEMURA, C. M.; MOURA, M. F.; MACHADO, L. S. C. **TopExtract – toponym extraction and disambiguation tool:** componente de software para extração e desambiguação de topônimos. Campinas: Embrapa Monitoramento por Satélite, 2013. 1 CD-ROM.

VAZ, G. J.; SILVA NETO, L. de F. da; LIMA, R. N.; MARQUES, F. A. M.; SANTOS, J. C. P. dos; OLIVEIRA, S. R. de M. Curadoria de dados de solos brasileiros por meio de um sistema especialista de classificação de solos. In: CONGRESSO BRASILEIRO DE CIÊNCIA DO SOLO, 37., 2019a, Cuiabá. **Intensificação sustentável em sistemas de produção:** resumos. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 2019.

VAZ, G. J.; SILVA NETO, L. de F. da; LIMA, R. N.; OLIVEIRA, S. R. de M. Uma API para a classificação de solos do Brasil. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 12., 2019, Jaguariúna, SP. **IoT na Agricultura:** anais. Campinas: Embrapa Informática Agropecuária, 2019b. SBIAgro 2019.

VAZ, G. J.; SILVA NETO, L. de F. da; OLIVEIRA, S. R. de M.; BOTELHO, F. P.; ARAUJO FILHO, J. C. de. Development of an expert system for classification of Brazilian soil profiles. In: WORLD CONGRESS OF SOIL SCIENCE, 21., 2018, Rio de Janeiro. Soil science: beyond food and fuel: abstracts. Viçosa, MG: SBCS, 2018. Não paginado. WCSS 2018. Publicado também em: WORLD CONGRESS OF SOIL SCIENCE, 21., 2018, Rio de Janeiro. Soil science beyond food and fuel: proceedings. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 2019. v. 1, p. 56-57.

VILAMIU, R. G. d'A.; TERNES, S.; LARANJEIRA, F. F.; SANTOS, T. S. Modelling the effect of an alternative host population on the spread of citrus Huanglongbing. **AIP Conference Proceedings**, v. 1558, p. 2504-2508, 2013.