DIGITAL AGRICULTURE

Research, development and innovation in production chains

Silvia Maria Fonseca Silveira Massruhá
Maria Angelica de Andrade Leite
Stanley Robson de Medeiros Oliveira
Carlos Alberto Alves Meira
Ariovaldo Luchiari Junior
Édson Luis Bolfe

Technical editors

Embrapa

Brazilian Agricultural Research Corporation
Embrapa Digital Agriculture
Ministry of Agriculture and Livestock

DIGITAL AGRICULTURE

Research, development and innovation in production chains

Silvia Maria Fonseca Silveira Massruhá
Maria Angelica de Andrade Leite
Stanley Robson de Medeiros Oliveira
Carlos Alberto Alves Meira
Ariovaldo Luchiari Junior
Édson Luis Bolfe

Technical editors

Translated by
Beverly Victoria Young
Karl Stephan Mokross

Embrapa
Brasília, DF
2023

Responsible for content
Embrapa Digital Agriculture
Av. André Tosello, 209, Barão Geraldo
13083-886 Campinas, SP
Telefone: (19) 3211-5700
www.embrapa.br
www.embrapa.br/fale-conosco/sac

Originally published as
"Agricultura Digital: pesquisa, desenvolvimento
e inovação nas cadeias produtivas"
ISBN 978-65-86056-37-2

Responsible for publication
Embrapa, Superintendência de Comunicação

Editorial coordination
Daniel Nascimento Medeiros
Nilda Maria da Cunha Sette

Editorial supervision
Cristiane Pereira de Assis

Text revision
Ana Maranhão Nogueira
Jane Baptistone de Araujo

Bibliographic standardization
Rejane Maria de Oliveira
Márcia Maria Pereira de Souza

Graphic design, layout and cover
Carlos Eduardo Felice Barbeiro

Cover
Unsplash with art forms by Marília Bastos Pícollo Silva

1st edition
Digital publication (2023): PDF

Cataloging-in-Publication (CIP) Data
Embrapa, Superintendência de Comunicação

Digital agriculture : research, development and innovation in production chains / Silvia Maria Fonseca Silveira Massruhá ... [et al.]; Beverly Victoria Young, Karl Stephan Mokross, translators. — Brasília, DF : Embrapa, 2023.
PDF (299 p.) : color. il.

Translation of: Agricultura digital: pesquisa, desenvolvimento e inovação nas cadeias produtivas, 2020.

ISBN 978-65-89957-72-0

1. Agriculture. 2. Digital agriculture. 3. Information technology. 4. Digital transformation. I. Leite, Maria Angelica de Andrade. II. Oliveira, Stanley Robson de Medeiros. III. Meira, Carlos Alberto Alves. IV. Luchiari Junior, Ariovaldo. V. Bolfe, Édson Luis. VI. Embrapa Digital Agriculture.

CDD (21. ed.) 004

Rejane Maria de Oliveira Cechinel Darós (CRB-1/2913)

Technical editors

Silvia Maria Fonseca Silveira Massruhá
Systems analyst, Ph.D. in Applied Computing, researcher at Embrapa Digital Agriculture, Campinas, SP

Maria Angelica de Andrade Leite
Civil engineer, Ph.D. in Computer Engineering, researcher at Embrapa Digital Agriculture, Campinas, SP

Stanley Robson de Medeiros Oliveira
Computer scientist, Ph.D. in Computer Science, researcher at Embrapa Digital Agriculture, Campinas, SP

Carlos Alberto Alves Meira
Mathematician, Ph.D. in Agricultural Engineering, researcher at Embrapa Digital Agriculture, Campinas, SP

Ariovaldo Luchiari Junior
Agronomist, Ph.D. in Agronomy and Soils, researcher at Embrapa Digital Agriculture, Campinas, SP

Édson Luis Bolfe
Forest engineer, Ph.D. in Geography, researcher at Embrapa Digital Agriculture, Campinas, SP

Authors

Adhemar Zerlotini Neto
Computer scientist, Ph.D. in Bioinformatics, researcher at Embrapa Digital Agriculture, Campinas, SP

Adriana Farah Gonzalez
Bachelor of languages, specialist in School Administration, technician at Embrapa Digital Agriculture, Campinas, SP

Alan Massaru Nakai
Computer scientist, Ph.D. in Computer Science, analyst at Embrapa Digital Agriculture, Campinas, SP

Alexandre Camargo Coutinho
Biologist, Ph.D. in Environmental Sciences, researcher at Embrapa Digital Agriculture, Campinas, SP

Alexandre de Castro
Physicist, Ph.D. in Science, researcher at Embrapa Digital Agriculture, Campinas, SP

Ana Paula Ribeiro
Biotechnologist, Ph.D. in Plant Biotechnology, postdoctoral fellow at the State University of Campinas and at the Genomics for Climate Change Research Center, Campinas, SP

André Fachini Minitti
Agronomist, analyst at Embrapa Digital Agriculture, Campinas, SP

Antonio Nhani Junior
Biologist, Ph.D. in Biochemistry and Molecular Biology, researcher at Embrapa Digital Agriculture, Campinas, SP

Ariovaldo Luchiari Junior
Agronomist, Ph.D. in Agronomy and Soils, researcher at Embrapa Digital Agriculture, Campinas, SP

Aryeverton Fortes de Oliveira
Economist, Ph.D. in Applied Economics, researcher at Embrapa Digital Agriculture, Campinas, SP

Balbina Maria Araujo Soriano
Meteorologist, Ph.D. in Agronomy, researcher at Embrapa Pantanal, Corumbá, MS

Breno Silva Beda de Assunção
Computer scientist, specialist in Digital Transformation, analyst at Embrapa, Brasília, DF

Carla Geovana do Nascimento Macário
Data processing technologist, Ph.D. in Computer Science, researcher at Embrapa Digital Agriculture, Campinas, SP

Carlos Alberto Alves Meira
Mathematician, Ph.D. in Agricultural Engineering, researcher at Embrapa Digital Agriculture, Campinas, SP

Carlos Cesar Ronquim
Agronomist, Ph.D. in Ecology and Natural Resources, researcher at Embrapa Territorial, Campinas, SP

Carlos Manuel Pedro Vaz
Physicist, Ph.D. in Science, researcher at Embrapa Instrumentation, São Carlos, SP

Cássia Isabel Costa Mendes
Lawyer, Ph.D. in Economic Development, analyst at Embrapa Digital Agriculture, Campinas, SP

Célia Regina Grego
Agronomist, Ph.D. in Energy in Agriculture, researcher at Embrapa Digital Agriculture, Campinas, SP

Celina Maki Takemura
Computer scientist, Ph.D. in Computer Science, researcher at Embrapa Territorial, Campinas, SP

Cleidson Nogueira Dias
Administrator, Ph.D. in Administration, analyst at Embrapa, Secretariat of Innovation and Business, Brasília, DF

Cornélio Alberto Zolin
Agricultural engineer, Ph.D. in Irrigation and Drainage, researcher at Embrapa Agrosilvopastoral, Sinop, MT

Cristina Aparecida Gonçalves Rodrigues
Zootechnician, Ph.D. in Plant Biology, researcher at Embrapa Territorial, Campinas, SP

Daniel de Castro Victoria
Agronomist, Ph.D. in Sciences, researcher at Embrapa Digital Agriculture, Campinas, SP

Debora Pignatari Drucker
Forest engineer, Ph.D. in Environment and Society, analyst at Embrapa Digital Agriculture, Campinas, SP

Edgard Henrique dos Santos
Computer scientist, analyst at Embrapa Digital Agriculture, Campinas, SP

Édson Luis Bolfe
Forest Engineer, Ph.D. in Geography, researcher at Embrapa Digital Agriculture, Campinas, SP

Eduardo Antonio Speranza
Bachelor's degree in Computer Science, Ph.D. in Computer Science, analyst at Embrapa Digital Agriculture, Campinas, SP

Eduardo Delgado Assad
Agricultural engineer, Ph.D. in Agrometeorology, researcher at Embrapa Digital Agriculture and professor at Fundação Getúlio Vargas, Campinas, SP

Enilda Coelho
Computer scientist, master's degree in Information Science, analyst at Embrapa Maize & Sorghum, Sete Lagoas, MG

Fábio César da Silva
Agronomist, doctor in Soils and Plant Nutrition, researcher at Embrapa Digital Agriculture, Campinas, SP

Fábio Danilo Vieira
Data processing technologist, master's degree in Agricultural Engineering, analyst at Embrapa Digital Agriculture, Campinas, SP

Fernanda Rausch Fernandes
Agronomist, Ph.D. in Phytopathology, researcher at Embrapa Digital Agriculture and at the Genomics for Climate Change Research Center, Campinas, SP

Flávia Bussaglia Fiorini
Public relations, specialist in Communication and Marketing Management, analyst at Embrapa Digital Agriculture, Campinas, SP

Francisco Ferraz Laranjeira Barbosa
Agronomist, Ph.D. in Phytopathology, researcher at Embrapa Cassava & Fruits, Cruz das Almas, BA

Geraldo Bueno Martha Júnior
Agronomist, Ph.D. in Agronomy, researcher at Embrapa Digital Agriculture, Campinas, SP

Giampaolo Queiroz Pellegrino
Forest Engineer, Ph.D. in Water and Soil, researcher at Embrapa Digital Agriculture, Campinas, SP

Gilberto Romeiro de Oliveira Menezes
Zootechnician, Ph.D. in Genetics and Breeding, researcher at Embrapa Beef Cattle, Campo Grande, MS

Glauber José Vaz
Computer scientist, master's degree in Computer Science, analyst at Embrapa Digital Agriculture, Campinas, SP

Goran Neshich
Physical chemist, Ph.D. in Molecular Biophysics, researcher at Embrapa Digital Agriculture, Campinas, SP

Guiomar Alessandra de Souza Telles
Bachelor of languages, technician at Embrapa Digital Agriculture, Campinas, SP

Gustavo Costa Rodrigues
Agronomist, master's degree in Plant Physiology, researcher at Embrapa Digital Agriculture, Campinas, SP

Helano Póvoas de Lima
Computer scientist, master's degree in Computer Science, analyst at Embrapa Digital Agriculture, Campinas, SP

Inácio Henrique Yano
Data processing technologist and economist, Ph.D. in Agricultural Engineering, analyst at Embrapa Digital Agriculture, Campinas, SP

Isabel Rodrigues Gerhardt
Agronomist, Ph.D. in Molecular Biology, researcher at Embrapa and at Genomics for Climate Change Research Center, Campinas, SP

Isaque Vacari
Technologist in data processing, master's degree in Computer Science, analyst at Embrapa Digital Agriculture, Campinas, SP

Ivan Mazoni
Technologist in data processing, Ph.D. in Bioinformatics, analyst at Embrapa Digital Agriculture, Campinas, SP

Ivo Pierozzi Júnior
Biologist, Ph.D. in Ecology, researcher at Embrapa Digital Agriculture, Campinas, SP

Jaderson Silveira Leite Armanhi
Armanhi Biologist, Ph.D. in Genetics and Molecular Biology, postdoctoral fellow at the State University of Campinas and at Genomics for Climate Change Research Center, Campinas, SP

Jayme Garcia Arnal Barbedo
Electrical engineer, Ph.D. in Electrical Engineering, researcher at Embrapa Digital Agriculture, Campinas, SP

João Camargo Neto
Electrical engineer, Ph.D. in Agricultural and Biosystems Engineering, analyst at Embrapa Digital Agriculture, Campinas, SP

João dos Santos Vila da Silva
Mathematician, Ph.D. in Agricultural Engineering, researcher at Embrapa Digital Agriculture, Campinas, SP

João Francisco Gonçalves Antunes
Mathematician, Ph.D. in Agricultural Engineering, researcher at Embrapa Digital Agriculture, Campinas, SP

Joice Machado Bariani
Bachelor's degree in Biological Sciences, master's degree in Genetics and Molecular Biology, analyst at Embrapa Digital Agriculture, Campinas, SP

José Eduardo Boffino de Almeida Monteiro
Agronomist, Ph.D. in Agrometeorology, researcher at Embrapa Digital Agriculture, Campinas, SP

Juliana Erika de Carvalho Teixeira Yassitepe
Agronomist, Ph.D. in Genetics and Plant Breeding, researcher at Embrapa Digital Agriculture and at Genomics for Climate Change Research Center, Campinas, SP

Juliano Souza de Albuquerque Maranhão
Lawyer, Doctor of Law, professor at the University of São Paulo, São Paulo, SP

Júlio César Dalla Mora Esquerdo
Agronomist, Ph.D. in Agricultural Engineering, researcher at Embrapa Digital Agriculture, Campinas, SP

Júlio Cesar Franchini dos Santos
Agronomist, Ph.D. in Sciences, researcher at Embrapa Soybean, Londrina, PR

Kleber Xavier Sampaio de Souza
Electrical engineer, Ph.D. in Electrical Engineering, researcher at Embrapa Digital Agriculture, Campinas, SP

Ladislau Marcelino Rabello
Electrical engineer, Ph.D. in Electrical Engineering, researcher at Embrapa Instrumentation, São Carlos, SP

Laurimar Gonçalves Vendrusculo
Electrical engineer, Ph.D. in Agricultural Engineering and Biosystems, researcher at Embrapa Digital Agriculture, Campinas, SP

Leandro Carrijo Cintra
Computer scientist, Ph.D. in Bioinformatics, analyst at Embrapa Digital Agriculture, Campinas, SP

Líbia Cristina Xavier Santos
Lawyer, master's student in Intellectual Property and Technology Transfer for Innovation, analyst at Embrapa, Brasília, DF

Luciana Alvim Santos Romani
Computer scientist, Ph.D. in Computer Science, researcher at Embrapa Digital Agriculture, Campinas, SP

Luciano Vieira Koenigkan
Computer scientist, master's degree in Electrical Engineering, analyst at Embrapa Digital Agriculture, Campinas, SP

Lúcio André de Castro Jorge
Electrical engineer, Ph.D. in Electrical Engineering, researcher at Embrapa Instrumentation, São Carlos, SP

Luís Gustavo Barioni
Agronomist, Ph.D. in Animal Science and Pastures, researcher at Embrapa Digital Agriculture, Campinas, SP

Luiz Manoel Silva Cunha
Statistician, master's degree in Computer Science and Computational Mathematics, analyst at Embrapa Digital Agriculture, Campinas, SP

Luiz Otávio Campos da Silva
Zootechnician, Ph.D. in Genetics and Breeding, researcher at Embrapa Beef Cattle, Campo Grande, MS

Magda Cruciol
Public relations, analyst at Embrapa Digital Agriculture, Campinas, SP

Márcia Divina de Oliveira
Biologist, Ph.D. in Limnology, researcher at Embrapa Pantanal, Corumbá, MS

Marcia Izabel Fugisawa Souza
Bachelor of librarianship, Ph.D. in Education, analyst at Embrapa Digital Agriculture, Campinas, SP

Márcio José da Silva
Biologist, Ph.D. in Genetics and Molecular Biology, researcher at the State University of Campinas and at Genomics for Climate Change Research Center, Campinas, SP

Marcos Cezar Visoli
Computer scientist, master's degree in Informatics and Systems, researcher at Embrapa Digital Agriculture, Campinas, SP

Maria Angelica de Andrade Leite
Civil engineer, Ph.D. in Computer Engineering, researcher at Embrapa Digital Agriculture, Campinas, SP

Maria do Carmo Ramos Fasiaben
Agronomist, Ph.D. in Economic Development, researcher at Embrapa Digital Agriculture, Campinas, SP

Maria Fernanda Moura
Statistician, Ph.D. in Mathematics and Computer Science, researcher at Embrapa Digital Agriculture, Campinas, SP

Marília Ieda da Silveira Folegatti Matsuura
Zootechnician, Ph.D. in Food Technology, researcher at Embrapa Environment, Jaguariúna, SP

Martha Delphino Bambini
Chemical engineer, Ph.D. in Science and Technology Policy, analyst at Embrapa Digital Agriculture, Campinas, SP

Mateus Batistella
Biologist, Ph.D. in Environmental Science, researcher at Embrapa Digital Agriculture, Campinas, SP

Maurício de Alvarenga Mudadu
Biologist, Ph.D. in Bioinformatics, researcher at Embrapa Digital Agriculture, Campinas, SP

Nadir Rodrigues Pereira
Journalist, master's degree in Education, analyst at Embrapa Digital Agriculture, Campinas, SP

Paula Giovanna Guimarães Ribeiro
Lawyer, specialist in Civil Law and Civil Procedure, analyst at Embrapa, Brasília, DF

Paula Regina Kuser Falcão
Physicist, Ph.D. in Applied Physics, researcher at Embrapa Digital Agriculture, Campinas, SP

Paulo Arruda
Biologist, Ph.D. in Genetics and Molecular Biology, professor at the State University of Campinas and at the Genomics for Climate Change Research Center, Campinas, SP

Poliana Fernanda Giachetto
Zootechnician, Ph.D. in Animal Production, researcher at Embrapa Digital Agriculture, Campinas, SP

Rafael Soares Correa de Souza
Biologist, Ph.D. in Genetics and Molecular Biology, postdoctoral fellow at the State University of Campinas and at the Genomics for Climate Change Research Center, Campinas, SP

Ricardo Augusto Dante
Biologist, Ph.D. in Plant Science, researcher at Embrapa Digital Agriculture and at Genomics for Climate Change Research Center, Campinas, SP

Ricardo Fonseca Araújo
Bachelor in Computer Science, master's degree in Information Technology Knowledge Management, analyst at Embrapa, Secretariat of Innovation and Business, Brasília, DF

Ricardo Yassushi Inamasu
Mechanical engineer, Ph.D. in Mechanical Engineering, researcher at Embrapa Instrumentation, São Carlos, SP

Roberto Hiroshi Higa
Electrical engineer, Ph.D. in Electrical Engineering, researcher at Embrapa Digital Agriculture, Campinas, SP

Sandra Aparecida Santos
Zootechnician, Ph.D. in Animal Science, researcher at Embrapa Pantanal, Corumbá, MS

Santiago Vianna Cuadra
Meteorologist, Ph.D. in Agrometeorology, researcher at Embrapa Digital Agriculture, Campinas, SP

Sérgio Aparecido Braga da Cruz
Electrical engineer, Ph.D. in Applied Computing, researcher at Embrapa Digital Agriculture, Campinas, SP

Shalon Silva de Souza Figueiredo
Psychologist, specialist in Psychoanalytic Theory, analyst at Embrapa, Secretariat of Innovation and Business, Brasília, DF

Silvia Maria Fonseca Silveira Massruhá
Systems analyst, Ph.D. in Applied Computing, researcher at Embrapa Digital Agriculture, Campinas, SP

Sílvio Roberto Medeiros Evangelista
Statistician, Ph.D. in Electrical Engineering, analyst at Embrapa Digital Agriculture, Campinas, SP

Sônia Ternes
Applied and computational mathematician, Ph.D. in Electrical Engineering, researcher at Embrapa Digital Agriculture, Campinas, SP

Stanley Robson de Medeiros Oliveira
Computer scientist, Ph.D. in Computer Science, researcher at Embrapa Digital Agriculture, Campinas, SP

Suzana Maria Salis
Biologist, Ph.D. in Plant Biology, researcher at Embrapa Pantanal, Corumbá, MS

Tércia Zavaglia Torres
Administrator, Ph.D. in Education, analyst at Embrapa Digital Agriculture, Campinas, SP

Thiago Teixeira Santos
Computer scientist, Ph.D. in Computer Science, researcher at Embrapa Digital Agriculture, Campinas, SP

Urbano Gomes Pinto de Abreu
Veterinary doctor, Ph.D. in Animal Science, researcher at Embrapa Pantanal, Corumbá, MS

Vinícius Milléo Kuromoto
Public relations, specialist in Marketing, analyst at Embrapa Digital Agriculture, Campinas, SP

Viviane Cristina Heinzen da Silva
Biologist, Ph.D. in Functional and Molecular Biology, postdoctoral fellow at the State University of Campinas and at Genomics for Climate Change Research Center, Campinas, SP

Walfrido Moraes Tomas
Veterinary doctor, Ph.D. in Ecology and Conservation, researcher at Embrapa Pantanal, Corumbá, MS

Acknowledgements

We would like to take the opportunity to thank all those who, through collaborative work and exchange of ideas, made it possible to publish Embrapa's initiatives in the implementation of digital agriculture in Brazil. It focus on concepts, technologies and applications that are available or are under development, as well as its future challenges and perspectives.

As editors of this work, and on behalf of Embrapa Digital Agriculture, we thank the Ministry of Agriculture, Livestock and Food Supply and the Ministry of Science, Technology and Innovation for the opportunity to participate in the movement of digital transformation in agriculture within the scope of Agro 4.0 Chamber.

We are immensely grateful to the Executive Board of the Brazilian Agricultural Research Corporation (Embrapa) for their constant support and for believing in our work. We would like to thank Embrapa's Decentralized and Administrative Units for their collaboration.

We would like to especially thank the State University of Campinas (Unicamp), which hosts us and with which we have numerous joint initiatives in research, development and innovation (RD&I). We also mention the Center for Research in Genomics Applied to Climate Change, the participation of our researchers as collaborators in undergraduate and graduate courses related to the activities of our Unit, in addition to the work of numerous students from this university in our RD&I projects, carried out through the process of training and qualifying Embrapa trainees and scholarship holders.

In the context of open innovation, Embrapa has also joined forces, together with other public and private institutions, in the International Hub for Sustainable Development, coordinated by Unicamp with the support of the municipality of Campinas and the government of São Paulo state.

We extend our thanks to the producers and agents in the production chains, our clients, responsible for the greatness of Brazilian agribusiness, for the use and validation of technologies produced by Embrapa, as well as to the funding agencies for financing our research. We would also like to thank all our partners, represented by public institutions at the municipal, state and federal levels, companies, startups, accelerators, cooperatives, science and technology institutes and universities.

All these actors participate in the digital agriculture innovation ecosystem and continually challenge us to develop digital solutions that support the processing of data, information and knowledge for the implementation of digital agriculture in Brazil.

The participation of Embrapa Digital Agriculture employees was decisive. The technical staff was diligently devoted to preparing the chapters of this book, under their responsibility, making an effort to disseminate the research and work carried out in this research center. In addition, we thank the reviewers of the Local Publications Committee (LPC) for the technical analysis; the Organizational Communication Nucleus (NCO) for the illustration and consolidation activities of this work; the Editorial and Production team (Epro/GCI), from the General Secretariat of Embrapa, for their support in the bibliographic standardization; the librarian and administrative staff for all their support.

Finally, we would like to thank all the co-authors of Embrapa's Administrative and Decentralized Units and the partners who joined us, bringing their technical and scientific experience on behalf of the digital transformation and sustainability of Brazilian agriculture for food security in the country and in the world.

Foreword

The expectation that the world population will reach 9 billion inhabitants in 2050, which will increase food demand, imposes a great challenge for agriculture, such as the need to increase productivity in the same planted area, reducing costs, while respecting the conservation of natural resources. At the same time, extreme weather events affect agricultural production and demanding consumers want more nutritious, functional and sustainably produced food. The population is increasingly concentrated in urban centers, while the displacement of individuals from the countryside to the cities grows, as well as the aging of the rural population. In addition, the coronavirus threat has affected all sectors, such as health, education, agribusiness, economy, among others. To overcome all these challenges, it is necessary to generate and use new technologies, adding more value in all stages of the production chains.

In this context, the digital transformation emerges, which consists in using information and communication technologies (ICT), combined with innovative technologies, to significantly increase the performance and reach of companies by changing the way business operations are carried out. Some of these innovative technologies identified as critical for digital transformation are cloud computing, internet of things, social media, Big Data and data science, artificial intelligence, augmented reality and virtual reality, robotics, ubiquitous connectivity, machine learning, digital twins, automation, biotechnology and bioinformatics, in addition to nanotechnology. These technologies, which work in a synergistic and complementary way, have transformative power that culminates in what has been pointed out as the fourth industrial revolution, also called Industry 4.0.

This scenario also brings new opportunities to apply these innovations in agriculture. For Brazil to guarantee, or expand, its production capacity with sustainability, while meeting the global demand for food and nutrition security, it will require modernization, technification and innovation throughout the agricultural production chain, converging to digital agriculture or Agriculture 4.0, an analogy to Industry 4.0, as a result of the digital transformation of the sector.

Embrapa delivers research, development and innovation solutions for the sustainability of agriculture, and as a world reference, it generates and supplies information, knowledge and technologies, hence contributing to the innovation and sustainability of agriculture and food security. As an institution driven by science, responsive to trends and employing the latest technologies in agricultural research and innovation, Embrapa established digital agriculture as one of its main areas in Brazil. Embrapa Digital Agriculture is a research center that focuses on working in a multidisciplinary manner in the areas of agroinformatics and bioinformatics to develop research and solutions for agriculture, applying methods, techniques and computational tools. It has set out to promote digital agriculture in Brazil, together with other research centers of Embrapa and partner institutions in public and private sectors.

This book will provide the reader the opportunity to learn more about the concepts, technologies, applications, and challenges identified by Embrapa in this area. The initial expectation is to disseminate the efforts of the institution and its partners in the implementation of digital agriculture and serve as a starting point to establish a knowledge base for further discussions, to allow the strengthening of partnerships, the exchange of experiences and the promotion of sustainable development of agricultural production chains in Brazil. Thus, we hope that this book contributes to consolidating digital agriculture increasingly in the Brazilian agricultural space, encompassing both large projects and family farming, and promoting the growth of the agricultural sector as a whole and the improvement of social, environmental and economic conditions of the rural environment.

Given the importance that digital agriculture assumes in the current context, this work is being translated into English. The aim is to disseminate the research disclosed in this book in order to reach a larger audience and expand our presence in the innovation ecosystem through new partnerships. It is also expected that our capillarity with national and international research institutions will be amplified, opening new doors for audiences linked to tropical agriculture and increasingly strengthening the growth of digital agriculture in Brazil and in the world

Silvia Maria Fonseca Silveira Massruhá
President of Embrapa

Preface

The new challenges in agriculture call for greater productivity and efficiency by means of optimizing the use of natural and environmental resources, which demand applying different digital technologies. The integration of these technologies allows developing solutions for automation, smart farms, animal and plant health, agricultural risk management, biotechnology, nanotechnology, climate change mitigation, bioeconomy, bio-inputs, certification and traceability, precision agriculture, low touch economy, among others.

Digital agriculture has been implemented in Brazil as a response to the digital transformation, which took place in all sectors of society, and lead to a better use of information and communication technologies, combined with disruptive technologies that advocate the new industrial revolution, Industry 4.0. In addition to Brazil being one of the world's largest producers and exporters of agricultural commodities, as well as one of the main contributors responsible for global food security, intensifying the technification of rural properties is critical to ensure the competitiveness of Brazilian agriculture.

In 2020, it was Embrapa Digital Agriculture's 35th anniversary, and the launch of this book aims to publicize its efforts to promote digital agriculture in Brazil. It was created in 1985, by Embrapa's Executive Board, as a research center focused on excellence in research and generating knowledge in information and communication technologies (ICT) for Brazilian agriculture. Since then, Embrapa Digital Agriculture has proven the transversality of ICTs through the execution of its research projects, applying methods, modeling and simulation techniques and tools, artificial intelligence, pattern recognition and geoprocessing, supported by information management and knowledge and the use of emerging technologies and open standards. The performance of the research and development area is guided by a strategic vision, mainly focused on solutions in the areas of agroinformatics and bioinformatics. Currently, this center is dedicated to integrating ICT with new digital technologies to promote digital transformation in agriculture through its four axes of action, namely: Bioinformatics and Computational Biology, Scientific Computing and Automation, Modeling Agroenvironmental and Geotechnologies, in addition to Information Engineering.

This book presents initiatives in digital agriculture carried out by activities and projects developed by Embrapa Digital Agriculture, together with other Embrapa Units, as well as partner institutions in the public and private sectors.

The chapters of the book address concepts, technologies, research applications and challenges, development and innovation in digital agriculture in the production chains. Chapters 1 and 2 present the context of digital agriculture in Brazil, as well as the definitions of digital technologies employed in the solutions discussed in later chapters.

Chapters 3 to 11 focuse on applications being developed within the scope of digital agriculture in the areas of agroenvironmental modeling, geotechnologies, scientific computing, computer vision, precision agriculture, information engineering, bioinformatics, structural biology and genomics applied to climate change.

Chapters 12 to 16 describe describes the innovation ecosystem in agriculture and future perspectives, covering the processes of innovation, digital law, communication, driving forces for agriculture and the challenges, trends and opportunities that unfold in this trajectory.

Considering the successful trajectory of digital agriculture in Brazil in response to the digital transformation in the field, this work has been translated into English. Our aim is to share our experience, report on key research challenges, show how digital technologies are being deployed at multiple scales and play a key role in science-powered agriculture. To effectively respond to these challenges, our participation in solid arrangements with the cooperation of the government, academia, the productive sector and civil society is essential. Besides that, we want to increase our international influence. These strengthened and expanded relationships will allow the insertion and expansion of digital technologies in agriculture, in a transversal way, as enablers of high-impact results, and will imprint the concept of digital agriculture to the most different links in the agricultural sector.

Stanley Robson de Medeiros Oliveira
Head of Embrapa Digital Agriculture

Contents

Photo: Gabriel Rezende Faria

1 Digital transformation in the field towards sustainable and smart agriculture

Silvia Maria Fonseca Silveira Massruhá | Maria Angelica de Andrade Leite | Ariovaldo Luchiari Junior | Sílvio Roberto Medeiros Evangelista

Introduction

Global agriculture has been challenged to ensure food security by providing, in a sustainable approach, food, fiber and clean energy. The predicted global scenario is critical: the world population reaching 9 billion in 2050; growing scarcity of land and water resources; climate change and extreme weather events; growing per capita income and urbanization levels; new digitized consumers demanding more nutritious and functional foods; and decreasing productivity gains in some countries. Projections based on patterns of population growth and food consumption indicate that agricultural production will have to increase by at least 70% to meet demands by 2050. Most estimates also indicate that climate change is likely to reduce agricultural productivity, production stability and income in some areas that already experience high levels of food insecurity. Therefore, the development of smart agriculture is crucial to achieving future goals of food security. (FAO, 2010).

In line with the sustainability of the planet, in 2015, the United Nations (UN) launched the 17 Sustainable Development Goals (SDGs) to promote a fairer society that respects the environment. The 17 SDGs constitute a universal appeal to protect the planet and ensure that all people have dignity, in order to lead governments, companies and societies towards a more sustainable and inclusive world.

They serve as a guide for countries to overcome the most pressing environmental, political, and economic challenges. Some of the 17 SDGs can be achieved by actions directly related to agriculture, as illustrated in Figure 1. Thus, SDG 2, Zero Hunger, can be minimized by increasing agricultural production. SDG 6, Clean Water and Sanitation, refers to the sustainable use of water in irrigation activities and agriculture in general. SDG 8, Decent Work and Economic Growth, can be met by promoting actions to improve the conditions of small-scale rural producers and family farmers and by expanding access to information. SDG 9, Industry, Innovation and Infrastructure, can be supported by improving the production chains. SDG 11, Sustainable Cities and Communities, is supported by greater integration between the field and the city. SDG 12, Responsible Consumption and Production, can be achieved by controlling crop losses and food waste. SDG 13, Climate Action, can be achieved by mitigating the risks of climate change and reducing the emission of greenhouse gases in livestock activities. SDG 14, Life Below Water, is supported through the improvement of aquaculture production. SDG 15, Life on Land, can be monitored by mapping land cover use and sustainable agricultural production. Finally, SDG 17, Partnerships for the Goals, is supported by increased information sharing among agricultural partners (Project Breakthrough, 2017).

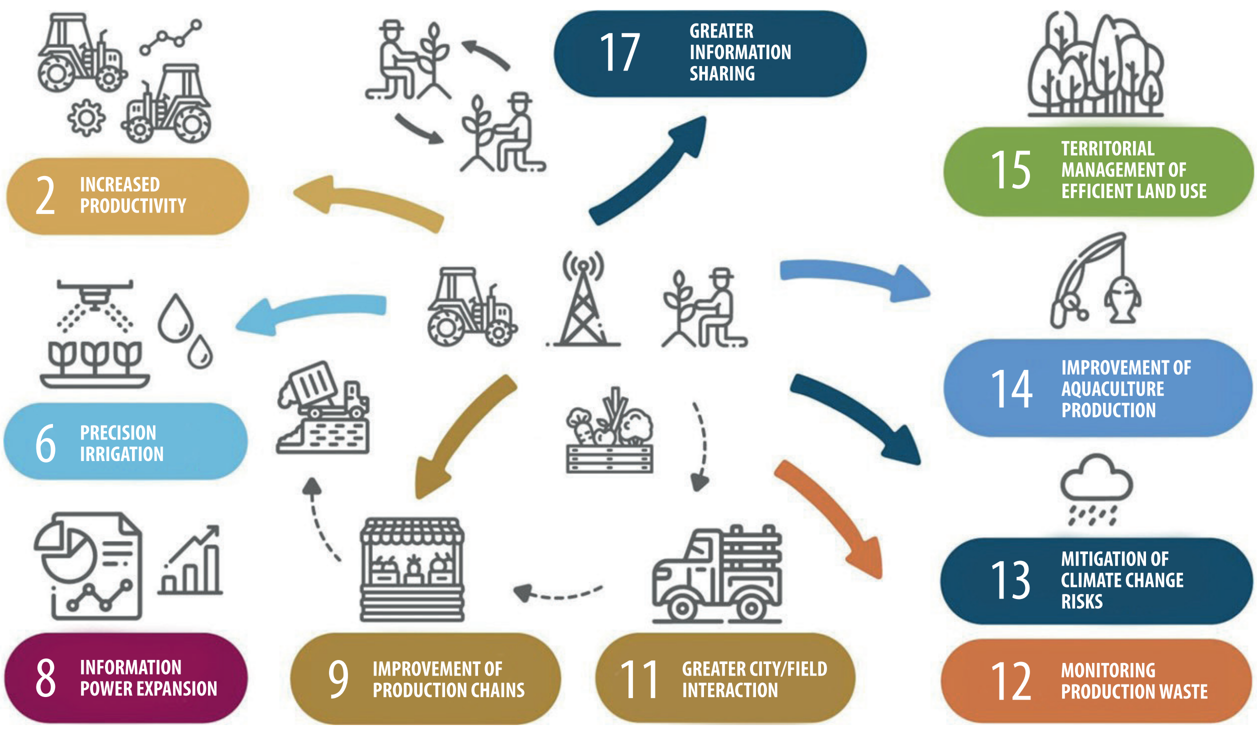


Figure 1. Sustainable development goals related to agriculture.

Source: Adapted from Project Breakthrough (2017).

Among all these challenges, the newest adversity has emerged: the pandemic caused by the coronavirus, which has impacted the health of millions of people, undermining all forms of social coexistence, interrupting education in schools and triggering serious damage to all sectors of the economy, including compromising the agribusiness production and distribution chain, and affecting the price of agricultural commodities. Prins (2020) reports that covid-19 is driving the transformation of agricultural data in three aspects: 1) increased digitization; 2) increased digital collaboration; and 3) visibility, mainly due to disruptions in the value chain, which make planning a fundamental tool in the process of supplying agricultural products. The damage caused by the pandemic is still being calculated, and new policies and strategies will be necessary, not only for risk mitigation, national integration and international cooperation, but also fiscal and economic incentives for the world to return to its development path.

Brazil is the world's largest exporter of soy, coffee, sugar, orange juice, sugarcane ethanol, beef and chicken. In 2019, agribusiness exports were US$ 96.8 billion, representing 43.2% of Brazil's total exports. Brazilian agriculture is based on more than 300 species of crops, and ships 350 types of products around the world, reaching 200 markets on the planet. Brazil is a large producer of grains, meat and fruits, and the agricultural sector accounted for 21.1% of Gross Domestic Product (GDP) and 20% of the workforce (Embrapa, 2019). The Brazilian 2019/2020 grain harvest is considerated as a new historical record, estimated at 250.5 million tons, 3.5% (or 8.5 million tons) higher than the 2018/2019 harvested in (Acompanhamento da Safra Brasileira [de] Grãos, 2020).

In Brazil, family farming is responsible for an important part of the national food production. Approximately 50% of family farming establishments are concentrated in the Northeastern region, 19% in the Southern region, 16% in the Southeastern region, 10% in the Northern region and 5% in the Midwestern region. Bahia is the state with the highest number of family establishments (15%), followed by Minas Gerais (10%). These two states also have the largest areas with family farming establishments, around 10 million and 9 million hectares, respectively (Embrapa, 2019).

In view of all these challenges in agriculture, mainly that of expanding agricultural production without significantly expanding the planted area, the increasingly intense use of new technologies is crucial to enable productivity gains in a sustainable manner. It is in this context that a new production factor has emerged, and which is changing the basis of economic growth for countries across the world: digital transformation. This is a new approach in which Information and Communication Technologies (ICT) play a key role in transforming the strategy, structure, culture and processes of organizations, using the internet power of connectivity.

Through new investments in technologies and business models, it is expected that the engagement of digital customers across all touchpoints in the life cycle of their experience will improve. Digital transformation consists of using ICT to significantly increase the performance and reach of companies by changing the way business is done. There are three elements of digital transformation: transformation of the customer experience, of business models and of operational processes (Transformação Digital, 2020).

Some of the technologies identified as critical in the digital transformation are: cloud computing, internet of things, social media, mobility, big data and data science, artificial intelligence, augmented reality and virtual reality, robotics, ubiquitous connectivity, machine learning, digital twins and automation, in addition to advances in biotechnology and bioinformatics and nanotechnology. These areas, in a synergistic and complementary way, have the power to transform the new world order, culminating in what has been called the fourth industrial revolution, or Industry 4.0. Furthermore, the lower cost of these advanced technologies plays an important role in accelerating innovation (World Economic Forum, 2017).

Disruptive technologies, combined with the latest innovations, hold the promise to leverage agricultural research. The convergence of the areas of Nanoscience, Biotechnology, ICT and Cognitive Science (NBIC) will provide a great qualitative leap in how the world of agriculture can be transformed. The evolution of systems approaches, mathematics and computation, with work in NBIC areas, allows, for the first time, to understand the natural world and cognition in terms of complex and hierarchical systems. Applied both to specific research problems and to the general organization of science and technology institutions, this complex systems approach provides holistic awareness and integration opportunities in order to achieve maximum synergy along the main directions of scientific and technological progress for agriculture (Kim et al., 2012).

The World Economic Forum (WEF) launched the Digital Transformation Initiative, in 2015, in collaboration with the company Accenture (2020), to serve as a focal point for new opportunities and emerging themes related to the latest developments in business digitization and in society. This initiative supports the Forum's actions on themes related to the Fourth Industrial Revolution. Since its inception, the initiative has analyzed the influence of digital transformation on various areas, such as: agriculture, aviation, travel and tourism, chemicals and advanced materials, mining and metals, oil and gas, professional services, retail, telecommunications, automotive industry, consumer sector, electricity, healthcare, logistics, and media (World Economic Forum, 2017).

Digital transformation gave rise to a proliferation of startups, whose most used definition is a group of people looking for a repeatable and scalable business model, working in conditions of extreme uncertainty, according to Yuri Gitahy, angel investor and company founder Aceleradora (2020) and board member of the Brazilian Association of Startups (ABStartups) (Associação Brasileira de Startups, 2019). Startups are highly flexible compared to traditional companies and have a clear goal and speed to adapt, change, create, re-strategize, see and generate new markets and new monetization possibilities. In 2020, ABStartups had more than 13 thousand startups among its affiliates (StartupBase, 2020).

This scenario gives rise to new opportunities to apply these innovations in agriculture. For Brazil to guarantee and expand its production capacity with sustainability, while meeting the global demand for food and nutritional security as a major exporter of agricultural commodities, it requires modernization, technification and innovation across the entire agricultural production chain, converging to digital agriculture as a result of the sector's digital transformation.

According to Embrapa's 2030 Vision document (Embrapa, 2018), ICT and its accelerated advances, such as social media and digital platforms, have transformed relationships, interactions and communication between companies and consumers. Increasingly accessible computers and cell phones, low-cost internet and wi-fi technology enable access to information and provide the consumer a larger role in decision-making when buying, as well as in sharing experiences and in controling products and brands. The advancement of the digital and collaborative economy increases the level of information, the consumers' skills and engagement, as well as the necessary conditions for them to have a leading role in making decisions about productive processes, promoting their empowerment (Gazzola et al., 2017).

It is not surprising that in this era of extensive and profound transformations brought about by ICTs, one of the main forces shaping the future vision of Brazilian agriculture is the influence exerted by new consumers, who are increasingly connected through social networks to promote their consumption choices, while influencing peers and agricultural production systems. Equipped with more information and greater knowledge about products and their prices, these economic agents increasingly become a determinant of the attributes they want, enhanced by digital opportunities and tools. Individuals ever more interconnected with their devices, always connected to the network (web) and with access to any type of information in real time, are strong opinion makers in their circles of influence. This demonstrates that the trust they place in organizations (suppliers) strongly affects their survival. (Embrapa, 2018).

The most recent Brazilian survey by ICT Households of the Brazilian Internet Steering Committee (CGI.br), carried out in 2019, shows that 50% of the Brazilian population has used a computer, 74% uses the internet and 99% has a cell phone (Comitê Gestor da Internet no Brasil, 2019). According to the Census of Agriculture the Brazilian Institute of Geography and Statistics (IBGE), carried out in 2017, the number of farmers who declared having access to the internet grew 1,900%, from 75 thousand in 2006 to 1,430,156 in 2017, with 659 thousand through broadband and 909 thousand through mobile internet (IBGE, 2019). In accordance with the study *The mind of the Brazilian farmer in the Digital Age*[1] by McKinsey & Company, young large-scale farmers, such as grain (32%) and cotton (62%), are the pioneers in adopting precision agriculture in Brazil and in learning about technologies. Among them, 47% use at least one precision farming technology, while 33% use two or more. Furthermore, according to the study, variable-rate

[1] Presentation of the study by Nelson Ferreira at the webinar Digital agriculture in the post-covid era, on May 7, 2020. Available at: https://www.insper. edu.br/agenda-de-eventos/a-agricultura-na-era-digital-pos-covid

application and drones are the most popular technologies, and many others use the internet of things, telemetry and remote sensing.

In another survey carried out in partnership with Embrapa, the Brazilian Micro and Small Business Support Service (SEBRAE) and the National Institute for Space Research (INPE) (Santin, 2020), 84.1% of the interviewed farmers use at least one digital technology in their production process. The main functions of digital technologies used by farmers are: obtaining information and planning activities on the property (66.1%); rural property management (43.3%); purchase and sale of inputs, products and production (40.5%); mapping and planning of land use (32.7%); and forecast of climatic risks such as frost, hail, *veranico*[2] and heavy summer rains (30.2%).

This chapter will explore the main aspects of digital transformation in agriculture. Section 2 will address the evolution of agriculture and automation in agriculture. Section 3 will discuss the path to fully automated 5.0 agriculture and the key factors involved. Section 4 will present initiatives to promote digital agriculture in Brazil. Section 5 will indicate how to incorporate the digital transformation megatrends in agriculture. Finally, section 6 will present the main contributions of this chapter.

Evolution of agriculture: from Agro 1.0 to Agro 4.0

Agriculture should be considered one of the greatest achievements of humanity. In the early 20th century, there was agriculture 1.0, in which the workforce was provided by the labor of families, using manual instruments, assisted by animal traction. It was low production agriculture. These farmers, in addition to cultivating for their own consumption, generated a food surplus that supported an ever-increasing number of people.

With the Industrial Revolution and the growth of the urban population, demand for food increased, requiring that the various agricultural production processes also evolve. At that time, scientific method and advanced technologies were applied to agriculture, and machines were being created and implemented to assist in the different stages of fertilization, planting and harvesting.

Brazilian agriculture was rudimentary in the middle of the last century, around 1950 and 1960. Manual labor prevailed in agricultural production. At that time, less than 2% of rural properties had agricultural machinery. Farmers suffered from a shortage of technology and information. As a result, there were low yields per hectare and little production. The expansion of agriculture required converting extensive natural areas into crops and pastures. Inappropriate practices caused severe environmental impacts, such as erosion and siltation. However, farms did not produce enough to meet the domestic demand. Inefficiency in the field created problems across the country. Brazil experienced a moment of strong industrialization, with growing cities and population, besides greater purchasing power. The context was that of food scarcity (Embrapa, 2020a).

The Green Revolution took place and introduced a series of technological innovations in the agricultural sector. These innovations aimed to increase productivity through genetically modified seeds, new soil fertilization techniques, the use of industrialized products (such as pesticides), and the intense use of machinery, which significantly shortened the time spent in harvesting process. Careful livestock rearing, crop rotation and better equipment, with the introduction of the combustion engine, helped to increase

[2] Dry period, accompanied by intense heat, strong insolation and low relative humidity during the rainy season or winter. Similar to Indian summer.

production. Field mechanization became a trend in the early 20th century. Nevertheless, it was only after World War II that manual traction was completely replaced by mechanical force in North American and European crops (Jacto, 2018). The use of all these new technologies culminated in the implementation of Agriculture 2.0.

At that stage, Brazil was a major food importer. Embrapa was created in 1973, and one of its main attributions was to ensure food security, investing in research to consolidate various production chains and to developed technologies to transform the Cerrado biome into productive land. Agriculture was predominantly based on monoculture and research had a monodisciplinary and adaptive vision. The answers emerged after years of research carried out by Embrapa, universities, state agricultural research institutions and, later, by the private sector. With genetic improvement techniques, suitable plants for the Brazilian soil and climate conditions were developed. They were cultivars less sensitive to long days and more tolerant to pests in the tropical world. Soil correction and fertilization were essential contributions. Research was carried out to optimize the use of correctives and fertilizers that allowed planting in Cerrado soils, which until then were considered unproductive. (Embrapa, 2020a).

Later, agricultural intensification was strengthened, and monoculture gave way to integrated and rotated production systems. These systems demanded more knowledge and involved multiple disciplines. Research became systemic, as it was important to understand the entire chain of production systems. (Pillon, 2017).

No-till, climate risk zoning, pest and weed management, mechanization, succession of up to three annual crops in the same area and integrating farming with livestock and forest are additional and valuable approaches and technologies. These results are directly related to investment in research, rural extension, public policies and entrepreneurship. (Embrapa, 2020a).

Since then, technologies have evolved in unimaginable ways, with machines and implements to increase the efficiency of field activities, a trend that became known as precision agriculture, initiating Agriculture 3.0. According to the International Society of Precision Agriculture, precision farming is:

> [...] a management strategy that gathers, processes and analyzes temporal, spatial and individual data and combines it with other information to support management decisions according to estimated variability for improved resource use efficiency, productivity, quality, profitability and sustainability of agricultural production (Springer, 2020).

Precision agriculture, through machines with built-in sensors, use of satellite images, remotely piloted aircraft, such as drones, and sensors implanted in animals and crops, made it possible to collect numerous types of data, such as information related to soil, climate, plants and animals, application of inputs, yield maps, among others. The large volume of data collected through precision agriculture constitutes a source of information obtained directly from the field (Bernardi et al., 2014).

Currently, driving forces point to technological aspects that consolidate clean production systems, with a positive carbon balance, based on sustainability; bio-based agriculture; advances in synthetic biology; demand for greater efficiency of water use in agriculture; operating in a new energy development cycle; technological disruptions; and increased demand for food, fiber and bioenergy with more efficient use of natural resources and environmental services. Systems become complex and involve many variables. It is the era of bioeconomy, which concerns the economic activity driven by research and innovation in biological sciences (National..., 2012). This includes everything from the production of renewable biological resources to the conversion of these resources and residues into food and non-food products, using the integration of knowledge and technologies generated in different areas (European Commission, 2012). It involves three major elements: advanced use of genes and complex cellular processes to develop new processes and products; use of renewable biomass and efficient bioprocessing to support production; integrating knowledge and applying biotechnology among economy sectors (Organization for Economic Co-Operation and Development, 2009).

Alongside these new agricultural demands, there is digital transformation, as discussed in the Introduction, which brought new disruptive technologies and which started to be used, causing the emergence of digital agriculture and leading to yet another phase of the technological revolution, in other words, Agriculture 4.0. Agriculture 4.0 is an analogy to Industry 4.0, resulting from the digital transformation of the agricultural sector through massive data collection to assist decision-making. Figure 2 illustrates the evolution of agriculture and its respective phases.

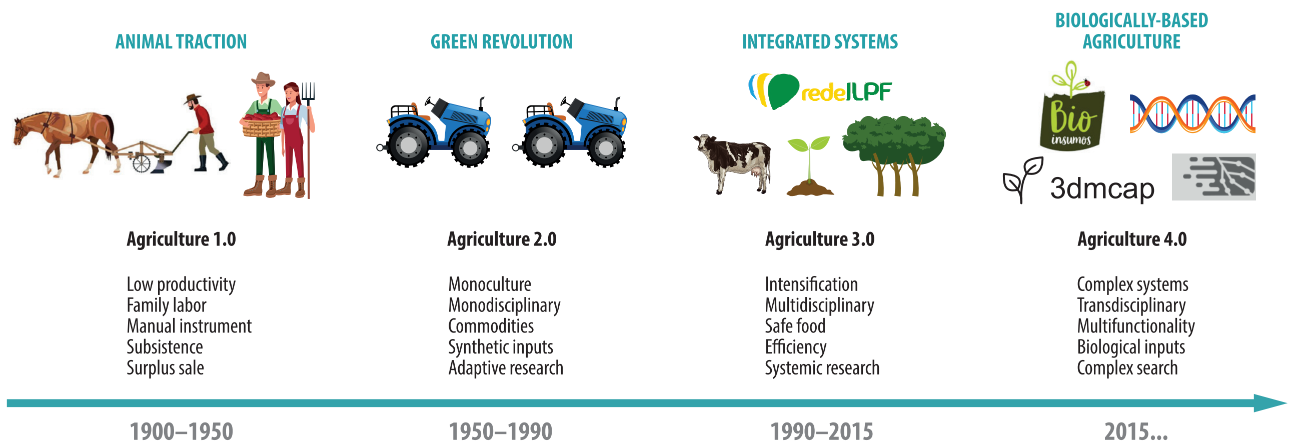


Figure 2. Phases of the evolution of agriculture.

Source: Adapted from Pillon (2017).

The digital transformation in the age of bioeconomy will combine technological advances in disruptive technologies with advancements in biotechnological areas. It will produce solutions for an agriculture that involves the study of complex systems, in which it is increasingly necessary to carry out analyses, monitor and predict, taking into account the social, biological, environmental and economic aspects related to the use of these new technologies.

Digital agriculture: from Agro 4.0 to Agro 5.0

Digital agriculture consists of inserting digital technologies in all stages of the value chain in order to promote competitive advantages and socio-environmental benefits. It is based on digital content, by processing the large volume of data produced in all stages of the production chain, from pre-production to the post-production phases, and covering production, as illustrated in Figure 3. Pre-production includes data for genetic improvement in plants and animals. The production phase involves data collected in precision agriculture by drones, satellites, sensors placed on plants, animals, soil, atmosphere, machines, equipment, and vehicles remotely connected to each other and to a data collection center. Finally, in the post-production phase, data comes from market analysis and from the stages of storage, distribution, logistics, traceability, consumption, among others.

In the pre-production phase, the use of data mining, high-performance computing, and modeling and simulation technologies, along with biotechnology and bioinformatics, will enable the discovery of genes that control complex features and their functions. Together with gene interaction studies, these

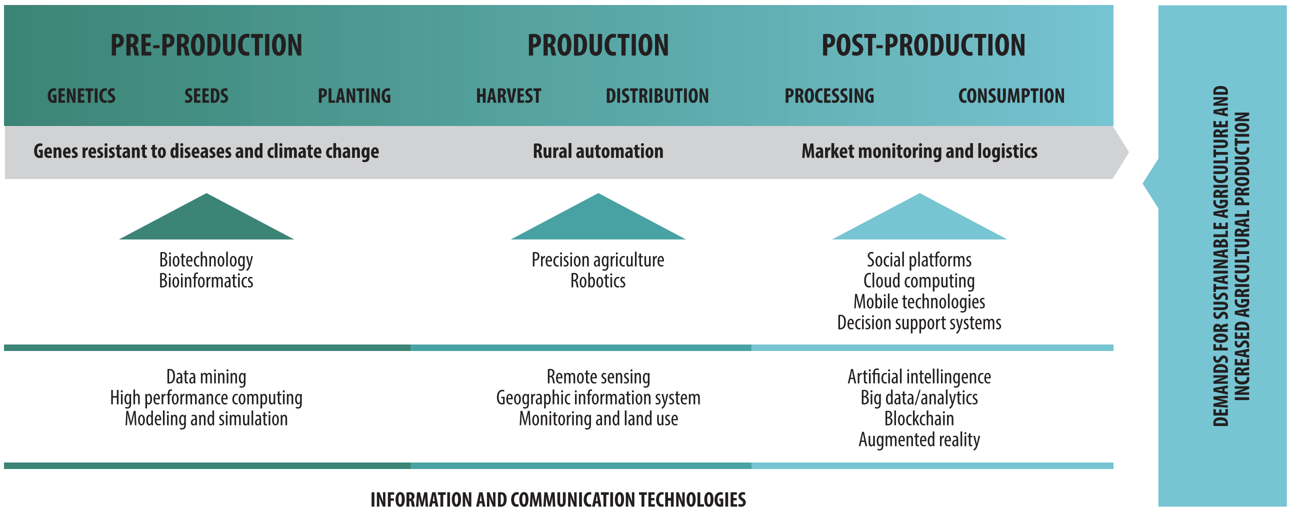


Figure 3. Digital agriculture in the production chain in the pre-production, production and post-production stages.

technologies will promote advances to impact several areas of animal and plant production, such as management, nutrition, resistance to diseases and water stress, health and genetic improvement, resulting in more sustainable products, with better nutritional quality and safety. Integrating the heterogeneous data and the large volume of information generated by the "omics" sciences is a major challenge in the area of integrative genomics. Dealing with the data stored in different places and formats, and combining this with the use of machine learning strategies, mathematics, computational algorithms and supercomputers, will make it possible to explore, in an innovative way, the data generated by different omics sciences (Boyle, 2013; May, 2014). This innovation occurs due to predicting biological functions and understanding biological mechanisms, such as those responsible for diseases, and defining characteristics of agronomic and productive interest.

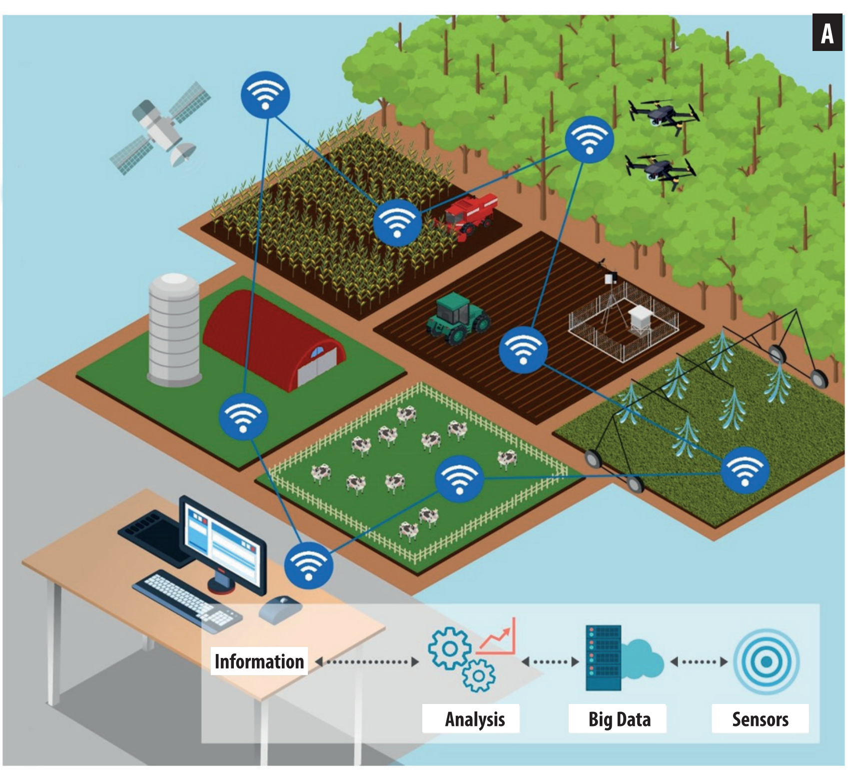
In this regard, bioinformatics emerged from the need to organize, manage, visualize and exchange biological sequence data. With this information, bioinformatics evolved towards the creation of tools for analysis, interpretation and modeling sequences, structures, genomes, metabolic networks, creating an increasingly complex network of information. Through bioinformatics, it is now possible to perform analyses at different levels of complexity, based on data sets that allow revealing aspects of the complex organization of biological systems through studies in genomics, transcriptomics, proteomics, metabolomics, in addition to the scale of phenotypic analysis of the most varied organisms (Varshney et al., 2014).

Biotechnology, on the other hand, brings innovations such as synthetic biology, which enables the design of an organism, allowing the creation of genetic machines with new properties and operations, for example the generation of plants as biomass raw materials for biofuels and bio factories to produce inputs for the industrial and pharmaceutical sector. Another technology is genome editing, which enables to carry out precise and specific genetic modifications in the DNA strands or generate genomic rearrangements to improve characteristics such as disease resistance and drought tolerance (Vasconcelos; Figueiredo, 2015).

Other areas related to the pre-production phase that will benefit from digital agriculture technologies are development and production of methods, equipment and inputs for laboratory analysis; chemical and biological inputs for managing the health and nutrition of plants and animals; seeds and seedlings; as well as financial services.

In the production phase, precision agriculture and robotics, supported by technologies such as remote sensing, geographic information system and monitoring of land use, enables the use of wireless sensors, located in the soil, in the plant, in the atmosphere or in machines and equipment, which together with data analysis software, enables more accurate field mapping. This mapping allows intelligent planting of seeds and optimized application of chemical or biological inputs for nutritional and sanitary crop management. Sensors that measure soil moisture indicate when irrigation is needed. Images of plants captured by cameras, drones and satellites can help detect pests, leading to the application of specific and adequate amount of pesticides. Devices can capture harvest information and map the productivity of each part of the land. Sensors embedded in agricultural machines can indicate the need for maintenance. Equipment installed in silos can indicate the storage conditions, avoiding storage losses. Sensors inserted in animals can help monitor their health, well-being and stress and predict calving dates, aimed at managing and improving performance.

In the context of digital agriculture, the production phase shows the emergence of digital farms or smart farms (Pivoto et al., 2018). On these farms, the agricultural establishment will be massively connected, monitored and automated in a fully integrated infrastructure, as illustrated in Figure 4. Through precision agriculture, sensors distributed throughout the property and interconnected to the internet (internet of things) will generate large volumes of data (Big Data) that will have to be filtered, stored (cloud computing) and analyzed. The human workforce will not be able to manage this amount of data and will count on algorithms further improved by computational intelligence techniques (analytics). After the analysis, the cycle will be closed by remote commands to tractors and agricultural implements, which, equipped with a global positioning system (GPS), will make specific interventions only when necessary to optimize cost, production and environmental impact. Society, through social networks, will be able to obtain detailed information about the production process, impacts and nutritional properties on their mobile devices. On smart farms, the current concept of precision farming is enhanced by context, situation and location recognition, data-rich ICT services, data integration, data communication,



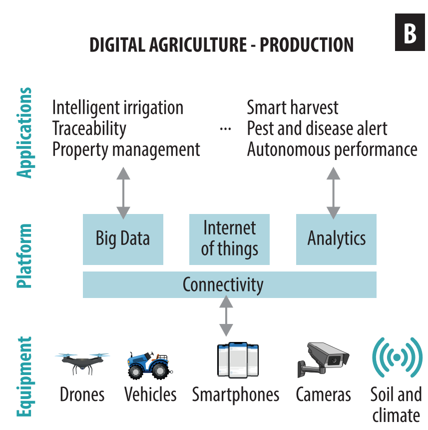


Figure 4. Digital agriculture in the production phase: the connected smart farm (A); equipment, technological platform and applications in the production phase (B).

standardization, signal processing and automation technologies, in addition to high-level automation planning and control (Sorensen, 2020).

In the post-production phase, new technologies will provide highly integrated communication and automation in the most varied activities of the agri-food and agro-industrial sectors. Prediction systems will forecast agricultural harvests and the risks involved. Advanced monitoring and control systems will inform consumers about food safety and sustainability. Traceability systems will provide production flow monitoring from the farm to the distribution centers, avoiding losses. Market information and economic variations will be processed and will guide the marketing processes. Storage, infrastructure and logistics will become more efficient, in addition to marketplaces, which will enable virtual connection between various actors in the production chains, offering negotiation and sales solutions. Packaging, environment and recycling, online restaurants and consulting are other areas that will be impacted.

In the context of digital agriculture, the data collection and management stage, through precision agriculture technologies, the internet of things and telematics, with the ensuing cloud storage, is expressed as Agriculture 4.0. Once the data is stored in the cloud, large analysis capacity is required, using artificial intelligence tools to process the large volume and extract relevant knowledge that not only helps decision making in property and production management but also conducts the performance of autonomous machines in the field (Saiz-Rubio; Rovira-Más, 2020).

The ability to use digital technologies to convert accurate data into knowledge to support and drive farmers' complex decision-making processes along the value chain will enable moving from precision farming to decision farming (Shepherd et al., 2018). The use of artificial intelligence and autonomous agricultural robots for agricultural work leads to a new phase, which is Agriculture 5.0 (European Agricultural Machinery Association, 2017). As the robots operate from the ground, the distance between the sensors and the target decreases to less than 2 m, increasing the accuracy of the captured data and allowing, for example, recording light intensity, moisture content of the soil, of the plant and the atmosphere and of disease severity, which will lead to a more specific action for the needs of each plant or animal (Saiz-Rubio; Rovira-Más, 2020).

In addition to including technological innovations, Agriculture 5.0 also needs to encompass characteristics such as: a) enable the production of more food on less land area and with fewer inputs; b) promote public policies and strategies to address the social and political aspects of agricultural systems; and c) contribute to reducing food losses along production and supply chains, including post-harvest losses and global food waste per capita in retail and consumption, besides it helping to understand consumer needs and their diets, in order to mitigate the impact related to the use of natural resources and the environment (Fraser; Campbell, 2019).

Brazil is already in tune with digital transformation in agriculture, especially through the incorporation of automation processes. Precision practices and processes, extensive use of sensors and sophisticated forecasting mechanisms and responding to climate change, for example, are among the improvements incorporated, opening spaces for Brazil in the global market, in strategic agricultural and bioeconomy sectors. The use of GPS-guided intelligent machines for planting, crop treatments and precision harvesting is growing in the country's most advanced production areas, with input savings, productivity gains and sustainability.

Initiatives for digital agriculture

Throughout its history, Embrapa Digital Agriculture has followed the evolution of ICT for the development of its applications, according to Figure 5. In the first phase, the systems were built as single-users and the software installed would run independently on desktop computers. It was when commercial internet was just beginning, and the main research centers and universities started connecting to the internet. In this phase, research had to adapt the existing models and solutions to the needs of Brazilian agriculture.

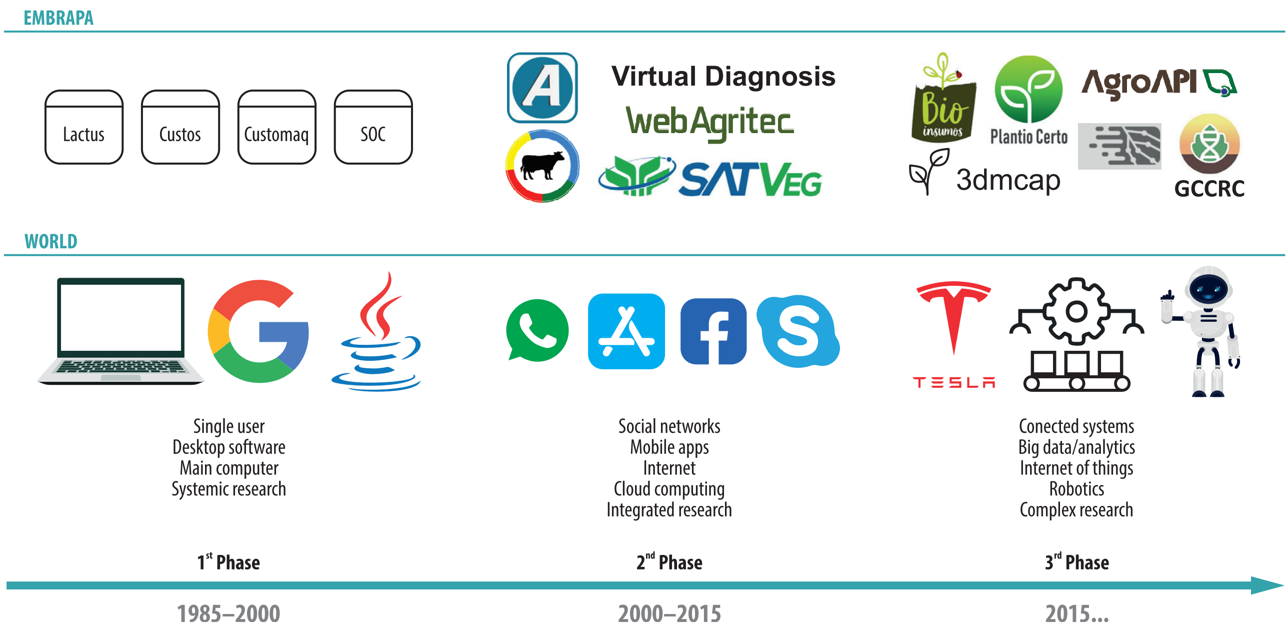


Figure 5. The evolution of Information and Communication Technologies and Embrapa's performance.

The second phase witnessed the appearance of mobile internet, which enabled using agricultural applications on cell phones, with data stored in clouds and social networks gaining global dimensions. Consequently, research gained an integrated dimension, since multidisciplinarity leads to aggregated solutions.

The third phase includes digital transformation, with highly automated agricultural activity through the constant evolution of precision agriculture and livestock systems, which are connected with all links in the production chain. The RD&I area will generate significant demands for new technologies in Brazilian agriculture. Some of the latest ICT innovations promise to leverage agricultural research. ICTs constitute the third pillar of scientific investigation, together with theory and experimentation, which allow simulating models of complex phenomena that could not be replicated in a laboratory. Mobile devices, cloud computing, Big Data, predictive analytics, wearable computing, cognitive computing, intelligent software systems, internet of things, advanced robotics, nanotechnology, biotechnology, integration of the omics sciences and the next generation genomics constitute the disruptive technologies that are transforming the way people live and work, through a new infrastructure in which the physical and digital worlds are fully interconnected.

In this scenario, in order to promote the sustainable and competitive development of the Brazilian economy, the National Plan for the Internet of Things was instituted by Decree no. 9,854, of June 25, 2019. This is an initiative of the Ministry of Science, Technology and Innovation (MCTI), the Ministry of Economy and the Brazilian Development Bank (BNDES), together with civil society – companies, academia, funding agencies and other bodies – to ensure that Brazil benefits from IoT technology. The plan defined four priority areas: industry, health, smart cities, and agriculture.

The potential impact and relevance of IoT for the country can be evidenced in its proposals, such as supporting pilot projects in these prioritized environments. In the rural area, it emphases initiatives such as "Tropical Farm 4.0", which increase the productivity and quality of Brazilian rural production by using data that, for example, help to accurately monitor biological assets (Produto 7C, 2017).

Within the scope of the National IoT Plan, the Agro 4.0 Chamber (Câmara Agro 4.0) was created as a technical cooperation agreement between the MCTI and the Ministry of Agriculture, Livestock and Supply (MAPA). The idea is to have a discussion group with the participation of government, companies and academia to build a strategy for connected farms that use solutions such as automation, interactivity, real-time monitoring, Big Data, among others. One of its actions is to promote connectivity in the countryside by expanding broadband internet in the rural environment. The Agro 4.0 Chamber is coordinated by the MCTI and the MAPA and the participation of actors from the private sector, academia and research institutes to debate and present solutions in the following areas: i) Development, Technology and Innovation; ii) Professional Development; iii) Productive Chains and Supplier Development; and iv) Field Connectivity.

To enable implementing the actions from the National IoT Plan, several development activities were created. In 2018, the BNDES Public Notice for IoT Pilots - Internet of Things was launched to finance project proposals for the implementation of pilots focused on the development of integrated IoT solutions through tests in real and controlled environments, whose impacts can be evaluated to allow its massification, commercial viability and interoperability.

Another initiative to promote digital transformation in the country is the applied research centers in artificial intelligence (AI), which will be created through cooperation between the MCTI, the São Paulo Research Foundation (FAPESP) and the Brazilian Internet Steering Committee (CGI.br). These centers will commit to the development of scientific, technological and innovation research, applied and oriented to problem solving through AI. The first four centers, two in São Paulo and two in other states, will focus on health, agriculture, industry and smart cities. The centers will be supported for 5 years and renewed for more 5 years depending on the results achieved (Arantes, 2019).

Along the same line, the Association for the Promotion of Brazilian Software Excellence (Associação para Promoção da Excelência do Software Brasileiro – SOFTEX), responsible for the Softex Priority Program of the Secretariat of Entrepreneurship and Innovation of the MCTI, launched the Softex Notice no. 01/2020 – Notice for Qualification of Institutions to Support the Research Process, Development, Innovation and Acceleration of IA²MCTIC Projects. The purpose of the Notice was to select and qualify pairs of Science and Technology Institutions (ICTs) and accelerators for joint action in the IA²MCTIC Program. The consortium formed by Baita Aceleradora, Eldorado Institute and Embrapa Digital Agriculture, was one of those certified in this initiative.

Embrapa, whose mission is to create research, development and innovation solutions to ensure the sustainability of agriculture, for the benefit of Brazilian society, is a protagonist in the technological modernization of agriculture. In the late 1990s, it created Embrapa's Precision Agriculture Network (PA Network) to provide guidance on the best and most appropriate use of PA and for research and development of new technologies. It currently involves 20 company research centers and more than 50 partners, such as companies, research institutions, universities and rural producers. The PA Network has the National Reference Laboratory for Precision Agriculture (LANAPRE), installed at Embrapa Instrumentation, in São Carlos, São Paulo state. The space, in a single location, is used to research and develop equipment, sensors, mechanical components and on-board electronics (Embrapa, 2020b).

Aware of the need to follow the global and national trends of the new economy and the world order and how these transformations impact agriculture, Embrapa, through its Strategic Intelligence System (Agropensa), prepared the document *Vision 2030: the future of brazilian agriculture* (Embrapa, 2018). In this process, the company and its network of partners prospected and analyzed the challenges and signs of new directions. These assessments gave rise to a group of seven megatrends: Socioeconomic and Spatial Changes in Agriculture; Intensification and Sustainability of Agricultural Production Systems; Climate Change; Risks in Agriculture; Adding Value in Agricultural Productive Chains; Consumer Protagonism; and Convergence of Technology and Knowledge in Agriculture. These integrated megatrends signal the agricultural challenges for the country.

Supported by the demands, opportunities and megatrends raised in Agropensa, in 2018, Embrapa created its project portfolios that set the challenges to direct its research focus. There are currently 33 portfolios, totaling 330 innovation challenges focused on various areas of agriculture, livestock, commodities and food production, as well as automation, precision and digital agriculture, climate change, biotechnology, nanotechnology and intelligence, territorial management and monitoring. In particular, the goal of Automation, Precision and Digital Agriculture Portfolio is to plan, promote and monitor the processes of development, adaptation and dissemination of knowledge and technologies in automation, precision agriculture and digital agriculture to increase productivity and sustainability of production systems. Moreover, it will provide support for generating assets that add value to agricultural products and processes. Through this portfolio, the company seeks to promote research that will contribute to the digital transformation of Brazilian agriculture.

Also within the scope of digital transformation in agriculture, the agricultural startups, the AgTechs, which are innovative technology-based firms focused on developing digital applications in agriculture. AgTechs play an important role in the implementation of digital agriculture in Brazil. According to the 2nd AgTech Census – Startups Brazil, carried out by AgTechGarage, the largest investments made by AgTechs are in the development of solutions for soy (46%), corn (41%), beef cattle (30%), sugar cane (35%), coffee (25%) dairy cattle (20%), citrus (18%), forestry (15%), fish (11%), swine (10%) and poultry (10%). In addition to these, there are also solutions for horticulture, fruit farming, cotton, organic and agroecological agriculture and equine production (Agtechgarage, 2020). In the study of Radar AgTech Brazil 2019: Mapping of Startups in the Brazilian Agro Industry, carried out in partnership with SP Ventures, Homo Ludens and Embrapa, within the Bridges for Innovation Program (Pontes para Inovação), it was found that there are currently a total of 1,125 startups related to agriculture in Brazil, with 196 in the pre-production phase, 397 in the production phase and 532 in the post-production phase (Dias et al., 2019). This number only tends to grow, given the importance of agribusiness for the Brazilian trade balance and the need to modernize and use new digital technologies so that the sector maintains its strength in the country's economy and in the global food supply.

Incorporating megatrends in agriculture: innovation ecosystem for digital agriculture

Reducing risk and vulnerability in agriculture and agricultural business, as well as increasing their resilience and adaptation to the new conditions imposed by climate change, are completely dependent on a structure for organizing and processing data and information, based on powerful computational platforms, generating knowledge for field management and public and private decision-making. The Digital Agriculture Innovation Ecosystem and its connected research institutions will have a powerful tool to support decision-making and propose public policies that involve all agents in the production chain, including the final consumer. It is essential for federal, state and municipal governments to meet both rural and consumption demands throughout the country. This collaborative environment, illustrated in Figure 6, will also enable solutions for research programs related to bioeconomy, biotechnology and climatology, facilitating the transformation of these research results into products and technologies for the agricultural sector. This, in turn, can generate new demands, feeding back into the process of research, development and innovation ecosystem.

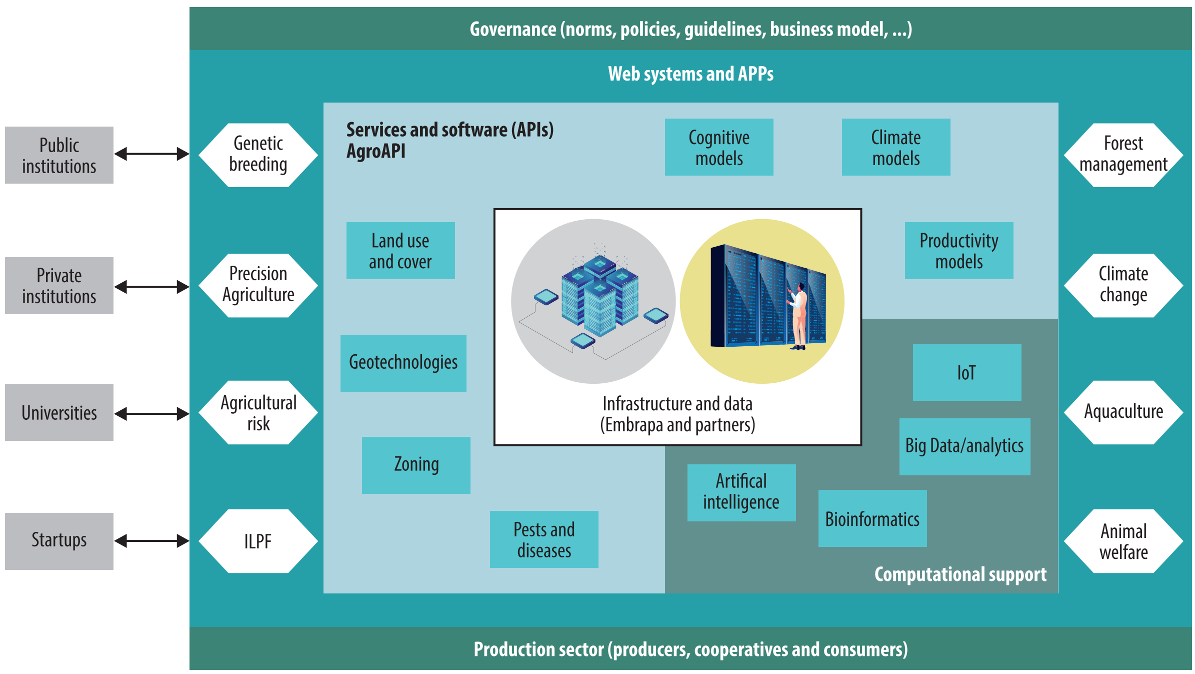


Figure 6. Innovation Ecosystem for Digital Agriculture.

Embrapa proposes and participates in this new Digital Agriculture Innovation Ecosystem, which is focused on the contribution of new disruptive technologies to add value to production, increase farmers profitability and food security. This reality imposes new challenges on entities linked to the sector, such as Embrapa, which must progressively act in cooperation, sharing expertise and knowledge for the development of new solutions, technologies and businesses. As a public research corporation, Embrapa can play a facilitating role in this environment of open innovation, bridging the gap between its various actors, which include rural producers, the public sector, research institutions, startups and companies in the ICT area and in the agricultural sector. In this ecosystem, Embrapa is ready to offer services and knowledge that can be shared by the entire agribusiness with a view to digital transformation in agriculture.

In the context of digital agriculture, which generates immense data and information, Embrapa proposes to implement a high-performance computational infrastructure (Data Center). This will provide support to improve the generation of relevant knowledge for the national agricultural policy and the integrated public and private risk management of agribusiness for the sustainable development of agriculture in Brazil. The new Data Center will have high capacity to store, organize and process data to generate information and knowledge that meets the demands of Agriculture 4.0. It will also allow the offer of three types of services: infrastructure, platform and software.

In the infrastructure service, Embrapa's partners will be able to use the managing services of large data volumes securely, including storage, high-performance processing, memory capacity and backup services for the data generated in their research.

In the software service, it will be possible to access several applications developed by Embrapa, such as Agrometeorological Monitoring System (Agritempo) (Agritempo, 2020), Agricultural Planning and Monitoring System (Webagritec) (Massruhá et al., 2008), Temporal Vegetation Analysis System (Satveg) (SATVeg, 2020), among others. Applications that can be developed both by Embrapa and its partners will be available. With regard to Embrapa, several applications related to the availability of information on bio-inputs, management of dairy farms, measurement of greenhouse gas emissions, production cost, indication of the best time to plant crops, provision of agrometeorological information for municipalities and Brazilian states, cattle production management, among others, are available at Embrapa's app store, both on Google Play and on Apple App Store.

In the platform service, Embrapa is already making available a pioneering tool in Brazil to serve the market of digital agriculture technologies called the AgroAPI platform (Vaz et al., 2017; Agroapi, 2020). AgroAPI provides information and models that can be used by companies and startups to create software, web systems and mobile applications for the agricultural sector, with lower cost and time reduction. The technology also enables an interface with mobile devices and embedded equipment that may emerge with the growth of the internet of things, which is fundamental for the digital transformation in rural areas. The AgroAPI platform will allow creating a supply and demand network for shared services that will benefit research networks and institutions in Brazil such as universities, startups, public and private institutions, since the data will be stored securely, and shared according to the interests of each institution. All these institutions can both consume the stored data and systems or make available the data and systems produced by them. In this initiative, the experimental fields of Embrapa Research Centers will act as testbeds to carry out experiments with disruptive technologies in the field, in partnerships with the public or private sector, working as showcases to demonstrate the implementation of digital agriculture.

Producers, cooperatives, farmers and technology and transformation companies will benefit from this entire infrastructure, as they will have access to vast aggregated, analyzed, and available information that will help their decision-making.

By developing collaborative work within the Digital Agriculture Innovation Ecosystem, Embrapa will work together with Unicamp's International Hub for Sustainable Development (HIDS) (Hub Internacional para o Desenvolvimento Sustentável, 2020). It focus on digital agriculture for the development and sharing of services for users of the agricultural production sector in the state of São Paulo, with the support of the municipality of Campinas and the state government. HIDS' vision is to contribute to the sustainable development process, aggregating national and international efforts to produce knowledge, innovative technologies and education for future generations, mitigating and overcoming the social, economic and environmental fragilities of contemporary society.

In this ecosystem, Embrapa Research Centers work on development and innovation focused on agriculture (agricultural, livestock, forestry and agro-industrial activities) and on the environment, integrating the demands of production systems with the needs for the conservation of natural resources and environmental preservation. Its research generates significant impact on public policies such as the Agricultural Climate Risk Zoning (ZARC), the Low Carbon Agriculture Plan (ABC Plan), the National Network for Research and Environmental Monitoring of Aquaculture in Union waters (Network) and the National Inventories of Agricultural and Waste Emissions. In addition to these, it is also involved in the preparation of Life Cycle inventories, which support environmental performance assessments in the different production chains that are important to Brazil, such as sugar-energy, as a result of the National Biofuels Policy (RenovaBio). In addition, Embrapa will be able to provide shared infrastructure of important multiuser laboratories, such as the Multiuser Bioinformatics Laboratory (LMB), the National Reference Laboratory of Precision Agriculture (Lanapre), the National Laboratory of Nanotechnology for Agribusiness (LNNA), the Multiuser Laboratory of Spectroradiometry, the Multiuser Laboratory of Chemistry of Natural Products (LMQPN), the Multiuser Laboratory of Biosafety for Livestock (BIOPEC), the Multiuser Laboratory of Molecular Biology (LMBM), the Sustainable System Analysis Laboratory (LASS) and the Multiuser Complex of Livestock Bio-efficiency and Sustainability (CMB), comprising four laboratories: Livestock Metabolism and Environmental Impacts, Biotechnology and Environment, Precision Livestock and Animal Health.

Another Embrapa contribution is the availability of experimental field structures that enable digital transformation in the field, through remote data collection, as well as management and decision-making, made possible by the interaction of electronically identified animals, equipment, actuators and sensors.

This experimentation structure will allow generating research data, for example in: 1) integrated systems (crop-livestock-forest integration, crop-livestock integration, livestock-forest integration); 2) milk production, milk quality and milk composition; 3) zootechnical data; 4) animal behavior; 5) physiological parameters; 6) animal consumption; 7) greenhouse gas emissions; and 8) edaphoclimatic data. Other expected deliverables are the development of applications for mobile devices for real-time monitoring, management and decision-making, and the development of digital platforms for information sharing and access to Embrapa's Spatial Data Infrastructure (Geoinfo) (Geoinfo, 2020). The spatial data can be related to Land Mapping, Agricultural Aptitude Zoning, Ecological-Economic Zoning, Land Use and Coverage Mapping and Monitoring, Land Use and Occupation Monitoring, Relief Data and Climatic Data.

This entire ecosystem will be governed by standards, policies and business models agreed between the partners and in accordance with the guidelines established by the federal government.

Final considerations

Brazil is one of the largest agricultural producers and exporters in the world, and the country needs to guarantee and expand its production capacity with sustainability while meeting the global demand for food and nutrition security. As a major exporter of agricultural commodities, modernization and innovation are needed throughout the agricultural production chain, thus converging to digital agriculture as a result of the digital transformation in the sector.

This work presented an overview of the evolution of ICT in agriculture and how digital transformation is driving the fourth industrial revolution and the emergence of Industry 4.0. Currently, this is inspiring the implementation of new technologies in digital agriculture and the consequent emergence of Agriculture 4.0 towards Agriculture 5.0.

Despite the growing interest and effort in implementing digital agriculture, there are challenges to be overcome, such as the difficulty in coordinating actions involving the various institutions and the business models to be practiced. In addition, the industry faces the lack and necessary amount of trained human resources; the need to guarantee information security; the definition of ownership when dealing with the generated data and information, as well as issues of data integration from different formats or different sources.

In this scope, some strategies for the full achievement of digital agriculture are conceptualized, such as: address the definition of rights and ownership of data; encourage the use of open standards protocols for data interoperability and communication between equipment; improve connectivity and broadband coverage for cell phones and the internet in rural areas. We also need to encourage research in order to support intelligent applications in agriculture, and lastly, we need to establish alliances between the public and private sectors to define strategies and policies for the implementation of digital agriculture in a collaborative way.

The challenges presented with the transformation of agricultural data, notably in relation to the digitization, digital collaboration, and sustainable development, make Embrapa one of the driving institutions in the implementation of digital agriculture in the country. One of the concrete initiatives to address these challenges is the creation of the Digital Agriculture Innovation Ecosystem, where Embrapa pursues the integration of the various segments and sectors by acting as a facilitator between companies interested in developing collaborative work.

References

ACCENTURE. Available at: https://www.accenture.com/br-pt. Accessed on: 18 June 2020.

ACELERADORA. Available at: http://aceleradora.net. Accessed on: 18 June 2020.

ACOMPANHAMENTO DA SAFRA BRASILEIRA [DE] GRÃOS: safra 2019/20: nono levantamento, v. 7, n. 9, p. 1-31, jun. 2020. Available at: https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra- de-graos/item/download/32042_910765cf5dc2d088c377e32cdef6fc72. Accessed on: 26 June 2020.

AGRITEMPO. **Sistema de Monitoramento Agrometeorológico**. Available at: http://www.agritempo.gov.br/agritempo. Accessed on: 29 May 2020.

AGROAPI. **A Plataforma de APIs da Embrapa**. Available at: https://www. agroapi.cnptia.embrapa.br/portal/index.html. Accessed on: 29 May 2020.

AGTECHGARAGE. **2º Censo AgTech Startups Brasil**. Available at: https:// www.agtechgarage.com/censo. Accessed on: 2 June 2020.

ARANTES, J. T. FAPESP criará oito centros de pesquisa em inteligência artificial com o governo federal. **Agência FAPESP**, 16 dez. 2019. Available at: http://agencia.fapesp.br/fapesp-criara-oito-centros- de-pesquisa-em-inteligencia-artificial-com-o-governo-federal/32196. Accessed on: 29 May 2020.

ASSOCIAÇÃO BRASILEIRA DE STARTUPS. **Tudo que você precisa saber sobre startups**. 2019. Available at: https://abstartups.com.br/o-que-e-uma-startup. Accessed on: 1 June 2020.

BERNARDI, A. C. de C.; NAIME, J. de M.; RESENDE, A. V.; BASSOI, L. H.; INAMASU, R. Y. (ed.). **Agricultura de precisão**: resultados de um novo olhar. Brasília, DF: Embrapa, 2014. 596 p.

BOYLE, J. Biology must develop its own big-data systems. **Nature**, v. 499, n. 7456, p. 7, 2013. DOI: 10.1038/499007a.

COMITÊ GESTOR DA INTERNET NO BRASIL. **TIC domicílios 2019**: principais resultados. 2019. 31 p. Available at: https://cetic.br/media/analises/tic_domicilios_2019_coletiva_imprensa.pdf. Accessed on: 2 June 2020.

DIAS, C. N.; JARDIM, F.; SAKUDA, L. O. (org.). **Radar AgTech Brasil 2019**: mapeamento das startups do setor agro brasileiro. Brasília, DF: Embrapa; São Paulo: VP Ventures: Homo Ludens, 2019. Available at: www.radaragtech.com.br. Accessed on: 2 June 2020.

EMBRAPA. **Automação e agricultura de precisão**. 2020b. Available at: https://www.embrapa. br/tema-mecanizacao-e-agricultura-de-precisao/nota-tecnica. Accessed on: 29 May 2020.

EMBRAPA. Secretaria Geral. Gerência de Comunicação e Informação. **Embrapa em números**. Brasília, DF: Embrapa, 2019. 140 p.

EMBRAPA. **Trajetória da agricultura brasileira**. 2020a. Available at: https://www. embrapa.br/visao/trajetoria-da-agricultura-brasileira. Accessed on: 29 May 2020.

EMBRAPA. **Visão 2030**: o futuro da agricultura brasileira. Brasília, DF: Embrapa, 2018. 212 p. Available at: https://www.embrapa.br/visao/o-futuro-da-agricultura-brasileira. Accessed on: 29 May 2020.

EUROPEAN AGRICULTURAL MACHINERY ASSOCIATION. **Digital Farming**: what does it really mean? And what is the vision of Europe´s farm machinery industry for Digital Farming? 13 Feb 2017. Available at: https://www.cema-agri.org/images/publications/position-papers/CEMA_Digital_Farming_-_Agriculture_4.0 13_02_2017_0.pdf. Accessed on: 20 May 2020.

EUROPEAN COMMISSION. **Innovating for sustainable growth**: a bioeconomy for Europe. Brussels, 2012. DOI: https://data.europa.eu/doi/10.2777/6462.

FAO. **"Climate-smart" agriculture**: policies, practices and financing for food security, adaptation and mitigation. Rome: FAO, 2010. The document was prepared as a technical input for the Hague Conference on Agriculture, Food Security and Climate Change, to be held 31 October to 5 November 2010. Available at: http://www.fao.org/fileadmin/user_upload/newsroom/docs/the-hague-conference-fao-paper.pdf. Accessed on: 29 May 2020.

FRASER, E. D. G.; CAMPBELL, M. Agriculture 5.0: reconciling production with planetary health. **One Earth**, v. 1, n. 3, p. 278-280, Nov. 2019. DOI: 10.1016/j.oneear.2019.10.022.

GAZZOLA, P.; COLOMBO, G.; PEZZETTI, R.; NICOLESCU, L. Consumer empowerment in the digital economy: availing sustainable purchasing decisions. **Sustainability**, v. 9, n. 5, 2017. DOI: 10.3390/su9050693.

GEOINFO. **Infraestrutura de dados espaciais da Embrapa**. Available at: http://geoinfo.cnpm.embrapa.br. Accessed on: 29 May 2020.

HUB INTERNACIONAL PARA O DESENVOLVIMENTO SUSTENTÁVEL. Available at: http://www.hids.depi.unicamp.br. Accessed on: 29 May 2020.

IBGE. Censo agropecuário 2017: resultados definitivos. **Censo Agropecuário**, v. 8, p. 1-105, 2019. Available at: https://biblioteca.ibge.gov.br/visualizacao/periodicos/3096/agro_2017_resultados_definitivos.pdf. Accessed on: 28 May 2020.

JACTO. Acompanhe a evolução dos implementos agrícolas. **Blog Jacto**, 29 maio 2018. Available at: https://blog.jacto.com.br/acompanhe-a-evolucao-dos-implementos-agricolas. Accessed on: 29 May 2020.

KIM, K.-H.; SHIM, W.; MOON, Y.-H.; KIM, K.-H.; SON, J.-K.; KWON, O. J. The structure of bio-information nano technology convergence from firms' perspective. In: PICMET: TECHNOLOGY MANAGEMENT FOR EMERGING TECHNOLOGIES, 12., 2012, Vancouver. **Proceedings** [...] New York: IEEE, 2012. p. 579-588.

MASSRUHÁ, S. M. F. S.; LIMA, H. P. de; EVANGELISTA, S. R. M.; PINTO, G. E. M. **Uma proposta de plataforma de software para integração e interoperabilidade de serviços Web - Webagritec**: estudo de caso. Campinas: Embrapa Informática Agropecuária, 2008. 23 p. (Embrapa Informática Agropecuária. Boletim de pesquisa e desenvolvimento, 19). Available at: https://ainfo. cnptia.embrapa.br/digital/bitstream/item/11850/1/bp19.pdf. Accessed on: 29 May 2020.

MAY, M. Big biological impacts from big data. **Science**, v. 344, n. 6189, p. 1298-1300, June 2014. DOI: 10.1126/science.344.6189.1298.

NATIONAL bioeconomy blueprint. Washington, DC: The White House, Apr. 2012. Available at: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ ostp/national_bioeconomy_blueprint_april_2012.pdf. Accessed on: 20 June 2020.

ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **The bioeconomy to 2030**: designing a policy agenda. Paris, 2009. 323 p. DOI: 10.1787/9789264056886-en.

PILLON, C. N. Dos pós de rocha aos remineralizadores: passado, presente e desafios. In: CONGRESSO BRASILEIRO DE ROCHAGEM, 3., 2017, Pelotas. **Anais** [...] Assis: Triunfal Gráfica e Editora, 2017. p. 16-23. Editado por Adilson Luis Banberg, Carlos Augusto Posser Silveira, Éder de Souza Martins, Magda Bergmann, Rosane Martinazzo e Suzi Huff Theodoro.

PIVOTO, D.; WALQUIL, P. D.; TALAMINI, E.; FINOCCHIO, C. P. S.; CORTE, V. F. D.; MORES, G. V. Scientific development of smart farming technologies and their application in Brazil. **Information Processing in Agriculture**, v. 5, n. 1, p. 21-32, Dec. 2018. DOI: 10.1016/j.inpa.2017.12.002.

PRINS, R. **3 ways COVID-19 is pushing Ag data transformation**. 2020. Available at: https://www.precisionag.com/digital-farming/3-ways-covid- 19-is-pushing-ag-data-transformation. Accessed on: 1 June 2020.

PRODUTO 7C: aprofundamento de verticais – rural. Brasília, DF: Banco Nacional de Desenvolvimento Econômico e Social: Ministério do Planejamento, Desenvolvimento e Gestão; Ministério da Ciência, Tecnologia, Inovações e Comunicações, dez. 2017. Available at: https://www.bndes.gov.br/ wps/wcm/connect/site/2fa8f7d1-9939-441d-b8ce-ed3459fcfd4d/relatorio-aprofundamento-das-verticais-rural-produto-7C.pdf?MOD=AJPERES&CVID=m3rPopG. Accessed on: 2 June 2020.

PROJECT BREAKTHROUGH. **Digital agriculture**: feeding the future. 2017. Available at: http:// breakthrough.unglobalcompact.org/disruptive-technologies/digital-agriculture. Accessed on: 1 June 2020.

SAIZ-RUBIO, V.; ROVIRA-MÁS, F. From smart farming towards agriculture 5.0: a review on crop data management. **Agronomy**, v. 10, n. 2, p. 1-21, Feb. 2020. DOI: 10.3390/agronomy10020207.

SANTIN, W. O campo em tempo real. **Globo Rural**, n. 418, p. 14-19, ago. 2020.

SATVeg: Sistema de Análise Temporal da Vegetação. Desenvolvido pela Embrapa Agricultura Digital. Ferramenta de acesso e visualização de perfis temporais dos índices vegetativos NDVI e EVI do sensor MODIS. Available at: https://www.satveg.cnptia.embrapa.br/satveg/login.html. Accessed on: 1 June 2020.

SHEPHERD, M.; TURNER, J. A.; SMALL, B.; WHEELER, D. Priorities for science to overcome hurdles thwarting the full promise of the 'digital agriculture' revolution. **Journal of the Science of Food and Agriculture**, v. 100, n. 14, Sept. 2018. DOI: 10.1002/jsfa.9346.

SORENSEN, A. G. Smart farming and digitalization. In: GIGR INTERNATIONAL CONFERENCE, 5., 2021, Quebec. **Proceedings**... [S.l.]: International Commission of Agricultural and Biosystems Engineering, June 2020. p. 2-4. Newsletter 121. 2020. Available at: http://cigr.org/documents/CIGR_NL121.pdf. Accessed on: 1 Jun. 2020.

SPRINGER. **Precision agriculture**. Available at: https://www.springer.com/ journal/11119/updates/17240272. Accessed on: 1 June 2020.

STARTUPBASE. **Encontre todas as startups do Brasil**. Available at: https://startupbase.com.br/home. Accessed on: 1 June 2020.

TRANSFORMAÇÃO DIGITAL. **O que é transformação digital?** Available at: https://transformacaodigital.com/o-que-e-transformacao-digital. Accessed: 29 May 2020.

VARSHNEY, R. K.; TERAUCHI, R.; MCCOUCH, S. R. Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. **PLoS Biology**, v. 12, n. 6, e1001883, June 2014. DOI: 10.1371/journal.pbio.1001883.

VASCONCELOS, M. J. V. de; FIGUEIREDO, J. E. F. **Tecnologia CRISPR-Cas para edição genômica**. Sete Lagoas: Embrapa Milho e Sorgo, 2015. 37 p. (Embrapa Milho e Sorgo. Documentos, 197).

VAZ, G. J.; APOLINÁRIO, D. R. de F.; CORREA, J. L.; VACARI, I.; GONZALES, L. E.; DRUCKER, D. P.; BARIANI, J. M.; EVANGELISTA, S. R. M.; ROMANI, L. A. S. AgroAPI: criação de valor para a Agricultura Digital por meio de APIs. 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. 59-68. SBIAgro 2017.

WORLD ECONOMIC FORUM. **Digital transformation initiative**: executive summary. Geneva, 2017. Available at: https://www.accenture.com/_acnmedia/Accenture/Conversion-Assets/ WEF/PDF/Accenture-DTI-executive-summary.pdf. Accessed on: 1 May 2020.

Photo: Have a nice day (AdobeStock)

2

Digital agriculture
Definitions and technologies

Kleber Xavier Sampaio de Souza | Stanley Robson de Medeiros Oliveira | Carla Geovana do Nascimento Macário | Julio Cesar Dalla Mora Esquerdo | Maria Fernanda Moura | Maria Angelica de Andrade Leite | Helano Póvoas de Lima | Alexandre de Castro | Sônia Ternes | Inácio Henrique Yano | Edgard Henrique dos Santos

Introduction

Advances in information processing and in the areas of nanotechnology, biotechnology, and cognitive science are promoting a convergence between sciences, which is currently called Nano-bio-info-cogno. The report commissioned by the National Science Foundation of the United States named *Converging Technologies for Improving Human Performance: Nanotechnology, Biotechnology, Information Technology and Cognitive Science* (Roco; Bainbridge, 2003), was prepared by more than 100 scientists and pointed out the synergy between nanotechnology, biotechnology, information technology, and cognitive science as the segment with the greatest potential for advancement in innovation. This report highlights that systemic approaches using mathematics and computing will, for the first time, enable us to understand the functioning of complex systems in the natural world such as the human mind, stellar explosions, social interactions, organs of the human body, and the natural phenomena involved with agriculture.

Agriculture works directly with three of these areas, nanotechnology, biotechnology, and information technology. In fact, it has been influenced and fueled by the staggering growth of data acquisition capacity coming from different sources. This data ranges from the cell scale, such as information obtained by analysis in the field of "omics" sciences (genomics, proteomics, transcriptomics and metabolomics), to the macroscopic scale, which includes socioeconomic data and data obtained through remote sensing devices, such as satellites. At the local scale, this can also come from farms equipped with agricultural equipment and sensors.

Digital agriculture is an increasingly connected and remote operation that surveys and processes large amounts of data collected in all links of the production chain: pre-production, production and post-production. This involves different types of digital technologies: sensors embedded in orbital, suborbital, airborne, or autonomous systems (drones, agricultural machines) which can be installed directly in the field or in different 'things' (Internet of Things – IoT) along the production chain. The technologies involve telecommunication systems, global positioning, control, management, analysis software (data analytics), and actuators.

Data from these technologies are now collected not only by conventional means, but also from collaborative platforms or social media (citizen science), among others. Their accumulation poses a challenge to storage, search, and retrieval systems, while also impacting processing and retrieval methods.

The abundance of data creates a great gap in terms of management and analysis capacity, and consequently in terms of the production of knowledge from them. This precipitates into a complex scenario, where transforming data into information and knowledge assumes a strategic role in all sectors of the economy, including agriculture, a strategic sector for Brazil. All these data need to be integrated, preprocessed, and analyzed so that the required knowledge to establish digital agriculture can be extracted.

This chapter presents the concepts related to digital technologies that are used throughout this book. This is done in a consolidated way in order to facilitate the readers' understanding and access.

Digital technologies

The digital technologies presented here are divided into five groups. In the first group there are technologies linked to the organization and representation of information. In the second there are the mathematical and statistical modeling techniques involving biological, social, and environmental phenomena. In the third, the application of artificial intelligence in agriculture. In the fourth group, sensor and robotics technologies. In the fifth and last group there are technologies in which applications interact with agriculture, such as cloud computing and blockchain.

Organization, representation, and information access

The volume of information and diversity of formats (DNA data, satellite images, sensor data) in which this information is presented represent an enormous challenge to its organization and reuse. It becomes necessary to annotate, classify, structure and provide access mechanisms so that information can be found and reused in the future, which is the purpose of the technologies in this section.

Thesaurus – According to ANSI/NISO Z39.19-2005 (National Information Standards Organization, 2010), thesauri are controlled vocabularies, arranged in such a way that the relationship between their terms is clearly identified and standardized. Terms are composed of one or more words and selected from natural language to be included in a thesaurus. In the National Agricultural Thesaurus (Thesagro), for example, the word *mite* is related to the word *Arachnid*, so that *Arachnid* is the broader term (BT) and mite is the more specific term (NT - Narrower Term). There are also other terms subordinate to *Arachnid* and which, therefore, are NT of *Arachnid*, such as *spider* and *scorpion*. BT and NT are forms of vertical relationship between terms used in thesauri; there are also horizontal associations between terms, expressed as a related term (RT - Related Term). For example, in Thesagro there are the terms *Acaricide* and *Tick* associated with the RT of *Mite*. The first is a counter agent for mites, belonging to the hierarchy that starts with *Pesticide*, and the latter belongs to the hierarchy of *Animal Parasite*. Thesauri, therefore, form a web of relationships between terms, and this web helps to find the information one is looking for. The terms and their hierarchy can be used to organize the content of websites on the internet and to expand the searches that are performed on a determined content. For example, when searching for *Acarus*, documents that speak of *tick* or *Acaricide* can also be retrieved.

Ontology – An ontology is formally defined as a common vocabulary for sharing information about a particular knowledge domain. The ontology includes machine-interpretable definitions of the basic concepts of this domain and the relationships between these concepts (Noy; Mcguiness, 2001).

In ontologies, the relationships between domain concepts are made explicit so that they can be interpreted by computers. Each concept contains its attributes, for which there are possible values. For example, the concept "vehicle", which is a class, contains the subclasses "car" and "motorcycle". A car usually has four wheels and a motorcycle has two. So, the attribute number of wheels would be four for a vehicle and two for a motorcycle. Both the motorcycle and the car have a manufacturer attribute and any other attributes one would want to add and enrich the information contained in the ontology for its reuse. As ontologies provide a common machine-processable language, an agent can automatically browse several websites that work on the same subject – for example car parts – and add the information provided by them for price comparison. This is greatly facilitated when multiple sites use the same ontology to describe their parts.

Big Data – The term Big Data includes data sets where sizes go beyond the capacity of data management systems to process them. It is usually data from various sources, such as mobile devices, body sensors, social media, emails, electronic medical records, genomics and geospatial sensor data, among many others. This variety of sources, the amount of data, and the speed at which the data arrives for processing generate what is called the "three Vs" of Big Data: volume, velocity and variety, to which sometimes "veracity" and "value" are also added. The definition encompasses structured, semi-structured, and unstructured data, although the treatment of unstructured data by systems that process Big Data is much more common (Dedić; Stanier, 2017). Big Data applications appear all the time: when analyzing posts on social networks about a certain subject to see their repercussion; when analyzing Google searches to identify outbreaks of flu pandemics. Given the inadequacy of traditional database management systems in processing Big Data, solutions were developed by companies that traditionally operate with large volumes of data, such as Google and Cloudera, which developed MapReduce, Flume, and Sqoop. MapReduce (Dean; Ghemawat, 2008) is an algorithm developed by Google with a free implementation developed by the Apache Foundation, called Hadoop (White, 2012). It operates by distributing large datasets to be processed on multiple computers in parallel (possibly thousands of computers) and then consolidating answers. Apache Flume[1] was originally developed by Cloudera to manage large volumes of log file data, but it has been extended to process events from web sources such as Twitter and Facebook. Apache Sqoop[2] is a tool to efficiently transfer data between structured, semi-structured, and unstructured data sources. It is an interesting tool for bringing data from external sources, such as relational databases, into Hadoop's distributed file system. MapReduce, Flume, and Sqoop are just examples of systems that were developed to handle Big Data, and they are not the only systems capable of handling data with large volume, variety, and velocity.

API – An Application Programming Interface (API) is a way for two applications to talk to each other. A requesting application triggers the execution of another so that its own task is completed, in other words, the requesting application needs the second one as a provider for its functioning. The communication intermediary between the two applications is the API, which defines protocols, routines and tools so that the message is delivered to the provider application and the response returns to the requesting

---

[1] Available at: http://flume.apache.org

[2] Available at: https://sqoop.apache.org

application. A Web API operates on the internet using the usual protocols for exchanging information, such as HTML, XML, or JSON. As an example of API applied to the agricultural area, we can mention the AgroAPI platform from Embrapa Agricultural Informatics (2020). It provides a series of information and models that can be coupled with software, web systems, or mobile applications from companies, including startups, public, and private institutions. Each API is allowed free use of up to 1,000 requests per month. An example is API Agritec, part of AgroAPI, which gathers information on planting time, fertilization, productivity, agricultural zoning and cultivars for five agricultural crops. Another example is the SATVeg API, which uses satellite data to generate the visualization of the evolution over time of NDVI and EVI vegetation indices for all of South America. These indices make it possible to observe variations in green biomass on the land surface, and may help implementing the Brazilian Forest Code (Código Florestal), or monitoring the cycle of certain agricultural crops, among other land cover and land use dynamics.

Mathematical modelling and statistics

The representation of natural phenomena through models is an integral part of the scientific method. This section is dedicated to showing the representation through models used in the scientific method. It also conceptualizes Data Science, which emerged from a confluence of various branches of expertise to extract knowledge from increasingly abundant masses of data.

Mathematical model – A model arises from the need to understand a phenomenon in the physical world and predict its behavior in a given situation. A model is always an abstraction of what happens in the real world, a simplification of what happens in reality so that a system can be understood and quantified (Torres; Santos, 2015). In Bassanezi (2002), a mathematical model consists of transforming reality into mathematical problems, which are solved and interpreted in light of what happens in the real world. Building a mathematical model involves several steps: a) conceptualization, which occurs after initial observations about a problem, formulation of hypotheses to explain its functioning, and a first selection of which variables, processes, and interactions are considered relevant. An important task occurs during the conceptualization, which is the simplification of the model in terms of variables and interactions that are essential for the representation of the problem, since the phenomena of the natural world, especially the biological ones, are extremely complex; b) mathematical formalization, which is the translation of the problem into mathematical language. There are many different approaches to this translation, such as differential equations, Bayesian equations, stochastic systems, finite difference equations, and agent-based systems, each with its advantages and limitations. Its choice depends on the nature of the problem that is being modeled; c) parameter estimation, which involves discovering which numerical values are guiding the elaborated mathematical formulation. These parameters can be obtained through experimental measures, and the adoption of experimental statistical techniques adds greater reliability; d) simulation and prediction, which is the moment when the system of equations is solved analytically, or the model is run on a computer. As biological problems usually involve control and regulation mechanisms, the analytical solution of the models is almost always impossible, which makes the computational approach the most frequent to solve a model; e) model validation, in which the system response is verified for each scenario of the input values of the model variables. This answer has to coincide both in terms of the trajectory of the system and with the values obtained in the experimental measurements. Therefore, it is at this point that one evaluates how close the model is representing reality while its accuracy is being measured. Another desired feature of the model is its capability to predict new facts and unknown relationships that can be verified in the real world; f) model refinement, where the validity of the model results is criticized in terms of the trajectories of the modeled system when confronted with the real world while its accuracy is evaluated. When models deviate from what was expected, it may be due to some hypotheses that either have not been considered or that are false. There may also have been an error in obtaining the data that fed the construction of the model or an error in its mathematical formulation. In this case, new hypotheses and/or new variables and a re-verification of the mathematical model may be proposed.

Statistical model – Statistics is the basis of the scientific method, which can be summarized as follows: a) definition of the problem to be studied; b) formulation of one or more hypotheses to be tested; c) conducting experiments to test the formulated hypotheses; d) statistical analysis of the data obtained; e) interpretation of results and obtaining conclusions, that is, obtaining a descriptive or inferential statistical model that proves or not the original hypotheses. As example (Snedecor; Cochran, 1967): the problem to be studied was the variability in the calcium concentration of turnips; the hypotheses concerned the behavior of the variability of calcium in plants and, specifically, in the leaves of each plant; in the experiment, four plants were randomly chosen, and then three leaves of each plant were randomly selected. Two 100 mg samples were taken from each leaf, determining the amount of calcium in each sample through microchemical processes ; the data were submitted to an analysis of variance according to the model of the formulated hypotheses; the analysis concluded that, statistically, at a 5% significance level, the variability in the leaves of each plant is more important than the variability in the whole plant, and that the ideal model (the hypotheses raised) statistically represents reality. Each of these variability effects is estimated according to the initial hypotheses, the calculated estimates show whether the model is adequate or not to the formulated hypotheses based on an accepted margin of error, which in this case was 5%. In general, biological processes are inherently complex and the variability of each observed factor needs to be estimated. That means that the number of observed variables is enormous, and sometimes not all variables are known. Those not known will introduce a greater error to the estimated model; remembering that the model is accepted after the estimates are statistically proven, and that the model as a whole has an error, which is also estimated. In this scenario, and in many others, such as general gas theory or natural selection (Fisher, 1934), the arguments are built on statistical grounds.

Data Science – Data Science is an interdisciplinary field focused on the processes and systems for extracting knowledge or insights from data in various forms, structured or not. It incorporates techniques and theories from the most diverse areas of knowledge such as computing, engineering, mathematics, statistics, economics, data mining, and artificial intelligence in order to collect, process, integrate, and analyze data so as to create data products and services (Amaral, 2016). Data Science is not restricted to analyzing large volumes of data (Big Data analytics). Small (Small Data) and large (Big Data) data repositories are important aspects of this research area. Small Data includes simple information, which is in the database of any company or small rural property. Small Data includes research results, consumer or rural producer data, data on agricultural properties, e-mails with information on management practices, data containing agricultural production volume per period, among others. It usually consists of structured data ready for analysis. Big Data, on the other hand, refers to (mainly) unstructured data, originating from multiple sources and which should be collected, aggregated, and analyzed in order to generate managerial information. Among the Data Science applications are digital marketing, which elaborates personalized advertisements based on information obtained through user profiles and their browsing history in companies. Further examples are recommendation systems, which are based on the pattern of pages visited or products purchased to suggest new products, and bank customer credit rating systems, which consults track record and existing scores in credit protection companies to calculate the likelihood of a customer defaulting.

Artificial intelligence

Pattern recognition and machine learning technologies, including deep learning, are an integral part of the many systems used today, such as autonomous cars and voice recognition systems. These new technologies analyze large data sets and learn patterns from them that allow, for example, to identify objects or anticipate the next word to be spoken in a sentence. On the other hand, fuzzy logic is used when the rules of a system are explained directly, without the use of machine learning, but still allows a certain degree of imprecision.

Pattern recognition – A pattern, as understood within the concept of pattern recognition, can be the representation of a handwritten number, a number written on a house, an orange, a car, a pronounced word, sequences of temperature measurements, pressure and rain, stock market value sequences, as well as many other things we want a computer system to learn to recognize. It is for this reason that many of the important pattern recognition problems can be characterized either as waveform classifications (sounds, temperature measurements, action values, etc.) or as classification of geometric figures, as with images (Fukunaga, 1990). Our brain is specially designed to recognize patterns. In the very early years of our existence, we learned to differentiate sounds, words, what is a cat and how it is different from a dog and so many other things. What pattern recognition in a computational system should achieve is to learn to differentiate the data presented to it, an activity that is computationally known as classification. To process this data, several techniques are used for pattern recognition, such as decision trees, random forests, k-nearest neighbors, support vector machines and neural networks (Bishop, 2006). The application of pattern recognition techniques can indicate, for example, that a given sequence of temperature values is within normality, that a stock listed on the stock exchange is on a downward trajectory, that the handwritten number on a paper is 3, or that the object in a certain position in an image is an orange.

Machine learning – This is a process closely related to pattern recognition (previous topic), as what is desired during machine learning is for the computer to learn from the patterns presented to it. According to Bishop (2006), machine learning and pattern recognition are two facets of the same field of knowledge, with pattern recognition originating from engineering and machine learning from computing. For this reason, the algorithms between pattern recognition and machine learning are also shared. Generally, it is possible to divide machine learning into supervised, when starting from a previously defined set of labeled data one wants to find a function that is capable of predicting unknown labels; and unsupervised, which seeks to identify groups or patterns from the data, without a specific objective to be achieved (Russel; Norvig, 2020). These two concepts are defined as follows.

Unsupervised learning – In this type of learning, the data set used does not have any type of label. The objective of this type of learning is to detect similarities and anomalies between the analyzed objects. The process of grouping objects into similar classes is called clustering. This procedure is also known as data segmentation, as it partitions large datasets according to the similarity between subsets. The objects that are more similar to the characteristics imposed by the domain must be allocated in the same group, while those less similar must be allocated in different groups. The similarity between objects must be obtained by algebraic measures, such as the Euclidean distance, for real values; or by simple correspondence, for nominal values. These algorithms can be divided into two more general classes, according to the heuristic used to construct the groups.

The first class refers to partitional algorithms, which, usually with linear execution computational cost, operate in an iterative way based on the previous definition of the desired number of groups and the definition of representative objects of each group, known as centroids. In each iteration, each object is associated with the centroid and, consequently, with its most similar group. The centroids of the groups are then recalculated for the next iteration. The algorithm reaches its point of convergence when the centroids are no longer changed between one iteration and another, that is, when the groups are well defined, considering the similarity measure used. The k-means is in this subclass of algorithms (Macqueen et al., 1967), and is considered one of the ten most influential algorithms in data mining (Wu, 2008).

The second class is hierarchical algorithms, which have a computational cost of execution that is normally quadratic and therefore do not require the identification of initial representatives or the desired number of groups. Thus, in a single execution, *n* nested partitions can be generated for the same set of *n* instances, containing from 1 to *n* groups each, constituting a cluster hierarchy (Han; Kamber, 2006). Two distinct strategies can be used to build this hierarchy: the agglomerative one, which initially considers each instance of the dataset as a group, merging pairs of groups in each iteration; and the divisive, which initially considers all samples belonging to a single group, dividing them into smaller groups in each iteration (Hastie et al., 2009).

Supervised learning – The process of supervised machine learning consists of presenting a large amount of previously classified data to a computer and making it learn from that data. Learning happens by modifying system parameters as more and more examples are presented to it. These parameters are numbers for which their values are unknown. So, the learning task is to find out which values make the system get it right most often. For each example, it is verified whether the system learned to correctly classify that example. If it is right, the system reinforces the parameters that allowed this correct classification through their weights. Otherwise, it calculates what correction the system must undergo so as not to make this error and negatively adjust the weights that led to that wrong answer. One can then imagine a system with many knobs or buttons that have to be turned just the right amount for that system to finally get the answer right. However, instead of turning the knobs ourselves, we have algorithms that do it in a controlled way in order to make learning happen. The main learning paradigms can be listed as follows:

a) **Symbolic (decision trees):** a decision tree is a flowchart-like structure in which each internal node represents a "test" in an attribute, each branch represents the test result, and each leaf node represents a class label (decision made after computing all attributes). The paths from root to leaf represent classification rules (Quinlan, 1986).

b) **Based on instances (*k*-NN or k nearest neighbors):** the main idea of *k*-NN is to determine the classification label of a sample based on the *k* neighbor samples from a training set. Among the *k* examples, there is the most frequent class. This class is attributed to the new example (Fukunaga; Narendra, 1975).

c) **Based on statistical learning (Support Vector Machines – SVM):** the simplest way to partition an *n*-dimensional Euclidean space is through hyperplanes. The SVM classifier is also based on this strategy, however, it uses a special type, the optimal separating hyperplane. It is a hyperplane that divides classes, maximizing the margin of separation between them (Vapnik, 1995, 1998).

d) **Committee-based:** it is the field of machine learning that builds a group of classifiers, called base classifiers, in order to be more accurate than the best elements of the group. The simplest approach based on this algorithm is simple majority voting, in which several classifiers are combined into one voting strategy. As a result, the response that receives the highest number of votes is considered the committee's response (Han; Kamber, 2006). Random Forest is an example of this type of approach. It is a classification and regression technique that consists of a set of decision trees combined to solve classification problems (Breiman, 2001).

e) **Connectionist (Artificial Neural Networks – ANN):** they are computational models inspired by the central nervous system (particularly the brain), capable of performing machine learning, as well as pattern recognition. An example of a connectionist model is the deep learning technique, detailed as follows.

Deep learning – The deep learning technique, or deep neural networks, is a machine learning technique in which the model chosen for the learning algorithm is an artificial neural network with many layers. Neural networks were inspired by the way neurons function in biological systems, operating in a parallel and decentralized way (Marblestone et al., 2016). Typically, a neural network can contain more than 100 layers, arranged one after the other or in parallel. Each of these layers is composed of one or more neurons, interconnected so that the result of neurons that are in one layer feeds the input of neurons that are in the posterior layer. The neural network training method often employs an algorithm called backpropagation. Since it is associated with each neuron, there is a weight that that neuron represents in the response, this algorithm compares the system response with the value that should have been and distributes the error by recalculating the values of the weights of the neurons backwards. There are many neural network architectures available, such as feedforward networks, convolutional networks, recurrent networks and restricted Boltzmann machines, among many others (Goodfellow et al., 2016). The architecture chosen for the network depends on the problem to be solved: forward connected networks are used in both classification and regression problems; convolutional networks, for image classification problems; recurrent networks, for problems involving sequences, such as natural language processing; and the restricted Boltzmann machines are applied for dimensionality reduction, a task with a large amount of input variables, and the most significant ones are identified. This list of problems for each network is not exclusive, it is only to serve as an example, as a constrained Boltzmann machine can be used to solve other problems, such as regression and classification, like with other neural network architectures. The deep learning area includes a great deal of art involved in selecting a given architecture for a problem, as well as in the parameterization of models.

Fuzzy sets and fuzzy logic – The classical set theory defines a class of objects with binary membership as a set, that is, each element may or may not belong to the set ($\in$ or $\notin$). Zadeh (1965) founded the concept of fuzzy sets (FS) as a class of objects in which each element has a continuous degree of membership, admitting any value between zero and one. This concept allows addressing real-world problems, where membership criteria and boundaries between classes are not precisely defined (that is, they are fuzzy). An element can have different "degrees of membership" for various sets. Analogous to the theory of classical set, a whole class of logical operations is derived from fuzzy sets, called fuzzy logic. Those that operate with fuzzy logic are called Fuzzy Rules Based Systems (FRBSs). They are inference systems whose logical components are expressed through FS. Typically, it is composed of a fuzzy database (input and output variables), an inference mechanism and a fuzzy rule base of the "IF A then B" type, whose linguistic terms are FS (Klir; Yuan, 1995).

Earth and study sensors

Sensors and actuators are at the heart of digital agriculture, as they enable to perceive what is happening in the environment, and take appropriate actions. Sensors can be orbital, such as satellites, which allow the collection of geospatial data, or proximal data, such as sensors installed on rural properties and linked to the Internet of Things devices. When computing is fully integrated with the sensors of an environment and distributed in that environment, we have Ubiquitous Computing.

Ubiquitous computing – The term Ubiquitous Computing was proposed by the scientist Weiser (1991), of Xerox's Palo Alto Research Center (PARC), to refer to a computing paradigm proposed for the 21st century. In this paradigm, computing should be everywhere, hence the term ubiquitous, and invisible to its user. To explain the concept, Weiser considers that written language was the first ubiquitous technology, because before it, information was restricted to people's memories. With his invention, anyone who knows how to read is able to understand what is written, therefore, independent of the memory of the person who wrote it.

The concept of computing everywhere is different from taking a notebook anywhere, because even at that point what you take with you is computational power and the focus therefore is on the computer. With ubiquitous computing, computers operate at a distance, with no physical contact with users. The interaction with these computers in the environment could be done by recognition of presence, voice and gestures by sensors installed in the environment, displays and projectors. Ubiquitous computing also implies more intelligence on the part of the computer, as its sensors would have to perceive what is happening in the environment and take actions to facilitate the task of users who are in it, activating services, without the user having required them. For example, when entering your office and looking for a certain document saved on paper, the system would point out where you have stored that document in the past. The system could also bring the project you were working on into a meeting room so it could be presented. Obviously, ubiquitous computing would present new challenges in terms of privacy and security, as the first example means that the system was watching your every step when you saved that document in the past, while the second means that the system would have access to all your files and would transfer only the files needed for the presentation. In addition to privacy and security, there are also other challenges, such as bringing together pieces of hardware and software from multiple manufacturers whose software would have to be integrated and driven by a larger system. Although there is no system that fully implements the idea of ubiquitous computing, some technologies try to approach this ideal, such as speaker systems that hear what is being said and perform tasks such as adjusting the lighting, playing a favorite song or perform an internet search. In agriculture, the concept of ubiquitous computing has been used in the application of agrochemicals. In this application, existing sensors close to the leaves would guide the electronics embedded in the sprayers in order to control the greatest possible coverage, using the least amount of liquid.

IoT – The internet of things (IoT) is defined by the International Telecommunication Union (ITU 2012) as a global infrastructure for information society, enabling advanced services through the interconnection of things (physical and virtual), based on interoperable information and communication technologies, whether these structures are in place or evolving. From the point of view of the internet of things, the ITU defines things as objects in the physical or virtual world that can be identified and integrated into communication networks. Virtual objects are included in the IoT through physical things linked to devices, which in turn have mandatory communication capabilities. Communication between devices can be performed through a communication network (with or without an intermediary gateway) or directly between devices, without a communication network, in the latter case, direct communication between devices is required. When communication between devices takes place through a gateway, it must provide at least two network technologies, either to integrate devices, such as ZigBee, Bluetooth, Wi-Fi or LoRa, or to integrate devices to the communication network, such as 2G, 3G, LTE, satellite networks or others. Devices also have the sleep mode and automatically return to save energy. This capability is especially important for sensors that are installed in remote locations that do not have direct connection to electricity, as is the case with some agricultural monitoring sensors. Objects connected to the IoT network can range from people or animals with RFID tags to pacemakers and other hospital devices for individual use, agricultural implements, cell phones, surveillance cameras, humidity and atmospheric pressure sensors, rain gauges, cars with embedded sensors and many others. All things connected to the IoT are required to have an internet address, that is, an IP address. With this address, things can be accessed by any machines connected to the internet. This access at any time makes things connected to the IoT vulnerable in two points: security and privacy. Concern with security creates the need to implement requirements to ensure the confidentiality and integrity of information, both in the data and in the services that process this data. The issue of privacy also needs to be supported by the IoT, as the data that travels through the IoT system can transit sensitive information linked to the owners or users of the connected things. Protecting the privacy of such data must take place during data transmission, aggregation, storage, processing and mining. In agriculture, RFID sensors have been used to identify and track animals in the field.

Robotics – The term robot comes from the word robota, which means servant in the Czech language. Josef Čapek proposed it to his brother Karel to be used in the fiction play Rossum's Universal Robots, published in 1920 (Szabolcsi, 2014). In this play, machines with human behavior and appearance perform work. Nowadays, robots take on various forms and functions. In manufacturing, they take the form of arms to perform repetitive tasks such as welding, or dangerous tasks such as decontamination in nuclear facilities. Military, agricultural and space exploration robots are often vehicles with wheels or wings. Robotics is a research area that combines efforts from multiple areas, such as computer engineering, information engineering, mechanical engineering, electronic engineering, biology, as well as in the social sciences as robots must assume behaviors suitable for human interaction. A robot's degree of autonomy can range from remote control to fully autonomous operation. Depending on the task or the degree of autonomy, the robot needs: to have computer vision to build a global representation of the environment it is in and the objects within the field of view; have a control system to perform the desired task, which may or may not include artificial intelligence; have actuators that will move the parts according to the control and implement a user interface. It may also need devices that implement the sense of touch, hearing and smell. There are several advanced robots today: the Asimo, developed by Honda, is one of the most evolved humanoid-looking robots. It can walk on uneven surfaces, talk to several people at the same time, open bottles and pour liquid into a glass, in addition to mastering several simultaneous conversations with different people. NASA created Robonaut2 and sent it to the International Space Station to help carry out dangerous or even mundane tasks. In agriculture, robots often take the form of an off-road vehicle, such as the See and Spray robot, developed by Blue River to detect weeds and selectively apply crop protection products, only on those plants, avoiding the planted crop.

Geospatial data – Also called geographic data, they belong to a particular class of data that describe facts, objects and phenomena of the terrestrial globe, associated with its location on the land surface, at a particular moment or period (Câmara et al., 1996). Geospatial data are fundamentally identified from others by their spatial component, which associates each entity or phenomenon with a location translated by a geodetic territorial reference system. Geotechnology is the name given to a special category of technologies used for the process of acquiring, visualizing, processing, analyzing and/or making available geospatial data. In this context, technologies such as remote sensing, the Global Positioning System (GPS), topography, Geographic Information Systems (GIS), geographic databases, among others, are classified as geotechnologies. When geospatial information is derived from one or more geotechnologies, it is called geoinformation. Finally, geoprocessing is the process of applying one or more geotechnologies to acquire, process, visualize, analyze and/or make available spatially referenced data, in order to generate geoinformation. Geospatial data are used in agriculture, for example, to monitor the crop of a particular commodity, in which a sequence of satellite images is analyzed over time in a given region to determine how much will be produced.

GIS – Geographic Information Systems (GIS) is one of the main technologies for visualization, analysis and treatment of geographic data. There are several definitions for GIS, from the most complex to the simplest. Pires et al. (1994) define GIS as a system that performs the computational treatment of geospatial data, storing, managing and retrieving information. These systems are widely used in decision environments, providing users with facilities to combine information from a given region. The main difference between GIS and a conventional information system is that GIS can store both the descriptive attributes of the data and the geometries of different types of geographic data. The following are the main characteristics of GIS: to input and integrate, in a single database, textual spatial information and other data sources, such as satellite images and GPS data; and offer mechanisms to combine the various information, through manipulation and analysis algorithms, as well as to consult, retrieve and visualize the contents of the geographic database. The approach traditionally used to organize geospatial data in a GIS is its distribution in layers, or information planes, where each layer addresses a different theme for a given geographic region. For example, a satellite image of a region is a layer, as well as the municipalities in that region, their geomorphology and their hydrology. Each layer is internally represented using logical structures specific to each GIS and is stored in different files, according to the format of the system used. In agriculture, GIS can be used to create a digital model of a rural property from the measurements made using GPS at various points on that property.

Converging technologies

Digital agriculture incorporates concepts that were originally developed for other areas, such as blockchain and cloud computing, which converge for the solution of agricultural problems. The reuse of these technologies came from the need to store data remotely, to better process the data, and also to meet a recurrent demand in agriculture, which is the traceability of its products and processes.

Blockchain – It is a type of distributed database with a storage model that allows permanent and inviolable record keeping. It is known worldwide as being the technology on which bitcoin cryptocurrency was developed, and its origin dates back to 2008, when its author, under the pseudonym Satoshi Nakamoto, published an article on the internet (Nakamoto, 2008) on the creation of a decentralized electronic payment system, secure and based on a peer-to-peer (p2p) network. Blockchain allows encoding the content of a variable-length message to fixed-length data via integrity and authentication protocols based on single-use ciphers, or one-way hashing, (Castro, 2017; Ethereum, 2019).

Each transaction can be understood as an action that can be traced, and which is certified by the network's nodes, and part or all of its content may be confidential. These transactions are grouped similar to a ledger, also used in accounting operations, and because of this characteristic, the set is called a ledger. Ledgers are the basis, within a framework of computational tools, for implementing transaction systems with blockchain technology in corporate environments.

Traceability systems via blockchain provide a secure and distributed way to provide information within an agricultural production chain, or any other agribusiness processes, allowing to track information such as the origin of the product and its inputs, the use of pesticides in farming, among others.

Cloud computing – It refers to a technology that allows to access programs, files and services through the internet, without the need to install programs or store data – hence the allusion to the "cloud". The term is generally used to describe data centers available to many users over the internet (Hayes, 2008). Once it is properly connected to the online service, it is possible to appreciate its tools and save all the work done and access it later, from anywhere, from any computer with internet access, regardless of the platform. The minimum requirement is a computer compatible with the resources available on the internet. For example, a personal computer becomes just a chip connected to the internet, which in this case would represent the "great cloud" of computers, requiring only input devices, keyboard, mouse and monitor.

Cloud computing can be understood as an infrastructure paradigm that allows establishing software as a service, a large set of web-based services, with the objective of providing features that, until then, required large investments in hardware and software, which works through a pay-as-you-go model (Buyya et al., 2009). A typical example of cloud computing is file sync services like Dropbox. When copying or moving a file to this space, it will be duplicated on the application server and on other computers that have the program installed and on which a user accesses his account.

Cloud computing offers several benefits, such as: a) cost reduction: either by reducing expenses with energy, no-break or generator, air conditioning and physical security of the equipment, or by purchasing software and hardware; b) saving space: from the moment the user connects/adheres to cloud services, the storage will be completely virtual; c) flexibility: the services are perfectly adaptable to the company's different needs. If this prediction is underestimated, it is easy to increase the service, readjusting it to the real demand; d) constant updating: technology advances and hardware quickly becomes outdated. When migrating to cloud computing, following technological development becomes a much less exhausting and costly task, since contracted services are constantly updated; e) storage capacity: the ability to backup a vast amount of data, instantly, is as important as the effortlessness of recovering this data at any time, at a considerably low cost; f) increased collaboration: by allowing multiple users to access the same file remotely, cloud computing encourages collaborative work. As updates are done in real time, the data exchange between members of the same team is much faster.

However, cloud storage can generate distrust, especially when it comes to security. After all, the proposal is to keep important information in a virtual environment, and not all companies and individuals are comfortable with this approach.

Final considerations

This chapter introduced the main concepts used in data management, processing and visualization of digital agriculture. It presented digital technologies linked to the organization and representation of information, mathematical and statistical modeling, artificial intelligence, sensors and robotics and convergent technologies such as cloud computing and blockchain. In the next chapters, these technologies are explored in the many applications, built by Embrapa Agricultural Informatics and its partners, in order to provide solutions for an increasingly dynamic and integrated agriculture, such as digital agriculture. As can be seen, based on the list of technologies conceptualized here, the tools used to solve agricultural problems are located at the frontier of technological knowledge.

References

AMARAL, F. **Introdução à ciência de dados**: mineração de dados e big data. Rio de Janeiro: Alta Brooks, 2016.

BASSANEZI, R. **Ensino-aprendizagem com modelagem matemática**. 4 ed. São Paulo: Contexto, 2002.

BISHOP, C. M. **Pattern recognition and machine learning**. Singapore: Springer Science+Business Media, 2006.

BREIMAN, L. Random forests. **Machine Learning**, v. 45, p. 5-32, 2001. DOI: 10.1023/A:1010933404324.

BUYYA, R.; YEO, C. S.; VENUGOPAL, S.; BROBERG, J.; BRANDIC, I. Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility. **Future Generation Computer Systems**, v. 25, n. 6, p. 599-616, June 2009. DOI: 10.1016/j.future.2008.12.001.

CÂMARA, G.; CASANOVA, M.; MEDEIROS, C. B.; MAGALHÃES, G.; HEMERLY, A. **Anatomia de Sistemas de Informação Geográfica**. Campinas: Ed. Unicamp, 1996. 193 p.

CASTRO, A. de. Quantum one-way permutation over the finite field of two elements. **Quantum Information Processing**, v. 16, article number 149, 2017. DOI: 10.1007/s11128-017-1599-6.

DEAN, J.; GHEMAWAT, S. MapReduce: simplified data processing on large clusters. **Communications of the ACM**, v. 51, n. 1, p. 107-113, 2008. DOI: 10.1145/1327452.1327492.

DEDIĆ, N.; STANIER, C. Towards differentiating business intelligence, big data, data analytics and knowledge discovery. In: PIAZOLO, F.; GEIST, V.; BREHM, L.; SCHMIDT, R. (ed.). **Innovations in enterprise information systems management and engineering**. Berlin; Heidelberg: Springer, 2017. p. 114-122. (Lecture Notes in Business Information Processing, n. 285). DOI: 10.1007/978-3-319-58801-8_10.

EMBRAPA AGRICULTURAL INFORMATICS. **AgroAPI**. Available at: https:// www.agroapi.cnptia.embrapa.br. Accessed on: 18 Jun. 2020.

ETHEREUM. 2019. Available at: https://www.ethereum.org. Accessed on: 22 May 2020.

FISHER, R. A. **Statistical methods for research workers**. 5th ed. Tweeddale Court: Oliver and Boyd, 1934.

FUKUNAGA, K. **Introduction to statistical pattern recognition**. 2nd ed. Boston: Academic Press, 1990. DOI: 10.1016/B978-0-08-047865-4.50007-7.

FUKUNAGA, K.; NARENDRA, P. M. A branch and bound algorithm for computing k-nearest neighbors. **IEEE Transactions on Computers**, v. C-24, n. 7, p. 750-753, July 1975. DOI: 10.1109/T-C.1975.224297.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. Cambridge: The MIT Press, 2016.

HAN, J.; KAMBER, M. **Data mining**: concepts and techniques. 2nd ed. San Francisco: Morgan Kaufmann, 2006. 770 p.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning**. New York: Springer, 2009. DOI: 10.1007/978-0-387-84858-7.

HAYES, B. Cloud computing. **Communication of the ACM**, v. 51, n. 7, p. 9-11, July 2008. DOI: 10.1145/1364782.1364786.

INTERNATIONAL TELECOMMUNICATION UNION. **ITU-T Y.2060**. Series y: global information infrastructure, internet protocol aspects and next-generation networks: next generation networks – frameworks and functional architecture models: overview of the Internet of things. 2012. Former ITU-T Y.2060 renumbered as ITU-T Y.4000 on 2016-02-05 without further modification and without being republished. Available at: https://www.itu.int/rec/T-REC-Y.2060-201206-I. Accessed on: 17 Apr. 2020.

KLIR, G. J.; YUAN, B. **Fuzzy sets and fuzzy logic**: theory and applications. Upper Saddle River: Prentice Hall, 1995. 574 p.

MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY, 5., 1967, Oakland. **Proceedings** [...]. Berkeley: University of California Press, 1967. v. 1, p. 281-297. Available at: https://projecteuclid.org/euclid.bsmsp/1200512992. Accessed on: 17 Apr. 2020.

MARBLESTONE, A. H.; WAYNE, G.; KORDING, K. P. Toward an Integration of deep learning and neuroscience. **Frontiers in Computational Neuroscience**, v. 10, n. 94, Sept. 2016. DOI: 10.3389/fncom.2016.00094.

NAKAMOTO, S. **Bitcoin**: a peer-to-peer electronic cash system. 2008. Available at: https://bitcoin.org/bitcoin.pdf. Accessed on: 22 Jan. 2020.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **ANSI/NISO Z39.19-2005 (R2010)**: guidelines for the construction, format, and management of monolingual controlled vocabularies. 2010. Available at: https://groups.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf. Accessed on: 18 Jun. 2020.

NOY, N. F.; MCGUINNESS, D. L. **Ontology development 101**: a guide to creating your first ontology. 2001. Available at: http://protege.stanford.edu/publications/ontology_development/ontology101.pdf. Accessed on: 17 Apr. 2020.

PIRES, M. F.; MEDEIROS, C. M. B.; SILVA, A. B. Modelling geographic information systems using an object oriented framework. In: BAEZA-YATES, R. (ed.). **Computer science 2**. Boston: Springer, 1994. DOI: 10.1007/978-1-4757-9805-0_18.

QUINLAN, J. R. Induction of decision trees. **Machine Learning**, v. 1, p. 81-106, 1986. DOI: 10.1007/BF00116251.

ROCO, M. C.; BAINBRIDGE, W. S. Overview converging technologies for improving human performance: nanotecnologia, biotecnologia, information technology and cognitive science. In: ROCO, M. C.; BAINBRIDGE, W. S. (ed.). **Converging technologies for improving human performance**. Dordrecht: Springer, 2003. p. 1-27. DOI: 10.1007/978-94-017-0359-8_1.

RUSSEL, S.; NORVIG, P. **Artificial intelligence**: a modern approach. 4th ed. New Jersey: Prentice Hall, 2020.

SNEDECOR, G. W.; COCHRAN, W. G. **Statistical methods**. 6th ed. Ames: The Iowa State University Press, 1967. 286 p.

SZABOLCSI, R. The birth of the term Robot. **Advances in Military Technology**, v. 9, n. 1, jun. 2014.

TORRES, N. V.; SANTOS, G. The (mathematical) modeling process in biosciences. **Frontiers in Genetics**, v. 6, n. 354, Dec. 2015. DOI: 10.3389/fgene.2015.00354.

VAPNIK, V. N. **Statistical learning theory**. New York: John Wiley and Sons, 1998.

VAPNIK, V. N. **The nature of statistical learning theory**. New York: Springer-Verlag, 1995. DOI: 10.1007/978-1-4757-2440-0.

WEISER, M. The computer of the 21st Century. **Scientific American**, v. 265, n. 3, Sept. 1991. DOI: 10.1038/scientificamerican0991-94.

WHITE, T. **Hadoop**: the definitive guide. Sebastopol: O'Reilly Media, 2012.

WU, X.; KUMAR, V.; QUINLAN, J. R.; GHOSH, J.; YANG, Q.; MOTODA, H.; MCLACHLAN, G. J.; NG, A.; LIU, B.; YU, P. S.; ZHOU, Z.-H.; STEINBACH, M.; HAND, D. J.; STEINBERG, D. Top 10 algorithms in data mining. **Knowledge and Information Systems**, v. 14, n. 1, p. 1-37, Jan. 2008. DOI: 10.1007/s10115-007-0114-2.

ZADEH, L. A. Fuzzy sets. **Information and Control**, v. 8, n. 3, p. 338-353, June 1965. DOI: 10.1016/S0019-9958(65)90241-X.

Photo: tostphoto (AdobeStock)

3

Agroenvironmental modeling and the digital transformation of agriculture

Santiago Vianna Cuadra | Daniel de Castro Victoria | Giampaolo Queiroz Pellegrino | Édson Luis Bolfe | José Eduardo Boffino de Almeida Monteiro | Eduardo Delgado Assad | Aryeverton Fortes de Oliveira | Maria do Carmo Ramos Fasiaben | Geraldo Bueno Martha Júnior | Mateus Batistella | Luís Gustavo Barioni | Alan Massaru Nakai | Fábio César da Silva | Marília Ieda da Silveira Folegatti Matsuura

Introduction

Agriculture is an activity with significant returns and risks. Therefore, production processes and businesses are increasingly benefiting from the availability and use of information. Technical, economic, social, and environmental dimensions are of interest to all agribusiness segments, which seek reliable indicators for operations and business. Financing, inputs, sales and insurance contracts, certifications and regulatory processes come to depend on the development of intelligent systems and information at reasonable costs. These become essential for competitiveness and sustainability from a viewpoint of strengthening the value chains in agriculture.

It is undeniable, for example, that the upheavals resulting from climatic instability, sanitary events, and fluctuations in markets or business environments stand out due to their impacts on agriculture, and the related data and information must be investigated in an interdisciplinary manner. A study supported by the World Bank indicates that Brazil loses more than R$ 11 billion annually due to risks that could be minimized (Arias et al., 2015). In many regions, more than 60% of the variability and risk of agricultural production is caused by climatic effects. Farmers exert little or no control over natural phenomena such as drought, frost, heat waves, windstorms, or hail (Rossetti, 1998), and to bypass such occurrences, agricultural producers depend on information derived from more complex analytical processes.

It is essential to understand and quantify the climate risks of different ecoregions[1] in order to reduce risks of exposure and ensure greater agroecosystem resilience[2] in Brazil. This task is complex, given the continental dimension of the country, the diversity of agricultural crops, production systems, the supply of natural resources, soil conditions, relief and climate. It requires specialized knowledge about the functioning of agroecosystems and the correct allocation of competences and resources, especially from Embrapa, considering its mission, dimension, and capillarity. Science advances the understanding of these complex biophysical and economic systems with the intensive use of data associated to robust analytical systems. As such, the formulation of policies, the creation of incentives, and regulation of economic activities progressively incorporated these advances and consequently came to depend on new information processes. Measures of risks and returns, as well as productive, economic, social, and environmental impacts have become an essential aspect of the innovation that Embrapa promotes in digital agriculture, also called Agriculture 4.0.

Dealing with this complexity requires the ability to process a large set of data and information in order to generate knowledge and support decision-making that is more informed and, consequently, more accurate.

Therefore, a large investment in information technology is necessary in order to guarantee the collection of basic and primary data, using field and remote sensors; storing, organizing, accessing, and interoperating multiple databasesas. Lastly, but key to the entire process is the adaptation and development of powerful processing tools for analyzing large data volumes. Thus, information and knowledge are generated and disseminated in an appropriate and understandable format for various audiences. The advancements in information and communication technologies (ICT) can contribute to developing knowledge, public policies, and investments in a highly technified agriculture backed by scientific knowledge, regardless of scale.

Considering this scenario, Embrapa Digital Agriculture invests in solutions by the development and application of agroenvironmental models. The Research Group on Agroenvironmental Modeling was established, and professionals with interdisciplinary skills have contributed to the development of research, processes, products, and services related to understanding and quantifying soil-plant-animal-atmosphere interactions. Another strategic dimension of the Unit's performance is to foster synergy and networking with other Embrapa Units, universities, national, and international research institutions. Agroenvironment modeling no only generates spatially explicit data and information that supports the stability and increase in agricultural productivity, but also decreases the use of natural resources, helping to achieve the Sustainable Development Goals (SDGs) for 2030 (United Nations, 2015) – especially in SDG 02, Zero Hunger; in SDG 12, Responsible Consumption and Production; and in SDG 13, Climate Action.

This chapter presents a brief history of agro-environmental modeling and its importance as an essential element in the digital transformation of agriculture, especially with regard to rural planning and strategic/managerial decision-making in the public and private sectors. Some of the generated knowledge and products are summarized here, which highlight the direct contributions of Embrapa Digital Agriculture. Also noteworthy is the support that the Unit offers to the challenges and risks posed by historical variability and climate change, as well as understanding the synergies between production and ecosystem services.

[1] Resilience, for agriculture, can be understood as the capacity of production systems to coexist with variability and risks, whether through the best selection of planting times, cultivars and the use of technologies to combat adversities (irrigation, use of pesticides etc.) or financial mechanisms to absorb shocks caused by adverse effects.

[2] Ecoregions are geographic units with similar physical and biological characteristics, whose boundaries are defined based on abiotic characteristics (altitude, relief, soil, geology, precipitation, flood cycle, tidal effects) and biotic (groups of plants and animals present).

The evolution of agroenvironmental modelling

Modeling of agricultural systems began to flourish in the mid-twentieth century, with the contributions of linear programming and economic modeling of agricultural systems as important milestones. At that time, the models explored the components of the agricultural system and their interactions in a more limited manner. From the 1960s onwards, many models were developed with a focus on forecasting and evaluating agricultural system performance in response to changes in their components and interactions (Jones et al., 2017).

In the 1980s and 1990s, modeling of agricultural systems started to be used progressively so as to expand our knowledge on fundamental aspects of the functioning of the soil-plant-animal-atmosphere system. It enabled to simulate variations in the state of the system, along with its components and their interactions in different spatiotemporal scales. This evolution allowed to incorporate physical processes (such as the balance of radiation, absorption by vegetation, and soil water movement), biophysical (such as photosynthesis and plant growth processes), biogeochemical (such as the soil carbon cycle) within a single numerical simulation model. At the same time, the development of numerical models by agrometeorology community had grown significantly, with the emergence of various software that included models that simulate the development and growth of agricultural crops. Based on these analyses, the models began to provide important elements to support the prediction of potential responses of the agricultural system to changes in the conditions of the environment, the production system and the management employed. Therefore, agroenvironmental modeling stands out as a tool to assess the responses of agricultural productivity to climatic conditions using empirical statistical models and models based on biophysical and socioeconomic processes, which simulate agricultural productivity and its interactions with the environment and management practices (Jones et al., 2017). The models used utilize mathematical formulations to represent the functioning of natural systems, be it simulating the growth of a plant in specific meteorological conditions, the need for food supplementation for cattle or another process to be evaluated.

The expanded knowledge brought about by modeling on the multiple facets of agricultural systems and their components has percolated more extensively into decision-making on rural properties and with decision makers in the political sphere. Today these models are now widely used in precision agriculture, irrigation management, soil fertility management, plant breeding, monitoring and forecasting of yield for crop management, agricultural insurance, impact assessment and subsequent adaptation to climate change., quantification of carbon sequestration in the soil, environmental impacts in land cover and land use changes, forecast of agricultural crop production, and disease risk assessments.

Agroenvironmental modeling products to support decision-making

Several technologies and products that make use of agro-environmental modeling applied to rural planning have been developed, directly or indirectly, by the Agroenvironmental Modeling Research Group, with the main lines of action presented below. These works can be grouped into four central areas: i) obtaining, organizing, storing and distributing basic data for agroenvironmental modeling; ii) quantification and analysis of climate risks and resilience of agricultural systems; iii) products that support territorial planning; and iv) integration of socio-economic analysis in agro-environmental modeling. Although these components are presented separately in order to be evidenced and specified, they complement each other, and at various times merge during analysis, modeling, and simulation development. This generation of knowledge and products are applied to rural planning in an integrated approach.

Databases for agricultural and environmental research

One of the prerequisites for using models is the collection, storage, use, and distribution of agrometeorological data. These must be suitable for their specific purpose with the appropriate spatial and temporal coverage, as well as having a known quality. Climatic data are essential for agro-environmental modeling. Thus, tools are developed for the acquisition, storage, processing, and availability of agrometeorological data for Brazil.

Agritempo

Agritempo[3], currently available on the portal and the Agritempo mobile and Agritempo GIS mobile applications has been offering free agrometeorological data that supports agricultural activities in rural properties, reducing risks related to climate and weather while also supporting public policies and allowing online actions for agrometeorological monitoring.

The main innovation offered by the system refers to the automation of tasks, enabled using ICT, in which the entire process of receiving data, incorporating it into the database, and building maps occurs automatically through the system without human intervention. This provides greater speed and accuracy, in addition to offering better quality to the database itself as the system automatically performs tests on the collected variables (Alencar et al., 2016).

Agritempo mobilizes a collaborative network with 40 institutions, involving the exchange of meteorological data, research in agrometheorology, generation of new technologies such as system modules, functionalities, and availability of information, such as studies and scientific publications. The system organizes and manages a set of more than 1,600 meteorological stations, a number that is constantly expanding. It also contains a database of at least ten years of satellite images that can be used to support research in agrometeorology.

In 2019 alone, the Agritempo portal (Figure 1) had 180,950 hits.

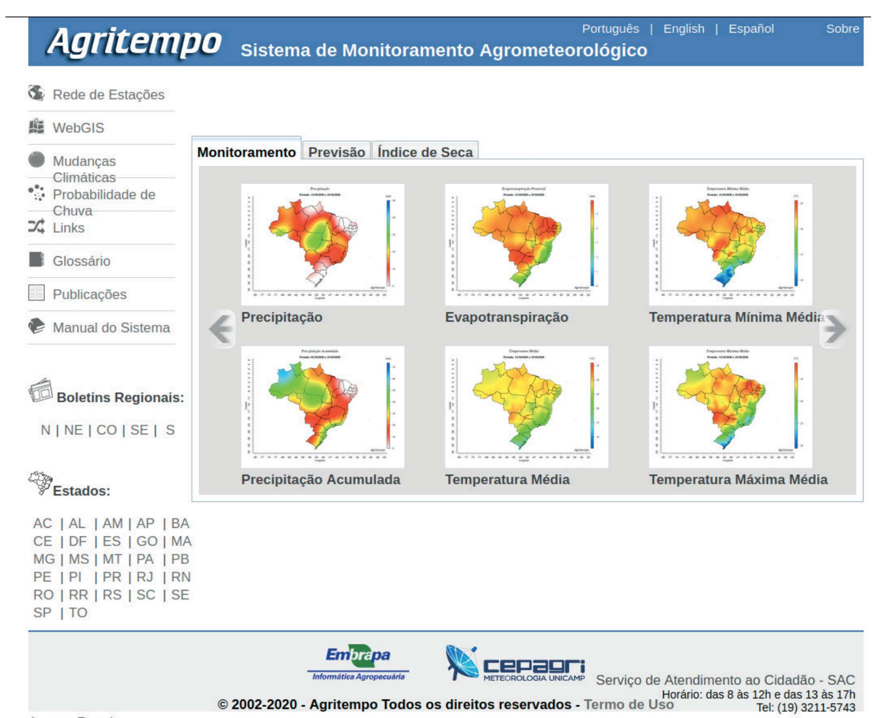


Figure 1. Interface of the Agritempo system – version 2.0.

Source: Agritempo (2020).

[3] Available at: http://www.agritempo.br

The Agritempo mobile and Agritempo GIS mobile applications recorded more than 10,000 installations, showing the demand and transfer of information contained in these systems.

Conprees

Currently, Embrapa Digital Agriculture has meteorological data integration and remote sensing systems from various sources. However, the low density of meteorological stations in vast regions of the country, in addition to inherent flaws in the operation and maintenance of existing stations, results in the lack of reliable data for many producing regions. More recently, CONPREES, an acronym for Consistent, Completed, and Spatialized meteorological data, will enter the testing and operation phase in 2020. This system uses a much larger number of data sources ranging from public and private meteorological stations and from different meteorological and remote sensing models. Thus, it is possible to develop a database with sufficient resolution and accuracy to monitor the occurrence of adverse events and agrometeorological accidents.

The integration of agrometeorological monitoring systems and remote sensing vegetation cover, such as System for Temporal Analysis of Vegetation (SATVeg) (detailed in chapter 4), will provide information on plant biomass that will allow for more effective monitoring of the areas. While systematic agrometeorological monitoring allows identifying unfavorable conditions (low temperatures, dry spells, water deficit, etc.), vegetation indices obtained from satellite images can indicate conditions related to management, such as planting time, planted area and vegetative vigor. In addition, vegetation indices synthesize the vigor of vegetation during the development process of an agricultural crop. As such, they also reflect the agrometeorological conditions in a cultivated area when analyzed through the harvest period.

Risk assessments and climate resilience evaluation

Climate is the main environmental factor associated with agricultural productivity variability. Climatic risks have the potential to cause significant or total losses to production, divided into two groups: i) extreme event (low and high temperatures, very intense rain, strong winds, among others); ii) cumulative event (prolonged droughts, temperatures limiting growth for long periods, etc.).

The adoption of good agricultural management practices is considered one of the most viable means for adding resilience to the production system. This occurs mainly by reducing exposure to climate risks, but it is also possible to reduce current productivity gaps. In this context of serving rural and agricultural planning, agroenvironmental modeling has been used both in the history of meteorological phenomena and in developing the productivity of agricultural crops. This risk assessment focus and promotion of climate resilience is leading to the simulation of future agricultural scenarios derived from climate change. As such, the use of modeling can guide both research priority and development themes or practices to be adopted and intensified in production systems. The guiding base for this development is the greater risk and vulnerability, or a pronounced reduction in adaptive capacity of a particular crop.

Agricultural Climate Risk Zoning

The Agricultural Climate Risk Zoning (ZARC), governed by the National Program of Presidential Decree No. 9,841/2019 is one of the most important examples of the agroenvironmental modeling application aimed at generating value with broad social, economic, and environmental benefits. It is an agrometeorological product that delimits regions and planting times according to the probabilities of production loss caused by adverse weather events. These studies are based on a broad knowledge base, agronomic, and meteorological data, as well as modeling techniques combined with large-scale processing systems. This allows for the generation and analysis of different scenarios per crop, enabling an assessment of climate impacts associated with the hydro-physical characteristics of the soil, crop cycles, planting dates, and crop sensitivity to climate effects at different plant stages, among others (Santos; Martins, 2016). The results translate into risk levels per ten days of planting for each municipality.

The objective of ZARC is to provide information for the management of climatic risks on rural properties, as well as for public managers (Ministry of Agriculture and Livestock – MAPA), Central Bank of Brazil – BACEN, National Monetary Council – CMN). Additional to this is the decision-making support in the Federal Government's insurance programs, such as the Agricultural Activity Guarantee Program (PROAGRO) and the Rural Insurance Premium Subsidy Program (PSR). This information is used to avoid excessive losses with compensation in areas or times of high agricultural risk, as well as to evaluate solutions for production systems less susceptible to climatic adversities. In order to be entitled to PROAGRO or PSR, and thus have access to Rural Credit, the producer must observe the ZARC recommendations. In addition, several financial agents in the private sector condition the granting of rural credit to the ZARC indicators.

Embrapa Digital Agriculture houses the processing cloud infrastructure, with large servers dedicated to ZARC. They use a workflow management system to automate the pre-processing (storage and processing of data used by agrometeorological models), processing (execution of simulations) and post-processing steps (probability calculations, results visualization and processing the final results), ending with its delivery to the Agricultural Policy Secretariat (SPA-MAPA).

The generated results are stored in a database, available to different teams in Brazil through Micura (Figure 2). This system enables the visualization and analysis of the ZARC results, allowing the technical teams to carry out validations and to investigate better ways to parameterize the models. After verification by the teams of each analyzed agricultural crop, the results are presented to the broad public at validation meetings in several states in Brazil, through Micura, with the participation of rural producers, technicians, agronomists, researchers from different institutions, public managers, financing and insurance agents, and cooperatives. If the results are not approved, the team identifies and proposes the necessary adjustments. New scenarios are then processed and presented, until results that are more coherent with the reality in the field are obtained.

The results are sent to MAPA to determine the recommended planting windows for more than 44 agricultural crops. These recommendations are the basis of insurance programs tied to the Federal Government, such as the PROAGRO, the Family Farming Agricultural Activity Guarantee Program (PROAGRO MAIS) and the PSR.

Plantio Certo (Sure Sowing)

As these are tools aimed for academic use, the results of the models often lack adequate interpretation. Thus, in order to facilitate access to ZARC indications, the Plantio Certo (Sure Sowing) mobile application was developed. It is available at Embrapa's app store[4]. With it, farmers, bank agents and people linked to rural insurance are able to consult the planting periods recommended by ZARC for different agricultural

[4] Available at: www.embrapa. br/applications

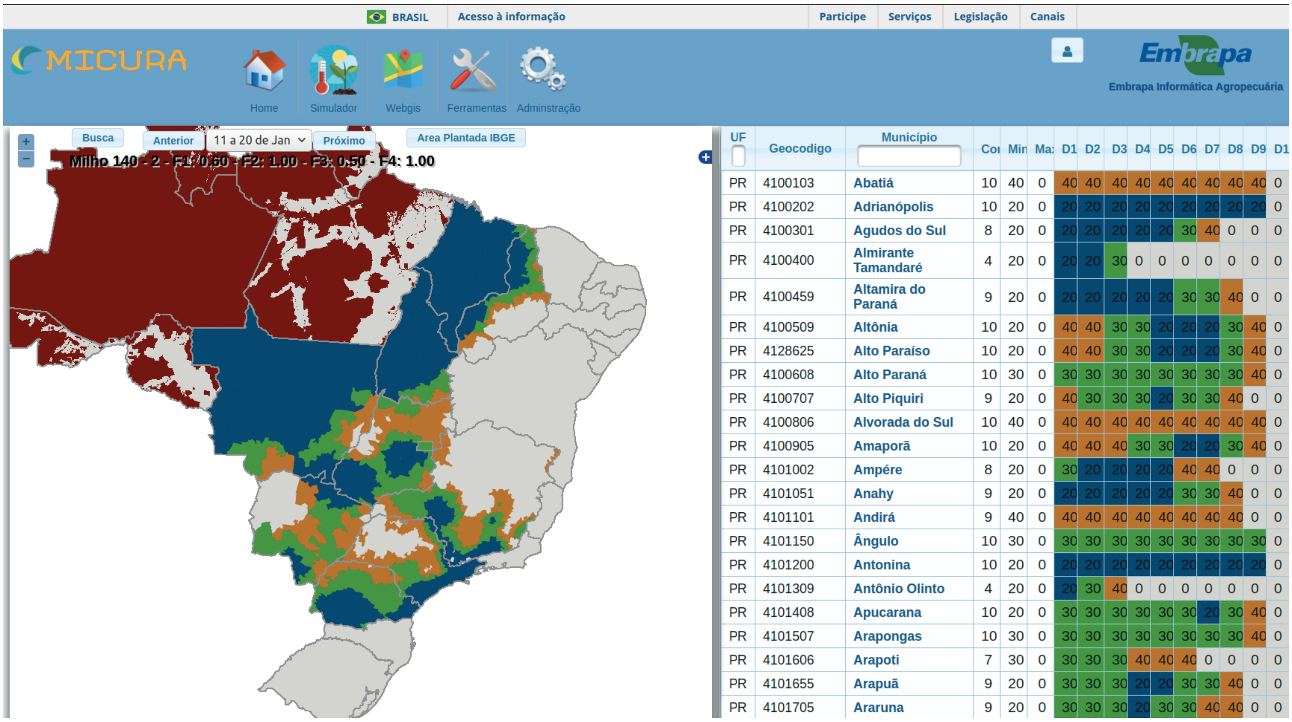


Figure 2. Vizualization of scenarios in the ZARC on the Micura system.

Source: Micura (2020).

crops in all Brazilian municipalities in a simplified way (Figure 3). The application also allows climate monitoring from the informed sowing date, returning information to the user on soil water storage, accumulated precipitation, number of days without rain, as well as minimum and maximum temperatures.

Figure 3. Interface of the Plantio Certo application. The example shows recommendations for cotton planting time in the municipality of Rio Verde, Goiás state, and agrometeorological monitoring of conditions at the location.

Source: Embrapa (2020).

Support for agricultural planning and monitoring

The availability of data from sensors located in the field or from remote platforms, along with time series obtained throughout the season and accessible as soon as they are collected, provides opportunities to monitor crops in real time and improve their management. This can be done directly from raw data or from derived indices such as rainfall measurements or vegetation indices (e.g., SATVeg, described in Chapter 4), or from more complex model results that make use of such data. The use of modeling can be done directly as the data are fed as input parameters in the models (e.g., temperature and precipitation data in a water balance model). This can also be done in the form of data assimilation, in that the results of a model are corrected throughout its execution, as new field observations are incorporated.

Helping in the planning of agricultural activities and monitoring field conditions are of interest to both producers and managers in various sectors, such as agricultural insurance, retail, processing industry, government agencies, among others. Embrapa has been working in this area with tools such as Agritempo and the Plantio Certo mobile application, SATVeg, Invernada, WebAgritec, in addition to other technologies under development.

Invernada

Invernada[5] is a support system for planning the production of beef cattle. It incorporates a database of climate and nutritional composition of pastures and supplementary foods. It also has dynamic models of pasture growth as a function of climate and soil water content, and able to estimate the seasonal distribution of forage production. In addition, it takes into account the selectivity of grazing animals, the growth of animals, and whether or not nutritional demands are met. Invernada incorporates algorithms for formulating diets with several optimization alternatives which are used in different aspects of decision-making, from pasture performance to animal management, and nutrition strategies. Additionally, it allows analyzing and comparing different management scenarios.

WebAgritec

WebAgritec makes several systems developed at Embrapa Agricultural Informatics available in the form of a website or a set of Application Programing Interface (APIs). APIs provide a set of functions and procedures that allow other software applications, internal or third-party, to access resources, data, and functionalities. The system is entirely developed by Embrapa Agricultural Informatics (detailed in chapter 12). The main modules available to users are: agricultural zoning, weather forecasting, crop selection, fertilizer and liming recommendation, identification of plant diseases, estimation of achievable productivity for soybean, corn, rice, bean and wheat, in addition to ancillary support with videos and recommendations. The productivity module uses meteorological data from the current crop as input into specific calibrated models. Thus, WebAgritec's main objective is to support rural extension services in Brazil, being used by institutions such as technical assistance and rural extension companies EMATER-GO and EMATER-MG. In addition, considering the functions of penalizing productivity, the system has been used by the National Supply Company (CONAB-MAPA) to assist in crop forecasting.

Climate change impact assessments and agricultural adaptation based on agroenvironmental models

By incorporating future climate projections into agro-environmental models it is possible to assess the impact of climate change on agricultural crops. These are obtained from projected scenarios using climate models, for example, these projections can be applied to ZARC models in order to assess whether a given crop will have more or less low-risk areas.

[5] Available at: www.invernada.cnptia.embrapa.br

Current and projected impacts for the coming decades are generally derived from trends observed in the present, which can be, for example, derived from observed climate series. These projections are extremely important for rural producers and for the territorial planning of production. On the other hand, projections of the impacts of climate change in the long term are extremely important for defining public policies, foreseeing trends and for the planning of adaptation and mitigation actions.

Adaptation generally seeks to reduce exposure to projected risks or increase the resilience capacity of production systems. The main objective of mitigation is adopting agricultural practices and crops that reduce emissions or increase carbon sequestration in agricultural systems. In agriculture, as a rule, good agricultural practices that promote resilience also have the co-benefit of mitigating climate change, either by emissions reduction and/or the improvement of the carbon balance in the agricultural system. In this context, Embrapa Agricultural Informatics has made a great contribution, through its leadership in Embrapa's Climate Change Portfolio, in the preparation of data and the execution of simulations. The results obtained have been used as the basis for important public policies, such as the definition of the strategic lines of the ABC Plan and the National Adaptation Plan – Agriculture Sector.

Climatic projections

In order to assess the impact of climate change on agriculture, the first step is to process and analyze climate projections. These are carried out by the Intergovernmental Panel on Climate Change (IPCC) at the global level and by the Brazilian Panel on Climate Change (PBMC) in the national context.

The processing and treatment of climate projections is a complex task that demands high processing and data storage capacity. Currently, the climate scenarios provided by the IPCC are in their sixth version. The Embrapa Agricultural Informatics team has traditionally made these data available for studies on the impacts of climate change. This step involves not only data processing, but also analyses and distribution to be used in agro-environment models. Figure 4 shows, as an example, the expected changes to the water balance, considering the results of climate projections for 76 different climate models.

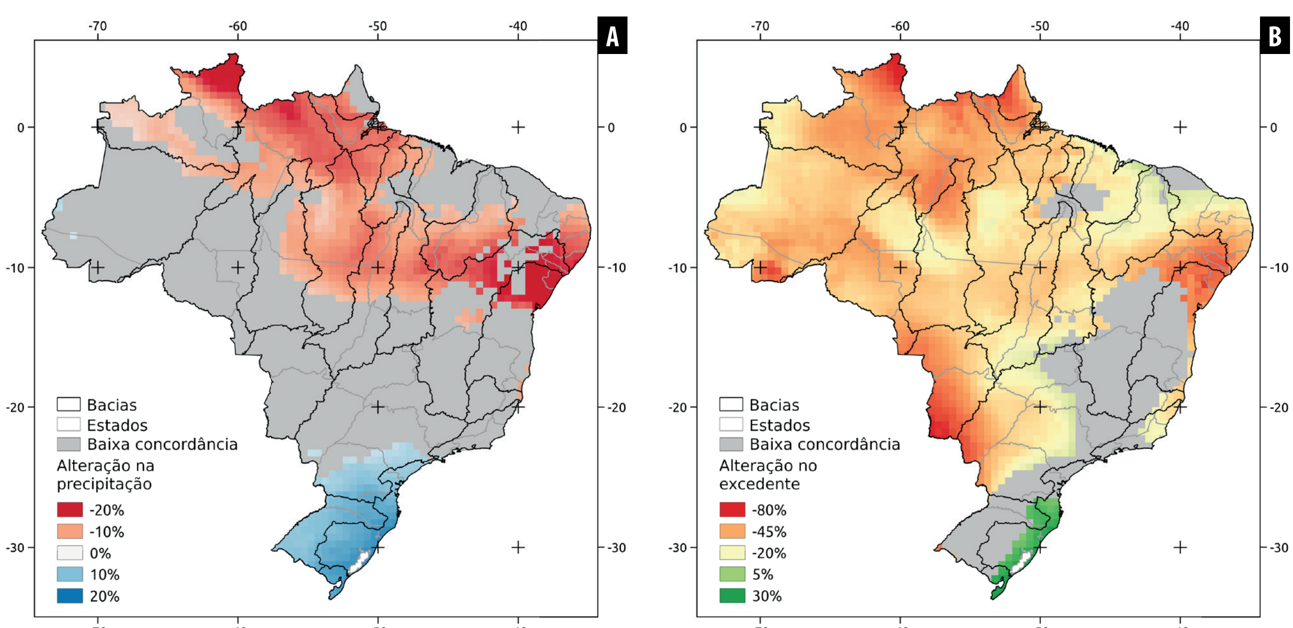


Figure 4. Change in annual precipitation (A) and water surplus (B), medians from 76 global climate model projections. Gray areas indicate less than 2/3 of agreement between the models.

Agricultural impacts

Several studies in Brazil elucidate that the materialization of climate change scenarios will severely impact Brazilian agriculture (Assad et al., 2016). Unlike high latitude regions, tropical regions, exporters of agricultural commodities are likely to experience more severe impacts of climate change on crop yields (Stevanović et al., 2016). The increase in global average temperatures could increase the occurrence of thermal and water stresses and, as a consequence, decrease productivity (Zhao et al., 2017). It is estimated that climate change is already reducing global agricultural production by 1 to 5% per decade over the past 30 years, and that it will continue to pose challenges in the coming decades (Challinor et al., 2014).

Embrapa Agricultural Informatics has actively contributed to the understanding, quantification, and proposal of adaptation measures against climate change. As an example Embrapa led the development of SCenAgri, which has already been used as a modeling basis in some studies on the subject (Assad; Pinto, 2008).

Simulation of future agricultural scenarios

The Agricultural Scenario Simulator (SCenAgri) is a system created to provide high-performance computing to support researchers working with climate change impacts on Brazilian agriculture. The system was developed based on the Bipzon model (Assad, 1986), which allows the simulation of future agricultural scenarios using data from different regionalized climate projection models. Its database includes more than 3,000 rainfall stations with daily data for at least 30 years, and is prepared to simulate climate risks for 20 annual and perennial crops. Several models of future climate projections are also incorporated into the simulator and made available by the Coupled Model Intercomparison Project Phases 5 and 6 (CMIP5 and CMIP6).

Territorial planning and land use

When considering rural planning, it is essential to understand aspects of the territory and its biophysical and human interrelationships. The integration and articulation between different territory scales and actors is complex and sensitive, requiring a series of preliminary considerations with implications at the local, municipal, and regional levels. It must be taken into account that rural planning has advanced beyond agricultural planning and needs to integrate elements from different disciplinary domains. For example, consider not just soil conservation, irrigation, and drainage, but the allocation of water resources and integrated watershed management, involving urban and rural populations. The existence of conflicting interests must also be recognized, as well as developing processes to deal with them.

Considering this scenario, it is essential to obtain social, economic, and environmental data, as well as additional information to support decision-making on land use at different territory scales. Studies and application of territorially based models allow to better understand the processes of expansion, retraction, transition, conversion, and agricultural intensification. These can support public policies associated with climate change and sustainable rural development in Brazil (Bolfe et al., 2016).

By combining agro-environmental modeling in a Geographic Information System (GIS) environment with the use of geographic data and remote sensing, it is possible to support territorial planning based on more complex spatial analyses, integrating information on soil, climate, vegetation, agriculture, water and socioeconomic resources. As an example, we can highlight the generation of models and simulations associated with the potential of land use, the analyses and projections on agricultural dynamics, agro-ecological and ecological-economic zoning, risk assessments and climate resilience.

As in most of the products and tools presented in this chapter, this spatial and temporal information supports managerial decision-making in public policies and private actions in the rural planning of properties, micro-watersheds, municipalities, states or biomes. Therefore, they benefit greater productive diversification and a more sustainable use of natural resourses in rural areas.

Agroideal

The Agroideal system, available at agroideal.org, developed by The Nature Conservancy (TNC), in partnership with Embrapa Agricultural Informatics and commercial companies in the agricultural sector, represents an example of the application of agri-environmental models in the decision-making of territorial occupation. Agroideal brings together information on logistics (location of storage silos), socioeconomic (occurrence of land conflicts), and environmental legislation (location of conservation units) with information on crop growth models (attainable productivity of soy). As such, decision makers can assess their operating strategy, identifying risks and opportunities in different regions of Brazil. The estimate of attainable soybean productivity (Figure 5) was only possible due to the organization of a large climatic database covering the entire territory along with a database of physical and water characteristics of Brazilian soils. These were then used in an agricultural crop growth model.

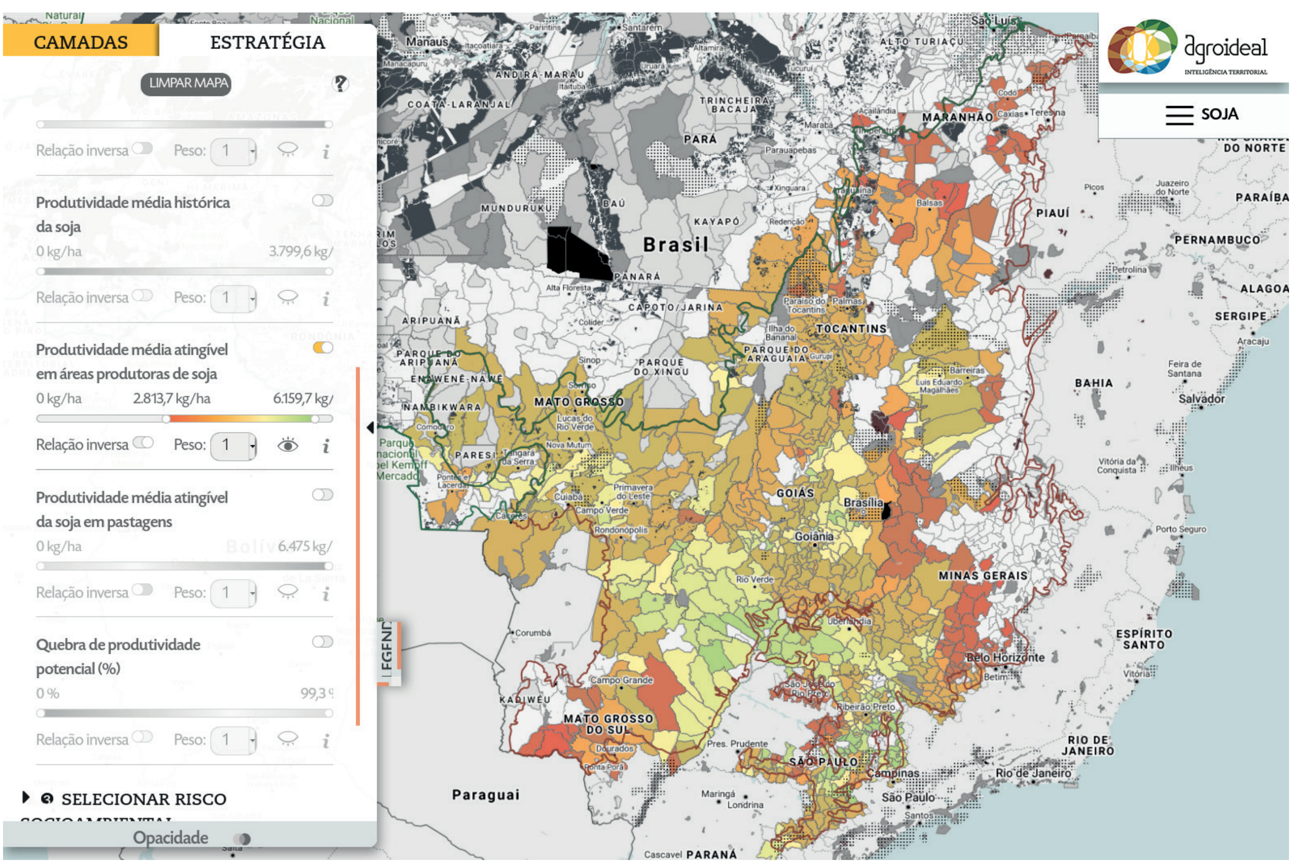


Figure 5. Potential soy productivity average in municipalities within the Cerrado biome (limited to productivity above 3,000 kg ha^{-1}).

Source: Agroideal (2020).

DINACER

Agricultural Dynamics in the Cerrado (DINACER) is another example of a database and information to support public and private decision-making in agricultural development on geospatial land use bases (Bolfe et al., 2020). This biome has strategic importance for the country's interests regarding food security, environmentally sustainable agriculture, and the preservation of biodiversity. DINACER, is carried out in collaboration with the National Institute for Space Research (INPE) and the Institute for Applied Economic Research (IPEA). They analyzed edaphoclimatic and vegetational aspects, public policies, research, innovations, technical assistance, agricultural dynamics, productivity, climate change, projections on the potential for expansion, and agricultural diversification of the biome. The analyses considered, whenever possible, the period corresponding to the past four decades up to 20-year projections. One of the analyses assessed the potential for expansion of agriculture in areas occupied by cultivated pastures (Figure 6). They identified that 44.5 million hectares of pastures present climatic and relief characteristics similar to the areas currently occupied by annual rainfed agriculture (Victoria et al., 2020).

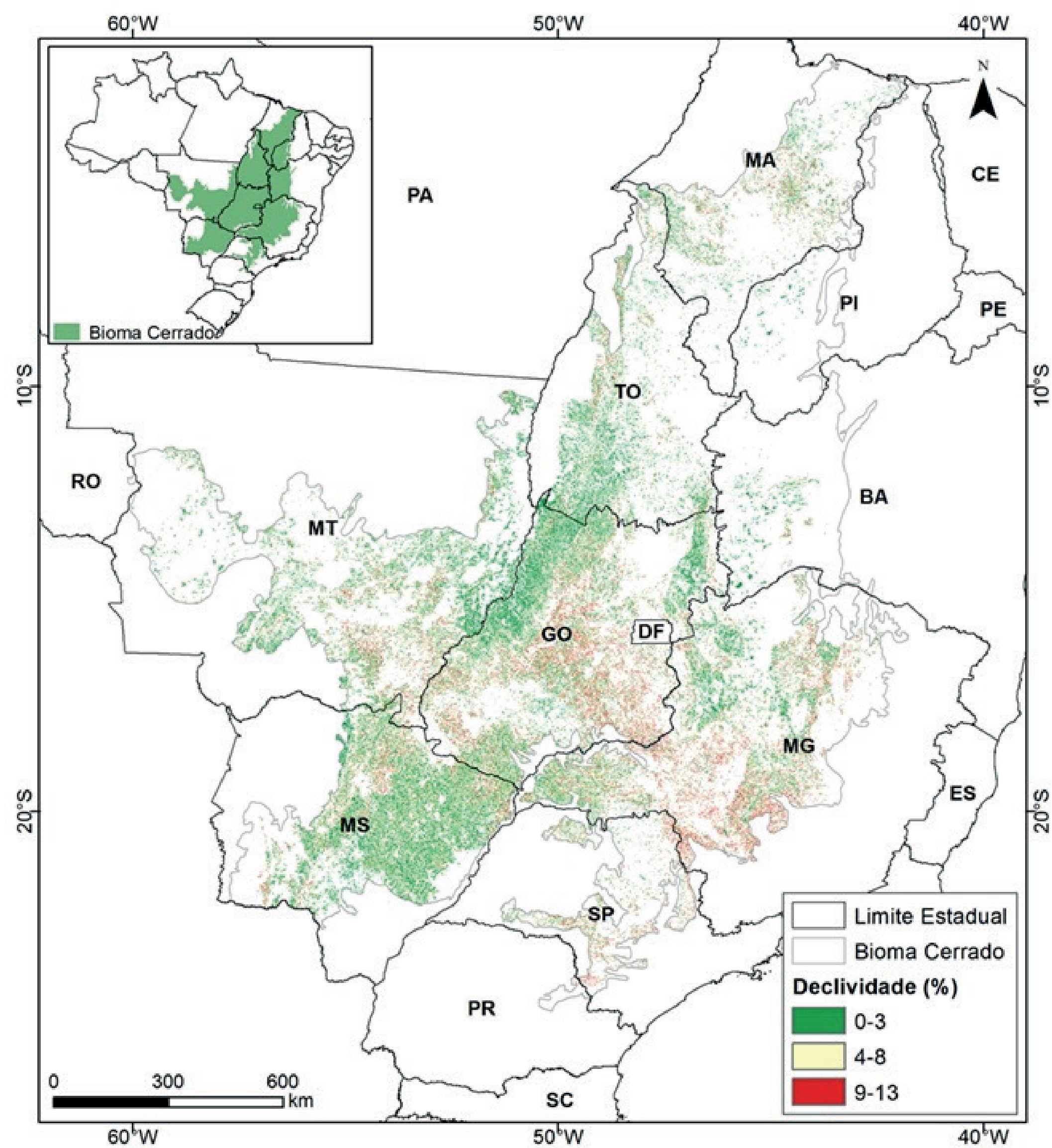


Figure 6. Pastures areas in the Cerrado biome with potential for annual agriculture according to water balance and broken down into slope ranges.

Source: Victoria et al. (2020).

Applications of agroenvironmental models for the conservation of ecosystem services

Agroenvironmental models are a powerful tool in the planning of a technified and digital agriculture based on robust knowledge about the functioning of agroecosystems. They consider other aspects in the relationship between agriculture and its means of production. These go beyond the generation of food, fibers and energy and analyze other benefits, such as the impacts of agricultural systems on the maintenance of water regulation (through the assessment of the water footprint), climate regulation, and other ecosystem services enjoyed by society.

Ecosystem services are the benefits directly and indirectly appropriated by man from the functioning of healthy ecosystems. Their importance for the economic system and for the well-being of future generations is increasingly recognized, as they provide essential goods (such as food) and services (such as the assimilation of waste).

According to Costanza et al. (1997), examples of ecosystem services, among many others, consist of: the carbon and nutrient cycle, the water cycle, soil formation, erosion control, climate regulation, conservation and evolution of biodiversity, the concentration of minerals, the dispersion or assimilation of contaminants, and the various usable forms of energy. The authors estimated the annual value of global flow from 17 services in 16 types of ecosystems. The results show that planet Earth's natural capital would annually yield an estimated average flow of US$ 33 trillion per year, about 1.8 times the gross world product at the time (US$ 18 trillion), at 1994 prices.

The Millennium Ecosystem Assessment (2005), conducted between 2001 and 2005, provided scientific bases for the sustainable management of ecosystems, allowing the continuous provision of the services they generate. This work demonstrates the recognition, by the international community, of the need and urgency of innovative measures to protect ecosystems, measuring their preservation for economic development (Andrade; Fasiaben, 2009).

Despite the importance of ecosystem services, these are currently not considered in economic transactions, as they are considered "free" or "gifts" from nature. The fact that they are not priced like other goods or services means that there are no incentives for their preservation, leading to overexploitation and often to total loss (Andrade; Fasiaben, 2009). However, such services and the stocks of natural capital that produce them are critical to supporting life on Earth. They contribute to human well-being, and therefore represent part of the planet's total economic value.

As natural capital and ecosystem services become more overexploited and scarcer, their value is expected to increase. Thus, studies related to its conservation are fully justified, in order to guarantee the provision of ecosystem services and support the formulation of policies that move in this direction. In this regard, agro-environmental modeling provides tools to support decision-making by public and private agents.

WebAmbiente

WebAmbiente[6] is an interactive system intended to facilitate the storage and search for information on technological solutions. These are involved in the use, recovery, and restoration of environments in legal reserves and areas for permanent preservation in all six biomes. The system was developed

[6] Available at: www.webambiente.gov.bz

by Embrapa and the Secretariat for Extractivism and Sustainable Rural Development of the Ministry of the Environment (MMA), in cooperation with specialists from partner institutions. It provides technical assistance and rural extension (ATER) multiplying agents with an information set aimed at environmental recovery, in particular a detailed database on native species, as well as articles, videos and a glossary that addresses various topics and techniques. These range from seed collection to seedling production, planting, and strategies for ecological restoration. The system provides the user with a friendly tool that helps generate a report containing suggestions for native species (Figure 7), recomposition strategies and good practices to be adopted based on the characterization of the rural property in terms of biome, above-soil vegetation and risk conditions. WebAmbiente, aligned with the National System of Rural Environmental Registry (SICAR), expands the integration with the Brazilian Forest Service (SFB), and one of its main functions is to support the implementation of the new Forest Code, by encouraging the use of the Rural Environmental Registry (CAR).

Figure 7. Native plants catalog from WebAmbiente.

Source: WebAmbiente (2020).

Hydric resources

Even though Brazil being considered a country with great availability of water resources, there are regional differences and variations throughout the year that render the study of hydrological regimes as very important. Human actions, such as changes in land use and coverage, and fluctuations in weather patterns can affect the availability of water resources, altering the natural flow in watercourses. These changes can affect the flow both in small basins (Bosch; Hewlett, 1982) and in large areas (Costa et al., 2003). The same applies to different agricultural systems and crops, with different characteristics of rain and soilwater interception. Such changes can modify the total flow rate of rainfall, generally increasing the portion drained into water bodies by anthropogenic changes (Lima et al., 2014).

Thus, biophysical models coupled with hydrological models enable evaluating human impacts or climate change on water resources. Such models vary in their degree of complexity and can be applied in the most different scales and situations. As an example, the integration of the results of the water balance with economic models of general equilibrium can be mentioned. These allow evaluating the effects of irrigation expansion on the water demand in relation to the water supply (Ferrarini et al., 2020).

Integration of socio-economic analysis in agro-environmental modeling

Projects conducted in partnership with other company units[7] and with the Brazilian Institute of Geography and Statistics (IBGE) improved the image of Brazilian agricultural production. Data from the IBGE Agricultural Census served as a basis to differentiate the types of beef and sugarcane production systems in use by producers throughout the country. Based on these works, the most representative types of production systems were chosen to be studied in depth. The project team held meetings with producers, technicians, and other agents linked to regional agriculture where they expanded the description of the technical and economic behavior of these products on rural properties. The information collected also allowed to associate the different forms of production with environmental impacts, such as the calculation of greenhouse gas emissions.

The obtained results were incorporated into mathematical models of land use – which explain how this activity's expansion takes place – which were the basis for the construction of Life Cycle Inventories (LCI), which resulted in studies of sugar cane Life Cycle Assessment (LCA) and derivatives of livestock production. Such works contribute to improving the environmental performance of products, enabling the reduction of environmental and socioeconomic impacts. The generated technical coefficients contributed to the results of several Embrapa projects with different partners[8], in addition to collaborating with the National Biofuel Policy (RenovaBio).

As a result, there is more reliable information on environmental performance of Brazilian agricultural products. This information was inserted by Embrapa in the most important international LCI databank, ecoinvent, through LCI AgroBR, which made available more than 400 LCIs of Brazilian agricultural products. Such actions help increase the competitiveness of agricultural products in the international market and promote the sustainability of Brazilian agriculture.

The characterization of sugarcane production systems was adopted by the National Bank for Economic and Social Development (BNDES) in the context of financing Flex Plants and by RenovaBio. The tool for accounting for the biofuels' carbon footprint was RenovaCalc, one of the pillars of RenovaBio, which was developed by Embrapa and partners. This policy strongly contributes to the adoption of a more sustainable model for the production of bioenergy and biofuels and to the reduction of greenhouse gas (GHG) emissions from the transport sector, contributing to reaching the national goals assumed in the Paris Agreement, as well as for national energy security.

Another aspect of the team's work makes use of models that integrate the biological and economic dimensions, with different degrees of complexity and regional outlines to assess the effects of intensification strategies on sustainability. More recently, more complex economic models have been developed to investigate the potential impacts of supply shocks, like productivity gains, and demand shocks, such as an increase in population, per capita income, changes in the Brazilian and global agriculture, equilibrium prices, and land use dynamics.

---

[7] Siscana Project, Socioeconomic Action Plan: Embrapa Environment, Embrapa Western Agriculture, Embrapa Coastal Tablelands; Economics Component of the PECUS Network Project: Embrapa Beef Cattle, Embrapa Dairy Cattle, Embrapa Eastern Amazon, Embrapa Pantanal, Embrapa Southeastern Livestock, Embrapa Southern Livestock, Embrapa Forestry and Embrapa Swine & Poultry.

[8] Among such projects, the following stand out: "Assessment of the Life Cycle of sugarcane and derivatives produced in the Center-South of Brazil, based on data, factors, and models adapted to national conditions" (LCI-cane, Embrapa); "Ethanol production through the integration of off-season corn to sugarcane mills: environmental-economic assessment and policy suggestions" (BNDES) – Available et: https://web.bndes.gov.br/bib/jspui/handle/1408/2496; "Life Cycle Inventories of Brazilian Agricultural Products: a contribution to the ecoinvent database" (ICVAgroBR, SECO, Switzerland).

Therefore, the capacity to investigate future expansion, competitiveness and sustainability scenarios for Brazilian agriculture is increased, as well as the potential impacts of some of the public policies of sectoral interest.

Applications for quantification and mitigation strategies for GHG emissions

Despite being affected by climate change, agriculture can contribute to reducing GHG emissions and mitigating the impacts of climate change. This is because the mitigation actions proposed for the sector also serve as adaptation modes, that is, by promoting greater carbon sequestration, they also result in lower nutrient losses in agroecosystems and improved physical structure and soil water availability, for example. Such actions also promote better productivity rates and better use of natural resources. Mitigating emissions result in more favorable GHG balances, helping the transition to low-emission agricultural production.

In defining the technologies contained in the Low Carbon Agriculture Plan (ABC Plan), estimates of the carbon balance were based on the difference between emissions and sequestration in production systems. These were based on data derived from experiments carried out by Embrapa and the use of simplified models for calculating how this balance could be more favorable with the adoption of good practices or technologies encouraged by the Plan. The comparison worked with the differences in relation to what is traditionally done in the management of agricultural, livestock and forestry systems. In technical jargon, the improvement of the balance in relation to what is usually done, *i.e.,* from business as usual, is called additionality. When focusing on ensuring sustainability in agriculture in all its aspects, only adaptive technologies and practices were considered. These, on top to being additional, bring greater efficiency, diversification, and profitability to the agricultural producer, along with the co-benefit of reducing emissions.

More complex studies on the carbon balance in agricultural, livestock, and forest systems, including integrated systems, were carried out by the Fluxus, Pecus, and Saltus projects, respectively. These enabled the improvement of carbon balance models in Brazilian production systems, knowledge about GHG emissions from these systems, and application in the national inventory of gases for the agricultural sector. This is part of the national communication to the UN and contains a national balance of how much is emitted in the sector.

As a way of monitoring the effectiveness of the ABC Plan and its actions, the Multi-institutional Platform for Monitoring the Reductions of Greenhouse Gas Emissions in Agriculture was created. The ABC Platform uses models based on georeferenced data and GHG emission parameters from different production systems in order to estimate emission reductions during the first ten years of the ABC Plan. This will enable the compliance assessment for the goals established for each different technology. Another tool that contributes in this regard are the GHG balance estimation protocols (GHG Protocol), internationally accepted as the best practices for the quantification of corporate, project, or product GHG emissions. These protocols offer specific technical guidelines for the national agricultural sector, constituting tools that measure and manage agricultural emissions, especially in the private sector, such as the program implemented by WRI Brazil[9].

The GHG Protocol has been used in the Carbono Araguaia project, hosted by the Liga do Araguaia[10]. The tool was adapted to the national context based on tropical agriculture and livestock parameters,

---

[9] Available at: https://wribrasil.org.br/pt/o-que-fazemos/projetos/ghg-protocolo-agropecuario

[10] Available at: http://www.ligadoaraguaia.com.br/projetos-da-liga

it was developed by Embrapa and Unicamp, and allows monitoring the GHG emission reductions from the adoption of intensification practices in 24 livestock farms in the region. These farms have a total 89,000 hectares of pastures.

For the biofuels chain, modeling contributes significantly by the development of RenovaCalc[11] in partnership with Embrapa Environment. RenovaCalc is a biofuels carbon footprint calculator, in other words, it measures the intensity of GHG emissions in the life cycle of biofuels. It generates the necessary estimates for the carbon market related to ethanol, biodiesel, and biomethane.

Another good practical example of this agro-environmental modeling derivation is the effort to estimate, monitor, better communicate, and highlight good practices in the meat sector in accordance with IPCC recommendations with Brazilian and international socio-environmental legislation. This involves the development of standards for voluntary certification of livestock products, the Certified Low Carbon Livestock Platform, led by Embrapa Beef Cattle de Corte and with the significant participation of researchers from Embrapa Agricultural Informatics. It is aligned with the ABC Plan, as is encourages and values the use of livestock systems in a more favorable carbon balance, such as ILP, ILPF, silvopastoral and intensified pasture systems. The following protocols are considered: Carbon Neutral (CCN) or Carbon Neutral Brazilian Beef (CNBB), Low Carbon Beef (CBC) or Low Carbon Brazilian Beef (LCBB), Native Carbon (CN), Low Carbon Calf (Bezerro-CN) and Neutral Carbon Leather (Couro-CN). The protocol developed by Embrapa allows partners to use their respective concept brands – or "environmental seals" – in their activities. This certification will allow consumers to recognize farmers' efforts to promote sustainable, low-GHG farming systems in an integrated landscape approach. Another important point is the dissociation of deforestation and the differentiation of meat sustainably produced in Brazil, which improves international acceptance in the international market, where the country is highly competitive.

Another important initiative to support the formulation of Brazilian public policies to mitigate GHG emissions was the development of the model Economic Analysis of Greenhouse Gases for Livestock Emissions (EAGGLE). It is a detailed optimization model that economically evaluates pasture recovery and GHG mitigation strategies in beef cattle production systems (De Oliveira Silva et al., 2017). It was developed in partnership between Embrapa, the State University of Campinas (UNICAMP), and the University of Edinburgh. The model explores complex scenarios focused on sustainable use of the production area by increasing productivity and the diversity of techniques and products. In the most technical jargon, sustainable intensification of animal production. Furthermore, it allows analyzing the optimization of the adoption rate in the main practices on animal performance efficiency (pasture supplementation, confinement) and pasture (direct and indirect restoration, irrigation), in order to mitigate emissions and save land. It was used by the Brazilian government to develop national policies aimed at emission mitigation actions, particularly in the submission of the Nationally Determined Contribution (NDC) (De Oliveira Silva et al., 2018), a document that records the main commitments of Brazil's contributions to the Paris climate agreement. Additionally, it allows for the consequential analysis on the intensification of meat production in Brazil (De Oliveira Silva et al., 2016).

All these initiatives have been carried out with the strong collaboration of the Agro-environmental Modeling Research Group and also the presidency of Embrapa's Climate Change Portfolio, under the responsibility of Embrapa Agricultural Informatics. They represent Embrapa's contribution to meeting the various demands of society with regard to the climate change and agriculture interface, using simulation tools to envision and positively influence trends for agriculture in the future climate.

---

[11] Available at: http://www.anp.gov.br/producao-de-biocombustiveis/renovabio/renovacalc

Final considerations

Models and simulators that can accurately and precisely depict the responses of agroecosystems demand quantity and quality data for their calibration, validation, and model fusion. In this "big data era", data derived from experimental stations are critical for the development of models and simulators, and must be complemented by massive data collection in the field. Obviously, this greater data collection capacity must be accompanied by compatible transfer capacity (IoT, cloud computing) and storage (Big Data). And with data in adequate quantity and quality, algorithms and analysis tools can be improved and/or developed, tested, validated. These can then be made available as part of the decision-making process within agricultural establishments from the production chains and in the formulation of public policies.

One of the great challenges of the digital age is the efficient integration of data, information, and knowledge into models and algorithms. This continuous incorporation of new knowledge into mathematical models and algorithms for data assimilation, decision analysis and optimization, and the integrated use of multi-source and multi-temporal spatial data through agro-environmental modeling are essential to promote competitiveness and the sustainable development of Brazilian agriculture. It is necessary to intensify public-private partnerships in order to enable massive data collection, in space and time, and faster research advances to assimilate these measures, generating analyses, products and services that can be used by rural producers and decision makers in the public and private spheres.

Producers will increasingly demand efficiency and sustainability, given the increase in production costs and the demands not only of agricultural products, but also of environmental and ecosystem services associated with them. Companies and service providers should be more interconnected with innovation ecosystems via research institutes, universities and rural extension. Consumer markets, national and international, tend to increasingly demand food, fiber and energy with certifications that guarantee sustainable production, in addition to quality. Therefore, digital technological solutions that integrate a wide knowledge spectrum will be essential to guide the various actors in Brazilian agribusiness.

As exemplified throughout this chapter, Embrapa Agricultural Informatics has been engaged in the direct development and support for new technologies, and also in making this technology accessible. This is a toolset capable of bringing innovation to all of society and the potential users of digital agriculture in the very near future.

References

AGRICULTURA DIGITAL: PESQUISA, DESENVOLVIMENTO E INOVAÇÃO NAS CADEIAS PRODUTIVAS DE OLIVEIRA SILVA, R.; BARIONI, L. G.; HALL, J. A. J.; MORETTI, A. C.; VELOSO, R. F.; ALEXANDER, P.; CRESPOLINI, M.; MORAN, D. Sustainable intensification of Brazilian livestock production through optimized pasture restoration. **Agricultural Systems**, v. 153, p. 201-211 2017. DOI: 10.1016/j.agsy.2017.02.001.

AGRITEMPO. Available at: https://www.agritempo.gov.br/agritempo/index.jsp. Accessed on: 25 May 2020.

AGROIDEAL. Available at: https://www.agroideal.org. Accessed on: 25 May 2020.

ALENCAR, J. R.; ROMANI, L. A. S.; MERLO, T. P.; EVANGELISTA, S. R. M.; OTAVIAN, A. F. Avaliação dos impactos do uso do Sistema de Monitoramento Agrometeorológico (Agritempo). **Revista de Política Agrícola**, ano 25, n. 1, p. 5-19, jan./fev./mar. 2016.

ANDRADE, C. A.; FASIABEN, M. do C. R. A utilização dos instrumentos de política ambiental para a preservação do meio ambiente: o caso dos Pagamentos por Serviços Ecossistêmicos (PSE). **Economia Ensaios**, v. 24, n. 1, 2009. Available at: https://ainfo.cnptia.embrapa.br/ digital/bitstream/item/23981/1/2223-29258-1-PB-1.pdf. Accessed on: 1 jun. 2020.

ARIAS, D.; MENDES, P.; ABEL, P. **Revisão rápida e integrada da gestão de riscos agropecuários no Brasil**. Brasília, DF: Banco Mundial, 2015. 76 p.

ASSAD, E. D. **Simulation de l'irrigation et du drainage pour les cultures pluviales de riz et de mas en sols de bas-fonds à Brasília**. Montpellier: Irat, 1986. 10 p. (IRAT. Memories et travaux, 13).

ASSAD, E. D.; OLIVEIRA, A. F. de; NAKAI, A. M.; PAVÃO, E.; PELLEGRINO, G. Q.; MONTEIRO, J. E. B. de A. Impactos e vulnerabilidades da agricultura brasileira às mudanças climáticas. In: MODELAGEM climática e vulnerabilidades setoriais à mudança do clima no Brasil. Brasília, DF: Ministério da Ciência, Tecnologia e Inovação, 2016. p. 127-188.

ASSAD, E.; PINTO, H. S. (coord.). **Aquecimento global e a nova geografia da produção agrícola no Brasil**. São Paulo: Embrapa: Unicamp, 2008.

BOLFE, E. L.; LOPES, D. B.; CONTINI, E. Territórios & políticas públicas rurais. **Clima com Cultura Científica**, v. 3, p. 1-15, 2016.

BOLFE, E. L.; SANO, E. E.; CAMPOS, S. K. (ed.). **Dinâmica agrícola no cerrado**: análises e projeções. Brasília, DF: Embrapa, 2020. v. 1, 308 p. Available at: http://ainfo.cnptia.embrapa.br/digital/ bitstream/item/212381/1/LV-DINAMICA-AGRICOLA-CERRADO-2020.pdf. Accessed on: 5 Aug. 2020.

BOSCH, J. M.; HEWLETT, J. D. A review of catchment experiments to determine the effect of vegetation changes on water yield and evapotranspiration. **Journal of Hydrology**, v. 55, n. 1-4, p. 3-23, Feb. 1982. DOI: 10.1016/0022-1694(82)90117-2.

CHALLINOR, A. J.; WATSON, J.; LOBELL, D. B.; HOWDEN, S. M.; SMITH, D. R.; CHHETRI, N. A meta-analysis of crop yield under climate change and adaptation. **Nature Climate Change**, v. 4, n. 4, p. 287-291, Apr. 2014. DOI: 10.1038/nclimate2153.

COSTA, M. H.; BOTTA, A.; CARDILLE, J. A. Effects of large-scale changes in land cover on the discharge of the Tocantins River, Southeastern Amazonia. **Journal of Hydrology**, v. 283, n. 1-4, p. 206-217, Dec. 2003. DOI: 10.1016/S0022-1694(03)00267-1.

COSTANZA, R.; D'ARGE, R.; DE GROOT, R.; FARBER, S.; GRASSO, M.; HANNON, B.; LIMBURG, K.; NAEEM, S.; O'NEILL, R. V.; PARUELO, J.; RASKIN, R. G.; SUTTON, P.; BELT, M. van den. The value of the world's ecosystem services and natural capital. **Nature**, v. 387, p. 253-260, May 1997. DOI: 10.1038/387253a0.

DE OLIVEIRA SILVA, R.; BARIONI, L. G.; HALL, J. A. J.; MATSUURA, M. F.; ALBERTINI, T. Z.; FERNANDES, F. A.; MORAN, D. Increasing beef production could lower greenhouse gas emissions in Brazil if decoupled from deforestation. **Nature Climate Change**, v. 6, p. 493-497, 2016. DOI: 10.1038/nclimate2916.

DE OLIVEIRA SILVA, R.; BARIONI, L. G.; PELLEGRINO, Q. G.; MORAN, D. The role of agricultural intensification in Brazil's Nationally Determined Contribution on emissions mitigation. **Agricultural Systems**, v. 161, p. 102-112, Mar 2018. DOI: 10.1016/j.agsy.2018.01.003.

EMBRAPA. **Aplicativo Zarc – plantio certo**. Available at: https://www.embrapa.br/busca-de-solucoes- tecnologicas/-/produto-servico/6516/aplicativo-zarc---plantio-certo. Accessed: 25 May 2020.

FERRARINI, A. dos S. F.; FERREIRA FILHO, J. B. de S.; CUADRA, S. V.; VICTORIA, D. De C. Water demand prospects for irrigation in the São Francisco river: Brazilian public policy. **Water Policy**, v. 22, n. 3, p. 449-467, Apr. 2020. DOI: 10.2166/wp.2020.215.

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. Brief history of agricultural systems modeling. **Agricultural Systems**, v. 155, p. 240-254, July 2017. DOI: 10.1016/j.agsy.2016.05.014.

LIMA, L. S.; COE, M. T.; SOARES FILHO, B. S.; CUADRA, S. V.; DIAS, L. C. P.; COSTA, M. H.; LIMA, L. S.; RODRIGUES, H. O. Feedbacks between deforestation, climate, and hydrology in the Southwestern Amazon: implications for the provision of ecosystem services. **Landscape Ecology**, v. 29, p. 261-274, 2014. DOI: 10.1007/s10980-013-9962-1.

MICURA. Available at: https://www.micura.cnptia.embrapa.br. Accessed on: 25 May 2020.

MILLENNIUM ECOSYSTEM ASSESSMENT. **Ecosystem and human well-being**: synthesis. Washington: Island Press, 2005.

NAÇÕES UNIDAS. **Agenda 2030**. 2015. Available at: https://nacoesunidas. org/pos2015/agenda2030. Accessed on: 1 June 2020.

ROSSETTI, L. A. Seguro rural e zoneamento agrícola no Brasil: novos rumos. **Revista de Política Agrícola**, ano 4, p. 33-43, out./dez. 1998.

SANTOS, W. G. dos; MARTINS, J. I. F. O Zoneamento Agrícola de Risco Climático e sua contribuição à agricultura brasileira. **Revista de Política Agrícola**, v. 25, n. 3, p. 73-94, dez. 2016.

STEVANOVIĆ, M.; POPP, A.; LOTZE-CAMPEN, H.; DIETRICH, J. P.; MÜLLER, C.; BONSCH, M.; SCHMITZ, C.; BODIRSKY, B. L.; HUMPENÖDER, F.; WEINDL, I. The Impact of High-End Climate Change on Agricultural Welfare. **Science Advances**, v. 2, n. 8, p. e1501452, Ago. 2016. DOI: 10.1126/sciadv.1501452.

VICTORIA, D. de C.; BOLFE, E. L.; SANO, E. E.; ASSAD, E. D.; ANDRADE, R. G.; GUIMARAES, D. P.; LANDAU, E. C. Potencialidades para expansão e diversificação agrícola sustentável do Cerrado. In: BOLFE, E. L.; SANO, E. E.; CAMPOS, S. K. (ed.). **Dinâmica agrícola no cerrado**: análises e projeções. Brasília, DF: Embrapa, 2020. v. 1, p. 229-258. Available at: http://ainfo.cnptia.embrapa.br/digital/ bitstream/item/212396/1/PL-Dinamica-agricola-cap8-2020.pdf. Accessed on: 1 June 2020.

WEBAMBIENTE. Available at: https://www.webambiente.gov.br. Accessed on: 25 May 2020.

ZHAO, C.; LIU, B.; PIAO, S.; WANG, X.; LOBELL, D. B.; HUANG, Y.; HUANG, M.; YAO, Y.; BASSU, S.; CIAIS, P.; DURAND, J.-L.; ELLIOTT, J.; EWERT, F.; JANSSENS, I. A.; LI, T.; LIN, E.; LIU, Q.; MARTRE, P.; MÜLLER, C.; PENG, S.; PEÑUELAS, J.; RUANE, A. C.; WALLACH, D.; WANG, T.; WU, D.; LIU, Z.; ZHU, Y.; ZHU, Z.; ASSENG, S. Temperature increase reduces global yields of major crops in four independent estimates. **Proceedings of the National Academy of Sciences**, v. 114, n. 35, p. 9326-9331, 29 Ago. 2017.
DOI: 10.1073/pnas.1701762114.


Photo: sakdam (AdobeStock)

4

Geotechnologies in digital agriculture

Carla Geovana do Nascimento Macário | Júlio César Dalla Mora Esquerdo | Alexandre Camargo Coutinho | Eduardo Antonio Speranza | João dos Santos Vila da Silva | João Francisco Gonçalves Antunes | Laurimar Gonçalves Vendrusculo | Sérgio Aparecido Braga da Cruz

Introduction

Geotechnologies applied in agriculture are increasingly used in Brazil. The monitoring of agricultural crops, surveying and characterizing natural resources, mapping land use and land cover, zoning and evaluating scenarios are some examples that show the presence and the use of geotechnologies. Geotechnology is a specific technology aimed at the acquisition, storage, processing, visualization and analysis of geospatial data, which, in turn permeate, directly or indirectly, a series of themes related to the dynamics of agricultural activity. Remote Sensing, Geographic Information Systems, Global Positioning Systems (GPS) of satellite navigation and Geospatial Database are examples of geotechnologies widely used by various sectors within the particularly diverse agricultural and environmental applications.

The production and availability of geospatial data have increased significantly in recent years. The spread of data repositories and geospatial services on the internet, advances in data processing and storage in computational cloud structures, the expanding use of standards for geoinformation representation, the increase in the number of various types of sensors and for different purposes, and the presence of GPS on mobile devices supporting the georeferenced collection of different types of data are examples that explain the growing use of geospatial data by the society in general. In addition to the vast volume of information, the current context in the production of these geographically located data is characterized by the speed it is produced. Furthermore, the technological advances, georeferenced information and data are increasingly accessible, becoming strategic for decision-making in agriculture and in matters related to territorial management.

Geotechnologies are extremely important in digital agriculture, as they enable verifying temporal and spatial variability, conducting the traceability of production in all links of the chain, as well as monitoring farms, among other functions.

Embrapa Digital Agriculture has worked on this theme over the past few years, including the use, application and development of geotechnologies through a multidisciplinary technical team of the Geotechnology Research Group. This chapter discusses some geotechnologies used by this team in R&D activities, particularly to meet government demands related to agriculture and environment, especially in land management and on the development of computational tools in order to support public policies. The basic concepts about geospatial data are presented, as well as their organization in computational environments, which is a vital and necessary activity in the development of tools for this specific type of data. Next, a brief characterization of an important source of geospatial information is presented: remote sensing, fundamental in various activities for monitoring the earth's surface, which is present in projects and in research initiatives using geotechnologies. Finally, some computational tools based on Geographic Information Systems on the Internet are described, the "WebGIS", intended for visualization and for spatial analysis of geospatial data, and their use in land management.

Applications of geotechnologies in agricultural and environmental monitoring

Geospatial data: basic concepts and their organization

Geospatial data are used to represent real-world phenomena associated with its location on the Earth's surface. This representation is accomplished through three characteristics: spatial, non-spatial and temporal. The spatial characteristic concerns the geographic location of the phenomenon that the data represent and its geometry; the non-spatial characteristic identifies the phenomenon and its properties; and the temporal characteristic informs when the phenomenon occurred. For example, a grain harvester equipped with a yield monitor and a global positioning system can record the amount of grain harvested, the geographic location in the field along the machine's trajectory, while recording the time of periods which the grain inflows into the machine.

The spatial characteristic provides the geospatial data topological and geometric properties. The former is based on the relative positions of objects in space, such as connectivity, orientation, adjacency and containment. The second one represents the geometry of the entities, obtained from their features such as points, lines and polygons. Such properties allow defining the relationships between geospatial data, which are fundamental in geographic applications.

In the agricultural context, geospatial data and their relationships can be used in a variety of activities. For example, on a regional scale, they can be used to represent natural resources, such as rivers, lakes and protected areas close to agricultural projects, and also to represent the dynamics of land use and land cover in the same region over time. At a local scale, on the other hand, they can represent specific characteristics in a farm field, such as the occurrence of weeds and pests, soil fertility or the occurrence of planting failures.

Using this type of data for answering questions related to agriculture requires their computational representation, which can be done through file structures. As the geographic data are spatialized and computationally represented, considering their characteristics and their spatial relationships, different analyses can be performed to provide the desired answers. Geospatial data can be computationally represented by two models: vector and raster, also called raster, as shown in Figure 1. In the vector

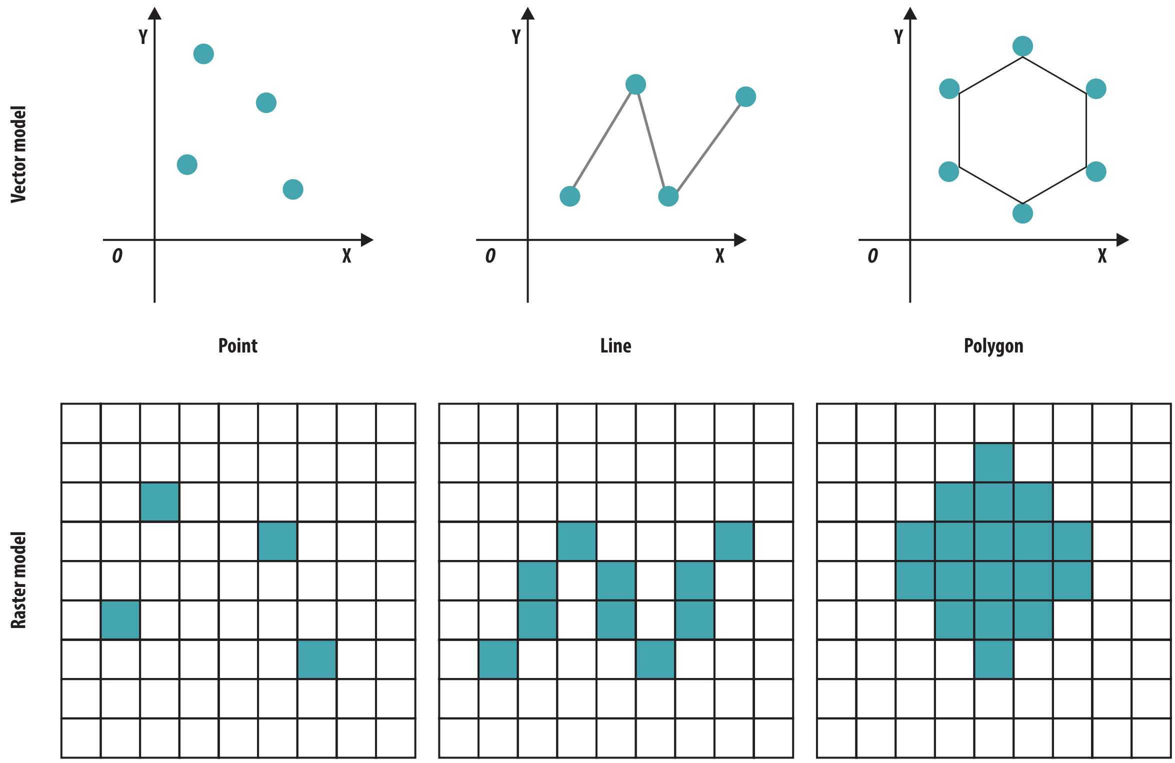


Figure 1. Vector and raster representation models.

Source: Adapted from Li and Yan (2020).

model, the spatial characteristic is reproduced by points, lines or polygons in two-dimensional space, defined in a precise manner with respect to its location, boundary, interior and exterior. Groups of points, lines or polygons can also be used to represent a single phenomenon. In the raster model, on the other hand, space is considered as a regular surface, divided into pixels (picture element) or fixed size cells, representing portions of a territorial area. An important factor in this model is its spatial resolution, calculated based on the size of each cell on the map and the terrain area that the cell can represent.

Large amounts of data, whether geospatial or not, need to be organized for efficient handling. In this regard, we have the concept of Database Management System (DBMS), which are sets of records arranged in a regular structure, which enables data organization and production of information. Geographic Databases work in the same way, and the difference between them is the ability to support geometric features in their tables. In general, traditional DBMS can manipulate geospatial data from spatial modules or extensions. Thus, information is georeferenced, its operations are spatial and, generally, its visualization is cartographic. Examples of spatial extension are PostGIS, present in the public and open-source DBMS PostgreSQL, and Oracle Spatial, present in the Oracle commercial package.

DBMS with spatial extension implements spatial types and operators, in addition to traditional operations, with independence and efficient access, allowing data sharing and redundancy reduction, integrity and security, uniform administration, short-time application development and concurrent access. These characteristics facilitate the developer's task of geographic applications, such as Geographic Information Systems (GIS), which will be discussed in item "Web-based Geographic Information Systems" of this chapter.

The first geographic DBMS efficiently treated only with data represented in the vector model, offering ways of storage and operations for their handling. On the other hand, spatial extensions for treating data in raster model, such as satellite images, were developed along the time and are now present in various geographic database systems. One example is Well-Known Text (WKT Raster), an extension of PostGIS, which allows for a more efficient and practical way to store and analyze raster data.

The computational organization of geospatial data, whether by a geographic database or by any other type of structure allows them to be handled by different available geotechnologies, such as GIS, WebGIS and remote sensing tools. These technologies will be addressed in the following items in this chapter.

Remote sensing

Several definitions for the term remote sensing can be found in the literature. The American Society for Photogrammetry and Remote Sensing (ASPRS), one of the most important societies that brings together researchers from all over the world on this topic, adopted Colwell's (1983) definition. According to this definition, remote sensing can be understood as the measurement or acquisition of information of some property of an object or phenomenon, by a recording device that is not in physical or close contact with the object or phenomenon under study. According to Jensen (2007), this measurement or acquisition can be performed *in situ* through equipment taken to the field, or more remotely, from equipment installed in suborbital air vehicles, such as balloons, planes and drones, and at orbital level, such as satellites.

Remote imaging sensors are those most commonly found onboard orbital platforms, at altitudes ranging between 200 and 36 thousand kilometers. These sensors transform the electromagnetic energy emitted or reflected by the Earth's surface into an analogical signal, which is processed and converted to build a digital image. Each element that forms the image, called pixel, has a value relative to the amount of energy reflected or emitted by a portion of the Earth's surface. The size of this portion of the observed surface area depends on the sensor's spatial resolution, ranging from centimeters to kilometers, depending on the type and purpose of each instrument.

In addition to the spatial resolution, other sensor characteristics are taken into account when choosing the type of image to be used in each application, such as: temporal resolution, which defines the time the sensor has to revisit the same location; the spectral resolution, which defines the widths and amounts of spectral bands that the sensor is able to "see"; and radiometric resolution, which defines the sensor's ability to distinguish different digital levels, that is, the system's efficiency in detecting and recording differences in the energy reflected or emitted by ground targets.

Currently, there are several sensors in operation onboard orbital platforms developed by governments or private companies, generating products with different surface details, with different passing frequencies and with different acquisition costs. In general, large-scale applications for government purposes make use of freely available public data by official repositories on the internet.

The most successful government program for the remote observation of land resources is the Earth Resources Technology Satellite (ETRS), implemented in the 1970s by National Aeronautics and Space Administration (NASA), the American space agency. This program, which was later renamed Landsat, is responsible for launching eight orbital platforms into space, equipped with imaging sensors of medium-spatial resolution. Launched in 2013, the Landsat-8 is the latest platform from this family of satellites, taking aboard the Operational Land Imager (OLI) sensor. These images are available at a spatial resolution of 30 meters and are processed to remove atmospheric influences (atmospheric correction), as well as terrain distortions effects (orthorectification), in order to improve its radiometric and geometric features, respectively. Another important NASA mission is carried out by the Moderate Resolution Imaging Spectroradiometer (MODIS), sensor aboard the Terra and Aqua satellites, in operation since the early

2000s and which provides data with great periodicity and low spatial detail. Its historical series, which covers a period of more than 22 years of images, is freely available in the form of terrestrial, atmospheric and oceanic products, pre-processed in three spatial resolutions (250 m, 500 m and 1,000 m), derived from 36 spectral bands.

Brazil is also among the countries that produce remote sensing data for monitoring land resources through the CBERS Program (China-Brazil Earth Resources Satellite), the result of the technical-institutional partnership with China, initiated at the end of the 1980s, which involved the National Institute for Space Research (INPE). Currently, the CBERS satellite series, started by CBERS-1 in 1999, has four other satellites, the last one launched in December 2019. CBERS-4A is part of the second generation of this family of satellites and features imaging sensors with different characteristics, producing detailed images between 2 and 60 meters of spatial resolution. The historical series of CBERS images is available at INPE's official repository, one of the forerunners of this policy of publicly and free availability of data through catalogs on the internet.

The availability of high spatial resolution public images has also grown in recent years. In 2014, European Space Agency (ESA) launched Sentinel-1, a satellite equipped with an active imaging sensor (radar). Currently, the Sentinel series has six satellites in orbit. In agricultural monitoring, the most used is Sentinel-2, launched in 2015, which has an onboard the MSI (Multispectral Instrument) sensor that generates images with spatial resolution of up to 10 m in the visible and infrared bands, with revisit every five days, combining good spatial detail and relative observation frequency.

The continuous advancement of technologies in the production of increasingly smaller and more efficient components has represented a trend in the geospatial market with the new generation of Earth observation nanosatellite constellations, such as the private PlanetScope satellites, which offer an unprecedented combination of images with daily temporal resolution and spatial resolution of three meters. Thus, nanosatellites can support the monitoring of agricultural crops that require greater spatial detail, such as coffee and citrus, as well as ILPF systems, which integrate temporary farming, livestock and planted forest in the same production area.

The following topics present initiatives and projects by Embrapa Digital Agriculture and its respective partners, which directly involved the use of remote sensing in the processes of monitoring the earth's surface.

Time series of satellite images and the development of the Temporal Vegetation Analysis System (SATVeg)

In recent years, satellite imagery time series have been increasingly used in a wide range of applications for monitoring the Earth's surface. In remote sensing, the spectro-temporal approach exploits the short revisit time of some orbital sensors aimed at the more frequent acquisition of spectral information from the Earth's surface, bringing advantages over the traditional approach, which is based on a restricted set of images.

In multitemporal analyses for monitoring terrestrial vegetation cover, vegetation indices are commonly used, derived from mathematical combinations between the sensors' spectral bands. These will enhance the presence and vigor of vegetation and reduce the influence of soil and atmospheric factors. The Normalized Difference Vegetation Index (NDVI) and the Enhanced Vegetation Index (EVI), proposed by Rouse et al. (1974) and Huete et al. (1994), respectively, are the most cited in the literature and show a high correlation with green biomass and leaf area. When organized and observed chronologically, these spectral indices can be used to produce temporal profile charts in order to represent the variations in vegetative vigor over time. Among the most used orbital sensors in multitemporal analysis, MODIS stands out, whose time series of more than 22 years of images has good radiometric and geometric consistency. Figure 2 illustrates a temporal profile example of NDVI, obtained from the MODIS sensor, and the respective interpretation of the land use and land cover dynamics between 2000 and 2018.

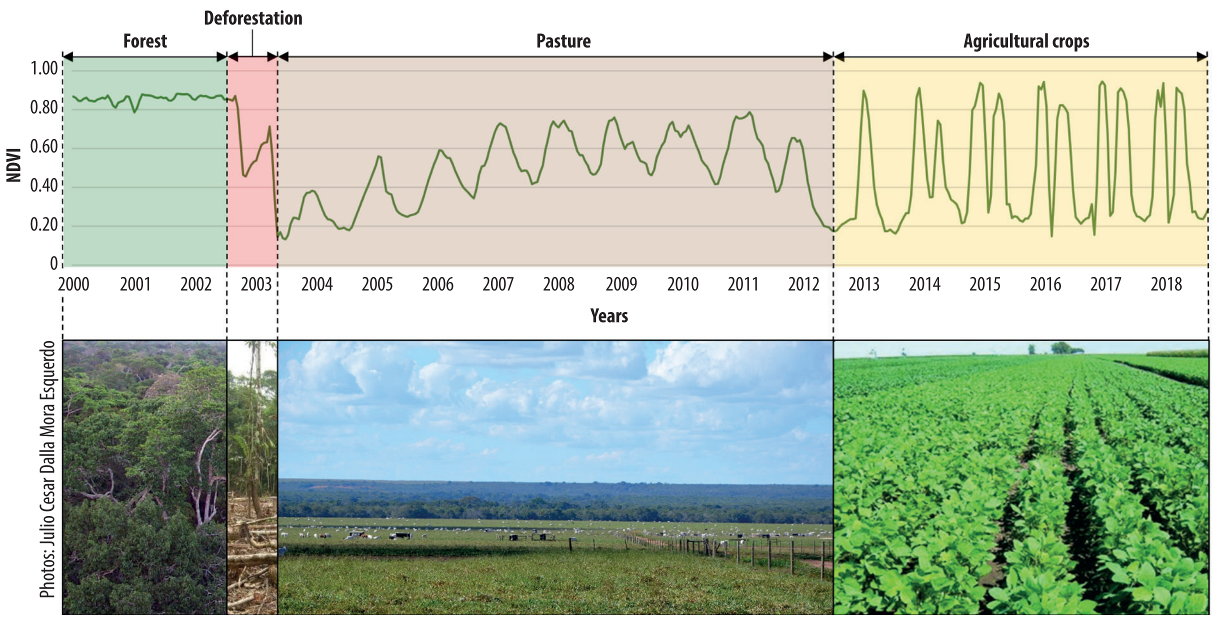


Figure 2. Normalized Difference Vegetation Index temporal profile showing different patterns from transitions in land use and land cover between 2000 and 2018.

Source: Adapted from Embrapa Agricultural Informatics (2020a).

This example enables to identify different behavior patterns of the NDVI temporal profile throughout the historical series, considering the spatial detail of the MODIS sensor images of 6.25 ha (250 m x 250 m). Such variations are due to changes in vegetation cover at a selected point in the state of Mato Grosso. Figure 2 shows that, originally, the area had a natural forest cover, which was deforested at the end of 2002 and occupied by pasture activity until the end of 2012, when it started to be used for temporary agricultural cultivation, with successive cycles of single and double crops. Each segment of the NDVI chart reflects the phenological variations which are characteristic of each type of vegetation cover, which allow to identify the breaks and changes in patterns as a result of the transitions over time.

Analyses based on time series require the acquisition and processing of a considerable volume of data derived from satellite images and involve robust computational activities. Thus, in 2011, Embrapa Digital Agriculture began to develop a Web platform capable of providing temporal profiles of the NDVI and EVI vegetative indices of the MODIS sensor for any location in the Brazilian territory, without the user having to perform the acquisition of images or the execution of any type of processing. Initially aimed at supporting current projects at Embrapa, this Web platform, which was later called the Temporal Vegetation Analysis System (SATVeg), was developed to meet the demand for the immediate supply of temporal profiles of vegetative indices by the internet, from an easy access and visualization platform (Esquerdo et al., 2020).

The system development process went through several phases and involved using different database models to identify the best way to store a large time series of satellite images and, at the same time, allow the instantaneous query and extraction of data from a user-supplied geographic location. The results achieved in these stages proved to be promising and, in 2014, the Embrapa team decided to make the system open to society, free of charge. SATVeg then started to be accessed by a wide-ranging audience, with different interests and purposes of use, stimulating demands for new functionalities. In 2016, with the objective of developing new features, expanding the system's coverage area throughout South America and creating a new visual identity, a partnership with the private sector was initiated, resulting in the current version of the system[1].

SATVeg works very intuitively and can be used by audiences unfamiliar with remote sensing. Figure 3 illustrates the main screen of the system, in which the user interacts with a Google Maps layer to select an area of interest and obtain the temporal profile of the chosen vegetation index (NDVI or EVI), which is shown below the reference map. The system also offers a set of features, such as curve smoothing filters, overlaying layers on the map, pattern library, how-to-use tutorial, among others.

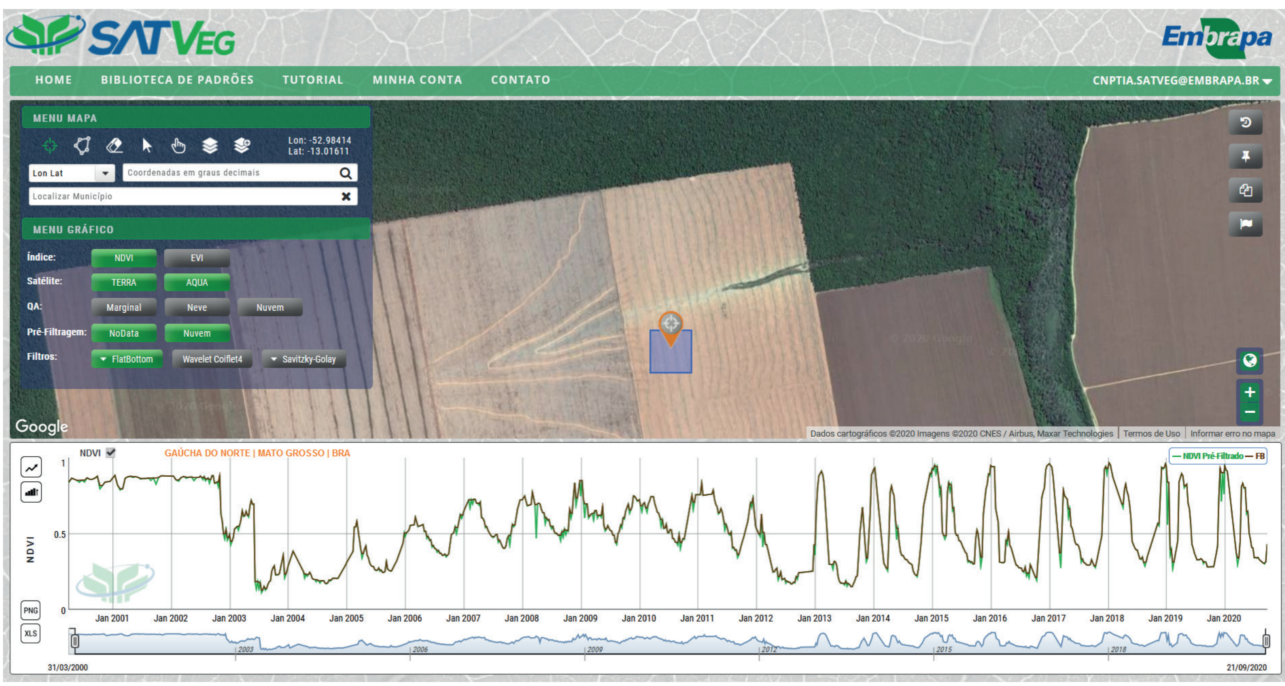


Figure 3. Temporal Vegetation Analysis System (SATVeg) main screen.

Source: Adapted from Embrapa Agricultural Informatics (2020a).

With more than 14 thousand registered users at present, SATVeg has supported several land surface monitoring activities, such as large-scale land use and land cover mapping projects, monitoring of the productive potential of agricultural crops, detection of deforestation, inspections, among others. SATVeg was recently included in the Central Bank of Brazil under Ordinance no. 4,796 as a remote tool to support proof of agricultural losses under the Agricultural Activity Guarantee Program (Proagro), as the system can indicate the conditions of green biomass from crops, and support decisions on issues related to agricultural insurance payment.

[1] Available at: www.satveg.cnptia.embrapa.br

Developed using free software, the SATVeg database has been periodically fed with new images from the MODIS sensor, and Embrapa Digital Agriculture has kept the software components updated. An Application Programming Interface (API) was also developed to programmatically provide the system data, in order to feed third-party systems that make use of the temporal profiles of the vegetation indexes in several other applications.

Land use and land cover monitoring – TerraClass Project

The theoretical bases of sustainable development presuppose public actions and policies to promote economic development, synchronized with guarantees for the preservation of natural ecosystems, their biodiversity and their services. In Brazil, a country with one of the largest tracts of unexplored arable land on the planet, the expansion of the agricultural frontier and the technological development of agricultural activities have been accompanied by intense national and international debates on this topic.

Concerned with the negative impacts on the agricultural commodities market, generated by the release of data which showed that agriculture was rapidly advancing on the Amazon forests, the Federal Government asked Embrapa and INPE, public institutions with accumulated knowledge and competence on the subject, to produce, based on scientific and impartial data, a new vision on the land use and land cover in the deforested areas of the Legal Amazon. As a result of this institutional mobilization, the TerraClass Amazon Project emerged (Almeida et al., 2016), to systematically produce maps related to land use and land cover in the region, allowing to monitor the impacts of actions, and public policies of the Federal Government focused both on the development and intensification of agricultural activity and the preservation of natural systems.

Developed and implemented by federal institutions, TerraClass generates official data on the dynamics of land use and coverage, fundamentally expanding the management capacity in the Amazon territory and reinforcing national sovereignty over a region with explicit disputes between sectors, agents and actors, national and international, whose interests are driven by economic, social and environmental issues, among others.

Currently, seven mappings are available, referring to the base years 1991, 2000, 2004, 2008, 2010, 2012 and 2014, carried out by the technical teams of Embrapa Digital Agriculture, Embrapa Eastern Amazon and the Regional Center of the Amazon (CRA/INPE). To organize, store and make them accessible to users, Embrapa Digital Agriculture has developed a free public access portal, presented in topic "Interactive Environmental Licensing Support System (SISLA)" of this chapter.

The area covered by the maps covers the entire extension of the Legal Amazon, although the identification of thematic classes is carried out by satellite imagery only in the deforested areas mapped by the Program for Deforestation Monitoring in the Brazilian Legal Amazon (PRODES). The thematic classes covered by TerraClass are presented in Table 1.

The mappings were carried out at a scale of 1:100,000 (each centimeter on the paper map is equivalent to one kilometer in the real world) from public images with varied spatial detail. The main image base is formed by scenes from the Landsat satellite, with a spatial resolution of 30 m. Time series of MODIS images, with spatial resolution of 250 m, were also used in the mapping process of the temporary agriculture class.

Different methods were used in the mapping activities of the thematic classes, according to the complexity of the targets and the technical knowledge of the institutions involved. These methods used different types of approaches, from the most complex, based on machine learning techniques, to the most simple and laborious approach, such as the visual interpretation of images. Temporary agriculture,

Table 1. Thematic classes mapped by TerraClass.

Classes	Description
Temporary agricultural crop	Agricultural crops that present one or more production cycles in the reference crop year mapping, such as soybeans, corn, cotton, among others.
Semi-perennial agricultural crop	Agricultural crops that present a larger production cycle than the reference crop year mapping, mainly represented by sugarcane.
Perennial agricultural crop	Permanent agricultural crops, with different stages of maturity and vegetation cover, such as coffee, citrus, rubber plantations, among others.
Cultivated arboreal shrub pasture	Pastures with a predominance of woody vegetation, composed of shrub or arboreal species, in addition to cultivated herbaceous species.
Herbaceous cultivated pasture	Pastures with predominance of herbaceous forage vegetation, composed of cultivated species.
Silviculture	Forest species of commercial interest, represented by monospecific tree formations, such as eucalyptus and pine.
Secondary natural forest vegetation	Natural vegetation in the process of regeneration, characterized by the densification of tree species that have already undergone total original vegetation suppression since the beginning of deforestation monitoring in the Amazon.
Mining	Mineral extraction areas characterized by the presence of bare soils and changes in the local landscape.
Urban area	Urban stains as a result of population concentration that form villages, towns or cities that present an infrastructure different from the rural area, presenting densification of streets, houses, buildings and other public facilities.
Other urbanized areas	Areas with differentiated infrastructure, with less density of streets, houses and other industrial equipment such as sheds, mills, warehouses, among others.
Water bodies	Natural water bodies such as rivers, lakes, dams and dams.
Others	Areas that do not fit into the other thematic classes, such as rocky outcrops, fluvial beaches, sand banks, among others.
Not observed	Areas not mapped in the satellite images used due to the presence of clouds or burned areas.
Deforestation of the year	Areas whose natural vegetation cover was removed during the reference year mapping.

Source: GeoPortal TerraClass (2020).

which presents high spectral dynamics and occurs in specific periods of the year, was mapped using the spectral-temporal approach, addressed in item "Time Series of Satellite Images and the Development of the Temporal Vegetation Analysis System" of this chapter. Here, machine learning-based classifiers were trained with field samples and applied to a time series of vegetation indices to detect targets with specific temporal behaviors. The secondary vegetation areas were mapped based on a Linear Spectral Mixture Model (LSMM) applied to the deforestation areas to identify the vegetation, soil and shade fractions and to separate the secondary vegetation areas from the others. As for other types of thematic classes, such as pastures, forestry, mining, urban areas, and others, identification and mapping were done based on visual interpretations of Landsat images, due to the wide variation in patterns and the difficulty to automate a classification method.

Due to the strategic nature of the information produced by TerraClass throughout the historical series of mappings for the Legal Amazon, the Federal Government requested expanding the coverage area of the mappings for the Cerrado biome, another territorial portion strongly pressured by agricultural expansion and monitored by national and international organizations. To meet this demand, a new institutional articulation, which involved research, development and innovation institutions, universities and government agencies, was consolidated to carry out the mapping of land use and land cover of the Cerrado biome, referring to the base year 2013. Currently, new versions of TerraClass Cerrado, referring to the base years 2018, 2020 and 2022, are being carried out to update the mappings and to form a historical series that allows analyzing the dynamics of land use and coverage in that biome.

Since the first demand initiated by the Federal Government, there has been a constant and significant advance in the area of information technology, both with regard to new sensors embedded in satellites, planes, drones, and others, as well as new storage devices, processing, analysis and distribution of digital data and information. Thus, a large part of the efforts and focus of the TerraClass coordination and execution team has been directed towards updating the data from the historical series of the Amazon and Cerrado mappings, expanding the mappings to other Brazilian biomes and promoting greater automation of mapping processes, in order to provide significant reductions in the cost and execution time.

GeoMS Project

From 2007 to 2012, the government of Mato Grosso do Sul state, Embrapa Digital Agriculture and partners developed the GeoMS project in order to structure a geo-referenced information system to monitor the rural space. This could help state governments improve their decision-making efficiency to implement strategic projects, using the state of Mato Grosso do Sul as a case study. This demand came from specific needs of the state, such as mitigating problems associated with climate change, avoiding the waste of natural resources and reducing deforestation and carbon emissions. The initiative also originated from the concern with issues inherent to sustainable development and conservation of the environment, as well as seeking to improve the state's environmental management system, making environmental licensing processes and procedures more efficient and transparent.

Throughout its term, the GeoMS teams developed and implemented the Interactive Environmental Licensing Support System (Sisla), a web-based geographic information platform intended to organize the geospatial data from the state of Mato Grosso do Sul and decision support, which is presented in item "Interactive Environmental Licensing Support System (SISLA)" of this chapter. Considering that the system was developed to provide reliable information to support environmental licensing, it was necessary to map the land use and land cover, as well as to detail the water resources data in the whole state, activities strongly based on the use of remote sensing.

The mapping of deforestation and the land use and land cover was carried out on a scale of 1:100,000 for the year 2007, considered ground zero in the monitoring process of this type of information in the state of Mato Grosso do Sul. The activities were carried out by Embrapa Digital Agriculture using CBERS satellite images from the High-Resolution Imaging Camera (Charge Coupled Device – CCD), with a spatial resolution of 20 m, aboard the CBERS-2B platform. The mapping was individualized into 64 classes (Silva et al., 2011a) at level III, according to the Brazilian vegetation classification system, which can be aggregated into 15 classes of phytophysiognomies, considered level II, namely: 1- Riparian Vegetation (tree, shrub, herbaceous); 2-Seasonal Semideciduous Forest; 3-Deciduous Seasonal Forest; 4-Savanna Cerrado; 5-Savanna Steppe (chaco); 6-Pioneer Formations (tree, shrub and herbaceous); 7-Floristic Contacts (ecotone and enclave); 8-Vegetational Refuges (relic communities); 9-Secondary Vegetation; 10-Annual Agriculture; 11-Agriculture; 12-Semi-perennial Agriculture; 13-Forestry; 14-Livestock (planted pasture); and 15-Other Anthropogenic Areas (urban influence, mining, occupied floodplains).

Special attention was given to the identification and mapping of cultivated pastures, which in the state of Mato Grosso do Sul represented 16 million hectares. Of these, about 57% were in different stages of degradation, which significantly affected the livestock economy, requiring a recovery process through revegetation work or enrichment of the area (Silva et al., 2011b). Thus, according to degradation, four levels of pasture were identified and mapped on a scale of 1:50,000, in sample areas: non-degraded (I), moderate degradation (II), strong degradation (III) and very strong (IV). This activity involved merging images of HRC sensors (High-Resolution Camera or High-Resolution Panchromatic Camera) with 2.5 m of resolution and CCD, both aboard the CBERS-2B satellite.

The information on water resources in the state of Mato Grosso do Sul was also updated and improved based on remote sensing data, at a scale of 1:100,000. This process included updating the drainage network and detailing the territorial limits of the Planning and Management Units of water resources, as well as creating slope maps, based on the classes defined by the National Council for the Environment (Conama). Such information, in addition to supporting the spatial analysis of the surroundings of the project whose licensing was requested, also helps other sectors or areas of the government. For example, altimetry and slope data assist in the application of the Brazilian Forest Code in relation to Permanent Preservation Areas (APP) on slopes and hilltops; the delimitation of watershed basins or planning and management units and the updating of the drainage network assist in applying the state water resources plan. Both are also inputs for the elaboration of the Ecological-Economic Zoning (ZEE) of the state.

Another important result of the GeoMS project included training the team at the Mato Grosso do Sul Environment Institute (Imasul), in order to use Sisla and to use satellite images and mapping of vegetation cover and land use. The results are widely used by managers, consultants, environmental inspectors, environmental analysts, entrepreneurs, researchers, teachers, students, among others.

Web-based geographic information systems

GIS has been used for many years to support environmental decisions, providing users with easy visualization, storage and analysis of geospatial data. This is its main difference in relation to conventional information systems: its ability to store both descriptive attributes and geometries of the different types of input data (Casanova et al., 2005). Through GIS it is possible to insert and integrate, in a single database, textual spatial information, satellite images, data obtained from GPS, vector grids and thematic maps, in addition to offering mechanisms to combine this information and provide spatial analysis, whose results can guide decision making in various applications. This tool is available in various computer packages, with commercial and free options, compatible with various raster and vector data.

Over the years, GIS has incorporated new technologies and adopted innovative architectures, following the evolution of the areas of computing and system development. It was from this evolutionary process that Web Geographic Information System (WebGIS) emerged, whose purpose is the same as conventional GIS, but which can be accessed through an internet browser. In a WebGIS, the information system is implemented in the client-server architecture, in which the "server" part is responsible for the processing and storage of geospatial data and the "client" part corresponds to a web browser, which presents an interface for visualizing the data and interaction with the system (Smith et al., 2020). An important feature of a WebGIS is the way communication is executed between the client's graphical interface and the server's geographic DBMS, which stores geospatial information. Therefore, an intermediary layer is used between the application and the geographic DBMS, which basically consists of map servers. These servers can be seen as APIs capable of collecting data directly from geographic DBMS and delivering them, in a standardized way, to the application accessed by the end user through services such as those implemented by the Open Geospatial Consortium (OGC). Thus, map servers allow the use of the same set of geographic data by different applications.

With the adoption of this client-server architecture, WebGIS enables access to its resources by any user connected to the internet, without having to install specific programs on their computer or acquire geospatial databases. A WebGIS can be characterized both as a tool for the generic treatment of geospatial data, similarly to a traditional GIS, or as an environment for exploring a pre-defined data set, which allows to carry out operations designed to meet specific user profiles.

Regarding this topic, we present some WebGIS solutions developed under the coordination of the technical team of Embrapa Digital Agriculture to meet government demands related to the monitoring of land use and land cover, environmental licensing and the organization of geospatial information for the ZEE of the Amazon states.

GeoPortal and WebGIS TerraClass

The current collection of maps of the TerraClass Project, presented in item "2.2.2 Land use and land cover monitoring – TerraClass Project", supports the formulation of public policies, territorial management and actions related to the preservation of national biodiversity, and the maintenance of the quality of ecosystem services from different sectors of civil society and the Federal Government. The mappings have been available on the internet since their release, but in raw format, which required users to download them to a computer and use a GIS to process and analyze them. Furthermore, considering the very diverse audience of users for this type of information, not always familiarized with geoprocessing, there was a growing demand for the development of user-friendly tools that could provide simple mechanisms to analyze territorial dynamics in the Legal Amazon.

Given the large volume of existing geospatial data, as well as operational difficulties that required previous computational processing to organize and adapt the data to the users' areas of interest, the GeoPortal TerraClass was developed (Figure 4). This digital platform consists of: a landing-page with a

Figure 4. Landing-page of GeoPortal TerraClass.

Source: GeoPortal TerraClass (2020).

general presentation of the TerraClass Project; description and characterization of the legend of thematic mapping classes; a mechanism for downloading the raster or vector data, in different geographic areas; and access to WebGIS TerraClass, which is a technology to facilitate analysis on the computer screen.

WebGIS TerraClass (Figure 5) is a web-based geoinformation system that provides interactive tools for visualization, analysis and interoperability of geospatial data on land use and land cover identified in deforested areas of the Legal Amazon. The system is intended for the common user up to the technician with advanced knowledge in geotechnologies, providing simple comparisons and complex spatial analysis across the entire historical series of available mappings. To use the tools, the user must select an object associated with the available land areas, which can be from the entire Legal Amazon, Federation Units, Municipalities or Water Planning Units (UPH).

The visualization tools allow to select the land use and land cover mapping year; compare side-by-side mappings between pairs of years by scrolling the map with a slider that shows each year on one side of a vertical line; and observe mapping details by increasing the approximation in a given area of interest.

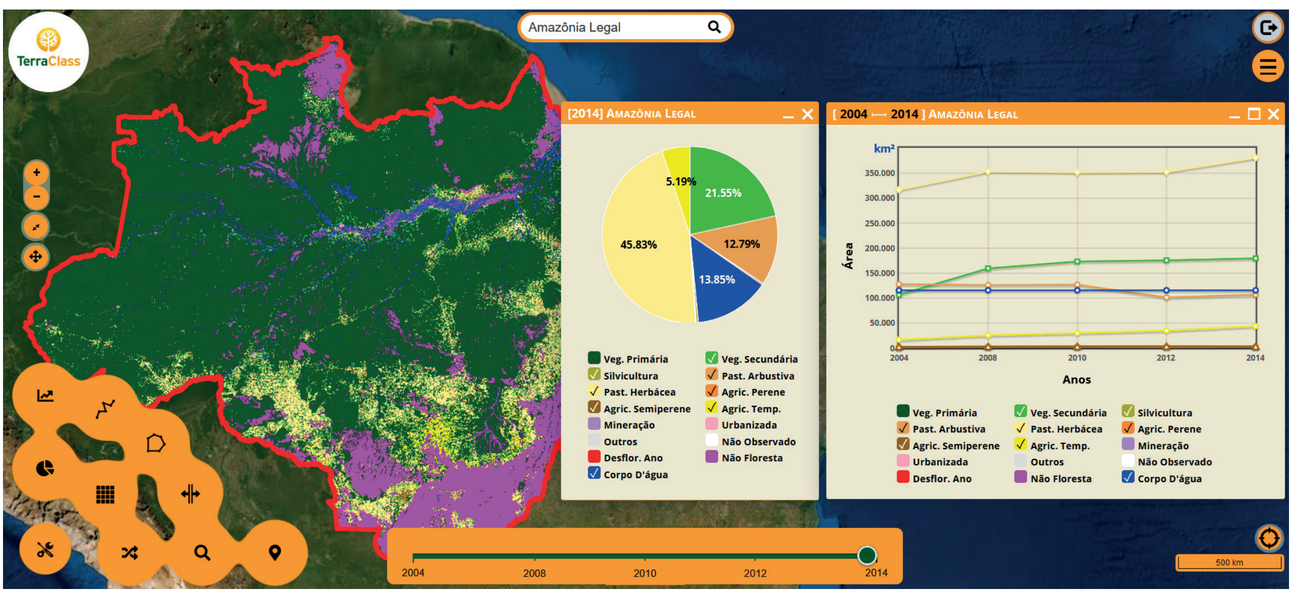


Figure 5. WebGIS TerraClass featuring the 2014 mapping pie chart and the 2004 to 2014 evolution chart for the Legal Amazon.
Source: GeoPortal TerraClass (2020).

The analysis tools generate graphs of map sectors that represent all classes of land use and land cover by slices proportional to their respective frequencies; transition matrices that presents the probabilities of land use and land cover classes to remain in the same classes or to change to other ones, showing the temporal dynamics in any combination between pairs of years; transition diagrams based on Sankey's flowchart (Schmidt, 2008), in which nodes constitute the area of a given thematic class in a year, and the edges represent the flow of area transitions between classes over the years, enabling the spatialization of changing areas on the screen ; and the evolution graph, that shows the variation of the area of all thematic classes across the historical series of mappings.

Figure 6 exemplifies the Sankey transition diagram for the state of Mato Grosso based on the selection related to the years 2004, 2010 and 2014, enabling a detailed analysis of land use and land cover transitions during the period. The transition values between thematic classes of interest can be observed by positioning the cursor over the lines. In this example, the line highlighted in cyan represents the

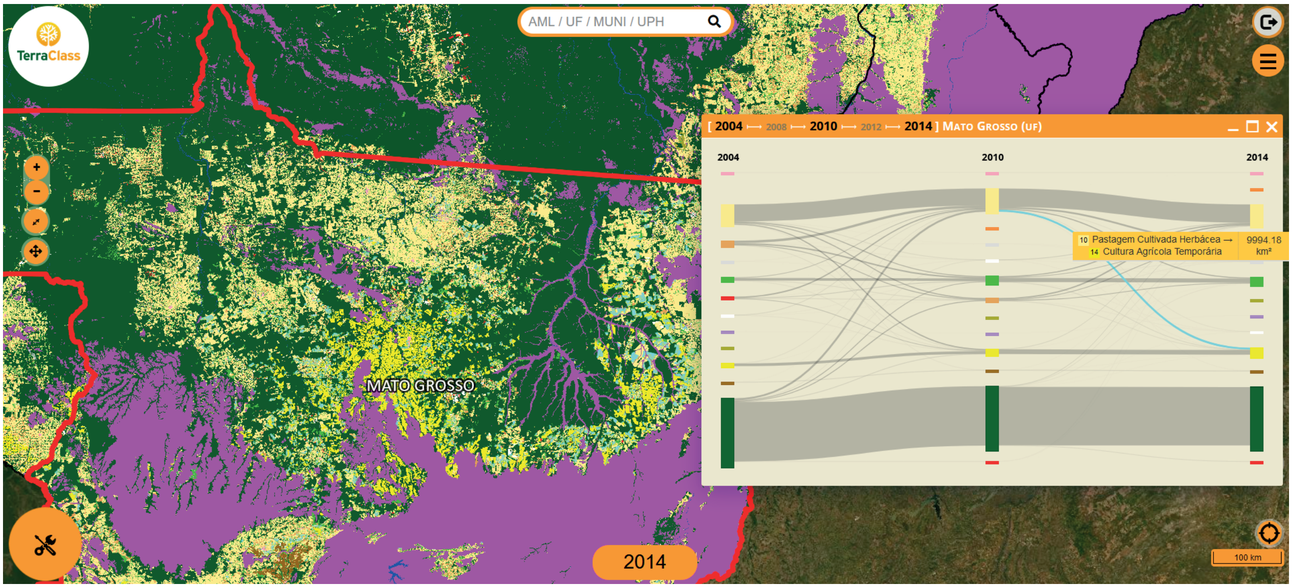


Figure 6. Sankey transition diagram generated for the years 2004, 2010 and 2014 in the state of Mato Grosso.

Source: GeoPortal TerraClass (2020).

change of 9,948.18 km^2 from the Herbaceous Cultivated Pasture thematic class in 2010 to the Temporary Agricultural Crop class in 2014, which also appears spatialized over the 2014 mapping with the same color, indicating the increase in agricultural areas associated with the conversion of areas previously destined for livestock.

The TerraClass WebGIS has interoperability with SATVeg, accessed by selecting a point in the map, which automatically activates an application integration bridge and which will display in SATVeg the temporal profile of MODIS vegetation index in the respective geographic location, expressing changes in plant green biomass over time.

The TerraClass GeoPortal[2] and its target audience is data users who need to know and analyze the dynamics of land use and cover in the deforested areas of the Legal Amazon, with a focus in territorial management such as public managers, researchers, teachers, students, self-employed professionals, extension workers, association of producers, rural unions, among others.

Interactive Environmental Licensing Support System (SISLA)

The Interactive Environmental Licensing Support System (SISLA) is a WebGIS application that allows the collection, organization, integration and management of georeferenced information related to environmental licensing processes by governmental agencies. The main objective of this application is to enable governmental agencies to fully monitor the environmental licensing process in its various stages, ranging from georeferenced analysis around the enterprise requesting the licensing, in order to support decision-making. The greatest benefit provided using applications such as SISLA is the ability to provide greater flexibility, transparency and security in the procedural protocols, as the analyses are based on official government data and are freely available in the application itself. Thus, the applicant requesting an

[2] Available at www.terraclass.gov.br

environmental licensing can check, via the internet, whether or not this enterprise is in geographical and environmental compliance, even before submitting the process to the responsible government agency.

SISLA[3] was implemented in Mato Grosso do Sul in October 2008, when the state started to use a WebGIS as a supporting tool to analyze processes involving environmental licensing requests for activities in sectors such as infrastructure, agropastoral, mining, tourism, industry, sanitation, fishing, fauna and forests.

Specifically for the agropastoral sector, activities such as construction of dams, establishing irrigation infrastructure, feedlots for cattle, horses, sheep and goats, and the construction of silos, warehouses and tanks for aquaculture are subject to environmental licensing in the state of Mato Grosso do Sul, as defined by the Mato Grosso State Institute of Environment (IMASUL).

The development of SISLA included the use of free software focused on the geotechnology theme, such as I3Geo (Speranza et al., 2011; Brazilian Public Software, 2020). This allows its adaptation, at low cost, to the demands and local needs of each government, and also its integration with corporate systems for the procedural protocols. The main module of the system – the analysis and generation of the project's surroundings report – allows the applicant requesting an environmental licensing to send its property's georeferenced maps, which contain the enterprise's boundaries, permanent protection areas and related issues, in accordance with the regulation by the government agency. Based on these maps, SISLA performs geospatial queries to verify the proximity or intersection of the project/enterprise boundaries with the areas protected by the government, such as indigenous lands and conservation units, in addition to slope analyses for the areas in the project (Figure 7), generating spatial analysis reports (Figure 8).

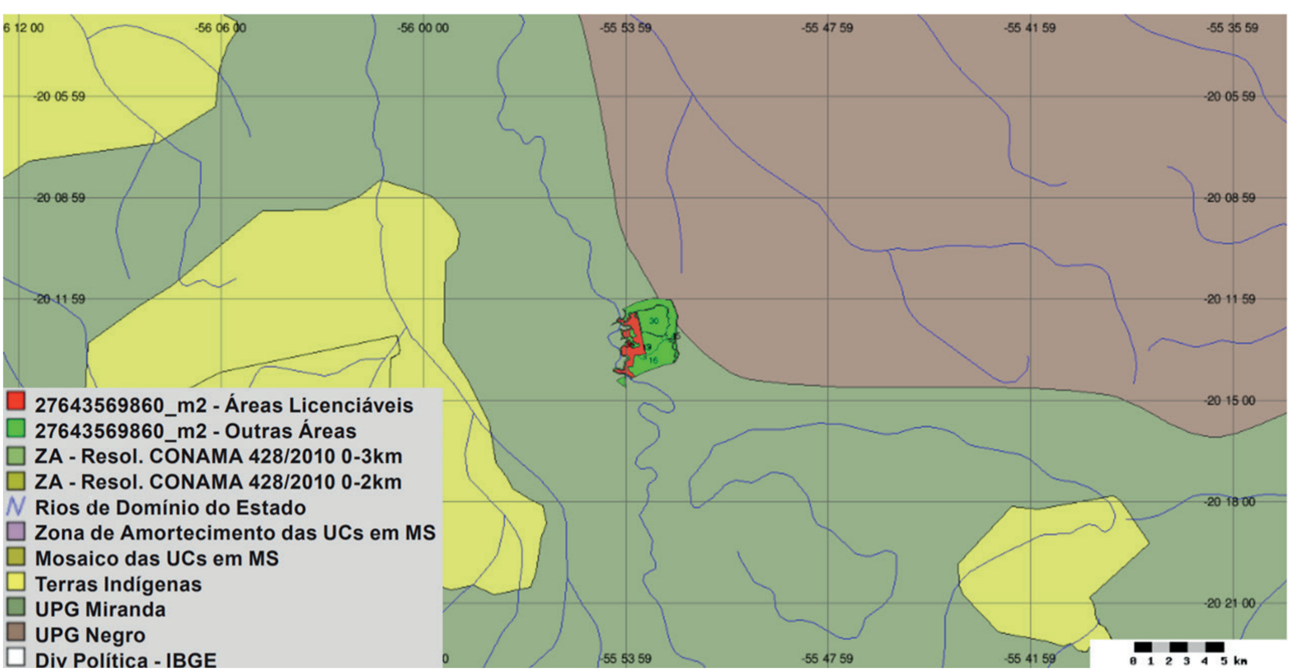


Figure 7. Map containing the property boundaries and the areas protected by the government.

Source: Silva et al. (2011b).

[3] Available at: sisla.imasul.ms.gov.br

Conservation Units, Indigenous Lands, Buffer Zones, ZAs (Conama Resolution 428 0-2km) and ZAs (Conama Resolution 428 0-3km) inside the property

	Name	Distance
Nothing found		

Conservation Units, Indigenous Lands, Buffer Zones, ZAs (Conama Resolution 428 0-2km) and ZAs (Conama Resolution 428 0-3km) that fully contain the property

	Name	Distance
Nothing found		

Conservation Units, Indigenous Lands, Buffer Zones, ZAs (Conama Resolution 428 0-2km) and ZAs (Conama Resolution 428 0-3km) close to 15 km from the property

	Name	Distance
Conservation Units		
Nothing found		
Indigenous Lands		
Taunay-Ipegue	36 - Project Area for Provisional Legal Reserve Registration (APTAP)	11.012 km
Taunay-Ipegue	36 - Project Area for Provisional Legal Reserve Registration (APTAP)	7.061 km
Buffer Zones		
Nothing found		
ZAs (Conama Resolution 428 0-2km)		
Nothing found		
ZAs (Conama Resolution 428 0-3km)		
Nothing found		

Figure 8. Example of a report generated by SISLA, with the identification of protected areas inside or close to the property.

Source: Silva et al. (2011b).

Other SISLA modules were developed for internal and exclusive access by technicians from government agencies, such as registration and consultation of the procedural progress per the activity requested; geospatial consultation for generating management reports containing the spatial distribution and the procedural processes; and the visual and individual technical analysis of processes that allows the specialist to issue an opinion for approving or not the process (Figure 9).

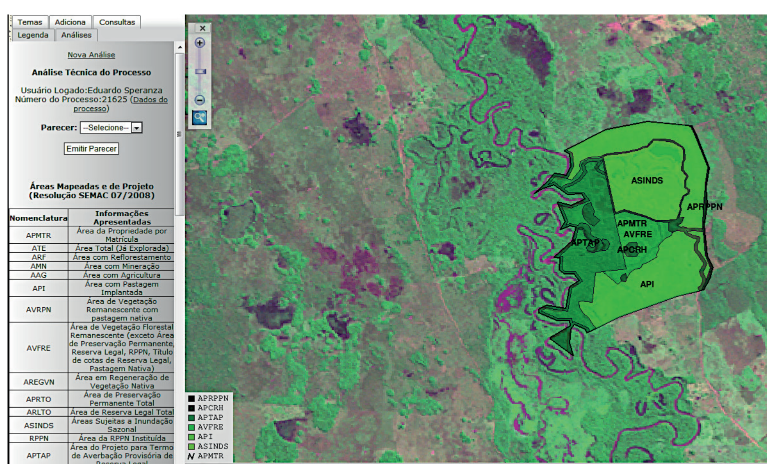


Figure 9. Example of visual and individual technical analysis of the environmental licensing process provided by SISLA.

Source: Silva et al. (2011b).

In almost 12 years of operation in the state of Mato Grosso do Sul, SISLA'S registered users include professionals from government agencies, private companies, consultants, universities and banks, who use the system to request environmental licenses, as well as to access its geospatial information. Several regulations published by the state government, related to environmental licensing, include analysis carried out by

SISLA as a prerequisite in all phases of the environmental licensing process. Thus, it is possible to affirm that its implementation in the state of Mato Grosso do Sul was successful and can be extended to other municipal, state and federal bodies.

Interactive Geospatial Analysis System for the Legal Amazon (SIAGEO Amazônia)

The Legal Amazon is a Brazilian region established by federal law with the objective of empowering economic planning and for defining public policies and protecting biodiversity in this part of the national territory. This region covers the states of Acre, Amapá, Amazonas, Mato Grosso, Pará, Rondônia, Roraima, Tocantins and part of the state of Maranhão, with a total area of 5,217,423 km^2, that is, almost 60% of the country's territory.

The occupation of this territory intensified in a disorderly manner from the 1970s onwards, through the implementation of production systems incompatible with the sustainability of the region's natural resources. The ZEE, an instrument of the National Environmental Policy, was then established by the Federal Government in order to achieve sustainable development by making socioeconomic development compatible with environmental protection. It was elaborated based on the diagnosis of the physical, biological, socioeconomic and legal-institutional means and on establishing exploratory scenarios for the proposition of guidelines for each identified territorial unit, including actions aimed at correcting existing harmful environmental impacts. The economic, social, environmental and cultural specificities, the vulnerabilities and potential of each state, resulted in obtaining different guidelines and procedures, in their ZEEs, for planning the occupation and use of their territories (Brasil, 2020).

The project Ecological-Economic Zoning Uniformization of the Legal Amazon and Integration with Agro-ecological Zoning of the Region (UZEE) was carried out by Embrapa with the objective of integrating the different state ZEEs and providing a global characterization of the Legal Amazon, capable of subsidizing the construction of macro-regional public policies consistently and independently of state boundaries. The leadership of this project was directed by Embrapa Eastern Amazon, with the participation of Embrapa Digital Agriculture, and with the support of the Ministry of the Environment (MMA) and the states covered by the Legal Amazon for the access and data validation used in the construction of the state ZEEs. The project had financial support from the Financier of Studies and Projects (FINEP) of the Ministry of Science, Technology and Innovation (MCTI).

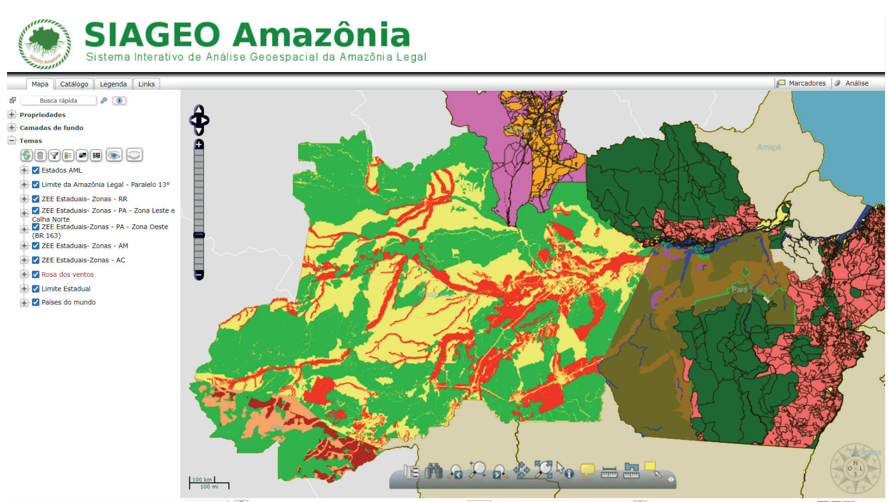


Figure 10. SIAGEO Amazon screen.

Source: Embrapa Agricultural Informatics (2020b).

To meet the need for organization, storage and availability of geospatial information of the project as well as the data from the state ZEEs and the respective basic data used to build them, the Interactive Geospatial Analysis System of the Legal Amazon (SIAGEO Amazônia) was developed, as shown in Figure 10.

SIAGEO Amazon[4] enables users to visualize and obtain interactive georeferenced maps and tabular data through the manipulation of different levels of information according to the user's interest and need. Similarly, it allows the user to perform spatial analysis based on a vector reference data, such as points, lines or polygons, and a set of thematic maps that are structured in WebGIS, in addition to providing access to technical documents and legal frameworks for each zoning initiative.

Currently, the system has a total of approximately seven thousand layers of geospatial data, both in vector and raster format, properly cataloged and organized so they can be viewed directly in the tool, downloaded for user consumption or used as parameters for spatial analysis on regions of interest.

SIAGEO was developed based on free software, and its basic functionalities are derived from the I3Geo software. The system has modules for generating two types of environment reports, which allow characterizing the neighborhood of a region of interest. The first one is the Spatial Analysis Module, which allows a user to send the geospatial data of their area of interest and obtain a neighborhood report regarding a set of themes selected from those already cataloged in the system. The Banking Module, on the other hand, offers a similar approach, but the themes used for the neighborhood analysis are pre-selected and placed in those determined by the banking institutions as relevant for the analysis of project financing.

Final considerations

In recent decades, Brazilian agriculture has experienced intense processes of structural, technological, economic, social and environmental transformations, which have reconfigured the activities of the sector and its territorial dynamics. In this chapter, some initiatives and products of Embrapa Digital Agriculture were presented. They directly or indirectly meet the demands related to the theme of land management in Brazil, in which agriculture undertakes a large role due to its dynamic characteristic and outstanding contribution to the outcome of the national Gross Domestic Product. All mentioned technological solutions, developed using free software, are public and with unrestricted use, guaranteeing their access and use by all of society.

Geotechnologies are essential tools for the generation of data and information that contribute to discussions and official positions of governments in different forums related to strategic management of the national territory, as well as facing crises and decisive questions for the defense and maintenance of national sovereignty.

New scientific challenges arise for the treatment, organization and availability of volumes of geospatial data, produced at increasing speed and quantity as a result of the evolution and emergence of new geotechnologies. Future perspectives indicate that, in the process of digital transformation, geospatial data will be increasingly present in people's daily lives, increasing the demand for geospatial services and solutions to improve production and decision-making processes related to territorial intelligence and the management of landscape dynamics.

References

ALMEIDA, C. A. de; COUTINHO, A. C. C.; ESQUERDO, J. C. D. M.; ADAMI, M.; VENTURIERI, A.; DINIZ, C. G.; DESSAY, N.; DURIEUX, L.; GOMES, A. R. High spatial resolution land use and land cover mapping of the Brazilian Legal Amazon in 2008 using Landsat-5/TM and MODIS data. **Acta Amazonica**, v. 46, n. 3, p. 291-302, July/ Sept. 2016. DOI: 10.1590/1809-4392201505504.

[4] Available at: www.amazonia.cnptia.embrapa.br

BRASIL. Ministério do Meio Ambiente. **Zoneamento ecológico-econômico**. Available at: www.mma.gov.br/gestao-territorial/zoneamento-territorial.html. Accessed on: 8 May 2020.

BRAZILIAN PUBLIC SOFTWARE. I3Geo. Accessed on: https://softwarepublico.gov.br/social/i3geo. Acesso em: 8 May 2020.

CASANOVA, M.; CÂMARA, G.; DAVIS, C.; VINHAS, L.; QUEIROZ, G. R. (ed.). **Bancos de dados geográficos**. Curitiba: MundoGeo, 2005. Available at: www.dpi.inpe.br/livros/bdados. Accessed on: 3 Oct. 2014.

COLWELL, R. N. (ed.). **Manual of remote sensing**: interpretation and applications. 2nd ed. Falls Church: American Society of Photogrammetry, 1983. v. 2, 2440 p.

EMBRAPA AGRICULTURAL INFORMATICS. **Sistema de Análise Temporal da Vegetação (SATVeg)**. 2020a. Available at: www.satveg.cnptia.embrapa.br. Accessed on: 4 Aug. 2020.

EMBRAPA AGRICULTURAL INFORMATICS. **Sistema Interativo de Análise Geoespacial da Amazônia Legal (SIAGEO Amazônia)**. 2020b. Available at: www.amazonia.cnptia.embrapa.br. Accessed on: 4 Aug. 2020.

ESQUERDO, J. C. D. M.; ANTUNES, J. F. G.; COUTINHO, A. C.; SPERANZA, E. A.; KONDO, A. A.; SANTOS, J. L. SATVeg: a web-based tool for visualization of MODIS vegetation indices in South America. **Computers and Electronics in Agriculture**, v. 175, 105516, Aug. 2020. DOI: 10.1016/j.compag.2020.105516.

GEOPORTAL TERRACLASS. **O projeto TerraClass**. Available at: www.terraclass.gov.br. Accessed on: 4 Aug. 2020.

HUETE, A.; JUSTICE, C.; LIU, H. Development of vegetation and soil indices for MODIS-EOS. **Remote Sensing of Environment**, v. 49, n. 3, p. 224-234, Sept. 1994. DOI: 10.1016/0034-4257(94)90018-3.

JENSEN, J. R. **Remote sensing of the environment**: an earth resource perspective. 2nd ed. Upper Saddle River: Prentice Hall, 2007. 592 p.

LI, Z.; YAN, H. Transformation in scale for continuous zooming. In: GUO, H.; GOODCHILD, M. F.; ANNONI, A. (org.). **Manual of digital earth**. Singapore: Springer Singapore, 2020. p. 279-324. E-book. DOI: 10.1007/978-981-32-9915-3_8.

ROUSE, J. W.; HAAS, R. H.; SCHELL, J. A.; DEERING, D. W. Monitoring vegetation systems in the Great Plains with ERTS. In: EARTH RESOURCES TECHNOLOGY SATELLITE SYMPOSIUM, 3., 1973, Washington, DC. **Proceedings**. Washington, DC: National Aeronautics and Space Administration, 1974. v. 1: technical presentations, p. 309-317. Available at: https://ntrs.nasa.gov/citations/19740022592. Accessed on: 4 Aug. 2020.

SCHMIDT, M. The Sankey diagram in energy and material flow management: part II: methodology and current applications. **Journal of Industrial Ecology**, v. 12, n. 2, p. 173-185, 2008. DOI: 10.1111/j.1530-9290.2008.00015.x.

SILVA, J. dos S. V. da; POTT, A.; ABDON, M. de M.; POTT, V. J.; SANTOS, K. R. dos. **Projeto GeoMS**: cobertura vegetal e uso da terra do Estado de Mato Grosso do Sul. Campinas: Embrapa Informática Agropecuária, 2011a. 64 p.

SILVA, J. dos S. V. da; SPERANZA, E. A.; VENDRUSCULO, L. G.; ESQUERDO, J. C. D. M.; MAURO, R. de A.; BIANCHINI, S. L.; FLORENCE, R. de O. **Projeto GeoMS**: melhorando o sistema de licenciamento ambiental do Estado de Mato Grosso do Sul. Campinas: Embrapa Informática Agropecuária, 2011b. 64 p.

SMITH, M. J. de; GOODCHILD, M. F.; LONGLEY, P. A. **Geospatial analysis – a comprehensive guide to principles techniques and software tools**. Available at: https://spatialanalysisonline.com. Accessed on: 28 May2020.

SPERANZA, E. A.; ESQUERDO, J. C. D. M.; SILVA, J. S. V.; ANTUNES, J. F. G.; LOURENCO, F. V.; CEZAR, V. M. SISLA- Interactive system for environmental licensing support. **Geografia**, v. 36, p. 57-72, 2011.

Scientific computing in agriculture

5

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 | Silvia Maria Fonseca Silveira Massruhá | Urbano Gomes Pinto de Abreu | Balbina Maria Araujo Soriano | Suzana Maria Salis | Márcia Divina de Oliveira | Walfrido Moraes Tomas

Introduction

Digital technologies have advanced incredibly fast, and its recent development has stimulated the acquisition of large volumes of different types of data, from the most varied sources. In agriculture, along its value generation chain, this data can include: a) omics data (genomics, proteomics, transcriptomics and metabolomics); b) acquired physicochemical attributes with spatiotemporal location through sensors; c) aerial and satellite images with spatiotemporal location; d) socioeconomic data; among others.

Similar to the data from more traditional sources, the use of this large data volume is analyzed through models and algorithms capable of extracting useful information for the decision-making process. This can occur in the development of a new biotechnological asset, land use monitoring, or in the control of a production process. Thus, scientific computing is understood as a collection of techniques, tools, and theories that encompass mathematics, statistics, physics, and computing. It also covers specific knowledge of certain sub-areas, such as applied statistics, econometrics, applied mathematics, computational intelligence, scientific visualization and biometrics. These will continue to be central in the development of new agricultural technologies, now in the context of the emerging Digital Agriculture. In recent decades, scientific computing has been identified as the third pillar of scientific research, along with experimentation and theory (Souza et al., 2017).

In the following sections, we present examples of applications that use scientific computation algorithms and techniques for the solution of problems in the agricultural sector. Section 2 presents two applications. They are based on the observation of a large raw dataset in order to recognize embedded patterns and derive knowledge and actions from such patterns to be used by an Expert System. Section 3 presents three applications based on the construction of mathematical and statistical models. These can carry out predictions and analyses from simulation scenarios in order to assist public decision-making. Through these different applications, it is possible to see that scientific computing is a research area that is eminently transversal to others.

Artificial intelligence

Artificial Intelligence is a broad area that began in the second half of the 1940s, when an artificial neural network was conceived, and which described how human neurons should learn to perform calculations. This area has undergone many modifications and has intersected with other disciplines, especially statistical modeling and various pattern recognition methods. These intersections compose a group of techniques known as intelligent systems, which are based on machine learning.

A machine learning model is supported by previously observed data coming from either databases, experiments, images, or texts. Data has attributes, which need to be described for each observation. For example, if we collect data from a pasture at different locations on the property, we will have the same attributes for each data collection, such as: location, grass type, date, pasture status (degraded, non-degraded, in degradation), geographic location, percentage of soil cover, soil type, pasture height, etc. With these attributes and the data collected, a classification model of the pasture's status can be built. If the observed data were texts, the attributes could be the words in the texts; if they were images, the attributes could be such images divided into very small pieces, or pixels, and could consider, for example, the color of each pixel.

Items "Automatic Soil Classification" and " SiBCS-based Expert System" present, respectively, examples of automatic soil classification and exploration of information in texts, making use of different artificial intelligence techniques.

Automatic soil classification

To classify a soil profile, the Brazilian Soil Classification System (SiBCS) considers a wide range of morphological, physical, chemical, and mineralogical attributes in addition to environmental aspects such as climate, vegetation, relief, parent material, hydric conditions, external characteristics and soil-landscape relationships (Santos et al., 2013).

To assist in this laborious process, Embrapa Digital Agriculture and Embrapa Soils designed two intelligent tools for automatic soil classification. The first is an Expert System that uses the SiBCs rules for soil classification. The second is aWeb system (SoloClass) for soil profile classification through a committee of intelligent solutions based on machine learning algorithms. These smart tools were developed within the scope of the project "Use of smart mobile devices in the classification of Brazilian soils – SmartSolos", led by Embrapa Soils. Both tools are presented in the following subsections.

SiBCS-based expert system

The SiBCS rules-based expert system simulates the reasoning of a domain expert when performing the classification of soil profiles. Thus, it can be used to classify soil profiles not yet classified and to validate previously classified profiles (Vaz et al., 2018).

Vaz et al. (2019a) used the expert system to analyze soil data provided by IBGE. They showed that this is an important tool for the curation of Brazilian soil data, as it allows it to be executed more efficiently and with fewer errors, benefiting soil governance in Brazil.

The advantages of making the expert system available through an API and the importance of this tool to facilitate soil data curation, while guiding a more adequate data recording, were also shown in Vaz et al. (2019b). Figure 1 shows that by making the expert system available through the API, the user can obtain the soil profile classifications from the expert system and compare them with previously known classifications. Thus, possible errors in soil data can be analyzed and corrected, making it a powerful tool for improving the quality of soil data in Brazil.

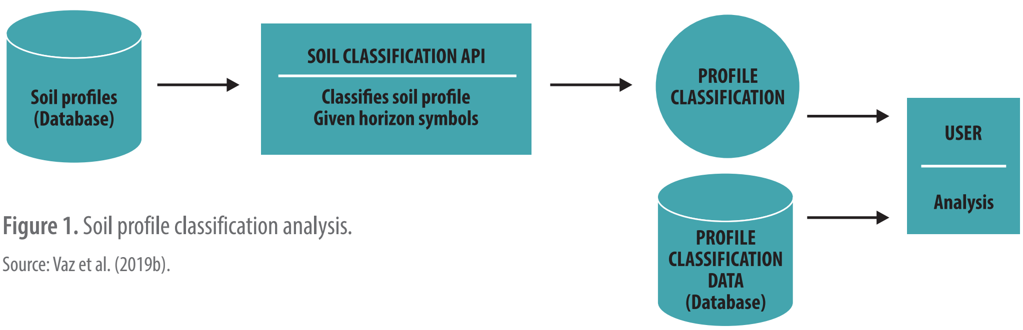


Figure 1. Soil profile classification analysis.

Source: Vaz et al. (2019b).

The great challenge of this system is codifying all SiBCS rules to treat its first four categorical levels. The classification taxonomy has more than a thousand classes between the first and fourth categorical level. In addition, the rules are quite complex, so joint work and great effort by computer and soil scientists are essential to make viable the development of such a system.

Although a specific application is being developed by Embrapa to use this soil classification API, partner institutions can also use it to create new solutions that rely on soil classification, provided their data is coded in accordance with the standards established by the system.

Regarding Brazilian soil data standards, there are different initiatives that seek to organize them. However, none of them has been consolidated as a standard, nor do they meet the needs of the expert system developed. As such, many observations could be made in relation to organizing these data in Brazil throughout this work. It is common, for example, to observe data redundancy in different fields, absence of fields necessary for recording important soil information for classification, and data representations that make difficult computational processing and data retrieval. The next step of this work is, therefore, to consolidate a series of recommendations for the structuring of Brazilian soil data in order to simplify computational manipulation, ensure higher quality of stored data, and facilitate the creation of new solutions that depend on them.

Future research is on the possibility of automating other processes that are normally time consuming or greatly increase the uncertainty of the data collected in the field. For example, color, texture, soil layer boundaries and other attributes are determined in a subjective way, according to personal interpretations made during fieldwork. The collection of this type of information can be facilitated and automated through computational tools that extract characteristics from images taken in the field.

Intelligent soil classification system

A promising alternative for automatic soil classification is the combination of machine learning (ML) algorithms with attribute selection methods. ML algorithms operate by building a model obtained from training samples to make data-driven predictions. Such data contain soil profile observations previously classified by pedologists. On the other hand, the attribute selection methods aim to find a subset of relevant variables related to the target task. It makes the learning process more efficient by simplifying the operating cost of the models, enabling to better understand the obtained results (Guyon; Elisseff, 2003).

SoloClass is an intelligent system developed for classifying soil profiles. This system allows a user to input a set of variables from one or more soil profiles, and then receives the classification of each profile according to SiBCS with a probability associated to the predicted class.

Five classes of ML algorithms were used for intelligent soil classification: a) symbolic: decision trees; b) based on instances: k-NN or k nearest neighbors; c) statistical learning: Support Vector Machines (SVM); d) bootstrap aggregation: Random Forest; and e) connectionism: Deep Neural Networks. All these algorithms were trained for the four categorical levels (orders, suborders, large groups, and subgroups) adopted by SiBCS.

The architecture of the SoloClass system is based on a classifiers committee, as shown in Figure 2.

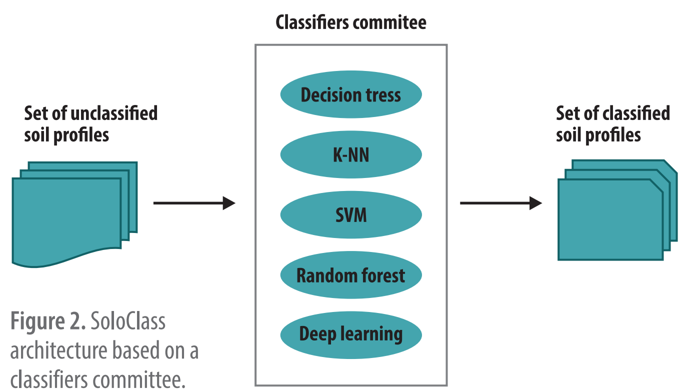


Figure 2. SoloClass architecture based on a classifiers committee.

Upon receiving a set of unclassified soil profiles, with different numbers of horizons, the user can select one or more classifiers that have been trained from a pre-classified database by pedologists (induction process). Subsequently, the system triggers the selected classifiers and stores the results presented individually. At the end of the deduction process, the classification committee (Figure 2) assigns the classification result to the soil profile by vote, that is, the classification associated with the profile is the one that obtained the highest frequency or majority vote.

This classifier committee-based architecture has some advantages such as: a) increase in the predictive power of the system due to the use of several classifiers adjusted to the data and combined for this purpose; b) reduction of variance and bias when compared to using only one machine learning method; c) extensible architecture, that is, other classifiers can be added.

The main benefits of the SoloClass system are: a) assisting national soil survey projects and programs, such as Pronasolos (Polidoro et al., 2016), acting as a facilitating tool in soil classification work; b) facilitate the understanding of soil classification for farmers, students, teachers, extension workers, and researchers; c) minimize possible human errors during the soil classification activity.

As it is a Web System, SoloClass[1] can be accessed through mobile devices or personal computers, without any operating system restrictions. This helps to expand access and the inclusion of a greater number of users. SoloClass has a responsive interface, that is, it can be characterized by the visual adaptation of a page or interface to any device on which it is viewed, without the need of a specific versions for each model.

[1] Available at: www.soloclass.cnptia.embrapa.br

Text mining in technical-scientific publications

The human learning process is based on observations, pattern formation, hypotheses, and inferences from these observations. Nowadays, there are plenty of observations, specifically, an excessive volume of data, both in databases and in published textual format. Data mining uses statistical analysis processes, in which algorithms are implemented in computer programs that can handle a large volume of data to find patterns and help formatting hypotheses and models that allow describing these patterns.

Text mining (TM) is a specialization of the data mining process. The main difference between the two processes is that, whereas conventional data mining works exclusively with structured data (pre-organized in databases or some representation, such as a spreadsheet), text mining inherently deals with unstructured data. Therefore, in TM, the first challenge is to structure the data with their respective attributes, based on the texts, so that data mining algorithms can be used.

The structuring of texts depends on the problem addressed. For example, if we want to know or relate which types of agricultural technologies are linked to the use of water in Brazilian agriculture, we can delimit a set of technical-scientific publications on the topic and extract this information. In this case, one option is the use of linguistic tools that allow identifying the vocabulary of interest (for example: irrigation, harvesting, water resources, pivot, etc.) and delimiting and disambiguating geographic locations (such as: São Francisco River, São Francisco Church, São Francisco City, etc.) in the texts.

Similar to this, in the methodology proposed by Moura et al. (2017), there is a semi-automated step-by-step process, which used software tools specifically developed for this purpose, and contained the following steps: 1) delimitation of publications of interest; 2) extraction and disambiguation of toponyms with the TopExtract tool (Takemura et al., 2013); 3) formatting a dictionary of terms of interest, manually performed by domain experts; 4) use of the ExtracTrans tool (Transaction Extraction tool) to: a) extract terms from texts by similarity and synonymy; b) creation of the transactions present in the texts (all the words of interest that appeared in the text); and c) elimination of redundant data, which does not contribute to the results; 5) pattern extraction, using machine learning algorithms, such as association rules, or even placing the results in an Excel spreadsheet or other similar software. For example, in Moura et al. (2017), 40 association rules were found for the Northeast region, among which:

If technologyClass = agricultural engineering & culture = grapes & cultureClass = fruit & region = NE → technology = irrigation.

Another application of very specific interest was to describe which quantitative and scientific computing methods were cited in Embrapa’s scientific publications. We searched among those considered of the highest level, and according to the Qualis CAPES indicator (A1, A2, B1 and B2), between the years 2000 and 2018. Embrapa has its own cataloging system for its publications and technologies (Embrapa, 2020) where the metadata´s keywords are audited. This in itself indicates its very high quality. However, two major problems are present for the study: a) the large number of publications in this interval, approximately 22,000 articles; and b) the fact that the keywords in the articles cover agricultural terms, and not necessarily quantitative methods and scientific computation terms. In other words, the keywords of interest in this data analysis were not part of the conventional keyword repertoire of these articles and, therefore, could not be located only by search results, let alone by reading each of the 22,000, which would be a very extensive task.

Thus, the methodology by Moura et al. (2017) was adapted as follows: 1) the articles of interest were already selected; 2) the geolocation process was not necessary and instead, a process was created to download the articles and convert them to plain text; 3) the domain specialists, in quantitative methods and scientific computation organized the necessary dictionary of terms for the area; 4) a tool was adapted (from ExtracTrans tool) to extract the words of interest from the text collection by similarity, and subsequently, put the data in a spreadsheet; and 5) from the data sheets in Excel format, the techniques of crossing dynamic tables, aggregation of data from other sources, grouping, selection, and filters were applied to facilitate the data exploration in different views.

Some exploratory applications on a large volume of texts make use of a process similar to that used by search engines, such as Google, Yahoo, etc. The textual collection is indexed, in which each text (data) corresponds to a row of a table and each word (attribute) to a column, it is not always necessary to know the language in which the text is written, much less if there are dependencies between words. In each cell, the frequency of a word in the text, or some derived measure, is placed. Therefore, as this table has an exaggerated number of columns and many cells with zero value, we try to reduce the number of columns, selecting the most statistically significant words or word compositions.

There are many techniques to reduce the number of columns in a table, all of which depend on what one wants to answer in relation to the collection of texts. To format a collection of texts in a table like this, we have the I-PreProc tool (Pereira; Moura, 2015). A common application for this formatting is to group texts with similar content so that they must correspond to specific topics, that is, subdivided into more related subjects, as carried out in the Compilation and Retrieval of Technical-scientific Information and Induction to Knowledge (CRITIC@) project. This initiative, developed by Embrapa, also uses other tools such as the previously mentioned TopExtract application.

In Figure 3 on the left, it is possible to see that based on a search expression in the publications database, the search results are organized hierarchically into documents groups from where statistically significant terms found in the group are considered "topics". In the middle, there is the distribution of accumulated frequencies for the group over time. These are represented by "*Tractor, Effect, Term, Difference, Applied, Leaf, Pruning*". To the right, the locations mentioned in these documents. This result of data exploration gives us clues as to: a) how these documents could be organized according to groups; b) what the topics or set of keywords of this group would be, for example "tractor, pruning, pruning applied to the leafs", that is, what an expert in the area considers most important in the presented result; and c) geographic location, more specifically, where these groups appear most significantly.

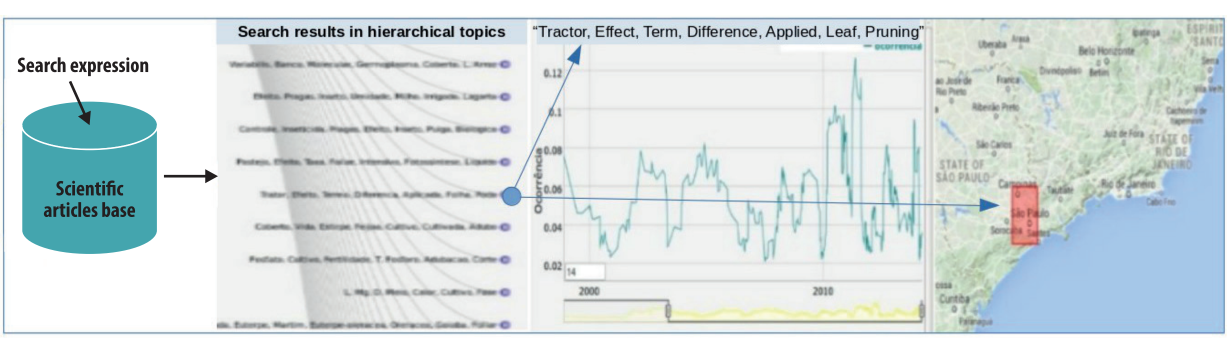


Figure 3. Example of a CRITIC@ project result.

As seen in the cited applications, text-mining processes, whatever the techniques, help the exploration and identification of information in a large volume of texts.

Mathematical and statistical modeling

Mathematical modeling is an even broader area than Artificial Intelligence. It uses a small framework of mathematical solutions, the same occurring with statistical modeling. The general idea for the modeling process is the simplified interpretation of a phenomenon, which is then described in mathematical language. Subsequently, it allows simulations to be carried out in a computer. Thus, the users of the model are positioned as experimenters in the real world, and based on the results of several computer simulations, they can understand details of the phenomenon in situations not experienced in practice. In agricultural research, for example, mathematical and statistical models are essential to complement biological experiments, allowing the study of disease dynamics in the field from computer simulations, that is, without environmental impact and with a great economy of resources. As an example, item "Modelagem da dinâmica de dispersão do HLB do citros" presents a simulation model for analyzing the intra-orchard dispersion of the disease known as citrus HLB.

In order to understand the difference between mathematical and statistical models, it can be considered a simple example, such as the mathematical equation that represents a straight line in a Cartesian plane (x, y), given by:

$$y = a + bx$$

In which a is the slope of this line and b the factor that correlates each value of x to exactly a value of y in this plane. On the other hand, if it is observed the weight and height values of a group of people, it is known a priori that the behavior of the observed points (weight, height) is linear, that is, it can be represented by a line, but it does not correspond exactly to the weight and height ratio of the entire population, that is, this set of points is just a sample of this population. A good sample should be randomized, so each person is randomly drawn for weight and height measures and has a statistically reasonable size. Thus, with this collected sample, the behavior of the population for the problem under study (weight, height) is estimated, which is a line composed of estimated values of the slope of the line (average of the observed values of weight), and the factor that correlates height with weight. This process considers model errors and estimates depend on probability distributions. In item "Genetic evaluation of livestock", a multivariate linear model is presented, that is, several dependent variables (which would replace weight) and several independent variables (which would replace height), and also: a) fixed effects, which are the of the factors that can be observed, such as the height in our example; and b) random effects, which are not observed in the sample collection, but need to be estimated by the model.

Another framework within mathematical modeling is inductive logic models, for example, "if A is a stable then A has horses". Among these models are fuzzy logic. For example, if we have a glass of water, it can be full, half full, half empty, or empty, according to the interpretation of each person looking at the glass. So, you can form rules, such as the glass is empty if it has 0 to 20 mL of water, it is half empty if it has 10 mL to 100 mL of water, it is half full if it has 50 mL to 160 mL of water, and it is full if you have more than 140 mL of water. To solve a classification of how each cup is, a system based on fuzzy rules is developed. Item "Sustainable Pantanal Farm" shows an application of systems based on fuzzy rules that assist decision making regarding sustainability in Pantanal farms, considering environmental, social, and economic values.

Modeling the citrus HLB dispersion dynamics

Brazil is the world's largest producer of oranges, with the 2020/2021 harvest being estimated at almost 288 million boxes (40.8 kg) (Fundecitrus, 2020). The disease known as huanglongbing (HLB) or greening, identified in the country in 2004, is currently the most important for the national citrus industry. Citrus HLB is caused by the bacterium *Candidatus Liberibacter asiaticus* and transmitted in Brazil mainly by the psyllid *Diaphorina citri*, which acquires the bacteria by feeding on the sap of infected plants, later transmitting them to healthy plants.

Due to its importance to the national economy, Embrapa has been developing biomathematical tools to assist in monitoring, sampling, detecting, and eradicating HLB from citrus since 2012. Initially, a deterministic compartmental mathematical model was developed (Vilamiu et al., 2013) to assess the impact on decreasing population levels of the insect vector *D. citri* in the Recôncavo Baiano region. More specifically in areas where citrus and alternative hosts are planted (orange jessamine – *Murraya paniculata*) in different proportions, aiming to collaborate with public policies for the sector. In this study, citrus and myrtle populations were divided into compartments (susceptible, exposed, infected and recovered plants), and the general characteristics of each compartment were expressed through mathematical equations in order to analyze HLB propagation temporal dynamics.

More recently, a new modeling approach based on simulation scenarios with different spatial configurations of orange jessamine and citrus was used to assess, among other aspects, the role of orange jessamine as a push or pull factor on vector insects in cultivated areas (Barbosa, 2015). For this purpose, individual-based modeling (IBM) (Grimm; Railsback, 2005) was used, and considers in the model the presence and particularity of each individual of the populations involved, while observing the final system as the result of interactions between the individuals of different populations. The IBM approach is adequate for the objectives of the study because it allows one to jointly explore the temporal and spatial aspects of the "host-insect vector-HLB" system in a more intuitive and flexible way than classical mathematical models such as those used in Vilamiu et al. (2013).

The IBM was developed in Python programming language and considers a standard agricultural landscape of the Recôncavo Baiano, containing 9 plots with 20 x 42 plants in each plot (total of 840 host plants per plot or 7,560 plants in the landscape), spacing between rows of 6 m and spacing between columns of 4 m, totaling an area of 120 m (width) x 168 m (length), just over 2 ha.

In order to analyze the intra-orchard dispersion of the insect vector and the propagation of HLB, 3 different landscapes were tested and compared: a) Scenario 1: only citrus; b) Scenario 2: citrus and myrtle around the entire area; c) Scenario 3: citrus and myrtle on the edges of each plot. Thus, the populations considered in the IBM and involved in the computer simulations are: a) main host plant (citrus); b) alternative host plant (myrtle) for testing the repulsion and attraction effect; c) *D. citri* insect vector in the nymph stage; d) adult vector insect.

In the execution of the model simulations, the user can choose different values for the following biological parameters obtained from studies and biological experiments conducted at the Embrapa Cassava & Fruits (Cruz das Almas, Bahia, Brazil) experimental fields:

- time of the disease incubation phase in the plant: 180 to 540 days;
- duration of the latency phase of the disease in the plant: 30 or 60 days;
- proportion of insects per plant: 0.41 to 5;

- simulation time: 1, 2, 5, 10 or 20 years;
- simulation mode: 1 (single) or 2 (multi);
- probability of primary infection (PIP), according to the incidence in the region: 0.01 (low), 0.15 (medium) or 0.30 (high);
- probability of detection of the disease in the field by the human inspection: 0 or 0.476.

The simulations start with all healthy plants and the arrival of a certain proportion of infective insects, according to PIP values. Populations evolve stochastically over time (according to the probability of occurrence) from processes such as birth and death of nymphs and adult insects, infection of host plants, acquisition of bacteria by nymphs and adult insects, reproduction, and flight of adult insects.

At the end of the computer simulations, two types of results are generated: a) "single" simulation type: at every 10 days of the simulation execution, a file with the status of populations in each position of the planting area is generated (type of host, infection status, number of insects in position); b) "multi" simulation type: at the end of 100 automatic executions (Monte Carlo process), graphs of the number of susceptible, infected and symptomatic plants are generated over time.

The MBI results are saved, and a software developed in Java language allows the visualization of the model results via Web. This is illustrated by the examples shown in figures 4, 5, and 6, related to the 3 simulation scenarios that represent different landscapes (different configurations and proportions of citrus and myrtle) for a "single" simulations type.

Figures 4, 5, and 6 show the dynamics of HLB spread, after a certain number of days from the start of the simulation, which occurs from the arrival of insects in random plants of the first two left columns of the

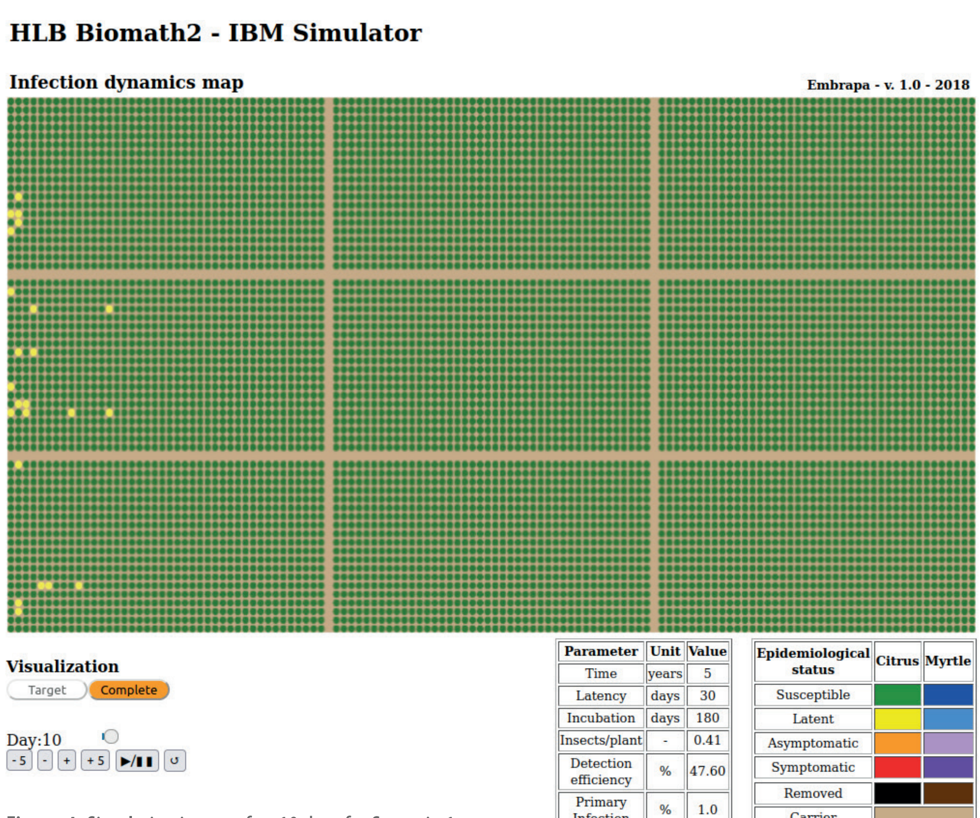


Figure 4. Simulation image after 10 days for Scenario 1.

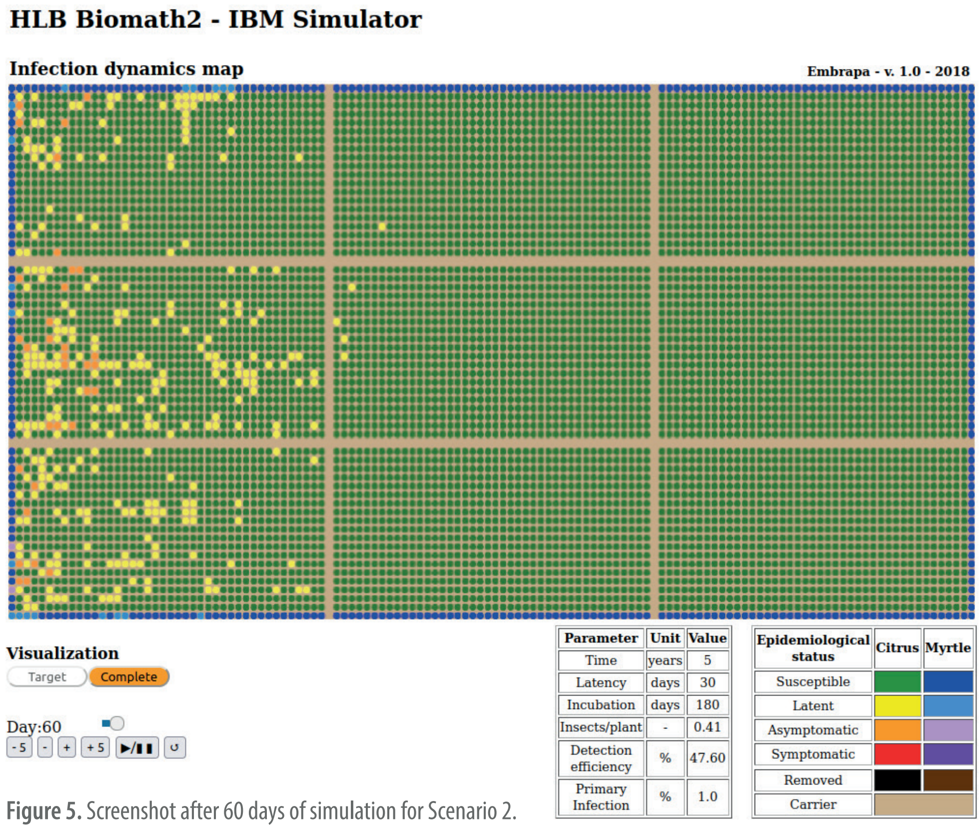


Figure 5. Screenshot after 60 days of simulation for Scenario 2.

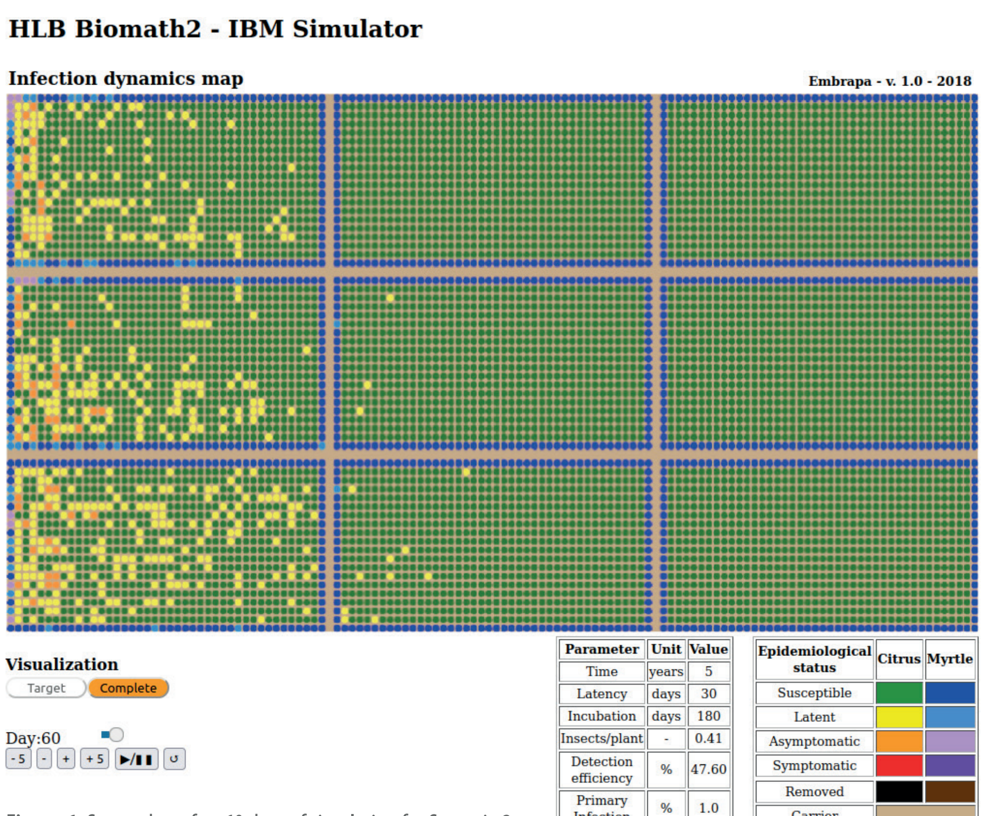


Figure 6. Screenshot after 60 days of simulation for Scenario 3.

fields. The amount of infective insects arriving in this area depends on the proportion of insects per plant and the PIP value chosen by the user. For example, for the proportion of 0.41 insects per plant, there are 1,033 insects at the beginning of the simulation, of which: a) for PIP = 0.1%: 1 infective insect; b) for PIP = 1%: 10 infective insects; c) for PIP = 15%: 154 infective insects.

Scenarios 1, 2, and 3 were tested separately in numerous combinations of the aforementioned parameters for the repulsion and attraction analysis. No visual differences were found in graphics generated by the “multi” execution, or the dynamics observed in the “single” executions. Following the analyses, scenarios were compared in a two by two scheme, and several simulations were performed for each scenario. Statistical tests were performed to compare the time of arrival of the disease in the target plot as well as in all comparisons between scenarios while considering different probabilities of primary infection. It was found that, statistically, there is no difference (comparisons between PIP equal to 1% and 15%) regarding the time of disease arrival in the target plot.

The results of the simulations prove observations made in field experiments: the primary infection has much more weight in the dynamics of disease propagation than the different spatial configurations of orange jessamine and citrus in the simulation scenarios .

Thus, the main conclusion obtained is that the simple presence of the alternative host (orange jessamine) does not significantly influence the epidemic process. This leads us to question how the interaction of the “HLB-insect vector-citrus” system would be with the use of vector population control methods, such as the application of insecticides, which could significantly affect the primary infection.

At the same time, the search for a threshold value for primary infection leads us to estimate the effort of regional management in order to stabilize the epidemic process. Furthermore, vector infectivity levels can be an indicator to be used in the future for the effectiveness of control measures in regional management. This indicator can be obtained more easily than extensive surveys with infected plants.

Currently (Barbosa, 2019) the MBI is evolving by the inclusion of new alternative hosts to evaluate in repulsion and attraction configurations, as well as testing periodic insecticide control strategies which minimize the effect of primary infection on landscapes.

From the spatiotemporal dynamics observed in the citrus HLB represented in the model, it is possible to simulate complex dissemination scenarios and perform the selection of more promising repulsion and attraction configurations to control the spread of the vector insect. This may be tested in future experiments along with obtaining indicators of effectiveness, with potential for more detailed studies in other projects.

Genetic evaluation of livestock

Animal breeding programs aim to genetically improve the population in terms of economic characteristics demanded by the market, adopting appropriate indices for the production system. In short, they consider the identification and genetic discrimination of individuals in the population, the selection of those with superior traits for replacement, either male or female, and the mating between them. An integral part of these programs are the genetic evaluation processes, which consist of continuously and cumulatively collecting biometric and genealogical data from the population undergoing improvement and periodically using a genetic-statistical model to predict the genetic values of each animal. The data include observation on the expression of physical or behavioral attributes of interest to the market. These attributes are called phenotypes and pedigree data, which in other terms means the relationships that define the genealogy of the population.

Currently, the methodology used in genetic evaluations of animals is based on the theory of mixed models (Henderson, 1963), known as BLUP (Best Linear Unbiased Prediction). It basically consists of the prediction of genetic values, adjusting the data simultaneously for fixed effects and an unequal number of observations per class (Lopes, 2005). Among the advantages of a genetic evaluation using BLUP are the inclusion of complete family information through a kinship matrix; comparison of individuals with different levels of fixed effects; and simultaneous evaluation of sires, females, and progenies. Lastly, there is the evaluation of individuals without observations, missed observations and with observations in only some characteristics (Lopes, 2005). BRBLUP (Higa, 2020) is a software for genetic evaluation of animals developed by Embrapa, based on the Python programming language and associated scientific computing libraries called Scipy/Numpy and PyTables. It supports the specification of mixed model equations so that different genetic-statistical models can be specified, including the multivariate animal model (MAM), which simultaneously evaluates fixed and random effects for a set of quantitative phenotypes while taking correlations between phenotypes or random effects into account, such as genetic origin effects.

As an example to illustrate the use of BRBLUP, (Example 5.4 of Mrode (2014) an animal model with two phenotypes (bivariate) is considered: a) FAT1: fat yield in lactation period 1; b) FAT2: fat yield in the lactation period 2. Associated with each phenotype is the presence of a fixed effect referring to herd-year-season (HYS1 and HYS2). The data set is shown in Table 1: there are eight animals, numbered from 0 to 7, and only those that have observed phenotypes (animals 0, 1 and 2) appear in the pedigree (columns Sire and Dam). The residual variances are 65 for the FAT1 phenotype and 70 for the FAT2 phenotype, with the covariance between them equal to 27; the genetic variances are 35 for the FAT1 phenotype and 30 for the FAT2 phenotype, with the covariance being equal to 28.

Table 1. Dataset (columns 0, 1, 2: pedigree – columns 0, 3, 4, 5, 6: observed data).

Animal	Father	Mother	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

Source: Adapted from Mrode (2014).

To solve the model, the BRBLUP software is executed through a command line, passing a configuration file as a parameter with the model specification. The result is stored in an output file.

Table 2 presents the contents of the generated output file. It contains 4 columns (Trait: column in the data file corresponding to a phenotype; Effect: specified effect in the model; Level: level of the effect in the data file; Sol: obtained solution). In this example, the first line of the file means that the solution for level 0 of effect 1 (HYS1) for the phenotype in column 3 is equal to 175.73126996362862. The seventh line means that level 1 (animal 1) for effect 0 (genetic value) for the phenotype in line 3 is equal to -2.999142788478058.

Accuracy values, which represent the reliability of the solution obtained for the genetic value, were not presented in this example, but are always used together. Finally, another aspect not addressed refers to the fact that, currently, animal genetic improvement programs are making efforts to include genomic information in the genetic evaluation process. This has direct implications for the construction and resolution of the genetic-statistical model used.

Sustainable Pantanal Farm

Table 2. Result of genetic evaluation.

Trait	Effect	Level	Solution
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
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

In recent decades, given the globalization of the economy and the creation of competitive markets, pressures to increase the productivity of farms in the Pantanal have intensified, compromising the sustainability of their production systems due to the fragility of its ecosystems. Given this scenario, a multidisciplinary group of researchers from Embrapa Pantanal, using previous experience on the characterization of Pantanal farms (Santos et al., 2017), developed a project in partnership with Embrapa Digital Agriculture. They aim to develop a tool to assess the sustainability of beef cattle production systems in complex and dynamic regions, such as the Pantanal, so that it would be possible to verify the system's weaknesses in order to seek good sustainability management practices.

The Pantanal biome is located in the Midwestern region of Brazil (80%), also covering part of Bolivia and Paraguay. It constitutes an extensive neotropical wetland that is seasonally flooded, with a temporal and spatial variability of diversity, which is controlled by the flood pulse. This makes the region a complex, dynamic and uncertain system (Santos et al., 2017). Because it has extensive areas of natural grasslands with a predominance of forage, the Pantanal has a vocation for the extensive beef cattle ranching with low use of external inputs which has contributed to its conservation for more than two centuries. This has been the main economic activity on Pantanal farms, making it an important socioeconomic sector at the regional and national level. Considering that farms comprise about 95% of the Pantanal plain, the main challenge for decision makers is to define beef cattle production systems that do not cause major environmental impacts while bringing economic and social benefits to the local population and ensuring the conservation and sustainable use of natural resources.

In order to understand Pantanal farms holistically, aspects and indicators were defined in a hierarchical manner at both ranch and regional level and assess the beef cattle production system (Figure 7). These aspects and indicators were selected due to their practicality in representing and simplifying complex and systemic phenomena. Some of

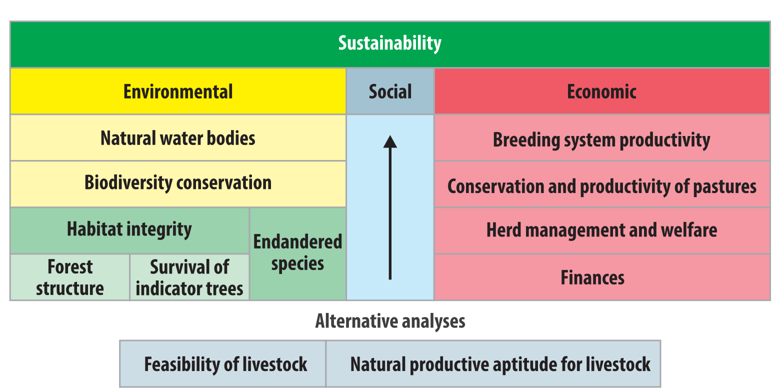


Figure 7. Hierarchical structure of the Sustainable Pantanal Farm.

the indicators were based on scientific studies carried out by the multidisciplinary team, while others were determined through several participatory workshops involving decision makers to validate the indicators. Some of these indicators must be evaluated directly in the field, while others can be studied through image analysis and mathematical calculations, or defined within the inference system adopted. To guide the field assessment and the collection of information necessary for the calculations, several protocols were developed and published (Soares et al., 2014; Santos et al., 2014a, 2014b, 2015; Abreu et al., 2015; Amâncio et al., 2016). This hierarchical process (Figure 7) enables assessing each aspect of sustainability individually and simultaneously.

The Sustainable Pantanal Farm software

Some problems arise in sustainability assessment, and it is necessary to take the level of abstraction involved in the concept into account, as well as the existence of natural variability in some phenomena. The synthesis provided by the indicators for a given "degree of sustainability" requires a robust methodology to deal with uncertainties, express complex interrelations, while at the same time, being interpretable and transparent to guarantee confidence in the assessment.

A mathematical and computational framework capable of dealing with these difficulties come from fuzzy set theory (FS), fuzzy logic, and fuzzy rule-based systems – FRBS. Such systems have been applied in areas such as engineering, modeling, and control, among others. Historically, its success is due to the ability to model knowledge based on natural language and good generalization capacity as well as the remarkable competence of FRBS in explaining the elaboration of the result based on the input values provided.

The Sustainable Pantanal Farm software was built as a decision support system based on models expressed in FRBS. Sustainability is evaluated by the environmental, economic, and social dimensions, at both ranch and regional level. Models were defined for each assessment (Figure 7), while input variables were the indicators themselves, with their scales defined in natural language (such as Good, Moderate and Bad). The relationships between indicators are expressed as a set of rules defined by domain experts. The evaluation results (indices and sub-indices), in addition to providing a comparative numerical value (1 to 10), have a corresponding qualitative output. Each model (index) feeds the more general models hierarchically further down, culminating in the farm's sustainability model.

The Sustainable Pantanal Farm interface to Internet[2] (Figure 8) is an interactive system where the user, given the indicator values, is able to infer qualitative concepts and numerical values, as well as compare how good these values are in relation to what is desired. It is also possible, through graphics, to visualize which indicators had more influence on the result. The rules that were used for the conclusion are shown to the user, ensuring interpretability and transparency. The system also allows the user to simulate scenarios in order to plan which ones lead to the level of sustainability one wants. The Sustainable Pantanal Farm software also has a second interface, aimed at mobile devices such as tablets and smartphones (Figure 9) using the Android operating system (available on the Google Play app store). Essentially, it provides the same functionalities, and it is based on the same mathematical models. Given a regional restriction of the Pantanal and farms in general, this version does not need an internet connection, as it has its own inference engine built into the application.

The Sustainable Pantanal Farm tool can be adopted by several decision makers (researchers, owners, technicians, politicians, legislators, certifiers, among others). Its main use is the diagnosis (degree of sustainability) of the beef cattle production system in the Pantanal through the assessment of environmental, social, and economic impacts of this activity, thus assisting in efficient management

[2] [2] Available at: https://www.fps.cnptia.embrapa.br

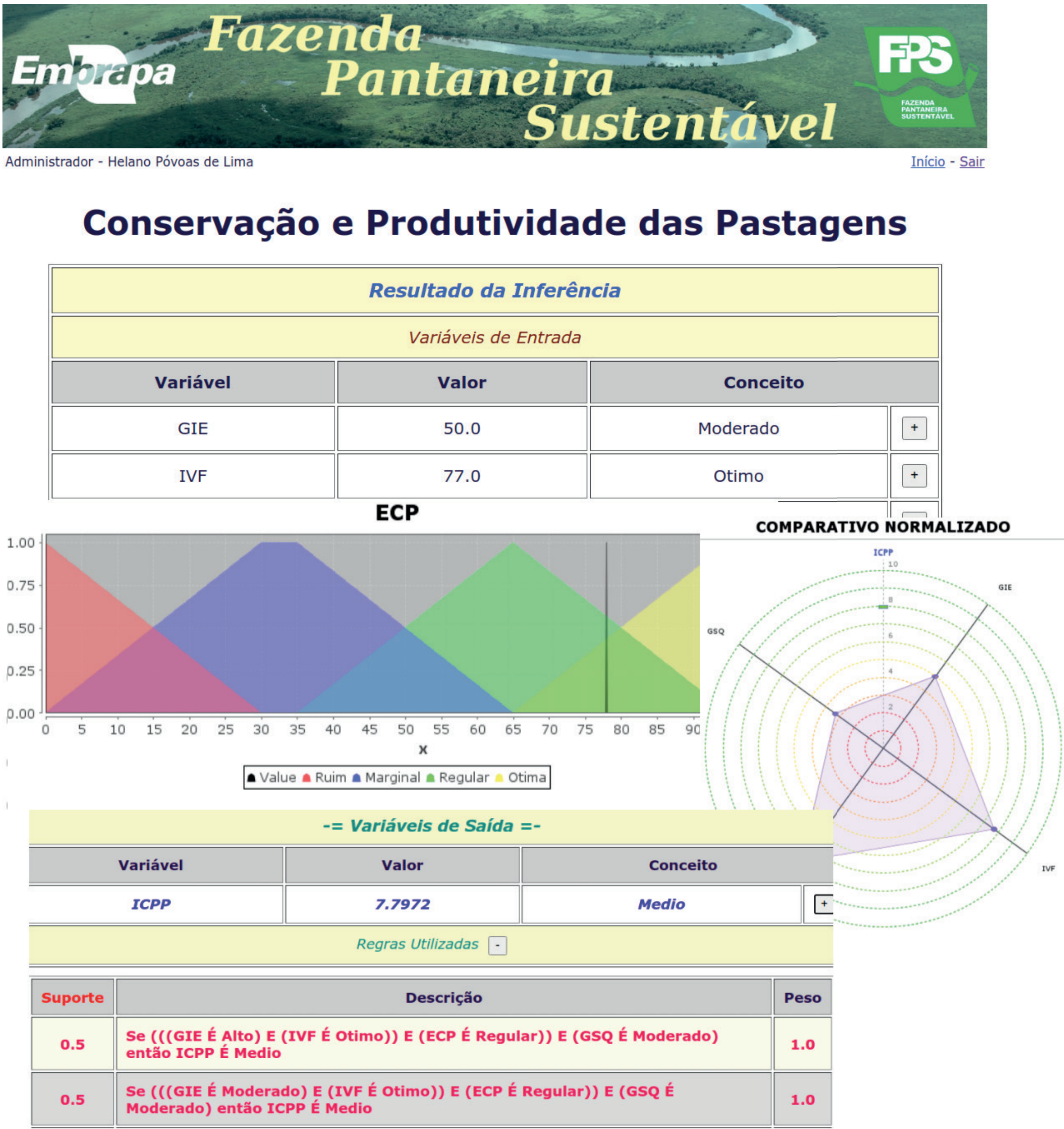


Figure 8. Internet Sustainable Pantanal Farm software interface elements.

Source: Sustainable Pantanal Ranch (2020).

through technology selection and good management practices. However, its application can be much broader for financing subsidy programs, certification, and marketing strategies that value products from the region. It may also offer necessary subsidies for the reformulation of current legislation and public incentive policies for sustainable production in the region. It is intended to insert the aspect of multifunctionality and ecosystem services in the future, something essential for the sustainability of production systems. The tool is being implemented in 15 farms in the Mato Grosso Pantanal with support from other agribusiness regional institutions, such as FAMATO, ACRIMAT, SENAR, IMEA, and rural unions,

Figure 9. Sustainable Pantanal Farm Android app interface.
Available at: https://play.google.com/store/apps/details?id=br.embrapa.cnptia.fps

as well as in six farms in the Pantanal of Mato Grosso do Sul, with support from FAMASUL, SENAR, and rural unions. Improvements will be incorporated over time, together with technicians, producers, and researchers.

Final considerations

In this chapter, several scientific-computing techniques applied in solving problems in the agricultural sector were presented. In the area of artificial Intelligence, classical logic techniques were applied to the development of an expert system for soil classification. The same problem was also addressed by a completely different technique using machine learning algorithms, which are fundamentally linked to statistics. Statistical analysis is also the basis of text mining techniques used to group documents with similar content in the agricultural area.

Another area of scientific computing, mathematical modeling, was explored in three different ways. In the first, Individual-Based Model provided a fully computational tool through a simulation system to compare three citrus and orange jessamine planting configurations in order to evaluate propagation control strategies for HLB in citrus. In the second application, linear predictor models, composed of classical mathematical equations, were used to assess the genetic values of livestock, with the objective of discovering which of them reinforce characteristics desired in the market. In the third model, the mathematical calculations were internally performed in a fuzzy logic inference-based system in order to assess sustainability in Pantanal farms. In this case, the advantage of fuzzy logic is combining natural language in the construction of a logical model where the answer is explainable to the decision maker.

Scientific computing techniques are essential for analyzing the large volume of data produced in this process of agricultural digital transformation. Through these techniques, it will be possible, from the collected data, to extract information and knowledge that will assist in the decision-making process in all links of the production chains, becoming central in the development of new agricultural solutions and technologies in Digital Agriculture. The applications presented in this chapter illustrate the variety of problems that can be addressed by the scientific computing methodological framework, including mathematical and statistical modeling, classical and fuzzy logic systems, simulation models, and machine learning models.

Considering these applications, it is worth emphasizing that the constant growth in data availability, technological advances and the expansion of the dimension and complexity of the demands of Brazilian society pose enormous challenges and opportunities for research and development in scientific computing applied to agriculture.

References

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**. Available at: https://www.bdpa.cnptia.embrapa.br. Accessed on: 19 May 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. Available at: https://www.fundecitrus.com.br/pdf/pes_relatorios/2020_05_11_Sumario-Executivo- da-Estimativa-da-Safra-de-Laranja-2020-2021.pdf. Accessed: 19 May 2020.

GRIMM, V.; RAILSBACK, S. F. **Individual-based modeling and ecology**. Princeton: Princeton University Press, 2005. DOI: 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 Agricultural Informatics, 2020. (Embrapa Agricultural Informatics. Documentos, 168). 2020.

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 Agricultural Informatics, 2017. p. 537-546. SBIAgro 2017.

MRODE, R. A. **Linear models for the prediction of animal breeding values**. 3rd ed. CABI, 2014. DOI: 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 AGRICULTURAL INFORMATICS, 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.

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.

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 Agricultural Informatics, 2017. p. 537-546. SBIAgro 2017.

SUSTAINABLE PANTANEIRA FARM. Available at: https://www. fps.cnptia.embrapa.br. Accessed on: 19 May 2020.

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. **IoT na Agricultura**: anais. Campinas: Embrapa Agricultural Informatics, 2019b.

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. Also published in: 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.

Photo: Monopoly919 (AdobeStock)

6

Computer vision applied to agriculture

Thiago Teixeira Santos | Jayme Garcia Arnal Barbedo | Sônia Ternes | João Camargo Neto |
Luciano Vieira Koenigkan | Kleber Xavier Sampaio de Souza

Introduction

Computer vision, in a simple and comprehensive definition, is a field of artificial intelligence dedicated to extracting information from digital images. In the context of digital agriculture, computer vision can be used in the detection of diseases and pests, in yield estimation and in the non-invasive evaluation of attributes such as quality, appearance and volume, and it is also an essential component in agricultural robotic systems. According to Duckett et al. (2018), field robotics could enable a new range of agricultural equipment: small and intelligent machines capable of reducing waste and environmental impact[1] and providing economic viability, thus increasing food sustainability. Also according to Duckett et al. (2018), there is considerable potential to increase the window of opportunity for interventions, for example, in wet soil operation, night operation and constant crop monitoring.

A class of problems addressed by computer vision are the alleged perceptual problems: the detection and classification of patterns in images that are associated with an object of interest, as for instance fruits (Sa et al., 2016; Santos et al., 2020), animals (Barbedo et al., 2019) or symptoms of diseases and pests (Ferentinos, 2018; Barbedo, 2019).

Constant and efficient monitoring can be carried out based on images captured by field teams or obtained by cameras attached to tractors, implements, robots or drones: the search for crop or livestock anomalies; the evaluation of crop spatial variability for intervention, according to the precepts of precision agriculture; and autonomous action by machines and implements. Figure 1 shows an example of detection of grape bunches in images obtained from vineyards.

[1] Due to the sparing and intelligent use of pesticides or simply mechanical intervention: the physical removal of pests

Photos: Thiago Teixeira Santos

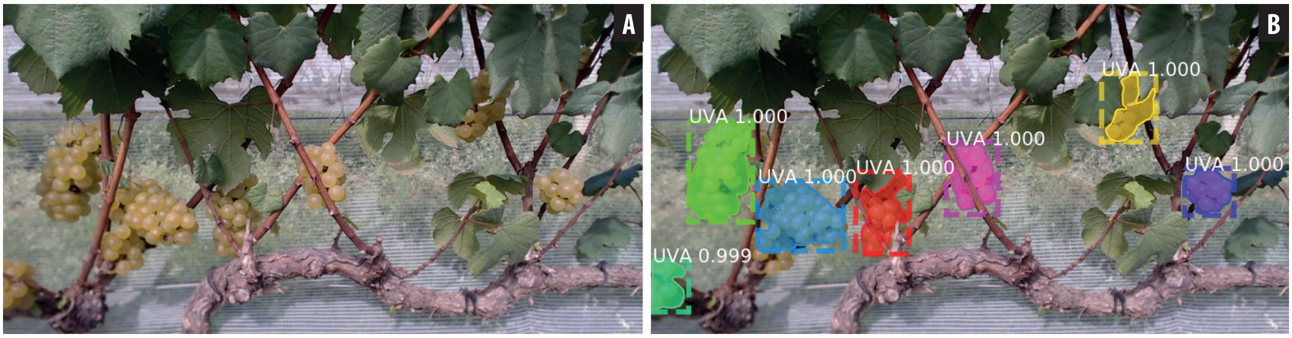


Figure 1. Examples of a perceptual task, the detection of grapes in images: image taken in a winery of a Chardonnay vine (A); detection result using a neural network (B).

Illustration: Thiago Teixeira Santos

Another class of problems are geometric ones. In forming an image, the light captured by the lens is projected onto a surface so that the three-dimensional scene produces a 2-D representation. Much of the scene structure is in the image, but depth information (the distance between the camera and the objects in the scene) is lost. One of the greatest contributions of geometric computer vision was the development of algorithms for recovering lost three-dimensional information from a set of images of the same scene. This is one of the most widely used computer vision applications in the market today: three-dimensional mapping and the production of maps from imagery obtained by Unmanned Autonomous Vehicles (UAVs – popularly known as drones, see Figure 2). Methodologies based on geometric computer vision have been employed in geological studies (Westoby et al., 2012), in pasture height assessment (Forsmoo et al., 2018) and in crop mapping (Comba et al., 2018), among other uses. Commercially, it is the core technology behind 3-D mapping and reconstruction services by UAVs extensively used in agriculture, such as Pix4D mapper and Agisoft PhotoScan/Metashape.

There is a growing number of computer vision applications in agricultural research. Consider, for example, the journal *Computers and Electronics in Agriculture*, which specializes in new software, hardware, and electronics applications in agriculture. A search for articles related to computer vision reveals that 23.7% of all works published in 2018 are associated with computer vision, rising to 29.1% in 2019. From January to June 2020, 115 of the 319 works (36.0%) published are related to computer vision. This volume of articles also translates into impact: of the 25 most cited works by June 2020, 14 are computer vision applications. Some simple factors explain this growth. Digital cameras are affordable and widely available devices in various configurations, easily integrated into larger systems (such as smartphones and UAVs). The advances in algorithms and hardware over the last ten years are reflected in the current dynamism of the area.

The next sections will present the recent innovations in the application of computer vision to agriculture, focusing on the contributions by Embrapa Digital Agriculture over the last 3 years. These advances are the result from both perceptual computer vision, the recognition of elements in the scene (Section 2), and geometric computer vision, the retrieval of three-dimensional information from images (Section 3). The combination of both fronts (Section 4) opens the way for systems that can perform highly complex operations, such as field robotics. Section 5 closes the chapter with some final remarks.

Perception: pattern recognition in images

Pattern recognition can be seen as the role of finding a representation for the pattern sought that is sufficiently versatile to cover observable variations, yet simple enough to be processed in a timely

Figure 2. UAV mapping: images are used to identify the three-dimensional structure of the area, and the position and orientation of the aircraft, displayed in red (A); the geolocated three-dimensional model is then projected onto a plane, forming a map. (B).

Illustration: Thiago Teixeira Santos

manner by the machine. In other words, it is an adequate pattern description to allow the machine to find it in the input data, yet succinct so that its interpretation is carried out within operating time constraints.

Visual patterns in natural images can be incredibly intricate, with regularities and variations that are difficult to describe. In agriculture, patterns assumed by fruits, leaves, grains, plants and symptoms of pathologies exhibit enormous variability, amplified by differences in lighting, position, occlusion and different sources of noise (dirty lenses, dust, interference, etc.). Figure 3 illustrates some of the difficulties a fruit detection system faces in real field growing conditions: severe occlusion between fruits, leaves and branches; color similarity between green fruits and the canopy; lighting variations between images; specular reflection (direct reflection of sunlight that saturates the camera sensor); and focus problems. Notwithstanding some success from the use of machine learning techniques (Gongal et al., 2015), pattern recognition in natural images began to reach high levels of accuracy with the arrival of convolutional neural networks (Lecun et al., 2015), quickly adopted for image recognition in agriculture (Kamilaris; Prenafeta-Boldú, 2018).

In neural networks, an architecture or model is a sequence of modules that perform simple operations on the data so that a module receives data from previous modules and propagates the result of its operations to the following modules. In computer vision, the most used neural networks are the convolutional neural networks (CNNs), in which the main operation employed is convolution, a linear combination of values in the vicinity of the input pixels. Neural networks are said to be deep if there is a large sequence of linked modules. The deeper the network, the greater its ability to learn representations for complex patterns, since each module is able to compose the representations of previous modules in a hierarchy. In the case of images, there is an intuitive interpretation for this behavior: the initial modules are able to find lines and edges of objects, the following modules compose these patterns into simple textures and structures like triangles and spots, which are then combined into other structures like parts of leaves, branches and berries. Finally, the final modules combine these elements into objects of interest: a plant, a bunch of grapes, an ox.

Photos: João Camargo Neto

Figure 3. Examples of the difficulties faced in fruit detection. In the images, we can observe problems of focus, specular reflection, severe occlusion by leaves, branches and other fruits, light variations and similarities in the color pattern between fruits and leaves.

The modules have parameters that need to be adjusted so that the joint operation of the entire network produces the expected results. A frequently used metaphor is to imagine that each parameter is adjusted by a dimmer. Adjusting a neural network would be performing the adjustment of millions of dimmers, each of which could affect the pattern recognition performance. Manually, however, this adjustment would be impractical and virtually impossible. The training of neural networks is an automated process for adjusting these parameters, so that the network "learns" the appropriate representations for the recognition problem in question.

In supervised learning of image patterns, this training is carried out using observations, images whose desired answer is known ("there is an orange in this image", "there are signs of coffee rust on this leaf"). This training requires thousands of observations, which is directly linked to the size of the network: more parameters require more observations, although it is difficult to determine an exact relationship between the number of parameters and the number of observations required. When the network processes the input image, the produced result is compared to the expected result, and their error is computed. The parameters are then adjusted to reduce the previous error, in a process known as backpropagation (Goodfellow et al., 2016). In practice, observations are grouped into batches, the network processes the batch and the observed error is computed. The backpropagation algorithm is used to adjust the parameters, starting with the final modules of the network and proceeding towards the parameters of the initial modules (hence the name of the procedure). Training proceeds with the next batch, and the procedure is repeated until the error reaches an observable minimum[2]. In short, deep neural networks automate the process of searching for adequate representations in pattern recognition problems, provided there is a sufficiently large set of observations for training in order to adequately represent the variability of the intended pattern. It is precisely this ability that makes the methodology so attractive to the intricate problems of recognition in agriculture.

Identification of plant diseases

The detection and classification to diagnose disease, pests and plant nutritional deficiencies in images are of great interest in agriculture. Automatic detection enables constant monitoring and searching for crop anomalies, based on images captured by field teams or obtained by cameras attached to tractors, implements, robots or UAVs. On the other hand, classification associates the detected anomalies to the disease, deficiency or pest, assisting the producer in the correct intervention. Neural networks can be used in both tasks, even simultaneously.

As seen above, thousands of observations are required before a neural network is able to produce accurate results. This need is amplified for plant disease recognition due to the large number of combinations resulting from the crossing between target cultures, pathologies, stage of disease development and imaging condition (manual collection, aerial monitoring by UAVs, capture at the ground level by machine, camera position, among others). This situation points to the need for large shared databases (Barbedo, 2018; Ferentinos, 2018), as considerable effort is required for their production.

The process of collecting and annotating the images, in other words, associating each image with the desired result for the supervised learning stage, is usually lengthy and costly. However, some strategies can be used to increase the number of observations. Barbedo (2019) showed that multiple lesions

[2] The ideal error would be zero, but there is no guarantee that an architecture will be able to achieve this. It is also an open problem to determine *a priori* what is the smallest error a network will be able to achieve for a given training set.

of the same pathologies which occur on the same leaf can be exploited to increase the number of observations from the same collection. Several examples of symptoms can be obtained from a single leaf or plant tissue sample, as seen in Figure 4. This strategy allowed an original database, containing 1575 observations (Barbedo, 2018), to be expanded to 46409 observations (Barbedo, 2019), producing gains in disease classification accuracy of, on average, 12%.

Photos: Jayme Garcia Arnal Barbedo

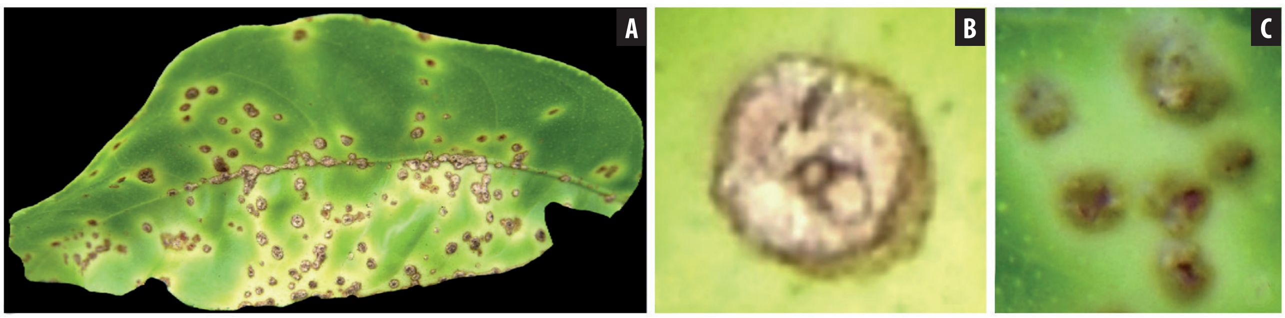


Figure 4. Examples of observations used in training systems for plant disease recognition: a sample of a diseased leaf, collected in the field (A); an observation of symptoms associated with the pathology (B); clusters of symptoms that also form a discernible pattern associated with the pathology (C).

Barbedo (2019) showed that a convolutional neural network, the GoogLeNet architecture (Szegedy et al., 2015), can be applied in the classification of many pathologies in different cultures, reaching accuracy values of 80% (passion fruit) up to 100% (cassava, cabbage, cotton, wheat, and sugarcane), as shown in Table 1. The database used, termed as Digipathos, was made publicly available[3]. Although the classification results are promising, there are still major challenges, especially with regard to detection ("are there symptoms present in the observation?"), which is crucial in autonomous monitoring for pest and disease management, but which still does not present the same classification accuracy ("what is the pathology for the observed symptom?"). In his experiments, Barbedo (2019) shows that accurate detections can be produced when symptoms are already severe, but not when the symptoms are still mild or do not occupy large portions of plant tissue, which is the ideal time for intervention by the farmer. False positive detection errors (healthy tissue detected as diseased) are often caused by factors such as the presence of dust, debris or even water droplets. It is also not clear yet what number of samples is needed so that

Table 1. Accuracy of the classification of pathologies in different cultures. For the cassava and kale images, the accuracy reached 100% in all tests.

Crop	Number of images	Accuracy (%)
Bean	3,079	94 ± 0.8
Cassava	895	100 ± 0.0
Citrus	1,868	96 ± 0.6
Coconut	1,504	98 ± 0.6
Corn	10,480	75 ± 4.4
Coffee	1,899	89 ± 1.9
Cotton	2,023	99 ± 0.3
Cashew	4,509	98 ± 0.5
Grape	2,330	96 ± 0.8
Kale	196	100 ± 0.0
Passion fruit	280	80 ± 4.2
Soy	13,733	87 ± 3.6
Sugar cane	2773	99 ± 0.4
Wheat	840	99 ± 0.5
Total	46,135	94 ± 2.0

Source: Adapted from Barbedo (2019).

[3] Available in: https://www.digipathos-rep.cnptia.embrapa.br

the characteristics of symptoms can be properly learned by neural networks (still an open question in computer vision in general).

Detection of animals in pastures

Barbedo et al. (2019) present an example of how UAV technologies and computer vision can be combined for monitoring large areas, for example detecting cattle in extensive livestock production. Given the dynamics of the animals and the enormous size of the pasture areas, the ranchers face great difficulties monitoring the herds in the pastures.

A database composed of 1853 images containing 8629 Canchim animals was produced based on images obtained by a commercially available quadrirotor[4]. Barbedo et al. (2019) tested 15 different neural network architectures at 3 distinct spatial resolutions (1, 2 cm/pixel and 4 cm/pixel), in order to analyze the performance resulting from different flight heights. The results showed that most of the tested architectures were able to reach high levels of accuracy, above 95%. The NasNet architecture (Zoph et al., 2018), a very deep network with great capacity to learn complex patterns, achieved accuracy close to 100%. These results are expressive, especially considering the complexity of the problem, as shown in Figure 5: several situations, from severe occlusion by trees and drinking fountains to differences in

Photos: Jayme Garcia Arnal Barbedo

Figure 5. Examples of situations observed in the detection of animals in pastures: animal in high pasture (A); dry pasture (B); exposed soil (C); tree occlusions (D); covering of drinking fountains (E) and electrical cables (F).

[4] In this case, a DJI Phantom 4 Pro vehicle.

lighting and pasture conditions, in addition to the position and disposition of the animals, all of which present highly variable situations. Even so, the accuracy of most of the architectures tested is expressive. Another particularly interesting effect from an operational point of view was that most models present better results at the 2 cm/pixel resolution and not at the maximum 1 cm/pixel resolution, which may be due to the resolution of the convolutional modules with which these architectures were originally designed. In practice, this enables flights at higher heights, which allows covering areas in less time.

Detection and counting of fruits

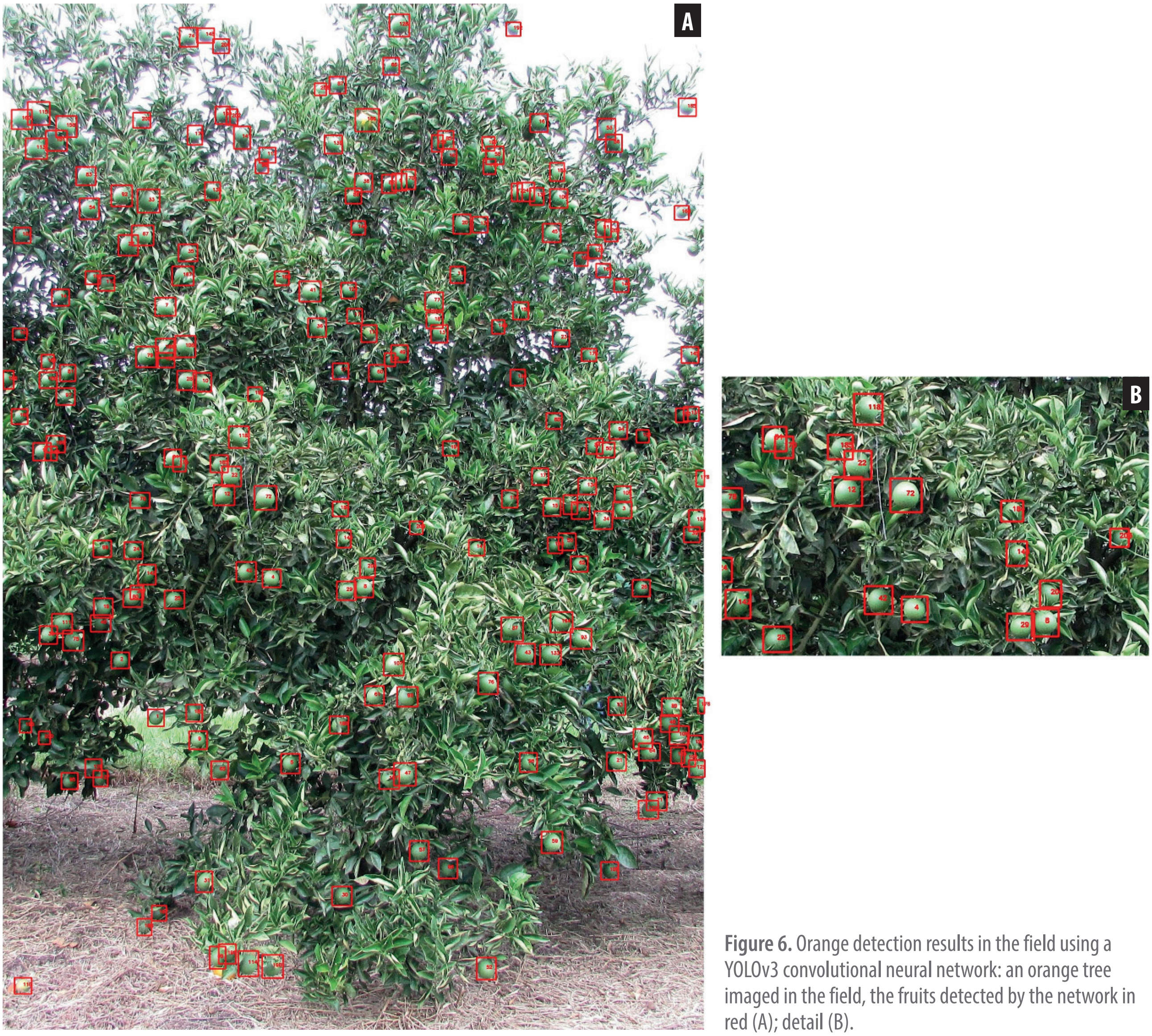
Automatic fruit detection is an enabling component for many agricultural applications. It can help estimate production, which is useful in logistical planning and in negotiations between rural producers and buyers. If detection is combined with precise spatial location, new applications can be developed in precision agriculture, assisting in the proper management of spatial crop variability. Fruit detection can also be a preliminary step in monitoring disease and nutritional deficiencies (see item "Identification of plant diseases"), restricting the areas in the images that should be inspected for symptoms. Given the decline in the agricultural workforce, fruit detection is also a technology that enables automated spraying and harvesting systems (Duckett et al., 2018; Xiong et al., 2020).

As discussed earlier, there are several factors that hinder the detection process, from occlusion by leaves and branches to camera focus and lighting issues (Figure 3). In some crops, the fruits also have various shapes, compactness and orientation, such as viticulture (Santos et al., 2020). Despite some success with other machine learning techniques (Gongal et al., 2015), fruit detection has recently gained traction with the improvements in convolutional neural networks (Sa et al., 2016; Bargoti; Underwood, 2017; Kamilaris; Prenafeta-Boldú, 2018).

Camargo Neto et al. (2019) produced a dataset with 3,066 images of oranges collected in the field, from different devices, such as cameras and smartphones. Most of the images were provided by the Crop Estimation Program (PES) of the Citriculture Defense Fund (Fundecitrus). The fruits, from different varieties of orange, had different levels of maturation, with a predominance of green fruits (Figure 3). From these images, a subset of 2036 observations was used in the training of a YOLOv3 neural network (Redmon et al., 2016; Redmon; Farhadi, 2018). The authors evaluated the network trained in the 1030 remaining images and verified the correct detection of more than 90% of the fruits, with an accuracy also above 90%, that is, less than 10% of the detections produced were false positives. Figure 6 shows an example of fruit detection in an orange tree image taken in the field.

Santos et al. (2020) showed that for the grapes in viticulture that present high variation in shape, color, size and compactness, bunches can be detected and segmented using architectures such as Mask-RNN and YOLO. The authors produced a new annotation tool that can speed up the process of associating pixels to fruits, discriminating exactly which pixels belong to which bunches. The generated dataset, named WGISD (Embrapa Wine Grape Segmentation Dataset) and publicly available[5], contains 4,432 bunches in 300 images, covering five wine varieties. The authors evaluated three different neural network architectures, YOLOv2 (Redmon; Farhadi, 2017), YOLOv3 (Redmon; Farhadi, 2018) and Mask-RCNN (He et al., 2017), the latter responsible for the most promising results. In a test base composed of 837 bunches, the network identified 87% of the bunches, with precision of 90.7%. Examples of the produced detections are shown in Figure 1 (B).

[5] Available at: https://doi.org/10.5281/zenodo.3361736.



Photos: João Camargo Neto

Figure 6. Orange detection results in the field using a YOLOv3 convolutional neural network: an orange tree imaged in the field, the fruits detected by the network in red (A); detail (B).

However, a complete fruit counting application needs a methodology that can integrate the detections reported in several images, so that fruits seen in more than one image are not counted multiple times. In other words, fruits (and objects of interest in general) observed in several images must be associated with each other. This data association task can be performed by integrating pattern recognition with geometric computer vision, as shown as follows.

Three-dimensional mapping and reconstruction

One of the greatest contributions of geometric computer vision was developing algorithms capable of recovering three-dimensional information from a set of images of the same scene. As results from decades of research in areas such as projective geometry and continuous optimization, these algorithms can transform even a simple webcam into a powerful 3-D scanner. Perhaps even more importantly, they allow a mobile agent, such as a UAV, not only to map the three-dimensional structure of the environment, but also to determine its precise location (Figure 2), paving the way for autonomous agents that can navigate and interact with its surroundings (Stachniss et al., 2016).

Images must be obtained from different positions, by multiple cameras or by a single camera moving through the scene. This is the meaning of the term structure from motion (SfM), used in computer vision to define the problem of recovering the three-dimensional structure of a scene and the position of the camera from a set of images. Figure 7 illustrates the process of projecting a point in the scene as the camera is moved to three different positions. If we can determine correspondences between points in different images, it is possible to determine, with the help of projective geometry techniques, the position of the camera at the time each image was captured, more precisely the location of its projection centers, represented in the Figure 7 for points C_1, C_2 and C_3[6]. Once the location of the projection centers has been determined, it is then possible to estimate the position of the point in three-dimensional space based on its projections on the images (the points x1, x2, and x3 in Figure 7), a process known as *triangularization*. A detailed description of the entire process can be seen in Hartley and Zisserman (2003). The determination of image correspondence is also obtained automatically, using algorithms specialized in finding visually salient points (the points x1, x2, and x3) and, by comparing the pixels in their neighborhoods, associating different image points (Lowe, 2004; Detone et al., 2018).

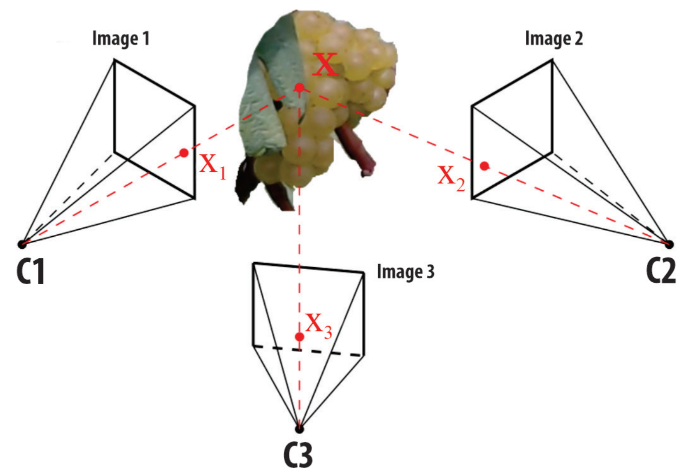


Figure 7. Structure from motion. An X point on a scene surface is projected onto the image plane at different positions as the camera is moved to positions C1,C2 and C3.

Illustration: Thiago Teixeira Santos

Santos et al. (2017) showed that an SfM system using a simple webcam can build accurate three-dimensional plant models in the field. Figure 8 shows an example, for a Chardonnay vine. As we will see in the next section, these three-dimensional models can be used to estimate 3-D attributes, such as

Figure 8. Three-dimensional reconstruction with SfM for a Chardonnay vine in the field: the red prisms indicate the camera (a commercial webcam) position and orientation, when each image was captured (A); same 3-D model observed from another angle (B).

Illustration: Thiago Teixeira Santos.

[6] Additional information, obtained by calibration methods, is needed to determine the correct scale, that is, the distance in a known unit such as meters or millimeters.

fruit volume and position. The 3-D system used and developed at Embrapa Digital Agriculture, named 3dmcap, is freely available[7] for non-commercial use.

The use of three-dimensional information in agriculture is expected to intensify in the coming years, not only through the use of the SfM technique (already commercially used by 3-D mapping services with UAVs), but also by the falling costs of stereo cameras, which provide depth information in the image, and by LIDAR sensors. Recent examples are the use of stereo cameras in vineyard phenotyping (Milella et al., 2019) and the detection of apples using LIDAR (Gené-Mola et al., 2020).

Figure 8 shows the three-dimensional reconstruction with SfM for a Chardonnay vine in the field: red prisms indicate the position and camera orientation (a commercial webcam, at the time of capture of each image (A); same 3-D model from another angle (B).

Combination of structure and recognition

If the SfM retrieves the three-dimensional structure from the scene and the imaging itself (the camera position(s) during the capture time), and the recognition identifies objects of interest in the scene, such as symptoms, fruits, plants or animals, the combination of the two pieces of information allows a broad assessment of the observed environment.

One of the uses of this combination is fruit mapping: the 3-D information combined with the detection of fruit in each image allows the spatial position of each fruit to be determined and that the same fruit is not counted more than once when it appears in multiple images.

Santos et al. (2020) used SfM to obtain a three-dimensional reconstruction of a row of vines in the field, based on the frames of a video sequence produced by a camera embedded in a service vehicle. A neural network was used to detect bunches of grapes in each image. By projecting the 3-D points of the scene onto the images, it was possible to associate detections with positions in the three-dimensional space and, therefore, determine the consistency between the bunches observed in one video frame and the bunches seen in the following frames[8] (Figure 9).

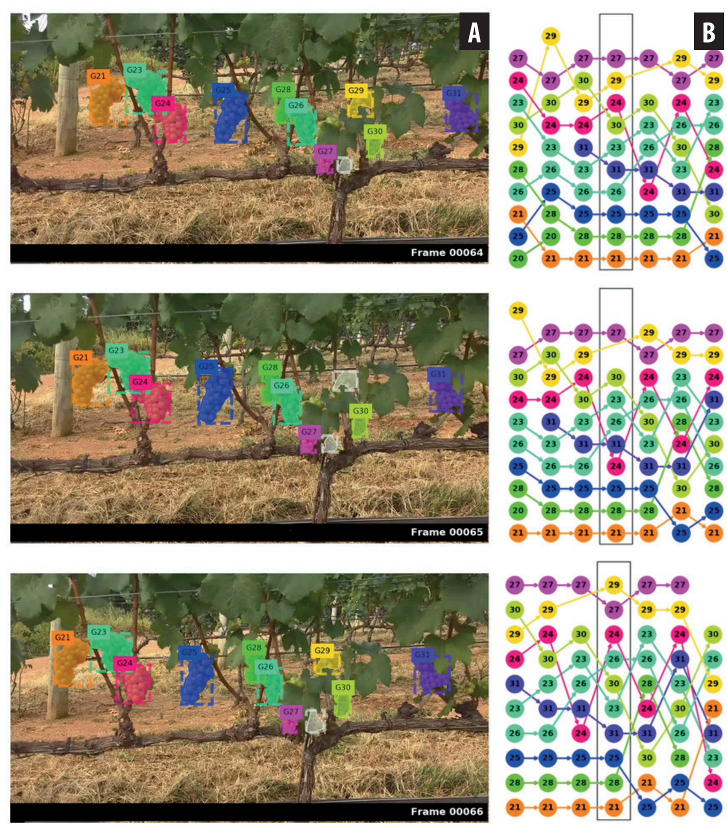


Figure 9. Tracking grape bunches in a video sequence obtained in the field: video frames were extracted and submitted to fruit detection by neural networks (A); the nodes represent bunches of grapes, in the order in which they were found by the neural network (each column of nodes represents a frame of the video sequence). The arrows inform the association between nodes from one frame to another, performed using 3-D information obtained by SfM (B).

Illustration: Thiago Teixeira Santos.

The joint use of 3-D models obtained by SfM and convolutional networks for fruit

[7] Available at: https://github.com/thsant/3dmcap

[8] 8 A video demonstrating the tracking of grape bunches is available at: https://www.youtube.com/watch?v=1Hji3GS4mm4

detection and counting was also explored by Liu et al. (2019) in mango orchards and by Häni et al. (2020) in apple orchards.

Attributes of great interest in agronomic applications can be extracted from three-dimensional information. Santos et al. (2017) used a machine learning algorithm to identify which regions of the three-dimensional vine models corresponded to the bunches of grapes, as shown in Figure 10 (A). The volume of bunches was then estimated based on these regions. Fruit volume has a strong correlation with its weight, as can be seen in Figure 10 (B). These computer vision-based systems can provide a non-invasive and non-destructive methodology for estimating fruit weight, without having to remove them from the plant. Such technology can be used to assess growth throughout the crop cycle, without the need to remove (collect) samples.

Figure 10 shows the estimation of fruit weight based on volume in three-dimensional models.

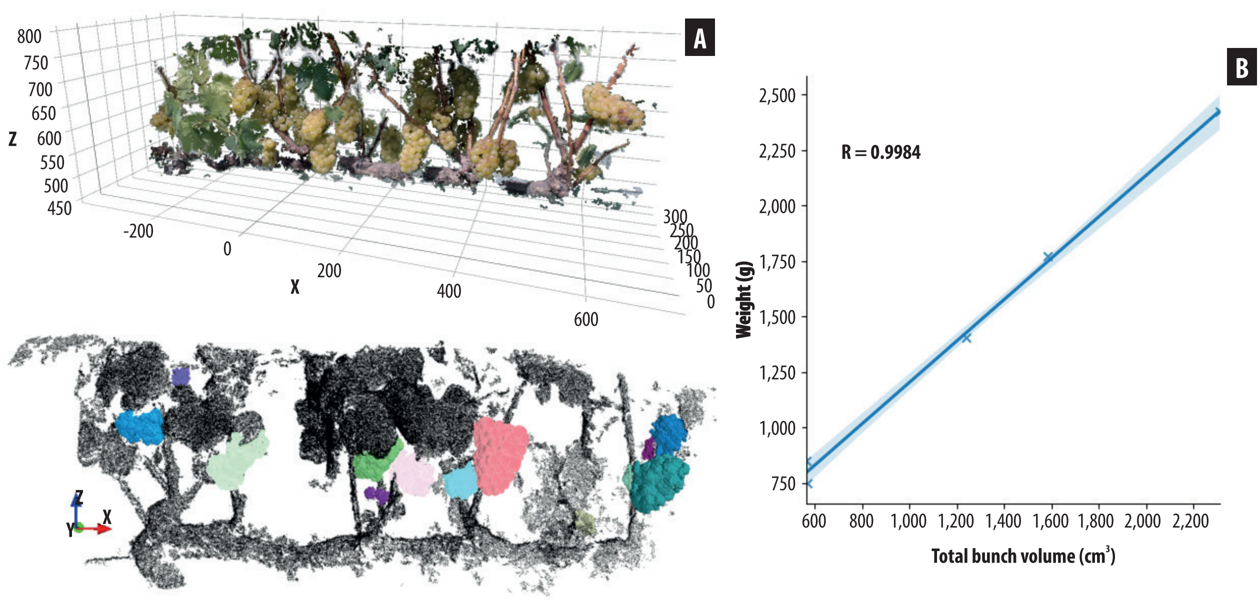


Figure 10. Estimation of fruit weight from the volume in three-dimensional models: grape bunches are identified (in colors) and separated from the rest of the plant (in black) (A); coefficient of determination between the estimated volume and the total weight of fruits in five different vines (B).

Source: Adapted from Santos et al. (2017).

Performance and intervention: field robotics

The combination of SfM and recognition is precisely one of the enabling technologies for one of the most challenging and impactful applications in agricultural automation: field robotics. Take, for example, a major challenge in agricultural robotics: automated fruit harvesting. While crops such as grains, sugarcane and coffee have their own machinery for automated harvesting, the same does not apply to horticulture and fruit cultivation – especially for the latter – due to the existing complexity in the structure of the orchards. Fruit harvesting depends on manual harvesting, which is unsettling considering the decreasing availability of labor in the field (Roser, 2013).

Automatic harvesting systems require two components of computer vision: the perceptual, for identifying fruits and obstacles, and the geometric, for the automatic positioning of the robot and its handlers. Several research groups have applied these two components in the development of automated harvesting systems. Taking apple farming as an example, Silwal et al. (2017) developed a robotic apple harvesting system, evaluated in a commercial orchard. Their computer vision system was accurate, taking an average of 1.5 s to locate each fruit. The system was successful in harvesting 85% of the fruits, with an average time of 6 s per fruit. In addition to pomiculture, other crops have been investigated for implementing robotic harvesting, such as peppers (Bac et al., 2017), lettuce (Birrell et al., 2020), strawberry (Xiong et al., 2020), kiwi fruit (Williams et al., 2020), among others.

Final considerations

Computer vision has enormous potential for application in the area of digital agriculture. Several products and services based on computer vision components are expected to reach producers in the coming years. However, many challenges still depend on research and development endeavors.

A major bottleneck is the need for large databases to train neural networks for perceptual tasks. Research in the area of semi-supervised and unsupervised learning is currently being conducted by the computer vision community. The idea is to be able to learn patterns of interest with few examples and obtain systems with good accuracy in order to detect patterns such as fruits, symptoms and animals.

In robotics, the challenge continues to be developing robust systems that are capable of autonomously operating in the field for long periods, but which are safe for people and animals circulating in the field. These systems need to map the environment quickly, respond promptly, accurately find the objects to be monitored, and carry out the interventions for which they are designed. Despite the immense challenges, the computer vision and robotics communities have made great advances in recent years, which will soon be reflected in various agricultural applications, from monitoring to performance.

Finally, the authors emphasize that the results in fruit detection were financed by the Embrapa SEG 11.14.09.001.05.04 and FAPESP 2017/19282-7 projects. The results related to disease detection were financed by projects FAPESP 2013/06884-8 and Embrapa SEG 02.14.09.001.00.00. The results related to animal detection experiments were funded by FAPESP 2018/12845-9 project. The images for citriculture research were provided by PES/Fundecitrus. In addition, the GPUs used to train the neural networks were donated by NVIDIA Corporation.

References

BAC, C. W.; HEMMING, J.; TUIJL, B. A. J. van; BARTH, R.; WAIS, E.; HENTEN, E. J. van. Performance evaluation of a harvesting robot for sweet pepper. **Journal of Field Robotics**, v. 34, n. 6, p. 1123-1139, Sept. 2017. DOI: 10.1002/rob.21709.

BARBEDO, J. G. A.; KOENIGKAN, L. V.; SANTOS, T. T.; SANTOS, P. M. A study on the detection of cattle in UAV images using deep learning. **Sensors**, v. 19, n. 24, article number 5436, Dec. 2019. DOI: 10.3390/s19245436.

BARBEDO, J. G. A. Plant disease identification from individual lesions and spots using deep learning. **Biosystems Engineering**, v. 180, p. 96-107, Apr. 2019. DOI: 10.1016/j.biosystemseng.2019.02.002.

BARBEDO, J. G. Factors influencing the use of deep learning for plant disease recognition. **Biosystems Engineering**, v. 172, p. 84-91, Aug. 2018. DOI: 10.1016/j.biosystemseng.2018.05.013.

BARGOTI, S.; UNDERWOOD, J. Deep fruit detection in orchards. In: IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION, 2017, Singapore. **Proceedings**... Singaropre: IEEE, 2017. p. 3626-3633. DOI: 10.1109/ICRA.2017.7989417.

BIRRELL, S.; HUGHES, J.; CAI, J. Y.; IIDA, F. A field-tested robotic harvesting system for iceberg lettuce. **Journal of Field Robotics**, v. 37, n. 2, p. 225-245, Mar. 2020. DOI: 10.1002/rob.21888.

CAMARGO NETO, J.; TERNES, S.; SOUZA, K. X. S. de; YANO, I. H.; QUEIROS, L. R. Uso de redes neurais convolucionais para detecção de laranjas no campo. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 12., 2019, Indaiatuba. **Anais [...]** Ponta Grossa: SBIAGRO, 2019. p. 312-321. Organizadores: Maria Fernanda Moura, Jayme Garcia Arnal Barbedo, Alaine Margarete Guimarães, Valter Castelhano de Oliveira. SBIAgro 2019. Available at: https://www.alice.cnptia.embrapa.br/ alice/bitstream/doc/1125722/1/PC-Redes-neurais-SBIAGRO-2019.pdf. Accessed on : 10 May 2020.

COMBA, L.; BIGLIA, A.; AIMONINO, D. R.; GAY, P. Unsupervised detection of vineyards by 3D point-cloud UAV photogrammetry for precision agriculture. **Computers and Electronics in Agriculture**, v. 155, p. 84-95, Dec 2018. DOI: 10.1016/j.compag.2018.10.005.

DETONE, D.; MALISIEWICZ, T.; RABINOVICH, A. SuperPoint: self-supervised interest point detection and description. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION WORKSHOPS, 2018, Salt Lake City. **Proceedings**... IEEE, 2018. p. 337-349. DOI: 10.1109/CVPRW.2018.00060.

DUCKETT, T.; PEARSON, S.; BLACKMORE, S.; GRIEVE, B. Agricultural robotics: the future of robotic agriculture. **UK-RAS Network**, June 2018. DOI: 10.31256/WP2018.2.

FERENTINOS, K. P. Deep learning models for plant disease detection and diagnosis. **Computers and Electronics in Agriculture**, v. 145, p. 311-318, Feb. 2018. DOI: 10.1016/j.compag.2018.01.009.

FORSMOO, J.; ANDERSON, K.; MACLEOD, C. J. A.; WILKINSON, M. E.; BRAZIER, R. Drone-based structure-from-motion photogrammetry captures grassland sward height variability. **Journal of Applied Ecology**, v. 55, n. 6, p. 2587-2599, Nov. 2018. DOI: 10.1111/1365-2664.13148.

GENÉ-MOLA, J.; GREGORIO, E.; CHEEIN, F. A.; GUEVARA, J.; SANZ-CORTIELLA, R.; ESCOLÀ, A.; ROSELL-POLO, J. R. Fruit detection, yield prediction and canopy geometric characterization using LiDAR with forced air flow. **Computers and Electronics in Agriculture**, v. 168, 105121, Jan. 2020. DOI: 10.1016/j.compag.2019.105121.

GONGAL, A.; AMATYA, S.; KARKEE, M.; ZHANG, Q.; LEWIS, K. Sensors and systems for fruit detection and localization: a review. **Computers and Electronics in Agriculture**, v. 116, p. 8-19, Aug. 2015. DOI: 10.1016/j.compag.2015.05.021.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. Cambridge; London: MIT Press, 2016.

HÄNI, N.; ROY, P.; ISLER, V. A comparative study of fruit detection and counting methods for yield mapping in apple orchards. **Journal of Field Robotics**, v. 37, n. 2, p. 263-282, Mar. 2020.

HARTLEY, R.; ZISSERMAN, A. **Multiple view geometry in computer vision**. 2nd ed. New York: Cambridge University Press, 2003. DOI: 10.1017/CBO9780511811685.

HE, K.; GKIOXARI, G.; DOLLÁR, P.; GIRSHICK, R. Mask R-CNN. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION, 2017, Venice. **Proceedings**... IEEE, 2017. p. 2980-2988. DOI: 10.1109/ICCV.2017.322.

KAMILARIS, A.; PRENAFETA-BOLDÚ, F. X. Deep learning in agriculture: a survey. **Computers and Electronics in Agriculture**, v. 147, p. 70-90, Apr 2018. DOI: 10.1016/j.compag.2018.02.016.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436-444, May 2015. DOI: 10.1038/nature14539.

LIU, X.; CHEN, S. W.; LIU, C.; SHIVAKUMAR, S. S.; DAS, J.; TAYLOR, C. J.; UNDERWOOD, J.; KUMAR, V. Monocular camera based fruit counting and mapping with semantic data association. **IEEE Robotics and Automation Letters**, v. 4, n. 3, p. 2296-2303, 2019. DOI: 10.1109/LRA.2019.2901987.

LOWE, D. G. Distinctive image features from scale-invariant keypoints. **International Journal of Computer Vision**, v. 60, n. 2, p. 91-110, Nov. 2004. DOI: 10.1023/B:VISI.0000029664.99615.94.

MILELLA, A.; MARANI, R.; PETITTI, A.; REINA, G. In-field high throughput grapevine phenotyping with a consumer-grade depth camera. **Computers and Electronics in Agriculture**, v. 156, p. 293-306, Jan 2019. DOI: 10.1016/j.compag.2018.11.026.

REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You only look once: unified, real-time object detection. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2016, Las Vegas. **Proceedings**... IEEE, 2016. p. 779-788. DOI: 10.1109/CVPR.2016.91.

REDMON, J.; FARHADI, A. YOLO v3: an incremental improvement [DB]. **arXiv preprint arXiv:1612.08242**, 2018.

REDMON, J.; FARHADI, A. YOLO9000: better, faster, stronger. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2017, Honolulu. **Procedings**... IEEE, 2017. p. 6517-6525. DOI: 10.1109/CVPR.2017.690.

ROSER, M. **Employment in agriculture**. 2013. Available at: https://ourworldindata. org/employment-in-agriculture. Accessed on: 19 May 2020.

SA, I.; GE, Z.; DAYOUB, F.; UPCROFT, B.; PEREZ, T.; MCCOOL, C. DeepFruits: a fruit detection system using deep neural networks. **Sensors**, v. 16, n. 8, p. 1222, Aug. 2016. DOI: 10.3390/s16081222.

SANTOS, T. T.; BASSOI, L. H.; OLDONI, H.; MARTINS, R. L. Automatic grape bunch detection in vineyards based on affordable 3D phenotyping using a consumer webcam. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 11., 2017, Campinas. **Ciência de**

dados na era da agricultura digital: anais. Campinas: Editora da Unicamp: Embrapa Informática Agropecuária, 2017. p. 89-98. SBIAgro 2017. Available at: https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1083291/1/AutomatcgrapeSBIAgro.pdf. Accessed on: 16 Oct . 2020.

SANTOS, T. T.; SOUZA, L. L. de; SANTOS, A. A. dos; AVILA, S. Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association. **Computers and Electronics in Agriculture**, v. 170, article 105247, p. 1-17, Mar. 2020. DOI: 10.1016/j.compag.2020.105247.

SILWAL, A.; DAVIDSON, J. R.; KARKEE, M.; MO, C.; ZHANG, Q.; LEWIS, K. Design, integration,and field evaluation of a robotic apple harvester. **Journal of Field Robotics**, v. 34, n. 6, p. 1140–1159, sept. 2017. DOI: 10.1002/rob.21715.

STACHNISS, C.; LEONARD, J. J.; THRUN, S. Simultaneous localization and mapping. In: SICILIANO, B.; KHATIB, O. (ed.). **Springer handbook of robotics**. Cham: Springer International Publishing, 2016. p. 1153-1176. DOI: 10.1007/978-3-319-32552-1_46.

SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCKE, V.; RABINOVICH, A. Going deeper with convolutions. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2015, Boston. **Proceedings**... IEEE, 2015. p. 1-9. DOI: 10.1109/CVPR.2015.7298594.

WESTOBY, M.; BRASINGTON, J.; GLASSERA, N. F.; HAMBREY, M. J.; REYNOLDS, J. M. 'Structure-from-Motion' photogrammetry: a low-cost, effective tool for geoscience applications. **Geomorphology**, v. 179, p. 300-314, Dec 2012. DOI: 10.1016/j.geomorph.2012.08.021.

WILLIAMS, H.; TING, C.; NEJATI, M.; JONES, M. H.; PENHALL, N.; LIM, J.; SEABRIGHT, M.; BELL, J.; AHN, H. S.; SCARFE, A.; DUKE, M.; MACDONALD, B. Improvements to and large-scale evaluation of a robotic kiwifruit harvester. **Journal of Field Robotics**, v. 37, n. 2, p. 187-201, Mar. 2020. DOI: 10.1002/rob.21890.

XIONG, Y.; GE, Y.; GRIMSTAD, L.; FROM, P. J. An autonomous strawberry-harvesting robot: design, development, integration, and field evaluation. **Journal of Field Robotics**, v. 37, n. 2, p. 202-224, Mar 2020. DOI: 10.1002/rob.21889.

ZOPH, B.; VASUDEVAN, V.; SHLENS, J.; LE, Q. V. Learning transferable architectures for scalable image recognition. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2018, Salt Lake City. **Proceedings**... IEEE, 2018. p. 8697-8710. DOI: 10.1109/CVPR.2018.00907.

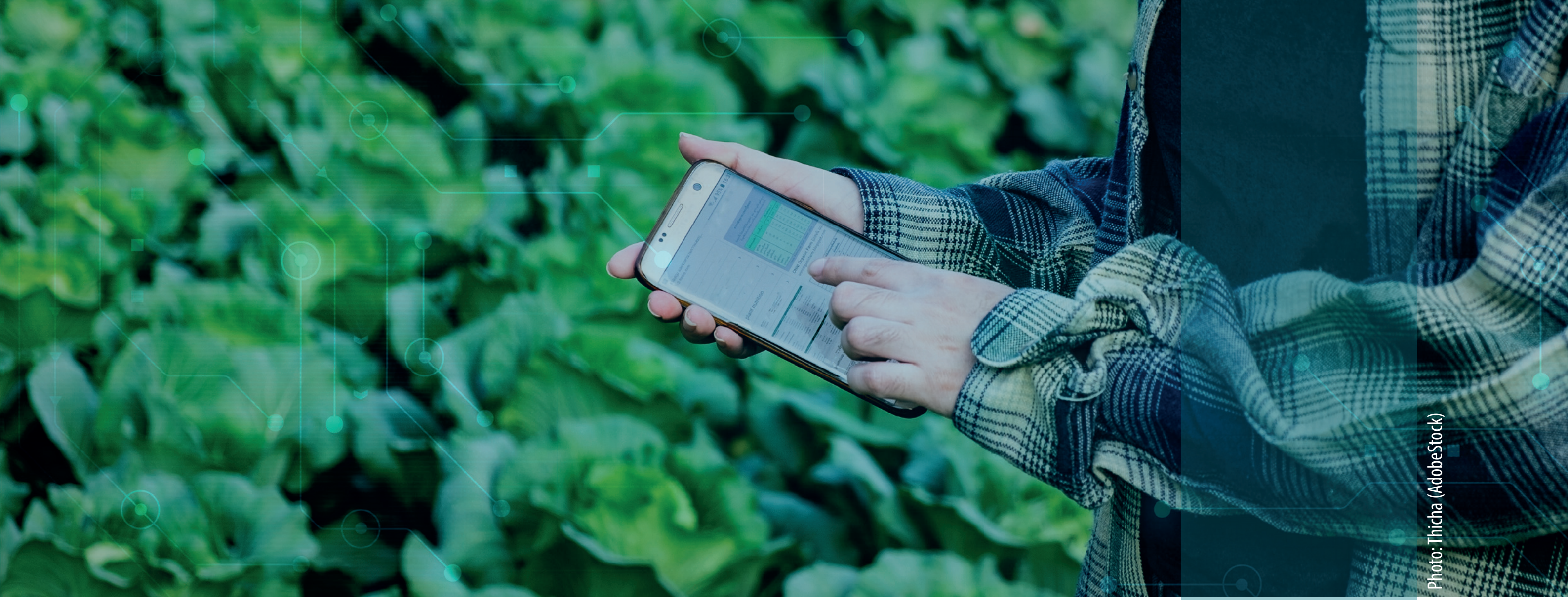


Photo: Thicha (AdobeStock)

7

Technologies developed in precision agriculture

Célia Regina Grego | Eduardo Antonio Speranza | Gustavo Costa Rodrigues | Ariovaldo Luchiari Junior | Laurimar Gonçalves Vendrusculo | Cristina Aparecida Gonçalves Rodrigues | Ricardo Yassushi Inamasu | Carlos Manuel Pedro Vaz | Ladislau Marcelino Rabello | Lúcio André de Castro Jorge | Cornélio Alberto Zolin | Júlio Cesar Franchini dos Santos | Carlos Cesar Ronquim

Introduction

Precision Agriculture (PA) is a management strategy that considers temporal and spatial variability to improve the sustainability of agricultural production (International Society of Precision Agriculture, 2020). Through it, temporal, individual, and spatial data are collected, then processed, analyzed, and combined with other information to support management decisions. These decisions are made according to estimated variability so as to improve the efficiency in resources use, productivity, quality, profitability, and sustainability of agricultural production.

Developing methodologies either in experimental fields or by inference in an ideal environment by agents without field experience can result in proposals that are very difficult for the productive sector to absorb. With that in consideration, we will develop the chapter as the first steps for implementing PA and focus on enabling technologies.

It is important to understand that the crop is not uniform and has variability. There are areas where crops are more prone to flood and others with good drainage, and plots that can vary from clayey to sandy soil, or from more acidic to less acidic soil, and so on. Such variation includes different characteristics that imply productivity variation in the same crop. It is intuitively understood that in areas that produce two to four times more, the need for input is different, and that perhaps there is an excessive application of input in one area, or productivity is being lost due to lack of this input in another. In both situations, the producer is losing an important economic return on his business and possibly an environmental return.

Despite the variability of productivity in the areas, conventional machines are rigid, their adjustments are mechanical and cannot be changed during a field operation. In the case of PA machines, as the adjustment is programmable, the input release can be adjusted according to a recommendation map. Its performance is flexible and is able to react to values obtained by sensors in real time.

Thus, when starting PA with machines able to map crop productivity and enable input application according to geographic coordinates, the cycle apparently closes – that is, the fundamental cycle of control: reading, analysis, and action, as illustrated in Figure 1.

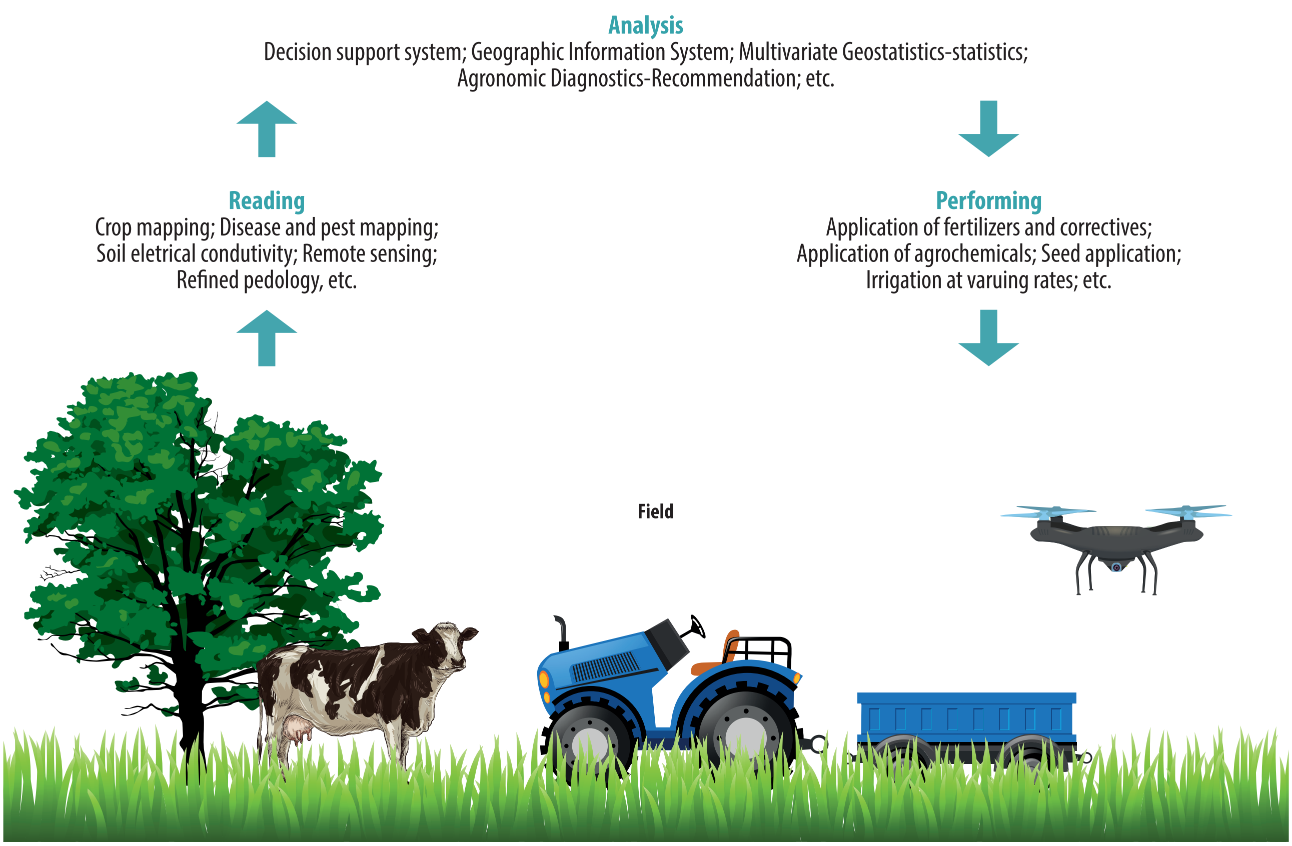


Figure 1. Control cycle used in Precision Agriculture.

It was at this point that PA moved forward, and likely surpassed the "inflated expectation" peak of the Gartner[1] hypercycle, plunging into the "disillusionment" stretch. It was often reported that the machines were being operated in the field with disabled electronic controls.

There were and still are operational problems. The various proprietary formats generated the lack of compatibility of digital systems, both for communication between equipment, such as file and data exchange, are problems recurring operations. But perhaps most important is the fact that the methodologies used in the analysis stage, the "heart" of the control cycle, still being planted in conventional agriculture. The methodologies of Conventional analysis does not take into account the variables that differentiate regions from the farm. The methodologies developed in this chapter, such as those created in onfarm system (procedures by experimentation according to planning inside the farm or on the farm), are an attempt to develop and surprise the absence of this information and methodologies.

[1] Cycle representing maturity, adoption, and social application of specific technologies.

Collecting, storing, and analyzing data in PA

An important part of the PA application cycle is obtaining data and relevant attribute maps that influence the spatial and temporal variability of crop yields. Productivity maps express natural and anthropic soil variations, such as topography, texture, fertility, compaction, and other variations, as well as plants, with responses to the attributes of soil, climate, and crop management.

There are, commercially, many equipment, sensors, techniques, and approaches that can be used to map crop variability in PA (Leroux; Tisseyre, 2019; Molin; Tavares, 2019). The number of information layers will depend on the variability level in the area and the user's interest. However, in general, a suitable approach should include mapping of the soil variability, another mapping of the plants, and at the end of the cycle, a harvest map. For the selection of the most suitable techniques and instruments, the production system, the availability of instrumentation, and the scale or dimension of the plots must be considered.

For soil mapping, the measurement of apparent electrical conductivity (ECa) has proved to be a very useful tool (Corwin; Plant, 2005), as it integrates mineralogical and physical factors such as texture, density, compaction, water retention; chemicals, Cation Exchange Capacity (CEC), and organic matter. Thus, it can identify global soil variability, which can lead to regions of higher (or lower) productivity within the stand. For plant mapping, Vegetation Indices (VI) obtained from satellite images, remotely piloted aircraft (RPA), and active canopy sensors (Lee et al., 2010) make it possible to trace crop vigor, both spatially and temporally (throughout the crop cycle). With these three attributes maps (CEa, IV and productivity) it is possible to establish management zones which can be used for more detailed sampling of additional attributes, such as diseases, pests, and soil compaction. Other variables of great importance in PA are plant nutrients (fertility) and soil texture, which can be obtained in a regular grid or by management zones defined from CEa, IV and productivity maps.

This section presents the procedures and instruments used in data collection in PA, as well as their storage and analysis.

Identification of soil spatial variability

Soil sampling

Soil sampling is an essential procedure in PA, as it enables knowledge on fertility and characteristics that influence productivity. The regular grid procedure is the most used. Therefore, it is recommended that the sampling density be one sample per stand in extensive areas of grain production, for example. In fruit-growing areas, commonly with plots of a few hectares, the sampling density can be a few dozen samples per hectare. Another criterion to be considered is prior knowledge and visual analysis of field variability. In plots with great variability, visually observed and recognized by the producer over the years, the sampling density must be greater than in those where little variability has already been verified. Figure 2A illustrates some real examples of sampling grids, where samples were collected for fertility and soil texture, both in a soil texture system, in soybean-cotton production areas in Mato Grosso (A), and soil potassium maps (B) of these two areas. The plot areas are 110 ha and 200 ha, and the numbers of samples collected were 70 and 135, respectively. That is, both with about 1a sample per hectare and a half. Figure 2B shows the potassium maps of these two areas, obtained by spatial interpolation.

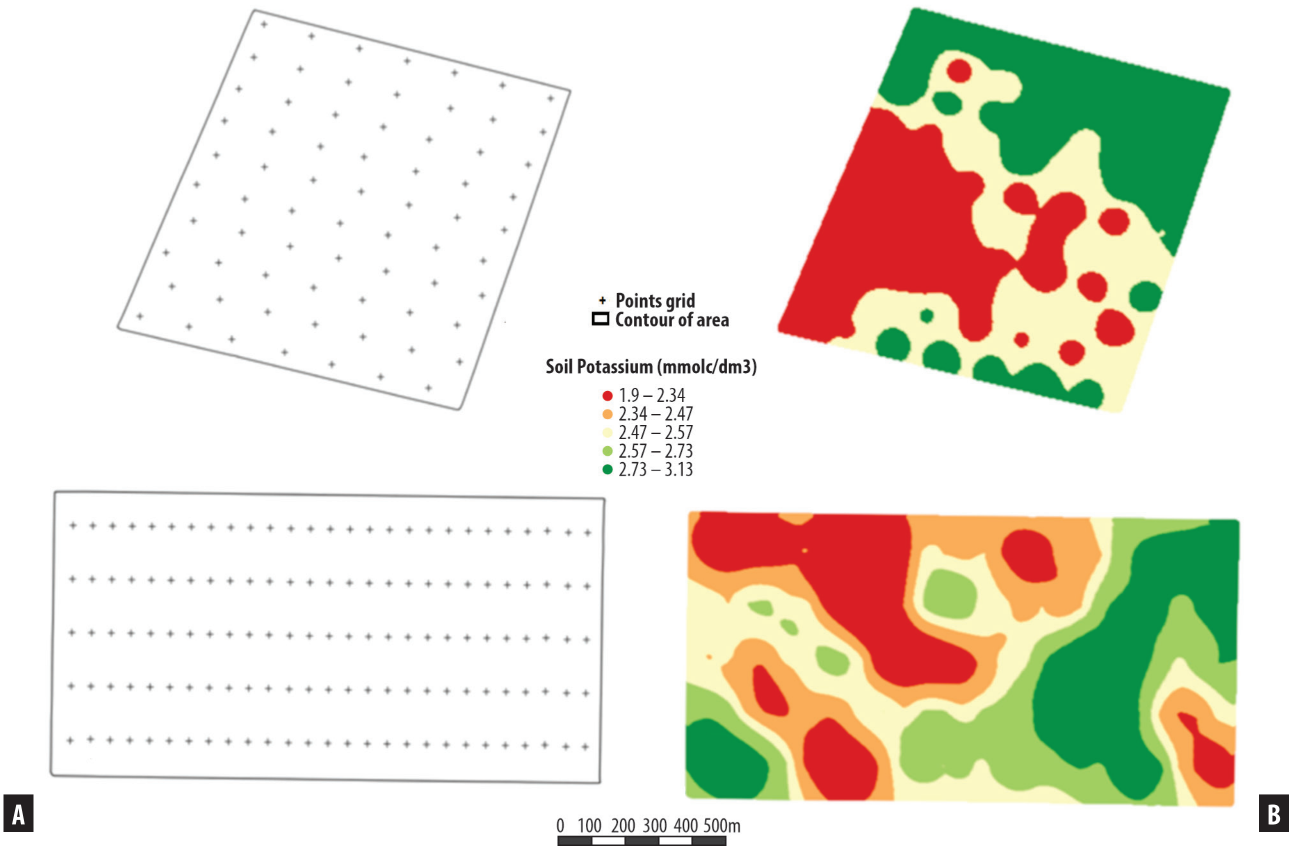


Figure 2. Examples of sample grids in Precision Agriculture, for the determination of fertility attributes and soybean-cotton production, in the municipalities of Pedra Preta, MT, and Sapezal,MT (A) and soil potassium maps (B).

Other important issues considered in PA sampling are related to the equipment used for collection and crop stage. In intensive production systems, where two or three crops are grown in a year, the time intervals between harvesting the first crop and planting the second crop are very short. For the use of quad bikes equipped with soil samplers, the sampling should be carried out after the first harvest (summer) or after the second harvest (winter), avoiding damage to plants when the quad bike is moving in the area. If collection is needed when the crop is in a well-developed phase, manual cutting should be used. In both cases, georeferencing of samples is always necessary. For collection with the quad bike, it is recommended to extract the soil in at least 9 subsamples, in a circle around the georeferenced point, at a depth of 0-20cm. As for sampling with Dutch augers, the number of sub-samples around the point can be reduced to five, since the amount of soil collected per hole is greater.

In the same way as for collecting soil samples, the soil compaction measurement can be done by penetrometers coupled to quad bikes or by manual devices. Equipment coupled to quad bikes are recommended for periods after the summer crop is harvested, since in the winter crop harvest, the soil will have very low moisture, which is inadequate for this type of determination. Figure 3 shows the penetration resistance maps, measured with a manual penetrometer and generated based on spatial interpolation, in plots with cotton crops in the municipality of Pedra Preta, MT, (Figure 3A) and Sapezal, MT, (Figure 3B).

Soil and root sampling for the quantification of phytonematodes should also be carried out during the crop cycle, therefore, the use of quad bike samplers is not recommended, but rather soil treatment and manual collection of roots. Figure 4 shows maps of *Rotylenchulus reniformis* phytonematodes in soil and root, obtained by kriging from 70 samples collected in a soybean production area, in Pedra Preta, MT.

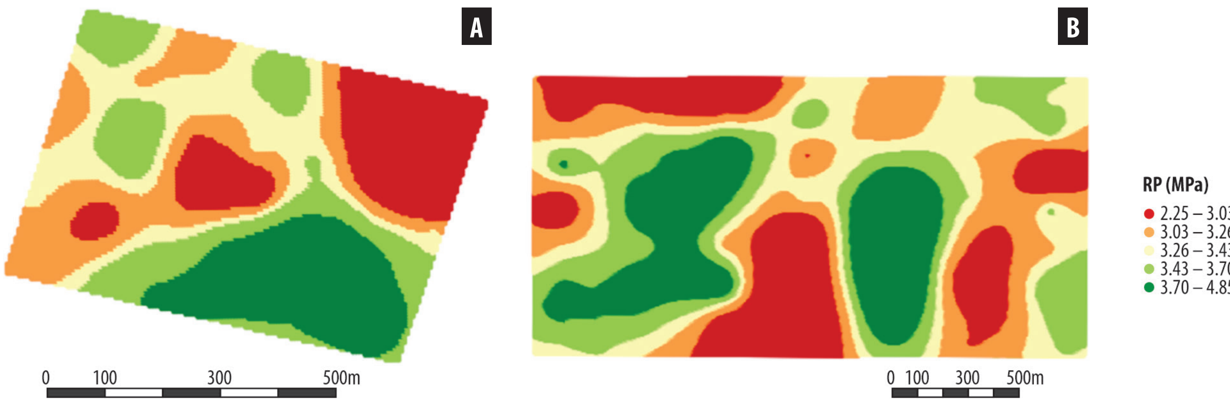


Figure 3. Penetration resistance maps in the 10-40cm layer in cotton planting area in Pedra Preta, MT (A) and Sapezal, MT (B).

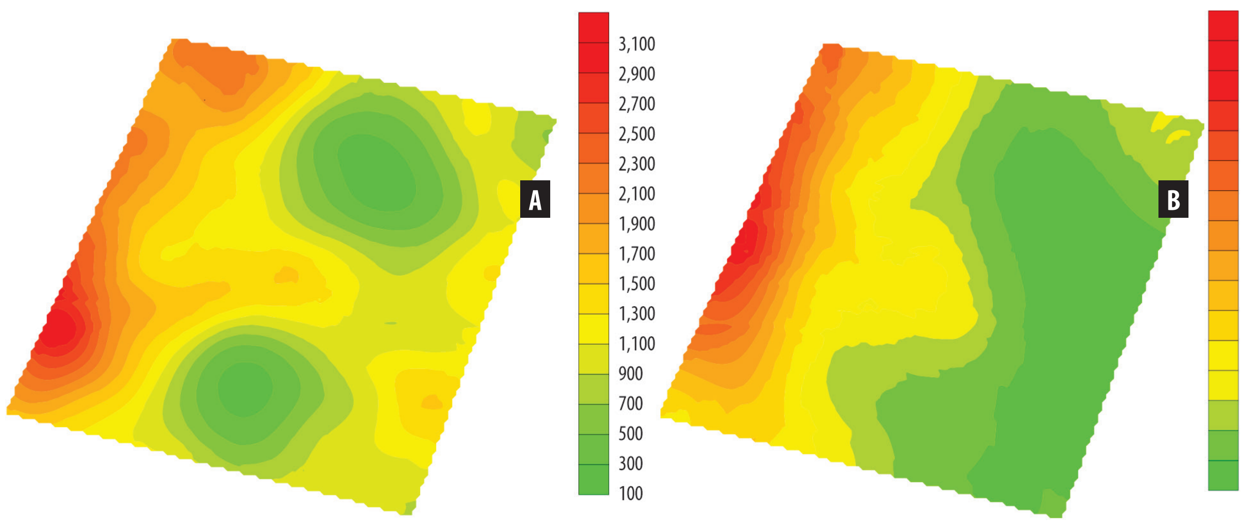


Figure 4. Nematode maps (*Rotylenchulus reniformis*) in soil (A) and in soybean root (B).

Electrical conductivity

Apparent soil electrical conductivity (AEC) is used at field scale to map the spatial variability of numerous edaphic properties, such as texture, salt concentration, and moisture. This tool is faster, more reliable, and easier to use compared to other techniques, and it is often correlated with crop yields. Therefore, it is widely used in PA research for the spatiotemporal characterization of edaphic and anthropogenic properties that influence crop productivity. AEC measurements are usually obtained from the method known as the four-point system (Smtis, 1958), which consists of using four metal electrodes sequentially aligned with known spacing (Figure 5).

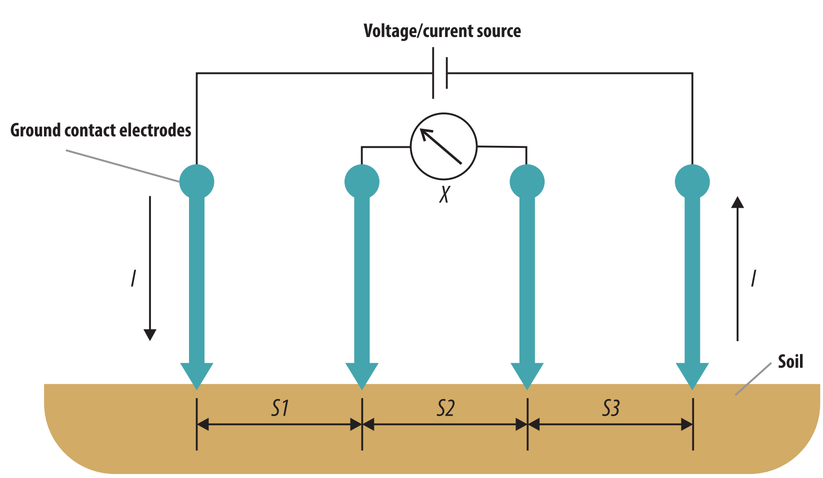


Figure 5. Four-point system for measuring apparent soil electrical conductivity.

Source: Rabello et al. (2014).

The market has systems that have already been developed for measuring AEC with the traditional four-point system and the magnetic induction system, both manufactured abroad. Embrapa has also developed an AEC measurement system based on the four-point system.

AEC provides high density data, enabling a quick overview of the area that allows dividing it into homogeneous regions, facilitating its interpretation and, consequently, management decision-making.

Identification of crop spatial variability

Productivity mapping

To analyze spatial variability of productivity using PA, a set of production values is registered in geographic coordinates and stored by the device during harvesting. System integration is conceptually simple. It consists of a mass sensor, a GNSS receiver, and a data logging system. The first system appeared in a grain harvester. Currently, a force sensor board (similar to electronic scales) located where the grains are loaded by elevators is used for mass measurement. The greater the grain flow, the greater the impact on the board. These values are accumulated in memory and, for each coordinate sent by the GNSS receiver, the accumulated values are integrated and recorded. Each manufacturer records this differently, and there is no standard. However, all of them store at least the geographic coordinate, the time, and the mass value. The force sensor is like an electronic scale and must be calibrated frequently. Some current harvester models perform self-calibration, but this process is commonly performed at each harvest.

Since each manufacturer has developed their own processes, many applications and file formats are proprietary. There are standardization efforts, and it is hoped that incompatibility between systems will no longer be a problem for the producer. There are difficulties in the field. In harvests carried out in fleets, which are common in exporting regions such as the Cerrado biome, the data composition is not possible if one or more machines do not have mapping capacity. The whole process hinders visualizing the production. Therefore, PA harvesting requires attention and, above all, dedication. In the case of the first crop harvest, this operation is tense, caused by waiting until grain moisture has reached the correct point, with the possibility of rain interrupting the operation (harvesting occurs in the middle of the rainy season), and the need to prepare the area to plant the following crop.

It should be noted that if the planting does not take place at the best time/moment, the crop will not express maximum productivity. There are other important factors to achieve differentiated productivity, but it is important to emphasize that the window of opportunity is narrow and time management is directly related to the efficiency of the crop. Therefore, it is understandable that a machine calibration operation may not be a priority for the vast majority of producers, especially in reduced, efficient and lean teams.

The first crops in which machine yield measurement took place commercially were grains such as corn, soybeans, and wheat. They are large machines. In addition to grains, coffee, and cotton also have harvesters capable of recording productivity – noting that, in the case of cotton, these are not impact sensors and are based on radars using infrared signals, and more recently, microwaves. For sugarcane, productivity sensors in machines are not yet commercial successes, since it is challenging and complex to measure this crop automatically. This is due to the difficulty in developing sensors that can accurately identify the biomass flow. In crop cultivars that do not have harvesting machines with a production sensor, maps have been obtained through sampling, similar to soil sampling. A collection protocol with the coordinates of each sample and its weighing is needed, and spatial dependence is also required to enable a precise interpolation process. Some academic works can be found, but not in the sugarcane crop, they are not yet available in the market. Camera sensing and 3D reconstruction are considered to be promising (Santos et al., 2017).

Pre-processing is essential for correct interpretation once harvest data are obtained. Figure 6 illustrates, in a geographic information system, actual cotton harvest productivity data collected from a machine. Each point illustrates the coordinate and the collected value, grouped into four classes, ranging from red (low production) to orange, light green and more intense green (maximum production). The first four Figures 6A, 6B, 6C and 6D, come from four different files, either collected from different machines or performed at different times; Figure 6E illustrates the composition of the four files in one. Harvesting is carried out interspersing planting lines. There is an interruption during the harvesting process, and maneuvering and displacement may also occur. In these cases, if the operator keeps the registration system turned on, there is zero productivity registration, so it is important to exclude these points.

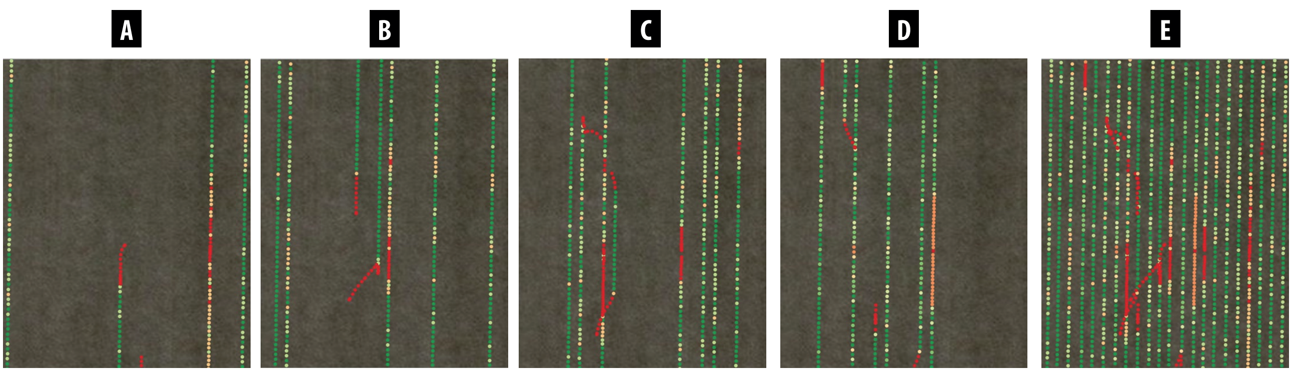


Figure 6. Illustration of harvest data (A, B, C, D) and its composition (E).

Proximal sensing

Various types of sensors have been developed for the acquisition of soil and plant monitoring data allowing efficient data generation a lower cost. These provide reliable estimates of crop development and improve the estimation of production potential. For the vegetative monitoring of plants, the vegetative indices (IV) NDVI and NDRE (Red Edge with normalized difference) can be obtained by reflectance sensors, such as Crop Circle® (Holland Scientific, Lincoln, USA).

Light Detection and Ranging (LiDAR) technology is a proximal remote sensing methodology based on optical concepts. The main objective is to identify distances to a target object, which are determined by time differences between the emission of a laser pulse to a target object and the detection of the signal reflected by that object (Reutebuch et al., 2005). Using a LiDAR sensor allows to quickly and accurately reconstruct three-dimension objects, which makes this technology feasible in various agricultural activities carried out by land implements equipped with automation. In PA, LiDAR sensors can be embedded in spray implements, together with reflectance sensors, for plant height detection and online application of growth regulators (Figure 7).

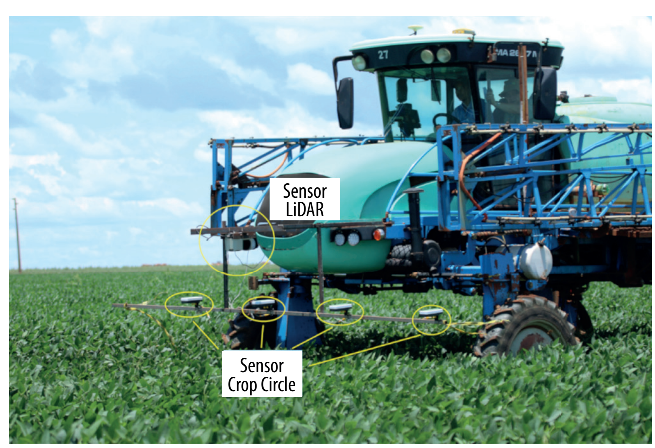


Photo: Ricardo Yassushi Inamasu

Figure 7. Use of LiDAR sensor in conjunction with Crop Circle® reflectance sensors, embedded in an implement for plant height detection.

Suborbital and remote sensing

Sensors with free images, available from the internet, with spatial resolutions between 10 m and 30 m, and temporal resolutions between 5 and 15 days, such as Sentinel-2/MSI and LandSat-8/OLI, can support various PA activities, mainly in more extensive plots. In Brazilian agriculture, these areas are concentrated in sugarcane and grain producing regions. For these regions, remote sensing can analyze stand spatial variability based on vegetation indices (VI), which can be obtained by combining the bands in the visible and infrared spectrum.

Analyses on biomass presence can be performed with VImaps obtained from combinations with infrared spectrum bands. Through these indices, it is possible to establish correlations with the availability of nitrogen and other nutrients in plants, monitor the evolution of crop growth, and perform productivity estimates in different regions within the same field (Candiago et al., 2015). The presence of water in plants can also be spatially assessed using index maps that combine visible spectrum bands with short-wave infrared spectrum bands, such as the Normalized Difference Water Index (NDWI) (Zhang et al., 2019). Combined with data related to soil variability and crop yield, VIs can be important for area subdivision strategies for specialized treatments, such as the design of management zones (Figure 8) and on-farm experimentation.

Remote sensing images with spatial resolutions below 10 m are not yet available for free. However, the need to obtain more accurate data for PA activities, which allow, for example, identifying pests, diseases, and the presence of invasive plants, and to estimate productivity, increased the use of multi and hyperspectral cameras onboard remotely piloted aircraft (RPA). The use of such devices has intensified in recent years (Jorge; Inamasu, 2014) due to the falling cost of this technology, gradually increasing its acceptance by producers.

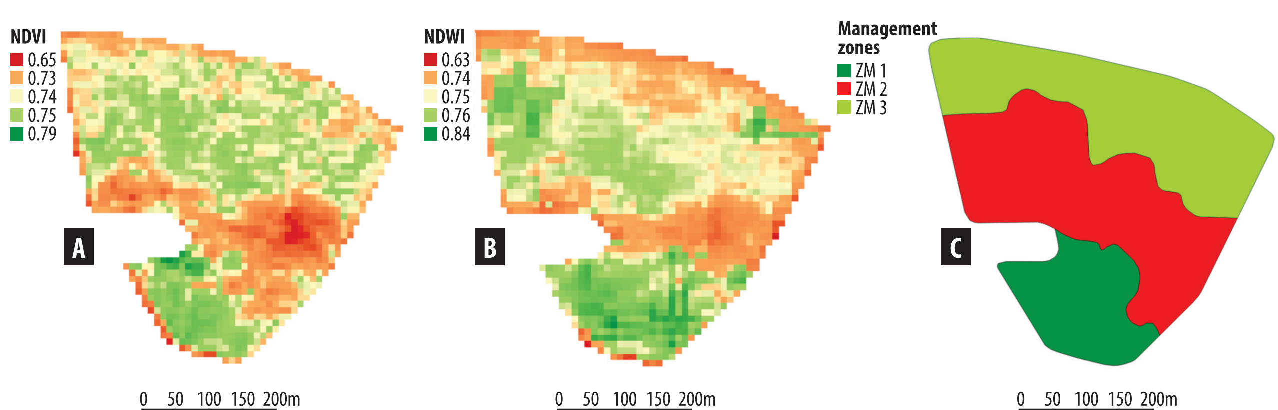


Figure 8. Sugarcane area maps: NDVI (A) and NDWI (B) composition from Sentinel-2/MSI satellite images between February and May 2017; also, management zones map (ZM) (C) from soil and crop attributes between 2012 and 2016.

In addition to reducing costs, the use of RPA in PA also gives the producer the advantage of planning the data collection time, thus avoiding problems with rainy days or high incidence of clouds that often occur with satellite data. Another important factor is the possibility of planning the height of the flight, which allows obtaining images with different spatial resolutions using the same camera. This allows to identify, for example, planting failures that remote sensing images cannot capture (Figure 9). RPAs that include performance functions also allow the producer to apply agricultural inputs and correctives in a localized manner, making them an interesting alternative to the conventional automation performed by traditional agricultural implements (Mogili; Deepak, 2018). However, the operationalization of RPA still needs to be performed by professionals specialized in most applications. Another negative point is the autonomy and imaging capacity of the equipment. A single RPA has the ability to generate images covering a few hundred hectares in a single day, images that must later be processed by high-performance computers in order to generate mosaics.

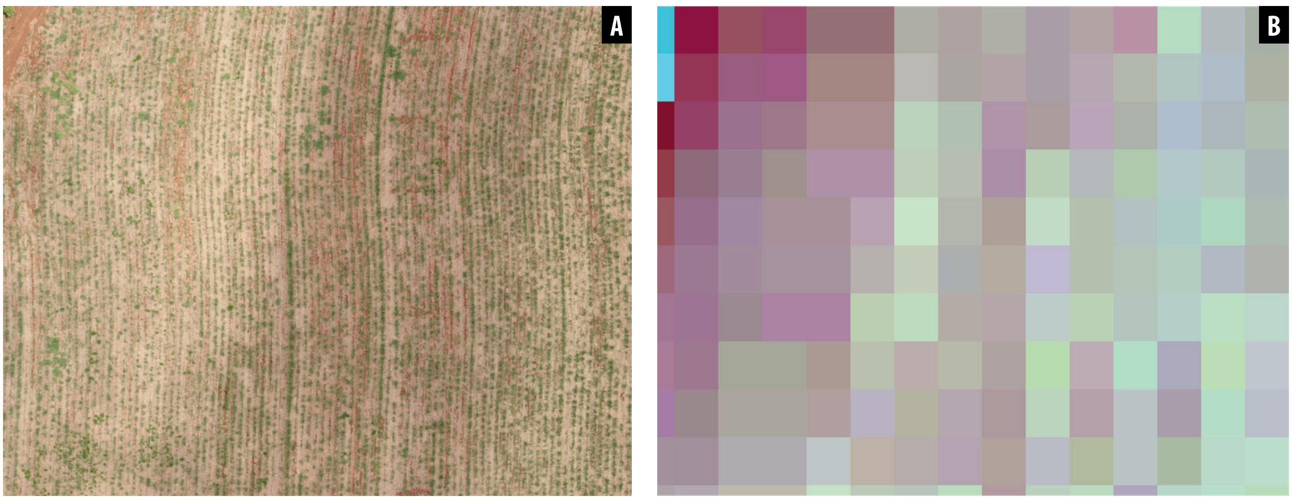


Figure 9. Example of cut-out images, with a sugarcane stand with an area of about 120 m × 120 m taken in October 2019 using: RPA with embedded camera, with a spatial resolution of 2 cm/pixel (A); and Sentinel-2/MSI satellite, with a 10 m/pixel resolution and cloud incidence (B).

Following the evolution of RPAs and multispectral cameras, aerospace technology private companies have placed satellites and, more recently, constellations of nanosatellites equipped with sensors capable of capturing images with a spatial resolution below 1 m, into orbit. The current market allows producer associations to different platforms for accessing these images, already in the form of ready-to-use products, usually at a high cost. However, unlike what happens with ARPs, it is possible to obtain images that cover large areas in one day. In summary, both satellite images with submetric spatial resolution and images captured by RPAs still require investments by the producer, which must make up a cost-benefit relationship, which depends on the application. In this relationship, issues such as the time when one wants to obtain the images and the need for spatial resolutions must be taken into account.

Cloud data storage

Embrapa's PA Network (Rede PA) is committed to raising awareness among researchers and partners in order to encourage the sharing of final data produced in their research. Thus, the GeoNode tool[2], a free and open-source software, was evaluated and adapted to absorb the specific requirements of the community, providing the development of a new version of the PA Network data repository[3]. Metadata cataloging is the main functionality of the repository to ensure data integrity over time. Based on the metadata, it is possible to verify authorship data, the equipment and methodologies used, the environmental conditions, and the difficulties encountered during data collection. This information is an important input for future reuse and analysis for any given dataset.

[2] Available at: http://geonode.org

[3] Available at: http ://www.redeap.cnptia.embrapa

Data analysis in PA

Data mining and pattern extraction

Computational techniques for data mining are extremely important for the data analysis process. The main objective of these techniques are to discover patterns in databases using tools made available by the use of artificial intelligence, machine learning, and statistics (Majumdar et al., 2017). In this context, unsupervised (grouping) and supervised (classification) learning methods are normally considered.

In PA, there is a growing popularization of algorithms derived from clustering and classification methods used for different analyses based on the data. Classifiers such as artificial and convolutional neural networks are used for the semi-automatic identification of various events that occur in crops, such as planting failures and the appearance of pests and invasive plants, based on images (Tang et al., 2017). The productivity of a crop can also be estimated from regressors and vegetation multispectral data (Al-Gaadi et al., 2016). Grouping methods, on the other hand, are fundamentally used for the subdivision of cultivated areas into regions with similar productive potential, known as management zones (Luchiari Júnior et al., 2000).

Regardless of the method used, this pattern extraction analysis is part of a broader process known as Knowledge Discovery in Databases (KDD). Thus, previous steps, such as filtering, cleaning, normalizing the data; as well as subsequent ones, such as the statistical validation of applied models must be performed before and after the use of data mining methods. The more complete the KDD process used to analyze the data, the greater the chance of obtaining more accurate and reality-consistent results.

Filtering tools and data cleaning

By means of data filtering and cleaning tools, data samples obtained in the field with positioning errors or associated outliers can be eliminated from the data set in order to be analyzed, thus avoiding interpretation errors in later steps. In PA, the greatest concern is with the productivity data obtained by harvesters, due to their high sampling density and the diversity of manufacturers. Thus, algorithms and software based on statistical methodologies and variability parameters informed by the user can be used for this task (Sudduth; Drummond, 2007; Vega et al., 2019). In addition to productivity data, any other dataset from proximal sensing subject to collection errors must go through this cleaning and filtering step before being used in the data-mining step.

Geostatistics and spatial interpolation

Geostatistical analyses applied to PA are essential to ensure better accuracy in the mapping of interpolated data. According to Vieira (2000), Geostatistics is a tool set that allows to analyze the degree of spatial dependence of varying data in space, whether in thousands of hectares or in a small plot, such as an experimental 30x30m-plot, as shown in Grego and Vieira (2005).

Geostatistics assumes that the greater the number of samples, the better the representation of the real spatial variability expressed by the tool. However, it is known that, in practice, it is necessary to meet the needs of the user, mainly considering available resources, labor, and operational time for sampling. To assist in this step, it is possible to use historical information about the area and data obtained from sensors and satellite images, thus optimizing the number of samples.

With the georeferenced data, an investigation is carried out as to the existence or not of spatial dependence, and, if so, it is possible to interpolate data by kriging, which guarantees minimum variance and non-bias in the interpolated values. The result is based on accurate maps in which spatial variability patches are observed, these can be correlated to form a spatial information platform during crop cycles. This mapping, as detailed in Bernardi et al. (2014), helps to identify differentiated management zones for localized application of inputs.

Outlining management zones

Under PA, activities like planting, interventions such as the application of inputs and irrigation can be uniformly managed, based on the delimitation of sub-areas known as management zones (MZ). A MZ can be defined as a portion of land that is stable over time, where the production potential, the efficiency of input use and the risk of environmental impact are essentially uniform (Doerge, 1999; Luchiari Júnior et al., 2000). In order to obtain MZs with these characteristics, the main prerequisite is using non-anthropogenic attributes related to the genesis and culture (Molin et al., 2015). Thus, factors such as relief, electrical conductivity, texture, physical attributes of the soil, biomass indices, and historical productivity must be used to support the delineation of ZMs (Kitchen et al., 2005; Li et al., 2007; Scudiero et al., 2013).

It is now possible to generate ZM maps by combining numerous datasets and machine learning algorithms, offering the producer greater precision and confidence. Due to the current high availability of data, it is necessary to adapt these algorithms so they can handle massive data sets and make the most of the computational processing capacity available in the environment in which they are executed. Despite this need, these efforts are at an embryonic stage.

On-farm experiments, spatial correlation, and recommendations

The on-farm experiment consists of defining virtual plots within an experimental cultivation area in order to evaluate different application rates with repetitions in interventions such as planting (population), nitrogen fertilization, and growth regulators. The plot procedures must be carried out according to the producer's planning and with the available agricultural implements – hence the use of the term on-farm (or on the farm) (Shiratsuchi et al., 2014). The experiments must be carried out for a few seasons, so when considering the crop yield as the final result, it is possible to establish adequate and spatially differentiated recommendations for population and input doses. Intervention recommendations can be obtained based on the use of spatial correlation analysis tools. In PA, the determination coefficient (Nagelkerke, 1991) is a result of a linear regression model that allows identifying the correlation and trend values (Figure 10).

The correlation measures between attributes are an important support for establishing recommendations in areas with specific characteristics. In soil homogeneous areas free from pests, for example, it is possible to infer adequate nitrogen rates that should be applied until reaching a maximum yield threshold for the crop under those conditions.

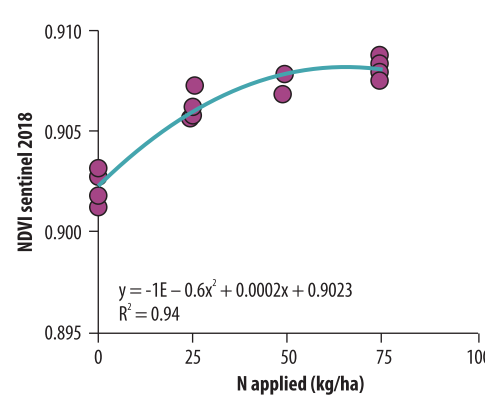


Figure 10. Example of correlation coefficient R^2, obtained from averages of NDVI and nitrogen (N) values, applied in on-farm experimental plots.

Practical applications in precision agriculture

Characterization of the spatial variability in southern Minas Gerais specialty coffee production systems

In order to spatially assess the representative areas of Arabica coffee (*Coffea arabica*) production, which is classified as special in southern Minas Gerais, Rodrigues et al. (2019a) used the vegetative and chlorophyll indices to identify the existence of spatial variability for the application of Precision Coffee Growing. In the 2017 post-flowering period, data related to biomass, chlorophyll, and plant altitude were collected in two farms in southern Minas Gerais. The occurrence of spatial variability was observed, and at Fazenda Santa Cruz, an inverse relationship was observed between altitude and NDVI, with the highest NDVI values found in the lower areas of the field. These facts indicate that management techniques must be carried out due to this variability. At the Morro Alto Farm, the right-side sun exposure was more uniform (Figure 11).

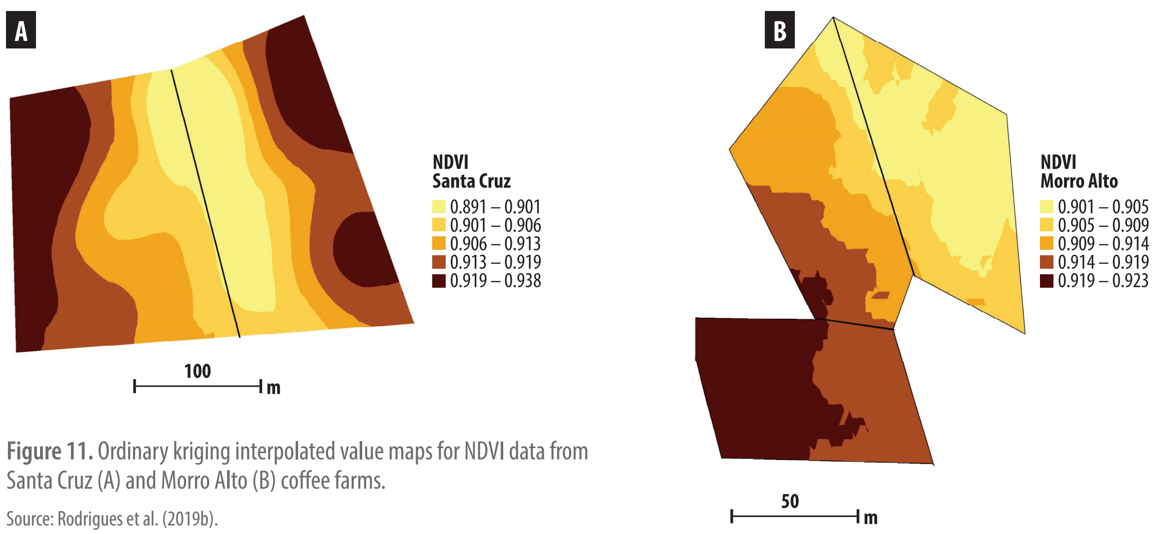


Figure 11. Ordinary kriging interpolated value maps for NDVI data from Santa Cruz (A) and Morro Alto (B) coffee farms.

Source: Rodrigues et al. (2019b).

In order to define which of the variables were able to assist in the delineation of the MZ in this study, Speranza et al. (2019a) correlated maps of ZM potentials with sun exposure maps in each area, allowing a differentiation of the coffee's quality. There was high correlation of sun exposure faces with the IRC index, and a better overall performance of the NDRE index in relation to the NDVI for this context. This difference can be explained by the fact that data were collected at the beginning of the reproductive period, right after flowering, when there was a mixture of fully expanded leaves and others with initial vegetative growth. Both Rodrigues et al. (2019b) and Speranza et al. (2019a) emphasize that the spatial responses of coffee in relation to vegetation and chlorophyll indices will be complemented in planned future analyses, considering the different phenological phases of the coffee tree. Productivity and beverage quality maps will also be generated and correlated with biophysical and microclimatic variables.

Spatial and spectral behavior in sugarcane and its correlation to soil electrical conductivity

Vegetation indices (VI), obtained from satellite images, are powerful tools that for years have been monitoring and providing near real-time information on agricultural crops, especially sugarcane. Soil AEC presents similar spatial behavior with VIs obtained both by remote sensing (Rodrigues et al., 2019a), by proximal and suborbital sensing (Speranza et al., 2019b), and also by sugarcane production (Sanches et al., 2019). In this context, Embrapa's Precision Agriculture Network, through a partnership with Usina Santa Cruz, belonging to the São Martinho group, develops on-farm experiments within a 15.7-hectare sugarcane plot (Grego et al., 2019; Rodrigues; Rodrigues et al., 2019a). Through geostatistics and machine learning algorithms, two simplified management zones (Mzs) were identified for the area, which reflect the soil texture: MZ 1 – clayiest area (55% of the area), in the lowest part, on the west side of the stand, but with greater variability in relation to AEC; and MZ 2 – sandy area (45%), in the highest part, on the east side of the stand and with less variability in relation to AEC (Figure 12). The difference between MZ 1 and MZ 2 for TCH was approximately 16 t/ha.

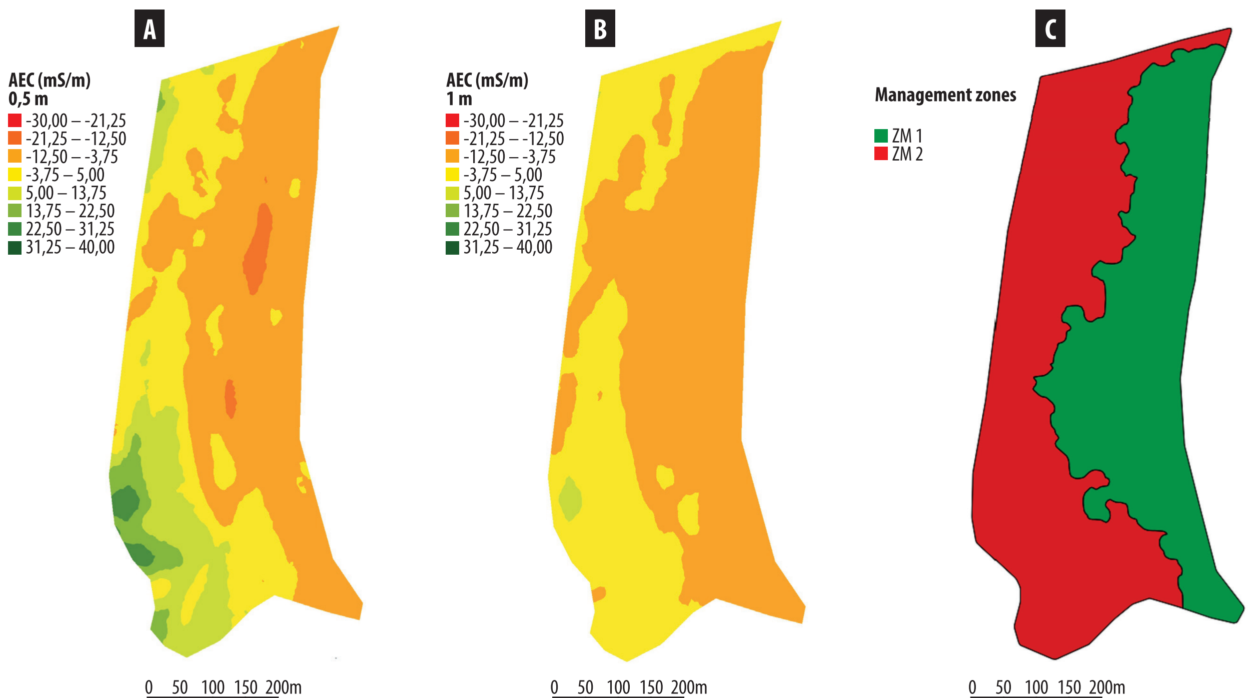


Figure 12. Apparent Electrical Conductivity (AEC) in two coil spacings: 0.5 m (A), and 1 m (B); delineated management zones (ZM 1 and ZM 2) (C).

Source: Adapted from Speranza et al. (2019b).

The MZ design for this area (Figure 12C) allowed programming more dense soil samples in MZ 1, with greater soil variability; and less density in MZ 2, with less soil variability, supporting the definition of onfarm experimentation. Measuring the AEC of the soil indirectly optimizes the spatial sampling of cultivated areas, promoting savings in the amount of collections and costs for soil analysis.

To correlate AEC to VIs from remote sensing images, 18 images were collected from the study area, between March 2018 (planting month) and July 2019, from the MSI/Sentinel-2A and 2B sensor. The results indicated a significant correlation between soil AEC (0.5 m and 1.0 m) and the VIs for most of the evaluated dates. The correlation between the VIs and AEC variables was positive from March to September 2018, that is, the VIs tended to move in the same relative direction (but not necessarily at a constant rate) until approximately 223 days after planting. From that period until the final date of sugarcane harvesting, the correlation was negative (Figure 13). There was similarity in spatial behavior and correlation values with the soil vegetation and AEC indices. Therefore, soil AEC can be an indicator for different managements in the same sugarcane crop.

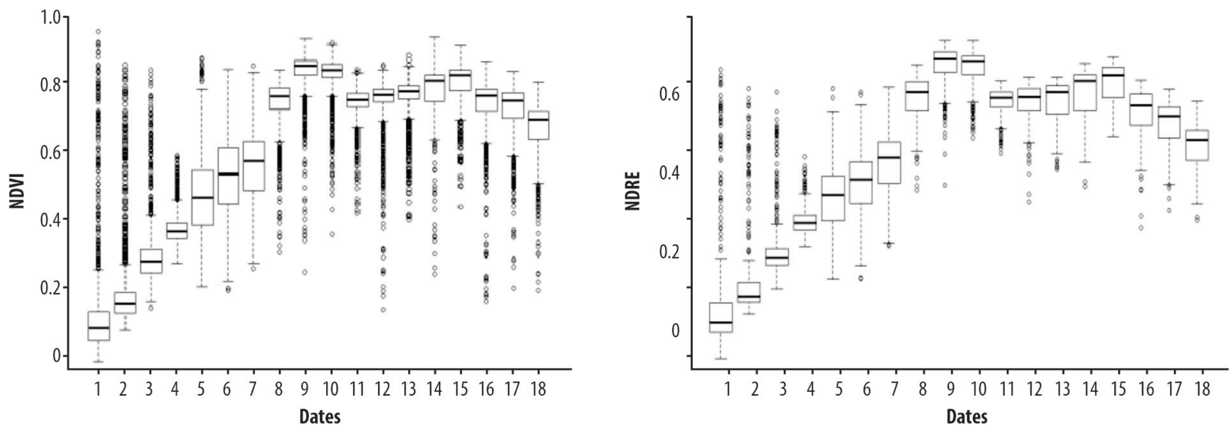


Figure 13. Boxplot of NDVI and NDre2 values on 18 observation dates of VIs in a sugarcane plantation in Ibaté, SP.

Source: Rodrigues et al. (2019a).

PA technologies in fiber and grain management systems in the state of Mato Grosso

The cotton-growing regions in the west and southeast of Mato Grosso are those that most use instruments: light bar, automatic pilot, and section control (sprinklers and seeders). The most used implements are the autopilot (61%) and the sprayer section control (58%). The average rate of PA instrumentalized techniques in this state is 42%, with the highest rate (54%) in the west region. The western region is also above average in the use of other management instruments under the PA approach. In it, 51% use fertility maps, 22% harvest maps, 27% maps of pests, diseases and invasive plants, 49% use variable rate application, and 49% applications use management zones.

Due to the large amount of equipment, there is an urgent concern to continuously train the field workforce in activities such as regulation, preventive, and corrective maintenance, machinery technology, and agricultural operations. After training, the mid-northern region of the state had the highest rate in efficiency (81%), compared to the average rate of 71% for the entire state. The study points out that most properties in the state use some of the PA techniques, however the lack of qualified labor not only for field operations, but also for data analysis represents a limitation for the maintenance and growth of this management approach.

Another important point refers to the systematic collection and storage of physicochemical soil analysis associated with stand productivity for soil management effectiveness. In Figure 14, the high correlation of CTC and organic matter, at both depths, reflects the importance of applying fertilizers in order to avoid sudden pH changes, which sometimes require higher doses of lime (Ronquim, 2010).

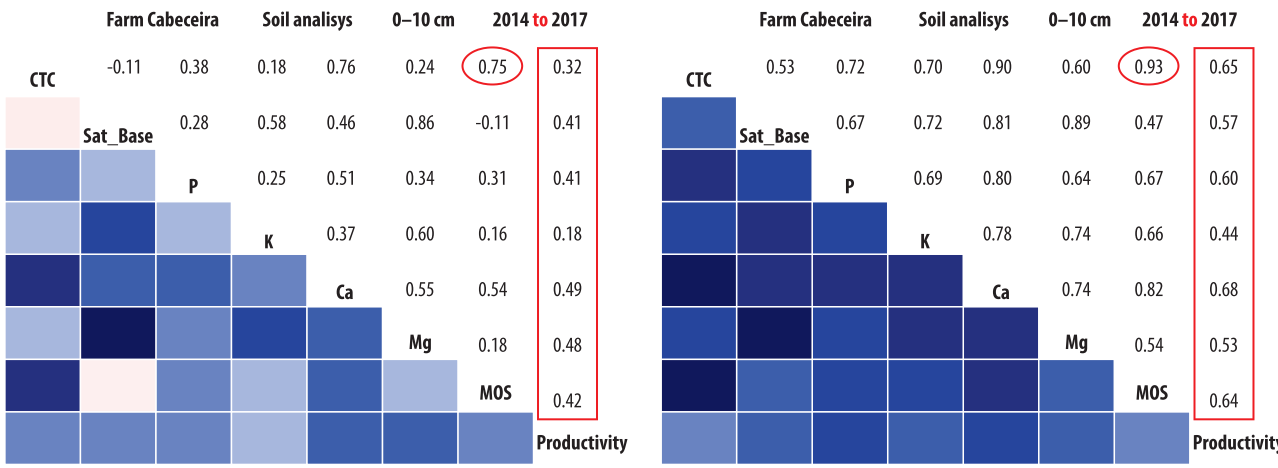


Figure 14. Correlation matrix of soil variables at depth of 0–10 cm (A) and 10–20 cm (B) in a property located in the north of Mato Grosso state.

As a way of optimizing the chemical control used for identification of weeds there is the use of processed images of ARPs, which allow specific use in areas with infestation of invasive plants (Figure 15). The use of herbicide application maps significantly reduces the number of shares, directly impacting the reduction of the cost of culture production.

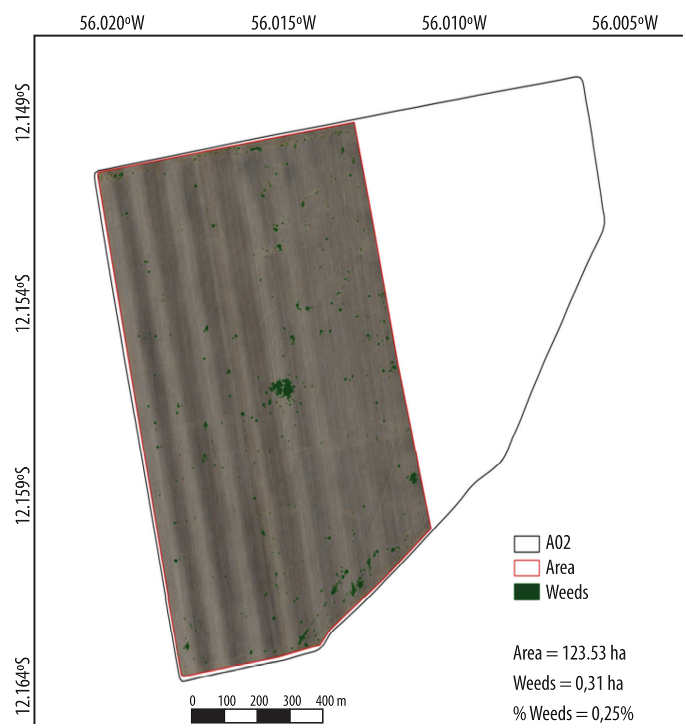


Figure 15. Processed ARP image for identification and geolocation of volunteer corn and bitter grass on a property located in northern Mato Grosso.

Despite having high-performance machines equipped with automatic pilot, managing variability for grain and fiber crops is not common. Flat areas have spatial variability that does not encourage producers to apply Precision Agriculture. Another factor that discourages producers relates to machine versions and models. In this region, two plantings are interspersed in a year. In the case of cotton, soybean is the crop used in the rotation. Each plot has an average size of 200 hectares. Soybean harvesting is carried out with a fleet of machines, and not all have a harvest monitor, therefore, maps are incomplete. This is not the case with cotton, where, due to good economic returns, it is possible to find similar models during harvest, allowing acquiring complete maps.

In 2018, Embrapa's Precision Agriculture Network began a partnership with the Instituto Mato-grossense do Algodão (IMAmt) for research in four large producer experimental plots in the municipalities of Sapezal and Rondonópolis, in the state of Mato Grosso. The 2018 and 2019 cotton crops were monitored by obtaining AEC, soil texture, fertility, phytonematode distribution; and productivity maps, IR maps came from remote and suborbital sensing. Currently, three of the four studied plots have adopted the on farm experimentation process, where it is now possible to visualize responses in relation to the application of different doses of nitrogen fertilization in the virtual plots (Figure 16).

The collected data are corrected, filtered, and made available in the cloud for access and analysis by the Embrapa and IMAmt work teams, as well as by the producers. Although the recommendations are applied in specific experimental plots, they can be extended to plots with similar characteristics, increasing the adoption of PA by producers in the region.

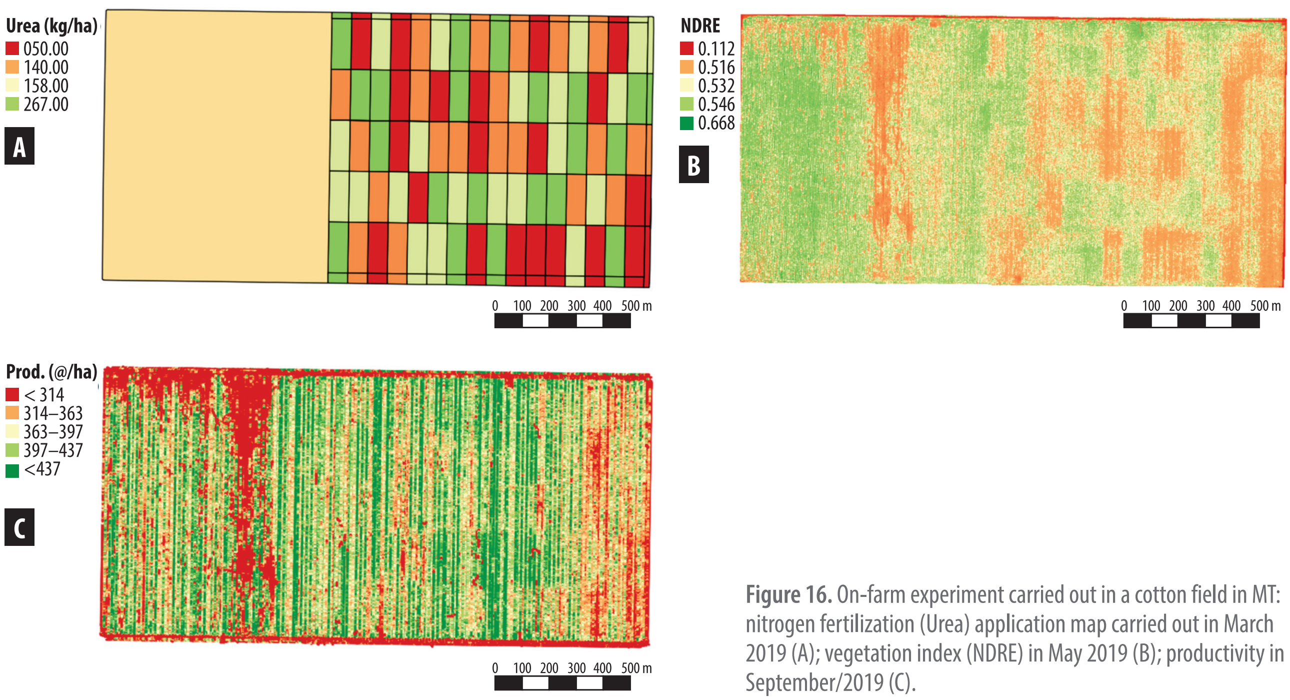
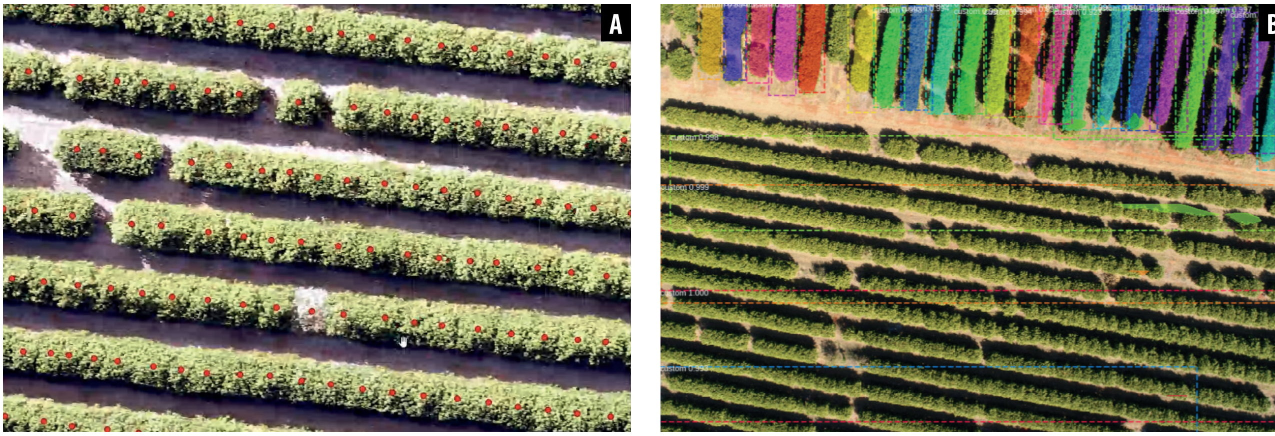


Figure 16. On-farm experiment carried out in a cotton field in MT: nitrogen fertilization (Urea) application map carried out in March 2019 (A); vegetation index (NDRE) in May 2019 (B); productivity in September/2019 (C).

RPA applications in different crops

The applications of RPAs in agriculture have increased with the advancement of technology and available sensors, highlighting the estimation of biomass and productivity, nutritional assessment, pest and disease detection, assessment of plant water requirement, and soil mapping with RGB, multispectral, hyperspectral, LIDAR, and thermal sensors, among others (Hatfield et al., 2008; d'Oliveira et al., 2020). By assessing reflectance values in certain regions of the electromagnetic spectrum it is possible to observe differences between plants and soil and between healthy green vegetation and vegetation with nutritional and water deficiency or attacked by pests and diseases (Jorge; Inamasu, 2014).

Some of the most prominent applications are the monitoring of plant vigor with vegetation indices, studies for nitrogen fertilization using RPAs equipped with sensors, the use of multispectral aerial images to assess the spatial variability of soil and biomass, as well as the cotton, soybean, and corn yield estimates. Also with the data obtained by RPAs equipped with image sensors and LiDAR, studies for plant counting stand out. Figure 17 shows an area of citrus with automatic counting by deep learning techniques, described by Osco et al. (2020a). Figure 18 shows the height determination of cotton plants by LIDAR in a Lidar-equipped RPA to develop a growth regulator application methodology.



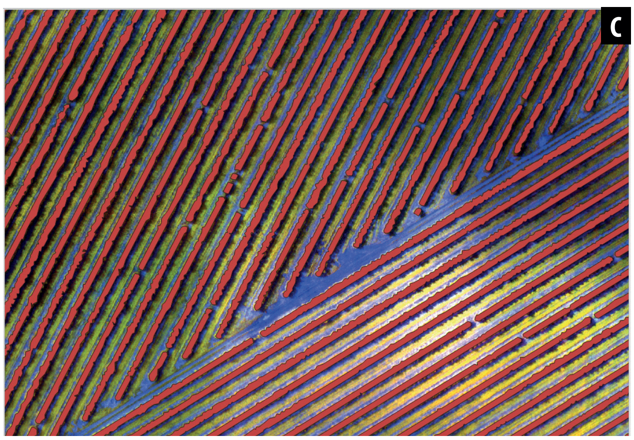


Figure 17. Plot of a dense citrus stand with identification of plants (A) and lines (B, C) by deep learning techniques.

Photo: Lucio André de Castro Jorge

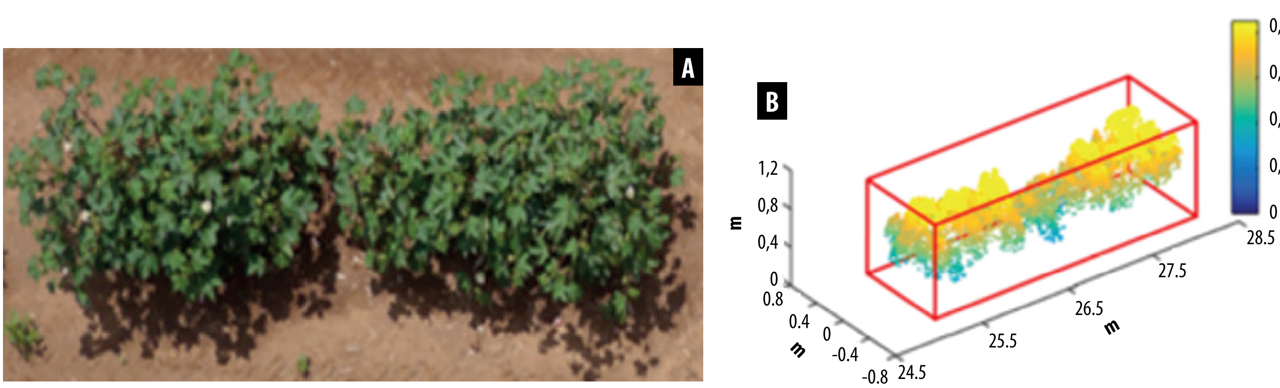


Figure 18. RPA image (A) and LIDAR point cloud (B).

Source: Adapted from Sun et al. (2018).

Sensor-equipped RPAs have evolved considerably and have enabled to study spectral crop signature based on hyperspectra, as in Figure 19A, while assessing the presence of pests in the crop right at the beginning of the infestation (Osco et al., 2020b). Figure 19B shows the result of the hyperspectral analysis when pod caterpillars (*Spodoptera eridania*) occur in the first few hours.

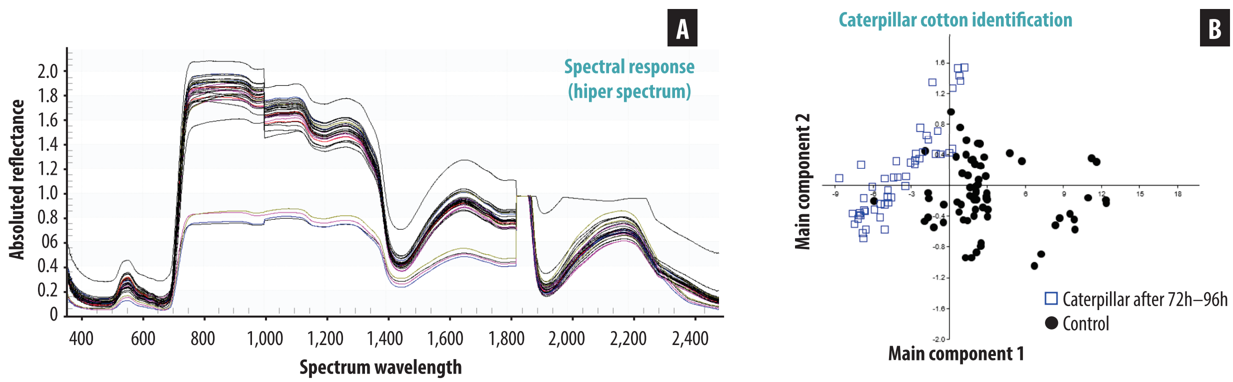


Figure 19. Pixel-by-pixel Hyperspectral RPA images (A) and hyperspectral analysis when pod caterpillars (*Spodoptera eridania*) occur in the first hours (B).

Final considerations

In summary, as regards PA, it is important to understand that crops are not uniform and, therefore, the spatial variability must be considered so that the producer has economic and environmental return on his property.

PA has advanced and exceeded expectations mainly in the use of field-operated machines. However, there are still operational problems due to the lack of digital system compatibility, both for communication between equipment and for exchanging files and data, rendering it still an operational challenge.

An important part of the PA application cycle is the stage of obtaining, storing, and analyzing relevant attribute data and maps that influence the spatial and temporal variability of crop productivity.

Yield maps express natural and anthropogenic soil variations. Various types of terrestrial, orbital, and suborbital sensors have been used to assist in the acquisition of soil and plant monitoring data, allowing efficient and low cost data generation, while providing reliable crop estimates to improve the production potential.

All these advances presented in this chapter lead to current agriculture innovations for detailed spatial management of the agricultural production system, in order to maximize economic returns and reduce environmental impacts.

The methodologies and results presented in this work were developed with the collaboration of several Embrapa Research Units, the management and technical teams of several partners such as the Gatto, 3D Engenharia, Amaggi, Scheffer, Sementes Petrovina, Sugarcane Plant Santa Cruz, Santa Cruz coffee farms and Morro Alto farm groups. In addition, we emphasize the technical field support of the Instituto Matogrossense do Algodão (IMAmt) and of the agronomist Guy Carvalho.

References

AL-GAADI, K. A.; HASSABALLA, A. A.; TOLA, E.; KAYAD, A. G.; MADUGUNDU; BANDER ALBLEWI, R.; ASSIRI, F. Prediction of potato crop yield using precision agriculture techniques. **PloS One**, v. 11, n. 9, 2016. DOI: 10.1371/journal.pone.0162219.

BERNARDI, A. C. de C.; NAIME, J. de M.; RESENDE, A. V.; BASSOI, L. H.; INAMASU, R. Y. (ed.). **Agricultura de precisão**: resultados de um novo olhar. Brasília, DF: Embrapa, 2014. 596 p.

CANDIAGO, S.; REMONDINO, F.; DE GIGLIO, M.; DUBBINI, M.; GATTELLI, M. Evaluating multispectral images and vegetation indices for precision farming applications from UAV images. **Remote sensing**, v. 7, n. 4, p. 4026-4047, 2015. DOI: 10.3390/rs70404026.

CORWIN, D. L.; PLANT, R. E. Applications of apparent soil electrical conductivity in precision agriculture. **Computers and Electronics in Agriculture**, v. 46, n. 1-3, p. 1-10, 2005. DOI: 10.1016/j.compag.2004.10.004.

D'OLIVEIRA, M. V. N.; BROADBENT, E. N.; OLIVEIRA, L. C.; ALMEIDA, D. R. A.; PAPA, D. A.; FERREIRA, M. E.; ZAMBRANO, A. M. A.; SILVA, C. A.; AVINO, F. S.; PRATA, G. A.; MELLO, R. A.; FIGUEIREDO, E. O.; JORGE, L. A. C.; JUNIOR, L.; ALBUQUERQUE, R. W.; BRANCALION, P. H. S.; WILKINSON, B.; OLIVEIRA-DA-COSTA, M. Aboveground biomass estimation in Amazonian tropical forests: a comparison of Aircraft- and GatorEye UAV-borne LiDAR Data in the Chico Mendes Extractive Reserve in Acre, Brazil. **Remote Sensing**, v. 12, n. 1, May 2020. DOI: 10.3390/rs12111754.

DOERGE, T. Defining management zones for precision farming. **Crop Insights**, v. 8, n. 21, p. 1-5, 1999.

FALKER. **SoloDrill amostrador de solo hidráulico**. Available at: https://falker.com.br. Accessed on: 3 Sept. 2020.

GREGO, C. R.; SPERANZA, E. A.; RODRIGUES, C. A. G.; NOGUEIRA, S. F.; BAYMA, G.; CIFERRI, R. R.; LUCHIARI JR., A. Definição de zonas de manejo em cana-de-açúcar usando séries temporais de NDVI derivados do satélite Sentinel-2A. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 19., 2019, Santos. **Anais**... São José dos Campos: Inpe, 2019. 3 p.

GREGO, C. R.; VIEIRA, S. R. Variabilidade espacial de propriedades físicas do solo em uma parcela experimental. **Revista Brasileira de Ciência do Solo**, v. 29, n. 2, p. 169-177, maio/abr. 2005. DOI: 10.1590/S0100-06832005000200002.

HATFIELD, J. L.; GITELSON, A. A.; SCHEPERS, J. S.; WALTHALL, C. L. Application of spectral remote sensing for agronomic decisions. **Agronomy Journal**, v. 100, n. S3, S117-S131, 2008. DOI: 10.2134/agronj2006.0370c.

INTERNATIONAL SOCIETY OF PRECISION AGRICULTURE. Available at: https://ispag.org. Accessed: 11 June 2020.

JORGE, L. A. de C.; INAMASU, R. Y. Uso de veículos aéreos não tripulados (VANT) em agricultura de precisão. In: BERNARDI, A. C. de C.; NAIME, J. de M.; RESENDE, A. V. de; BASSOI, L. H.; INAMASU, R. Y. (ed.). **Agricultura de precisão**: resultados de um novo olhar. Brasília, DF: Embrapa, 2014. p. 109-134.

KITCHEN, N. R.; SUDDUTH, K. A.; MYER, D. B.; DRUMMOND, S. T.; HONG, S. Y. Delineating productivity zones on claypan soil fields using apparent soil electrical conductivity. **Computers and Electronics in Agriculture**, v. 46, n. 1-3, p. 285-308, 2005. DOI: 10.1016/j.compag.2004.11.012.

LEE, W. S.; ALCHANATIS, V.; YANG, C.; HIRAFUJI, M.; MOSHOU, M.; LI, C. Sensing technologies for precision specialty crop production. **Computers and Electronics in Agriculture**, v. 74, n. 1, p. 2-33, 2010. DOI: 10.1016/j.compag.2010.08.005.

LEROUX, C.; TISSEYRE, B. How to measure and report within-field variability: a review of common indicators and their sensitivity. **Precision Agriculture**, v. 20, p. 562-590, 2019. DOI: 10.1007/s11119-018-9598-x.

LI, Y.; SHI, Z.; LI, F.; LI, H.Y. Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. **Computers and Electronics in Agriculture**, v. 56, n. 2, p. 174-186, 2007. DOI: 10.1016/j.compag.2007.01.013.

LUCHIARI JR., A.; SHANAHAN, J.; FRANCIS, D.; SCHLEMMER, M.; SCHEPERS, J.; LIEBIG, M.; SCHEPERS, A; PAYTON, S. Strategies for establishing management zones for site specific nutrient management. In: INTERNATIONAL CONFERENCE ON PRECISION AGRICULTURE, 5., 2000, Bloomington. **Proceedings**... Bloomington: American Society of Agronomy, 2000. p. 1-10.

MAJUMDAR, J.; NARASEEYAPPA, S.; ANKALAKI, S. Analysis of agriculture data using data mining techniques: application of big data. **Journal of Big Data**, v. 4, n. 1, article number 20, 2017. DOI: 10.1186/s40537-017-0077-4.

MOGILI, U. R.; DEEPAK, B. B. V. L. Review on application of drone systems in precision agriculture. **Procedia Computer Science**, v. 133, p. 502-509, 2018. DOI: 10.1016/j.procs.2018.07.063.

MOLIN, J. P.; AMARAL, L. R. do; COLAÇO, A. **Agricultura de precisão**. São Paulo: Oficina de Textos, 2015. 224 p.

MOLIN, J. P.; TAVARES, T. R. Sensor systems for mapping soil fertility attributes: challenges, advances, and perspectives in Brazilian tropical soils. **Engenharia Agrícola**, v. 39, p. 126-147, set. 2019. Número especial. DOI: 10.1590/1809-4430-eng.agric.v39nep126-147/2019.

NAGELKERKE, N. J. D. A note on a general definition of the coefficient of determination. **Biometrika**, v. 78, n. 3, p. 691-692, Sept. 1991. DOI: 10.1093/biomet/78.3.691.

NO-TILL FARMER. **Sensor EM38**. Available at: https://www.no-tillfarmer. Accessed on: 3 Sept. 2020.

OSCO, L. P.; ARRUDA, M. dos S. de; MARCATO JUNIOR, J.; SILVA, N. B. da; RAMOS, A. P. M.; MORYIA, É. A. S.; IMAI, N. N.; PEREIRA, D. R; CRESTE, J. E.; MATSUBARA, E. T.; LI, J.; GONÇALVES, W. N. A convolutional neural network approach for counting and geolocating citrus-trees in UAV multispectral imagery. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 160, p. 97-106, Feb. 2020a. DOI: 10.1016/j.isprsjprs.2019.12.010.

OSCO, L. P.; RAMOS, A. P. M.; PINHEIRO, M. M. F.; MORYIA, É. A. S.; IMAI, N. N.; ESTRABIS, N.; IANCZYK, F.; ARAUJO, F. F. de; LIESENBERG, V.; JORGE, L. A. de C.; LI, J.; MA, L.; GONÇALVES, W. N.; MARCATO JUNIOR, J.; CRESTE, J. E. A machine learning framework to predict nutrient content in Valencia-orange leaf hyperspectral measurements. **Remote Sensing**, v. 12, n. 6, p. 906, 2020b. DOI: 10.3390/rs12060906.

RABELLO, L. M.; BERNARDI, A. C. de C.; INAMASU, R. Y. Condutividade elétrica aparente do solo. In: BERNARDI, A. C. de C.; NAIME, J. de M.; RESENDE, A. V. de; BASSOI, L. H.; INAMASU, R. Y. (ed.). **Agricultura de precisão**: resultados de um novo olhar. Brasília, DF: Embrapa, 2014. p. 48-57.

REUTEBUCH, S. E.; ANDERSEN, H.-E.; MCGAUGHEY, R. J. Light detection and ranging (LIDAR): an emerging tool for multiple resource inventory. **Journal of Forestry**, v. 103, n. 6, p. 286-292, 2005.

RODRIGUES, C. A. G.; GREGO, C. R.; SANCHES, G. M.; SPERANZA, E. A.; RONQUIM, C. C.; SILVEIRA, H. L. F.; LUCHIARI JUNIOR, A. Índices de vegetação (Sentinel-2) da cana-de-açúcar relacionados com a condutividade elétrica aparente de solo. In: SIMPÓSIO NACIONAL DE INSTRUMENTAÇÃO AGROPECUÁRIA, 4., 2019, São Carlos, SP. **Ciência, inovação e mercado**: anais. São Carlos, SP: Embrapa Instrumentação, 2019a. p. 151-155.

RODRIGUES, G. C.; GREGO, C. R.; LUCHIARI, A.; SPERANZA, E. A. Caracterização espacial de índices de vegetação índice relativo de clorofila em áreas de produção cafés especiais no sul de Minas Gerais. In: SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 10., 2019, Vitória. **Pesquisa, inovação e sustentabilidade dos cafés do Brasil**. Brasília, DF: Embrapa Café, 2019b. 6 p.

RONQUIM, C. C. **Conceitos de fertilidade do solo e manejo adequado para as regiões tropicai**s. Campinas: Embrapa Monitoramento por Satélite, 2010. 26 p. (Embrapa Monitoramento por Satélite. Boletim de pesquisa e desenvolvimento, 8).

SACI SOLUÇÕES. **SACI TRAIL**: amostrador de solo da Saci Soluções. Available at: https://www.sacisolucoes.com.br. Accessed: 3 Sept. 2020.

SANCHES, G. M.; PAULA, M. T. N. de; MAGALHÃES, P. S. G.; DUFT, D. G.; VITTI, A. C.; KOLLN, O. T.;BORGES, B. M. M. N.; FRANCO, H. C. J. Precision production environments for sugarcane fields. **Scientia Agricola**, v. 76, n. 1, p. 10-17, Jan./Feb. 2019. DOI: 10.1590/1678-992x-2017-0128.

SANTOS, T. T.; BASSOI, L. H.; OLDONI, H.; MARTINS, R. L. Automatic grape bunch detection in vineyards based on affordable 3D phenotyping using a consumer webcam. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 11., 2017, Campinas. **Ciência de dados na era da agricultura digital**: anais. Campinas: Ed. Unicamp: Embrapa Informática Agropecuária, 2017. v. 11, p. 89-98.

SCUDIERO, E.; TEATINI, P.; CORWIN, D. L.; DEIANA, R.; BERTI, A.; MORARI, F. Delineation of site-specific management units in a saline region at the Venice Lagoon margin, Italy, using soil reflectance and apparent electrical conductivity. **Computers and Electronics in Agriculture**, v. 99, p. 54-64, Nov 2013. DOI: 10.1016/j.compag.2013.08.023.

SUN, S.; LI, C.; PATERSON, A. H.; JIANG, Y.; XU, R.; ROBERTSON, J. S.; SNIDER, J. L.; CHEE, P.W. In-field high throughput phenotyping and cotton plant growth analysis using LiDAR. **Frontiers in Plant Science**, v. 9, p. 16, Jan. 2018. DOI: 10.3389/fpls.2018.00016.

SHIRATSUCHI, L. S.; BRANDÃO, Z. N.; VICTORIA, D. de C.; DUCATI, J. R.; OLIVEIRA, R. P. de; VILELA, M. de F. Sensoriamento remoto: conceitos básicos e aplicações na agricultura de precisão. In: BERNARDI, A. C. de C.; NAIME, J. de M.; RESENDE, A. V.; BASSOI, L. H.; INAMASU, R. Y. (ed.). **Agricultura de precisão**: resultados de um novo olhar. Brasília, DF: Embrapa, 2014. p. 58-73.

SMTIS, F. M. Measurement of sheet resistivities with the four-point probe. **The Bell System Technical Journal**, v. 37, n. 3, p. 711-718, May 1958. DOI: 10.1002/j.1538-7305.1958.tb03883.x.

SPERANZA, E. A.; GREGO, C. R.; JORGE, L. A. C.; RODRIGUES, C. A. G.; LUCHIARI JUNIOR, A.; RONQUIM, C. C.; SANCHES, G. M. Delineamento de zonas de manejo em cana-de-açúcar a partir de atributos do solo e da cultura e imagens georreferenciadas. In: SIMPÓSIONACIONAL DE INSTRUMENTAÇÃO AGROPECUÁRIA, 4., 2019, São Carlos, SP. **Ciência, inovação e mercado**: anais. São Carlos, SP: Embrapa Instrumentação, 2019a. p. 175-179.

SPERANZA, E. A.; GREGO, C. R.; RODRIGUES, G. C.; LUCHIARI, A. Influência das diferentes faces de exposição ao sol nos índices vegetativos e relativo de clorofila em cafés especiais. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA ,12., 2019, Indaiatuba. **Anais** [...]. Indaiatuba: Fatec; Campinas: Embrapa Informática Agropecuária; Ponta Grossa: Ed. UEPG, 2019b.

SUDDUTH, K. A.; DRUMMOND, S. T. Yield editor: software for removing errors from crop yield maps. **Agronomy Journal**, v. 99, n. 6, p. 1471-1482, 2007. DOI: 10.2134/agronj2006.0326.

TANG, J. L.; WANG, D.; ZHANG, Z. G.; HE, L. J.; XU, Y. Weed identification based on K-means feature learning combined with convolutional neural network. **Computers and Electronics in Agriculture**, v. 135, p. 63-70, 2017. DOI: 10.1016/j.compag.2017.01.001.

VEGA, A.; CÓRDOBA, M.; CASTRO-FRANCO, M.; BALZARINI, M. Protocol for automating error removal from yield maps. **Precision Agriculture**, v. 20, n. 5, p. 1030-1044, 2019. DOI: 10.1007/s11119-018-09632-8.

VERIS. **Sensor Veris Techonologies**. Available at: http://www.veristech.com. Accessed on: 3 Sept. 2020.

VIEIRA, S. R. Uso de geoestatística em estudos de variabilidade espacial de propriedades do solo. In: NOVAIS, R. F.; ALVAREZ V., V. H.; SCHAEFER, C. E. G. R. (org.). **Tópicos em ciência do solo**. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 2000. v. 1, p. 1-54.

ZHANG, T-X.; SU, J.-Y.; LIU, C.-J.; CHEN, W. H. Potential bands of Sentinel-2A satellite for classification problems in precision agriculture. I**nternational Journal of Automation and Computing**, v. 16, n. 1, p. 16-26, 2019. DOI: 10.1007/s11633-018-1143-x.

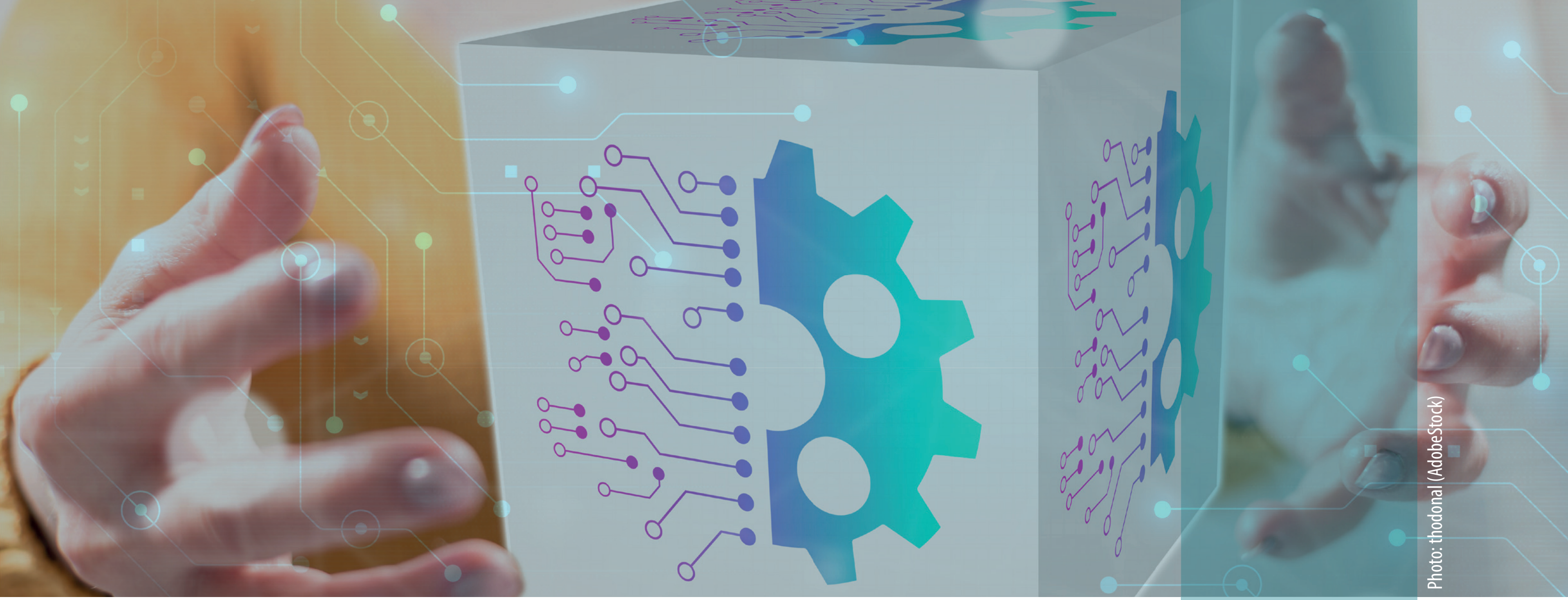


Photo: thodonal (AdobeStock)

Information engineering
Contributions to digital agriculture

8

Ivo Pierozzi Júnior | Marcos Cezar Visoli | Marcia Izabel Fugisawa Souza | Luiz Manoel Silva Cunha | Isaque Vacari | Tércia Zavaglia Torres

Introduction

Digital agriculture comprises the modeling of agricultural phenomena and processes in the environmental, economic and social dimensions through computational artifacts and Information and Communication Technologies (ICT) to bring to the agricultural sector organization, access, use, sharing, and dissemination facilities, as well as the application of scientific knowledge.

> There are multiple challenges in a globalized world, and it is complicated to even reach a consensus on what the priorities should be. However, one of them stands out as the most important due to its structuring effects – making knowledge accessible to everyone. At no other time in history has the production of knowledge been as intense as it is today, and at no other time has its application assumed such a preeminent role. Hence the importance of knowledge management, because between its production and its use there is a chain of complex procedures that may or may not determine its operational success. For some experts like Manuel Castells, the application of knowledge is at the center of the conceptual and operational revolution driven by science and technology advances which are operating in contemporary societies, and that reach all sectors of human life at unprecedented speed. It is therefore important to think about the use of knowledge, pave the way for its various uses and ensure its social and ethical dimension. (Defourny, 2006, p. 7).

The term Information Engineering, as presented by Martin and Finkelstein (1989), is expressed in three different definitions, but converges conceptually. In two of them, the word “automated” stands out:

1) The application of an interconnected set of formal techniques for planning, analyzing, designing and building information systems about an organization as a whole or in one of its main sectors.

2) An interconnected set of automated techniques in which organization models, data models and process models are built into a comprehensive knowledge base used to create and maintain data processing systems.

3) A set of automated disciplines at the organization level used to provide the right information, to the right people, at the right time.

It is based on this perspective and this context that Information Engineering is presented in this chapter, according to how this subject is understood, as a means of mapping, organizing and representing agricultural knowledge and in the context of digital agriculture.

From the perspective of Knowledge Management (KM), the delivery itinerary of scientific knowledge and the guarantee of its effectiveness and efficiency in response to society's demands, including those affecting agriculture (mainly food, energy and fibers) are made possible by Information Engineering. Embrapa has already addressed the relationship between data, information and knowledge (Pierozzi et al., 2017) to enable the use of theoretical and conceptual concepts that align these three levels of organization of human perception about the real world and its consequent technological transformation.

The application of knowledge, that is, to apprehend and use it as a solution to problems and challenges, goes through the decision-making process. There is no best decision to make. There is a possible decision regarding the ability to identify, gather, process and combine the greatest possible amount of information on a given subject. Thus, as a research, development and innovation discipline, Information Engineering is positioned at the central point to combine data, originated from the practice of agricultural research, knowledge, which represents the offer and use of Embrapa's performance results, modeled on ICT. In addition to the aforementioned challenges, the dynamic and massive pace of knowledge production and supply, accelerated by the advancement and support of ICTs.

Therefore, another challenge emerges, simultaneously, in which the quality of the knowledge offered is also configured as a social demand: the knowledge that is sought begins to be demanded as environmentally sustainable, economically viable and socially fair. It is no different in the context of the pragmatics of scientific knowledge and, in particular, in the context of agricultural knowledge.

Embrapa has expressed, in its constant strategic planning efforts, in which it periodically reviews its mission, vision, objectives and goals (Embrapa, 2018), a constant concern with the delivery of technological knowledge to society, using premises of quality and effectiveness. Its recent incursion into the implementation of innovation-oriented research management models corroborates the convergence and coherence of this intention, especially as it is a company that generates knowledge and competences and, therefore, a company that learns and evolves scientifically, technologically, and organizationally (Garcia; Salles Filho, 2009).

In light of this reality, Embrapa Digital Agriculture inserts its contribution, given the efforts it has invested to develop and innovate methodologies and technologies and produce knowledge within its competences in Computing and ICT.

It is in this context that the paths of digital agriculture are aligned and explored, a concept and term that in the scope of Research, Development & Innovation (RD&I) and Science and Technology (S&T) is not only as a trend but, in particular, a new socioeconomic paradigm of the agricultural sector, since it uses ICT to move research data and transform it into information, knowledge and technology for the producer, through the Internet of Things (IoT), Big Data, Cloud Computing, Machine Learning, etc. This paradigm, inherent to digital agriculture, is combined with other contemporary paradigms, such as the Information Economy or Knowledge Economy, Data Science and Open Science (Porat; Rubin, 1977; Powell; Snellman, 2004; Pordes et al., 2007; Aalst, 2016).

The word "engineering" has been associated with computing (software, data, knowledge engineering, etc.), as a way to express the processes of "construction" of computational artifacts that represent things (entities, phenomena, processes) from the real world to machine language. A possible explanation for this linguistic phenomenon of interdisciplinary conceptual and terminological recombinations is the understanding that the practical use of knowledge is an endless, continuous and dynamic process of conceptual recombination, analysis, synthesis and re-signification in permanently emerging contexts. Hence the metaphorical meaning of the word "engineering".

At the same time, in the same itinerary of knowledge construction and development, another conceptual reflection has associated the words "data", "information" and "knowledge" (D-I-K), creating several representation models of this relationship and thus giving new meaning to the term "engineering". Recently, in the proposition of a Knowledge Data and Information Governance model at Embrapa, a conception of a model related to this relationship, different from the conventional ones, was presented to facilitate its organizational and operational implementation in order to provide support to corporate processes of data, information and knowledge management (Pierozzi Junior et al., 2017). The model has as theoretical and conceptual references, in addition to the notion of the D-I-K relationship, the life cycles of data, information and knowledge, conceived in a conjugated and aligned way and represented as a mandala.

This model also supports an ontological approach (Mol, 2008), which serves as an itinerary for the construction of the ontic, that is, the entity itself. The term "entity" is used as a reference to the computational objects or artifacts to be engineered (software, applications, information systems, etc.), since these are the objects that operationally implement scientific and multidisciplinary knowledge, enabling its application in the solution of problems or in response to demands from the agricultural sector.

Based on this general conception, Information Engineering, as an area of knowledge, discipline or proposal for a productive process of technologies and innovation, was configured as an attractive conceptual and terminological option to bring together competences, technologies and solutions executed and produced by Embrapa Digital Agriculture throughout its history and, mainly, as an appropriate option to systematize the process of transforming scientific data into pragmatic knowledge.

Therefore, a conceptual metamodel is being elaborated to organize worldviews in the environmental, agricultural, social and economic spheres (Figure 1), for the development of computational products in response to challenges and opportunities in digital agriculture. Another metamodel (Figure 2) brings together conceptual, methodological and technological approaches that are aligned, based on the concept of Information Engineering, as an integrative construct of knowledge pragmatics that are inherent to various sciences such as Cognition, Information and Computing.

In the following sections, research actions and results will be presented, discussed and contextualized, which in the context of Information Engineering, are being developed at Embrapa Digital Agriculture.

Knowledge organization and representation systems

To make knowledge accessible and usable, whether by human agents or technological agents, it has to be organized (Soergel, 2009). Thus, including technological perception, Information Engineering can be understood as an area or discipline of knowledge that allows the construction of an operational itinerary, supported by computing and ICT, so that knowledge becomes accessible and usable.

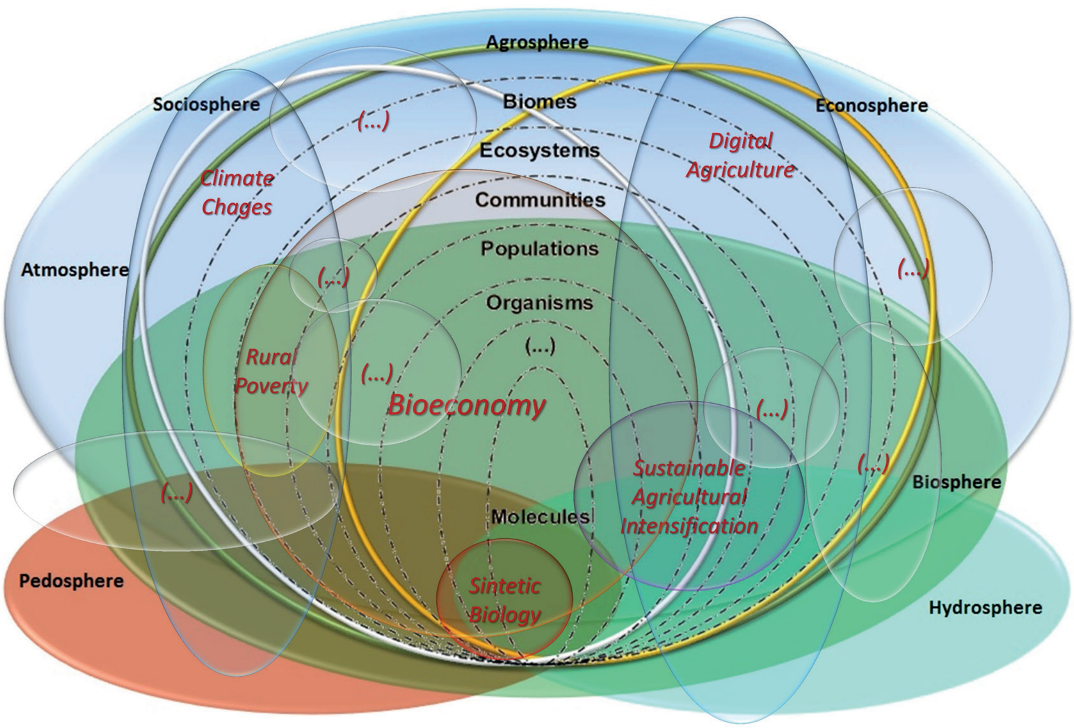


Figure 1. Multi, inter and transdisciplinary conceptual representation of agriculture.

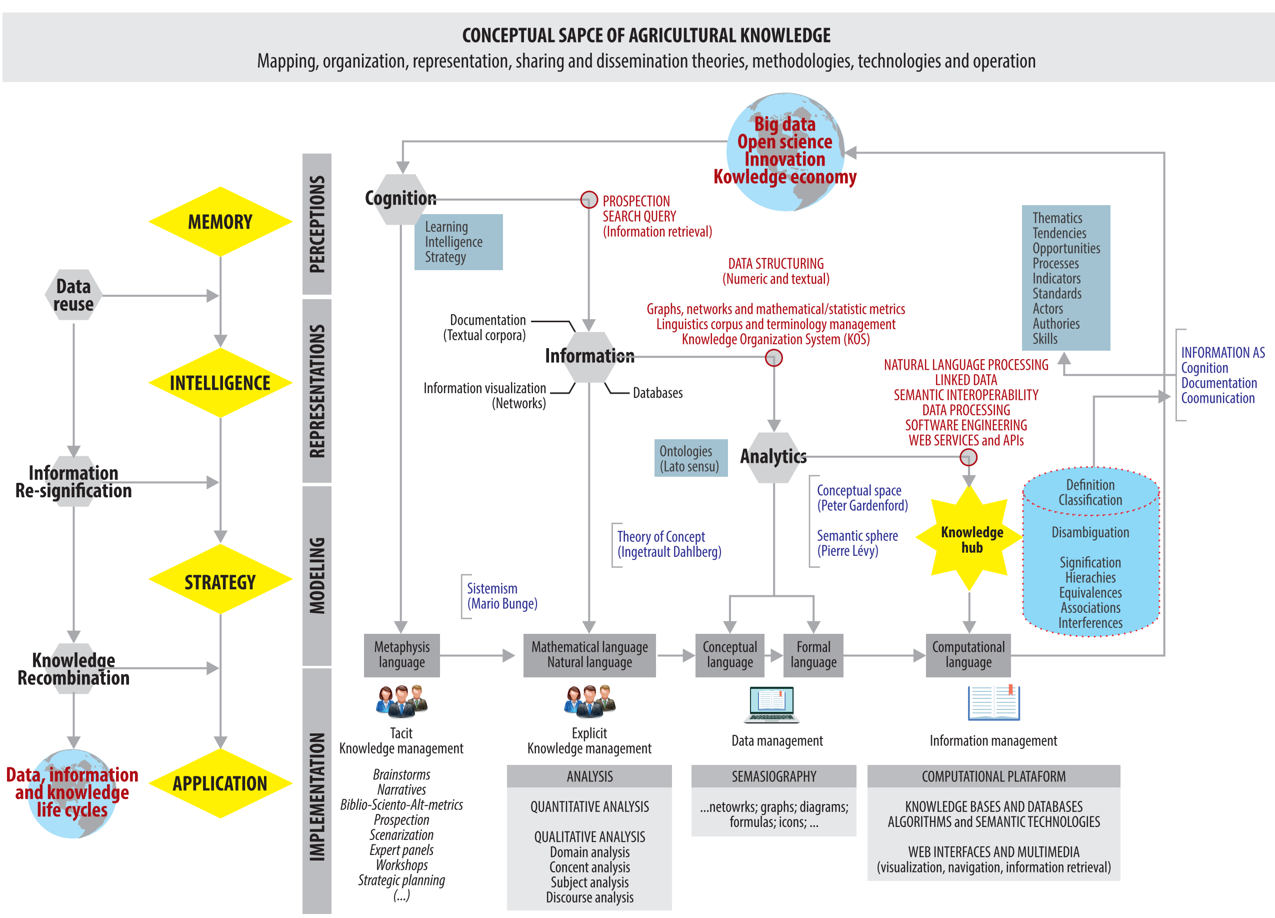


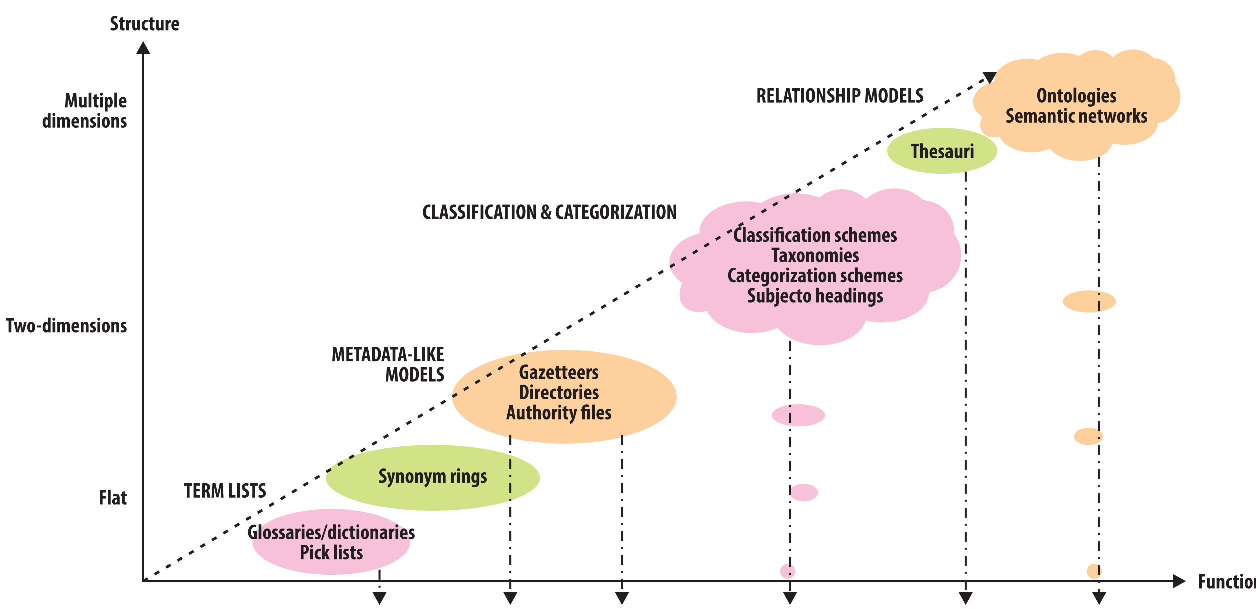
Figure 2. Agricultural Information Engineering Metamodel.

The Knowledge Organization Encyclopedia defines Knowledge Organization Systems (KOS) as:

> [...] a generic term used to refer to a wide range of items (e.g., subject titles, thesauruses, classification schemes, and ontologies) that were conceived with regard to different purposes at different historical times. They are characterized by different structures and specific functions, different ways of relating to technology, and used in a plurality of contexts by various communities. However, what they all have in common is that they are designed to support the organization of knowledge and information to facilitate management and retrieval (Mazzocchi, 2019, p. 1).

They can also be defined as "[...] semantically structured conceptual systems that include terms, definitions, relationships and properties of the concepts" (Carlan; Medeiros, 2011, p. 54, own translation). The term Knowledge Organization System (KOS), was proposed by the Networked Knowledge Organization Systems Working Group, at the 1st Conference of the ACM Digital Libraries, in 1998, in Pittsburgh, Pennsylvania (Carlan; Medeiros, 2011 p. 54).

Figures 3 and 4 illustrate how KOS can be understood and indicate how they can be learned and used in the context of Information Engineering.



Eliminating ambiguity	xxx		xxx	xx	xxxx	xx
Controlling synonyms		xxxx	xxx	xx	xxxx	xx
Establishing relationships: hierachical			x	xxxx	xxx	xxx
Establishing relationships: associative					xxxx	xxxxx
Presenting properties						xxxxx

Figure 3. Types of Knowledge Organization System.

Source: Zeng (2008).

The process of organizing and representing knowledge has always been part of human history (Martins; Moraes, 2015). This process is configured as a multi and interdisciplinary action inherent to all sciences and cultures. Thus, when Aristotle[1] defined the ten categories of 'being' as metaphysical categories, which classify words in relation to our knowledge of 'being', and today in the plural modernity in which

[1] 1 The ten categories of knowledge are used in the classification and representation of human thought, that is, thoughts became the starting point for representations.

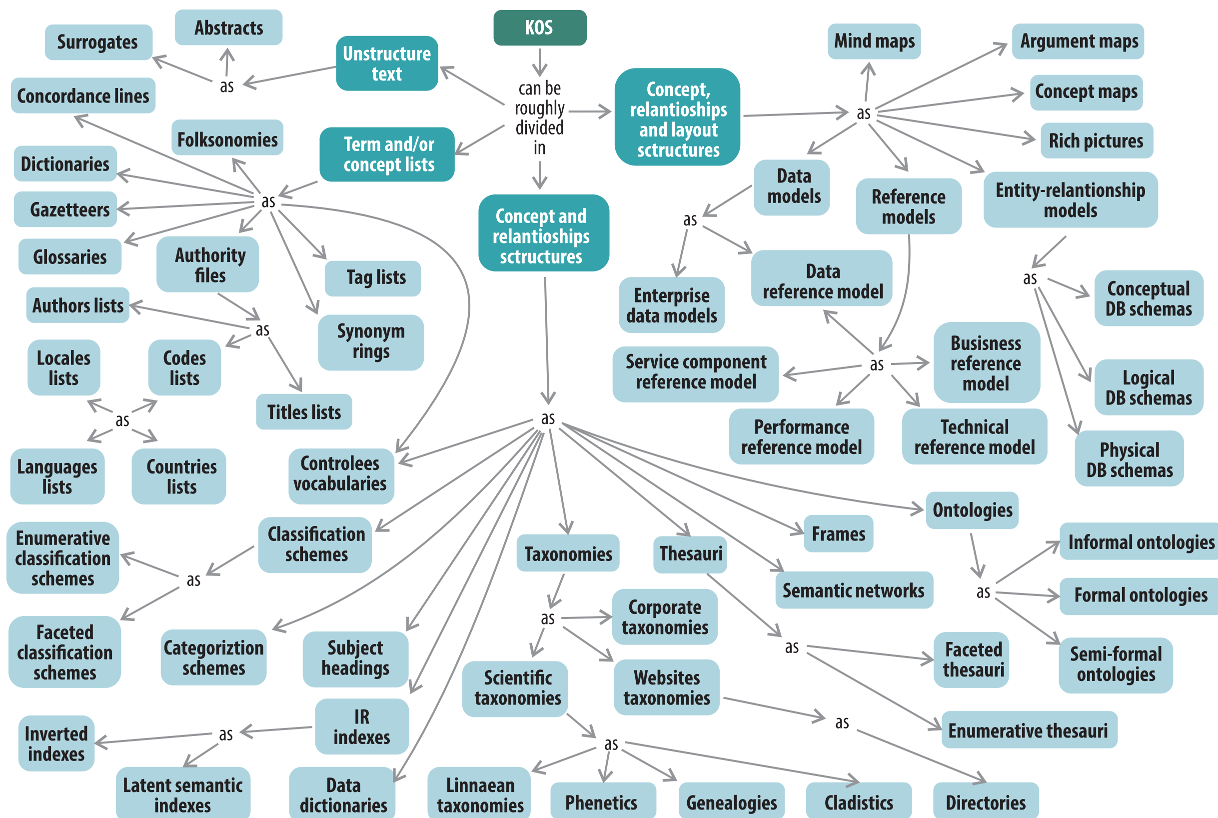


Figure 4. Classification of Knowledge Organization System.

Source: Souza et al. (2012).

we are in, we need methods of organization and representation of knowledge to understand the world around us (Mendonça, 2005). It so happens that to represent knowledge, before taking any action, it has to be classified, in other words, we need to organize it in a system so that we can understand it (Martins; Moraes, 2015).

The process of organizing objects is a classificatory and relational process, which requires our cognitive ability to associate ideas, create order and meaning in our experiences, using the interpretation of the world, the attribution of meanings and the structuring of ideas. The representation of knowledge is, therefore, part of a totality that, through perception and reason, looks to formulate abstract concepts about the reality to which it belongs (Martins; Moraes, 2015).

It can be argued that the world as an object of human knowledge exists as an interpreted world that is completely infused with meaning. Human cognition cannot see simple facts without these being part of its structure of meaning (Tuomi, 1999). Hence the importance of knowledge organization and representation systems, especially when they fall within the scope of the developing Information and Communication Technologies (ICT) solutions for digital agriculture.

In the context of the Information Engineering Research Group of Embrapa Digital Agriculture, these systems are understood as "performed" objects, that is, constructed and executed in line with their contexts, as objects that insert multiple realities, responding directly to the peculiarities of modeling complex systems. In the context of Brazilian agricultural knowledge, the KOS are presented as organizational and representative systems of knowledge in this domain, describing associations between the multiple interdependent elements that act in the production of knowledge. (Latour, 2012).

They must be perceived as spaces that integrate the social, cultural, environmental and economic contexts of Brazilian agriculture in interdependent relations that, in addition to action, consider the process of knowledge production within the scope of an articulated and hybrid reality, with complex capillarity involving human and non-human agents.

In this sense, the system of representation and organization of knowledge at Embrapa is an open system, which is configured more as a constitutive map of a network of actors that influence themselves by being in permanent interaction – redesigning new routes – than as a closed, limited and static territory.

As open systems that have agents (human and non-human) that articulate and transform each other, the KOS at Embrapa have been constructed and employed in different dimensions of knowledge domains, resulting in fragmented and dispersed ontologies in time and in space, whose representation potential is impaired due to the absence of higher-level perceptions. Thus, the use of these conceptual artifacts as they have been conceived and used, to support computational modeling of agricultural knowledge, still faces difficulties of coherence and convergence, as KOS should represent multiple ontologies resulting from various methods and practices used by researchers, who move the knowledge of Brazilian agriculture at Embrapa (Baum et al., 2020). Therefore, KOS conceived from the perspective of ontological policies (Mol, 2008) are understood as performative objects with greater adherence to the paradigm of complexity related to Brazilian agriculture.

Thus, agriculture as an object of knowledge is not a passive subject waiting to be perceived from the point of view of an endless series of perspectives. On the contrary, it is a living and open object constituted through scientific practices, through which it is manipulated to be better understood (Baum et al., 2020).

This rationality justified the need to engineer a system of knowledge organization and representation for Embrapa that had a pluralist ontological conception, which implies that it is an oriented system and "[...] favorable to the coexistence of a variety of explanations, assumptions, methods, methodologies, approaches, theories" (Baum et al., 2020, p. 14), which together perform agriculture as an open object, exposed to its multiple relationships, interactions, capillaries and, consequently, to its own entirety.

Agroterms: controlled vocabulary at Embrapa

Knowledge is a personal intellectual experience and, thus, the term "transfer of knowledge" may have a conceptual meaning, but in practice it has no operational meaning. Such transfer, in fact, happens through a process of encoding brain energy in natural language, and is manifested via communication between a "sending" agent and a "receiving" agent. Humanity has performed this process so naturally that at times it is not even considered that there are other possible means of encoding knowledge, such as symbols, sounds, smells, textures, etc. The truth is that, fundamentally, almost all human knowledge has been encoded in spoken or written natural language, and more recently, digitally, which still retains its original nature. This human naturalness of representation is indeed based on the preponderance of visual perception over other senses.

Thus, a good part of KOS is conceived, built and executed in this manner: based on the lexicon, which technologically, can be modeled through Natural Language Processing (NLP) methods and tools, and therefore, graphically encoded.

Controlled vocabularies are KOS that collect and organize words or terms, in the field of scientific specialties. Based on Concept Theory, terms denote concepts, but they are only one of the vertices of the triangle that represents a given concept. Another vertex is the referent, that is, what the human mind perceives in the real world. Finally, the third vertex refers to the properties that can be attributed to the referent and that, finally, guide the choice of a lexical element in natural language that best synthesizes the perception of the referent (Dahlberg, 1978). According to the same author, concepts are considered units of knowledge. Thus, the role of controlled vocabularies is contextualized as facilitating resources in the management of institutions, which are confronted with the production, access and sharing of D-I-K, including volume scales and Big Data flow. And in the same logic, currently controlled vocabularies greatly benefit from Information Engineering to be conceived, managed and maintained as open and dynamic systems, in accordance with the premises of best practices of knowledge representation and their pragmatic applications.

Until recently, Embrapa used externally controlled vocabularies in its corporate processes for managing D-I-K, conceived and managed in contexts not perfectly aligned with its own range of contents related to tropical agriculture, which greatly hindered the alignment of this collection of knowledge at various stages of the D-I-K lifecycles, and even more so when deployed on global scales. At the level of cataloging, indexing, retrieving, accessing and disseminating D-I-K processes, problems of consistency, ambiguity and lack of interoperability between information systems accumulated in the same proportion this collection grew exponentially.

The Agroterms was then built, through a combination of Portuguese language terminologies found in national and international agricultural thesauruses, with a methodological and technological basis in the Global Agricultural Concept Space (GACS) initiative (Research Data Alliance, 2020). As a result, Embrapa was recognized as content curator in Portuguese, for the Brazilian variant, by the editor group of Agrovoc (FAO, 2020) – the controlled vocabulary of the Food and Agriculture Organization (FAO) of the United Nations (UN).

Currently, Agroterms is composed of approximately 245,000 terms. Through Information Engineering, using NLP methodologies and tools, Corpus Linguistics and semantic modeling, it is being prepared to expand its technological functionality as a terminological resource to a level of conceptual space for Brazilian agricultural knowledge. In order to achieve this potential, Agroterms is being addressed organizationally by a permanent working group from Embrapa, the Gtermos, responsible for its conception, curation and management, within the context of Data Governance and Information for Knowledge Policy, already implemented at Embrapa and which contributes to the intention and efforts in order to bring digital agriculture to the reality of the Brazilian agricultural sector.

Research data management

The first decades of the 21st century have been characterized by an explosive growth in human capacity to acquire, store and communicate digital data. From a scientific perspective, the concept of "data-intensive science" or "e-Science" (Borgman, 2007; Gray, 2009), has been consolidated as a reality in numerous fields of knowledge, many of them relevant to agricultural research.

Technological advances have allowed for greater accuracy and coverage in data acquisition. Some examples are Internet of Things (IoT) applications, which are making the use of sensors a reality in the field; the growing possibilities of imaging rural areas using drones (also known as Unmanned Aerial Vehicles - UAVs); other geotechnology applications; and the evolution of bioinformatics and nanotechnology areas.

The current situation has also been called the "Big Data" Era, characterized by the 5 Vs: Volume, Velocity, Variety, Veracity and Value (Mcafee; Brynjolfsson, 2012). For this large amount of data to be useful, it must be well managed, retrievable, and accessible, understandable, and integrated. This scenario has led to transformations in how data, information and knowledge are created and used: to intelligently and quickly deal with the data "flood", new skills are required to ensure the preservation, integration and reuse of the data.

Research data management (RDM) is a discipline that brings together a set of activities that are essential to the planning, implementation and execution of strategies, procedures and practices aimed at effective data management. There are several approaches to understanding RDM; one of them refers to different conceptions of data life cycle management (DLM) models, which provide a view of the dynamics of data, from its generation to its reuse.

Data lifecycle is defined by DataONE (2020b) as a high-level representation of the stages involved in management and preservation of data for use and reuse (Figure 5). In this chapter, this definition of DataONE is taken as a reference, due to the suitability for adaptation and the versatility of interpretation of its definitions and concepts. This life cycle is composed of eight stages: planning, collecting, ensuring quality, describing, preserving, discovering, integrating and analyzing, which are briefly described, based on Sayão and Sales (2015), Strasser et al. (2015), Araújo et al. (2019) and DataONE (2020b). These steps involve cyclical actions that enable: a) to construct a data management plan aimed at meeting the data policy of the research institution; b) data collection for guaranteeing its usability and long-term reuse; c) ensured guarantee to data sets so they can be used and are reproducible; d) precise and detailed description of the data, adopting a standard of metadata, taxonomies and controlled vocabularies; e) data preservation through proper storage in data centers in order to ensure interoperability, recovery and search; f) discovery of potentially useful data, which, as described by metadata, can be easily found; g) integration of data from different sources, which after combined, generate new sets of data that can be used; h) data analysis provided by the new datasets created in the integration, in order to provide relevant information for future research.

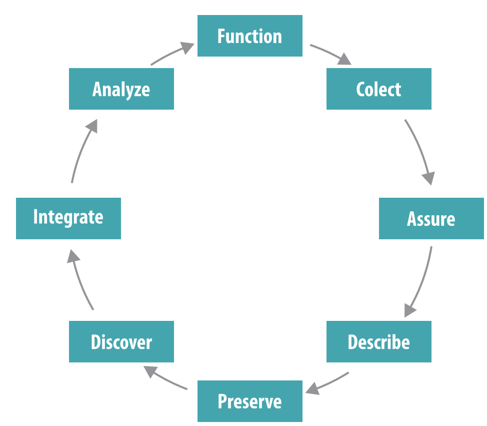


Figure 5. Data Lifecycle.

Source: DataOne (2020b).

The RDM based on the data life cycle refers to adopting best practices, which are defined as "[...] methods or approaches that are recognized by a community as being correct or more appropriate for acquisition, management, analysis and data sharing" (Sayão; Sales, 2015, p. 81, our translation). Such practices guide people on how to effectively work with their data at each stage of its life cycle, thus helping to map the processes involved in DML (DataONE, 2020a). Examples of best practices in the planning stage, as those recommended by DataONE it recommends: creation, management and documentation of data; definition of the types and format of data to be produced, etc. For the step to describe, for example, it recommends: to name the files so they describe and reflect their content; describe format for geospatial and temporal location of the data; adopt standard taxonomies for describing any datasets; adopt specialized controlled vocabulary; define set of metadata elements, etc.

However, only the recommendation of best practices does not guarantee efficiency and effectiveness in data management and, according to Veiga (2019, p. 15, own translation), "it is not enough to share data, they need to be FAIR". It is therefore necessary to associate such best practices with the FAIR principles (Findable, Accessible, Interoperable, Reusable), so that research data, in addition to being well managed, can also become findable, accessible, interoperable and reusable (Wilkinson et al., 2016). The four FAIR principles are made up of 15 elements that contribute to complement and enrich the information essential to the eight stages of data lifecycle, expanding the possibilities of location, access, interoperability and reuse.

The FAIR principles apply to any research objects, so that they become available and understandable to humans and machines, ensuring transparency, reproducibility and reuse, in addition to providing the proper and adequate citation of information generated by data-intensive science (Wilkinson et al., 2016). The FAIR principles also guided the design of Embrapa's Data, Information and Knowledge Governance Policy (Embrapa, 2019) to bring researchers and other research subjects closer to the datasets available in data repositories and platforms.

Metadata and data cataloging

The objective of cataloging and metadata in this subject is largely data. Data is a word with several meanings, including generic and specialized ones, depending on the context in which it is used. According to Semeler and Pinto (2019, p. 113, own translation), "[...] data means a single piece of information, "while" research data are the result of any systematic investigation that involves research processes of observation, experimentation or simulation of scientific research procedures."

Metadata and data cataloging are necessary so that research data can be "[...] identifiable, citable, visible, retrievable, interpretable, contextualizable, interoperable and reusable when considering consistency and origin" (Semeler; Pinto, 2019, p. 116). Also noteworthy is the need to consider the context of research data management and the data life cycle in which metadata and data cataloging are inserted, which, in addition to being two important sub-steps of the description stage, should be aligned with the FAIR principles.

Given the need to expand data and information representation mechanisms to better manage them, and the consequent complexity involved in defining their attributes, metadata can no longer be defined as only "data about data". Currently, this definition is considered an expression that does not help to understand what exactly metadata means (Sayão; Sales, 2015). What expands this understanding and expands its application domain is the definition given by Riley (2017, p. 1), who considers metadata as the information we create, store and share to describe things, and which allows us to interact with these things to obtain the necessary knowledge.

Metadata allow exploring other dimensions and facets of the data, which, when revealed by cataloging, contribute to improving management and quality, favoring the discovery of data collections for the scientific community. Such dimensions bring to light the need to create new metadata elements that are capable of expanding and enriching the adopted metadata scheme. Metadata are essential so that in the future digital content can be accessed and interpreted. Without metadata, according to Gray (2009 cited by Sayão and Sales 2016, slide 83), the users

> [...] will not know the details of how the data was obtained and prepared: 1) how the instruments were designed and built; 2) when, where and how data were collected; and 3) will not have a description of the processes that led to the derived data, which are typically used for scientific analyses.

Metadata are also essential for technical and semantic interoperability, so that without them, data repositories and platforms will not be able to exchange data and information. Metadata consists of well-defined descriptive elements, for example: author, title, description, subject, keyword, identifier, producer, types of data, access conditions, terms of use of collections, etc., and based on the data cataloging, formulating a body of information capable of contextualizing the data in terms of provenance, history, nature, purpose and other aspects.

The implementation of enriched metadata brings direct benefits to data management, positively impacting archiving and preservation, as well as interoperability and retrieval of research datasets. Data will only be useful for analysis if it has been described by quality metadata, and for that to happen, the best recommendation is to use FAIR principles when cataloging it.

Also with regard to metadata, and according to Veiga (2019, p. 18-22), it is necessary to briefly highlight the key elements that should guide the implementation of FAIR principles, especially regarding the descriptive aspect of metadata: a) metadata elements for single and persistent identifiers for both the data and the dataset; b) dataset using metadata enriched with a wide range of precise and relevant attributes; c) metadata element that clearly and explicitly indicates persistent identifiers, both from the dataset and from the metadata itself in the data repositories and platforms; d) metadata registered or indexed in identification resources that offer search capability; e) metadata using standardized communication protocols to facilitate data retrieval via metadata, including; f) availability of access to metadata, even if the data is not accessible and available; g) metadata element for the representation of knowledge through formal language and the use of taxonomies and controlled vocabularies according to FAIR principles, specialized and standardized by specific area of the domain; h) metadata element for qualified dataset references and other derived research objects, which interconnect, ensuring semantic interconnections between them, and which are linkable to other datasets; i) metadata with a wealth of attributes and high level of detail to allow researcher to evaluate the possibility of reuse and relevance to their needs; j) metadata element with unambiguous information, clearly defining who can have access to the data, for what purpose and under what conditions; k) metadata element that specifies the origin of the data, supporting the researcher when deciding on the usefulness of the data or metadata and when attributing credit to the data producer; m) implementation of metadata must be aligned with relevant and specific standards of the community and the research area.

In the context of digital agriculture, metadata described in accordance with FAIR principles will directly contribute to the discovery and reuse of data by other researchers and research institutions.

Dataverse platform

The report entitled *Open access to research data in Brazil: technological solutions – 2018 report* (Rocha, 2018) presents the results of the Brazilian Research Data Network (RDP Brazil) research project, which identifies, explores and analyzes in depth three technological solutions (Dataverse, DSpace and CKAN) for building an Open Access to Research Data repository.

Based on the Open Access and Scholarly Information System (OASIS) model – composed of 56 criteria, classified in Repository Environment Representation, Data Set Representation, Data Set Description and Documentation, Data Set Production, Long Storage Deadline and Planning for Preservation, Access and Use of Data Sets and Use, Development and Maintenance of the Software –, it is concluded that the Dataverse and DSpace technologies have resources for configuring various types of data repository, including organizational, thematic and hierarchies distinct data policies for research groups or units, with metadata schemas and support for usage licenses. However, CKAN software is a good alternative when used as a publishing and access service, with submission and digital preservation performed by other repository environments.

The Dataverse Platform is an open-source software application for storing, publishing and sharing data (Dataverse, 2020b). It provides facilities to represent scenarios composed of several hierarchical entities, such as universities, institutions, laboratories, research groups and departments, so as to autonomously implement the details of data management, such as defining who can create, authorize the publication or access data sets, establish licenses, as well as define that the data can only be used upon request.

The platform uses metadata schemas (compatible with DDI Lite, DDI Codebook, Dublin Core, DataCite, VORResource, ISA-Tab), manages dataset versions, uniquely identifies datasets (considering versions) in a universal and persistent way (DOI or Handle System), provides citation metadata and a citation structure that involves checking the immutability of the cited material. It also enables the storage of complementary documents together with a dataset, adds visualization and data exploration tools, allows the customization of its interface, and provides the collection of metadata from the Open Archives Initiative Protocol for Metadata protocol Harvesting (OAI-PMH).

Additional functionality can be incorporated, such as support for storing large volumes of data; support for data visualization and exploration; improvement in the indexing and search engine and in user authentication systems.

Metadata on the Dataverse platform: the experience of Embrapa Digital Agriculture

As mentioned before, the Dataverse Platform is a web application dedicated to the sharing, preservation, citation, exploration and analysis of research data, which hosts several dataverses composed of datasets, which are processed through metadata (Dataverse, 2020b). The Dataverse Platform was chosen by Embrapa Digital Agriculture to support the management of research datasets, with the Embrapa Bioinformatic Multi-user Laboratory (LMB) as a pilot project (Embrapa Digital Agriculture, 2020). Therefore, it is in this environment of the Dataverse Platform dedicated to the LMB that the experience reported for metadata takes place. (Dataverse, 2020a).

As in several platforms and data repositories, the datasets in the Dataverse Platform are inserted using a registration form, which contains numerous fields, corresponding to metadata elements. Some of these fields are mandatory, but most are optional. In addition, there are additional metadata sets that can be added for specific data domains. Metadata sets follow established standards, ensuring interoperability with other platforms.

Based on the FAIR principles, studies and tests were conducted that involved users and producers of the LMB pilot project data, searching to define metadata elements (terms) to meet the specificities of the pilot project users. Based on this work, definitions were obtained regarding: a) basic metadata, some mandatory, as well as complementary, non-mandatory metadata; b) tools for metadata description: description norms, taxonomy, thesaurus and controlled vocabularies. These definitions are important for generating enriched metadata, which are those that use domain-specific description tools, such as taxonomy, thesaurus and controlled vocabulary, associated with provenance information, bringing semantic clarity when compared to basic (original) metadata. According to Lira (2014, p. 43):

> A set of enriched metadata must present features such as: (i) greater number of semantic attributes...; (ii) ease of interpretation and processing of dataset content; (iii) associated standard vocabulary terms [...].

Metadata on the Dataverse Platform, based on experience with the creation of dataverses and datasets for managing research data in the LMB, are mirrored in the FAIR principles, especially to make them rich in highlights and with numerous precise and relevant attributes.

Data cataloging on the Dataverse platform

The process of cataloging the datasets of the LMB pilot project begins with the self-deposit activity, which is carried out by the researcher, the dataset owner or another designated person. The time of self-deposit of the dataverse or dataset on the Dataverse Platform corresponds to the pre-cataloging phase, filling in the fields corresponding to the basic metadata elements and mandatory information: title, author, contact, description and subject.

The cataloging takes place in the next phase, which begins by reviewing the previous filling in of the contents related to the basic metadata elements. The catalog description of the dataverse and dataset essentially fills in the complementary metadata elements, which will be done by the domain specialist, preferably by the dataset owner. However, this activity requires knowing librarianship norms and standards, giving greater meaning and quality to datasets.

The use of cataloging techniques to describe and represent data or any research objects, based on a structured set of FAIR metadata elements, is essential to ensure technical and semantic interoperability, sharing, use and reuse research data, with the technical responsibility of the librarian and/or the information scientist.

Data quality

The high standard of quality in data archiving and maintenance is widely recognized by different segments of society. In digital agriculture, data quality is particularly important for a high level of assertiveness in the decision-making process, planning activities, and others. Based on this finding, and on the impact on different types of business, the data quality (DQ) theme is seen as a strong pillar in the data management process. This has increasingly called the attention of researchers from different areas of knowledge to investigate and expand studies on the subject.

As shown in the literature, there are different definitions for DQ. These variations are related to the context that discusses DQ and the degree of demand the user addresses quality.

It is important to highlight the definition of Data Management Body Knowledge (DMBOK) for DQ:

> "[...] the planning, implementation and control of activities that apply data quality management techniques, in order to ensure they are suitable for the consumer and meet the needs of data consumers" (Knight, 2017, p. 1).

For an organization, whether public or private, focused on business or research, the archiving, maintenance and recovery of high quality data bring opportunities to formulate better business strategies, decision-making facilities for high level of success, and for achieving better competitive advantage conditions. The lack of this quality level, in addition to making it difficult to achieve the aforementioned advantages, contributes to added data processing costs. In addition, the customer's level of satisfaction decreases when he/she receives the result of a service or a required information as a response to a research action. (Jaya et al., 2017).

The DQ theme should not be treated independently, as poor-quality data leads to misleading and costly conclusions, which can lead to the data team's distrust or loss of credibility. Problems found in the organizational culture of an institution also affect the data collection and quality process.

DQ can be developed under qualitative and quantitative approaches (Vancauwenbergh, 2019); in the qualitative approach, categories (for example, measurement, contextualization, representation and access) are determined, and dimensions/characteristics are associated with each one of them. In turn, in the quantitative approach, DQ is guided by the adequacy of data to serve a purpose in a given context, that is, in decision-making and/or planning operations.

Measures for evaluating data quality

Data quality assessment is a crucial process within data quality management, as it comprises different stages that involve several groups of people in an organization. The purpose of this type of assessment is to identify data containing some type of error and measure the impact of various data-driven business processes.

The process can be started in the data collection phase, including minimum guidelines for quality assurance, helping to reduce the amount of work performed in the data qualification/preparation phase (Pyle, 1999) and thus providing better conditions for carrying out data analysis.

DQ can be evaluated using subjective measures and/or determined by computational calculations. In practice, when a data quality assessment process is initiated, basic measures are used, such as the number of missing data, amount of data with typographical error, amount of non-standard data, basic statistical measures and visual analyses. Adopting these measures provides a preliminary notion of how good the level of data quality is. In some situations, the use of a single measure is not enough to achieve the desired result, hence using other measures together.

In addition to the measures mentioned for evaluating data quality, others are described by Cichy and Rass (2019). They are related to frameworks available for establishing this type of evaluation. These measures are used when there is an interest in expanding the assessment level. The higher the quality level, the greater the accuracy of the results generated, which brings greater possibilities for more assertive decision-making. Another important contribution of having quality data is contributing to the discovery of knowledge in a database, knowledge that is inserted in the data and not visualized, at first, by the user (Fayyad et al., 1996).

Data quality management

DQ is one of the crucial problems to correctly measure and analyze science, technology and innovation, which allows the adequate monitoring of research efficiency, productivity and also strategic decision-making. (Fan; Geerts, 2017).

Typically, data has some kind of inconsistency – some are duplicated, incomplete, inaccurate and obsolete. To enable them to produce quality results, after being processed and analyzed, the Data Quality Management (DQM) process is used.

The main objective of DQM is to remove any and all problems found, raising the quality of the data and enabling them to contribute to the addition of value to the business processes and/or produce qualified answers to the questions addressed (Vancauwenbergh, 2019).

DQM involves performing various tasks, defining parameters and assigning values to them, and establishing a workflow. All this must be easily recorded, updated and retrieved. To support all this work, different frameworks are available, with emphasis on: DAMA DMBOK´s Data Governance Model (Barbieri, 2013); EWSolutions' EIM Maturity Model (Smith, 2009); and Oracle's Data Quality Management Process (Oracle, 2009).

These models are centered on three basic elements, which are the metadata associated with the data, the processes for recording, organizing and (re)using data, and the organizational context in relation to the data. The quality of each element and the interaction between them, will ultimately determine the quality and therefore the true value of an organization's data assets.

Permission to describe metadata that is understandable by the entire organization and aligned with the processes, strategies and business objectives of that organization is a resource available in these models. In addition, these models provide the means to report critical success factors, which are useful elements for developing effective DQ management strategies.

Critical success factors for implementing Data quality management

According to Milosevic and Patanakul (2005, p. 183), critical success factors (CFS) are "[...] characteristics, conditions or variables that can have a significant impact on the success of an organization or a project when appropriately sustained, maintained or managed." Santos (2015) presents 20 CFS applicable to the DQM which, according to Milosevic and Patanakul (2005), form four large groups: a) operational; b) management; c) governance; and d) qualification. The operating group focuses on the operational processes involved in data collection, storage, analysis and security, all of which are highly interdependent. The management group brings together the management processes, which originate from the operational group, mainly aligning data quality with the organization's goals in relation to data and the results of data analysis. The third group, governance, involves the governance processes associated with DQM. These processes can be presented by the organization's senior management as a priority commitment for the implementation of DQM, stimulating a culture change throughout the organization focused on this subject. Finally, the qualification group is deemed essential to invest in a DQM program, even if the company has employee training structure for operational, management and governance actions. The main objective of this group is to inform people about the importance of qualitative data for the organization. In addition to training for the systematic implementation of DQ, throughout the organization, it must institute a continuous monitoring action of the qualifications. This will allow quick adjustments, if errors and adjustments in the business rules are identified.

Final considerations

The chapter presented an account of the research actions and the results that were carried out and being developed at Embrapa Digital Agriculture, in the context of Information Engineering. Thus, the objective is to align this work with the actions of digital agriculture and also to add value to the pragmatics of scientific knowledge, offering technologies more easily perceived and assimilated by its potential users.

These actions, in progress in the Information Engineering Research Group, take place in the domain of three natures of information (cognitive, documentary and communicative) and through computational artifacts that operationalize the processes that constitute the data life cycles of information and knowledge (D-I-K). Computer Science is the main source generating these actions, able to combine and complement methodological and technological contributions originating in other fields of knowledge, producing inter, multi and transdisciplinary developments with Agronomy, Ecology, Mathematics, Economics, Sociology and the full range of imaginable intersections. Inserted and articulated in this universe of interactions between different areas of knowledge, Information Engineering is an alternative for the operationalization of strategies, aiming at greater alignment between the actions developed in the areas of Research and Development (R&D) and Embrapa's innovation process.

Embrapa Digital Agriculture, by inaugurating the Information Engineering research line, reorganizes, guides and rescues its competences towards a repositioning of its RD&I actions, considering the perspective of the innovation process implemented in the company. In particular, with regard to facing the current research challenges to consolidate the digital transformation in agriculture, Information Engineering is capable of effectively contributing to conceptual, methodological, procedural contributions and, especially, to support the development of quality artifacts, objects and computational tools, according to an engineering process designed within a pluralistic ontological conception. In other words, Information Engineering expands the possibilities of the various actors that circumscribe the phenomenon of "Brazilian agriculture" to perceive it as an object that admits multiple explanations, assumptions, methods, methodologies, approaches, theories, etc. Furthermore, the efforts and initiatives in Information Engineering, thus far, are aligned with the trends and contemporary opportunities for development and computational applications and ICT in digital agriculture.

Based on the heuristics made possible by Information Engineering, the computational artifacts of D-I-K representation can be used beyond their immediate functionalities (Pierozzi Junior et al., 2018). However, the contributions of Information Engineering can be translated and materialized under different perspectives and exemplified in the form of repositories and databases that enable collaborative work; in the cataloging, indexing and intelligent retrieval of information; use, reuse and data management; in the redefinition of information and interoperability with other systems; in discovering knowledge; in facilitated and controlled access, communication, sharing, learning and collective intelligence. Since digital agriculture is fundamentally based on digital content, from data obtained through the Internet of Things, it is envisaged that Information Engineering will promote facilities and improvements in the construction of computational artifacts that meet the interests of users in different segments of Brazilian agriculture.

Another reading is possible: a) when the data is addressed based on the perspectives of classification, meaning and access, it is said that what is being worked on are its cognitive properties; b) when information is worked on based on the perspectives of cataloging, indexing and retrieval, it is said that what is being worked on are its documentary properties; c) when knowledge is worked on based on visualization perspectives or machine languages, it is said that what is being worked on are its communication properties (dissemination). In addition to these properties, those that are worked on and inherited from the preceding levels of data and information, respectively, must be considered. Thus, the result is a continuous, cyclic feedback movement, which occurs when the communicated knowledge returns as insight for a new round of data and information cycles.

References

AALST, W. van der. **Processing mining**: data science in action. Berlin: Springer-Verlag, 2016. 467 p. DOI: 10.1007/978-3-662-49851-4.

ARAÚJO, D. G. de; ALMEIDA LLARENA, M. A.; SIEBRA, S. de A.; DIAS, G. A. Contribuições para a gestão de dados científicos: análise comparativa entre modelos de ciclo de vida dos dados. **Liinc em Revista**, v. 15, n. 2, p. 32-51, nov. 2019. DOI: 10.18617/liinc.v15i2.4686.

BARBIERI, C. **Uma visão sintética e comentada do Data Management Body of Knowledge (DMBOK)**. Belo Horizonte: Fumsoft, 2013. 46 p.

BAUM, C.; MARASCHIN, C.; MARKUART, E. N. Política ontológica como abordagem para as relações intercientíficas. **Psicología, Conocimiento y Sociedad**, v. 9, n. 2, p. 8-30, nov. 2019; abr. 2020. Available at: http://www.scielo.edu.uy/pdf/pcs/v9n2/1688-7026-pcs-9-02-6.pdf. Accessed on: May 2020.

BORGMAN, C. L. **Scholarship in the digital age**: information, infrastructure and internet. Cambridge, MA: MIT Press, 2007. DOI: 10.7551/mitpress/7434.001.0001.

CARLAN, E.; MEDEIROS, M. B. B. Sistemas de organização do conhecimento na visão da Ciência da Informação. **Revista Ibero-Americana de Ciência da Informação**, v. 4, n. 2, p. 53-73, ago./dez. 2011. DOI: 10.26512/rici.v4.n2.2011.1675.

CICHY, C.; RASS, S. An overview of data quality frameworks. **IEEE Access**, v. 7, p. 24634-24648, Feb. 2019. DOI: 10.1109/ACCESS.2019.2899751.

DAHLBERG, I. Teoria do conceito. **Ciência da Informação**, v. 7, n. 2, p. 101-107, dez. 1978. Available at: http://revista.ibict.br/ciinf/article/view/115/115. Accessed on: 19 May 2020.

DATAONE. **Best practices**. Albuquerque, NM: University of New Mexico, 2020a. Available at: https://www.dataone.org/best-practices. Accessed on: 27 May 2020.

DATAONE. **Data life cycle**. Albuquerque, NM: University of New Mexico, 2020b. Available at: https://www.dataone.org/data-life-cycle. Accessed on: 2 May 2020.

DATAVERSE. **GenClima**. Campinas: Embrapa Informática Agropecuária, 2020a. Available at: https://www.dataverse-h.cnptia.embrapa.br/dataverse/umip. Accessed on: 2 May 2020.

DATAVERSE. **Harvard Dataverse**. Cambridge, MA: Harvard College, 2020b. Available at: https://dataverse.harvard.edu. Accessed on: 2 May 2020.

DEFOURNY, V. Apresentação. In: TARAPANOFF, K. (org.). **Inteligência, informação e conhecimento em corporações**. Brasília, DF: Ibict: Unesco, 2006. p. 7.

EMBRAPA DIGITAL AGRICULTURE. **Laboratório Multiusuário de Bioinformática**. Campinas, 2020. Available at: https://www.embrapa.br/en/agricultura-digital/lmb. Accessed on: 2 June 2020.

EMBRAPA. Política de Governança de Dados, Informação e Conhecimento da Embrapa. **Boletim de Comunicações Administrativas**, v. 45, n. 16, p. 1-19, abr. 2019. 19 p. (Manual de normas da Embrapa).

EMBRAPA. **Visão 2030**: o futuro da agricultura brasileira. Brasília, DF, 2018. 212 p. Available at: https://www.embrapa.br/visao/o-futuro-da-agricultura-brasileira. Accessed on: 15 May 2020.

FAN, W.; GEERTS, F. Foundations of data quality management. **Synthesis Lectures on Data Management**, v. 4, n. 5. p. 1-227, July 2017. DOI: 10.2200/S00439ED1V01Y201207DTM030.

FAO. **AGROVOC**. Rome: FAO-AIMS, 2020. Available at: http://aims.fao.org/ vest-registry/vocabularies/agrovoc. Accessed on: 20 May 2020.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI Magazine**, v. 17, n. 3, p. 37-54, Fall 1996.

GARCIA, A. E. B.; SALLES FILHO, S. L. M. Trajetória institucional de um instituto público de pesquisa: o caso do Ital após 1995. **Revista de Administração Pública**, v. 43, n. 3, p. 661-693, maio/jun. 2009. DOI: 10.1590/S0034-76122009000300007.

GRAY, J. Jim Gray on eScience: a transformed scientific method. In: HEY, T.; TANSLEY, S.; TOLLE, K. (ed.). **The fourth paradigm**: data-intensive scientific discovery. Redmond, WA: Microsoft Research, 2009. p. xvii-xxxi.

JAYA, I.; SIDI, F.; ISHAK, I.; AFFENDEY, L. S.; JABAR, M. A. A review of data quality research in achieving high data quality within organization. **Journal of Theoretical and Applied Information Technology**, v. 95, n. 12, p. 2647-2657, 2017.

KNIGHT, M. What is data quality? In: DATAVERSITY. **Dataversity.net**. Studio City, CA; Dataversity Education, 2017. Available at: https://www.dataversity.net/what-is-data-quality. Accessed on: 19 May 2020.

LATOUR, B. Reagregando o social: uma introdução à Teoria do Ator-Rede. Salvador: Edufba; Bauru: Edusc, 2012. 399 p. Resenha de: SEGATA J. **Ilha Revista de Antropologia**, v. 14, n. 2, p. 238-243, jul./dez. 2012. DOI: 10.5007/2175-8034.2012v14n1-2p238.

LIRA, M. A. B. de. **Uma abordagem para enriquecimento semântico de metadados para publicação de dados abertos**. 2014. 95 f. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Pernambuco, Centro de Informática, Recife.

MARTIN, J.; FINKELSTEIN, C. **Information engineering**. Englewood Cliffs: Prentice Hall, 1989.

MARTINS, G. K.; MORAES, J. B. E. Organização e representação do conhecimento: institucionalização como disciplina científica no âmbito da Ciência da Informação. In: ENCONTRO NACIONAL DE PESQUISA EM PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO - ENANCIB, 16., 2015, João Pessoa. **Anais**... João Pessoa: Ed. UFPb, 2015. Available at: http://www.ufpb.br/evento/index. php/enancib2015/enancib2015/paper/viewFile/3162/1030. Accessed on: 16 May 2020.

MAZZOCCHI, F. Knowledge organization system (KOS). In: HJORLAND, B.; GNOLI, C. (ed.). **Encyclopedia of Knowledge Organization**. Alberta: University of Alberta, 2019. Available at: http://www.isko.org/cyclo/kos. Accessed on: 16 May 2020.

MCAFEE, A.; BRYNJOLFSSON, E. Big data: the management revolution. **Harvard Business Review**, v. 90, n. 10, p. 61-68, Oct. 2012.

MENDONÇA, E. S. A organização e a representação do conhecimento no tempo. **Revista de Ciências Humanas**, n. 38, p. 277-94, out. 2005.

MILOSEVIC, D.; PATANAKUL, P. Standardized project management may increase development projects success. **International Journal of Project Management**, v. 23, n. 3, p. 181-192, Apr. 2005. DOI: 10.1016/j.ijproman.2004.11.002.

MOL, A. Política ontológica: algumas ideias e várias perguntas. In: NUNES, A.; ROQUE, R. (ed.). **Objectos impuros**: experiências em estudos sobre a ciência. Porto: Ed. Afrontamento, 2008. p. 63-106.

ORACLE. **Oracle® Warehouse Builder**: user's guide - 11g release 1(11.1). Redwood City, CA, 2009. 764 p. Available at: https://docs.oracle.com/cd/B31080_01/doc/owb.102/b28223.pdf. Accessed on: 16 May 2020.

PIEROZZI JUNIOR, I.; BERTIN, P. R. B.; MACHADO, C. R. de L.; SILVA, A. R. da. Towards semantic knowledge maps applications: modelling the ontological nature of data and information governance in a R&D organization. In: THOMAS, C. (ed.). **Ontology in information science**. Rijeka: InTech, 2017. p. 83-104. DOI: 10.5772/67978.

PORAT, M. U.; RUBIN, M. R. **The information economy**. Washington, DC: US Govt. Print Off, 1977.

PORDES, R.; PETRAVICK, D.; KRAMER, B.; OLSON, D.; LIVNY, M.; ROY, A.; AVERY, P.; BLACKBURN, K.; WENAUS, T. The open science grid. **Journal of Physics**: conference series, v. 78, p. 1-15, 2007. DOI: 10.1088/1742-6596/78/1/012057.

POWELL, W. W.; SNELLMAN, K. The knowledge economy. **Annual Review of Sociology**, v. 30, p. 199-220, Aug. 2004. DOI: 10.1146/annurev.KOS.29.010202.100037.

PYLE, D. **Data preparation for data mining**. San Francisco, CA: Morgan Kaufmann, 1999. 560 p.

RESEARCH DATA ALLIANCE. **Global Agricultural Concept Space (GACS)**. [S.l.]: European Commission; National Science Foundation, 2020. Available at: https://agrisemantics.org/#GACShome. Accessed on: 20 May 2020.

RILEY, J. **Understanding metadata**: what is metadata, and what is it for? Bethesda, MD: NISO Press, 2017. 49 p. Available at: https://groups.niso.org/apps/group_public/ download.php/17446/Understanding%20Metadata.pdf. Accessed on: 1 May 2020.

ROCHA, R. P. da. (coord.). **Acesso aberto a dados de pesquisa no Brasil:** soluções tecnológicas - relatório 2018. Porto Alegre: UFRGS, 2018. 75 p. Available at: http://hdl.handle.net/10183/185126. Accessed on: 14 May 2020.

SANTOS, M. P. da C. dos. **Fatores críticos de sucesso na gestão da qualidade dos dados**. 2015. 55 f. Dissertação (Mestrado em Gestão de Sistemas de Informação) – Lisbon School of Economics & Management, Lisboa, Portugal.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN, 2015. 93 p.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa**: [minicurso]. Rio de Janeiro: CNEN, 2016. 196 slides.

SEMELER, A. R.; PINTO, A. L. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados. **Ciência da Informação**, v. 48, n. 1, p. 113-129, jan./abr. 2019.

SMITH, A. Enterprise information management maturity: data governance's role. **EIMInsight Magazine**, v. 3, n. 1, jan. 2009. Available at: http://www.eiminstitute.org/library/eimi-archives/ volume-3-issue-1-january-2009-edition/EIM-Maturity. Accessed on: 16 May 2020.

SOERGEL, D. Digital libraries and knowledge organization. In: KRUK, S. R.; MCDANIEL, B. (ed.). **Semantic digital librarie**s. Berlin: Springer, 2009. p. 9-39. DOI: 10.1007/978-3-540-85434-0_2.

SOUZA, R. R.; TUDHOPE, D.; ALMEIDA, M. B. Towards a taxonomy of KOS: dimensions for classifying knowledge organization systems. **Knowledge Organization,** v. 39, n. 3, p. 179-192, 2012. DOI: 10.5771/0943-7444-2012-3-179.

STRASSER, C.; COOK, R.; MICHENER, W.; BUDDEN, A. **Primer on data management**: what you always wanted to know. [S.l.]: California Digital Library, 2015. 12 p. (DataONE best practices primer).

TUOMI, I. Data is more than knowledge: implications of the reversed knowledge hierarchy for knowledge management and organizational memory. **Journal of Management Information Systems**, v. 16, n. 3, p. 103-117, 1999. DOI: 10.1080/07421222.1999.11518258.

VANCAUWENBERGH, S. Data quality management. In: KUNOSIC, S.; ZEREM, E. (ed.). **Scientometrics recent advances**. London: Intechopen Limited, 2019. p. 1-15. DOI: 10.5772/intechopen.86819.

VEIGA, V. Gestão de dados de pesquisa FAIR: dando um JUMP em seus dados. In: ENCONTRO DA REDE SUDESTE DE REPOSITÓRIOS INSTITUCIONAIS, 1., 2019, Rio de Janeiro. **Anais** ... Rio de Janeiro: Fiocruz/Icit/ UFRJ, 2019. 59 p. Available at: https://www.arca.fiocruz.br/handle/icict/33343. Accessed on: 27 Apr. 2020.

WILKINSON, M. D.; DUMONTIER, M.; AALBERSBERG, J. J.; APPLETON, G.; AXTON, M.; BAAK, A.; BLOMBERG, N.; BOITEN, J.-W.; SANTOS, L. B. da S.; BOURNE, P. E.; BOUWMAN, J.; BROOKES, A. J.; CLARK, T.; CROSAS, M.; DILLO, I.; DUMON, O.; EDMUNDS, S.; EVELO, C. T.; FINKERS, R.; GONZALEZ-BELTRAN, A.; GRAY, A. J. G.; GROTH, P.; GOBLE, C.; GRETHE, J. S.; HERINGA, J.; HOEN, P. A. C T.;

HOOFT, R.; KUHN, T.; KOK, R.; KOK, J.; LUSHERM, S. J.; MARTONE, M. E.; MONS, A.; PACKER, A. L.; PERSSON, B.; ROCCA-SERRA, P.; ROOS, M.; SCHAIK, R. van; SANSONE, S.-A.; SCHULTES, E.; SENGSTAG, T.; SLATER, T.; STRAWN, G.; SWERTZ, M. A.; THOMPSON, M.; LEI, J. van der; MULLIGEN, E. van; VELTEROP, J.; WAAGMEESTER, A.; WITTENBURG, P.; WOLSTENCROFT, K.; ZHAO, J.; MONS, B. The FAIR guiding principles for scientific data management and stewardship. **Scientific Data**, v. 3, article number 160018, 2016. DOI: 10.1038/sdata.2016.18.

ZENG, M. L. Knowledge Organization Systems (KOS). **Knowledge Organization**, v. 35, n. 2-3, p. 160-182, 2008. DOI: 10.5771/0943-7444-2008-2-3-160.

Photo: your123 (AdobeStock)

9

DIPN
A dictionary of the internal proteins nanoenvironments and their potential for transformation into agricultural assets

Ivan Mazoni | Goran Neshich

Introduction

Proteins play a vital role in supporting life. They are macromolecules resulting from the combination, through peptide bonds, of these 20 amino acids: alanine, arginine, aspartate, asparagine, cysteine, phenylalanine, glycine, glutamate, glutamine, histidine, isoleucine, leucine, lysine, methionine, proline, serine, tyrosine, threonine, tryptophan, and valine. Considering a linear combination between these 20 amino acids, the number of possible variations is 20^n, in which n is the amount of amino acid residues in the protein (as the amino acids lose some atoms when forming the peptide bond, it is common to call them amino acid residues, since they are part of a polypeptide chain). For example, for a protein with 100 amino acid residues, the number of possible combinations will equal $20^{100} = 1.27 \times 10^{130}$. In comparison, the estimated total number of atoms in the Universe is 9×10^{78} (Villanueva, 2009). Each organism, animal or vegetable, has thousands of different proteins. Among their various functions: structural, transport, protection, defense, control and regulation of expression, catalysis, movement, and storage stand out as some examples. For a better understanding of the relationship between the aminoacid sequence in a protein, its three dimensional structure, and its function, came the proposition for analyses of the proteic nanoenvironment. It is also known as a proteic district or functional region, and it is where biologicaly functional elements are located.

The hypothesis that motivated the work of the Embrapa Digital Agriculture Computational Biology Research Group (CBRG) in Campinas (SP), during the 2010s, was an approach that assumed the existence of a "sign", or that is, a variation in the values of the physico-chemical and structural descriptors that distinguish a specific site (or a protein substructure). This is where a certain element of secondary structure (or an active site, an interface, etc.) is inserted in the framework of a whole protein. Understanding how the subordinate structural elements to the biologically functional structure are formed and later maintained will open the way for us to understand how proteins assume their final structure and, consequently, their function. In our work we use STING_RDB, a unique database in the world, produced and maintained by the Embrapa CBRG, which gathers in a single repository more than 1300 physicochemical and structural descriptors of all amino acid residues for each chain of all protein structures deposited in the PDB (Protein Data Bank – a world public repository where all macromolecular structures deciphered so far were deposited).

Based on the obtained results, we conclude that a given nanoenvironment can be described not by a single descriptor, but by a set of descriptors, and that this set of descriptors varies according to the element of the protein structure selected from a hierarchically superior one. This differentiates a given nanoenvironment from the rest of the protein and even from other nanoenvironments in the same protein. The knowledge acquired from the study of different nanoenvironments allows specialists in different areas, such as experts in plant improvement, in search of new pesticides, or researchers in search of more sustainable fuels, to advance their work with greater molecular introspection and use of more precise and refined tools, working at the most fundamental level (molecular-atomic) of all biologically relevant processes for medicine, agriculture, livestock, etc.

Protein nanoenvironments and their characteristics

The local structural environment of proteins, here called the nanoenvironment (Neshich et al., 2015), characterizes the functional purpose of different protein districts, also known as "structural sites" in proteins. It is therefore suggested that the local environment at each protein point and/or region reflects not only its structural role, but also its contribution in providing the necessary characteristics for the functional purpose of each protein. For example, protein-protein communication is performed via protein interfaces: amino acid residues at the same site have some particular characteristics that not only differentiate them from other residues on the free surface of the protein, but also allow specific and selective binding between proteins and the realization of their biochemical function (Moraes et al., 2014). Similarly, the function of an enzyme is normally related to the activity of its catalytic amino acid residues (Catalytic Site Residues – CSR). These very peculiar residues are inserted in a very specific nanoenvironment, also defined by the contribution of the CSRs themselves. Consequently, the enzymatic function can be described by the characteristics of the CSRs and their surroundings (Salim, 2015). Based on these considerations, and assuming that the local nanoenvironment defines the protein function, this is a concept that can be used to obtain specific metrics to quantify and describe other nanoenvironments.

The exploration of nanoenvironments properties of can be done through a method that is both self-explanatory and intuitive. Suppose it is possible to insert an imaginary probe anywhere in a protein structure and obtain as a result, a diagnosis describing the characteristics of the environment in which the probe is inserted. This type of physical intervention cannot be carried-out, and therefore the probe needs to be replaced by calculating values, metrics, and forces that we want to quantify at each particular site/point. This approach resembles the GRID method for calculating molecular interaction fields in drug development (Goodford, 1985; Von Itzstein et al., 1993), but with a different focus. Its advantage is that any amino acid residue, or any of its main or side chain atoms, can serve as the center for the probe. With this selected point, the interactions of all forces can be estimated, cataloged, and stored in an appropriate relational database – in our case, the STING_RDB (Oliveira, 2007). Once stored, the attributes and their respective values can be mapped back to the protein structure, the protein sequence, or even the nucleotide sequence of the gene that encodes that protein, and can be used for visual inspection or statistical and/or numerical analyses. Our hypothesis is that any specific environment (the nanoenvironment) has a precise tuning of the specific physicochemical and structural writers for the performance of its function and, thus, can be identified and classified accordingly. For example, interfaces for protein contacts, which are specific areas of the protein occupying part of their surface, can be expected to have characteristics sufficiently different from the amino acid residues found in free surface areas (Moraes et al., 2014). In fact, we consider such an assumption to be part of the biological requirements for performing a specific function: in this example, the function is actually a kind of "communication" between very specific protein partners. Therefore, a nanoenvironment is accurately characterized by its physicochemical and/or structural descriptors and their corresponding values, making it possible to distinguish it from the rest of the protein structure. It is also possible to predict the coordinates of these districts in other proteins (homologous or not) that have not yet been chemically and functionally characterized through computational techniques and machine learning statistics.

Among the most studied protein nanoenvironments, ten stand out, as follows:

1) Protein interfaces: These are intersections of protein surfaces, where the two proteins approach and touch, building a macromolecule homo or heterocomplex (Moraes et al., 2014).

2) Antibody and antigen interfaces: as in case 1, but the two proteins in question are an antibody and an antigen (Viart et al., 2016).

3) Protein surface hot spots: locations delimited from the surface area of the protein, obligatorily located at its interface, and with identified hydrophobic amino acids prone to interact with similar residues from the complementary interface of the other protein (Pereira, 2012).

4) Interfaces between proteins and DNA: as in case 1, but with the two molecules in question being a protein and a DNA molecule.

5) Interfaces between proteins and ligands: as in case 1, here the two molecules in question are a protein and a ligand (Borro et al., 2016).

6) Interfaces between proteins and membranes.

7) Amino acid residues from catalytic sites: identifying the amino acid residues that form the enzymes catalytic site, determining their function (Salim, 2015).

8) Allosteric sites: usually located on the protein surface. When occupied by a particular molecule, they control the speed of a chemical reaction that the protein performs, using, as a rule, its set of CSRs as part of its function.

9) Secondary structure elements: physicochemical and structural characterization of α-helices (Mazoni et al., 2018), β-sheets and turns.

10) The depth of range of local sensing between amino acids: a measure often used to delimit the distance over which atoms, with their charges (and other characteristics), still exert some influence in remote locations, but within the aforementioned limit (Silveira et al., 2009).

Items: 1 to 6 describe the interfaces in general; 7 and 8 describe chemical activity of proteins; and 9 and 10 describe structural characteristics of proteins in general.

List of physicochemical and structural descriptors that characterize specific nanoenvironments

Currently, the Blue Star STING (BSS) (Neshich et al., 2006) has 32 independent physicochemical and structural protein descriptor types or classes (Table 1) (Neshich et al., 2005), and a total of 1,307 variations of these descriptors are pre-calculated (using different parameterizations) and stored in the STING_RDB database (Oliveira, 2007). On May 18, 2020, the STING_RDB had 151,711 structures, with 467,038 chains and 95,148,233 amino acid residues. For each, 1,307 parameters were pre-calculated, totaling 12×10^9 records in the database. Among these, some were chosen to be used in the nanoenvironment characterization and in the composition of their dictionary, considering only those that are more likely to be associated with pattern recognition processes in the selected proteins. For an adequate definition of the catalytic residues nanoenvironment and which is generally valid also for the other mentioned nanoenvironments, based on the physicochemical and structural descriptors, the descriptors referring to the conservation of amino acids were initially discarded, since these parameters are a measure of a set of homologous proteins and do not reflect any feature present in the protein structure (Salim, 2015).

Table 1. List of the 32 physicochemical descriptors classes and Blue Star STING structures.

Blue Star STING Descriptor Classes	
1. ResBoxes	17. Hot spots
2. Intra-chain atomic contacts [ITC]	18. Sequence conservation [HSSP]
3. The inter-chain atomic contacts [IFC]	19. Sequence conservation [SH_2Q^s]
4. ITC contacts energy	20. Solvent accessibility
5. IFC contacts energy	21. Dihedral angles
6. Interface area [IF]	22. Pockets/cavities
7. Water contacting [WC]	23. Electrostatic potential
8. Ligand pocket forming [LP]	24. Hydrophobicity
9. Surface forming [SF] residues	25. Curvature
10. Prosite	26. Distance from the N-/C-terminal
11. ProTherm	27. Density
12. Secondary structure indicator [PDB]	28. Sponge
13. Secondary structure indicator [DSSP]	29. Order of cross presence
14. Secondary structure [STRIDE]	30. Order of cross link
15. Multiple occupancy	31. Rotamers
16. Temperature factor	32. Space clash

Contributions

What does knowledge about protein nanoenvironments entail?

The protein's structure defines its functionality. However, how this is performed and which structural features contribute to their function remains to be fully deciphered. To answer this question, it is necessary to consider the structural elements (also called protein districts or nanoenvironments) rather than considering the structure as a whole. These elements, on the other hand, must be understood based on the physicochemical and structural characteristics from the amino acid residue properties, which interact with each other and create a new hierarchical structural element. Only by considering these elements in the structural hierarchy can we understand that the functionality of proteins can be broken down into communication elements, such as interfaces, constructive elements, secondary structure, and elements of chemical activity. The latter normally give rise to the functionality and specificity of the protein as a whole. Following this reasoning, each element in the structural hierarchy has its distinctive local characteristic and, consequently, its local function. It is clear that a general and detailed knowledge about protein nanoenvironments is, basically, a dictionary with which we can construct complex expressions to describe the structural-functional protein relationships.

A dictionary of nanoenvironment descriptors will impact the variety of research aimed at innovation in areas such as agriculture, medicine, and biology in general

A compilation of the results of work done since 1998 – when the STING platform was launched in the US as an integral part of platforms offered for protein structural analysis at the Brookhaven National Laboratory, the headquarters of the Protein Structures Database (PDB) – it resulted in a website called: "Dictionary of Internal Protein Nanoenvironments" (DIPN)[1].

Figures 1 to 3 show the general interface of the new CBRG offered by Embrapa Digital Agriculture. It is an introductory page, with a general description of the purpose of this platform with detailed elements listed in a functional order.

Figure 1 shows the entry page of the Dictionary of Internal Protein Nanoenvironments (DIPN) platform, indicating the purpose of this product, the options for access, the site organization logistics, and the list of the ten most studied protein nanoenvironments.

In Figure 2 we have a Dictionary of Internal Protein Nanoenvironments (DIPN) platform page showing six of the ten available nanoenvironments, with a short description and access to details of the entry of each option: a) protein-DNA interfaces, b) protein interfaces-membrane, c) elements of secondary structure.

Figure 3 presents a Dictionary of Internal Protein Nanoenvironments (DIPN) platform page showing three more of the ten available nanoenvironments, with a short description and access to details of each entry option: a) residues from the catalytic site, b) allosteric sites, and c) depth of local sensing range between amino acids.

[1] Available at: https://www.proteinnanoenvironments.cnptia.embrapa.br/index.html

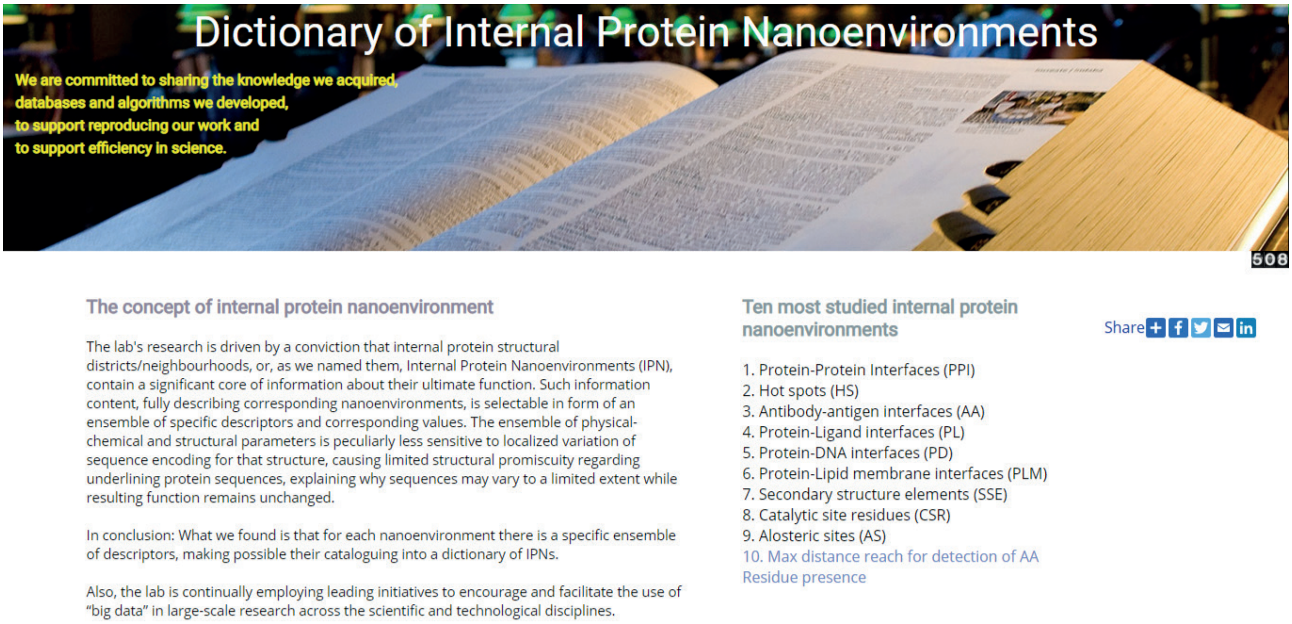


Figure 1. Dictionary of Internal Protein Nanoenvironments (DIPN) platform entry page.

Source: Embrapa (2020).

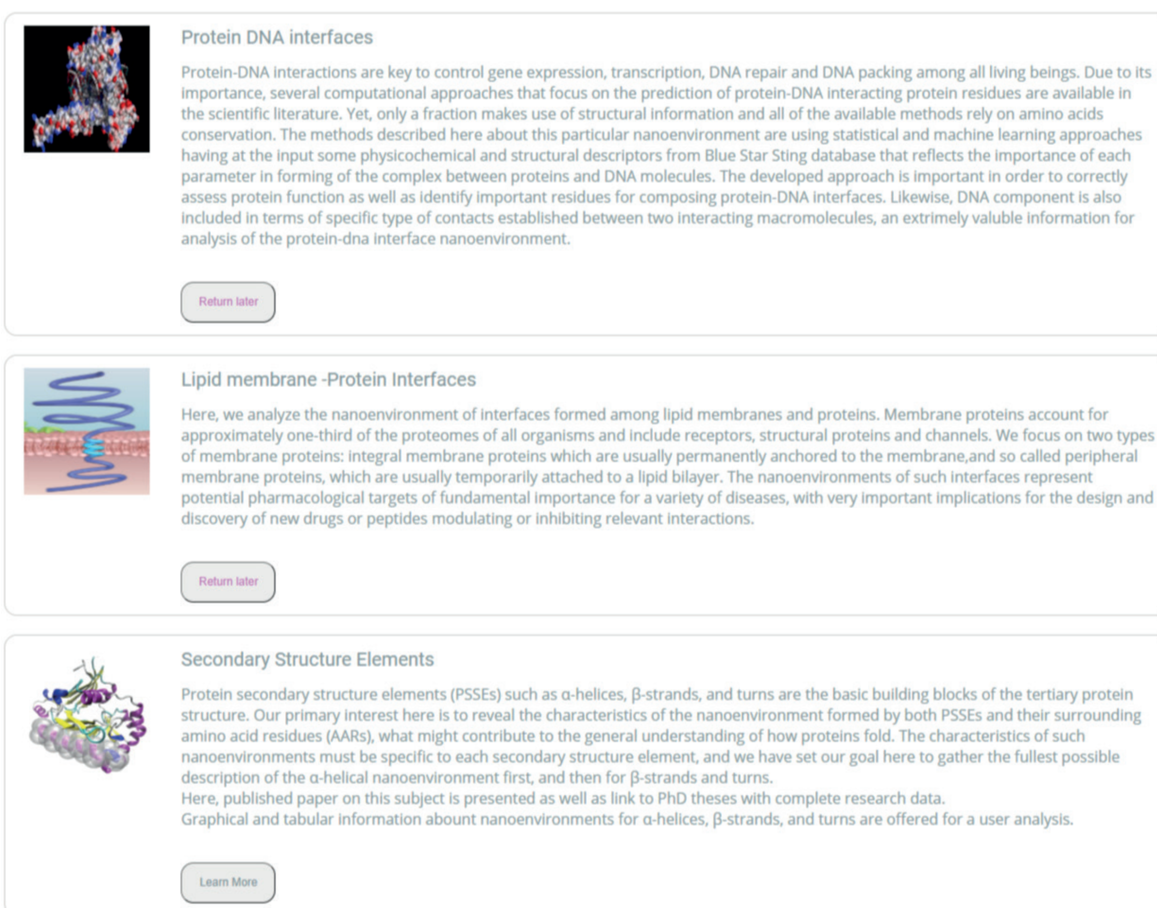
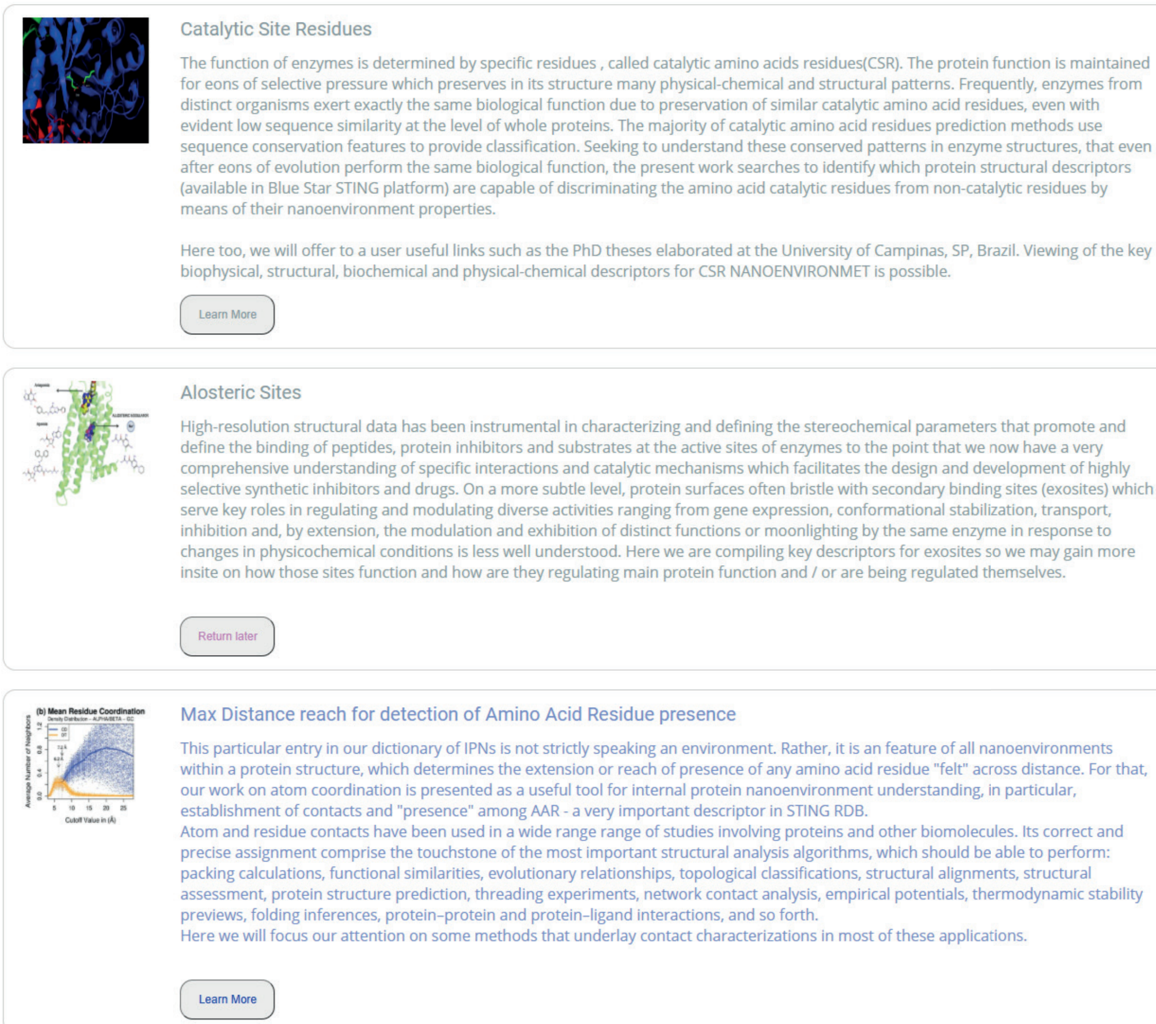


Figure 2. Dictionary of Internal Protein Nanoenvironments (DIPN) platform page.

Source: Embrapa (2020).


Catalytic Site Residues

The function of enzymes is determined by specific residues , called catalytic amino acids residues(CSR). The protein function is maintained for eons of selective pressure which preserves in its structure many physical-chemical and structural patterns. Frequently, enzymes from distinct organisms exert exactly the same biological function due to preservation of similar catalytic amino acid residues, even with evident low sequence similarity at the level of whole proteins. The majority of catalytic amino acid residues prediction methods use sequence conservation features to provide classification. Seeking to understand these conserved patterns in enzyme structures, that even after eons of evolution perform the same biological function, the present work searches to identify which protein structural descriptors (available in Blue Star STING platform) are capable of discriminating the amino acid catalytic residues from non-catalytic residues by means of their nanoenvironment properties.

Here too, we will offer to a user useful links such as the PhD theses elaborated at the University of Campinas, SP, Brazil. Viewing of the key biophysical, structural, biochemical and physical-chemical descriptors for CSR NANOENVIRONMET is possible.

Learn More

Alosteric Sites

High-resolution structural data has been instrumental in characterizing and defining the stereochemical parameters that promote and define the binding of peptides, protein inhibitors and substrates at the active sites of enzymes to the point that we now have a very comprehensive understanding of specific interactions and catalytic mechanisms which facilitates the design and development of highly selective synthetic inhibitors and drugs. On a more subtle level, protein surfaces often bristle with secondary binding sites (exosites) which serve key roles in regulating and modulating diverse activities ranging from gene expression, conformational stabilization, transport, inhibition and, by extension, the modulation and exhibition of distinct functions or moonlighting by the same enzyme in response to changes in physicochemical conditions is less well understood. Here we are compiling key descriptors for exosites so we may gain more insite on how those sites function and how are they regulating main protein function and / or are being regulated themselves.

Return later

Max Distance reach for detection of Amino Acid Residue presence

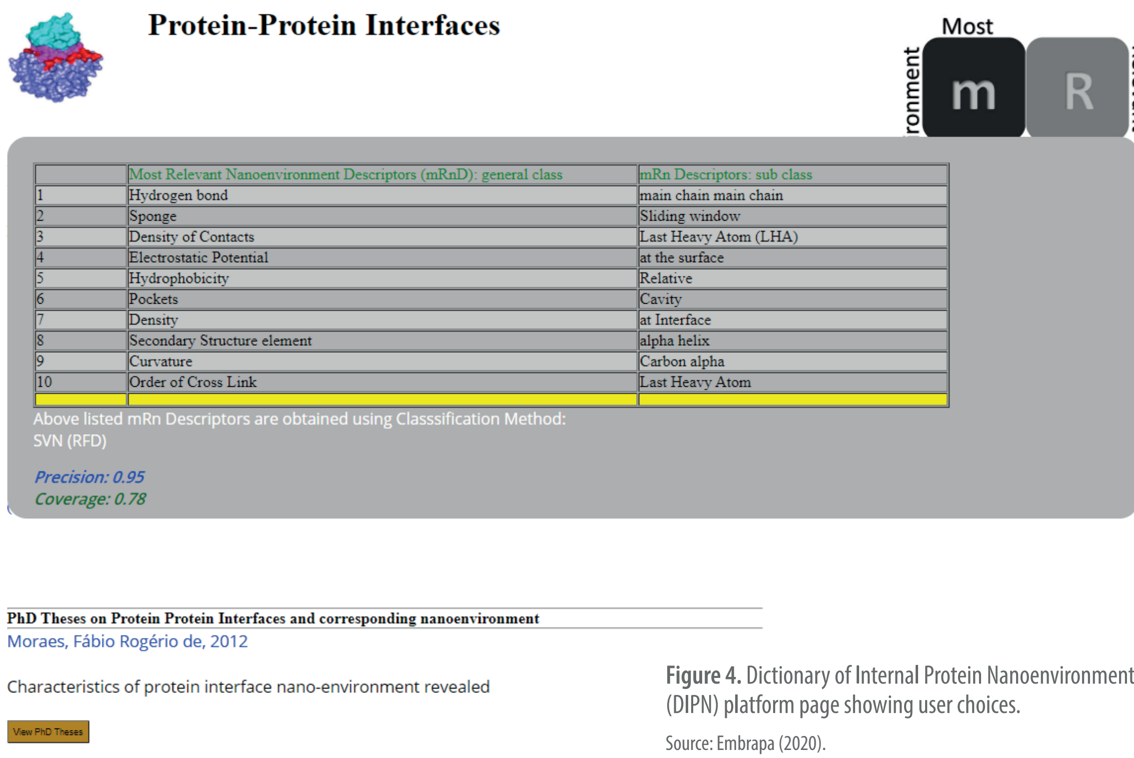
This particular entry in our dictionary of IPNs is not strictly speaking an environment. Rather, it is an feature of all nanoenvironments within a protein structure, which determines the extension or reach of presence of any amino acid residue "felt" across distance. For that, our work on atom coordination is presented as a useful tool for internal protein nanoenvironment understanding, in particular, establishment of contacts and "presence" among AAR - a very important descriptor in STING RDB.
Atom and residue contacts have been used in a wide range range of studies involving proteins and other biomolecules. Its correct and precise assignment comprise the touchstone of the most important structural analysis algorithms, which should be able to perform: packing calculations, functional similarities, evolutionary relationships, topological classifications, structural alignments, structural assessment, protein structure prediction, threading experiments, network contact analysis, empirical potentials, thermodynamic stability previews, folding inferences, protein–protein and protein–ligand interactions, and so forth.
Here we will focus our attention on some methods that underlay contact characterizations in most of these applications.

Learn More

Figure 3. Dictionary of Internal Protein Nanoenvironments (DIPN) platform page.

Source: Embrapa (2020).

In Figure 4, the user can see the presentation details of one of the nanoenvironments: protein interfaces. The purpose of DIPN is to provide the user with information that indicates which are the most relevant descriptors that, with their specificity and broad coverage, describe the nanoenvironment selected for analysis. At the bottom of Figure 4 there is a table with the ten descriptors of the most relevant protein interfaces. These are: 1) main-chain main-chain hydrogen bonds; 2) spongicity (in a sliding window mode); 3) contact density between amino acids (centered on the last heavy atom of the amino acid side chain); 4) electrostatic potential on the protein surface; 5) hydrophobicity (on the relative scale); 6) structural pockets (cavity type); 7) atomic density on the surface; 8) element of the secondary structure present (α-helix); 9) curvature from α-carbon; and 10) the order of cross-linking (starting from the last heaviest atom in the side chain). These descriptors can be understood as main requirements that demand their inclusion so that a set of amino acids, not necessarily contiguous in the primary sequence, build a set that can be considered apt to form an interface with another protein. Then, the platform informs which statistical classification method was used to obtain this ranking of the importance of the descriptors (in this case: Support Vector Machine and Random Forest), and also informs with what precision and coverage the conclusions were reached. In this case, 0.95 and 0.78, respectively. On that



	Most Relevant Nanoenvironment Descriptors (mRnD): general class	mRn Descriptors: sub class
1	Hydrogen bond	main chain main chain
2	Sponge	Sliding window
3	Density of Contacts	Last Heavy Atom (LHA)
4	Electrostatic Potential	at the surface
5	Hydrophobicity	Relative
6	Pockets	Cavity
7	Density	at Interface
8	Secondary Structure element	alpha helix
9	Curvature	Carbon alpha
10	Order of Cross Link	Last Heavy Atom

Above listed mRn Descriptors are obtained using Classsification Method:
SVN (RFD)

Precision: 0.95
Coverage: 0.78

PhD Theses on Protein Protein Interfaces and corresponding nanoenvironment
Moraes, Fábio Rogério de, 2012

Characteristics of protein interface nano-environment revealed

View PhD Theses

Figure 4. Dictionary of Internal Protein Nanoenvironments (DIPN) platform page showing user choices.

Source: Embrapa (2020).

same page, you will find a variety of additional information, such as links to the doctoral thesis that generated the results and publications describing pertinent work to the subject (in Figure 5 we are illustrating the abstract of this publication). Lastly, there is a link so that the user can access the software if one wants to generate new data for a set of proteins for biological interest.

In Figure 4, we have the Dictionary of Internal Protein Nanoenvironments (DIPN) platform page showing user options once the protein interfaces item is selected. At the top of this figure, there is an indication of relevant publications to the subject and a list of software. Next, an abstract of the main publication can be seen describing our work with the nanoenvironment of protein interfaces, with a corresponding pointer to the original publication. On the upper right side, there is an icon with the title: MRND (Most Relevant Nanoenvironment Descriptors). By hovering over this icon, a window is opened with information indicated in the icon's title.

In Figure 6, we present the available items for accessing the software page which helps the user to prepare a list of descriptors for a set of proteins of interest. In Figure 7 we have the two main options for preparing protein interface data: LDA methodology (linear models for inferring the list of the most relevant descriptors of protein interfaces) and SHI option, an alternative methodology that determines the hydrophobicity index on the surface protein, an accurate interface indicator. The user can find a tutorial to find details about the software, datamart description for defining benchmarks and description of the complexes used in the training of the method using both homo and heteroprotein complexes. In Figures 1 to 7 we show only the most crucial entries of the DIPN platform.

The platform is complex and requires the knowledge of a trained computer biologist to process the data for a set of selected proteins. However, molecular biology specialists interested in knowing which descriptors are most relevant for each nanoenvironment listed in the DIPN platform can do so in a reasonable time, with minimal training, and know which characteristics of these nanoenvironments

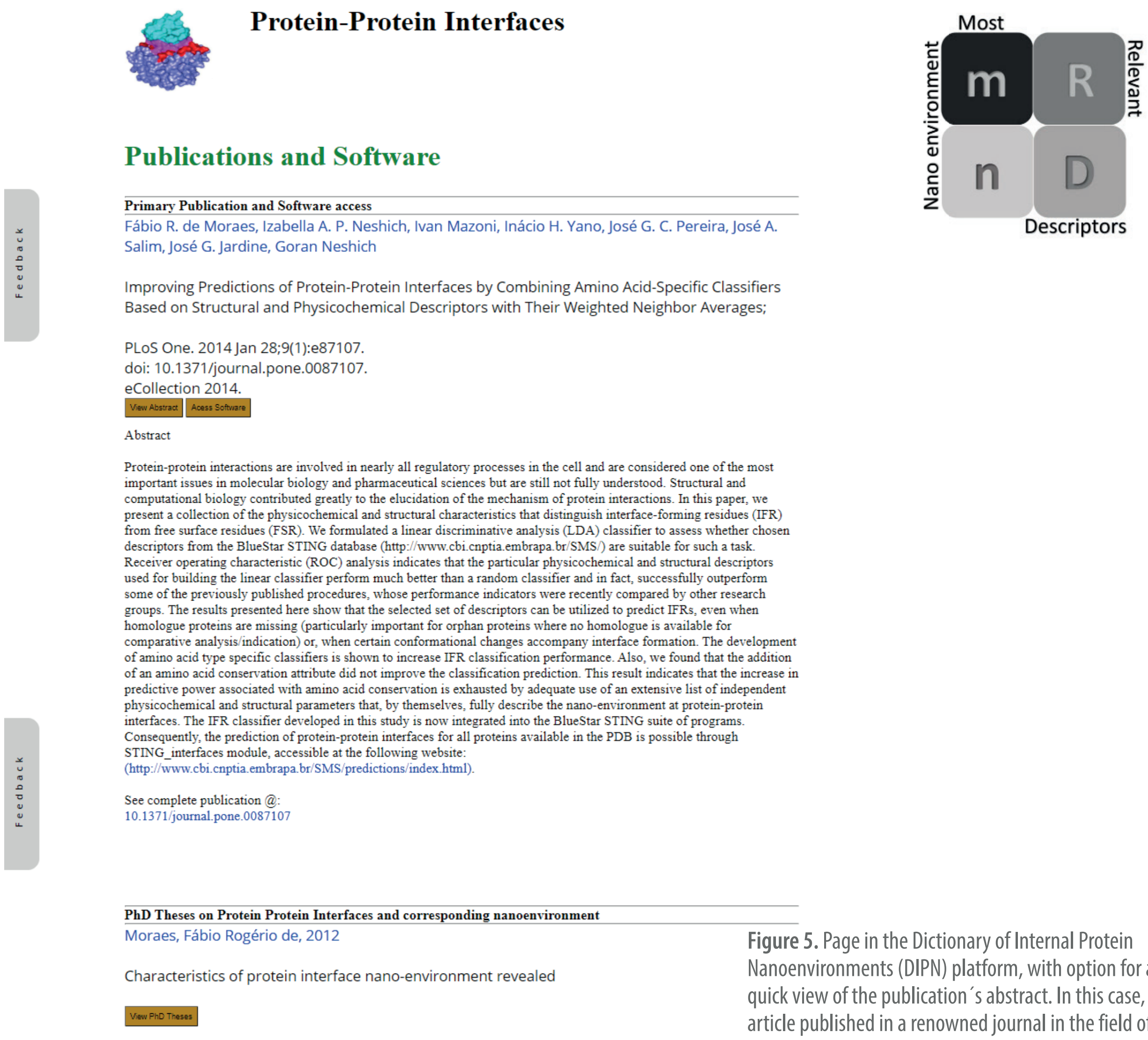


Figure 5. Page in the Dictionary of Internal Protein Nanoenvironments (DIPN) platform, with option for a quick view of the publication´s abstract. In this case, an article published in a renowned journal in the field of computational biology about nanoenvironment.

Source: Embrapa (2020).

are crucial. Thus, there are candidates that cannot perform modifications, for example in attempts that require site-directed mutations in the proteins of interest. The algorithm options for using or even accessing the source code are provided in order to offer a complete work environment, including for those computational biologists who wish to adapt the algorithms to their own requirements. This allows the sharing of work already carried out by Embrapa, and may be modified by colleagues in other laboratories for specific purposes.

Final considerations

With a dictionary of descriptors of the main protein nanoenvironments, a reality is built that guides researchers and enables advancement in areas aiming to intensify innovation in agriculture, medicine, and biology in general. It is understood that a compilation of the essential descriptors of the 10 most studied

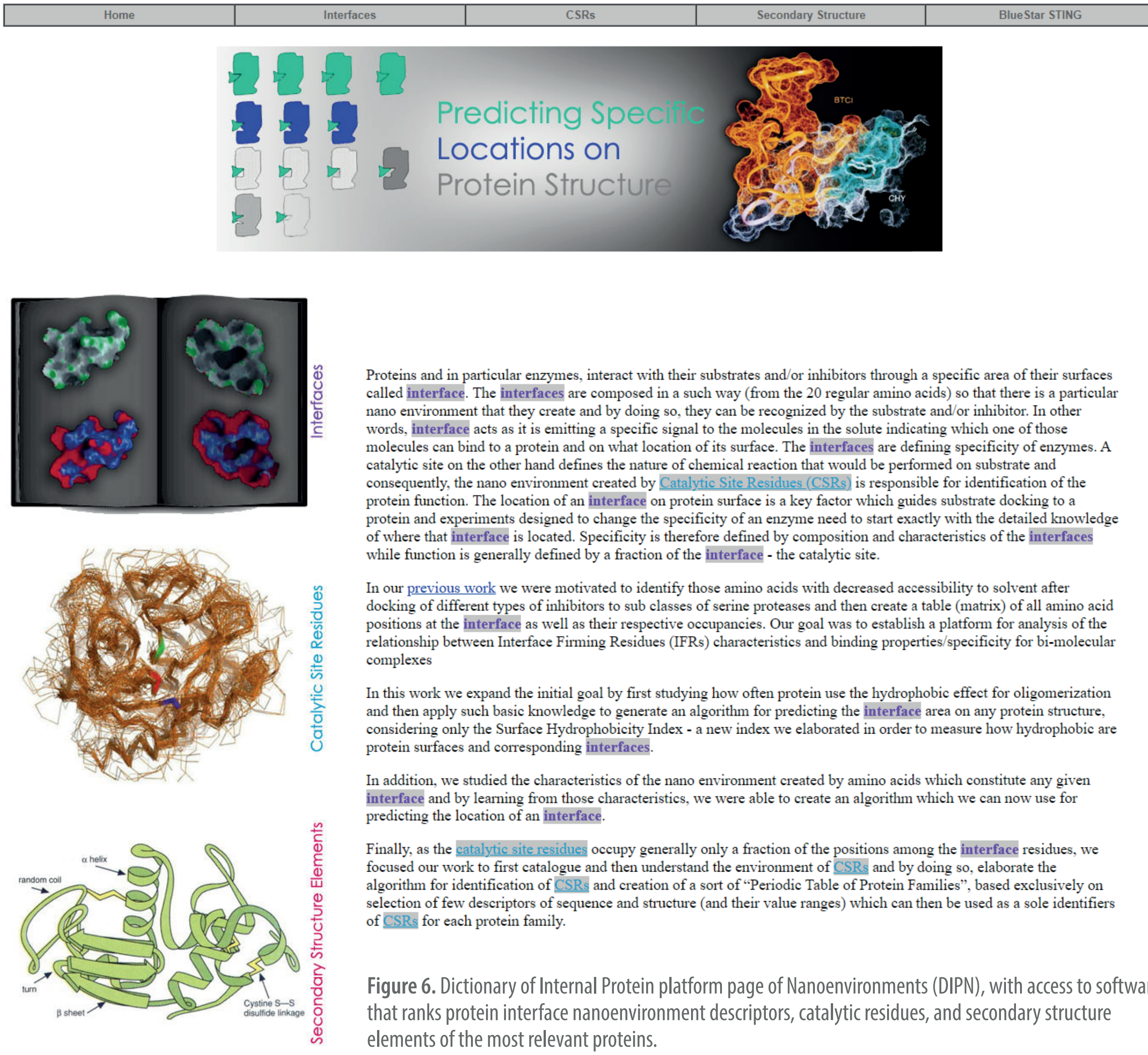


Figure 6. Dictionary of Internal Protein platform page of Nanoenvironments (DIPN), with access to software that ranks protein interface nanoenvironment descriptors, catalytic residues, and secondary structure elements of the most relevant proteins.

Source: Screen captured from the DIPN platform (available at: https://www.proteinnanoenvironments.cnptia.embrapa.br/index.html)

protein nanoenvironments may provide an optimized condition for the most accurate, effective, and effective design of new drugs, pesticides, vaccines, inhibitors, catalysts, and antibodies. We can use as an example, the applicability of the content presented in this chapter and mention some of the technologies which the CBRG of Embrapa Digital Agriculture managed to file. This resulted in the application for four patents over the years, focusing mainly on understanding, learning, and in the analysis of protein nanoenvironments which were crucial to the solution of some biologically relevant demands. The research group also focused on a path to the necessary impacts in the field for producers who needs to use the technology in order to avoid losses and improve its effectiveness. Some of them are listed below:

Fungicide: a method for designing a new fungicide by computationally designing new compounds with potential inhibitory function on the endopolygalacturonase enzyme, involved in invasion processes in plant cells. (Neshich et al., 2013a)

Biodiesel: method for predicting mutants that increase the surface hydrophobicity index of proteins. (Neshich et al., 2013b)

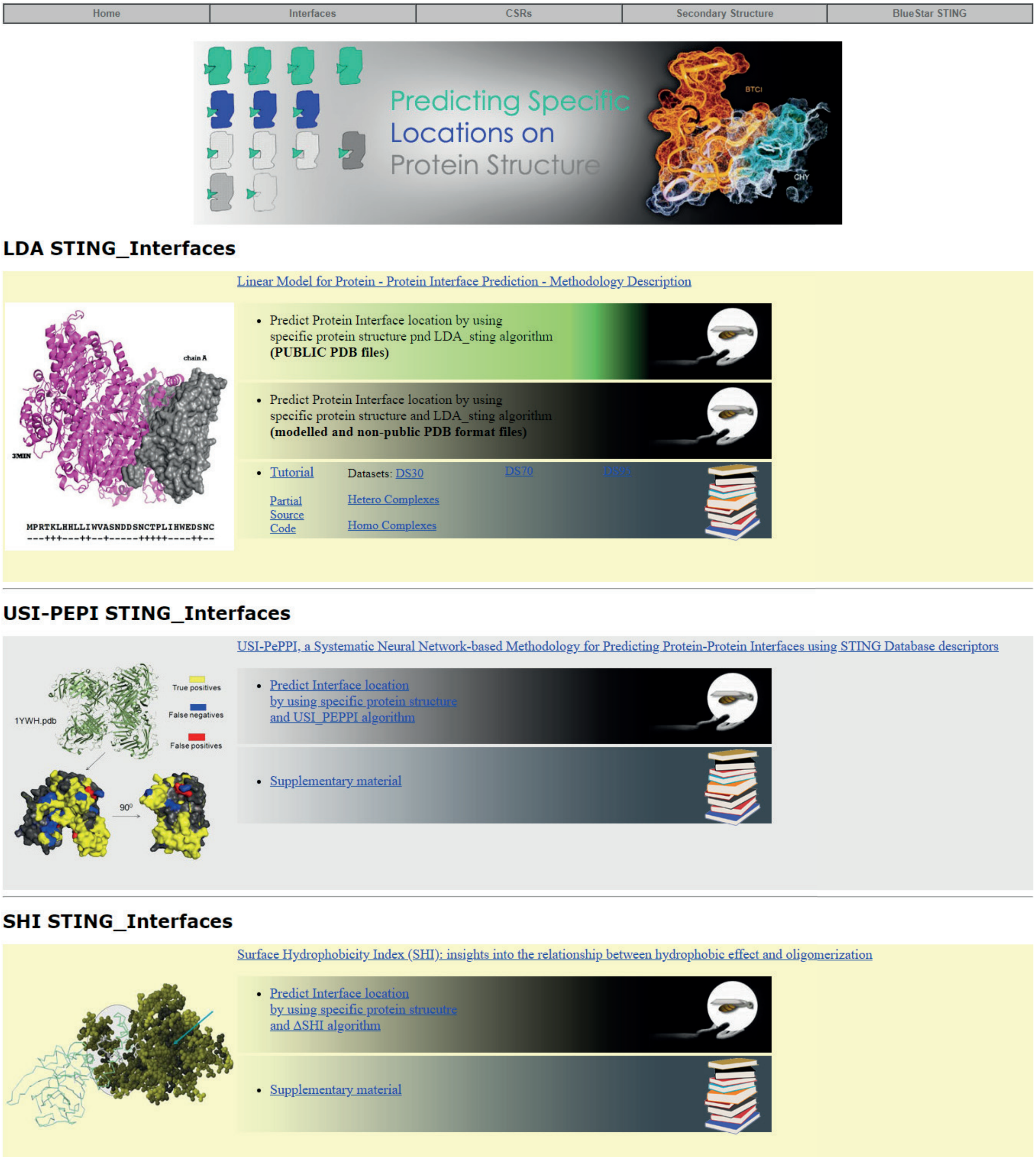


Figure 7. Page of the Dictionary of Internal Protein Nanoenvironments (DIPN) platform, with options for users who want to rank the most relevant protein interfaces descriptors by using a set of proteins of interest for a biological problem that requires their engagement.

Insecticide: computational design for new alpha-amylase inhibitors. (Neshich et al., 2013c)

Bactericide: identification of therapeutic targets for computational design of drugs against bacteria possessing the pilt protein. (Neshich et al., 2012)

These four technologies reflect the strong interdependence between the demands of modern agriculture and knowledge, calling for an innovative, interdisciplinary, and molecular approach, interconnected with mathematics, computation, and statistics for advances in the increasingly complex needs of the productive sector. The example of the CBRG at Embrapa Digital Agriculture is a manifestation of national possibilities for the potential of technological development at the highest and most competitive level. The research carried out by the CBRG at Embrapa Digital Agriculture drew the attention of international collaborators and colleagues from the most renowned universities, such as Oxford, Cambridge, MIT, followed by companies with great digital impact, such as Microsoft Research and companies in the field of agricultural pesticides, such as Bayer and BASF. Half a hundred publications in scientific journals with an average impact factor of 3, and several with impact factors above 11. There were hundreds of lectures and seminars, international courses, and workshops, as well as international meeting organized here in national territory with the participation of several Nobel Prize-winning scientists. Fifty software packages were published and made available for the scientific community, as well as dozens of databases in the field of computational structural biology, including the STING_RDB. Twenty-six projects were approved (90%) by external sources to Embrapa, with funding approaching 4 million dollars and total deliverables approaching 500 million. This entire library of results and professional awards was a stepping-stone for us to transform our acquired knowledge into something applicable to the production chain and developing these solutions into products for national and international markets. Therefore, the platform called Dictionary of Internal Protein Nanoenvironments was developed while considering the applications from our knowledge, but with patience and determination to stay on the path that requires time, learning, and basic science, since scientific applications do not exist without the former.

References

BORRO, L.; YANO, I. H.; MAZONI, I.; NESHICH, G. Binding affinity prediction using a nonparametric regression model based on physicochemical and structural descriptors of the nano-environment for protein-ligand interactions. In: STRUCTURAL BIOINFORMATICS AND COMPUTATIONAL BIOPHYSICS, 2016, Orlando. **Proceedings**... Orlando: [s.n.], 2016. p. 116-117.

EMBRAPA. Computational Biology Research Group. **Dictionary of Internal Protein NanoEnvironment**s. Available at: https://www.proteinnanoenvironments. cnptia.embrapa.br/index.html. Accessed on: 18 May 2020.

GOODFORD, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. **Journal of Medicinal Chemistry**, v. 28, n. 7, p. 849-857, July 1985. DOI: 10.1021/jm00145a002.

MAZONI, I.; BORRO, L. C.; JARDINE, J. G.; YANO, I. H.; SALIM, J. A.; NESHICH, G. Study of specific nanoenvironments containing α-helices in all-α and (α+ β)+(α/β) proteins. **PLOS One**, v. 13, n. 7, p. 1-25, 2018. Artigo e0200018. DOI: 10.1371/journal.pone.0200018.

MORAES, F. R. de; NESHICH, I. A. P.; MAZONI, I.; YANO, I. H.; PEREIRA, J. G. C.; SALIM, J. A.; JARDINE, J. G.; NESHICH, G. Improving predictions of protein-protein interfaces by combining amino acid-specific classifiers based on structural and physicochemical descriptors with their weighted neighbor averages. **PLOS ONE**, v. 9, n. 1, p. 1-15, 2014. DOI: 10.1371/journal.pone.0087107.

NESHICH, G.; JARDINE, J. G.; NESHICH, I. A.; SALIM, J. A.; MAZONI, I. **EUA Patente Nº WO2013/110147Al**, 2013c.

NESHICH, G. E. A.; BORRO, L. C.; HIGA, R. H.; KUSER, P. R.; YAMAGISHI, M. E.; FRANCO, E. H.; KRAUCHENCO, J. N.; FILETO, R.; RIBEIRO, A. A.; BEZERRA, G. B.; VELLUDO, T. M.; JIMENEZ, T. S.; FURUKAWA, N.; TESHIMA, H.; KITAJIMA, K.; BAVA, A.; SARAI, A. TOGAWA, R. C.; MANCINI, A. L. The diamond STING server. **Nucleic Acids Research**, v. 33, n. 2, p. W29-W35, July 2005. Supplement. DOI: 10.1093/nar/gki397.

NESHICH, G.; MAZONI, I.; OLIVEIRA, S. R. M.; YAMAGISHI, M. E. B.; KUSER-FALCÃO, P. R.; BORRO, L. C.; MORITA, D. U.; SOUZA, K. R. R.; ALMEIDA, G. V.; RODRIGUES, D. N.; JARDINE, J. G.; TOGAWA, R. C.; MANCINI, A. L.; HIGA, R. H.; CRUZ, S. A. B.; VIEIRA, F. D.; SANTOS, E. H.; MELO, R. C.; SANTORO, M. M. The Star STING server: a multiplatform environment for protein structure analysis. **Genetics and Molecular Research**, v. 5, n. 4, p. 717-722, 2006.

NESHICH, G. E. A.; NESHICH, I. A. P.; MORAES, F.; SALIM, J. A.; BORRO, L.; YANO, I. H.; MAZONI, I.; JARDINE, J. G.; ROCCHIA, W. Using structural and physical-chemical parameters to identify, classify, and predict functional districts in proteins – the role of electrostatic potential. In: ROCCHIA, W.; SPAGNUOLO, M. (ed.). **Computational electrostatics for biological applications**: geometric and numerical approaches to the description of electrostatic interaction between macromolecules. Cham: Springer, 2015. p. 227-254. DOI: 10.1007/978-3-319-12211-3_12.

NESHICH, G.; JARDINE, J. G.; NESHICH, I. A.; SALIM, J. A.; MAZONI, I. **EUA Patente Nº WO2012/031343A2**, 2012.

NESHICH, G.; JARDINE, J. G.; NESHICH, I. A.; SALIM, J. A.; MAZONI. **EUA Patente Nº WO2013097012A1**, 2013a.

NESHICH, G.; JARDINE, J. G.; NESHICH, I. A.; SALIM, J. A.; MAZONI, I. **EUA Patente Nº WO2013/016794Al**, 2013b.

OLIVEIRA, S. R. de M.; ALMEIDA, G. V.; SOUZA, K. R. R.; RODRIGUES, D. N.; KUSER-FALCÃO, P. R.; YAMAGISHI, M. E. B.; SANTOS, E. H. dos; VIEIRA, F. D.; JARDINE, J. G.; NESHICH, G. Sting_RDB: a relational database of structural parameters for protein analysis with support for data warehousing and data mining. **Genetics and Molecular Research**, v. 6, n. 4, p. 911-922, 2007.

PEREIRA, J. G. D. C. **Caracterização dos aminoácidos da interface proteína-proteína com maior contribuição na energia de ligação e sua predição a partir dos dados estruturais**. 2012. 106 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Genética e Biologia Molecular, Instituto de Biologia, Universidade Estadual de Campinas, Campinas.

SALIM, J. A. **Aplicação de técnicas de reconhecimento de padrões usando os descritores estruturais de proteínas da base de dados do software STING para discriminação do sítio catalítico de enzimas**. 2015. 214 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Engenharia Elétrica, Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas.

SILVEIRA, C. H. da; PIRES, D. E. V.; MINARDI, R.; RIBEIRO, C.; VELOSO, C. J. M.; LOPES, J. C. D.; MEIRA JÚNIOR, W.; NESHICH, G.; RAMOS, C. H. I.; HABESCH, R.; SANTORO, M. M. Protein cutoff scanning: a comparative analysis of cutoff dependent and cutoff free methods for prospecting contacts in proteins. **Proteins**: structure, function, and bioinformatics, v. 74, n. 3, p. 727-743, Feb. 2009. DOI: 10.1002/prot.22187.

VIART, B.; DIAS-LOPES, C.; KOZLOVA, E.; OLIVEIRA, C. F. B.; NGUYEN, C.; NESHICH, G.; CHÁVEZ-OLÓRTEGUI, C.; MOLINA, F.; FELICORI, L. F. EPI-peptide designer: a tool for designing peptide ligand libraries based on epitope–paratope interactions. **Bioinformatics**, v. 32, n. 10, p. 1462-1470, May 2016. DOI: 10.1093/bioinformatics/btw014.

VILLANUEVA, J. C. How many atoms are there in the Universe. **Universe Today**, July 30, 2009. Available at: https://www.universetoday.com/36302/atoms-in-the-universe. Accessed on: 18 May 2020.

VON ITZSTEIN, M.; WU, W.-Y.; KOK, G. B.; PEGG, M. S.; DYASON, J. C.; JIN, B.; VAN PHAN, T.; SMYTHE, M. L.; WHITE, H. F.; OLIVER, S. W.; COLMAN, P. M.; VARGHESE, J. N.; RYAN, D. M.; WOODS, J. M.; BETHELL, R. C.; HOTHAM, V. J.; CAMERON, J. M.; PENN, C. R. Rational design of potent sialidase-based inhibitors of influenza virus replication. **Nature**, v. 363, p. 418-423, 1993. DOI: 10.1038/363418a0.

Photo: DC Studio (AdobeStock)

10

Applications of bioinformatics in agriculture

Adhemar Zerlotini Neto| Antonio Nhani Junior | Fábio Danilo Vieira | Leandro Carrijo Cintra | Maurício de Alvarenga Mudadu | Paula Regina Kuser Falcão | Poliana Fernanda Giachetto

Introduction

Biotechnology has been fundamental for the progress observed in Agriculture over the last 30 years. Bioinformatics, the multidisciplinary area responsible for analyzing the large volume of data resulting from genomic technologies, was essential in this progress. With the arrival of next-generation sequencing technologies, an extraordinarily large volume of genomic data that needed to be analyzed was produced. In the era of digital transformation, the ability to generate biological data more rapidly, more affordably and in greater volume, produces an enormous amount of data, Big Data. This large and growing volume of data requires solutions in at least three spheres: scalable infrastructure, data management and intelligent use of that data.

Bioinformatics uses computational tools to answer complex biological questions and contribute to innovative results. The theme involves the use of high-performance computing infrastructure and tools to organize, analyze, integrate, process, simulate and store large volumes of data derived from *in vivo* and *in vitro* experiments. A challenge for bioinformatics is to integrate the heterogeneous data generated by the "omics" sciences (both with each other and with the data generated by traditional sciences), allowing discoveries that go beyond what is possible in each of the individual disciplines. Several new layers of omics, such as analysis of genomes, metabolomes, transcriptomes, or interactomes, have become important for research advances. The integration of all this information allows making discoveries and improving the knowledge of biological systems.

Access to high storage and processing capacity, with powerful indexing algorithms, as well as machine learning applications, is crucial for the execution of bioinformatics activities. More importantly, a trained and constantly updated team to assist in planning data generation processes, data analysis and extraction/acquisition of new knowledge from Big Data is what will enable Embrapa to be a relevant actor in this area of knowledge.

In this context, in 2011, Embrapa's Multiuser Laboratory of Bioinformatics (LMB) was created to provide bioinformatics support to RD&I projects aligned with Embrapa's strategic objectives. Since its creation, the LMB has already contributed in a broad portfolio of projects, within three operating guidelines:

- **Access to the computing park**, hence its high-performance infrastructure.
- **Consulting in the analysis of biological** data that require high-performance computing, whether due to the volume of data or the complexity of the analyses.
- **Training** for multiplying skills through courses and other training actions.

The LMB has performed research projects at Embrapa and partner institutions that involve more than 20 crops and livestock systems studied in more than 50 research projects. An important aspect in bioinformatics is that each project is unique, and the LMB team works to meet these demands. This work in bioinformatics is based on the following areas: analysis of gene expression, assembly and analysis of genomes, identification of molecular markers, analysis of transcriptomes and metagenomes, evolution studies, modeling of biological systems, prediction of protein structures and molecular interaction, interaction or inhibition of molecules, among other activities.

LMB computational infrastructure to support bioinformatics projects applied to agriculture

Bioinformatics projects require a differentiated computational infrastructure, and most of them are very difficult or even impossible to carry out using common computational equipment. These requirements can be understood considering the computational complexity of the algorithms and the volume of biological data analyzed.

The objective of this session is to present the computational infrastructure used for storage and processing of large volumes of data produced by the biotechnology research projects of Embrapa and its partner institutions. This infrastructure focuses on making available processing and memory capacity as well as storage of large volume of data.

To deal with the various algorithms with high computational complexity in bioinformatics, it is standard to use computational clusters of computers for data processing. For those less familiar with the field of high-performance computing, a computer cluster is a set of computers connected in a network with a central coordination node that work together to solve computational problems. The main advantage of a cluster is to provide computing power of tens, hundreds and, in some extreme cases, thousands of processing nodes in a transparent way for the user, that is, without the user having to interact and trigger data analysis in each of the machines individually. The jobs to be executed in the system are activated from a management node that remain in one or more execution queues and are automatically sent to a suitable processing node, when available.

With the arrival of multicore computing, each processing node in modern clusters has a few dozen cores; in some exceptional situations, each node can reach hundreds of processing cores. Therefore, a very important question for processing in bioinformatics is: how much memory should each processing node have? The answer requires careful consideration regarding that as the amount of memory is directly proportional to the number of cores in the processing node. In addition, it should be considered that this proportion has increased with the development of new biological investigation techniques, which generate significantly increasing amounts of data. Thus, until recently, it was recommended that each processing node should have 8 Gb of RAM for each available core. With the significant increase in the volume generating biological data, this amount was updated, and new processing platforms for bioinformatics activities are being developed with 16 Gb of RAM for each available CPU core in the computing node.

Another relevant issue in biological data processing platforms is related to data storage and preservation. Basically, the most significant bottleneck that has to be addressed is the amount of data to be stored. The speed of data accessing does not significantly impact the performance of the platforms, as in general, the tools and programs executed to perform the analyses will load the data into memory and execute the analyses for a significant amount of time. A delay in the initial load does not considerably impact the total execution time of the task. However, a restriction on the storage capacity of the computing environment will have a wide range of negative occurrences. It is not possible to execute several projects at the same time, as they commonly demand a few hundred gigabytes, and can reach a few tens of terabytes for raw data storage for some exceptional projects. During the analyses, it is necessary to store intermediate data, possibly up to an order of magnitude of the original data size. Therefore, currently the platforms for processing biological data commonly use storage systems with capacity of a few petabytes.

The processing environment available today has a cluster with a head node and 14 processing nodes. Of these, 13 have 64 cores and 512 Gb of RAM each. There is also a special node that is used to perform jobs that require a large amount of memory. This node has 2 Tb of RAM and 160 processing cores. In total, the cluster provides 992 processing cores. For managing tasks in the cluster, a queue management system is used, initially developed by Sun Microsystems, known as Sun Grid Engine (SGE). For bioinformatics analyses, a high performance cluster is specially useful as the analyses, in general, involve multiple datasets to be processed in pipelines consisting of multiple stages, enabling the execution of computing tasks in parallel on separate machines. Computational analyses with such characteristics are ideal to be executed in clusters of computers.

The following data storage servers are available: an SGI Infinite storage with 150 Tb capacity in a RAID 6 configuration and an IBM DS3412 storage capable of storing 51 Tb in a RAID 5 configuration. In addition to primary storage, it is critical to have a backup policy that ensures data security on the platform. Due to the volume of data constantly received and generated, the most cost-effective methodology for backup involves the use of LTO tapes. Currently, the platform has a tape library available with capacity for 44 LTO6 drives. As each LTO6 tape provides, on average, 6.25 Tb of data storage, the total library is capable of handling up to 275 Tb of online backup.

This type of computational infrastructure is essential for carrying out data analysis of bioinformatics research projects in agriculture.

Applications

Bioinformatics and the tambaqui production chain

Embrapa's first strategic objective is "to develop knowledge and technologies for the adequate management and sustainable use of Brazilian biomes." Historically, Embrapa has always been concerned with regional development, actively performing on front lines where scientific or economic risks were discouraging factors for the private sector. This exceptional role played by Embrapa has guaranteed the use of the Cerrado biome for agriculture, bringing development and wealth to the region. The North region of Brazil has a fish production chain with an annual native fish production of 290 thousand tons, according to the 2019 Pisciculture Yearbook, and the main product is tambaqui (*Colossoma macropomum*). To promote the development of this important production chain, among other equally relevant objectives, Embrapa, through the BRS Aqua[1] project, identified critical points for increasing the production that, if properly resolved, would increase the competitiveness and sustainability of the tambaqui production chain.

One of the critical points identified by Embrapa in the tambaqui production chain[2] was the occurrence of crossbreeding between related matrices. Many fish farmers do not know this, but the simple choice of matrices for crossbreeding can, if wrongly done, reduce the final weight of fish by 10% to 30%. In other words, using the same amount of feed in the food, the producer could lose up to 30% of food conversion. In the scientific literature, this phenomenon is known as inbreeding depression, and few fish producers are aware of this. To measure the size of the problem, let us observe the case of native fish, which are highly appreciated in the North region. As mentioned, in 2019, the production was 290 thousand tons, assuming a conservative estimate, as the inbreeding of related matrices may have negatively impacted production by at least 30 thousand tons.

In addition to inbreeding depression, the inbreeding between related matrices causes yet another harmful phenomenon, scientifically known as lethal alleles. In any population, lethal alleles are rare; however, when they occur in homozygosis, they impair embryo development. That is, these alleles cause deformities in embryos or abort their development when inherited from both the father and the mother. Hence the recommendation to avoid consanguineous pairings. If these alleles are rare in the population as a whole, within families carrying these alleles the occurrence of homozygosity is significantly more frequent, reaching up to 25%. That is, in consanguineous breeding, up to 25% of embryos can be lost or have birth defects. Both inbreeding depression and lethal alleles are critical problems in the fish production chain.

In addition to inbreeding depression and lethal alleles, another critical point is the existence of fertile hybrids in the breeding stock. In biology classes, one learns that when two different species interbreed, the result is an infertile animal. Unfortunately, this is not always true for fish. For example, the tambaqui can crossbreed with the pacu (*Piaractus mesopotamicus*), and the hybrid is a fertile animal. However, many producers crossbreed tambaqui with pacu because the hybrids gain more weight than purebred animals and the flavor of the meat is not significantly affected. In the literature, this phenomenon is known as "hybrid vigor", and it is widely used in grain production, for example. The problem occurs when hybrids are wrongly chosen to compose the breeding stock. While this choice may seem unlikely at first, it occurs because selection is often based on external characteristics, and because of hybrid vigor it is not uncommon for a hybrid to be wrongly selected because it weighs more, for example. In this case, as the hybrids are fertile, the error of this choice will only be discovered during crossbreeding, when the producer observes the natural segregation that entails a great deal of variability in the economic interest characteristics, such as slaughter weight. Producers who sell fingerlings for fattening may have their credibility affected by selling low quality animals, as the segregation variability greatly affects fattening.

[1] The BRS Aqua project is financed by the BNDES/Funtec Technological Fund, the Fisheries and Aquaculture Secretariat (SAP) of the Ministry of Agriculture, Livestock and Supply (MAPA), CNPq, FAPDF and Embrapa. In this part of the BRS Aqua project, the following Units largely participated: Embrapa Genetic Resources and Biotechnology, Embrapa Fisheries and Aquaculture and Embrapa Agricultural Informatics.

[2] This critical point occurs in all fish production chains where it is not possible to identify the relationship between the matrices.

Once these problems were identified, Embrapa researchers developed two DNA chips that solve such issues in an innovative, efficient and low-cost way. These chips have molecular markers, known as single nucleotide polymorphisms, or simply SNPs, that can provide enough information to determine the degree of relatedness and purity of the species. In the case of kinship, the markers must have considerable variability in the population studied. Mathematically, this means requiring the Minor Allele Frequency (MAF) to be close to 0.5. The principle is exactly the same as a paternity test, except that this application can identify any degree of kinship to avoid inbreeding, reducing inbreeding depression and minimizing the occurrence of lethal alleles. The scientific challenge is to precisely choose such SNPs molecular markers. In the case of tambaqui, the lack of a publicly available reference genome was the first obstacle to be overcome. This led Embrapa to carry out an internal Tambaqui Genome Project, and the LMB was responsible for assembling the Tambaqui Genome. The genome contains approximately 1.3 billion nucleotides divided into 27 chromosomes (or linkage groups). Once the genome was ready, the next step was to select a representative subpopulation of the tambaqui population and then sequence the DNA from the pool of that subpopulation. The result of this sequencing was mapped to the reference genome, and finally the discovery of SNPs was carried out. Although a minimum coverage of 150X was required, more than 2 million SNPs were identified (Ianella et al., 2019). The task of selecting 96 SNPs to compose the kinship chip took into account the MAF, the spacing within the chromosomes, the functional annotation and, finally, the absence of genomic variations in the flanking regions of the candidate SNP. As noted, the bioinformatics work was very intense in order to carry out all these tasks, which justifies the need for an infrastructure like the LMB. After the validation phase of the SNPs in a different population from that used in the previous phase, the validated SNPs were incorporated into the chip, which proved to be extremely efficient in determining the degree of relatedness and is currently being used in the tambaqui production chain. In other words, the producer already has an innovative tool to eliminate inbreeding depression and lethal alleles, thus avoiding silent damage caused by inbreeding.

The DNA chip for purity determination, however, required more complex analyses. This is because it was necessary to include in the analysis two more species that crossbreed with tambaqui and produce fertile hybrids, namely, the pacu and the caranha (*Piaractus brachypomus*). As none of these species has a reference genome, it was necessary to use the tambaqui genome as a reference. This procedure is important because, in addition to intraspecies variations, there are also interspecies variations (tambaqui x pacu / tambaqui x caranha), which increases the degree of complexity of the analyses. Even at the stage of mapping the reads in the reference genome, the similarity requirement had to be reduced due to interspecific differences. Differently from relatedness SNPs, SNPs to measure the purity of the species must be "fixed", that is, they must not show species variation, that is, MAF = 0.

An example can help to better understand the problem. If at a specific position in the genome there is an "A" nucleotide fixed on the tambaqui, and in that same position there is the "C" nucleotide fixed on the pacu, then this genome position is a serious candidate to compose the purity chip of species, because, in a DNA test, an "A" result would mean "tambaqui" and a "C" would mean pacu. And to further reduce costs, genomic markers capable of simultaneously separating the tambaqui from the other two species were explored. In the previous example, this would mean that the caranha also had a "C" fixed to that same genomic position[3]. Thus, with a single DNA chip, it is possible to assess the purity of tambaqui in relation to the two main species that produce hybrids[4]. Once again, using allelic frequency, physical spacing in the genome and functional annotation, 96 SNPs were selected to compose the chip, and after the validation

[3] SNPs are biallelic markers, which enable separating one species from two others simultaneously. There are triallelic SNPs, but they are very rare, and therefore it is not possible to produce a single genotyping chip that separates the three species two by two simultaneously.

[4] From what has already been shown, this purity chip does not separate pacu from caranha.

phase in independent populations, the validated SNPs were incorporated into the purity measurement chip. This genomic tool enables us to eliminate all the hybrids that were wrongly chosen to compose the breeding stock.

Economic impact studies carried out by Embrapa, assuming an average production of 150 thousand tons of tambaqui, forecast additional gains between US$ 1,8 million and US$ 5,5 million for producers[5]. Each sample analysis for purity and relatedness currently costs US$ 12.00. For a producer with 100 matrices, this would be equivalent to an investment of US$ 2,4 thousand. As each matrix has a useful life of three years, this amount is amortized over an equal period. These two technologies, named TambaPlus[6], have already been adopted by producers in five states: Mato Grosso, Tocantins, Roraima, Amazonas and Rondônia, and more than 1,500 tests have already been performed. The TambaPlus is so important that the technology was selected to compose a select group of technologies that were highlighted at the 47th Anniversary of Embrapa[7].

Research on the tambaqui production chain will continue. There is still plenty of room to improve fish production. In any genetic improvement program, there are two main phases, namely, the Selection and the Crossing phase. Tambaqui is still at an earlier stage, known as pre-breeding. The main concern was to first avoid inbreeding and the presence of hybrids in the breeding stock.

Bioinformatics in vaccine development: reverse vaccinology

In animal production, the use of vaccines is an effective and low-cost alternative for preventing or reducing the severity of diseases that affect livestock. Vaccination contributes to maintaining animal health and welfare, increasing the efficiency of food production and reducing the transmission of zoonoses. Compared to other forms of control, for instance the use of antibiotics and pesticides, vaccines have advantages, such as the non-contamination of the environment and animal products (meat, milk and eggs).

Following conventional vaccine development methodology, the pathogen is cultivated in vitro in the laboratory and used in its attenuated form (in which it loses the ability to cause disease) or killed to elicit a protective immune response in the host. Alternatively, purified components of the pathogen can also be used as antigens in the subunit vaccines (Rappuoli; Covacci, 2003).

Although conventional obtained vaccines are among humanity's most important inventions, which comprise a powerful tool in the fight against disease-causing biological agents, not all pathogens can be cultivated *in vitro* and used in the development of vaccines, in its conventional form. In addition, conventional methods are quite time-consuming, and it may take five to 15 years to obtain an effective vaccine (Vernikos, 2008).

Reverse vaccinology, a methodology first published by Rappuoli (2000), emerged as an alternative strategy for the discovery of protective antigens for developing vaccines based on the analysis of the target pathogen's genome. Made possible by large-scale gene sequencing, along with the development

---

[5] News provided in a video conference entitled TambaPlus®: Genomic tools for the analysis and management of tambaqui matrices for the production of fingerlings, available on the Agrotins platform. Available at: https://agrotins.to.gov.br/programacao/tambaplus-ferramentas-genomicas-para-analise- e-gestao-de-matrizes-de-tambaqui-destinadas-a-produca.html.

[6] Available at: https://www.embrapa.br/busca-de-noticias/-/noticia/46203188/ferramentas-genomicas-ajudarao-a-evitar-cruzamentos-consanguineos-entre-matrizes-de-tambaqui

[7] Available at: https://www.embrapa.br/47-anos/solucoes-tecnologicas-em-destaque?link=47-anos

of bioinformatics tools, reverse vaccinology uses *in silico* prediction tools to identify targets (antigens) for developing vaccines. Through these tools, genomes, transcriptomes and proteomes are examined *in silico*, predicted proteins are selected based on desirable attributes – which can induce an immune response capable of protecting against a given disease, and the targets are then identified. Based on them, different types of vaccines can be designed and developed within an interval of 1 to 2 years.

Commercial vaccines obtained through this methodology are already a reality. A vaccine developed against invasive meningococcal disease, caused by the bacterium *Neisseria meningitidis* serogroup B, was released for use in Europe in 2014 (Andrews; Pollard, 2014). In this vaccine, the immune response is triggered by epitopes – specific sequences of amino acid residues present in the antigen that directly participate in the interaction with antibodies, which were identified using bioinformatics tools. Epitopes have been considered particularly interesting in vaccine development, as it has been shown that vaccines composed of these peptides can optimize or exceed the protection potential induced by the cognate native protein (Kao; Hodges, 2009). In contrast to live attenuated vaccines, a vaccine containing a synthetic epitope is not able to reverse the virulence of a pathogen (Palatnik-De-Sousa et al., 2018). Furthermore, epitope-based vaccines are more specific, do not induce undesirable immune responses, are capable of generating long-lasting immunity and are less expensive than conventional vaccines (Ahmad et al., 2016).

In the reverse vaccinology approach, the protein sequences of an organism are analyzed using *in silico* prediction programs. These proteins, however, are mostly predicted from the sequencing of genomes and transcriptomes, using bioinformatics tools. This is because large-scale genetic sequencing, made possible by new technologies that have dramatically reduced the cost of generating sequences, as well as exponentially increasing the number of sequences generated from a sample, has accumulated an unprecedented amount of genomic and transcriptomic data. On the other hand, a technological advance that would allow a large-scale development of protein sequencing techniques with high sensitivity has not yet taken place. Methodological progress for obtaining expressed gene sequences caused the subsequent evolution in analysis methodologies. A list of programs can be accessed on the "List of RNA-Seq bioinformatics tools" page (Wikipedia, 2020). We will now provide a brief commented description of the methodology applied to obtain differentially expressed genes in the salivary gland of the bovine tick (Andreotti et al., 2018). All tools cited are obtained through an academic license or government research institution or are freely distributed.

In order to better understand the host-parasite interaction and identify possible genes and mechanisms involved, a study initiated in 2015, funded by Embrapa, generated more than 600 million sequences from RNA sequencing (using the RNA-Seq methodology) from larvae, nymphs, salivary gland, intestine and ovaries of the cattle tick, *Rhipicephalus (Boophilus) microplus* (Andreotti et al., 2018).

In addition to the characterization of transcriptomes from different tissues through *de novo* assembly, our research group also identified the differentially expressed genes (DEG) between ticks grown in resistant cattle (Nellore), susceptible cattle (Holstein) and crossbred animals with intermediate resistance to the parasite (Nellore x Holstein). The analysis of this dataset, using tools that inform the function of proteins predicted from DEG and the biological pathways in which they act, brought new discoveries about the cattle tick interaction and pointed out potential candidates that can be used as antigens in the development of vaccines to control the cattle tick (Giachetto et al., 2020).

The first step in RNA-Seq analysis is to check the quality of the generated sequences. Tools like FastX Toolkit (FastX-GitHub, 2020) and FastQC (FastQC-GitHub, 2020) check several parameters, highlighting the following:

- Average quality of bases and average quality per sequence. For a good result, the sequence must have a "Phred score" greater than 30.

- GC content (%GC). The percentage of the presence of Guanine and Cytosine nucleotide bases in the sequence must be close to the normal distribution, since the very high GC content prevents the synthesis and, often, the clustering of sequences during the acquisition and assembly processes.
- Number of indeterminate bases (%N). Indeterminate bases make the contingency process difficult. They can occur at the beginning of the sequencing, where there is a saturation of reagents; in the end, by decreasing the concentration of reagents; or in a region with high %GC, which hinders reading the region by the polymerase.
- Presence of adapters. Adapters are short nucleotide sequences used for library preparation and sequencing. Its presence impairs contingency, giving rise to chimeric sequences. To eliminate them, tools such as Trimmomatic (Bolger et al., 2014) and Trim Galore (TrimGalore-GitHub, 2020) are often used.

As we deal with a large number of sequences, a great tool to group and visualize the data obtained in the quality analysis (and subsequent steps) is the MultiQC (Ewels et al., 2016), which organizes the results obtained in a Web page.

After verifying the quality of the sequences, we proceeded to obtain the transcriptome, through sequence-to-sequence comparison and their contingency by similarity. Several tools can be used in this step, such as QUAST (Gurevich et al., 2013), which is recommended for the analysis of metagenomes. The tool of choice for analyzing this work was the Trinity program (Grabherr et al., 2011). This tool is, in fact, a pipeline that brings together, through scripts developed in the programming languages Perl[8] and Python[9], various analysis tools for quality, sequence contingency and statistics to identify DEGs, with the differential of being able to identify isoforms (the same as transcribed) of the same gene, resulting from alternative splicing. Different tissues can express different isoforms in different amounts. Identifying the locally expressed isoform allows to better understand the expression of a particular gene in a particular metabolic pathway or tissue.

Once the transcriptome is obtained, the next step is to verify the quality of the assembly. An initial approach is to map the sequences used for assembly back to the transcriptome obtained. In a good setup, more than 80% of the sequences map the transcriptome. A second assessment consists of identifying and quantifying complete sequences, through similarity analysis against curated databases, such as SwissProt or TrEMBL (The UniProt Consortium, 2019), or searching for orthologs present in the closest classification of the studied organism, in this case, the arthropods, using the BUSCO software (Seppey et al., 2019).

Several factors influence the experimental design of an RNA-Seq assay for the identification of DEGs:

- In the preparation of samples, from extracting total RNA to obtaining libraries for sequencing, the batch effect may occur, in which they are included from using different solutions (made on different days) to the person who prepares them (Conesa et al., 2016).
- The sequencing depth (the number of generated sequences), which influences the number of sequences obtained and, therefore, the quantification of the number of identified DEGs (Conesa et al., 2016; Lamarre et al., 2018).

[8] Available at https://www.perl.org

[9] Available at: https://www.python.org

- The number of technical replications (how many times the same sample is sequenced), which influences the statistical power for detecting DGEs, recommending no less than three repetitions (Conesa et al., 2016), although a higher number (about six repeats) can increase the representativity of transcriptome sequences (Lamarre et al., 2018). An assay with triplicates is commonly accepted, as the increase in replicates implies an increase in assay costs.
- The preparation of a biological repetition. Conesa et al. (2016) point out that biological variability is particular to each assay, and although difficult to control, it is important for a study involving populations, suggesting that the biological sample be done in triplicate. Lamarre et al. (2018) point to the detection of up to 20% of DEGs due to biological variability, which may not justify raising the costs of the assay.

The correlation between the samples used in the assay is also an important measure of the quality of the assembly and the libraries constructed. The principal component analysis allows visualizing correlations between technical and biological replicates, which should preferably form not too distant clusters. A discrepancy between samples from the same group may indicate contamination, sample mixture, sequencing error or batch effects, which must be considered for discarding that sample. Also important is the fact that without a technical triplicate, a biological duplicate must be discarded, impairing the entire analysis.

With a good quality transcriptome, the differentially expressed sequences were identified. Trinity incorporates several statistical tools for this purpose. In this case, we chose to use RSEM (Li; Dewey, 2011), which estimates the quantity of each transcript by realigning the sequences of each library (or experimental treatment) to the generated transcriptome – a reason for the importance of quality and the relationship between replicates - and edgeR (Robinson et al., 2010), a package developed in the statistical program R (R Core Team, 2020) and part of the Bioconductor Project (Huber et al., 2015) for the analysis of biological data, which performs the pairwise comparison sequences generated between all samples and identifies those with differential expression.

The penultimate step is the annotation (or identification) of each differentially expressed sequence, through similarity analysis in nucleotide and protein sequence databases, looking for homology to already known sequences, and in databases of metabolic pathways that inform in which one the gene participates. This is followed by a manual analysis of each result, the bibliographic basis seeking the role of such a gene in the development of the tick's life cycle, and the selection of possible targets for vaccine development.

The existence of commercial vaccines available for the control of bovine ticks demonstrates they can act effectively in the control of infestations, reducing the application of acaricides. However, the adoption of these vaccines has been limited, mainly because they are not effective against all life stages of the parasite, in addition to their low efficacy against some regional strains of *R.* (*B.*) *microplus* (Andreotti, 2006). The results obtained in a test conducted by Embrapa with a regional tick isolate showed an efficacy of 46.4% and 49.2%, respectively, for the TickGARD® and GavacTM vaccines (Andreotti, 2006). Thus, based on the database described above, our team is currently coordinating a study that foresees the identification of candidate immunogenic epitopes for the development of vaccines against cattle ticks, using the methodology of reverse vaccinology, based on the predicted proteins of the transcriptomes of the parasite. Executing a pipeline containing a series of analysis tools, candidate target genes for vaccine production are analyzed for the presence of epitopes that can interact with the bovine immune system for the production of antibodies, helping to fight to tick infestation.

A highly effective vaccine, integrated in cattle tick control strategies, can considerably reduce herd infestations and the implications related to the use of acaricides, which include, in addition to cost and environmental contamination, a growing concern with food safety, which has increasingly led to the consumption of food free of chemical residues, obtained from sustainable production systems. Also, by validating the pipeline we are proposing, the LMB will be able to apply the reverse vaccinology methodology in the identification of targets for the control of other problems of interest in agriculture.

Bioinformatics tools

As advocated by digital agriculture, to be transformed into useful knowledge, information generated from biological experiments must be accessible and, when possible, made available on the Internet. Bioinformatics and computational biologists have been dealing with this scenario for more than a decade in an environment with adequate infrastructure like the one described above, and have implemented software libraries, toolkits, platforms and databases to achieve success in this matter.

Several data analysis tools are used in Embrapa's LMB, and a search for a data integration solution became necessary. Analysis results are carefully stored in a directory structure and reports are generated. Some tools generate results in a format already available for the Internet or they can be executed directly online. Two tools under development have greatly contributed to the integration of the generated data and the transformation of these data into information.

Machado: a genomic data integration framework

In 2017, the PlantAnnot project was started to discover candidate proteins to use in pipelines for the development of transgenic plants (Prado et al., 2014; Napier et al., 2019) that are resistant to abiotic stresses. It aimed to develop a bioinformatics system applied to the discovery of genes related to abiotic stresses in plants, focused on climate change. In this project, a large volume of genomic data was extracted from public databases. The extracted dataset corresponds to 53 plant genomes, totaling more than 1.8 million genes and more than 2.3 million proteins. These data were used to perform computational analyses in order to select 72,000 proteins of interest for the pipelines. One of the project goals was to store and make available the data and analyses performed.

To solve this problem, the Machado open-source software was developed. Machado is a genomic data integration framework written in Python[10] that allows research groups to store genomic data, and also offers interfaces for navigation, searches, and visualization. Machado uses the BioPython library (Cock et al., 2009) which supports the vast majority of file formats and programs used in bioinformatics. In addition, Python has become one of the main programming languages in the data sciences area (Millman; Aivazis, 2011), and Machado can also benefit from the tools in this area. This framework uses the Chado database schema and therefore should be very intuitive for current developers to adopt or execute Machado on existing databases.

GMOD's biological relational database schema, Generic Model Organism Database Project[11], known as Chado (Mungall; Emmert, 2007), is one of the few open-source initiatives that has achieved relative success in the community. Many software systems can connect to it, such as Gbrowse (Stein et al., 2002), Jbrowse (Skinner et al., 2009) and Apollo (Lee et al., 2013), which are important tools for visualization and

[10] Available at: https://www.python.org

[11] Available at: http://www.gmod.org

annotation of genomes. There are some data integration tools that use Chado as a database schema or that can extract data from that schema, but they were developed in programming languages not often used in bioinformatics (Kalderimis et al., 2014; Spoor et al., 2019).

Machado has several data loading tools for genomic data and for analysis results from known software in the biological environment (BLAST, InterproScan etc.) (Altschul et al., 1990; Quevillon et al., 2005), and its web interface contains a powerful search tool that allows users to quickly filter and sort the results.

Within the scope of the PlantAnnot project, a tool called Plant Co-expression Annotation Resource was created using Machado to store and make available the data of the project[12]. This tool is an implementation of Machado, which is an example of its usefulness for researchers who need to store and make accessible a large volume of genomic data.

As an example, one of the uses of the Plant Co-expression Annotation Resource is to enable navigation through the genome of 53 species of angiosperm plants, which allows visualizing details about genes, proteins and RNA through the JBrowse genome browser. This tool is also used to perform keyword searches and use filters. The user can then perform simple searches for genes, proteins and RNA, using keywords of interest. Furthermore, it can also add more complex filters to the search results, producing more specific result lists. For example, a set of proteins with no known function, candidates for the creation of transgenic plants resistant to abiotic stresses, such as drought, heat, cold, among others.

Machado is meant to be a modern object-relational framework that uses the latest Python modules to produce an effective open-source program for genomic research and can be an engaging project for new developers, contributors and users. Thus, we created a corporate account for LMB on GitHub, which we believe is the first Embrapa account on this platform[13]. A demo version of the system was also created[14].

Machado will undergo improvement phases for ongoing projects at Embrapa, such as the project "The Hologenome of Nelore: Implications for Meat Quality and Food Efficiency" which is focused on genomic improvement of cattle, led by Embrapa Southeast Livestock. This project intends to identify molecular mechanisms related to meat tenderness, therefore, several data sets that need to be integrated were produced, such as genomes, transcriptomes, proteomes, genotyping, among others.

DBGAP: web system for retrieving information on pedigree, phenotypes and genotypes

The development of large-scale genotyping technologies of molecular markers such as the Single Nucleotide Polymorphisms (SNP) – to estimate the genomic profile of animals – has enabled developing genome-wide association studies – GWAS at a genomic scale, as well as the introduction of genomic selection technology in genetic improvement programs. Current technologies for generating molecular data are capable of genotyping tens to hundreds of thousands of SNP markers, in a single assay for each individual, with enormous speed and automation (Caetano, 2009).

On the other hand, this situation implies the need to store an enormous volume of data, not only of genotypes, but also the phenotypes and pedigree of an increasing number of animals. Therefore, performing the proper storage and extracting useful knowledge from this amount of data is a major challenge. Given the volume of data stored, an important matter to consider when developing a computational solution is the suitability of database modeling for the desired application, as this will

[12] Available at: https://www.machado.cnptia.embrapa.br/plantannot

[13] Available at: https://github.com/lmb-embrapa

[14] Available at: https://www.machado.cnptia.embrapa.br/demo_machado

directly impact the query and writing times in relational database management systems (RDBMS) where this information will be stored.

Therefore, in order to provide a solution that would be efficient both in storage and in the integration and querying of this high volume of data, the Database of Pedigree, Phenotypes and Genotypes (DBGAP) system was developed. The purpose of this system is to integrate the data sent in various formats, so they can be analyzed in genetic/genomic evaluation software. The DBGAP was initially developed using a data diagram proposed by Higa and Oliveira (2015). This diagram was redesigned in such a way as to allow implementing the JavaScript Object Notation (JSON) type. With the implementation of JSON and text types in some tables, it was possible to use the Not Only SQL[15] (NoSQL) approach to store part of the data, streamlining queries that would need to perform joins with other tables.

To develop the system, information technology components were chosen within the philosophy of using free software. The database management system chosen was PostgreSQL[16], as it is a reliable DBMS, widely used in the market. The version control software, GitLab[17], hosted at Embrapa, was used. The programming language chosen was Java[18] and its components of the Java Enterprise Edition (Java EE) technology.

Among the Java EE technologies available and used by DBGAP, the Java Server Faces (JSF) framework stands out. The architecture of the JSF framework employs the MVC model (Model, View, Controller), which separates the presentation and application layers. The application server chosen to host the DBGAP system was WildFly[19].

The system development project used some concepts from Scrum, which is an agile framework to perform complex projects. Scrum combines monitoring and feedback activities, generally through quick and daily meetings with the entire team, in order to identify and correct any deficiencies in the development process. In addition, the Scrum method is based on fundamentals such as: small teams, unknown requirements and short iterations, these are called sprints (Schwaber, 2004).

The DBGAP system has many features implemented and is currently in the process of user's approval. Through its web interface, it is possible to query and import phenotypic, genotypic and pedigree data of various animal species. When accessing it, the login page will be displayed (Figure 1):

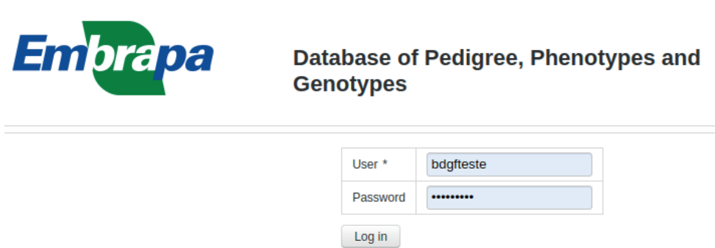


Figure 1. BDPFG20 system login screen.
Available at: http://www.dbGaP.cnptia.embrapa.br

[15] Available at http://nosql-database.org

[16] Available at: https://www.postgresql.org

[17] Available at: https://gitlab.com

[18] Available at: https://www.oracle.com/br/java

[19] Available at: http://wildfly.org/downloads

One of its important features concerns the visualization of animal data (Figure 2). On this screen, the user finds various information about the individual, such as individual identifier code, original name, father, mother, date of inclusion in the population, population and other information contained in the JSON variables related to the type of individual (beef cattle, poultry, etc.). However, it is important that the variables of the phenotypes related to the species considered by the system must be previously registered, which are imported from the Embrapa Experiments System - SIEXP (Apolinário et al., 2016), where they were defined for the species the user will work with in the user group (e.g., beef cattle, poultry, etc.).

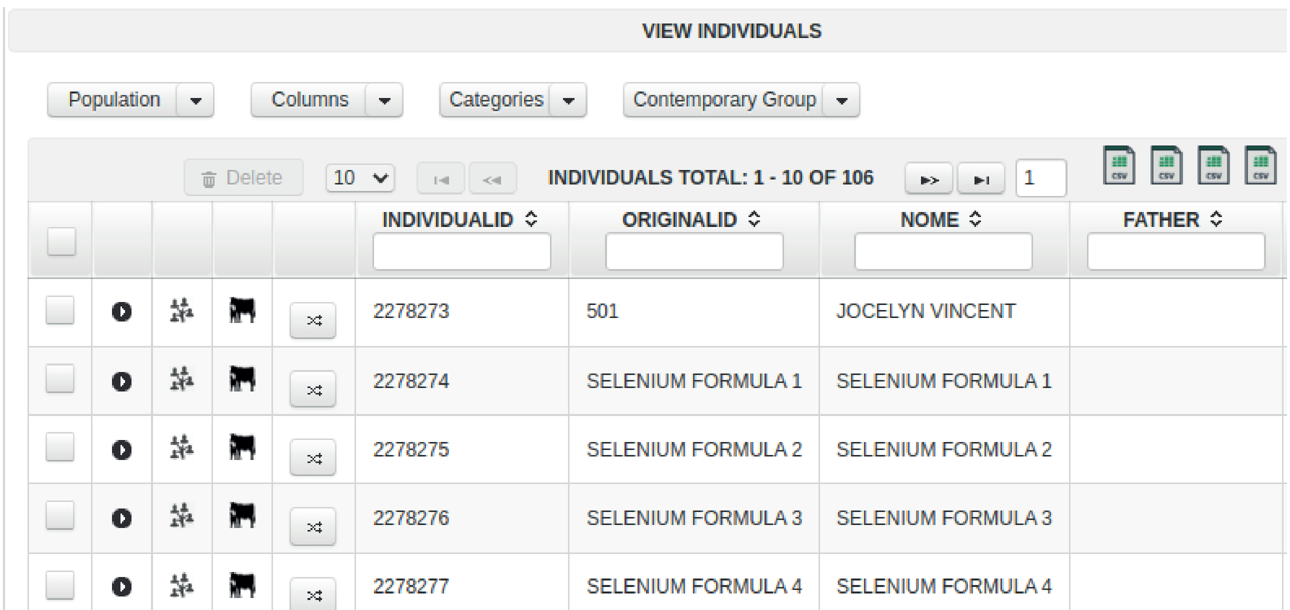


Figure 2. Screen showing registered individuals in the system.
Available at: http://www.dbGaP.cnptia.embrapa.br

One can also import data from files with columns separated by tabs (TSV). These files must follow a standardized format. After importing the data, the pedigree of an animal listed on the animal view page can be viewed. The pedigree window can be expanded to facilitate viewing the animals and their relatives.

The database provides several filters so that the user can check the data that has been uploaded and then export it to the format of the evaluation software. Generally, the data are exported in tabular format to be analyzed in the R program, as they are extensive tables with measurements of animal characteristics. It is also possible to export the data of these animals (phenotypes, pedigree) to files in CSV format and operate them in Excel. Existing filters allow queries by population, category, animal name, father's name, mother's name. Another tool, perhaps the most important in the system, is the one for identification of duplicated animals, allowing the user to associate duplicated animals in a single animal.

The DBGAP system is part of a computational solution proposed in other Embrapa projects (MaxiDep and MaxiPlat). The goal of these projects was to combine efforts to structure a computing solution (of which DBGAP is one of the components) to support routine genetic evaluation of beef cattle breeding programs, within the scope of the Embrapa-Geneplus program. This effort included both the development of assets to support the organization of data used in genetic evaluations (DBGAP system) and the development of a national solution for the resolution of genetic-statistical models (brBlup software).

A comparison by a search in other software systems with web interface developed by Embrapa Digital Agriculture (Vieira, 2012a, 2012b), with functionality to store genotypes and phenotypes and that includes basic queries to molecular data (SNPs), shows that a simple query in about 800 animals and 700 thousand SNP markers took at least an hour to be processed in the other software systems developed. A similar query performed in the DBGAP database takes less than a minute, as using the JSON type fields and text in the tables removes part of the necessary normalization of the traditional model, speeding up the searches.

Final considerations

The research reported in this chapter is ongoing and will continue to other stages. Regarding the research on tambaqui, with the progress of production in the near future, it will be possible to start the genetic improvement itself. The genomic tools presented in this chapter may evolve to assist in the matrix selection stage, focused on improving some characteristic of economic interest, like for instance, the slaughter weight. In the beef production chain, genomic selection is already a reality, and the results are excellent. The same can occur with the fish production chain. With the growing status of fish protein on the world menu, perhaps the Amazon region may soon become a major producer and, possibly, even an exporter of native fish. There is still a long way to go, but Embrapa has already made a significant contribution by showing and opening the way, and bioinformatics plays a fundamental role.

Validating a methodology that includes the identification of antigens through a reverse vaccinology pipeline and obtaining a multi-epitope vaccine is underway at Embrapa, with the participation of the LMB, which is aimed at controlling the bovine tick. The infestation of cattle herds by this parasite is considered today one of the most significant problems in livestock farming in economic terms, affecting all countries with tropical and subtropical climates.

In Brazil alone, annual losses due to tick infestation are in the order of US$3.24 billion (Grisi et al., 2014). Obtaining an effective vaccine will certainly contribute towards controlling the parasite, reducing the application of acaricides, as well as the environmental and economic damage resulting from this practice. Moreover, once validated, there are several possible applications of the methodology, including the identification of targets for the control of other problems of interest to agriculture involving animal health and welfare.

Machado tool will assist other ongoing projects at Embrapa. There is already a program for its use in the Genomics Applied to the Optimization of Genetic Improvement Programs for Tropical Forage Species, led by Embrapa Cerrados, with a focus on forage plant improvement. This project predicts the sequencing of reference genomes for six tropical forage species, with the characterization of broad sets of genomic variants, and Machado will probably be used as a basis for the implementation of a portal to access the generated genomic data.

DBGAP database is being structured to allow its use in other data collections, with some specific changes for each project.

As shown in the research reported here, bioinformatics has become fundamental and will be even more important in the innovation agendas towards the digital transformation of agriculture. The existence of multi-user structures is crucial to support research projects that do not have the necessary structure for complex analyses, allowing better use of the resources. With bioinformatics relying on the availability of a specialist team and adequate infrastructure, the management of the structure that supports research projects must be focused on keeping both aspects up to date.

References

AHMAD, T. A.; EWEIDA, A. E.; SHEWEITA, S. A. B-cell epitope mapping for the design of vaccines and effective diagnostics. **Trials in Vaccinology**, v. 5, p. 71-83, 2016. DOI: 10.1016/j.trivac.2016.04.003.

ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403-410, Oct. 1990. DOI: 10.1016/S0022-2836(05)80360-2.

ANDREOTTI, R. Performance of two Bm86 antigen vaccine formulation against tick using crossbreed bovines in stall test. **Revista Brasileira de Parasitologia Veterinária**, v.15, p. 97-100, 2006.

ANDREOTTI, R.; GIACHETTO, P. F.; CUNHA, R. C. Advances in tick vaccinology in Brazil: from gene expression to immunoprotection. **Frontiers in Biosciences**, v. 10, p. 127-42, Jan. 2018. DOI: 10.2741/s504.

ANDREWS, S. M.; POLLARD, A. J. A vaccine against serogroup B Neisseria meningitidis: dealing with uncertainty. **The Lancet Infectious Diseases**, v. 14, n. 5, p. 426-434, May 2014. DOI: 10.1016/s1473-3099(13)70341-4.

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.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114-2120, Aug. 2014. DOI: 10.1093/bioinformatics/btu170.

CAETANO, A. R. Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. **Revista Brasileira de Zootecnia**, v. 38, p. 64-71, 2009. Número especial. DOI: 10.1590/s1516-35982009001300008.

COCK, P. J. A.; ANTAO, T.; CHANG, J. T.; CHAPMAN, B. A.; COX, C. J.; DALKE, A.; FRIEDBERG, I.; HAMELRYCK, T.; KAUFF, F.; WILCZYNSKI, B.; DE HOON, M. J. Biopython: freely available Python tools for computational molecular biology and bioinformatics. **Bioinformatics**, v. 25, n. 11, p. 1422-1423, June 2009. DOI: 10.1093/bioinformatics/btp163.

CONESA, A.; MADRIGAL, P.; TARAZONA, S.; GOMEZ-CABRERO, D.; CERVERA, A.; MCPHERSON, A.; SZCZEŚNIAK, M. W.; GAFFNEY, D. J.; ELO, L. L.; ZHANG, X.; MORTAZAVI, A. A survey of best practices for RNA-seq data analysis. **Genome Biology**, v. 17, article number 13, 2016. DOI: 10.1186/s13059-016-0881-8.

EWELS, P.; MAGNUSSON, M.; LUNDIN, S.; KÄLLER, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. **Bioinformatics**, v. 32, n. 19, p. 3047-3048, Oct. 2016. DOI: 10.1093/bioinformatics/btw354.

FAStQC-GitHub. Available at: https://github.com/s-andrews/FastQC/releases. Accessed on: 7 May 2020.

FAStX-Github. Available at: https://github.com/agordon/fastx_toolkit. Accessed on: 7 May 2020.

GIACHETTO, P. F.; CUNHA, R. C.; NHANI JUNIOR, A.; GARCIA, M. V.; FERRO, J. A.; ANDREOTTI, R. Gene expression in the salivary gland of Rhipicephalus (Boophilus) microplus fed on tick-susceptible and tick-resistant hosts. **Frontiers in Cellular and Infection Microbiology**, v. 9, p. 477, Jan. 2020. DOI: 10.3389/fcimb.2019.00477.

GRABHERR, M. G.; HAAS, B. J.; YASSOUR, M.; LEVIN, J. Z.; THOMPSON, D. A.; AMIT, I.; ADICONIS, X.; FAN, L.; RAYCHOWDHURY, R.; ZENG, Q.; CHEN, Z.; MAUCELI, E.; HACOHEN, N.; GNIRKE, A.; RHIND, N.; DI PALMA, F.; BIRREN, B. W.; NUSBAUM, C.; LINDBLAD-TOH, K.; FRIEDMAN, N.; REGEV, A. Full-length transcriptome assembly from RNA-seq data without a reference genome. **Nature Biotechnology**, v. 29, n. 7, p. 644-652, 2011. DOI: 10.1038/nbt.1883.

GRISI, L.; LEITE, R. C.; MARTINS, J. R. de S.; BARROS, A. T. M. de; ANDREOTTI, R.; CANÇADO, P. H. D.; LEÓN, A. A. P. de; PEREIRA, J. B.; VILLELA, H. S. Reassessment of the potential economic impact of cattle parasites in Brazil. **Revista Brasileira de Parasitologia Veterinária**, v. 23, n. 2, p. 150-156, Apr./June 2014. DOI: 10.1590/S1984-29612014042.

GUREVICH, A.; SAVELIEV, V.; VYAHHI, N.; TESLER, G. QUAST: quality assessment tool for genome assemblies. **Bioinformatics**, v. 29, n. 8, p. 1072-1075, Apr. 2013. DOI: 10.1093/bioinformatics/btt086.

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 Informática Agropecuária, 2015. 30 p. (Embrapa Informática Agropecuária. Documentos, 133). Available at: https://ainfo.cnptia. embrapa.br/digital/bitstream/item/138127/1/Doc133.pdf. Accessed on: 7 May 2020.

HUBER, W.; CAREY, V. J.; GENTLEMAN, R.; ANDERS, S.; CARLSON, M.; CARVALHO, B. S.; BRAVO, H. C.; DAVIS, S.; GATTO L.; GIRKE, T.; GOTTARDO, R.; HAHNE, F.; HANSEN, KD.; IRIZARRY, R. A.; LAWRENCE, M.; LOVE, M. I.; MACDONALD, J.; OBENCHAIN, V.; OLE'S; A. K.; PAG'ES, H.; REYES, A.; SHANNON, P.; SMYTH, G. K; TENENBAUM, D.; WALDRON, L.; MORGAN, M. Orchestrating high-throughput genomic analysis with Bioconductor. **Nature Methods**, v. 12, n. 2, p. 115-121, Jan. 2015. DOI: 10.1038/nmeth.3252.

IANELLA, P.; YAMAGISHI, M. E. B.; VARELA, E. S.; VILLELA, L. C. V.; PAIVA, S. R.; CAETANO, A. R. Tambaqui (Colossoma macropomum) single nucleotide polymorphism discovery by reduced representation library deep sequencing. In: AQUACULTURE, 2019, New Orleans. **Abstracts**. [S.l.: s.n.], 2019. p. 491. Available at: https://ainfo.cnptia.embrapa.br/digital/ bitstream/item/208846/1/CNPASA-2019-Aqua2.pdf. Accessed on: 7 May 2020.

KALDERIMIS, A.; LYNE, R.; BUTANO, D.; CONTRINO, S.; LYNE, M.; HEIMBACH, J.; HU, F.; SMITH, R.; ŠTĚPÁN, R.; SULLIVAN, J.; MICKLEM, G. InterMine: extensive web services for modern biology. **Nucleic Acids Research**, v. 42, n. W1, p. W468-W472, July 2014. DOI: 10.1093/nar/gku301.

KAO, D. J.; HODGES, R. S. Advantages of a synthetic peptide immunogen over a protein immunogen in the development of an anti-pilus vaccine for Pseudomonas aeruginosa. **Chemical Biology & Drug Design**, v. 74, p. 33-42, 2009. DOI: 10.1111/j.1747-0285.2009.00825.x.

LAMARRE, S.; FRASSE, P.; ZOUINE, M.; LABOURDETTE, D.; SAINDERICHIN, E.; HU, G.; LE BERRE-ANTON, V.; BOUZAYEN, M.; MAZA, E. Optimization of an RNA-Seq differential gene expression analysis depending on biological replicate number and library size. **Frontiers in Plant Science**, v. 9, article 108, Feb. 2018. DOI: 10.3389/fpls.2018.00108.

LEE, E.; HELT, G. A.; REESE, J. T.; MUNOZ-TORRES, M. C.; CHILDERS, C. P.; BUELS, R. M.; STEIN, L.; HOLMES, I.H; ELSIK, C.G.; LEWIS, S.E. Web Apollo: a web-based genomic annotation editing platform. **Genome Biology**, v. 14, n. 8, article number R93, Aug. 2013. DOI: 10.1186/gb-2013-14-8-r93.

LI, B.; DEWEY, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. **BMC Bioinformatics**, v. 12, n. 1, article number 323, Aug 2011. DOI: 10.1186/1471-2105-12-323.

MILLMAN, K. J.; AIVAZIS, M. Python for scientists and engineers. **Computing in Science & Engineering**, v. 13, n. 2, p. 9-12, Mar. 2011. DOI: 10.1109/MCSE.2011.36.

MUNGALL, C. J.; EMMERT, D. B. A Chado case study: an ontology-based modular schema for representing genome-associated biological information. **Bioinformatics**, v. 23, n. 13, p. i337-i346, July 2007. DOI: 10.1093/bioinformatics/btm189.

NAPIER, J. A.; HASLAM, R. P.; TSALAVOUTA, M.; SAYANOVA, O. The challenges of delivering genetically modified crops with nutritional enhancement traits. **Nature Plants**, v. 5, n. 6, p. 563-567, June 2019. DOI: 10.1038/s41477-019-0430-z.

PALATNIK-DE-SOUSA, C. B.; SOARES, I. da S.; ROSA, D. S. Epitope discovery and synthetic vaccine design. **Frontiers in Immunology**, v. 9, p. 826, 2018. DOI: 10.3389/978-2-88945-522-5.

PRADO, J. R.; SEGERS, G.; VOELKER, T.; CARSON, D.; DOBERT, R.; PHILLIPS, J.; COOK, K.; CORNEJO, C.; MONKEN, J.; GRAPES, L.; REYNOLDS, T.; MARTINO-CATT, S. Genetically engineered crops: from idea to product. **Annual Reviews of Plant Biology**, v. 65, n. 1, p. 769-790, Apr 2014. DOI: 10.1146/annurev-arplant-050213-040039.

QUEVILLON, E.; SILVENTOINEN, V.; PILLAI, S.; HARTE, N.; MULDER, N.; APWEILER, R.; LOPEZ, R. InterProScan: protein domains identifier. **Nucleic Acids Research**, v. 33, p. W116-W120, July 2005. Issue suppl_2. DOI: 10.1093/nar/gki442.

R CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2020. Available at: https://www.R-project.org. Accessed on: 7 May 2020.

RAPPUOLI, R. Reverse vaccinology. **Current Opinion in Microbiology**, v. 3, n. 5, p. 445-450, Oct. 2000. DOI: 10.1016/s1369-5274(00)00119-3.

RAPPUOLI, R.; COVACCI, A. Reverse vaccinology and genomics. **Science**, v. 302, n. 5645, p. 602, Oct. 2003. DOI: 10.1126/science.1092329.

ROBINSON, M. D.; MCCARTHY, D. J.; SMYTH, G. K. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. **Bioinformatics**, v. 26, n. 1, p. 139-140, Jan. 2010. DOI: 10.1093/bioinformatics/btp616.

SCHWABER, K. **Agile project management with scrum**. United States: Microsoft Press, 2004. 163 p.

SEPPEY, M.; MANNI, M.; ZDOBNOV, E. M. BUSCO: assessing genome assembly and annotation completeness. In: KOLLMAR, M. (ed.). Gene prediction. New York: Humana, 2019. p. 227-245. (Methods in molecular biology, v. 1962). DOI: 10.1007/978-1-4939-9173-0_14.

SKINNER, M. E.; UZILOV, A. V.; STEIN, L. D.; MUNGALL, C. J.; HOLMES, I. H. JBrowse: a next-generation genome browser. **Genome Research**, v. 19, n. 9, p. 1630-1638, 2009. DOI: 10.1101/gr.094607.109.

SPOOR, S.; CHENG, C. H.; SANDERSON, L. A.; CONDON, B.; ALMSAEED, A.; CHEN, M.; BRETAUDEAU, A.; RASCHE, H.; JUNG, S.; MAIN, D.; BETT, K.; STATON, M.; WEGRZYN, J. L.; FELTUS, F. A.; FICKLIN, S. P. Tripal v3: an ontology-based toolkit for construction of FAIR biological community databases. **Database**, v. 2019, 2019. DOI: 10.1093/database/baz077.

STEIN, L. D.; MUNGALL, C.; SHU, S.; CAUDY, M.; MANGONE M.; DAY, A.; NICKERSON, E.; STAJICH, J. E; HARRIS, T. W.; ARVA, A.; LEWIS, S. The generic genome browser: a building block for a model organism system database. **Genome Research**, n. 516, p. 1599-1610, 2002. DOI: 10.1101/gr.403602.

THE UNIPROT CONSORTIUM. UniProt: a worldwide hub of protein knowledge. **Nucleic Acids Research**, v. 47, p. D506-D515, Jan. 2019. Issue D1. DOI: 10.1093/nar/gky1049.

TRIMGALORE-GITHUB. Available at: https://github.com/FelixKrueger/TrimGalore. Accessed on: 7 May 2020.

WIKIPEDIA. **List of RNA-Seq bioinformatics tools**. 2020. Available at: https://en.wikipedia. org/wiki/List_of_RNA-Seq_bioinformatics_tools. Accessed on: 7 May 2020.

VERNIKOS, G. S. Genome watch: overtake in reverse gear. **Nature Reviews Microbiology**, v. 6, n. 5, p. 334-335, 2008. DOI: 10.1038/nrmicro1898.

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.

Photo: MPIX-TURE (AdobeStock)

Genomics applied to climate change

Biotechnology for digital agriculture

11

Juliana Erika de Carvalho Teixeira Yassitepe | Ricardo Augusto Dante | Isabel Rodrigues Gerhardt | Fernanda Rausch Fernandes | Rafael Soares Correa de Souza | Jaderson Silveira Leite Armanhi | Viviane Cristina Heinzen da Silva | Ana Paula Ribeiro | Márcio José da Silva | Paulo Arruda

Introduction

The most conspicuous manifestations of climate change are higher global atmospheric temperatures, which are already observed in several regions of the world. The projected climate scenarios for the coming decades show increases in the frequency and intensity of extreme events, such as prolonged periods of heat and drought, heavy rainfall, floods, and others (Mbow et al., 2019). Food production is particularly at risk in tropical, subtropical, and semi-arid regions, such as those in South America, Asia, and Africa. Between 1981 and 2010, reductions in the world average productivity were attributed to climate change, especially in these regions. Corn, wheat, and soybeans were reported to lose 4.1%, 1.8%, and 4.5%, respectively (Iizumi; Ramankutty, 2016). Impacts on fruit, vegetable, and animal production are also predicted for these same environments. Climate change not only jeopardizes world food security but also reduces food production and availability. Climate affects several biological processes important to the growth and development of plants and animals, and changes in these mechanisms can alter growth and reproduction rates, as well as nutrient quality and content (Damatta et al., 2010; Lara; Rostagno, 2013). Reductions in food supply and quality may impact consumers globally but will especially impact low-income consumers – up to about 183 million people could go hungry under projected climate change scenarios (Mbow et al., 2019).

Responding to current and future climate change scenarios requires two possible and necessary approaches, mitigation and adaptation. Applying better agricultural practices and developing more adapted and tolerant varieties to this new climate reality are essential and urgent for the sustainable increase of agricultural production in the coming decades. Increased tolerance to high temperatures and long periods of water restriction are selection criteria that should be applied in breeding programs to develop new cultivars, particularly in the most sensitive phases of the crop development cycle. Biotechnology tools such as molecular markers, gene editing, transgenesis, and microbiome, as well as more accurate large-scale phenotyping techniques, can and should be employed to accelerate the availability of genotypes adapted to specific regional conditions modified by recent climate changes.

A reduction in public funding for breeding programs has been observed worldwide compared to private sector companies, despite the undeniable importance of public research in agricultural production and food security, especially in medium and long-term scenarios (Alson et al., 2009). This decline is accompanied by Intellectual Property (IP) protection practices and increased private investments. This change has been observed mainly in seed companies, which went through a series of acquisitions and mergers, resulting in greater and concentrated participation in the world market and technological domain (Gutiérrez et al., 2014; Ray et al., 2015; OECD, 2018). Until the mid-1990s, the participation of national companies, including Embrapa, in the Brazilian soybean and corn seed markets was 70% and 30% (Silva et al., 2015). However, with the creation of the Patent (1996) and Plant Variety Protection (1997) Laws, biotechnology multinationals massively introduced proprietary seeds with biotechnological traits into the Brazilian market (Castro et al., 2006). As they do not invest in technological development on the same scale, the share of public companies in the seed market was reduced to less than 10% (Silva et al., 2015).

High investments and innovation capacity enable these large multinational companies to continuously develop new cultivars with specific genetic modifications through research and development pipelines integrating improvement and biotechnology. Traits incorporated include herbicide and pest resistance and, more recently, drought tolerance (Eisenstein, 2013; Rippey, 2015). A pipeline is the sequential process of research and development phases in which technologies move, broadly speaking, through discovery, validation, optimization, and commercial launch. As pipelines work in a continuous flow, at any given time, different technologies are in different phases of technological maturation throughout their development. Many of these biotechnological traits are also combined in the same cultivar or even licensed to competing companies. However, as they almost exclusively develop new biotechnology-derived crop traits in the countries of origin where their research and development centers are located, the maximum performance of these technologies is not achieved in global consumer markets where new discoveries are incorporated or adapted to local research and development programs. Therefore, it is strategic for the Brazilian agricultural sector, currently responsible for a quarter of the Gross Domestic Product (GDP), that public and private national institutions strengthen their scientific and technological production to contribute to the national development of appropriate technologies and varieties to our demands.

Embrapa is recognized for its remarkable track record in improving agricultural crops, a wide network of test sites, and qualified multidisciplinary human resources. In response to current demands, the operation of a biotechnology pipeline is emerging in this institution. The objective is to provide new biotechnology research in developing varieties adapted to the new and complex conditions imposed by climate change. This initiative requires long-term funding, highly coordinated approaches, and cross-disciplinary partnerships, including those between public and private companies. Public-private partnerships have been successful in discovering, developing, and commercializing biotechnological traits. The multinationals Bayer, BASF, Corteva, Syngenta, and their respective public and private partners have developed cultivars with increasingly advanced biotechnological traits. In two cases, genes introduced by genetic engineering were able to increase corn grain yields by 15% and 120% under intense water stress in a wide range of tested sites (Castiglioni et al., 2008; Nuccio et al., 2015). Cultivars of economically important crops generated by gene editing are already entering the North American seed market. In early 2019, soybean cultivars generated by the transcription activator-like effector nucleases (TALEN) system with high oleic acid content were released for commercial use. In addition, in 2019, corn hybrids with high amylopectin content generated through CRISPR-Cas9 were in the pre-release phase, with prospects of being available in 2020 (Kim; Kim, 2019; Gao et al., 2020). Likewise, improvement of drought resistance traits mediated by CRISPR-Cas9 is expected (Shi et al., 2016). In Brazil, similar initiatives to implement agricultural biotechnology pipelines are incipient, both in the public and private sectors, especially those involving the initial phase of discovering new genes of biotechnological importance.

At the end of 2017, a partnership between Embrapa, UNICAMP, and FAPESP created the Genomics for Climate Change Research Center (GCCRC), bringing together the first two institutions in agricultural biotechnology. The Center's mission is to develop biotechnological assets that will increase crop drought and heat tolerance over the next 10 years while transferring the developed technologies to the productive sector. Biotech assets under development can fit into different intellectual protection strategies that balance value and access to technology. These include (but are not limited to) genes, alleles, and gene constructs – which can be appropriately developed into traits by third parties – microbial inoculants, synthetic microorganism communities, new support technologies such as gene expression regulatory methods and elements, and regulatory patent know-how.

The GGCRC is the consolidation and expansion of the Mixed Unit for Research in Genomics Applied to Climate Change (UMiP GenClima), a technical-scientific cooperation agreement between Embrapa and UNICAMP signed at the end of 2012. Researchers and analysts from both institutions compose the GCCRC, where activities follow steps from a pipeline similar to large biotechnology companies, although smaller. National and international partners, public and private, contribute to the GCCRC team to achieve its mission.

In the digital transformation scenario, information and communication technologies can be integrated into biotechnology by adding computational tools in a research pipeline, including sensors and cameras for monitoring and data capture. Furthermore, these data will require mathematical models and statistical analyses in order to process the large volume of data generated by "omics". Altogether, these approaches will provide research advances for plant genetic improvement. The GCCRC contributes to the implementation of digital agriculture in Brazil through biotechnology and molecular biology research in its pre-production phase, allowing the development of new biotechnological assets for agribusiness. In this chapter, the GCCRC research pipeline will be presented, showing the steps involved in generating biotechnology-derived crop traits and illustrating how digital technologies help researchers obtain results.

The GCCRC research pipeline

The main research activities of the GCCRC are carried out through a research and development pipeline in biotechnology, which spans from the discovery phase to the proof-of-concept phase under field conditions (Figure 1). The species chosen as the target of the research work was corn, one of the most important agricultural crops in Brazil and in the world, which has a wide availability of genetic and genomic resources.

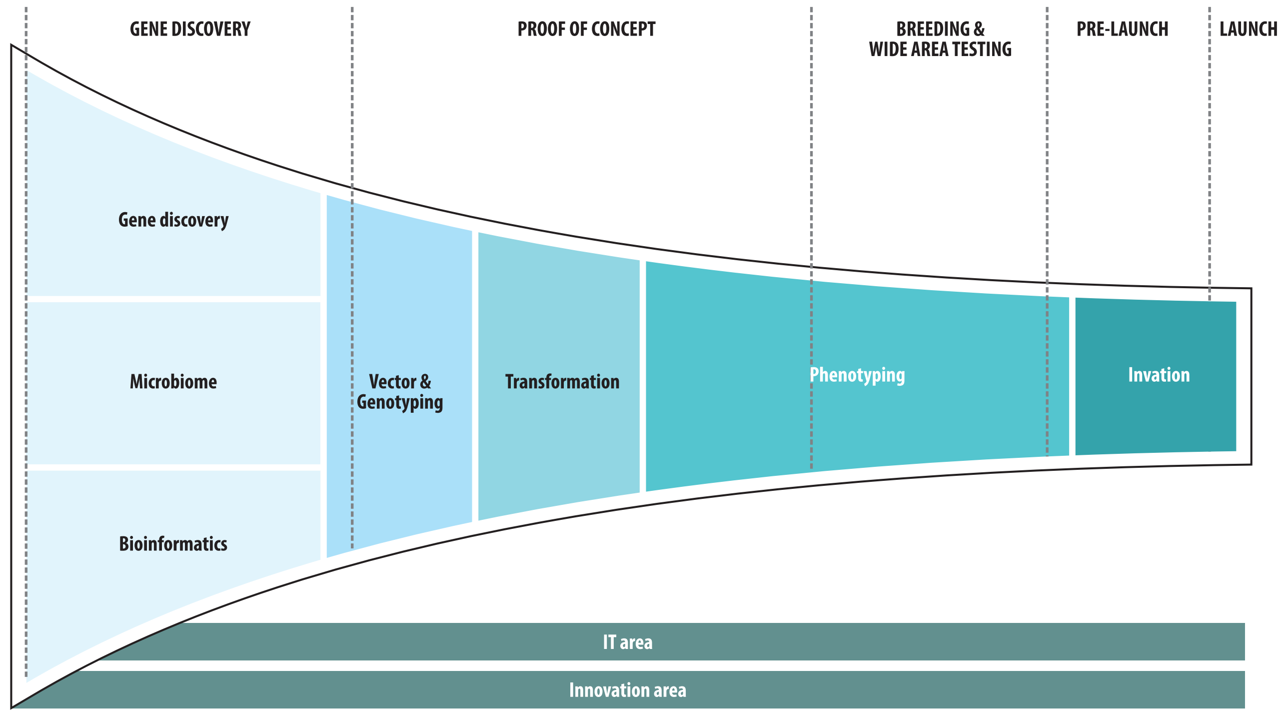


Figure 1. Genomics for Climate Change Research Center search pipeline phases.

The research pipeline has five phases. The first three concentrate most of the efforts of our team:

1) **Discovery:** when new genes and microorganisms are identified and indicated for introduction into the pipeline after the analysis of intellectual property and biosafety.

2) **Proof of concept:** genetic constructs and inoculants are developed, transgenic and edited plants are generated, with the first tests under controlled conditions (growth chambers and greenhouse), and on a small scale in the field. These are carried out for initial assessment of the strategy's effectiveness.

The subsequent phases are carried out in partnership with other organizations by way of collaboration and/or licensing, namely:

3) **Breeding and large-scale testing:** After discovery and selection, these transgenes, edited alleles, and/or inoculants are tested in large-scale field experiments at various locations and at different times. Promising events are being assimilated into elite corn strains.

4) **Pre-release:** development of commercial cultivars containing the technologies.

5) **Launch:** the technologies developed by the center are launched on the agricultural market.

A dedicated infrastructure was built for the effective operation of the research pipeline. The physical structure is composed of a molecular biology laboratory (Figure 2), a plant genetic transformation laboratory (Figure 3), a phenotyping laboratory under controlled environment conditions (under construction), and a modern greenhouse (Figure 4). The molecular biology laboratory houses all the activities of the discovery stage and a large part of the research team (Figure 2). The plant genetic transformation laboratory is equipped with a complete infrastructure for corn transformation, which includes two growth chambers designed for regeneration and acclimatization of transformed plants

(Figure 3). A modern greenhouse was built for growing corn dedicated to embryo production (explants used in genetic transformation), cultivation of transgenic and edited events from the pipeline, generational advancement, introgressions into elite material, inoculant tests, and initial screening experiments in controlled environments. This structure has five environments with temperature control, LED light supplementation, and a screened nursery to accommodate other species under study. All environments have internet access, and environmental conditions are constantly monitored (Figure 4).

Photos: Ana Paula Ribeiro

Figure 2. Molecular Biology Laboratory. Laboratory entrance (A); meeting room (B); internal view of work benches (C); and office (D).

Photos: Ana Paula Ribeiro

Figure 3. Laboratory of Genetic Transformation of Plants. External view of the laboratory (A); internal view of the laboratory (B and C); plant regeneration room (D); and plant acclimatization room (E).

Photos: Juliana Erika de Carvalho Teixeira Yassitepe

Figure 4. Vegetation house. Front view of the greenhouse (A); plants growing in the greenhouse and (C); plants growing with supplemental LED lighting (D); exterior night view of the greenhouse (E).

All pipeline routines and processes are registered in a Laboratory Integrated Management System (LIMS), ensuring that information is stored, managed, and tracked correctly, mainly for the purposes of intellectual property, biosafety, integrity of data and procedures (registration of activities, routines, protocols, reports, related documents, etc.).

Discovery of genes and microorganisms

The discovery phase is based on two fronts which focus on the identification and characterization of new (i.e., little or not studied) candidate genes and microorganisms with biotechnological potential to promote increased stress tolerance. Both fronts are structured, to a large extent, in multidisciplinary approaches for exploring the diversity of agricultural and wild plant species. There is special attention given to adaptations for limiting environmental conditions characterized by the incidence of one or more stresses. These approaches demand intensive use of bioinformatics and computational tools due to the large volume of analysis data produced by genomic technologies and related sciences (trancriptomics, metabolomics and metagenomics).

Global climate change, associated with population growth and competition for land will increasingly shift food and bioenergy production to marginal environments (Backlund et al., 2008; Ornella et al., 2012). These environments are characterized by one or more abiotic stresses, such as suboptimal levels of temperature (heat or cold) and water availability (drought or flood), unfavorable soil physical properties, and very low nutrient availability which impose limitations on productivity (Belaid; Morris, 1991). Therefore, the challenge posed by global climate change requires developing new adapted and more productive agricultural genotypes in stress-prone environments which naturally limit plant growth. Thus, understanding the adaptation of plant species to limiting environments and using a series of morphological, physiological, biochemical, and molecular changes in response to stresses that negatively affect growth and productivity can contribute to global food and bioenergy production in the next decades. Investigating wild species (not only those that are evolutionarily close, but also those distant from cultivated species) provides knowledge to guide the development of new genotypes capable of thriving in marginal environments (Mccouch et al., 2013). Among these species are the extremophiles and those tolerant to desiccation.

Extremophile organisms inhabit severely limiting environments, such as those characterized by extremes of temperature, water or nutrient availability and high salinity, stresses that occur alone or simultaneously (Oh et al., 2012). In turn, desiccation-tolerant species can survive long and/or severe periods of drought, supporting dramatically low levels of relative water content in vegetative tissues (Bartels; Hussain, 2011). Data sets and approaches derived from "omics" are a growing resource for the discovery of new genetic characters, where biotechnological use can contribute to abiotic stress adaptations. Many of these characters are unique to individual species (or a small group of related species) or belong to gene families present in many plant species that are functionally diversified through duplication and adaptive selection (Gollery et al., 2006; Horan et al., 2008). In such perspective, genes of unknown function represent 20-40% of the genes in each new sequenced genome, the majority constituting species-specific differences (Gollery et al., 2006, 2007), and are potentially associated with adaptive mechanisms, including stress tolerance (Mittler; Blumwald, 2010).

The Velloziaceae family of angiosperms contains the most desiccation tolerant species (approximately 200 of its 270 species). More than 80% of species in this family occur in South America, where the greatest morphological diversity is also found. The largest genus, *Vellozia*, comprises both tolerant and sensitive species to desiccation, offering an excellent model to study the evolution of desiccation and drought tolerance characters. *V. nivea* and *V. intermedia*, respectively tolerant and desiccation-sensitive species, are both drought-tolerant, endemic to Brazilian rupestrian grasslands, and highly adapted to its extreme conditions. These environments are characterized by a prolonged dry season, typically between late autumn and early spring, high solar radiation, and rocky, shallow soils which are poor in nutrients, particularly phosphorus. Unlike most model plant species, which originate from environments where nitrogen is the main limiting nutrient, the genus *Vellozia* evolved in an environment where phosphorus is the most limiting nutrient, making it a valuable model for crops grown in tropical soils, where the very low availability of this mineral prevails. The group has been exploring *V. nivea* and *V. intermedia* genomes, transcriptomes and metabolomes, in addition to other species of the same family. The resulting knowledge will help to identify genes and pathways underlying the adaptation of these species to their limiting environments, generating future agricultural genotypes with increased production capacity in marginal environments.

Plant survival under stressful conditions involves a combination of adaptive mechanisms that go beyond the unique contribution of their genomes (Rodriguez et al., 2008; Lau; Lennon, 2012). Microorganisms associated with plant tissues play a role in the adaptation to biotic and abiotic stresses, and play a key role in plant phenotypic plasticity (Woodward et al., 2012; Coleman-Derr; Tringe, 2014). Furthermore, recent advances have shown the existence of an unexplored microbial community with a significant impact on their hosts (Bulgarelli et al., 2013; Souza et al., 2016). These findings make microbiome research an important source of genetic and biological resources for biotechnological use in improving plant adaptation to stressful conditions.

Traditionally, research on microorganisms associated with plants is based on culture-dependent techniques, which are based on the isolation and cultivation of microorganisms. However, the restricted use of these culture methodologies can bias the microbiota sampling, since only microorganisms capable of growing in the culture media are recorded. Furthermore, these methodologies do not provide information about the real abundance or real functional contribution of an isolate in its original habitat. More recently, large-scale sequencing tools have allowed access to microbial diversity in a crop-independent manner, enabling more accurate mapping of the phylogenetic and functional microbiota profile associated with plants. However, although sequencing techniques elucidate vital issues, the isolation of microorganisms is still necessary for biotechnological applications. Despite being complementary, however, both strategies are rarely used together.

Differently from traditional approaches, the microbiome investigation pipeline makes use of dependent and independent cultivation techniques concomitantly. The use of genomic investigation tools provides information about diversity, colonization patterns and functions performed by the microbiota in association with the plant. These data enable us to identify the most efficient microorganisms in association with plants that promote plant growth. Based on this information, synthetic microbial communities are designed with the collection of isolated microorganisms (Armanhi et al., 2018). Synthetic communities are validated in inoculation experiments to assess their ability to increase stress tolerance and maintain plant productivity even under unfavorable conditions.

In a complementary and synergistic approach to the exploration of plant species in rupestrian fields, this approach has been applied in the pipeline to investigate the strategies by which microorganisms contribute to plant survival in the stressful conditions of these habitats. This is based on the assumption that microbial communities associated with plant species that evolved in environments historically exposed to drought and nutritional scarcity are more likely to promote tolerance to these stresses in the plant than microorganisms originating in environments where such resources are not limiting (Rodriguez et al., 2008; Redman et al., 2011; Lau; Lennon, 2012). These studies are ongoing and allow mapping the composition, abundance, and diversity of bacterial and fungal communities associated with native plants adapted to limiting environments; creation of a comprehensive collection of microorganisms associated with these species; the investigation of plant and microorganism interactions related to plant growth under stressful conditions; and extensive analysis of microbiome genomes (metagenomes, produced from DNA directly recovered from samples) in search of stress tolerance gene functions.

The microbiomes associated with Velloziaceae and other species from rupestrian fields had not been characterized until recently. A first study carried out by the GCCRC, the Joint Genome Institute (JGI, USA) and partners described the identification of great bacterial and fungal diversity and novelty in the microbiomes of two endemic Velloziaceae that inhabit soil and rock in rupestrian fields in Serra do Cipó (MG) (Camargo et al., 2019). Microbial diversity and abundance in epiphytic (external) and endophytic (internal) compartments of roots, stems, leaves and substrates were evaluated by sequencing molecular markers. The root and substrate metagenomes of each species were also sequenced. The results compose the first microbiome databases associated with endemic Velloziaceae species in rupestrian fields. These findings will significantly support the discovery of new microorganisms and, consequently, the potential for obtaining new inoculants.

Proof of concept

The proof-of-concept stage ranges from the design of the genetic constructs that will be inserted into corn plants for the development of transgenic or edited events, as well as the preparation of microbial inoculants for the testing of developed technologies in controlled environments and on a smaller scale, in the field.

After defining the gene constructs containing candidate genes and regulatory sequences, the Vector Construction and Genotyping teams build the vectors and begin the recombinant DNA molecule validation phase with sequencing and other molecular tools, as shown in Figure 1. Once the validated gene constructs are made available, the Transformation team is called upon to carry out the genetic transformation of the target species, corn. The Transformation team performs corn transformation using locally optimized protocols and appropriate corn genotypes. The transformation platform is designed to routinely transform immature embryos using weekly gene constructs provided by the Vector Construction and Genotyping teams. The Transformation team is constantly improving corn transformation protocols through established national and international partnerships. Alternatively, for strategies based on candidate genes (or their related genes) that are identified in corn, gene editing methods are being used to carry out specific modifications. This approach has been used in parallel, while aiming to obtain biotechnological assets. Regenerated plants are evaluated for number of copies and expression level, and edited plants are evaluated for the presence of the edited allele by means of sequencing. Once the genetic transformation or editing is confirmed, the plants are considered transgenic and edited events, respectively, and are later transferred to the greenhouse for crossing (or self-fertilization) and phenotyping by the Phenotyping team.

Microorganisms with a potential role in plant tolerance to abiotic stresses discovered by the Discovery team are organized into synthetic microbial communities. Inoculants with different microorganism combinations and microbial communities are prepared and used in experiments, validating their effectiveness. This is carried out by the Phenotyping team based on the assessment of their ability to promote tolerance to abiotic conditions.

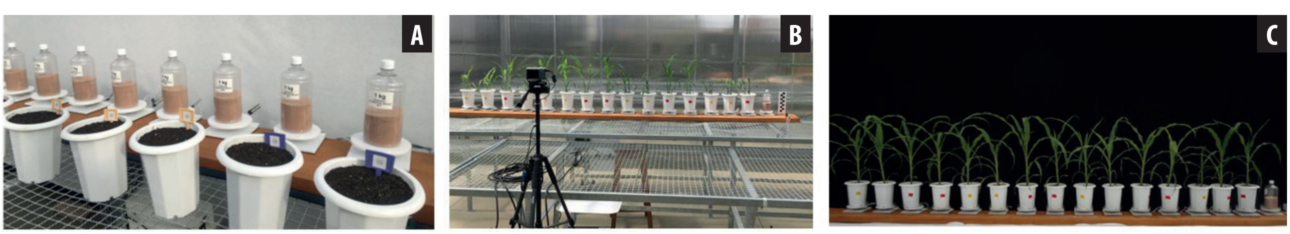
The phenotypic evaluation of plants is one of the most important phases in any cultivar development program, as it will define which genotypes will be eliminated and which will proceed to the next steps. In biotechnology pipelines, this phase is even more important, especially for characters that tolerate abiotic stresses, such as water and heat stress. These are complex features, and are often the effect of the transgene, edited allele, and/or inoculant conferring tolerance. These may be significant, but difficult to separate from the effect of the plant's genetic background. In a pipeline where hundreds of transgenes, edited alleles, and/or inoculants, and thousands of plants must be evaluated, a fast and reliable selection procedure is necessary in order to eliminate unpromising discoveries. The instruments conventionally and routinely used in laboratories to assess plant physiological condition are reliable, but often require destructive sampling, in addition to only allowing specific assessments. Appropriate instruments for continuous and real-time phenotypic assessment allow for a more detailed evaluation of plant physiological responses to environmental variables and treatments. They can also provide additional information with the potential to improve the understanding of the phenotypic response. Several non-invasive and non-destructive technologies have emerged in the field of plant phenotyping in recent years, including spectroscopy, fluorescence, thermography, and digital image capture. These new technologies are currently being used to increase the quantity, quality and plurality of measured characters and allow the distinction of phenotypic effects with the support of modern statistical analyses.

In most biotechnology pipelines, the evaluation of transgenic, edited events and/or microbial inoculants is carried out in three stages: a) initial screening in a growth chamber and/or greenhouse; b) detailed characterization, in a greenhouse; and c) phenotyping in field assays. Controlled environments, such as growth chambers and greenhouses, have a low-cost phenomics platform, which uses sensors and chambers to monitor the environments and plant responses to applied treatments. In the initial screening phase, the plants are evaluated for their stress resistance to short heat and drought cycles during the vegetative stage. Characters are measured in seedling aerial parts and roots according to the expected effect of the event or the tested inoculant. In the detailed characterization phase, plants are evaluated throughout the development cycle, including the reproductive stage and grain production. Various biometric and physiological assessments are performed at different developmental stages, and promising events are assessed for copy number, expression levels, protein and metabolomic profiles, among others, in order to characterize and understand the effect of the gene/construction and microorganisms applied in the plant.

The low-cost phenomics platform developed at the Center has sensors and cameras that continuously monitor the plants' phenotypic response in real time (Armanhi, 2018). Raspberry Pi and Arduino microcontrollers automatically control sensor readings that monitor the environment (light intensity, relative air humidity, and temperature) and individual plant response (leaf temperature, substrate moisture, and water loss from the pot-plant system). In addition, other parameters can be indirectly obtained, such as the vapor pressure deficit (VPD), a parameter that indicates the plant's propensity to lose water to the environment, and evapotranspiration through loss of water from the pot-plant system. The recorded data is statistically processed, stored, and sent to a local server. An internally developed website allows the graphical visualization of all mentioned parameters in real time.

Microcomputers also automatically control photographic cameras, which record the plants at different angles, at the desired frequency, and send the images to the local server. The entire time image series is accessed remotely, which is used for biometric assessments through available image analysis software. The images can also be used to construct time-lapse image videos, useful in visualizing the continuous response over time, in addition to the observation of small variations throughout the day, such as the expansion and rolling movement of the leaves in response to variations in light intensity and ambient temperature, for example.

Figure 5 illustrates some aspects of the phenomics platform installed in the greenhouse. Scales are used to monitor vessel weight throughout the experiment (Figure 5A). Before the beginning of each experiment, a calibration of the scales is carried out to verify its functioning and the quality of the measurements. A camera is installed to record plant growth in real time (Figure 5B). Cameras and sensors installed in the plants, in the pot and in the environment constantly monitor the weight of the pot-plant system, soil/substrate moisture, leaf temperature, in addition to temperature, relative humidity and ambient light intensity (Figure 5C).



Photos: Jaderson Silveira Leite Armanhi

Figure 5. Some aspects of the phenomics platform installed in the greenhouse. Scales and sensors installed for continuous monitoring of the weight of the pot-plant system and soil/substrate moisture (A); camera installed to continuously record plant growth (B); corn plants monitored continuously in the experiments carried out (C).

After the evaluation phases in controlled environments, promising events and inoculants are selected based on increased resistance to drought and heat stresses (e.g., higher growth rate, lower leaf temperature, lower water loss, higher photosynthetic efficiency, among others) compared to control plants. These are subsequently moved on to the field evaluation phase.

In the field phenotyping phase, the tested events are evaluated in experiments with water restriction, in the reproductive phase, in at least three environments and two different times. Agronomic characters and grain production are evaluated. Events that demonstrate superiority over controls, in more than one location and time, are selected and proceed to the next steps in the pipeline, where building optimizations, elite germplasm introgression, and large-scale testing are performed.

Enhancement, large-scale testing, pre-release and release

The research pipeline stages after the proof of concept are carried out in partnership with public and private institutions that show interest in advancing the technologies for later commercialization.

After the proof-of-concept phase when compared to control treatments, transgene, edited allele, and/or microbial inoculant showing superior field effect on drought and heat tolerance, can be explored in product development pipelines. Edited genes and alleles can be incorporated into improvement programs as an additional source of variability for tolerance to abiotic stresses. Large-scale tests in various locations and years should be carried out as part of the routine for selecting superior genotypes in improvement programs. These may indicate the potential for gains that the introduction of the edited gene or allele could generate in new cultivars. The team works together with partners to ensure an optimal evaluation of the events developed by the center, following all the required biosafety standards.

Similarly, microbial inoculants that present superior performance in field tests in the proof of concept phase should be investigated in extended tests, mainly to evaluate the effects of genotype x inoculant interaction and performance in different locations and seasons. In addition to agronomic efficacy, the development process of a commercial inoculant involves a series of tests aimed at identifying the best formulation. The dose to be applied, the application and storage conditions, and others follow the recommendations of the Ministry of Agriculture, Livestock and Supply (MAPA).

The innovation team will develop actions for: a) collaboration and prospecting for partnerships; b) technology assessment; c) mapping and monitoring of potential markets for our technologies. All these activities are incorporated into the Strategic Plan, which works as a guideline for the pipeline and composes the technology showcase. At this point, establishing partnerships with private companies can provide information and demands to guide the development of applicable technological solutions, in alignment with market demands, and facilitate the transfer in future businesses.

Technology transfer must consider the design of new commercial models, which justify exploring new varieties of market technologies and/or cultural hybrids. Since the objective is the advancement of technologies to the proof-of-concept phase, some of these models may consider licenses and commercial benefits contemplating investments made by a licensee in regulatory and administration processes, and product development.

Final considerations

The GCCRC has corn crop as a research target, but the developed technologies could potentially be transferred to other agricultural crops. The GCCRC has built a modern infrastructure to meet the pipeline's demands, with new greenhouses and laboratories for plant transformation, molecular biology, bioinformatics, and phenotyping. The latter, particularly, has technology that incorporates several low-cost, high-precision sensors, and information systems developed locally for the collection of large numbers of phenotypic data in real time. The first scientific and technological results are already being achieved. Unexplored genes and sometimes of unknown function, associated with responses to abiotic stresses, were discovered, and the first ones are in the proof-of-concept phase in corn and in field tests in sugarcane. The team already masters the technology of gene editing in corn, and edited plants are continuously generated. Synthetic microbial communities composed of beneficial microorganisms that increase corn yield under stressful conditions have been discovered and tested under controlled conditions and in the field. Recent efforts in the sequencing and assembly of genomes and microbiomes of plants in rupestrian fields open a new path to be explored, in search of new genes and microorganisms adapted to water and nutritionally limiting environments. Following the pipeline rationale, new genes and microorganisms are continually being discovered and tested.

The research carried out within the GCCRC leverages the digital transformation in agriculture, thus promoting the development of new cultivars with genetic modifications that incorporate tolerance to drought and other stresses, thus contributing to the country's ability to sustainably grow and saving natural resources.

References

ALSON J. M.; BEDDOW, J. M.; PARDEY, P. G. **Agricultural research, productivity, and food prices in the long run. Science**, v. 325, n. 5945, p. 1209-1210, Sept. 2009. DOI: 10.1126/science.1170451.

ARMANHI, J. S. L. **Construção de uma comunidade sintética bacteriana promotora do crescimento vegetal oriunda do microbioma de cana-de-açúcar**. 2018. 148 f. Tese (Doutorado em Genética e Biologia Molecular) – Universidade Estadual de Campinas, Instituto de Biologia, Campinas.

ARMANHI, J. S. L.; SOUZA, R. S. C.; DAMASCENO, N. B.; ARAÚJO, L. M.; IMPERIAL, J.; ARRUDA, P. A Community-based culture collection for targeting novel plant growth-promoting bacteria from the sugarcane microbiome. **Frontiers in Plant Science**, v. 8, article 2191, 2018. DOI: 10.3389/fpls.2017.02191.

BACKLUND, P.; JANETOS, A.; SCHIMEL, D. **The effects of climate change on agriculture, land resources, water resources, and biodiversity in the United States**: synthesis and assessment product: report by the U.S. Climate Change Science Program and the Subcommittee on Global Change Research. Washington, DC: Environmental Protection Agency, Climate Change Science Program, 2008. 240 p.

BARTELS, D.; HUSSAIN, S. Resurrection plants: physiology and molecular biology. In: LÜTTGE, U.; BECK, E.; BARTELS, D. (ed.). **Plant desiccation tolerance**. Berlin: Springer, 2011. p. 339-364. (Ecological studies. Analysis and synthesis, v. 215). DOI: 10.1007/978-3-642-19106-0_16.

BELAID, A.; MORRIS, M. L. **Wheat and barley production in rainfed marginal environments of West Asia and North Africa**: problems and prospects. Mexico, D.F.: CIMMYT, 1991. (CIMMYT economics working paper 91/02).

BULGARELLI, D.; SCHLAEPPI, K.; SPAEPEN, S; VAN THEMAAT E. V. L.; SCHULZE-LEFERT, P. Structure and functions of the bacterial microbiota of plants. **Annual Review in Plant Biology**, v. 64, p. 807-838, 2013. DOI: 10.1146/annurev-arplant-050312-120106.

CAMARGO, A. P.; SOUZA, R. S. C.; DE BRITTO COSTA, P.; GERHARDT, I. R.; DANTE, R. A.; TEODORO, G. S.; ABRAHAO, A.; LAMBERS, H.; CARAZZOLLE, M. F.; HUNTEMANN, M.; CLUM, A.; FOSTER, B.; FOSTER, B.; ROUX, S.; PALANIAPPAN, K.; VARGHESE, N.; MUKHERJEE, S.; REDDY, T. B. K.; DAUM, C.; COPELAND, A.; CHEN, I. A.; IVANOVA, N. N.; KYRPIDES, N. C.; PENNACCHIO, C.; ELOE-FADROSH, E. A.; ARRUDA, P.; OLIVEIRA, R. S. Microbiomes of Velloziaceae from phosphorus-impoverished soils of the campos rupestres, a biodiversity hotspot. **Scientific Data**, v. 6, article number 140, p. 1-11, 2019. DOI: 10.1038/s41597-019-0141-3.

CASTIGLIONI, P.; WARNER, D.; BENSEN, R.J.; ANSTROM, D. C.; HARRISON, J.; STOECKER, M.; ABAD, M.; KUMAR, G.; SLAVADOR, S.; D'ORDINE, R.; NAVARRO, S.; BACK, S.; FERNANDES, M.; TARGOLLI, J.; DASGUPTA, S.; BONIN, C.; LUETHY, M. H.; HEARD, J. E. Bacterial RNA chaperones confer abiotic stress tolerance in plants and improved grain yield in maize under water-limited conditions. **Plant Physiology**, v. 147, p. 446-455, 2008. DOI: 10.1104/pp.108.118828.

CASTRO, A. M. G. de; LIMA, S. M. V.; LOPES, M. A.; MACHADO, M. dos S.; MARTINS, M. A. G. **O futuro do melhoramento genético vegetal no Brasil**: impactos da biotecnologia e dos direitos de proteção do conhecimento. Brasília, DF: Embrapa Informação Tecnológica: Embrapa-Secretaria de Pesquisa e Desenvolvimento, 2006. 506 p.

COLEMAN-DERR, D.; TRINGE, S. G. Building the crops of tomorrow: advantages of symbiont-based approaches to improving abiotic stress tolerance. **Frontiers in Microbiology**, v. 5, article 283, 2014. DOI: 10.3389/fmicb.2014.00283.

CONCENTRATION in seed markets: potential effects and policy responses. Paris: OECD Publishing, 2018. 236 p. DOI: 10.1787/9789264308367-en.

DAMATTA, F. M.; GRANDIS, A.; ARENQUE, C. C.; BUCKERIDGE, M. S. Impacts of climate changes on crop physiology and food quality. **Food Research International**, v. 43, n. 7, p. 1814-1823, Aug 2010. DOI: 10.1016/j.foodres.2009.11.001.

EISENSTEIN, M. Discovery in a dry spell. **Nature**, v. 501, p. S7-S9, Sept 2013. DOI: 10.1038/501S7a.

GAO, H.; GADLLAGE, M. J.; LATIFFE, H. R.; LENDERTS, B.; YANG, M.; SCHRODER, M.; FARRELL, H.; SNOPEK, K.; PETERSON, D.; FEIGENBUTZ, L.; JONES, S.; CLAIR, G. S.; RAHE, M.; SANYOUR-DOYEL, N.; PENG, C.; WANG, L.; YOUNG, J. K.; BEATTY, M.; DAHLKE, B.; HAZEBROEK, J.; GREENE, T. W.; CIGAN, A. M.; CHILCOAT, N. D.; MEELEY, R. B. Superior field performance of waxy corn engineered using CRISPR-Cas9. **Nature Biotechnology**, v. 38, p. 579-581, 2020. DOI: 10.1038/s41587-020-0444-0.

GOLLERY, M.; HARPER, J.; CUSHMAN, J.; MITTLER, T.; GIRKE, T.; ZHU, J. K.; BAILEY-SERRES, J.; MITTLER, R. What makes species unique? The contribution of proteins with obscure features. **Genome Biology**, v. 7, article R57, 2006. DOI: 10.1186/gb-2006-7-7-r57.

GOLLERY, M.; HARPER, J.; CUSHMAN, J.; MITTLER, T.; MITTLER, R. POFs: what we don't know can hurt us. **Trends in Plant Science**, v. 12, n. 11, p. 492-496, Nov 2007. DOI: 10.1016/j.tplants.2007.08.018.

GUTIÉRREZ, A. P.; ENGLE, N. E.; DE NYS, E.; MOLEJÓN, C.; MARTINS, E. S. Drought preparedness in Brazil. **Weather and Climate Extremes**, v. 3, p. 95-106, June 2014. DOI: 10.1016/j.wace.2013.12.001.

HORAN, K.; JANG., C.; BAILEY-SERRES, J.; MITTLER, R.; SHELTON, C.; HARPER, JF.; ZHU, J. K.; CUSHMAN, J. C.; GOLLERY, M.; GIRKE, T. Annotating genes of known and unknown function by large-scale coexpression analysis. **Plant Physiology**, v. 147, p. 41-57, 2008. DOI: 10.1104/pp.108.117366.

IIZUMI, T.; RAMANKUTTY, N. Changes in yield variability of major crops for 1981–2010 explained by climate change. **Environmental Research Letters**, v. 11, n. 3, article 034003, 2016. DOI: 10.1088/1748-9326/11/3/034003.

KIM, J.; KIM, J. New era of precision plant breeding using genome editing. **Plant Biotechnology Reports**, v. 13, p. 419-421, 2019. DOI: 10.1007/s11816-019-00581-w.

LARA, L.; ROSTAGNO, M. Impact of heat stress on poultry production. **Animals**, v. 3, p. 356-369, 2013. DOI: 10.3390/ani3020356.

LAU, J. A.; LENNON, J. T. Rapid responses of soil microorganisms improve plant fitness in novel environments. **Proceedings of the National Academy of Science of the United States of America**, v. 109, p. 14058-14062, 2012. DOI: 10.1073/pnas.1202319109.

MBOW, C. C.; ROSENZWEIG, L. G.; BARIONI, T. G.; BENTON, M.; HERRERO, M.; KRISHNAPILLAI, E.; LIWENGA, P.; PRADHAN, M. G.; RIVERA-FERRE, T.; SAPKOTA, F. N.; TUBIELLO, Y. XU. Food security. In: SHUKLA, P. R.; SKEA, J.; BUENDIA, E. C.; MASSON-DELMOTTE, V.; PÖRTNER, H. O.; ROBERTS, D. C.; ZHAI, P.; SLADE, R.; CONNORS, S.; VAN DIEMEN, R.; FERRAT, M.; HAUGHEY, E.; LUZ, S.; NEOGI, S.; PATHAK, M.; PETZOLD, J.; PEREIRA, J. P.; VYAS, P.; HUNTLEY, E.; KISSICK, K.; BELKACEMI, M.; MALLEY, J. (ed.). **Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems**. [S.l.]: Intergovernmental Panel on Climate Change, 2019. p. 437-550.

MCCOUCH, S.; BAUTE, G. J.; BRADEEN, J.; BRAMEL, P.; BRETTING, P. K.; BUCKLER, E.; BURKE, J. M.; CHAREST, D. C.; CLOUTIER, S.; COLE, G.; DEMPEWOLF, H.; DINGKUHN, M.; FEUILLET, C.; GEPTS, P.; GRATTAPAGLIA, D.; GUARINO, L.; JACKSON, S.; KNAPP, S.; LANGRIDGE, P.; LAWTON-RAUH, A.; LIJUA, Q.; LUSTY, C.; MICHAEL, T.; MYLES, S.; NAITO, K.; NELSON, R. L.; PONTAROLLO, R.; RICHARDS, C. M.; RIESEBERG, L.; ROSS-IBARRA, J.; ROUNSLEY, S.; HAMILTON, R. S.; SCHURR, U.; STEIN, N.; TOMOOKA, N.; KNAAP, E. VAN DER; TASSEL, D. VAN; TOLL, J.; VALLS, J.; VARSHNEY, R. K.; WARD, J.; WAUGH, R.; WENZL, P.; ZAMIR, D. Feeding the future. **Nature**, v. 499, p. 23-24, July 2013. DOI: 10.1038/499023a.

MITTLER, R.; BLUMWALD, E. Genetic engineering for modern agriculture: challenges and perspectives. **Annual Review of Plant Biology**, v. 61, p. 443-462, 2010. DOI: 10.1146/annurev-arplant-042809-112116.

NUCCIO, M. L.; WU, J.; MOWERS, R.; ZHOU, H. P.; MEGHJI, M.; PRIMAVESI, L. F.; PAUL, M. J.; CHEN, X.; GAO, Y.; HAQUE, E.; BASUL, S. S.; LAGRIMINI, L. M. Expression of trehalose-6-phosphate phosphatase in maize ears improves yield in well-watered and drought conditions. **Nature Biotechnology**, v. 33, p. 862-869, 2015. DOI: 10.1038/nbt.3277.

OECD. **Concentration in Seed Markets**: Potential Effects and Policy Responses. Paris: OECD Publishing, 2018.

OH, D.-H.; DASSANAYAKE, M.; BOHNERT, H. J.; CHEESEMAN, J. M. **Life at the extreme**: lessons from the genome. Genome Biology, v. 13, article number 2747, 2012. DOI: 10.1186/gb4003.

ORNELLA, L.; CERVIGNI, G.; TAPIA, E. Applications of machine learning for maize breeding for stress. In: VENKATESWARLU, B.; SHANKER, A. K.; SHANKER, C.; MAHESWARI, M. (ed.). **Crop stress and its management**: perspectives and strategies. Netherlands: Springer Netherlands, 2012. p. 163-192. DOI: 10.1007/978-94-007-2220-0_5.

RAY, D. K.; GERBER, J. S.; MACDONALD, G. K.; WEST, P.C. Climate variation explains a third of global crop yield variability. **Nature Communication**, v. 6, article number 5989, p. 1-9, 2015. DOI: 10.1038/ncomms6989.

REDMAN, R. S.; KIM, Y. O.; WOODWARD, C. J. D.; GREER, C.; ESPINO, L.; DOTY, S. L.; RODRIGUEZ, R. J. Increased fitness of rice plants to abiotic stress via habitat adapted symbiosis: a strategy for mitigating impacts of climate change. **Plos One**, v. 6, issue 7, e14823, 2011. DOI: 10.1371/journal.pone.0014823.

RIPPEY, B. R. The U.S. drought of 2012. **Weather and Climate Extremes**, v. 10, part A, p. 57-64, Dec 2015. DOI: 10.1016/j.wace.2015.10.004.

RODRIGUEZ, R. J.; HENSON, J.; VAN VOLKENBURGH, E.; HOY, M.; WRIGHT, L.; BECKWITH, F.; KIM, Y.-O.; REDMAN, R. S. Stress tolerance in plants via habitat-adapted symbiosis. **The ISME Journal**, v. 2, p. 404-416, 2008. DOI: 10.1038/ismej.2007.106.

SHI, J.; GAO, H.; WANG, H.; LAFITTE, W. H.; ARCHIBALD, R. L.; YANG, M.; HAKIMI, S. M.; MO, H.; HABBEN, J. E. ARGOS8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. **Plant Biotechnology Journal**, v. 15, p. 207-216, July 2016. DOI: 10.1111/pbi.12603.

SILVA, F. F.; BRAGA, M. J.; GARCIA, J. C. Concentração nos mercados de sementes de milho, soja e algodão: uma abordagem ECD. **Agroalimentaria**, v. 21, p. 133-150, 2015.

SOUZA, R. S. C.; OKURA, V. K.; ARMANHI, J. S. L.; JORRÍN, B.; LOZANO, N.; DA SILVA, M. J.; GONZÁLEZ-GUERRERO, M.; ARAÚJO, L. M.; VERZA, N. C.; BAGHERI, H. C.; IMPERIAL, J.; ARRUDA, P. Unlocking the bacterial and fungal communities assemblages of sugarcane microbiome. **Scientific Reports**, v. 6, article number 28774, June 2016. DOI: 10.1038/srep28774.

WOODWARD, C.; HANSEN, L.; BECKWITH, F.; REDMAN, R. S.; RODRIGUEZ, R. J. Symbiogenics: an epigenetic approach to mitigating impacts of climate change on plants. **HortScience**, v. 47, n. 6, p. 699-703, June 2012. DOI: 10.21273/HORTSCI.47.6.699.

12

Innovation ecosystem in agriculture

Embrapa's evolution and contributions

Luciana Alvim Santos Romani | Martha Delphino Bambini | Joice Machado Bariani | Debora Pignatari Drucker | André Fachini Minitti | Adriana Farah Gonzalez | Vinícius Milléo Kuromoto | Guiomar Alessandra de Souza Telles | Ricardo Fonseca Araújo | Cleidson Nogueira Dias | Breno Silva Beda de Assunção | Shalon Silva de Souza Figueiredo | Ariovaldo Luchiari Junior | Carlos Alberto Alves Meira

Introduction

There are several challenges facing digital transformation of agriculture, and according to Simões et al. (2017, p. 52),

> [...] implementing an adapted and less expensive digital agriculture model using national technology could be the great revolution in the field, as small farmers and large grain producers search for innovation, thus making their crops more efficient and sustainable.

The concept of open innovation reflects the collaboration scenario, which is important to operate in a context as dynamic as that of digital innovation (Chesborough et al., 2003), in which knowledge, experience and capabilities are dispersed among various organizations. Gomes et al. (2018) consider the perspective of joint value creation and innovation in an ecosystem, with cooperation and competition processes.

The work of Teixeira et al. (2017) highlights that acting in innovation ecosystems involves reciprocity between public and private actors (from corporations to technological startups), organizations that support the creation of enterprises (such as incubators, business accelerators, associations, coworkings, hubs innovation), investors, services such as the Brazilian Micro and Small Business Support Service (SEBRAE) and entrepreneurship support organizations such as Endeavor.

The "ecosystem" concept was developed in the 1930s with the intention of creating a clearer and simpler nomenclature for biological systems in the field of ecology (Golley, 1991), and was later used in other study areas. Bambini and Bonacelli (2019) highlight other uses of the term "ecosystem" in the social sciences, referring to the entrepreneurial environment. Some approaches emphasize the entrepreneurial ecosystem developed around large universities (Fetters et al., 2010), while others emphasize the role of entrepreneurs in influencing their ecosystem (Feld, 2012). Malecki's (2018) approach considers that the entrepreneur is a central actor in an innovation ecosystem, always integrated with other equally important organizations and institutions, with a role of interdependence and complementarity.

One of the origins of the innovation ecosystem concept is precisely the systemic approach to innovation (Suominen et al., 2019), developed in the early 1990s by authors such as Lundvall (1992) and Nelson (1993), who consider that innovation is the result of relationships established between various actors to produce, disseminate and implement new knowledge, economically useful, within a nation.

The innovation systems approach quickly spread in the academic environment, as well as in regional and sectoral perspectives (Edquist, 2006). The focus of sectoral innovation systems considers the specificities of each economic sector in relation to innovation processes, based on their sources of scientific knowledge and technological opportunities, the research fields involved, actors, relationships, performance of institutions and policies, market, and others (Malerba, 2006).

The analysis of innovative processes in agriculture has been developed with a systemic focus and, more recently, also under the innovation ecosystem approach, according to Pigford et al. (2018). The difference of this approach is that the ecosystem actors interact with each other and with the ecosystem as a whole, creating a value that would not be generated without the relationships and complementarity of the other actors involved. The ecosystem concept is associated with a shared environment of evolution for the creation of value and innovation, with different roles and competences, marked by the sharing and mutual strengthening of the dynamic capabilities of the participating actors (Teece, 2009; Suominen et al., 2019).

There are several categories of actors involved in agricultural innovation, namely: a) agricultural producers; b) education and training system; c) agricultural research system; d) research and innovation agencies; e) credit agencies; f) rural extension and technical assistance system; g) companies supplying inputs, equipment and services; h) producer and business organizations; i) agroprocessors; j) exporters; k) government institutions; and l) end consumers (Rajalahti, 2012). Also part of the "enterprise generation mechanisms", such as business incubators, accelerators, coworkings and open laboratories (Audy; Piqué, 2016), and, with a central role are the agricultural technology-based startups, called agtechs, agritechs or agrifoodtechs (AgFunder, 2020). It is understood that agtechs have more agility, knowledge and audacity so that the new technologies reach the field, as they have a more agile and disruptive modus operandi and mentality (Cook, 2020).

In this new innovative context, Embrapa Digital Agriculture, a research unit of the Brazilian Agricultural Research Corporation (Embrapa) that works in digital technologies, has been contacted by companies from different segments, with emphasis on information technology, bank insurance, communication technologies, investors and accelerators, and also by Non-Governmental Organizations (NGOs). Researchers at Embrapa Digital Agriculture have offered technical mentoring in several startup development programs, strengthening companies and the sector's human capital.

This chapter reports the performance of Embrapa Digital Agriculture in terms of strengthening the Brazilian agricultural innovation ecosystem. The next section describes the Brazilian contextual characteristics and the actions developed by Embrapa nationwide. Section 3 details initiatives developed in the state of São Paulo, considering the type of actor involved and the type of relationship established. Section 4 describes the AgroAPI initiative, an example of an innovative public-private partnership model for digital agriculture. At the end of the chapter, the final remarks are presented.

Embrapa's performance record in agricultural innovation ecosystems

Agricultural innovation ecosystem in Brazil: actors, resources and regional differences

Economic geography is a field that seeks to understand dynamics and competitiveness, analyzing the spatial agglomeration of economic activity as a source of growing benefits. Feldman and Kogler (2010) point out that innovation is geographically concentrated, with differences between locations depending on their innovation capacities and configuration, defined by historical, cumulative and evolutionary processes developed over time. The evolutionary perspective[1] of economic geography also considers the influence of historical events, whether close or remote, random or not, on the trajectory and results of economic changes that occur over time (Davi, 1985; Arthur, 1994).

Sotarauta's (2004) work analyzes the capabilities of each region to use and create resources, developing a strategic regional development model based on dynamic capabilities, a concept originally established by Teece et al. (1997) to understand the process of acquisition of new competitive advantages y the firm, inserted in fast-changing environments.

Agricultural technology-based startups (agtechs) play a central role in offering and disseminating technologies and innovations to producers, using new business models, and in interacting with educational institutions, research centers, investors, large corporations and other innovation support organizations.

Bambini and Bonacelli (2019) identified several Brazilian organizations that have a relevant role in agricultural innovation with national capillarity: Brazilian Micro and Small Business Support Service (SEBRAE); SENAI-SESI-IEL System, formed by the National Service for Industrial Training (SENAI), the Social Service of Industry (SESI) and the Euvaldo Lodi Institute (IEL); Federal Institutes of Education, Science and Technology; Research Centers of the Brazilian Agricultural Research Corporation (Embrapa), with strong regional influence; National Service for Rural Learning; and the Public Rural Extension Network, linked to state governments and present throughout Brazil.

Oliveira Junior et al. (2019) highlight that both technology-based entrepreneurship and innovation environments and leading Brazilian universities are located in the Southeast and South regions of the country, which are more industrialized and account for about 75% of the Brazilian Gross Domestic Product (GDP). Table 1 corroborates this information, based on the analysis of recent reports on the distribution of innovation environments, such as technology parks, and mechanisms for generating enterprises, as defined by Audy and Piqué (2016).

Table 2 presents the main cities that concentrate agtechs in Brazil, totaling 55% agtechs. The São Paulo cities listed represent 36.4% of the agtechs mapped. This concentration corresponds to the new dynamic region of the state of São Paulo, identified by Marighetti and Sposito (2009). The state of São Paulo concentrates 58% of startups mapped by Dias et al (2019).

[1] Evolutionary economics, developed in the 1980s based on the seminal works of Nelson and Winter (1977, 1982), was adopted by economic geography to better understand the geography of technical progress; the dynamics of competitive advantage; economic restructuring; and economic growth, according to Boschma and Martin (2010).

In terms of resources and innovation capabilities, the regions of Campinas, São José dos Campos, São Carlos and Ribeirão Preto stand out as development hubs related to Information and Communication Technologies (ICT) and Telecommunications. In relation to agrarian sciences, a recent study identified 154 Research and Education organizations of the state of São Paulo operating in this area, including agricultural Research Center (50%), public colleges (30%), technology colleges (15%) and private colleges (7%), with 30% of these institutions concentrated in the cities of Campinas, Piracicaba and São Paulo (Firetti et al., 2016).

Table 1. Percentage of technology parks, mechanisms for generating new enterprises and agtechs in the Brazilian regions.

Region	Technology park (%)	Incubator (%)	Accelerator (%)	Agtech (%)
Midwest	9.7	10.7	6.9	6.5
Northeast	8.7	16.8	12.0	3.4
North	5.8	8.5	1.7	1.1
Southeast	39.9	36.4	57.0	66.0
South	35.9	27.6	22.4	23.0

Source: Brasil (2019), Dias et al. (2019) and National Association of Entities Promoting Innovative Enterprises (2019).

The city of São Paulo concentrates 23% of agtechs. Even though it is a large urban metropolis, the concentration of agrarian technology-based startups is justified by the resources and innovation capabilities offered by the state capital, considered the largest center of innovation and entrepreneurship in Latin America. The culture of startups and entrepreneurship in the capital is rapidly emerging, according to Oliveira Júnior et al. (2019). A 2019 startup ecosystems ranking (StartupBlink, 2019) classified the city of São Paulo as the 23rd startup ecosystem in the world, the only one in Latin America ranked in the list of the 25 most relevant ecosystems, taking into consideration the number of startups, quality ecosystem and business environment. Table 3 presents an international classification of entrepreneurial ecosystems based on the survey by StartupBlink (2019).

Table 2. Main Brazilian cities where agtechs are located.

	City	Number of agtechs	State	Participation percentage (%)	Accumulated percentage (%)
1	São Paulo	262	SP	23.3	23.3
2	Piracicaba	41	SP	3.6	26.9
3	Campinas	38	SP	3.4	30.3
4	Ribeirão Preto	37	SP	3.3	33.6
5	Curitiba	36	PR	3.2	36.8
6	Rio de Janeiro	35	RJ	3.1	39.9
7	Porto Alegre	29	RS	2.6	42.5
8	Belo Horizonte	24	MG	2.1	44.6
9	Florianópolis	21	SC	1.9	46.5
10	Uberlândia	19	MG	1.7	48.2
11	Goiânia	17	GO	1.5	49.7
12	São José Dos Campos	17	SP	1.5	51.2
13	Londrina	15	PR	1.3	52.5
14	Campo Grande	14	MS	1.2	53.8
15	São Carlos	14	SP	1.2	55.0

Source: Dias et al. (2019).

The United States is emphasized in this ranking with several relevant ecosystems. The country is an important agricultural producer, and the sector aggregates more than two million companies, generating significant income and around 11% of the country's jobs (Australian Trade and Investment Commission, 2018). The Land Grant Colleges had the role of working with communities, developing new agricultural technologies with experimental stations and rural extension services. Lyons et al. (2018) consider that the transfer of technologies generated in universities and the sharing of resources, such as extension services and experimental stations, can contribute to develop the opportunities identified with agtech entrepreneurs, in addition to supporting the capture of investments.

Table 3. Startup Ecosystem Ranking in the year of 2019.

Ranking	Country
1	United States
2	United Kingdom
3	Canada
4	Israel
5	Australia
6	Netherlands
7	Sweden
8	Switzerland
9	Germany
10	Spain

Source: StartupBlink (2019).

The United States is responsible for 35% of the volume of venture capital for the agtech sector, according to AgFunder (2020). The Austrade report (Australian Trade and Investment Commission, 2018) highlights the following American states as important agtech clusters: California, North Carolina, Missouri, Colorado, and Illinois. The document also mentions emerging clusters: Minnesota, Indiana, and Wisconsin.

It is worth noting, in California, the ecosystem of Salinas, a city located in Monterey County, on the central coast of the state, approximately 100 km from Silicon Valley. The city acts as an important economic hub in the region, which has a relevant agricultural industry both in terms of production (with horticulture, production of strawberries and wines) as well as the presence of large agricultural production enterprises (Myrick; Deloffre, 2017). The resources established in the locality to support this project were: a) startup acceleration program; b) a business incubator, with work and collaboration spaces, as well as research initiatives applied to real cases; c) programs to encourage young entrepreneurs; and d) partnerships between secondary schools in the region to strengthen the training of young people.

Another prominent country in agricultural innovation is the United Kingdom, which ranks fourth in risk investments in agtechs, according to AgFunder (2020). The British innovation ecosystem attracts many entrepreneurs and investors and accounts for 45% of European venture investment in the agtech sector (AgFunder, 2019). Another important region in terms of agtech venture investment is Israel, the 5th country in terms of invested resources (AgFunder, 2020). Known as the "startup nation", Israel is a global innovation center with a culture based on interdisciplinary capabilities, technological skills and entrepreneurial spirit (Israel Innovation Authority, 2019).

Brazil has attracted the interest of investors and fostered its agricultural innovation ecosystem, based on the actions of several actors. As identified in the study by Dias et al. (2019), Brazil has 1125 agtech startups headquartered in its territory. Operating in various technological areas, startups play an important role in disseminating new technologies to agricultural producers, especially new digital tools.

Relationships established by Embrapa with ecosystem actors

Focusing on open innovation, Embrapa's innovation model seeks partnerships for the different stages of creating technological assets. In all types of partnerships, those carried out with agtechs, technology companies applied to agriculture at the initial or medium stage of maturity, have gained increasing prominence. The agtech ecosystem is considered fundamentally important, as it uses new operating concepts that have contributed to developing technological solutions capable of increasing the sustainability and competitiveness of Brazilian agribusiness, such as agile management; lean startup; gamification; self-managing teams; and others.

In several agricultural chains, the interaction between actors from science and technology institutions (STI), private companies, rural producers and consumers is still incipient, therefore each link operates individually. Ideally, each part could interact with the others, so that rural producers and consumers present their needs to the ICTs, related to technological research, and to private companies, in relation to its capacity to complement the development of solutions and the process of making them available to the producer and/or consumer market.

Thus, it is noticeable that the Brazilian innovation and entrepreneurship environment has changed rapidly in recent years, especially the strengthening of joint initiatives between private companies, startups, development agencies and risk fund managers (Venture Capital). When it comes to entrepreneurship and rapid growth of technology-intensive companies, Silicon Valley, California (USA), is the reference, and one of the key points in this process is the existence of financing sources for venture capital. The allocation of financial resources of this nature – venture capital – is essential for these companies to have the financial conditions to operate in the early stages of their innovation and development process.

The approach of entrepreneurs and research centers is enhanced with the inclusion of the financial link in the process of building innovative companies. Therefore, Embrapa interacts with Venture Capital companies in the agribusiness startup segment (AgriTechs), such as Cedro Capital, SP Ventures and NTagro, so that companies with Embrapa technology receive financial resources to accelerate their business.

One of the ways to expand this interaction and promote innovation ecosystems can be through actions aimed at the development and strengthening of startups. Some of the actions with the greatest impact are the innovation challenges, such as hackathons, demodays, business rounds, matchmaking events and bootcamps.

Interacting with different sources of knowledge is a fundamental condition for a company to innovate and incorporate new solutions. This movement in favor of innovation, exploring disruptive technologies, was also intensively supported at Embrapa, which has actively followed the initiatives developed by its Decentralized Units (UDs) and by its partners.

Some of the possibilities that Embrapa has accessed external knowledge and created new partnerships to implement its open innovation model, are the initiatives presented in Table 4.

During the events presented above, startups received mentoring from experts in agribusiness, technology and business; had opportunities to present their ideas to representatives of the production sector and investors, receiving feedback on their strengths and weaknesses; participated in awards and matchmaking actions with large companies in the production sector, innovation hubs, accelerators and seed and venture capital investors.

As a result, Embrapa and key companies in the agricultural sector had opportunities for growth, and contributed to increase the effect of technologies generated by institutions in the agricultural research sector, co-developed, adopted or in phase of adoption by the private companies installed in the country. In each innovation initiative, Embrapa established cooperation agreements for the development of technological solutions and assets, catalyzing open innovation and getting financial returns for the federal government, through royalty payments or profit sharing in the commercialization of the solutions created.

The open innovation actions undertaken by Embrapa supported the consolidation of the Brazilian agricultural innovation ecosystem through the interaction between companies, universities, agricultural research institutes and the productive sector, through the presentation of new technological solutions for the promotion of technology-based entrepreneurship in agriculture. In addition, in its 2019–2023

Table 4. Embrapa's open innovation initiatives.

Event	Objective
Avança Café	Avança Café is a pre-acceleration program for startups that aims to encourage the development of technological solutions for the coffee sector.
Caminho Startups Seminar	Seminar to discuss opportunities and challenges for startups in São Carlos, with the presentation of startups linked to Embrapa Instrumentation to the companies AgroRobótica and Fine Instrument Technology (FIT).
Soja Open Innovation	Open Innovation Soybean is a public call for the selection of startups interested in developing open innovation projects having Embrapa Soybean as a technical partner for the development or improvement of solutions in areas that adhere to priority research lines indicated in the guidelines of the call.
TechStart AgroDigital	TechStart Agro Digital is an acceleration program created by Embrapa Digital Agriculture and Venture Hub®, with support from the National Association of Entities Promoting Innovative Enterprises (Anprotec), to help startups, large companies and institutions to accelerate businesses and technologies for the agribusiness.
Ideas for Farm	Ideas for Farm is an innovation challenge that seeks technological solutions for Brazilian agribusiness, focusing on the Mid-North region of Brazil.
Pitch Deck AgTechs	The Pitch Deck refers to a quick and visual presentation used to attract the attention of investors and show the public the main differentials in the food, environment, waste management, pest control, phenotyping and livestock solutions segment.
Ideas for Milk	Ideas for Milk is a startup challenge that creates opportunities for young entrepreneurs to validate and present their ideas and solutions, connecting investments from large corporations that value innovation and that boost incorporating digital technology into the world of milk. The objective is to increase the level of innovation in the milk chain, increasing efficiency from the farm to the relationship with the final consumer, with respect for animals, the environment and society in general.
Vacathon	It is a hackathon whose objective is to debate ideas for the development of software and hardware aimed at solving problems in the milk production chain.
InovaPork	InovaPork is the first challenge of ideas in pig farming. The objective is to foster meaningful innovation in pig farming and attract innovative individuals with ideas at any stage of maturity, helping them to become businesses and solutions for the swine production chain.
InovaAvi	InovaAvi is the first ideas challenge in poultry farming. The proposal is to foster meaningful innovation in poultry farming and attract innovative individuals with ideas at any stage of maturity, helping them to become businesses and solutions in the poultry production chain.
Camp de EcoInovação Agrotech	It is a challenge of ideas/startups focused on eco-innovation, promoted by the UN Environment, SEBRAE and Embrapa, which seek eco-innovative solutions for agribusiness. In the first edition, the challenge was for the grain chain; for the next edition, the theme will be "food waste".
Gado de Corte 4.0	The Beef Cattle Event 4.0 was an innovative action for the beef cattle chain in Brazil. Based on real demands identified together with companies in the chain, it promoted a call for proposals, open to startups and Science and Technology Institutes interested in working for the chain.
Pontes para Inovação	Pontes para Inovação is a public call developed in partnership between Embrapa and Cedro Capital, which aims to connect agritechs with investors, partners and customers, in order to allow them to have access to resources that could accelerate their business.
Hackathon Embrapa	The National academic Hackathon Embrapa is a contest for the participation of teams of students and graduates, with the objective of choosing the best technological solutions in the development of mobile applications, hardware solutions, Internet of Things (IoT) solutions, educational pieces or games, with a focus on technological innovation of agricultural interest.

Continue...

Table 4. Continuation.

Event	Objective
Agritech Semiárido	It is an innovation challenge conducted in order to promote the development of innovative solutions through startups for agribusiness problems facing the Brazilian semiarid region, promoting mentoring with experts in agriculture, technology and business and enabling the connection with the production sector.
Inova AgroBrasília	Inova AgroBrasília is the first challenge of technological solutions carried out by the Department of Agriculture of the Federal District (DF), Emater-DF, Embrapa, AgroBrasília and Coopa-DF. Its objective is to attract business men, academics or entrepreneurs with innovative ideas at any stage of maturity and collaborate for these ideas to become businesses with the potential to solve problems experienced by the sector.
InoveAqua	The purpose of InoveAqua is to provide a favorable environment for transferring knowledge to university students, to the community and to the professionals in the areas that are related to the many segments of the aquaculture chain. It aims to develop skills and promote innovations for the development of Brazilian aquaculture, contributing to increase the production and provide enhanced competitiveness, sustainability and innovation in the production chain.
Horta & Escola	The purpose of this contest is to promote a competition between elementary, secondary and technical education students from schools in the Federal District and surrounding cities of the state of Goias, inspiring them to work as a team and to create businesses, processes, products, services and innovative solutions, with social and economic impact. Therefore, it seeks to promote the practice of innovation and the dissemination of an entrepreneurial culture.

Source: Embrapa (2020).

business plan, Embrapa continues to prioritize the goal of implementing 25 innovation initiatives, nine of which have already been carried out in 2019.

It is noteworthy that the actions carried out by Embrapa were recognized within the federal government, with led to the inclusion of an Embrapa representative in the National Committee of Support Initiatives for Startups, created by Decree No. 10,122/2019, to articulate the Executive Power actions for innovative startups.

Relationships of Embrapa Digital Agriculture with the innovation ecosystem

Embrapa Digital Agriculture, one of Embrapa's Decentralized Unit, has the mission of enabling research, development and innovation solutions in digital agriculture, which has been heavily demanded in the last five years by various actors in the agricultural innovation ecosystem. Thus, in order to disseminate and potentialize its research, development and innovation initiatives in the production sector, this Research Unit has traditionally established collaborative research contracts with educational and research institutions and private companies. Figure 1 presents two graphs with the percentage of agreements, licensing and non-disclosure terms with public and private institutions over the last five years. Forty four technical cooperation agreements were signed with public institutions and 28 software licenses were granted, in particular the Ainfo software for library management. With a greater insertion of Embrapa Digital Agriculture in the innovation ecosystem, through different initiatives such as hackathons, participation in innovation programs, organization of workshops and fairs, among others, and consequently larger exposure of its research lines, there was an increase of private companies' interest in

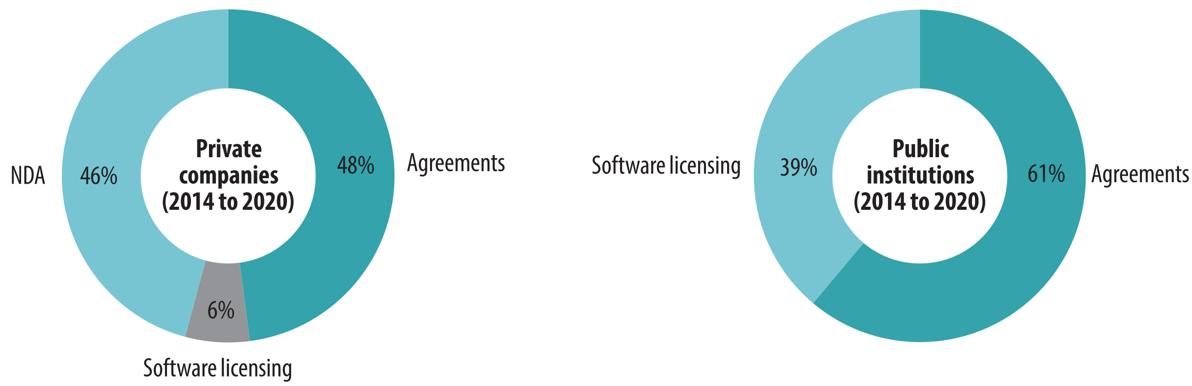


Figure 1. Percentage of legal instruments signed between Embrapa Digital Agriculture and public institutions and companies over the last 5 years.

partnerships. Negotiations with private institutions are usually preceded by Confidentiality Contracts or Non-Disclosure Agreement (NDA). Over the past few years, 22 NDAs were signed, many of which led to an Agreement, totaling 23 at the beginning of 2020.

As shown in Figure 2, the number of agreements, NDAs and software licensing has increased over the past few years as a result of the Unit's insertion into the innovation ecosystem. As 2020 could not be fully accounted for, it was not included in the graph shown in Figure 2. However, until May 2020, six cooperation agreements, five NDAs and one software licensing were signed.

Embrapa Digital Agriculture has established partnerships with institutions throughout the national territory, with emphasis on the Southeast and Midwest regions, as shown in Figure 3.

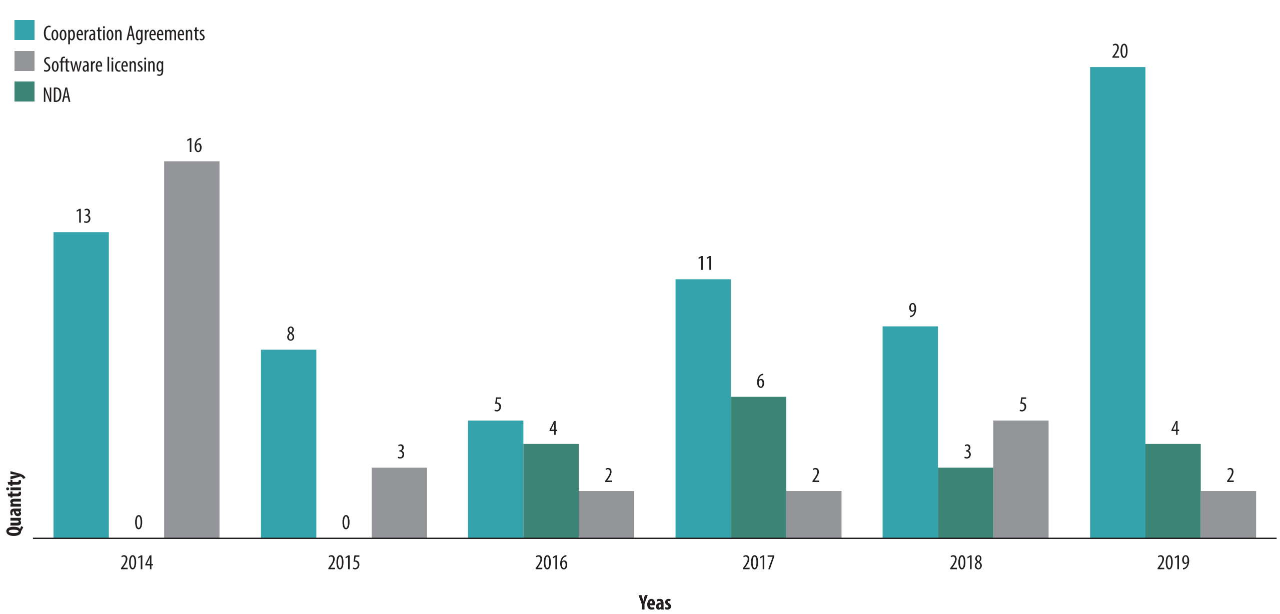


Figure 2. Legal instruments signed by Embrapa Digital Agriculture from 2014 to 2019.

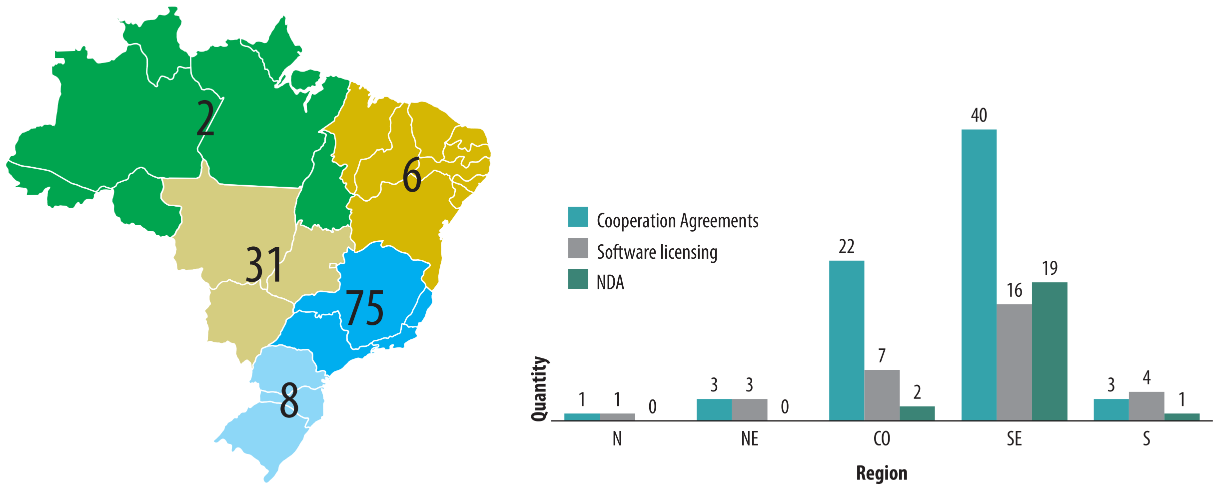


Figure 3. Research Center partnerships in the national territory.

Embrapa's contribution to adopting and evaluating the effect of technologies in agriculture

As a public agricultural research institution with a mandate for the entire Brazilian territory, Embrapa receives most of its resources from the federal government. Every year the Company evaluates the adoption rates and effects of the technologies and innovations generated, considering its various stakeholders.

The concern regarding the availability of systematized information about the adoption and impacts of technologies generated by Embrapa, mainly in the economic and social dimensions, dates back to the first edition of the Company's Social Balance, published in 1997. In its introduction, Alberto Duque Portugal, the president at the time, believed that the various experiences of rural development carried out by Embrapa, many of them in partnership, led to positive social impacts through the adoption of technologies developed and transferred to society, benefiting the entire country, which he called social profit.

This process was developed with the y training the Company's technical staff associated to methodological developments. In 2008, a reference methodology called Evaluating the Effects of Technologies Generated by Embrapa was published. Currently, all the Company's Decentralized Research Centers follow the same impact assessment model, based on the one proposed in 2008, but with the necessary adjustments to the present context. The dimensions currently analyzed are: a) economic; b) socio-environmental; c) employment; and d) institutional development.

Every year, all of Embrapa's Research Units report the impacts of their main technologies, and this effort to systematize information is consolidated in a Social Balance, which is published in digital media (internet) and printed as mechanism for accountability and transparency. From these data, the relationship between social profit and operating income is calculated, which generates an index that shows the return that each Brazilian Real invested in the Company offer to society. Over the last 2 years, according to data reported in the 2018 and 2019 in Embrapa´s Social Balance Reports, these rates were higher than USD 12.00 returned for each dollar invested. This data gains importance as the federal government invests billions of dollars in the Company, that is, this large capital contribution is remunerated by a counterpart

12 times greater than the amount invested. Thus, social profits in 2018 and 2019 were USD 8,44[2] billion and USD 9,02 billion, respectively.

In terms of the federal administration, the Impact Assessment Reports measure the Company's effectiveness, since the disclosure of its results contributes to promoting its accountability for control organizations, becoming an important indicator of Embrapa's operational viability.

As one of the 42 Decentralized Units of the Company, Embrapa Digital Agriculture is also involved in the preparation of impact assessment reports. As a reference center in agroinformatics for the entire Company, it develops digital technologies applied to various problems in Brazilian agriculture and livestock. This variety of themes is reflected in the impact reports of the technologies generated by this research center over the last ten years. Table 5 presents each reported technological asset, its respective evaluation period and the theme related to the initiative.

Table 5. Technologies analyzed by Embrapa Digital Agriculture through impact assessment reports over the last 10 years.

Technology	Period analyzed	Site	Theme associated with technology	Economic impact (BRL[(1)]) base year 2019
Computerized System for the Management of Printed and Digital Library Collections (Ainfo)	2012–2019	www.ainfo.cnptia. embrapa.br	Library management and knowledge availability	15,054,680.91
Embrapa Information Agency (Ageitec)	2010–2019	www.agencia. cnptia.embrapa.br	Availability of knowledge	6,279,104.22
Vegetation Temporal Analysis System (SATVeg)	2019	www.satveg. cnptia.embrapa.br	Geotechnologies	1,057,989.27
Interactive Environmental Licensing Support System (SISLA)	2013–2018	www.sisla.imasul. ms.gov.br	Geotechnologies	1,564,350.84 (data from 2018)
Agrometeorological Monitoring System (Agritempo)	2014–2019	www.agritempo. gov.br	Agrometeorology	2,491,920.01
Agricultural Climate Risk Zoning (ZARC)	2017–2019	-	Methodology applied to public policy	4,661,047,163.73
Virtual Diagnosis	2010–2012	www.diagnose. cnptia.embrapa.br	Remote diagnosis of plant diseases	332,941.00 (2012 data)
Free Software Network for Agriculture (Agrolivre)	2010–2011	www.agrolivre. gov.br	Free software repository	118,156.02 (2011 data)

[(1)] Average dollar exchange rate for 2019: USD 1 = BRL 3.9233.

[2] Average dollar exchange rate for 2020: USD 1 = BRL 5.1575.

Developing the agricultural innovation ecosystem in the state of São Paulo

Embrapa Digital Agriculture has been developing important actions in its surroundings in order to strengthen the local and state agricultural innovation ecosystem. This section details the initiatives developed in the state of São Paulo, considering the type of actor involved and the form of relationship established.

The state of São Paulo is responsible for about 32% of the Brazilian Gross Domestic Product (GDP). In 2019, agribusiness represented 12% of the state's GDP. The city of São Paulo stands out for being the largest innovation and entrepreneurship center in Latin America, concentrating a large amount of resources and innovation capabilities, such as research centers, universities, technology parks, business incubators, open laboratories, innovation hubs, coworkings and various events dedicated to technology, entrepreneurship and innovation. Furthermore, it concentrates the headquarters of numerous corporations, financial institutions and risk investors.

In addition to the capital, several cities in the state have important innovation ecosystems, especially Campinas, the third city in São Paulo in terms of population and GDP, where Embrapa Digital Agriculture is located. The Metropolitan Region of Campinas, which includes the city of Campinas and 20 surrounding municipalities, corresponds to 18% of the state's GDP and about 7% of its population (Sistema Estadual de Análise de Dados, 2020). In recent years, the region has gained and consolidated an important economic position in the state and national scenarios, concentrating many technological industries, technological and scientific research centers, as well as private universities and colleges, an important structure for agricultural research and a significant agro-industrial production (Agência Metropolitana de Campinas, 2020).

The strengthening of an agricultural innovation ecosystem involves relationships, partnerships and interactions established with new technology-based companies, such as the acceleration program for agricultural technology-based startups called TechStart AgroDigital (TSAD). Created in partnership with Venture Hub, accelerator and creator of new businesses, and the National Association of Entities Promoting Innovative Enterprises (Anprotec), the program is based on processes for identifying, selecting and offering support to innovative agribusiness ventures (startups) for a period of 6 months, offering them various development activities. The program had eight topics of interest a) biotech; b) precision livestock; c) field automation and robotization; d) nutrition and animal health; e) identification and detection of pests and diseases; f) agricultural risk management; g) fruit and vegetable chain; and h) management and monitoring of water, soil and plants. During the program, which took place in the 2nd semester of 2019, technical and business mentoring were conducted to support the development and validation of selected technologies. In the first cycle of the TSAD, more than 90 startups signed up and went through a process that selected 13 startups to participate in the program, 11 of which graduated in early 2020. The program contributes to solving several problems in the agribusiness production chain, meeting the expectations of customers, beneficiaries and users in the program's eight themes.

It is understood that the TSAD, due to its closer relationship with the selected startups, enabled Embrapa to gain greater knowledge and agility to interact with this new type of actors (startups), as well as with an accelerator for new business. In addition, the program represented an institutional strengthening opportunity for Embrapa, making it better known in the Brazilian agricultural innovation ecosystem and increasing its role in the agricultural digital transformation scenario. In this program, colleagues from other Embrapa Research Centers actively participated in mentoring or in the suggestion of content and lectures.

They are: a) Secretariat for Innovation and Business (SIN); b) Embrapa Agricultural Instrumentation (São Carlos); c) Embrapa Maize & Sorghum (Sete Lagoas); and d) Embrapa Soybean (Londrina).

Other recent actions related to the interaction with startups were:

- In 2017, at the *Open Innovation Business Round – 100 Open Startups and 100 Open Techs* – the Embrapa Digital Agriculture´s evaluation team was recognized as one of the most demanded by participating startups, helping 14 startups in a single day of the round, held in Campinas.
- Embrapa Digital Agriculture has contributed with ongoing mentoring offered by the São Paulo Research Foundation (FAPESP) to companies in its Small Businesses Innovative Research program, called PIPE. Some of PIPE's objectives are: a) to support research in science and technology as an instrument for technological innovation, development and business competitiveness of small companies; and b) enable the association between companies and academic researchers in research projects aimed at technological innovation. FAPESP offers training to the participating companies to improve their business models and therefore counts on the mentors' support. Embrapa Digital Agriculture employees have been mentoring the program since 2018.
- The Unit has supported, since 2018, the pre-acceleration program for digital startups (Startup SP) carried out by the SEBRAE-SP, which supports the development of innovative companies that use software or information technology services as the central point of its business model. In Piracicaba, the program's focus has been on startups linked to the agribusiness value chain. During the program, companies participate in workshops, seminars, individual and collective mentoring, and have the opportunity to interact with investors and accelerators – activities that help validate the product or service developed and its arrival on the market. The program takes place over four months, from April to July, and Embrapa Digital Agriculture participates in the mentoring stages with a team composed of researchers and analysts in the areas of Technology Transfer (TT) and Research and Development (R&D).
- Embrapa and its partners organized and carried out, during the *XI and XII Brazilian Congress of Agroinformatics* (SBIAgro), in 2017 and 2019, the SBIAgroConect@, in order to promote interaction and the formation of qualified market networking between institutions, ICT companies, users, accelerators, investors and developers related to the subject of data science and digital agriculture. The dynamics involved lectures by institutions and companies, presenting innovation initiative programs and relationship conversations, in order to enable integrated, advanced or differentiated solutions to be offered to the market. In 2017, the event had 100 participants and achieved the objective of promoting an environment for connections and networking qualified in technological solutions for agribusiness.
- In 2017, Embrapa Digital Agriculture promoted the XI SBIAgro Innovation Challenge, held at the State University of Campinas (UNICAMP). The challenge included encouraging young students and professionals to develop innovative technological solutions, in the form of mobile apps, aimed at solving problems faced by Brazilian agriculture. Teams of up to five members submitted proposals within the theme "Data Science in the Age of Digital Agriculture," the same as the conference. Each proposal involved a technology solution implemented in a mobile app, a one-page article describing it, and a video of up to 120 seconds.

The following criteria were considered by the evaluation committee: a) relevance of the problem to be solved; b) design quality; and c) correct functioning of the application. At the end of the event, seven proposals were classified to participate in the final stage, the Pitch Competition. It can be said that the SBIAgro 2017 Innovation Challenge promoted the approximation of agroinformatics research with real problems in agriculture, providing the teams' interaction with a judging panel composed of professionals, and raising the interest of investors in the presented technological solutions, which involved the following themes: a) classification of pests that attack crops; b) agroclimatic zoning; c) forecast of banana harvest and crop profitability; d) evaluation of the environments thermal conditions; e) monitoring wild boar spotting; f) identification of fruit fly species; and g) assessment of animal welfare during beef cattle transport. The articles presented by the teams were published in the annals of the event.

Embrapa Digital Agriculture has participated in Agropolo Campinas-Brasil, an initiative that started in 2015, to carry out projects in order to promote the development of the bioeconomy in the region of Campinas, enabling the proximity of other research institutions with the production sector. The following participated in Agropolo Campinas-Brasil: a) Secretariat of Agriculture and Supply (SAA) of the state of São Paulo, through the Agronomic Institute of Campinas (IAC); b) Institute of Food Technology (ITAL); c) Biological Institute (IB); d) Institute of Animal Science (IZ); e) State Secretariat for Economic Development, Science, Technology and Innovation; f) UNICAMP; g) Municipal Government of Campinas; h) Associtech Techno Park Campinas; and i) Agropolis International Association, mediated by Embrapa. Based on the concept of "collaborative innovation", as an inter-institutional platform, Agropolo Campinas-Brasil started its activities by promoting a series of work meetings, through workshops and events, for the selection of thematic areas that are the focus of actions, with the participation of representatives of Embrapa Digital Agriculture. With the approval of the project Agropolo Campinas-Brasil: Roadmap for Identification of Strategic Research Areas for the Creation of a World Class Bioeconomic Ecosystem, financed by FAPESP, a new plan of activities was started, with the promotion of workshops on several topics related to bioeconomy and agriculture, between 2016 and 2018.

Another important relationship and communication initiative in the Metropolitan Region of Campinas is the participation of Embrapa Digital Agriculture in the Inova Campinas (Tradeshow) event, occurring more intensively in 2017. The InovaCampinas event is promoted by the Campinas Forum Innovative Foundation (FFCi) during 2 days, bringing together companies, startups, research institutions, universities, incubators, accelerators and science and technology parks in the same space. The goal of the event is to present the technological potential of the region and new trends and initiatives of the ecosystem to the public, as well as promoting interactions with business roundtables and networking among the participants. Embrapa Digital Agriculture participated in 2018, with its own stand at the event, promoting presentations of its own technologies, AgTech pitches and a meeting of Biotech Hacking Campinas, a professional group for exchanging information on biotechnology, an initiative of the Venture Hub accelerator. In 2019, there was a shared stand called Inova#Agro, with the participation of partners such as Bayer and Venture Hub. Several accelerated startups in the TechStart Agro Digital Program participated by speaking about their solutions and technologies and interacting with the entrepreneurs, investors and researchers present at the event. Participating in InovaCampinas became a way to create an opportunity to connect Embrapa Digital Agriculture with startups from various areas, small and medium-sized companies, investors and other professionals, in order to communicate their activities and to establish new relationships.

Innovative public-private partnership models for digital innovation in agriculture

Faced with a scenario of digital transformation, including in agriculture, there is an urgent need to propose new business models. In this context, the AgroAPI initiative was created by Embrapa Digital Agriculture to boost value creation in agriculture by offering data and services via APIs. The acronym

API stands for "Application Programming Interface" and is often translated as "application programming interface," defined as specifications that govern the interoperability between applications and services (Vukovic et al., 2016). Therefore, they are deemed fundamental in the process of digital transformation in organizations, as they facilitate the integration of information systems, reducing cost and time.

APIs represent a set of patterns and programming languages that allow, in an automated way, the communication between different systems. Although invisible to the common user, they are responsible for the operation of several resources in mobile applications, e-commerce sites and social networks, among other market niches.

Due to the great demand for Embrapa's technologies and intelligence by public and private partners, the AgroAPI platform was conceived as an innovation and business strategy focused on the market of technologies in digital agriculture, enabling information and models generated by Embrapa to be accessed through APIs, in an agile, reliable and wide-ranging way, given that the same API can be useful for numerous purposes and customers, thus allowing the creation of solutions to support decision-making in the field, in real time.

The WSO2 API Manager tool has been used for the management of AgroAPI APIs (Vaz et al., 2017). The main components of the tool are: a) API publisher: user interface for API creators to develop, document and version APIs; b) API store or developer portal: collaborative interface for developers to host and publish APIs for consumers to use in a secure, protected and authenticated way. The portal is used for users to register, discover and evaluate the APIs, as well as to register to use them; c) API gateway: protects, manages and schedules calls to APIs; and d) other components for key management, traffic management and data analysis (WSO2 Inc., 2017). The AgroAPI Platform was launched in 2019 with two APIs initially published for use by external partners: API Agritec and API SATVeg.

The Agritec API arranges useful information for managing the production of agricultural crops and is based on the web version of the service called WebAgritec (Massruhá et al., 2008). The API includes the offer of data and models about the ideal planting time for dozens of crops, based on agricultural zoning of climatic risk; the list of the most suitable cultivars, for 12 different crops; the indication of fertilization and soil correction, for five crops, according to the previous soil analysis result; the yield forecast, also for five crops; and about the climatic conditions before and during harvest (water balance). The information provided by API Agritec can be used in solutions to support decision-making in the planning, monitoring and management stages of agricultural production. The distinct data made available by API can benefit farmers, cooperatives, representatives of technical assistance and rural extension and other agents, such as banks and insurance companies (Embrapa, 2020a).

However, SATVeg API is derived from the Temporal Analysis of Vegetation System (SATVeg) (Esquerdo et al., 2016), a web tool developed by Embrapa Digital Agriculture for generating and displaying temporal profiles of vegetation indexes of the Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) for Brazil and all of South America, with the objective of supporting territorial management and agricultural and environmental monitoring activities. The vegetative indices are provided by the MODIS sensor coupled to NASA's Terra and Aqua satellites, and include data produced from 2000 until the last date made available by its official repository. The NDVI and EVI indices are correlated with biophysical variables, such as leaf area and biomass that can indicate the presence and vigor of vegetation in a given area of interest. The time series of these indices allow monitoring, over time, the behavior of vegetation in these locations. Thus, it is possible to identify what is an urban area, annual planting, sugar cane, pasture or forest, for example, in addition to monitoring the cycle of an agricultural crop and deforestation and reforestation processes. The data made available can be used for activities related to environmental mapping and monitoring and for evaluating agricultural production, as well as verifying losses (Embrapa, 2020a).

As of May 2020, the Agritec API has been signed by 274 customers, and more than 111,300 requests were made, while the SatVeg API was signed by 118 customers and more than 1,700 requests made. To enable conducting business with commercial exploitation of APIs with a monetization profile, a marketing plan and business model were elaborated for the provision of services through the AgroAPI platform. This model comprises some legal instruments that also involve a research support foundation.

The AgroAPI platform facilitates integrating information systems, with cost and time reduction, improves the interface with mobile devices, expands the ability to obtain and disseminate agricultural data and information, enables savings in computational resources and sharing data and services, facilitates establishing agreements with other organizations and enables greater reach of the results of the company and its partners (Vaz et al., 2017). The currently available APIs and the partnerships signed so far have demonstrated that this strategy benefits numerous partners and, consequently, the end customers, contributing to the solution of real agricultural problems. The platform's expansion plan is under development, with new APIs to be published based on prospecting demands.

Final considerations

This chapter describes the characteristics of the new ecosystem of Brazilian agricultural innovation, specifically presenting the case of the state of São Paulo, emphasizing Embrapa's performance in this scenario. The actions developed by Embrapa Digital Agriculture to consolidate this ecosystem were described in the chapter, highlighting the actions developed in Campinas.

This chapter emphasized the strategies carried out to establish and strengthen relationships with the actors in the segment. The promotion of events has been important to externally present to Embrapa the challenges identified in the Brazilian agricultural sector and to promote the search for results together with students and entrepreneurs.

In this regard, the promotion of business roundtables with companies and startups is relevant to the innovation challenges. Organizing longer programs, aimed at AgTech startups, is another very important line of action in order to insert Embrapa in this context. These programs – focused on various agricultural production chains – offer startups several possibilities such as pre-acceleration, acceleration, establishing partnerships and greater exposure and dissemination of their projects to relevant actors. In these programs, Embrapa has partnered with venture capital companies, startup accelerators, government agencies, and others.

It is noteworthy that, as regards relationships, Embrapa Digital Agriculture has a tradition of establishing technical cooperation agreements, NDAs and technology licenses, legal instruments in order to regulate signed partnership initiatives regarding confidentiality, objectives, stages, duration, resources and expected results. These are partnerships with various actors in the agricultural innovation ecosystem, with emphasis on organizations located in the Southeast and Midwest regions.

The Unit has also worked to strengthen the ecosystem of agricultural innovation in its surroundings, within the state of São Paulo and, in particular, in the region of Campinas. With regard to partnerships with scientific research institutions, more traditional in the context of Embrapa, the scientific partnership to promote the innovation ecosystem in the Metropolitan Region of Campinas, Agropolo Campinas-Brasil, which started in 2015, stands out. The action was led by the Department of Agriculture and Supply

(SAA) of the state of São Paulo, through various agricultural research institutes in the state, such as the Agronomic Institute of Campinas (IAC), mediated by Embrapa, with the participation of local actors such as the Municipality of Campinas; the UNICAMP, Associtech Techno Park Campinas; and the Agropolis International Association. Several events on topics related to bioeconomy and agriculture were promoted between 2016 and 2018. This action enabled establishing relationships between Embrapa Digital Agriculture and municipal actors, promoting the relationship of the agricultural innovation ecosystem in Campinas.

A highlight in 2019 was the promotion of the TechStart AgroDigital (TSAD) program, in partnership with Venture Hub and Anprotec, aimed at accelerating the startups registered, which received more than 90 applications. It is worth pointing out that the closer interaction of Embrapa Digital Agriculture´s team with the universe of agtechs, investors and the acceleration environment enabled greater knowledge in this context, requiring more interactive flexibility with this type of actor and with the acceleration process of startups. Other actions involving mentoring for startups, both under FAPESP's PIPE and SEBRAE-Piracicaba programs, and business roundtables established at the 100 Open-Startups (2017) and SBIAgroConect@ (2017 and 2019) were also instrumental to bring the Embrapa Digital Agriculture teams closer to the business environment and, in particular, to startups.

Participation in the *Inova Campinas* event (Tradeshow), since 2017, proved to be very important to strengthen the image of Embrapa Digital Agriculture in the context of the Metropolitan Region of Campinas, presenting the Company and its technological potential, as well as providing opportunities for interactions with companies, organizations, investors and the press. All these events and local programs offered a new perspective on the startup environment for Embrapa teams, influencing culture and behavior, contributing towards strengthening the presence of the Unit in the agricultural innovation ecosystem of Campinas and São Paulo, as well as increasing its leading role in the area of digital agriculture, which is a technological driver of the Campinas ecosystem.

In this context, it is noteworthy that Embrapa Digital Agriculture has been demanded by companies from different segments, with emphasis on information technology, in order to establish partnerships for collaborative development and product validation. The AgroAPI platform is an example of technology designed to promote value creation in agriculture by offering data and services via APIs. The business model is based on the use of this technology (API) to make Embrapa data, information and models available to its partners with an agile, reliable and comprehensive approach.

Embrapa's actions aimed at strengthening the Brazilian agricultural innovation ecosystem are expected to continue, which were prioritized in its 2019–2023 Business Plan, the organization's medium-term strategy.

References

AGÊNCIA METROPOLITANA DE CAMPINAS. **RMC**: o que é. Available at: http://www.agemcamp.sp.gov.br/rmc. Accessed on: 27 May 2020.

AGFUNDER. **Agrifood tech funding report**: 19´ year in review. 2020. Available at: https://agfunder.com/research. Accessed on: 20 Mar. 2020.

AGFUNDER. **AgriFood tech**: 2018 in review. 2019. Available at: https://agfunder.com/research. Accessed on: 29 Apr. 2019.

ARTHUR, W. B. **Increasing returns and path dependence in the economy**. Ann Arbor: University of Michigan Press, 1994. p. 49-67. DOI: 10.3998/mpub.10029.

AUDY, J.; PIQUÉ, J. **Dos parques científicos e tecnológicos aos ecossistemas de inovação**: desenvolvimento social e econômico na sociedade do conhecimento. Brasília, DF: Anprotec, 2016. 26 p. Available at: https://anprotec.org.br/site/publicacoes-anprotec/ebooks. Accessed on: 20 May 2020.

AUSTRALIAN TRADE AND INVESTMENT COMMISSION. **Agtech**: advancing agribusiness and food processing - US Clusters. San Francisco: Austrade, 2018. 29 p.

BAMBINI, M. D.; BONACELLI, M. B. M. Ecossistemas Agtech no Brasil: localização, caracterização e atores envolvidos. In: WORKSHOP ANPROTEC; INNOVATION SUMMIT BRASIL, 2019, Florianópolis. **O futuro dos ambientes de inovação**: anais: chamada de trabalhos 2019. Brasília, DF: Anprotec, 2019. p. 789-802.

BOSCHMA, R.; MARTIN, R. Introduction - the new paradigm of evolutionary economic geography', the aims and scope of evolutionary economic geography. In: BOSCHMA, R.; MARTIN, R. (ed.). **The handbook of evolutionary economic geography**. Cheltenham: Edward Elgar, 2010. p. 3-42. DOI: 10.4337/9781849806497.

BRASIL. Ministério da Ciência, Tecnologia, Inovações e Comunicações. **Estudo de projetos de alta complexidade**: indicadores de parques tecnológicos. Brasília, DF: Centro de Apoio ao Desenvolvimento Tecnológico/UnB, 2019. 107 p.

CHESBOROUGH, H. The era of open innovation. **MIT Sloan Management Review**, v. 44, n. 3, p. 35-41 2003.

COOK, D. Innovation clusters and the rural powerhouse. **Blog Innovate UK**. 25 Jan. 2016. Available at: https://innovateuk.blog.gov.uk/2016/01/25/ innovation-clusters-and-the-rural-powerhouse. Accessed on: 6 May 2020.

DIAS, C. N.; JARDIM, F.; SAKUDA, L. O. (org.). **Radar AgTech Brasil 2019**: mapeamento das startups do setor agro brasileiro. Brasília, DF: Embrapa; São Paulo: SP Ventures e Homo Ludens, 2019. Available at: www.radaragtech.com.br. Accessed on: 30 Sept. 2019.

EDQUIST, C. Systems of innovation: perspectives and challenges. In: FAGERBERG. J.; MOWERY, D.; NELSON, R. R. **The Oxford book of innovation**. Oxford: Oxford University Press, 2006. p. 180-208. DOI: 10.1093/oxfordhb/9780199286805.003.0007.

EMBRAPA. **Ecossistema de inovação**. 2020. Available at: https://www. embrapa.br/ecossistema-de-inovacao. Accessed on: 30 Jun. 2020.

ESQUERDO, J. C. D. M.; ANTUNES, J. F. G.; COUTINHO, A. C.; KONDO, A. A.; SANTOS, J. L. dos. **Sistema de Análise Temporal da Vegetação - SATVeg**. Versão 3.0. Campinas: Embrapa Informática Agropecuária, 2016.

FELD, B. **Startup communities**: Building an entrepreneurial ecosystem in your city. Hoboken: John Wiley & Sons, 2012. DOI: 10.1002/9781119204459.

FELDMAN, M. P.; KOGLER, D. F. Stylized facts in the geography of innovation. In: HALL, B. H.; ROSENBERG, N. (ed.). **Handbook of the Economics of Innovation**. Oxford, IK: North-Holland, 2010. p. 381-410. DOI: 10.1016/S0169-7218(10)01008-7.

FETTERS, M.; GREENE, P. G.; RICE, M. P. (ed.). **The development of university- based entrepreneurship ecosystems**: global practices. Cheltenham, UK: Edward Elgar Publishing, 2010. 203 p. DOI: 10.4337/9781849805896.

FIRETTI, R.; OLIVEIRA, E. C. de; BONACELLI, M. B. M. Identificação e Mapeamento de Instituições de Ensino Superior e/ou Pesquisa Agrícola no Estado de São Paulo. **Blucher Engineering Proceedings**, v. 3 n. 4, p. 641-652, 2016. DOI: 10.5151/engpro-1enei-036.

GOLLEY, F. B. The ecosystem concept: a search for order. **Ecological Research**, v. 6, n. 2, p. 129-138, 1991. DOI: 10.1007/BF02347157.

GOMES, L. A. V.; FACIN, A. L. F.; SALERNO, M.; IKENAMI, R. K. Unpacking the innovation ecosystem construct: evolution, gaps and trends. **Technological Forecasting & Social Change**, v. 136, p. 30-48, 2018. DOI: 10.1016/j.techfore.2016.11.009.

ISRAEL INNOVATION AUTHORITY. **State of Innovation in Israel 2019**. Jerusalem: Israel Innovation Authority, 2019. Available at: https://innovationisrael.org.il/sites/default/files/ISRAEL%20 Innovation%20Authority%20report%202019%20eng_0.pdf. Accessed on: 5 May 2020.

LUNDVALL, B-A. **National systems of innovation**: towards a theory of innovation and interactive learning. London: Pinter Publishers, 1992.

LYONS, T. S.; MILLER, S. R.; MANN, J. T. A new role for land grant universities in the rural innovation ecosystem? **Journal of Regional Analysis & Policy**, v. 48, n. 2, p.32-47, 2018. Available at: https://jrap.scholasticahq.com/article/3775-a-new-role-for-land- grant-universities-in-the-rural-innovation-ecosystem. Accessed on: 30 June 2020.

MALECKI, E. Entrepreneurship and entrepreneurial ecosystems. **Geography Compass**, v. 12, n. 3, p.1-21, 2018. DOI: 10.1111/gec3.12359.

MALERBA, F. Sectoral systems: how and why innovation difffers across sectors. In: FAGERBERG, J.; MOWERY, D.; NELSON, R. R. **The Oxford book of innovation**. Oxford: Oxford University Press, 2006. p. 380-406. DOI: 10.1093/oxfordhb/9780199286805.003.0014.

MARIGHETTI, A.; SPOSITO, E. S. A. Formação dos polos tecnológicos e seu papel no processo de desenvolvimento territorial no município de São Carlos/ SP. **Geografia em Atos**, v. 1, n. 9, 2009. DOI: 10.35416/geoatos.v1i9.282.

MASSRUHÁ, S. M. F. S.; LIMA, H. P. de; EVANGELISTA, S. R. M.; PINTO, G. E. M. **Uma proposta de plataforma de software para integração e interoperabilidade de serviços Web - Webagritec**: estudo de caso. Campinas: Embrapa Informática Agropecuária, 2008. 23 p. (Embrapa Informática Agropecuária . Boletim de pesquisa e desenvolvimento, 19).

MYRICK, A.; DELOFFRE, R. Planting the seeds for an Agtech Innovation Ecosystem. **Economic Development Journal**, v. 16, n. 4, p. 5-10, 2017. Available at: https://search.proquest.com/openview/eac2ee702fb0e5b8a14ed9ba0416e950/1?pq-origsite=gscholar&cbl=44217. Accessed on: 20 May 2020.

NELSON, R. R. (ed.). **National nnovation systems**: a comparative analysis. New York: Oxford University Press,1993. 541 p.

NELSON, R. R.; WINTER, S. G. **An evolutionary theory of economic change**. Cambridge: Mass; London: Belknap Harvard, 1982.

NELSON, R. R.; WINTER, S. G. In search of a useful theory of innovation. **Research Policy**, v. 6, p. 36-76, 1977. DOI: 10.1016/0048-7333(77)90029-4.

OLIVEIRA JÚNIOR, M. de M.; CAHEN, F. R.; BORINI, F. M. (ed.). **Startups and innovation ecosystems in emerging markets**. Cham: Springer, 2019. p. 1-14. DOI: 10.1007/978-3-030-10865-6.

PIGFORD, A. A. E.; HICKEY, G. M.; KLERKX, L. Beyond agricultural innovation systems? Exploring an agricultural innovation ecosystems approach for niche design and development in sustainability transitions. **Agricultural Systems**, v. 164, p. 116-121, July 2018. DOI: 10.1016/j.agsy.2018.04.007.

RAJALAHTI, R. Sourcebook overview and user guide. In: WORLD BANK. **Agricultural innovation systems**: an investment sourcebook. Washington, DC: The World Bank, 2012. p. 1-13. DOI: 10.1596/9780821386842_FM.

SIMÕES, M.; SOLER, L.; PY, H. Tecnologias a serviço da sustentabilidade e da agricultura. **Boletim Informativo da SBCS**, p. 50-53, maio/ago. 2017. Available at: https://ainfo.cnptia.embrapa.br/digital/bitstream/item/167339/1/2017-044.pdf. Accessed on: 27 May 2020.

SISTEMA ESTADUAL DE ANÁLISE DE DADOS. **Portal de estatísticas do Estado de São Paulo**. Available at: https://www.seade.gov.br. Accessed on: 27 May 2020.

SOTARAUTA, M. **Strategy development in learning cities**: from classical rhetoric towards dynamic capabilities. [Tampere]: University of Tampere, 2004. 54 p. (SENTE working papers, v. 8). Available at: https://people.uta.fi/~atmaso/verkkokirjasto/sotarauta_strategy.pdf. Accessed on: 5 May 2020.

STARTUPBLINK. **Startup ecosystem rankings 2019**. 2019. 158 p. Available at: https://www.startupblink. com/blog/startup-ecosystem-rankings-report-2019-by-startupblink. Accessed on: 5 May 2020.

SUOMINEN, A.; SEPPÄNEN, M.; DEDEHAYIR, O. A bibliometric review on innovation systems and ecosystems: a research agenda, **European Journal of Innovation Management**, v. 22, n. 2, p. 335-360, 2019. DOI: 10.1108/EJIM-12-2017-0188.

TEECE, D. J.; AUGIER, M. The foundations of dynamic capabilities. In: TEECE, D. J. **Dynamic capabilities and strategic management**: organizing for innovation and growth. Oxford: Oxford University Press, 2009. p. 82-112.

TEECE, D. J.; PISANO, G.; SHUEN, A. Dynamic capabilities and strategic management. **Strategic management journal**, v. 18, n. 7, p. 509-533, 1997. DOI: 10.1002/ (SICI)1097-0266(199708)18:7<509::AID-SMJ882>3.0.CO;2-Z.

TEIXEIRA, C. S.; TRZECIAK, D. S.; VARVAKIS, G. (org.). **Ecossistema de inovação**: alinhamento conceitual. Florianópolis: Perse, 2017. 24 p. Recurso eletrônico.

VAZ, G. J.; APOLINÁRIO, D. R. de F.; CORREA, J. L.; VACARI, I.; GONZALES, L. E.; DRUCKER, D. P.; BARIANI, J. M.; EVANGELISTA, S. R. M.; ROMANI, L. A. S. AgroAPI: criação de valor para a agricultura digital por meio de APIs. 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. 59-68. SBIAgro 2017.

VUKOVIC, M.; LAREDO, J.; MUTHUSAMY, V.; SLOMINSKI, A.; VACULIN, R.; TAN, W.; NAIK, V.; SILVA-LEPE, I.; KUMAR, A.; SRIVASTAVA, B.; BRANC, J. Riding and thriving on the API hype cycle. **Communications of the ACM**, v. 59, n. 3, p. 35-37, 2016. DOI: 10.1145/2816812.

WSO2 INC. **WSO2 API manager documentation**: key concepts. Available at: https://docs.wso2.com/display/AM210/Key+Concepts. Accessed on: 25 May 2017.on **ecosystem and future perspectives**

Photo: Sebastian Duda (AdobeStock)

13

The law related to the digitization of agriculture

Cássia Isabel Costa Mendes | Silvia Maria Fonseca Silveira Massruhá | Juliano Souza de Albuquerque Maranhão | Paula Giovanna Guimarães Ribeiro | Líbia Cristina Xavier Santos

Introduction

The exponential technological evolution at the end of the last century and beginning of the 21st century, within a globalized economy, has produced profound changes in social relations. Such changes have brought new dynamics to capitalism, based on information and data processing. The phenomena behind these transformations are related to the intensification of the use of information and communication technologies (ICT), the democratization of access and production of information brought on by the internet, the growing relevance of analyzing large data volumes in the economy (Big Data), the dissemination of connected devices (internet of things) and artificial intelligent agents.

Modern-day world economy is strongly characterized by computational processing of information to generate knowledge, produce goods and services and generate value, changing the notion of wealth from a material asset to an intangible asset (Mendes et al., 2015).

The term "digital economy" emerged within the scope of these phenomena and is characterized by the central role of science, technological development and the use of digital technologies as leverage instruments for countries and economic agents to strategically and competitively position themselves in the international geopolitical scenario and in the market (Soares; Prete, 2018). The main factor of digital economy generating wealth is the transmission, processing and sharing of information. If at the initial stage the production of this information its processing was concentrated in large companies as the main productive agents, within a society of organizations, in the second stage, at the turn of the millennium, this production and consumption of information and data, from which new information is extracted, began to be decentralized, given the possibility of direct communication and economic peer-to-peer interactions through online platforms. Thus, the digital economy started to organize itself through the network society, which is defined by Castells (2006, p. 20) as a social structure based on networks operated by communication and information technologies based on microelectronics and on digital computer networks that generate, process and distribute information from accumulated knowledge.

Information and communication technologies, combined with a new social organization, enabled interconnecting factors on a global scale, with new production models, in which individuals can collaborate in common projects or directly establish relationships, reducing information costs through digital platforms, blurring the difference between production and consumption of information (Benkler, 2006). Given that the provision of technological infrastructure for communication became key, there was a strong increase in the value and economic power of providers, which began to mark the economy through a new model, known as "platform economy." In this model, each platform connects two groups of agents in a "two-sided market," offering low-cost or sometimes free services to one of the sides in order to collect data and process it to generate value to be marketed to the other side of the platform (example: social networks or search engines that collect data and content generated by free use, in order to economically exploit the advertising market).

This model centralizes the economy by collecting and processing data on one side of the platform, while generating intelligence to be exploited to obtain economic gains from the other side. Hence, massive investments in data analysis and artificial intelligence tools to increasingly encourage the use of the platform and enhance data collection in order to feed this value generation cycle.[1]

The network society, under which the new digital economy is structured, has produced profound institutional, economic, social, technological, cultural and behavioral changes, which raises questions about the role of regulations in relation to new types of conflict in this new economic order.

Such transformation in productive relations is reflected in the most varied economic sectors, including the agricultural sector. As evidenced in the preceding chapters of this book, there is extensive use of information and communication technologies in agriculture based on digital content. It is also observed that the data generated by technology consumption also generate value to its providers, which can increase the efficiency of their services, which is worthy, but can also create the concentration of power or questions about the autonomy of the producer, from which data is extracted. New legal implications arise in this new scenario, in view of the intervention of several new agent-based production activities, which are reflected in the sphere of copyright and civil liability, the protection of personal data, access to goods and services in consumer relations, as well as in labor. Thus, in a complementary and transdisciplinary approach, this chapter analyzes aspects of digital law in the context of digital agriculture.

Given the above, the objective of this chapter is to analyze digital agriculture from the perspective of the law. Thus, the chapter is divided into five sections, including this introduction and the conclusion. With thematic focus on the agricultural economic segment, the following section addresses some of the main elements that characterize digital agriculture to serve as a backdrop for the analysis carried out. The third section reviews the approaches to digital law and the stage of knowledge development related to the subject, in addition to tracing a global and Brazilian panorama and analyzing the legal implications in digital agriculture. With the presentation of the legal framework of digital law, the constituent elements of digital agriculture are addressed. The next section, with a more empirical approach, discusses the legal support related to Embrapa's performance – as a digital economic agent – regarding the application of its technologies from information to agriculture, presenting legal instruments that support Embrapa's business with digital assets, particularly those that regulate the relationship between Embrapa and its users, as well as the services provided by the company, through its websites and mobile applications. The end of the chapter presents the closing remarks by way of conclusion.

---

[1] Given the emphasis on data collection, often personal data, Zuboff (2019) calls this model "surveillance capitalism."

Digital agriculture: object of regulation by digital law

As shown in the preceding chapters of this book, there is an unprecedented agricultural technological revolution taking place.

The evolution of total factor productivity (TFP)[2] confirms the rise of the central role of technology in the growth of agricultural production and the reduced importance of land. According to Gasques et al. (2019), total factor productivity has been the main source of growth in agricultural production. The current phase of Brazilian agrarian development is characterized by a change in the pattern of accumulation in agriculture, as the role of land has decreased and the role of investment in technology, the use of knowledge and the application of capital has increased (Mendes, 2015).

Within the scope of innovative technologies applied to agriculture, the advancement of ICT in the rural area plays an essential role for the growth of agricultural production. ICT has contributed to several areas of knowledge, allowing the storage and processing of large volumes of data, the automation of processes and the exchange of information and knowledge (Massruhá, 2020).

Considering the relevance of agriculture to the country's economy, it needs to be able to absorb and use innovations and information technologies to increase the dynamic competitiveness of the agricultural sector (Mendes et al., 2014).

As described in Chapter 1, there is a continuous technological evolution in the field. Currently, a new era of agricultural technology is being consolidated, called digital agriculture, and 5.0 agriculture is in progress (intensive use of artificial intelligence tools – AI).

According to Cema (2017), Agriculture 4.0 is moving towards Agriculture 5.0. While agriculture 4.0 is characterized by the evolution of several technologies such as sensor networks, machine sensors, drones, satellite image processing, cloud-based information technology systems, analysis of large volumes of data (big data), mobile applications and autonomous tractors, the current threshold of 5.0 agriculture is primarily based on artificial intelligence, robotics, 3D and 4D printing, synthetic biology and vertical agriculture.

Therefore, Agriculture 4.0 has already paved the way for the next agricultural evolution, Agriculture 5.0, which consists of autonomous decision systems, unmanned vehicles, robotics and artificial intelligence (Cema, 2017).

It is observed, therefore, that one of the predominant characteristics of Agriculture 5.0 refers to the expanded use of artificial intelligence tools. artificial intelligence is a broad concept that encompasses studies on autonomous vehicles, machine learning, which gives the computer the ability to perceive the surrounding environment and identify patterns. The evolution of Agriculture 1.0 to 5.0 is represented in Figure 1.

As reported in the previous chapters, Embrapa Digital Agriculture is a public institutions and part of the agricultural innovation ecosystem that develops technologies to advance digital agriculture. However, despite the extended offer of digital technologies for agriculture by several public and private institutions – such as Embrapa – as well as the progress from Agriculture 4.0 to Agriculture 5.0, primarily characterized

[2] The concept of Total Factor Productivity (TFP) is defined as the relationship between the aggregate product and the inputs used in production (Gasques et al., 2019).

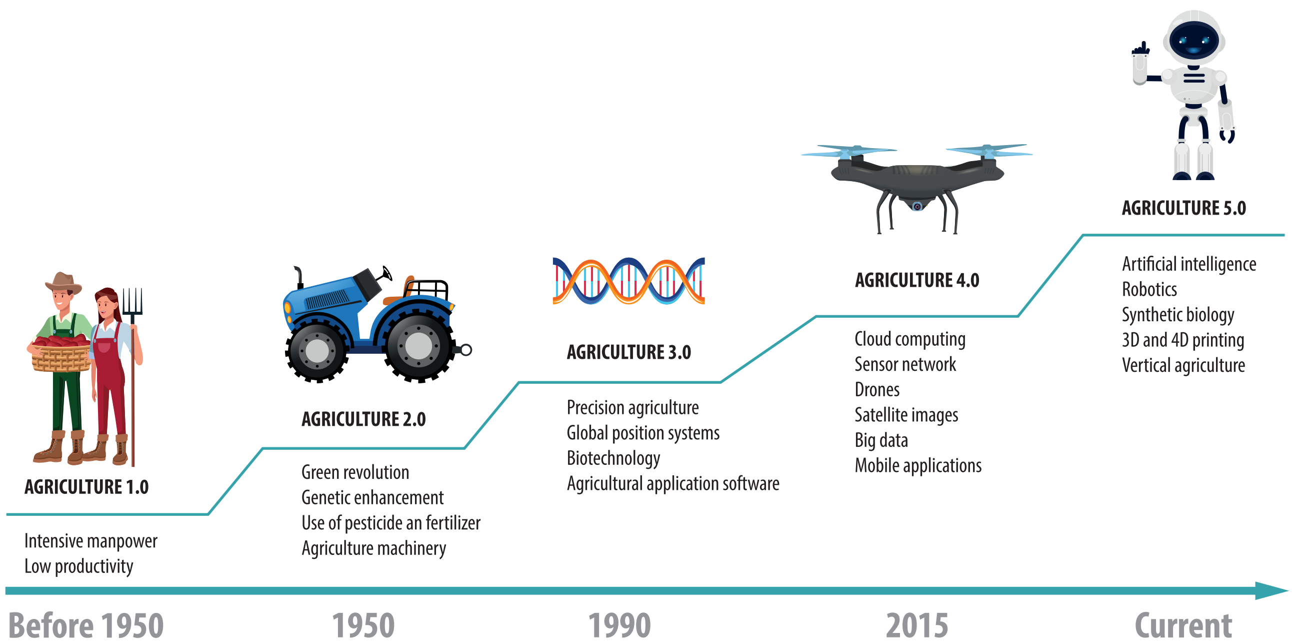


Figure 1. Evolution of Agriculture 1.0 to Agriculture 5.0.

Source: Adapted from Cema (2017), Melgar (2018) and Massruhá (2020).

by the intensive use of data, autonomous systems, unmanned vehicles, robotics and artificial intelligence, digital agriculture has also raised controversy. In a recent study by the European Union, Schimpf (2020) discusses some elements of these controversies, namely: the merging and market concentration of large agribusiness companies in digital agriculture; the social, ethical and legal implications of digital agriculture; and the need to define a legal framework to regulate the rights, ownership and privacy of agricultural data. Table 1 presents such controversies.

With regard to the capitalist movement of mergers and incorporations in the digital agriculture sector, Schimpf (2020) warns of a "digital arms race," possibly culminating in the control of digital tools by global agrochemical companies, as observed in the seeds and pesticides sectors.

The data is the fuel – or oil – of the 21st century. However, its added value depends on the ability to analyze, generate information and knowledge to support decision-making processes. In data-based digital agriculture, there are farmers who generate data from their farm, there are those who are able to collect and process the data through digital machines and devices, and there are those who are able to analyze the data, usually agribusiness companies. Therefore, the agricultural producers or farmers using smart machines generate data about their agricultural property – sometimes even personal data – and can maintain the rights to their data. However, adding value to data depends on analysis models to generate agronomic recommendations that can be marketed through digital platforms.

The study of the European Union on digital agriculture indicates the risk that the digitization of agriculture is driven only for profit and by the availability of tools and technologies, instead of being directed towards meeting specific demands identified in agriculture, in the environment or in society. The risk, pointed out in the literature, concerns the excess of economic concentration, fed back by a concentration of data, generating a competitive advantage for large agricultural groups that will barely be challenged by competitors or new agents.This can have important implications for agriculture and livestock and for the protection of natural resources and biodiversity, as the data holder can control food, agricultural producers and the rural area (Schimpf, 2020). Hence the need for action by antitrust

Table 1. Digital Agriculture: controversies raised in the European Community.

Dimension	Elements
Mergers and concentration in the digital agriculture market	
Monsanto and Bayer	The merger of Monsanto and Bayer (in 2018) will allow the companies to combine their digital agricultural acquisitions with their seeds, Genetically Modified Organisms and chemicals businesses, creating an unprecedented digital platform across the entire agricultural chain. Integration allows companies to extract data from agricultural producers and use it to drive their product choices, thus technologically dependent on the company's value chain. Create one-stop platforms, offering agricultural producers an inclusive package of services and decision-making guidance throughout the year.
John Deere and global seed and pesticide companies	John Deere (agricultural machinery company) is investing in digital agriculture. It has partnered with global seed and pesticide companies such as: Bayer/Monsanto, Syngenta/ ChemChina, Corteva (Dow, Dupont, Pioneer) and BASF. It developed its own platform for digital agriculture, automation and data.
Global companies investing in agricultural digitization	Cargill (mainly grain) invested in the digitization of the livestock sector, including dairy products. Companies from other segments invest in digital agriculture projects: Sony, Philips, Orange, Uber, Bosch, Siemens, Google and Microsoft.
Social, ethical and legal implications of digital agriculture	
Collect and store agricultural data	The risk of misusing the collected data. Anti-competitive practices, including price discrimination and commodity speculation. It can affect food security.
Yield and performance in agricultural data	Information related to crop yields and performance of crop or animal management contained in the collected data can provide a market advantage for the seed and fertilizer companies that own them. Agricultural data transmitted to large agribusiness companies can influence input prices.
Data rights, ownership and privacy	
Rules for use and access to agricultural data.	An agricultural group in the European Union has published a code of conduct to define the rights to use data. The code recommends license agreements between farmers as data owners and agribusiness companies. The agricultural producers must retain their right to decide who can access and use their data, including monetary compensation for its use.
Data protection and governance	Europe is close to allowing the centralization and concentration of data at an unprecedented scale, in the absence of any regulation. The power of large agribusiness companies to centralize and concentrate data is likely to give them decision-making power over agricultural producers throughout the entire production process, from seed to harvest. The large agribusiness companies that hold the data are in the central position of power, creating added value and earning a large part of the income generated in digital agriculture. In the absence of a legal framework for digital agriculture, weaker parties (farmers) will lose their data to platforms belonging to large corporation.

Source: Adapted from Schimpf (2020).

and data protection authorities to prevent abuses of market power or abuse in the collection and use of personal data beyond its intended purpose. To address these concerns, it is also effective to fill legislation gaps, by specifically regulating digital agriculture, taking into account proposals to mitigate the effects of economic concentration and data monopoly, in addition to directing the use of data and artificial intelligence for jobs that are socially beneficial, while ensuring ownership, data governance, and the privacy of farmers/agricultural producers.

Therefore, the relevance of having legal protection and regulation regarding the use and governance of agricultural data – collected, processed and analyzed through digital agriculture tools – refers to the interconnection and relationships arising from three elements, represented in Figure 2.

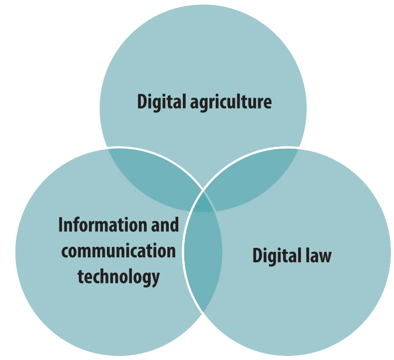


Figure 2. Interconnection between Digital Agriculture, Information and Communication Technology and implications of Digital Law.

Source: Mendes (2020).

Digital law: introductory lines

Within the scope of legal theory in Brazil, several authors have studied the transformations of legal relations stemming from advances on information technology, such as Leite (2016), Maranhão (2018), Novais and Freitas (2018), Nogueira and Nogueira (2019) and Abrusio (2020).

The network society, at the heart of the digital economy, is permeated by new types of conflicts in social relations within virtual environments or with the actions of artificial agents, presenting a series of difficult questions. For instance, what is the responsibility of online platforms that offer technical infrastructure, when the communication by third parties produces damages to individual rights? If personal data is the main source of value in the digital economy, should its collection be paid for? How to reward rural producers for the data generated with the use of digital products?

And how to ensure that the producer can control the use of his personal data? Is personal data a type of property? And who owns the inferences obtained from the aggregation of personal data? How to assess the market power of digital platforms in an economy constantly pressured by innovations? Can there be an electronic signature consent by artificial intelligence? What does "consent" mean in relation to artificial intelligence? Does an application that facilitates the interaction between drivers and passengers, receiving financial return, create an employment relationship with the driver? If the artificial intelligence wrongly indicated the harvesting period, harming producers and compromising investments, who will be responsible?

In general, several regulations, codes and judicial precedents created in the last century had social and economic relations as prototypical instances in the physical world. Its application to conflicts in the digital environment may face a series of gaps or indeterminacy and inadequacy of concepts. Such difficulties, as shown in the above questions, are manifested in the different branches of law (labor, competition, registration, contractual, civil liability, etc.).

Therefore, it is normal to question the legal nature of conflicts arising from the virtual environment. On the one hand, some will advocate the creation of a new branch of law, called digital law, cyber law or even computer law. According to Pinheiro (2019), digital law is the evolution of the rule of law and encompasses the fundamental principles and legal institutes[3] of the law in effect and currently applied, as well as introducing new institutes. This aspect is fed by the creation of specific regulations, such as the civil rights framework *Marco Civil da Internet*, which deals with the responsibilities of connection providers

[3] Legal institutions are a set of regulatory norms for the creation of specific laws/regulations, with their own characteristics, constituting an autonomous legal entity, which serves the interests of private or public order (Jusbrasil, 2020).

and online applications, as well as the development of new concepts or the recognition of new types of fundamental rights, such as the right to informative self-determination.[4]

On the other hand, it is recognized that digital issues and conflicts are transversal, affecting different branches of law, which could lead to concluding that there is no new branch of law, but only the application of different branches to a new object. Pimentel (2018, p. 37) argues, in a conciliatory line, that "Digital Law covers all areas of Law, in a transversal way, and brings together new elements to settle conflicts that have arisen with technology, especially the internet, and regulate the relations of the so-called information society".

Maranhão (2018), however, emphasizes the aspect of conceptual reconstruction driven by new conflicts in the digital environment, which affects all manifestations of the rule of law. This reconstruction is bidirectional: not only the legal concepts formulated for the physical world, in the different branches, are adapted to a possible "digital universe", but also the new concepts developed in the digital sphere affect their application in the physical world. Along these lines, Maranhão (2018) argues that society is facing a transformation of the law, which may affect several legal branches, in a new reconfiguration of its fundamental concepts such as responsibility, property, employment relationship, legal contracts, etc.[5]

Thus, it is possible to circumscribe a set of themes, including specific legislation, typical for a branch of "digital law," such as neutrality of networks, databases, electronic commerce, protection of personal data, artificial intelligence, obligations and responsibilities of internet connection and application providers, but without losing sight of the fact that the concepts developed in this field, due to their transversal nature, bring transformative implications for the law as a whole.

It is therefore necessary to undertake a "bottom-up" analysis, that is, to understand the characteristics and impacts of a given technology applied in a specific domain, such as agricultural production, to then identify its implications on rights and duties in possible conflicts, and understand whether the legal concept or institute, generally formulated for the physical world, can be applied or needs to be adapted. The adaptation effort cannot be isolated, and all concepts relevant to the domain must be considered: for example, to what extent can an eventual extension or restriction of civil liability affect liability for environmental damage and vice versa? A possible conservative interpretation, in a given court decision, that assigns civil liability for damage caused by a specific choice of an artificial intelligence to its developer, would also lead to liability for environmental damage, creating potential liability capable of discouraging the development and use of this technology. Hence the need to coherently elaborate legal concepts, considering all implications for the legal system and its consequences on economic activity.

The application of the law must, therefore, broadly assess the interests at stake in order to understand the new challenges that confront us when machines, soils, animals and other information on rural property are monitored by companies that generate an immense volume of data arising from rural activities and now hold precious and qualified information that will be processed by Big Data. This valuable information can be used by developing companies to eventually induce behaviors related to production and consumption (Leite, 2016). The massive collection of information from rural activities and its use by companies that develop computerized systems integrated to digital agriculture are challenges for

[4] In this regard, see the decision of the Federal Supreme Court (STF), in the judgment of the Provisional Measure in the Direct Action of Unconstitutionality 6,387 - Federal District, which suspended the effects of the Measure Provisional No. 954/2020 and recognized the fundamental right to informative self-determination (Brasil, 2020b).

[5] The chapter does not intend to exhaust such a profound discussion on the genesis of digital law. To expand the debate, in addition to the aforementioned works by Pinheiro (2019-), Maranhão (2018) and Pimentel (2018), see also Hoeschl (2011), Madalena (2016) and Costa and Pendiuk (2020).

public policy makers and public agricultural research institutions, such as Embrapa, which develop these computerized systems. Knowing how to deal with legal issues that arise from the generation and use of this large volume of data is a relevant factor for Embrapa.

This avant-garde theme motivated the creation of several research centers to support the growth of digital law, in a multidisciplinary perspective that includes computer science, engineering and law (Maranhão, 2017). Table 2 lists some of these centers, not exhaustively, but as an example.

Table 2. Digital Law research centers [(1)].

Research Center	Institution/Country	Website
The Stanford Center for Legal Informatics (CodeX)	Stanford University, United States	https://law.stanford.edu/ codex-the-stanford--center- for-legal-informatics/
Center for Research in Legal informatics (Cirsfid)	University of Bologna, Italy	http://www.cirsfid.unibo.it/
Intelligent Systems Program	University of Pittsburgh, United States	http://www.isp.pitt.edu/
Centre for Technology, Ethics, Law and Society	King's College London, England	https://www.kcl.ac.uk/law/ research/centres/ telos
Institute for Ethics in Artificial Intelligence	Technical University of Munich	https://ieai.mcts.tum.de/
Lawgorithm[(1)]	University of São Paulo, Brazil	https://lawgorithm.com.br/
International Association for Artificial Intelligence and Law		http://www.iaail.org/

[(1)] Lawgorithm is an association for research in artificial intelligence applied to the law, created in 2017 at the University of São Paulo, and brings together professionals from law, engineering, computing, Escola Politécnica and USP's Institute of Mathematics and Statistics (Maranhão, 2019).
Source: Maranhão (2017) cited by Mendes (2020).

These centers address two perspectives on the interaction between information technology and artificial intelligence and law: a) artificial intelligence law, which seeks to technically understand digital agents (such as AI tools developed by Embrapa) and reflect on what the social impacts are and the new legal issues arising from them; b) artificial intelligence in law – the application of artificial intelligence in legal practice (to predict decisions, perform intelligent jurisprudence searches, automatically generate legal documents, use chat bots on legal topics, etc.) (Maranhão, 2017).

Based on the first perspective of artificial intelligence law, Maranhão (2017) highlights that its applications bring new types of conflict and new issues, at least, to the following areas of law:

a) **Intellectual property:** the use of artificial intelligence for the creation of intellectual works – such as software, utility models, brands and industrial designs – raises the questions: who are the holders of the author's patrimonial and moral rights? Would the owner be the software developer or the company that invested in the development of the program? This first topic may have relevant implications for the activities of Embrapa and agribusiness, as artificial intelligence[6] can be used in the creation of new cultivars, among other forms of intellectual property.

[6] The generic term "artificial intelligences" refers to computer systems or programs that incorporate some machine learning or knowledge representation methodology or technique.

b) **Civil responsibility:** systems that employ artificial intelligence could, eventually, violate third-party rights, as systems based on machine learning make autonomous decisions based on the analysis of Big Data. This aspect will undoubtedly be relevant for the application of artificial intelligence in the production process, simply considering an investment chain for a given crop, pointed out by artificial intelligence, that proves to be wrong, or an artificial intelligence that uses a specific agricultural pesticide at an inadequate dose.

c) **Data protection:** artificial intelligence systems collect data for future decision making at each interaction. The question is how are these data collected, processed and used? The processed data, if skewed, can generate automatic decisions that interfere with individual rights. Therefore, concerns are related to the fact that the AI system can extract knowledge of decisions based on complex machine learning algorithms and, also, the necessary regulation and guarantees of people's rights – natural or legal – who are affected by such automated decisions. Here, too, there may be relevant issues, as artificial intelligences start to collect data from rural producers so as to draw profiles for the supply of consumer goods or, also, to influence their decisions on what, when and how to produce. Conflicts may also arise regarding collecting data from rural workers to create profiles and monitor their work.

d) **Impacts on employment:** there may be new labor issues, related to the hiring of workers, based on profiles created using artificial intelligence systems or automated contracts that may involve the rural producer monitoring the worker. Although not strictly legal, the impact of the use of artificial intelligence on employment in agricultural activities must also be analyzed and weighed to enable relocation and training programs for farmers, so they can be able to deal with Agriculture 5.0.

e) **Environmental law:** artificial intelligence systems are used to increase efficiency in a given activity. The focus on increasing agricultural productivity, a natural motivator for these investments, can neglect and bring risks to the environment, which can bring new issues about responsibility for environmental damage.

In addition to the legal implications emphasized, the use of artificial intelligence has produced a series of ethical questions. There are two types of risk observed in the discussion, that of overutilization, when such systems can have negative impacts on human rights, and that of underutilization, when the fear of artificial intelligences may fail in taking advantage of their potential benefits to humanity (Floridi et al., 2018).

In recent years, in response to concerns about the use of artificial intelligence, especially those based on machine learning, documents have been produced by government bodies, research associations and private organizations, proposing ethical parameters for the development and application of AI systems. The various documents show there is some convergence around the principles of transparency (it must be clear to the user who interacts with an artificial system), explanatory information (information disclosure to the interested party, allowing the user to understand the decision-making criteria), non-discrimination (preventing systems from incorporating biases that may offend fundamental rights), non-maleficence (AI systems cannot harm humans), accountability and privacy/data protection, although there are differences regarding its meaning and form of implementation (Jobin et al., 2019). As they are common, vague and potentially conflicting, their implementation is difficult.[7]

[7] This difficulty can be seen in a recent report by the Berkman Klein Center (Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI, 2020), associated with Harvard University, which identified 36 sets of potentially conflicting principles.

The European Union created the High-Level Expert Group of AI, which produced two reports, one to define artificial intelligence, indicating its potential benefits and risks, and another to establish ethical standards for artificial intelligence.[8] Although there are fears regarding regulatory intervention in a constantly changing environment (Maranhão; Coutinho, 2019), in early 2020 the European Commission released the White Paper "On Artificial Intelligence - A European Approach to excellence and trust", which points to regulation, mainly areas considered to be at risk (health, transport, energy and part of public services, in addition to applications that affect workers' rights and remote biometric identification). There is no detailed suggestion of regulation or exclusion of methodologies, but a convenient indication for AI developers to adopt internal transparency, that is, mandatory documentation of the entire decision-making process of software development (design, training, launching, monitoring), as well as the inclusion of reports that assess the aggregated outputs, which are easily accessible in audits.

In Brazil, there is an initiative by the Ministry of Science and Technology to publish a National AI Strategy, which will guide the investment targets for this technology in the country, in addition to creating ethical parameters (Brasil, 2020a). There are also two bills pending in the Senate regarding the subject to make sure the contribution of the Lawgorithm institute defends the need to define "bottom up" ethical parameters, that is, take into account the peculiarities of each application sector.[9] Therefore, for example, the specific ethical parameters for its application in agricultural production must be different from those applied to medicine or law.

Finally, from the perspective of artificial intelligence applied to law, although it is not the field of agricultural application, another research frontier is noteworthy. Given the ubiquity of AI systems and the impossibility of human oversight of all possible decision-making and actions by intelligent digital agents, it is imperative that artificial intelligence incorporates intelligent ethical/legal agents capable of processing law or standards that are computationally moral (see the CompuLaw project of the European Community[10]).

Therefore, the ethical rules defined for each application sector should be computable. A regimental programming model, in which the programmer prevents certain ex-ante actions, is not enough, as AI systems adapt their behavior to the circumstances. Thus, AI systems will need to process and apply ethical and legal rules when choosing their course of action, considering the particularities of the context.

In the agricultural field, artificial intelligence that makes decisions about cultivation and pesticides must incorporate a digital legal agent, which ensures compliance with the environmental rules in force. The creation of intelligent ethical/legal agents or the development of a "computable law" depends on research investments, which are currently one of the vanguard areas in artificial intelligence and law.[11] The perspective of applying artificial intelligences in agriculture should be attentive to these developments in order to develop systems that take into account ethical and legal compliance in their decision-making.

For Embrapa – as a digital agent that integrates the ecosystem of agricultural innovation and that develops and provides digital tools for agriculture – it is pertinent to analyze the legal aspects arising from ICT applied to agriculture and explore issues that are at the interface between the rule of law and the technological development of digital agriculture.

---

[8] AI-HLEG A definition of AI: Main capabilities and scientific disciplines e AI-HLEG Ethic Guidelines for a Trustworthy AI.

[9] Available at: https://lawgorithm.com.br/estrategia-nacional-de-inteligencia-artificial

[10] Available at: https://cordis.europa.eu/project/id/833647

[11] In Brazil, only the Faculty of Law of the University of São Paulo offers subjects related to computable law, in the postgraduate course.

The digital tools for agriculture, developed by Embrapa, may be subject to regulation, through legal instruments that support Embrapa's business with digital assets, to discipline the relationship between Embrapa and its users, in the services provided by the company, through their websites and mobile apps.

Digital assets for agriculture: legal support for Embrapa's performance

Following this reality of the agricultural technological revolution, Embrapa has been enriching its portfolio of digital assets by expanding the offer of services and products electronically, as research and innovation in agribusiness are increasingly associated with digital tools. Embrapa has a total of 292 software and web services[12], covering a wide range of topics so as to meet society's demands.

Among the mobile apps that were made available, the champion of downloads is "Roda da Reproduction", which already has 18,828 active installations[13]. This tool was developed to assist managing dairy herds, enabling to monitor the productive and reproductive stages of a herd in a simple manner. The name is due to the application's display, displaying the herd on a wheel that allows to quickly visualizing these stages, with color scales and placements. The use of technology brings enormous ease and simplification to the work of rural producers, hence abandoning the use of paper records to monitor the herd. Figure 3 shows the mobile applications made available by Embrapa.

In the entire universe of Embrapa's digital tools, more than half (165) are available on Embrapa's Portal, which are suitable for technology transfer. It should also be noted that the qualification of 119 digital assets[14] has already been carried out, using the criteria of the TRL/MRL scale[15].

Immersed in this reality and digital need for agribusiness, and aware of its legal status as a federal public company and all the social and technical responsibility involved, Embrapa's performance is focused on the multidisciplinary work of its team, in order to ensure that the development of technologies is always surrounded by legal protection.

Ultimately, it is up to the law and its operators – lawyers, judges, members of the Public Ministry – to face the challenges posed by digital technology, promoting the due legal protection not only to developers and owners, but also to users. One of the main challenges of the jurist in the world today is to consider the repercussion of the

Mobile application	Active installations	Score on Google Play
Roda de produção	18,828	4.34
Zarc - Plantio Certo	8,218	3.92
Guia InNat	3,562	4.25
Doutor Milho	2,955	3.98
AppLeite	2,115	4.29
Suplementa Certo	2,113	4.52
AgroPragas Maracujá	1,865	4.69
Custo Fácil	1,647	4.45
+Leite	1,550	4.40
SAC Gado de Corte	1,535	4.32

Figure 3. Digital assets: mobile applications provided by Embrapa (June 2020).

Source: Embrapa (2020).

[12] The total of 292 software and web services are from June 2020.

[13] Active installations in June 2020.

[14] Digital asset data in June 2020.

[15] Technological maturity level scale, TRL/MRL scale – Technology Readiness Levels/Manufacturing Readiness Levels.

law in view of the entirely new circumstances that are now presented, bearing in mind the paths for its transformation (Lemos, 2005).

Focusing on its research purpose, addressing innovation, Embrapa's lawyers are responsible for regularly analyzing the disruptive application of the law aimed at public administration, to monitor the technological evolutions promoted by the Company's technical staff based on the needs of the agribusiness.

Bureaucracy mirrors the functioning model of the hierarchical and standardized industrial society, while research requires management flexibility to meet its objectives in search of the unknown and transforming it into new goods or services. If research requires management flexibility, applying the law at Embrapa demands the same requirement (Peregrino, 2018).

Thus, in order to promote due legal protection of its digital assets, Embrapa, guided not only by its Innovation Policy, aligning national legislation on science, technology and innovation and intellectual property, but also by the General Data Protection Law – LGPD (Law No. 13,709/2018) (Brazil, 2018) and Access to Information (Law No. 12,527/2011) (Brazil, 2011), which edited the main contractual instruments that provide legal support to the Company's digital business management.

The LGPD emphasizes, in its article 50 (Brazil, 2018), the importance of formulating rules of good practices and governance, which set the conditions of organization, the operating regime and procedures adopted in the processing of data, security norms, technical standards, specific obligations for the various actors involved, educational activities, internal mechanisms for the supervision and mitigation of risks and other aspects related to processing personal data and handling complaints and doubts of the data owners.

Thus, these model documents were prepared to promote the legal security of Embrapa's digital assets, in addition to being a mechanism to standardize the relationship between Embrapa and its customers and users, in order to safeguard the Company's macro process of innovation efficiently and effectively.

The following legal instruments were created: Term of Use; Privacy Policy; Service Level Agreement (SLA) and Technical Support Guidelines. These instruments are atypical electronic agreements that promote legal certainty for digital services, with the ability to delimit the responsibility of both Embrapa for providing the service and the user for its satisfaction, explaining the conditions of operation of the asset.

The electronic contract is characterized by using an electronic means for its execution or by being related to a bilateral legal transaction that results from the assembly of two declarations of will, as agreed through the electronic transmission of data. (Finkelstein, 2004 cited by Pinheiro, 2019).

It can also be defined as an electronic transaction in which declarations of will are manifested by electronic means and may also be manifested automatically by a computer (automated computer system), or through a public offer on a website and acceptance by the consumer through a click (Lorezetti, 2006 cited by Pinheiro, 2019).

With regard to the Terms of Use and Privacy Policy instruments, as there are no legal definitions, there is no doctrinal consensus on the concepts, and it is common among authors to permeate the content of one into the other.

This interposition exists between the aforementioned instruments because they both clarify how the digital asset is used, determining obligations and clarifying doubts about its functioning. On the other hand, the SLA Agreement and the Technical Support Guidelines, which are more specific, do not undergo this emission.

Within the scope of Embrapa, the Term of Use is the adhesion contract that allows determining the conditions for accessing and using the website or mobile application and which must be observed by users. This instrument lists important information, through which the service or product is described. The essential clauses are: the adopted nomenclatures, the obligations of the user and Embrapa, how the application or website works, the cost of the service, the hypotheses of any pauses and termination of the service and in what way Embrapa handles third-party information.

The Privacy Policy deals with the terms and security conditions that will guide the relationship to be established between Embrapa and users, especially the privacy of users' personal information, in order to offer the due credibility and transparency to users regarding the use of its websites and applications.

Through this document, Embrapa proposes to communicate how the information the user entered on the website or in the application will be used, such as registration data, and those resulting from the tool to capture information, posted items, stored messages, also informing that the information may be shared with partner companies or used for research, in order to improve the performance of the site or application, as well as if there will be transfer of information to third parties and how this transfer can take place.

The rights and duties of the user were also outlined, with a specific chapter on sharing information with third parties allowed by the user, in order to clarify the exceptions that allow the transfer of information to third parties, such as by court order, legal determination, etc.

It should be noted that all these topics involved legal effort to reconcile the laws applicable to Embrapa as a public company. And as already mentioned, it is necessary to pay attention not only to laws affecting intellectual property, innovation and protection of personal data, but also to legislation relevant to Public Administration, such as the Access to Information Law (Law 12,527/2011) (Brasil, 2011), which simultaneously provides the duty to give access to public information and the confidentiality of private information.

The Service Level Agreement (SLA) instrument is the document required in any IT contractual relationship, which measures the performance and quality with which a service is effectively delivered, through objective criteria.

The purpose of the SLA is to be a tool for monitoring and controlling compliance with the standard established in the service agreement contracted between the parties, allowing for clear and unambiguous customer expectations and the supplier's obligations and limits of responsibility (Pinheiro, 2019). Such control requires overt monitoring, the stipulation of fines for insufficient performance, co-sourcing (having more than one supplier) to avoid concentration, guarantees and insurance, if applicable (Pinheiro, 2019).

The SLA designed by Embrapa defines the main technical terms, presents the Calculation of Monthly Activity and Service Levels, and outlines the limitations that are not applicable to the SLA, clarifying that they do not apply to any performance or availability issues.

Embrapa also provided for any compensation to the user, in the event of extrapolation of the service's downtime, which will only be carried out through a service credit, whose compensation cannot be made unilaterally by the user in their Applicable Monthly Service Fees.

Finally, the Technical Support Guidelines outline the supplier's responsibility to maintain the stability of the service provided, whether by offering technical support, clarifying doubts or performing preventive and corrective maintenance, among other support activities.

The objective of the technical support guidelines is to promote customer and user satisfaction in the services provided by Embrapa, through its websites and mobile applications, by efficiently meeting the respective demands, as quickly as possible, as well as by maintaining the proper functioning of the services provided by such channels. They also aim to correct any stoppages or loss of quality, disparate doubts, complaints, requests for new services and requests for changes to services or configuration items.

It should be noted that the edition and consolidation of these legal instruments, by the CID and CSJ of SIN, for the due legal support to Embrapa's digital business management, were also aware of the fact that the regulations applied to this type of business must be "globalized", so that they are effective not only internally, but also abroad. After all, digital activity breaks boundaries and there must be compatibility with globally established guidelines.

Acting as a legal support for the innovation and implementation of research, for the realization of agriculture fueled by science, Embrapa's statutory role, it is essential for its legal body to also operate in an innovative manner in terms of regulation/law. It is necessary to emphasize the need to apply the law with due attention to the impact and indispensability of technology, in order to offer adequate legal security, supporting the promotion and access to knowledge, science and technology. More specifically, considering that it is a state-owned company subjected to legislation affects Public Administration, it is imperative to link its obligations without impeding innovation, quite the contrary, helping it to succeed.

Final considerations

Science driven agriculture is a reality in Brazil. Embrapa has made and is still making relevant contributions to accomplish this – together with partner institutions of the National Agricultural Research System – through the development of RD&I solutions for the sustainability of agriculture, to benefit Brazilian society.

For Brazil to continue to be a world competitor in food exports, as well as a supplier to meet the domestic demand for food, it is essential to train agricultural agents and appropriate the most advanced digital technologies. This technological appropriation qualifies Brazilian agriculture to face the challenges of feeding Brazil, improving the performance of agribusiness' participation in the trade balance and increasing the sector's competitiveness in relation to competitors.

Agricultural innovation based on digital content lacks participatory governance and multidisciplinary approaches, such as the one presented in this chapter, which directs the legal views to digital agriculture. For the full advance of digital agriculture in Brazil to take place, it is imperative for its activities to focus on: a) solving Brazilian agricultural problems and developing the production system; b) meeting the goals of sustainable development to promote food security in the country; c) promote the training and ownership of digital technological innovations by farmers; d) advance digitalization in the field and rural areas with innovations that value and respect people, the climate, biodiversity and the environment.

It is important to observe the warning about the risk of digital agriculture being controlled and structured by a few giant companies of global agribusiness, considering the merger and market concentration movement, with the priority to detain and monopolize the oil of the 21st century – the data with added value – to obtain extraordinary profits.

Therefore, the legal framework for data use, governance and privacy needs to be improved and applied in the context of digital technologies, whether Agriculture 4.0, Agriculture 5.0 and its successive waves of technological advances, to regulate the legal relationships of the parties involved in the collection, the processing and analysis of agricultural data, in order to avoid or minimize the potential effects of inducing production and consumption behavior. The role of public agricultural research institutions, such as Embrapa, is essential to promote a sense of balance in the availability and socialization of digital technologies, in order to promote technological equity among farmers and agricultural producers.

The perspectives for the world and for Brazil in relation to digital law as regards digitization of agriculture – envisioned by the authors of this chapter – are to expand and qualitatively advance the debate on the subject, to improve the regulatory framework of digital law, expand State incentives for scientific development, research, technological training and innovation, as recommended by the Brazilian Federal Constitution.

References

ABRUSIO, J. **Proteção de dados na cultura do algoritmo**. São Paulo: Ed. D´Placido, 2020.

BENKLER, Y. **The wealth of networks**: how social production transforms markets and freedoms. New Haven: Yale University Press, 2006.

BRASIL. Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º , no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. **Diário Oficial da União**, 18 nov. 2011. Edição extra.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados (LGPD). **Diário Oficial da União**, 15 ago. 2018. Edição extra.

BRASIL. Ministério da Ciência, Tecnologia e Inovações. **Estratégia brasileira de inteligência artificial**. Available at: https://www.mctic.gov.br/mctic/opencms/inovacao/ paginas/politicasDigitais/Inteligencia/Artificial.html. Accessed on: 6 July 2020a.

BRASIL. Supremo Tribunal Federal. **Medida cautelar na Ação Direta de Inconstitucionalidade 6.387 - Distrito Federal**. Available at: http://www.stf.jus.br/arquivo/cms/noticiaNoticiaStf/anexo/ADI6387MC.pdf. Accessed on: 16 June 2020b.

CASTELLS, M. A. Sociedade em rede: do conhecimento à política. In: CASTELLS, M.; CARDOSO, G. (org.). **A sociedade em rede**: do conhecimento a acção política. Lisboa: Imprensa Nacional: Casa da Moeda, 2006. p. 17-30.

CEMA. **Digital Farming**: what does it really mean? And what is the vision of Europe´s farm machinery industry for Digital Farming? Brussels: CEMA aisbl - European Agricultural Machinery, 2017. Available at: https://www.cema-agri.org/images/publications/position-papers/CEMA_Digital_Farming_-_Agriculture_4.0__13_02_2017_0.pdf. Accessed on: 17 Feb. 2020.

COSTA, R. R. S.; PENDIUK, F. **Direito Digital**: o marco civil brasileiro da internet e as inovações jurídicas no ciberespaço. Available at: http://publica.fesppr.br/index.php/publica/article/viewFile/129/38. Accessed on: 12 May 2020.

EMBRAPA. **Ativos digitais**: aplicativos móveis disponibilizados pela Embrapa. Secretaria de Inovação e Negócios da Embrapa: Brasília, DF, 2020. Apresentação em PowerPoint.

FINKLSTEIN, M. E. R. **Aspectos jurídicos do comércio eletrônico**. Porto Alegre: Síntese, 2004.

FLORIDI, L.; COWLS, J.; BELTRAMETTI, M.; CHATILA, R.; CHAZERAND, P.; DIGNUM, V.; LUETGE, C.; MADELIN, R.; PAGALLO, U.; ROSSI, F.; SCHAFER, B.; VALCKE, P.; VAYENA, E. AI4People - An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. **Minds and Machines**, n. 28, p. 689-707, Nov. 2018. DOI: 10.1007/s11023-018-9482-5.

GASQUES, J. G.; VIEIRA FILHO, J. E. R.; BASTOS, E. T. Produtividade da agricultura brasileira: crescimento e inovação. In: VIEIRA, P. A.; CONTINI, E.; HENZ, G. P.; NOGUERIA, V. G. C. (ed.). **Geopolítica do alimento**: o Brasil como fonte estratégica de alimentos para a humanidade. Brasília, DF: Embrapa, 2019.

HOESCHL, H. C. **Elementos de direito digital**. Available at: https://siabi.trt4.jus.br/biblioteca/ direito/doutrina/livros/elementosdedireitodigital.pdf. Accessed on: 12 May 2020.

JOBIN, A.; IENCA, M.; VAYENA, E. Artificial intelligence: the global landscape of ethics guidelines. **Nature Machine Intelligence**, v. 1, p. 389-399, Sept. 2019. DOI: 10.1038/s42256-019-0088-2.

JUSBRASIL. **Instituto jurídico**. Available at: https://www.jusbrasil.com.br/topicos/297191/instituto-juridico. Accessed on: 24 Feb. 2020.

LEITE, G. C. Revolução digital no setor agribusiness pode gerar controvérsias jurídicas. **Consultor Jurídico**, set. 2016. Available at: https://www.conjur.com.br/2016-set-15/guilherme-leite-questoes-juridicas-revolucao-digital-agribusiness. Accessed on: 17 Feb. 2020.

LEMOS, R. **Direito, tecnologia e cultura**. Rio de Janeiro: Ed. FGV, 2005. Available at: http://bibliotecadigital.fgv.br/dspace/handle/10438/2190. Accessed on: 28 May 2020.

LORENZETTI, R. L. **Comércio eletrônico**. São Paulo: Revista dos Tribunais, 2006.

MADALENA, J. Regulação das fronteiras da internet: um primeiro passo para uma teoria geral do direito digital. **Revista dos Tribunais Online**, v. 974, p. 81-110, Dec. 2016.

MARANHÃO, J. A evolução da Inteligência Artificial aplicada ao Direito no Brasil. **Olhar Digital**, 30 jul. 2019. Available at: https://olhardigital.com.br/colunistas/juliano_maranhao/post/a_evolucao_ da_inteligencia_artificial_aplicada_ao_direito_no_brasil/88576. Accessed on: 17 Feb. 2020.

MARANHÃO, J. A pesquisa em inteligência artificial e Direito no Brasil. **Consultor Jurídico**, 9 dez. 2017. Available at: https://www.conjur.com.br/2017-dez-09/julianomaranhao-pesquisa-inteligencia-artificial-direito-pais. Accessed on: 17 Feb. 2020.

MARANHÃO, J. Inferências metafóricas e a reconfiguração do Direito na era digital. **Revista de Direito e as Novas Tecnologias**, v. 1, ano 1, p. 15-30, 2018.

MARANHÃO, J.; COUTINHO, D. R. Inteligência artificial: melhor investir do que regular. **Correio Braziliense**, 25 mar. 2019. Available at: http://sampaioferraz.com.br/melhor-investir-do-que-regular. Accessed on: 13 Oct. 2020.

MARANHÃO, J.; MARQUES NETO, F. A.; COZMAN, F. G. **Perspectivas para a Inteligência Artificial e Direito**, 5 dez. 2019. Available at: http://sampaioferraz.com.br/perspectivas-para-a-inteligencia-artificial-e-direito. Accessed on: 17 Feb. 2020.

MASSRUHÁ, S. M. F. S. Entrevista. **Revista Presença Internacional do Brasil (PIB)**, ano 12, n. 39, dez. 2019/jan. 2020.

MELGAR, M. Digital agriculture or agriculture 4.0. **SugarJournal**, Oct. 2018.

MENDES, C. I. C. **Inteligência artificial na agricultura 5.0**: uma análise sob a perspectiva do Direito Digital. São Paulo, 2020.

MENDES, C. I. C. **Transferência de tecnologia da Embrapa**: rumo à inovação. 2015. 386 f. Tese (Doutorado em Desenvolvimento Econômico) – Universidade Estadual de Campinas, Campinas.

MENDES, C. I. C.; BUAINAIN, A. M.; FASIABEN, M. C. R. Uso de computador e internet nos estabelecimentos agropecuários brasileiros. In: MASSRUHÁ, S. M. F. S.; LEITE, M. A. de A.; LUCHIARI JUNIOR, A.; ROMANI, L. A. S. (org.). **Tecnologias da informação e comunicação e suas relações com a agricultura**. Brasília, DF: Embrapa, 2014. p. 39-52.

MENDES, C. I. C.; RIBEIRO, P. G. G.; BUAINAIN, A. M. Direito autoral na agropecuária: relato da experiência da Embrapa. In: BUAINAIN, A. M.; BONACELLI, M. B. M.; MENDES, C. I. C. (org.). **Propriedade Intelectual e Inovações na Agricultura**. Rio de Janeiro: IdeiaD, 2015. p. 263-284.

NOGUEIRA, J. H. M.; NOGUEIRA, S. M. **Direito digital e cibernético**: legislação específica. Joinville: Clube dos Autores, 2019.

NOVAIS, P.; FREITAS, P. M. **Inteligência artificial e regulação de algoritmos**. Diálogos União Europeia e Brasil. Universidade do Minho. 2018. Available at: https://www. mctic.gov.br/mctic/opencms/inovacao/paginas/politicasDigitais/assuntosCiberneticos/inteligencia-artificial-e-regulacao-de-algoritmos.html. Accessed on: 17 Feb. 2020.

PEREGRINO, F. Questões sobre a burocracia e as sociedades industriais e do conhecimento. In: SOARES, F. M., PRETE, E. K. E (org.). **Marco regulatório em ciência, tecnologia e inovação**: texto e contexto da Lei nº 13.243/2016. Belo Horizonte: Arraes Editores, 2018. p. 1-19.

PIMENTEL, J. E. de S. Introdução ao Direito Digital. **Revista Jurídica da Escola Superior do Ministério Público de São Paulo**, v. 13, n. 1. p. 16-39, 2018. Available at: https://es.mpsp.mp.br/ revista_esmp/index.php/RJESMPSP/article/view/352/340340364. Accessed on: 13 Oct. 2020.

PINHEIRO, P. P. **Direito digital**. São Paulo: Saraiva, 2019.

SCHIMPF, M. **Digital farming**: can digital farming really address he systemic causes of agriculture´s impact on the environment and society, or will it entrech them? Belgium: Friends of the Earth Europe, 2020. Available at: http://www.foeeurope.org/sites/default/ files/gmos/2020/foee-digital-farming-paper-feb-2020.pdf. Accessed on: 25 May 2020.

SOARES, F. de M.; PRETE, E. K. (org.). **Marco regulatório em ciência, tecnologia e inovação**: texto e contexto da Lei nº 13.243/2016. Belo Horizonte: Arraes, 2018. 222 p.

ZUBOFF, S. **The age of surveillance capitalism**. Manhattan: Hachette, 2019.

14

Innovating communication in the age of digital agriculture

Nadir Rodrigues Pereira | Flávia Bussaglia Fiorini | Magda Cruciol | Vinícius Milléo Kuromoto

Introduction

New media revolutionized human communication, producing new social habits. With easy access to information, society has become more aware of its rights and much more demanding regarding their needs and expectations. In addition, this new and modern consumer profile and user of products and services encourages companies to be more concerned about accountability to the public, the transparency of their actions and social responsibility.

In a highly competitive environment, knowledge has been the driver of economic and social development. In this context, communication plays a fundamental role in the complex environment of public research, development and innovation companies. They are faced with the challenges to more efficiently contribute to the dissemination of the scientific knowledge produced, and also bringing to society, in an innovative way, the results of research and technologies.

The 21st century is marked by digital communication; society is increasingly eager for fast information, in real time. Furthermore, it is an agent that helps to build the contents that constitute collective knowledge, in other words, it is no longer limited to receiving information, in one-way communication. This society wants to participate, speak, listen and be heard.

As a result, communication tends to be increasingly horizontal in a participatory system, at all levels, highlighting the two-way symmetrical communication model, characterized by the balance between the interests of organizations and their audiences. Thus, it plays the role of enabling organizational change processes, as it goes beyond borders, provides greater access to information and facilitates dialogue, establishing strategies to deal with environmental changes.

The overwhelming global transformations caused by the pandemic of the new coronavirus imposed on companies and science and technology institutions the pressing need to adapt to the new scenario, and this reconfiguration is based on efficient communication with society and its strategic audiences. Thus, we witnessed a rapid increase in the participation of these entities in interactive portals, virtual communities and social networks, among other channels that enable greater interaction with individuals and consumers.

Communication is also considered a strategic instrument to awaken public motivation for science, to promote interaction and encourage the exchange of knowledge, contributing to socioeconomic development, democratization and social inclusion. These contributions are inherent to the mission of public institutions involved in research, development and innovation (RD&I), such as the Brazilian Agricultural Research Corporation (Embrapa), linked to the Ministry of Agriculture, Livestock and Supply (MAPA).

How to talk to this new audience profile; what language to use; what channels can be built; how to meet interaction needs; how can sounds, images and content be added to arouse interest and facilitate communication, given the many options and offers available on the web and social networks? How to make a difference and stand out? These are just some of the issues that communication professionals from RD&I institutions have been addressing in order to innovate communication and the relationship with society in the current context.

At Embrapa Digital Agriculture, an Embrapa Unit that is a research reference in the area of information technology, professionals from the Organizational Communication Nucleus (NCO) also face these challenges, in order to develop differentiated communication actions that bring innovative experiences in the relationship with the different audiences that the Unit interacts with.

Numerous solutions generated by agricultural research are made available on a daily basis in different formats to broadly disseminate the knowledge produced. The objective is that the actions contribute not only to more effectively bringing the research results, technologies, products and services offered to society, but also to facilitate interaction with the public and capture their demands.

This chapter presents some of the results obtained by the Unit. They are the outcomes derived from integrated communication based on adequate and consistent planning, which includes institutional, market and internal communication actions. Initiatives to support the innovation process and strategic management of Embrapa Digital Agriculture are also addressed, as well as participation in projects supported by an educommunicative vision, based on the subjects' autonomy and critical view.

New technologies and science communication at Embrapa

Companies and institutions began to be more intensely concerned with organizational communication in Brazil from the second half of the 1980s onwards. Political changes and, later, the phenomenon of globalization imposed the need for greater transparency in the relations of organizations with governments, non-governmental organizations, workers, unions, suppliers, the press and communities (Oliveira, 2013).

Since its creation, in 1973, Embrapa has been concerned about disclosing its administrative actions to society. For this reason, that same year, then-president José Irineu Cabral hired the first journalist to work at the Company (Duarte; Barros, 2003). In 1996, Embrapa implemented a Communication Policy, which helped to reorganize all its operation areas. Revised and updated in 2002, the policy defines that strategic and integrated communication[1] is a business intelligence system whose fundamental responsibility is to manage actions aimed at promoting the institution's relationship with the internal and external environments (Embrapa, 2002).

This policy created precepts to support the development of actions and programs to manage the dissemination of information to strategic audiences and to strengthen the institution's image. In addition to supporting actions aimed at popularizing scientific knowledge, at Embrapa communication is engaged with the transfer of research, development and innovation results, in order to improve the scientific literacy of the Brazilian population.

> In a research institution, acknowledging Business Communication as strategic implies including this competence in all instances of the organization, whether for prospecting demands or building scenarios, or when interacting with stakeholders involved in projects and solutions forwarded by company, or in search of the necessary interface with society (Embrapa, 2002, p. 15).

Communication actions, in addition to benefiting the Company and government, essentially play a role on behalf of society, as they facilitate disseminating research results generated at Embrapa and access to its products, services and technologies. The Company is committed to "make viable research, development and innovation solutions for the sustainability of agriculture, for the benefit of Brazilian society" (Embrapa, 2015, p. 8).

The arrival of new information and communication technologies (ICT) created a challenging scenario for companies and institutions to promote the dissemination of their actions, technologies, products and services to their strategic audiences, with an innovative approach. In this context, knowledge tends to be increasingly built using a collective, democratic and shared approach, so that everyone can benefit from the obtained results and therefore make their own decisions.

> The evolving scenario of both information technology and communication resulted – at least for organizations more attentive to new competitive requirements – in a converging scenario. A vital transformation in organizational communication seems to be underway, considering the expressive investments in integration and relationship solutions, which enable the intense participation of the entire chain of business agents in the dialogue and shared actions (Cardoso, 2020, p. 31).

In the 21st century, it is no longer possible for the scientific dissemination process to be a "one-way" model, built under a unidirectional vision and without the participation of all interested agents. The web tools provide interactivity with the public, which is no longer a mere consumer of information, but can become a co-author and participant in the process, playing the role of an active agent that acts and transforms its reality.

The impact of changes can be seen in all human activities, leading institutions to rethink their policies and create strategies to remain competitive, in addition to being socially responsible. New technologies also promote changes in the agricultural production process with the introduction of methods and tools that modernize agriculture.

In the case of Embrapa, there is a permanent focus on monitoring the demands of the external environment in order to align its performance. Embrapa's VI Master Plan (PDE) is a directive document

---

[1] Strategic and integrated communication is understood as that which is part of the organization's philosophy, guiding and integrating all communication actions generated in the company, seeking equilibrium between the organizational interests and those of its audiences (Kunschk, 1997).

that establishes the main guidelines for the activities to be developed at Embrapa, from 2014 to 2034, in line with the changes in the global scenario.

During the preparation of the VI PDE, the Company identified the need to expand its efforts to foresee the challenges in order to ensure the sustainability of Brazilian agriculture, which is greatly affected by technological intensification. The progress of ICT also offers enormous potential to revolutionize the agricultural sector, as new tools and technological solutions emerge for automating agricultural processes and impact business models.

> Amid these profound transformations, research, development and innovation (RD&I) activities in the agricultural sector have a central role. Technological transformations occur quickly, with the introduction of new products and processes, in which the control of genes and atoms becomes the center of change. There is a clear tendency to increase the complexity of this market with the expansion of various technologies, such as precision agriculture, biotechnology, nanotechnology, biological nitrogen fixation, biopesticides, biorefineries and intelligent packaging, etc. The concept of innovation no longer refers only to products and processes, but also to innovation in business models, logistics, services associated with products, distribution and marketing, management and organization (Fonseca Júnior et al., 2009, p. 87).

One of the current challenges of Agricultural research also includes making agriculture ever more connected (Rodrigues et al., 2020). Digital technologies represent significant commitments for the transformation of Brazilian agriculture, based on digital content, cutting-edge technology and connectivity, which characterize the digital age and Agriculture 4.0.

The focus of activities of Embrapa Digital Agriculture and Embrapa, "the so-called Agriculture 4.0 was already one of the priorities in the Company's Research, Development and Innovation (RD&I) programming, but with the pandemic, it will certainly grow even more to meet the demands of the productive sector" (Diniz, 2020).

The technological convergence, allied to the intense generation of data and information, have provided disruptive technologies an immense potential for applications in all activities, including planting, handling, harvesting and post-harvest. Furthermore, communication must also pay attention to all these innovations to help disseminate and appropriate these technological solutions.

> It is important to highlight that communication planning must be strategic, foreseeing scenarios, enabling the participation of company members, paying attention to expectations of the public and expressing the results desired by the company, in the short, medium and long term (Galerani, 2006, p. 51).

Digital revolution and network communication

The origin of communication as "convergence", ensuing from the understanding of the communicative process as a result of social interactions, replaces the diffusionist logic guided by the communication of "many for few" and now constitutes a new model, centered on the idea of "all for everyone" (Lévy, 1999). The new digital technologies make it possible to break the barriers of time and space, bringing together and stimulating the exchange of knowledge, cooperation and collective creation in the network.

In the virtual space interactions can be enhanced, and subjects are active, with the right to express themselves and interact, bringing proposals that benefit thousands of people who are interconnected, while also sharing solutions with those who do not have access to the network. Therefore, it is necessary to innovate in the sense of no longer treating the subject as a passive being, who receives a ready-made technology and needs to adapt to use it.

The digital revolution has brought, as well as to several other areas, a new horizon for organizational communication. The restriction of space and time is no longer an obstacle, and digital communication began to permeate spaces that were previously unnoticed or ignored. Universal digitization forced rethinking, not only for formal communication vehicles, but also for society, which discovered new ways to send news, receive information, seek updates and also be present. "Cyberspace presupposes a mixture of subject and object, unthinkable in the process of interpersonal and mass communication" (Santos, 2016, p. 4).

With new media, people can exchange information and share global solutions, contributing to the development of "knowledge networks". This model centered on a "everyone for everyone" type of communication (Lévy, 1999) assumes that anyone is able to be, at the same time, a consumer and a producer of information.

The dissemination actions are supported by internet resources – a form of "many-for-many" communication that, with the possibility of interaction, revolutionized human communication and designed a new paradigm for the socialization of information (Pereira et al., 2010, p. 4).

In this context, communication must also reinvent itself, operating in a network, as seen in research production based on the participation of researchers in collaborative virtual networks, which expand the possibilities of exchanging information and generating knowledge. A new dynamic is formed in these networks, stimulating collaboration in order to obtain the best results for their members.

> In these networks, scientists need to have a communication structure that facilitates, interactively, the exchange of information, knowledge, skills, competences, experiences, knowledge and abilities that allow them to simultaneously integrate them using an interdisciplinary and transversal approach, favoring the construction of new knowledge and solutions that add value to society (Torres et al., 2012, p. 3).

Technological convergence and multimedia resources offer potentials that allow researchers and scholars to organize in these networks, where participation and collaboration are encouraged both locally and internationally. This is a model that promotes collaborative creation, with the use of free and open technological resources that are available to multiple institutions, which benefit from ideas and collective improvements. As an example, we can mention the several collaborative initiatives currently underway in universities and research institutes for the production of a vaccine against covid-19 caused by the new coronavirus.

Characteristic of the information age and knowledge-based society, digital communication presents a new and complex scenario, in which the speed of obtaining information, the increased volume of content, not always from secure sources, and the overexposure of people and brands encourage thinking about how to position institutions, companies and actions in a relevant, clear, appropriate and attractive manner.

However, what is this digital communication and what makes it so different from traditional communication? It can be considered that digital communication is based on the communication strategy and actions carried out on the web, social networks and mobile devices, including the digital ecosystem and the digitization of information media. Supported by four pillars – presence, content, relationship and engagement – digital communication involves the relationships between connected human beings and the influence on corporate dynamics.

A segment of social communication, enhanced by technological advances, digital communication is a set of practices and forms of dissemination, interaction, reception and dialogue between sender and receiver on online platforms – accessible through devices such as computers, notebook, tablet, cell phone etc.

As it is comprehensive, it can be applied to diverse audiences, having started within corporations and expanded to the social world, in which legal and physical persons began to integrate and interact in a maze of information. With diversity increasingly present, the plurality of voices only tends to enrich relationships, professional or not. For companies and institutions, digital communication has brought valuable opportunities. Among them are the different forms of language, access, greater proximity and audience.

Today it is clearer that companies that are able to control their own digital communication can not only lead thoughts, but also shape behaviors, generating competitive advantages in a world where speed of delivery conquers not only consumers and business partners, but also followers. However, it is important to emphasize that communication must be intrinsically allied to corporate governance, for an effective contribution in this extremely competitive scenario, "seeking a more flexible, creative, collaborative environment and which, consequently, will bring more competitiveness and sustainability to organizations" (Sabbatini, 2010, p. 155).

Only after establishing dialogical relationships integrated to a collaborative and transparent communication proposal, which considers the citizen as an active subject in this process, will organizations be able to stand out and remain sustainable.

> Moreover, communicative actions need to be guided by a philosophy and an integrated communication policy that consider the demands, interests and demands of strategic audiences and society. This means there must be full integration between internal, institutional and business communication in the pursuit of organizational effectiveness, efficiency and efficacy for the benefit of the public and society as a whole, and not just solely the company. Studying, understanding and practicing organizational communication, therefore, is much more complex than one might imagine (Kunsch, 2009, p. 80).

Communication for innovation

The changes brought about by the 21st century, which impact both organizations and human beings, producing new social demands, also influence organizational communication within a complex context of relationships that require more interactive and collaborative systems.

Cajazeira and Cardoso (2009, p. 1) highlight the central role played by communication in the innovation process. However, they emphasize that the "complexity of the internal and external relations of organizations, and of individuals among themselves, combined with the growing competitive demand for innovation, poses unprecedented challenges on the way of thinking and acting in organizational communication."

Communication plays a fundamental role in the innovation process, as well as adequate information management. Strategies that encourage innovation attitudes have social interaction processes as one of their strongest allies, as communication, a reciprocal action, provides a favorable exchange environment for information to circulate and knowledge to be discussed, validated and possibly, adopted by the target audience.

For Wolton (2010, p. 121), "communicating is less and less transmitting, rarely competing, mostly negotiating and, finally, living together." Structured dialogue and the transmission of ideas can leverage the creative process, a precursor to innovation.

Therefore, communication and the creative process interact in complex but complementary ways, enabling the growth of the innovative potential in organizations. Having information is one of the first components for stimulating creativity and strategic thinking.

In this context, and to deliver more value to society through its products and services, Embrapa began implementing, in 2018, a new Innovation Policy. Among the six main guidelines, two are directly based on communication: promoting the culture, practices and internal environment for innovation; and expand the Company's participation and protagonism in the innovation market (Embrapa, 2018a).

Following this same philosophy of promoting innovations in its processes, the Company's communication strategies have already undergone reformulation, with the creation and use of differentiated communication tools to serve different audiences, with content and language suitable for each one. The main purpose has always been to improve how to communicate to society what Embrapa does and to talk more effectively with the innovation ecosystem, promoting dialogical and collaborative relationships.

The culture of organizations plays a preeminent role in stimulating, developing and disseminating innovations. It is essential for the internal environment of a RD&I institution to support the generation and sharing of ideas, which will be reflected in the research design, in the development of solutions and in the achievement of results to meet the demands of its strategic audiences.

> To overcome the limits of traditional business communication and the instrumental focuses of organizational communication, it is necessary to understand communication as a strategic process for action in a plural, dynamic and complex reality, which aims to elicit innovative, creative and dynamic behavior from a strategic point of view and which works, in a democratic way, as a disseminator of objectives and cultural values of the company for internal and external audiences. [...] These are economic changes with significant transformations for the markets and for the relationships between human beings inside and outside the company (Cardoso, 2006, p. 1.127).

To support this innovation process, it is essential that organizational communication is supported by a proposal for collaborative communication, which encourages dialogue and reciprocity. Communication for innovation plays a strategic role in the interpretation of the internal and external environments, identifying internal strengths and weaknesses, in addition to the characteristics and trends of the macro-environment.

Therefore, it comprises a multidisciplinary area, as it includes all forms of communication used by the organization to relate and interact with its audiences. It is important to highlight the work of public relations professionals in building the corporate innovation agenda, since "it is through public relations that the organization's philosophy contributes to strengthening and consolidating a solid and favorable corporate image and identity before the public of interest" (Kunsch, 2003, p. 164-165).

Thus, both communication and innovation need to be aligned so as to produce impactful results for companies and institutions, since the complexity of human relations makes this innovation process extremely challenging, which breaks the vertical structures of power and relationships. It is healthy for organizations to adopt more flexible and open styles of management that are reflected in their relationships with employees and stakeholders – including all the publics with which they relate.

The internal communication model must also be supported by participatory management, focused on collaborative work, which encourages autonomy and integration among teams. The Oslo Manual, published by the Organization for Economic Co-operation and Development (OECD), defines that among the factors that strongly influence the learning capacity of companies, vital for innovation, is knowledge management, including "policies and strategies, leadership, knowledge acquisition, training and communications" (Manual..., 2005, p. 32). "Corporate governance (legal, planning and public relations)" is also referenced among the administration and management actions for innovation (Oslo, 2018, p. 73).

In addition to all the disruptions caused by ICTs, the social changes that have taken place in the 21st century determine new relationship configurations that are based on pluralism and interdependence, "which require a new way of thinking about communication" (Cajazeira; Cardoso, 2009, p. 8). The challenges of this new context demand that strategic organizational communication, allied to corporate governance, advance within a relational perspective.

> The dialogic interaction is a new paradigm in this area, which breaks the mechanical model of information and adopts the posture of dialogue as the best way to resolve conflicts, make agreements, and, seek consensus in relation to a practice, thus understanding communication beyond technical rationality (Marchiori, 2011, p. 29).

In this scenario, it is essential to have an environment that will stimulate creativity and sharing ideas, with communication supporting the decision-making process of institutions and corporate governance. Seeing that innovation is an intrinsic process to organizational skills, it is essential to develop policies, programs and actions that favor cooperation, dialogical relations and the pluralism of opinions.

Thus, internal communication is also a vital characteristic in the decision-making process and in the construction of a participatory organizational environment, facilitating the integration and exchange of information. This will benefit enhancing the institutional image and strengthening the company's culture, producing positive results in a competitive and innovative business environment.

Therefore, it is important that leaders also develop skills in interpersonal relationships, face-to-face communication and management of information flows, establishing channels open to dialogue that are capable of supporting the construction of collaborative relationships, supported by ethics and the respect for the internal public.

> Internal communication cannot be isolated from the combination of integrated communication and from the other activities of the organization. Its effectiveness will depend on teamwork between the communication and human resources areas, the board of directors and all the employees involved. It will basically depend on adequate and consistent planning, and then it has to find support in the information obtained from the strategic planning, so that the programs to be developed correspond to the demands of the environment (Kunsch, 1997, p. 129).

Communication results and challenges in the digital age

Embrapa seeks to "be a world reference in the generation and offer of information, knowledge and technologies, contributing to innovation and sustainability in agriculture and food security" (Embrapa, 2015, p. 8). Among the 12 strategic objectives defined in its VI Master Plan 2014–2034, the development, adaptation and dissemination of "knowledge and technologies in automation, precision agriculture and information and communication technologies to enhance sustainability of production systems and add value to agricultural products and processes" are highlighted (Embrapa, 2015, p. 12).

For Embrapa Digital Agriculture, it is a priority to disseminate these innovations in Brazilian agriculture. Since the mid-1990s, the Unit has had a communication area dedicated to supporting the dissemination of research and technology transfer actions, in addition to managing the communication channels and flows with its stakeholders.

Communication is also among the strategic objectives of Embrapa's VI Master Plan. It is the responsibility of the Company to "develop and disseminate information products and communication strategies that will promote agricultural research and expand society's support for Brazilian agriculture." (Embrapa, 2015, p. 13).

The Organizational Communication Nucleus (NCO) is a sector directly linked to the head of the research center, and its performance is included in the strategic planning of the Unit.

> Based on this perspective, communication has the potential to become an instrument and also an intelligence process, a source of value generation and competitive advantage. After all, as it permeates all organizational dimensions – human, economic, marketing, cultural and social – communication is inextricably linked, whether acknowledged or not, to corporate performance as a whole (Mello, 2010, p. 200).

Communication actions are built at the Unit in a planned manner and then integrated with Embrapa's communication plan, which outlines strategies with corporate reach, aimed at the institution's various stakeholders.

> When preparing a communication plan, professionals must take care that their objectives and goals are not merely related to their productions – preparing publications, holding events, preparing reviews, and other productions. These can actually be means to reach nobler ends, such as effects on the relationships between the organization and its audiences (Galerani, 2006, p. 54).

The research center has been improving its performance in the communication area, investing in the composition and training of its team of professionals, so they can quickly respond to the new challenges imposed by the digital transformation. One of the main results of these investments is the consolidation of Embrapa Agricultural Informatics as a reference center in the field of digital agriculture, recognized by society and opinion makers.

The Unit has had a strong presence in the press in ICT reports, Agriculture 4.0, internet of things and development of technological solutions for the field. In 2019, the repercussion of news citing Embrapa Digital Agriculture increased by 123% compared to the previous year, with more than 1,450 news retrieved, which were published in local, regional, national and international media. Figure 1 shows some examples of inclusion in the media.

Figure 1. Covers of the magazines Pesquisa Fapesp, Dinheiro Rural and Globo Rural, which cited Embrapa Digital Agriculture.

In addition to producing podcasts for the radio program Prosa Rural by Embrapa[2] and video reports for the television program *Dia de Campo on TV*[3] about the technologies developed, the Unit is present on social networks on Embrapa's channels on Facebook[4], Instagram[5], Flickr[6], Twitter[7] and Youtube[8]. Figure 2 shows the dissemination of posts and tweets on social networks about research and developed technologies.

The Internet Portal[9] is one of the channels that Embrapa Digital Agriculture uses to publicize research and its results, in addition to presenting technological solutions, institutional and technical-scientific publications, available products and services. Aware of the external demands of its audience, the Unit is concerned with the permanent updating and revision of content to facilitate access for the population. Furthermore, the intense participation in agricultural events and exhibitions is to disseminate technologies, products and services, always seeking to approach and strengthen relationships with rural and urban audiences.

The challenge is to continue creating innovative strategies in the relationship with this society, which has a more participative and dynamic new profile. It's not just the tools that evolve, but also the organizational culture. It is important to have a receptive and open attitude to dialogue, which contributes to this closer proximity and interaction with the institution's stakeholders.

> We can add that in a complex environment, communication can only fulfill its role as a strategic management tool when the company creates the true channels for communication to fulfill its basic social principle, that is, its democratic character so as to enable all

Figure 2. Posts on artificial intelligence research and technological solution.

[2] Available at: www.embrapa.br/prosa-rural

[3] Available at: www.embrapa.br/dia-de-campo-na-tv

[4] Available at: fb.com/embrapa

[5] Available at: instagram.com/embrapa

[6] Available at: flickr.com/embrapa

[7] Available at: twitter.com/embrapa

[8] Available at: youtube.com.br/embrapa

[9] Available at: www.embrapa.br/agricultura-digital

> individuals to share ideas, behaviors, attitudes and, above all, the organizational culture. This democratic character is expressed through dialogue and the production of meanings (Cardoso, 2006, p. 1.135).

Therefore, the Organizational Communication Nucleus of Embrapa Digital Agriculture is also guided by an action aligned with Embrapa's innovation process. Hence, it develops strategies to promote connections with the Unit's various stakeholders, including institutions, companies and partners, in order to strengthen relationships, in addition to reinforcing the image of an innovative company in the agricultural sector.

Support is highlighted in the pioneering spirit, especially from 2018 onwards, with the organization of events such as Embrapa's first hackathon – programming marathon – with the theme of automatic diagnosis of diseases in agricultural crops, and conducting meetings related to the theme of data science and digital agriculture. Among these events, SBIAgro Conect@ stands out, focused on promoting qualified networking between institutions, companies, accelerators, investors, developers and users of ICT.

Communication professionals also supported the construction of the methodology for the first acceleration program for startups that work with agricultural technologies (agtechs), the TechStart Agro Digital. The relationship and interaction with the program's startups were facilitated precisely by the work of these professionals, from the program's conception up to the selection and interview phases, in addition to mentoring specialized in communication techniques, with the use of digital media and the creation of exclusive relationship channels between members of startups and communicators.

The communication area also plays an important role for supporting various events aimed at innovation and for conducting mentoring. Among the programs carried out with the participation of communicators are Sebrae Startup SP, Samsung Creative Startups, InovaPork, carried out by Embrapa Swine and Poultry, and the Bridges for Innovation, organized by Embrapa's Department of Innovation and Business (SIN).

Digital media was one of the strategies adopted, which has greater reach and faster and more qualitative delivery. It is noteworthy that traditional communication vehicles, which have a limited reach and are usually aimed at the technical-scientific audience of these events, were not replaced, but digital and interpersonal communication, guided by a closer relationship, gained focus.

As shown, the communication actions are created by the NCO with a planned and integrated approach, permeating all institutional processes and the different areas of activity of the Unit, in order to effectively achieve the institution's objectives. Also noteworthy are the relationship strategies with the internal public, for which the NCO develops targeted communication actions and supports events, in line with the personnel management area.

> We understand integrated communication as a philosophy that directs the convergence of different areas, enabling a synergistic action. It presupposes a combination of institutional communication, marketing communication, internal communication and administrative communication, which form the mix, the combination of organizational communication (Kunsch, 2003, p. 150).

Educommunication to support collective creation

With technological progress, the convergence and integration of new media, the production and distribution of information are marked by significant changes. Although throughout its history man has used instruments to communicate, the universalization of the means and resources of the contemporary world is now especially unique, pressing the communication means and ICTs to configure a new model of Man and Society (Gómez; Aguaded, 2011, p. 4).

Society organized in "virtual networks" is configured based on new spaces that can favor the process of sharing and creating content, with a collaborative and more participative approach. From the connectivity, mobility and portability resources of the web, any citizen can become a producer and consumer of information (Pereira, 2013, p. 1).

In the information society, this new paradigm of collective construction guided by the convergence of media, which enables organizing in networks, makes subjects to no longer be mere consumers-receivers, making them individuals-consumers. This new approach, based on constructivist learning, takes into account the numerous resources provided by ICT as possibilities for active appropriation, based on the individual's autonomy and cooperation, including creation, authorship, human development and innovation.

Technology offers enormous potential for interactivity, but there is considerable complexity in these mutual interactions mediated by technological resources, which include "reciprocal action, cooperation and collective creation" (Primo, 2008, p. 148) which cannot be ignored. In this regard, the exchange of knowledge in a digital world can be enriched by a pedagogical proposal designed to support the collective construction of knowledge and encourage the subject's critical view and autonomy.

The communication/education interrelationship constitutes a field of social intervention, called educommunication, characterized by a political action for the contribution of an ethical conscience and a pragmatic approach aimed at transforming society. This action is based on the formation of critical, participative citizens who are part of the social environment and the implementation of social utopias of quality education and participatory and democratic communication (Schaun, 2002).

> Education is a science concerned with the formation and constitution of the human being as a subject, in other words, a being who thinks about his reality, reflects and acts on it, transforming the environment in which he lives. The proximity of the fields of education, communication and technology favors multiple views on the human condition and development, enabling the shared construction of information, knowledge and experiences in a context of exchanges and social interactions that can encourage the exercise of citizenship (Pereira, 2013, p. 2).

According to the Communication and Education Center (NCE) of the University of São Paulo (USP)[10], studies carried out on the interrelationship between communication and education point to the emergence of a social intervention field characterized by offering theoretical-methodological support that allows social agents to understand the importance of communication actions for human coexistence, the production of knowledge and the elaboration and implementation of collaborative projects for social change.

> The concept of educommunication actually proposes the construction of open, dialogical and creative communicative ecosystems in educational spaces, breaking the hierarchy of knowledge distribution, precisely because of the recognition that all people involved in the flow of information are producers of culture [...]. Therefore, the goal of educommunication is to construct citizenship, based on the basic assumption of everyone exercising the right to expression and communication (Núcleo de Comunicação e Educação, 2012).

From the point of view of interactions, it can be said that communication is a social process focused on expanding the capacity of individuals to interrelate as active agents in the environment in which

[10] NCE – USP. Available at: http://www.usp.br/nce/onucleo

they live, promoting changes in their reality based on these interactions. Educommunication emerges from this conception, which is based on the communication/education interrelation. It is about adopting a perspective of educational communication that is designed as a dialogical relationship of educommunicative action, defined as a "field of dialogue, space for critical and creative knowledge, for citizenship and solidarity" (Soares, 2000, p 12).

For subjects to effectively take ownership of productive processes, it is important for learning environments to be characterized by the constructivist approach and that they favor intellectual exchanges, the development of thought, cooperation, leading to reflection on actions, and the awareness that determines their moral and intellectual autonomy (Piaget, 1998). Therefore, technological resources must be incorporated into a transformative critical-reflective pedagogical proposal, which promotes articulations between the educator's knowledge and their practice, favoring learning based on interaction, collaboration and cooperation between students and educators.

Thus, the communicators at Embrapa Digital Agriculture have also supported technology transfer projects that are guided by a vision supported by the concepts of educommunication, considering the autonomy and critical perception of the subjects involved.

Dialogical communication to support sustainable development and popularization of science

The United Nations (United Nations, 2015) established, in the 2030 Agenda, the 17 Sustainable Development Goals (SDGs) to support the construction and implementation of public policies worldwide. Embrapa understands that agricultural research plays an important role in achieving the 169 goals of the Agenda aimed at human development, as "food production in line with the generation of sustainable innovation in the field contributes to improving the quality of people's lives, to reduce the price of basic food and to export Brazilian products" (Embrapa, 2020a).

To align its work with the international commitment, Embrapa carried out a comprehensive evaluation of its agricultural research and innovation program, mapping how the Impact Axes and the 12 Strategic Objectives described in its VI Master Plan are related to the 17 SDGs. The Company understands that this is a way of being accountable to society and showing alternatives for an increasingly sustainable agriculture, serving as a model for other countries.

Also with the objective of contributing to the SDGs and subsidizing strategic actions in science, technology and innovation, Embrapa established the Strategic Intelligence System, Agropensa, responsible for widespread monitoring of the external environment focused on capturing signals and trends to elaborate scenarios and future visions for Brazilian agriculture (Embrapa, 2018b). The leading role of consumers is one of the megatrends indicated in the document Vision 2030: the future of Brazilian agriculture:

> The exponential growth of Information and Communication Technologies (ICT) applications means that individuals have much more power to influence food production chains, and their food consumption decisions are based on continuous interactions with production agents, which, together with the expanding market niches, consubstantiate this megatrend. In this context, the convergence of accelerated global movements intensify the use of digital platforms in consumer relations, the co-creation of products and services and the growing access to information by digital means. Safe, traceable, healthy and produced through sustainable processes will be increasingly valued (Embrapa, 2018b, p. 12).

At Embrapa, numerous solutions generated by agricultural research are made available on a daily basis in different formats, so as to broadly disseminate the knowledge produced. Moreover, the Company's communication is inserted in this context. Embrapa is concerned not only with production, but with sustainable consumption that provides a better quality of life for the population:

> Brazilian agricultural research faces many challenges regarding sustainable development, including systematizing all the knowledge generated, standardizing and integrating methods, translating knowledge into solutions to be directly appropriated by society, sufficient financial resources, proximity of scientists and decision makers, and other challenges. The mission of the Brazilian Agricultural Research Corporation (Embrapa), based on the results of its research, is to contribute to the sustainable development of agriculture (Palhares et al., 2018, p. 7).

The Company is also aware of the technological revolution of the last decades, which is marked by an accelerated computerization and digitalization process of analog procedures and by the development of new information and communication technologies (Antunes et al., 2018, p. 77). As a promoter for generating technical-scientific knowledge, Embrapa needs to make it accessible to the different segments of its stakeholders, from rural producers to consumers.

Several lines of study theorize about communication forms that are more compatible with the appropriation of knowledge and learning processes. Both are essential to translate scientific advancement and influence the life of rural producers, especially for those who belong to the uncomfortable statistics verified by the Brazilian Institute of Geography and Statistics (IBGE) on education in the countryside: approximately 80%, that is, the vast majority of rural producers in Brazil have primary education or have never attended school[11].

Considering that 77% of agricultural establishments are in the category of family farming[12], one of the challenges seen is understanding how communication can support the process of translating knowledge into solutions to be appropriated by the beneficiaries. And, educommunication is exactly one of the lines that seeks to bring together education, communication and technology for the shared construction of knowledge.

Based on the concept of a dialogical and transforming education, focused on solidary construction and knowledge sharing, it breaks with the vertical model of disseminating and transferring content to a liberating education, grounded on a process of analysis and reflection, in which subjects learn to think and, thinking, they are capable of promoting changes in their reality (Freire, 1982).

At Embrapa, some experiences and perceptions for ensuring the appropriation of knowledge are identified, even if they are incipient, given the wide range of actions provided by organizational communication. One of the examples related to collective production resulted in the publication of Colegio Povos e Comunidade Tradicionais, launched in 2017, which brings together reports on works carried out with rural communities and their traditional knowledge (Antunes et al., 2018, p. 77).

Another initiative also took place within the scope of the Pedagogical Production Methodology of Multimedia Materials with an Agroecological Focus on Family Farming (Pedagroeco[13]). Coordinated by Embrapa, with the participation of four Decentralized Units (Semi-arid, Coastal Tablelands, Middle-North

---

[11] According to data from the 2017 Agricultural Census, the rural producer has higher education in only 5.58% of agricultural establishments in Brazil. In high school, adding the numbers of the scientific article,EJA, and high school technician, they represent 14.95% (IBGE, 2017).

[12] In the IBGE classification for the Agricultural Census, family farming has different dynamics and characteristics from non-family farming. In it, property management is shared by the family and agriculture is the main source of income (IBGE, 2017).

[13] To learn about the project, go to (Embrapa, 2020b).

and Cotton), the project was carried out in partnership with civil society organizations. The objective was to develop a methodology to encourage young rural students to use ICT, in the context of family farming and agroecology.

The Organizational Communication Nucleus of Embrapa Digital Agriculture coordinated actions to train multipliers in the state of Piauí. In all, more than 200 young people were trained in five states in the Brazilian semiarid region. Organized as workshops, as shown in Figure 3, the training included the Griô Pedagogy as a policy-methodological reference (Pacheco, 2014), which conducted the actions throughout the entire process of training young people.

Photos: Magda Cruciol e Rosival Dias

Figure 3. Young students participate in multimedia production workshops.

The methodological approach[14] ensured participatory processes anchored in the transformative and autonomy perspective advocated by Freire (2011), and the material produced resulted in 18 videos. The experience and interactions among Pedagroeco participants showed how communication and the use of ICT, anchored in the Griô Pedagogy action model, played an expressive role in the knowledge appropriation processes and in the affirmation of identity of the young people involved in the project.

This perspective of network communication, with its characteristics that break with traditional notions of time and space, allows the reconfiguration of public powers, since "only looking at the technological bias in educational processes can be related to emptying the cultural dimension in question, as well as a fragmented thought of knowledge" (Ferreira, 2019, p. 10).

[14] The Griô Pedagogy is the pedagogy of experiencing affective and cultural rituals that facilitate dialogue between ages, groups and communities, through an enchanting, experiential, dialogic and shared method for the development of knowledge (Pacheco, 2006). Discover the history of the Griô Pedagogy. Available at: www.graosdeluzegrio.org.br.

Final considerations

The progress of information and communication technologies (ICT), especially from the 20th century onward, transformed human communication, democratizing access to the new media. As regard communication and the process of scientific dissemination, new media are accessible instruments that enhance the generation of knowledge in a collaborative way and facilitate the popularization of science.

Currently, a unidirectional communication process is no longer accepted, as citizens increasingly want to hear and be heard. In this context, the individual is at the center of decisions, and at the same time is a producer and consumer of information and knowledge, since democratizing the means of communication allows anyone to be a source of information.

This reality impacts the way in which companies and research institutions relate to society and their strategic audiences, in a disruptive way, since the technological resources of connection and interactivity expand social interaction and collective construction in the so-called cyberspace. Thus, there is growing concern with a more interactive digital communication for a more effective sharing of information and scientific dissemination.

With the new media, content production has grown exponentially. However, the process of scientific dissemination and technology transfer requires ensuring the quality of information, as they strongly impact the lives of citizens. Therefore, it is up to research institutions to find innovative ways to interact with society and its strategic audiences, implementing new mechanisms for scientific dissemination that include interactivity and collective participation, with social responsibility.

The progress of the means of communication, driven by technological development, has produced changes in habits and behavior. This requires a more educated and thoughtful attitude with the media based on inclusion, ethics and citizenship, supported by a pedagogical approach with the media. The mediation of the fields of education and communication, known as educommunication, seeks to encourage integration, reflection and the production of ethical contents that promote social transformations for the subjects involved in communicative processes.

Public research, development and innovation institutions are committed to promoting the dissemination of the knowledge produced as well as transferring the technologies generated, to promote the development and scientific literacy of society. Aware of its external recognition as reference in tropical agriculture research and of the excellent results in communication with society, Embrapa encourages innovation in its communication practices and communication channels focused on new media and technological resources.

Therefore communication must contribute to the dissemination of science and its results through innovative actions that consider a new way of talking to society, interacting through the exchange of knowledge, taking advantage of the potential of new technologies, which are increasingly within reach for the greatest number of people.

Thus, several actions have been developed to share information and promote a better relationship with the rural and urban public. Information and communication technologies can be used to help restructure channels that manage information and communication flows, in order to enhance the development of new social media for scientific dissemination.

Among the actions carried out, the most effective use of new technologies stands out, which add interactivity resources and benefit network performance, allowing the public not only to know but to interact and contribute to the production of knowledge, reflecting on their role as interactive user under a new perspective of participation and collective construction.

This can help to better understand strategic audiences, through the dissemination of research results, technologies, products and services, in an accessible and interactive approach. Embrapa already has several strategies to bring science to the population through events, radio and television programs and internet portals. However, it is also necessary to study new methods to adapt the language and facilitate people's access to scientific knowledge, promoting interactivity, in addition to capturing their demands, in a dialogical relationship.

Embrapa Digital Agriculture develops several initiatives to bring the results of research and technologies, products and services to the public knowledge, contributing to the dissemination of ICT, especially in the rural sector. With the production of publications, radio and television programs, events and participation in agricultural fairs and exhibitions, the research center stands out as a reference in the matter of digital agriculture, including a strong presence in the media.

In the research area, the Unit has been expanding its collaborative operations with the private sector, through joint projects to develop technological solutions. Regarding technology transfer, partnerships were also expanded, especially with companies and startups in the digital ecosystem. It is up to the communication area to find new ways to build and strengthen more open and integrated relationships with the institution's audiences, supporting actions for a more connected and digital agriculture.

The challenge for communication professionals is to discover innovative ways to disseminate the science produced to a society that increasingly demands for real-time information. It is expected that public research institutions will be able to promote innovations in their communication processes, which will help to improve the relationship with society and contribute to expanding the knowledge and participation of their audiences in relation to the development of research and its results.

References

ANTUNES, E.; OLIVEIRA, V. B. V. de.; RAMOS, R. G. C.; ESCOBAR, J.; COSTA, J. R.; SOUZA, R. B. de. Informação para a ação cidadã e promoção do desenvolvimento sustentável. In: PALHARES, J. C. P.; OLIVEIRA, V. B. V.; FREIRE JUNIOR, M.; CERDEIRA, A. L.; PRADO, H. A. do. **Consumo e produção responsáveis**: contribuições da Embrapa, 2018. p. 75-87. E-book. Available at: https://ainfo.cnptia.embrapa.br/digital/bitstream/ item/184241/1/ODS-12-Consumo-e-producao-responsaveis.pdf. Accessed on: 18 May 2020.

CAJAZEIRA, J. E.; CARDOSO, C. Comunicação e inovação: correlações e dependências. In: ABRAPCORP, 3., 2009, São Paulo. **Anais**... São Paulo, 2009. Available at: http://www. abrapcorp.org.br/anais2009/pdf/GT2_Cajazeira.pdf. Accessed on: 20 May 2020.

CARDOSO, C. (org.). **Comunicação organizacional hoje**: novas tecnologias, novas perspectivas. Lauro de Freitas: UniBahia, 2002.

CARDOSO, O. O. Comunicação empresarial versus comunicação organizacional: novos desafios teóricos. **Revista de Administração Pública**, v. 40, n. 6, p. 1123-44, nov./dez. 2006. DOI: 10.1590/S0034-76122006000600010.

DINIZ, F. **Pandemia na agricultura**: porteira aberta para a digitalização. Brasília, DF, maio 2020. Available at: https://www.embrapa.br/busca-de-noticias/-/noticia/52363786/pandemia-na- agricultura-porteira-aberta-para-a-digitalizacao?link=agencia. Accessed on: 19 May 2020.

DUARTE, J.; BARROS, A. T. (ed.). **Comunicação para ciência, ciência para comunicação**. Brasília, DF: Embrapa Informação Tecnológica, 2003. 338 p.

EMBRAPA. Assessoria de Comunicação Social. **Política de comunicação**. 2. ed. Brasília, DF, 2002.

EMBRAPA. **Metodologia de produção pedagógica de materiais multimídia com enfoque agroecológico para a agricultura familiar**. 2020b. Available at: https://www.embrapa.br/ busca-de-projetos/-/projeto/212110/metodologia-de-producao-pedagogica-de-materiais- multimidia-com-enfoque-agroecologico-para-a-agricultura-familiar. Accessed on: 6 Jul. 2020.

EMBRAPA. **Objetivos de Desenvolvimento Sustentável**. 2020a. Available at: https://www. embrapa.br/objetivos-de-desenvolvimento-sustentavel-ods. Accessed on: 18 May 2020.

EMBRAPA. Secretaria de Gestão e Desenvolvimento Institucional. **VI Plano Diretor da Embrapa**: 2014-2034. Brasília, DF, 2015.

EMBRAPA. Secretaria de Inovação e Negócios. **Política de inovação da Embrapa**. Brasília, DF, 2018a.

EMBRAPA. **Visão 2030**: o futuro da agricultura brasileira. Brasília, DF, 2018b.

FERREIRA, D. L. de F. Pedagroeco e identidade: uma análise de conteúdo audiovisual em rede. In: CONGRESSO NACIONAL DE EDUCAÇÃO, 6., 2019, Fortaleza. **Anais**... Fortaleza [s.n.], 2019. Available at: http://www.editorarealize.com.br/revistas/conedu/trabalhos/ TRABALHO_EV127_MD1_SA19_ID14901_03102019015710.pdf. Accessed on: 18 May 2020.

FONSECA JÚNIOR, W. C. F.; FRAGALLE, E. P.; SILVA, H. D.; FILHO, R. C. P. A comunicação na Embrapa: do difusionismo à comunicação como inteligência organizacional. **Prisma.com**, n. 8, p. 77-92, 2009. Available at: https://www.brapci.inf.br/index.php/res/v/71567. Accessed on: 19 May 2020.

FREIRE, P. **Extensão ou comunicação?** São Paulo: Paz e Terra, 2011.

FREIRE, P. **Pedagogia do oprimido**. Rio de Janeiro: Paz e Terra, 1982.

GALERANI, G. S. M. **Avaliação em comunicação organizacional**. Brasília, DF: Embrapa/Assessoria de Comunicação Social; Londrina: Embrapa Soja, 2006. 142 p.

GÓMEZ, A. H.; AGUADED, J. I. Recomendaciones para el desarrollo de la alfabetización mediática em Brasil: propuestas desde la experiencia europea. **RESGATE**, v. 19, n. 22, p. 3-15, jul./dez. 2011. DOI: 10.20396/resgate.v19i22.8645713.

IBGE. **Censo agro 2017**. Available at: https://censos.ibge.gov.br/agro/2017/templates/censo_agro/resultadosagro/index.html. Accessed on: 21 May 2020.

KUNSCH, M. M. K. (org.). **Comunicação organizacional**: histórico, fundamentos e processos. São Paulo: Atlas, 2009.

KUNSCH, M. M. K. **Planejamento de relações públicas na comunicação integrada**. São Paulo: Summus, 2003. 422 p.

KUNSCH, M. M. K. **Relações públicas e modernidade**: novos paradigmas na comunicação organizacional. São Paulo: Summus, 1997.

LÉVY, P. **Cibercultura**. Rio de Janeiro: Editora 34, 1999.

MANUAL de Oslo: proposta de diretrizes para coleta e interpretação de dados sobre inovação tecnológica. 3. ed. [S.l.]: Organização para Cooperação e Desenvolvimento Econômico, 2005. Available at: https://www.finep.gov.br/images/apoio-e-financiamento/manualoslo.pdf. Accessed on: 21 May 2020.

MARCHIORI, M. **Cultura e comunicação organizacional**: um olhar estratégico sobre as organizações. São Caetano: Difusão Editora, 2011.

MELLO, S. F. M. **Comunicação e organizações na sociedade em rede**: novas tensões, mediações e paradigmas. 2010. 271 f. Dissertação (Mestrado em Ciências da Comunicação — Universidade de São Paulo, São Paulo.

NAÇÕES UNIDAS. **17 Objetivos para transformar nosso Mundo**. 2015. Available at: https://nacoesunidas.org/pos2015. Accessed on: 6 July 2020.

NÚCLEO DE COMUNICAÇÃO E EDUCAÇÃO. Available at: http://www.usp.br/nce/onucleo. Accessed on: 10 Feb. 2012.

OLIVEIRA, F. I. **Ciência e tecnologia na comunicação social de instituições governamentais**. 1998. 168 f. Tese (Doutorado em Ciências da Comunicação — Universidade de São Paulo, Escola de Comunicações e Artes, São Paulo.

OLIVEIRA, I. L. A Abrapcorp e o impulso da pesquisa em Comunicação Organizacional e Relações Públicas. In: LEMOS, E.; FARIAS, L. A. de; SCROFERNEKER, C. (org.). **Margarida Maria Krohling Kunsch**: consolidação da comunicação organizacional e das relações públicas no brasil. São Paulo: Intercom, 2013. p. 59-67.

OSLO manual 2018: guidelines for collecting, reporting and using data on innovation. 4th ed. Paris: Eurostat; Luxembourg: OECD Publishing, 2018. (The measurement of scientific, technological and innovation activities). DOI: 10.1787/9789264304604-en.

PACHECO, L. Dossiê pedagogia Griô: escritas Griô, 2006. **Revista DIVERSITAS**, ano 2, n. 3, set. 2014/mar. 2015.

PACHECO, L. **Pedagogia Griô**: a reinvenção da roda da vida. Lençóis: Grãos de Luz e Griô, 2006.

PALHARES, J. C. P.; OLIVEIRA, V. B. V.; FREIRE JUNIOR, M.; CERDEIRA, A. L.; PRADO, H. A. do. (ed.). **Consumo e produção responsáveis**: contribuições da Embrapa. Brasília, DF: Embrapa, 2018. E-book. (Objetivos de desenvolvimento sustentável, 12). Available at: https://ainfo.cnptia.embrapa.br/digital/bitstream/item/184241/1/ODS-12-Consumo-e-producao-responsaveis.pdf. Accessed on: 18 May 2020.

PEREIRA, N. R. **Educomunicação na Pedagogia**. 2013. 102 f. Dissertação (Mestrado em Educação) — Universidade Estadual de Campinas, Instituto de Educação, Campinas.

PEREIRA, N. R.; TORRES, T. Z.; SANTOS, A. D. dos. Comunicação corporativa na era da televisão digital interativa. In: CONGRESO INTERNACIONAL COMUNICACIÓN 3.0, 2., 2010, Salamanca. **Libro nuevos medios, nueva comunicación**. Salamanca: Universidad de Salamanca, 2010. p. 1-12. Available at: https://ainfo.cnptia.embrapa.br/digital/bitstream/item/23608/1/072.pdf. Accessed on: 19 May 2020.

PIAGET, J. **Sobre pedagogia**. São Paulo: Casa do Psicólogo, 1998.

PRIMO, A. **Interação mediada por computador**: comunicação, cibercultura, cognição. Porto Alegre: Sulina, 2008.

RODRIGUES, N.; MAIO, A.; ROSSO, G.; GONÇALVES, D.; SILVA, J.; FRAGALLE, E. P. **Pesquisa contribui para transformação digital da agricultura brasileira**. Brasília, DF: Embrapa, 2020. Available at: https://www.embrapa.br/busca-de-noticias/-/noticia/51706860/pesquisa- contribui-para-transformacao-digital-da-agricultura-brasileira. Accessed on: 19 May 2020.

SABBATINI, J. F. **Comunicação organizacional e governança corporativa**: uma intersecção possível? 2010. 214 f. Tese (Doutorado) — Universidade de São Paulo, São Paulo. DOI: 10.11606/T.27.2010.tde-05112010-113946.

SANTOS, T. C. A comunicação nas teorias das mídias digitais no Brasil e nos Estados Unidos. **Revista Famecos**, v. 23, n. 1, jan./abr. 2016. DOI: 10.15448/1980-3729.2016.1.21490.

SCHAUN, A. **Educomunicação**: reflexões e princípios. Rio de Janeiro: Mauad, 2002.

SOARES, I. de O. Educomunicação: um campo de mediações. **Revista Comunicação & Educação**, p. 12-24, set./dez., 2000. DOI: 10.11606/issn.2316-9125.v0i19p12-24.

TORRES, T. Z.; PEREIRA, N. R.; GÂMBARO, B. **Comunicação Digital nas Redes de Pesquisa**. In: CONGRESSO BRASILEIRO DE PESQUISADORES EM COMUNICAÇÃO ORGANIZACIONAL E RELAÇÕES PÚBLICAS, 6., 2012, São Luís. **Anais**... São Luís: Abrapcorp, 2012. Available at: https://ainfo.cnptia.embrapa.br/digital/bitstream/item/75338/1/MT27.pdf. Accessed on: 22 May 2020.

WOLTON, D. **Informar não** é comunicar. Porto Alegre: Sulina, 2010. 96 p.

Photo: Joa Souza (AdobeStock)

15

Driving forces for Brazilian agriculture in the next decade

Implications for digital agriculture[1]

Geraldo Bueno Martha Júnior

Introduction

Brazilian agriculture is not isolated in an economic vacuum. Agriculture influences and is influenced by its surroundings. Given Brazil's status as an important player in global agricultural markets, these dual avenues of influence expand to include regional and global dimensions.

In the last decade, different aspects of digital agriculture and its sectoral applicability were presented to the productive sector. These advances include sensors, images from mobile devices, drones and satellites, internet of things (IoT), Big Data, computer vision, simulators, optimization algorithms, and artificial intelligence. The integration of these technologies has the potential to transform agriculture and livestock production. This change would be translated into improvements in management and decision-making processes, as well as efficiency gains at different stages of production.

The expansion of digital agriculture's relevance in production processes in the next decade seems, therefore, inevitable. This reflects the potential benefits these digital technologies and services can add to the agricultural production chain, at gradually lower costs, and with increasing efficiency. However, despite

[1] This chapter is a revised version of the article originally published as: MARTHA JÚNIOR, G. B. Forças motrizes para a agropecuária brasileira na próxima década: implicações para a agricultura digital. In: MASSRUHÁ, S. M. F. S.; LEITE, M. A. A.; OLIVEIRA, S. R. M.; MEIRA, C. C. A.; LUCHIARI JUNIOR, A.; BOLFE, E. L. (ed.). **Agricultura digital**: pesquisa, desenvolvimento e inovação nas cadeias produtivas. Brasília-DF: Embrapa, 2020. p. 358-379.

the growing availability of solutions in information technology (IT) presented to the market, actors in the private sector sometimes point to the lack of objectivity and applicability of the information presented. They frequently point to the lack of economic analysis supporting the proposed digital solution.

These findings suggest that general approaches to digital agriculture have been prioritized. Although interesting for the dissemination of technologies, that approach does not provide objective elements for the producer's decision-making. Digital solutions need to be analyzed based on the grand challenges and opportunities that Brazilian agriculture faces in order to be effective in the real world. These important issues, roughly speaking, can be grouped into micro, macro, biological, and management dimensions. The environmental dimension is transversal to these other dimensions.

This chapter proposes to briefly reflect on these topics, emphasizing the key driving forces on the demand and supply sides that shape and strongly interact with major challenges and opportunities observed in Brazilian agriculture. In addition to this introduction, the second section will discuss the main relevant driving forces to agriculture, outlining situations of interest for the potential insertion of digital technologies in agricultural systems. The third section presents digital agriculture in the context of the production chain. The fourth, and final, section explores some perspectives that are beginning to take shape in the domestic and international environments and that may bring about pressures, but also opportunities, for Brazilian agriculture (and for digital solutions) in the near future.

Relevant driving forces to agriculture and its digital transformation

Thinking about the future, whether in the short term (operational level) or in the long term (strategic level), is an intrinsic human trait (Harari, 2015). Recent studies have synthesized future opportunities and challenges as regard the sectoral, regional and temporal scope (Embrapa, 2014, 2018; Boumphrey; Brehmer, 2017; FAO, 2017; The Economist Intelligence Unit, 2019; Global, 2019; Kirova et al., 2020). Terms such as driving forces (or drivers), trends and megatrends are intertwined, sometimes get confused, and have been used to describe potential forces and impacts, either positive or negative, to the public and private sectors.

An interesting contribution to understanding these terms (and their application) was presented by Boumphrey and Brehmer (2017). According to these authors, the driving forces shape the megatrends, which are, thus, a second-order phenomena. Megatrends are well-established forces of medium- to long-term influence (5 to 15 years) in a world explained by the driving forces, whose evolution permeates different sectors of the economy and takes different forms over time. For a force to be considered a megatrend it must have multi-sector relevance. If this force is sector-specific, despite any potential to influence the sector, it is not considered a megatrend in the strict sense of the concept, instead, it is considered a trend.

In many situations, the description of how these megatrends and trends are relevant to planning within its different time horizons is often too vague, which does not contribute to decision-making process. Alternatively, reflections on future opportunities and challenges, based on driving forces, are generally simpler and more objective, in addition to the advantage of having available a robust set of quantitative tools. Thus, when appropriate, credible and verifiable analyses to support planning and decision-making at their different levels can be carried out.

The driving forces refer to changes induced by factors of natural or human origin that occur in (agro) ecosystems. These changes can have a direct or indirect effect (Nelson, 2005). The driving forces of direct action are climate change, land use changes (such as deforestation), plant's nutrient use efficiency, and the incidence of pests and diseases. Those with indirect action operate more diffusely, altering at least one direct driving force, which will eventually influence the (agro)ecosystem processes. The most important indirect driving forces are those of a demographic, economic, socio-political, scientific-technological, cultural, and religious nature (Nelson, 2005).

Driving forces act at global, national and local levels and present interactions between these different scales (Hazell; Wood, 2008; Embrapa, 2014). Awareness of these driving forces, their importance, potential impacts, and interactions was strongly stimulated from the 1990s onwards, with communication possibilities opened by the internet along with the declining costs in IT equipment and services.

On a global scale, factors such as climate change, globalization, foreign trade, international prices of food and inputs for agricultural production (including energy), policies (agricultural, environmental, etc.), among others, interact with factors at the national level. At the country level, there are other influencing factors, such as those of macroeconomic and political nature, legal issues, market channels, per capita income, and urbanization. The national scale mainly, but also the global scale, further interacts with the local level, in which factors such as infrastructure, access to markets, agroecological zoning (and environmental restrictions), employment, incentives (public policies) and disincentives (taxes) are shown to be active (Hazell; Wood, 2008).

Determining the relative importance of each force, direct or indirect, as well as their interactions on the impacts observed at a given location is not a trivial task. Nevertheless, strategies to respond to eventual undesirable changes will be motivated by the ability of local agents to influence these drivers. Thus, notwithstanding the difficulties, it is important to focus on the factors that most influence opportunities and challenges at the local level (Hazell; Wood, 2008).

At the local level, a farmer's perceptions about the relative importance of these forces in terms of pressures on available resources, restrictions and opportunities to business, subject to his/her individual values, guide the course of decision-making. The farmer's perspectives regarding opportunity costs and risks involved in decision making are unique to a given farmer-farm combination. This happens because the quantity and quality of resources (land, labor, physical and human capital) and inputs, as well as the relative prices involved, vary on a case-by-case basis.

Therefore, analyzing these driving forces provides elements that support reflections and actions in the public and private sector, thus guiding the definition of plans, goals, and objectives. For research and technology transfer, they indicate the direction of priorities and strategies for programs and projects. The growing possibilities offered by information and communication technologies (ICT) allow greater interaction between relevant actors and agents with interest in agriculture. This interaction has increasingly included other important stakeholders, from Brazil and abroad, in discussions related to agriculture and its production environment.

An interesting way to analyze the driving forces and their impacts is through the economy's supply and demand model. This analysis framework incorporates many of the signs provided by the different driving forces. Thus, the main factors related to the increase in demand for agricultural products are variations in population growth and per capita income. Factors such as the urbanization rate and the preferences of individuals, subject to cultural, socio-political, and religion, also influence the shift in demand.

Demand-side factors, which reflect society's main preferences, influence decisions regarding the agricultural production process (supply). The main driving forces related to variations in agricultural supply are the availability of technologies and the costs of production. Digital solutions (technological supply) can contribute favorably to efficiency gains in management and in the production process. This makes room for reducing production costs, in addition to creating opportunities for the expansion of income for farmers and, more broadly, for the actors involved in the production chain.

Key-driving forces on the demand side

The two main shifters of the demand curve are population growth and per-capita income. Hertel and Baldos (2016) estimated that between 1961 and 2006 a population growth rate of 1.7% per year in the period explained 83.7% of the variation in agricultural demand. The variation in income in the period, 1.4% per year, accounted for the remaining 16.3%.

The most recent projections from the United Nations indicated that the world population in 2050 will be in the range of 9.4 billion to 10.1 billion, representing an average expansion of 0.8% per year compared to the population of 2019, of 7.7 billion (United Nations, 2019). Per capita income will grow in importance as a driver explaining the expansion of demand in the decades ahead (Guillemette; Turner, 2018).

Hertel and Baldos (2016) were able to capture these developments using as reference the period from 2006 to 2051. In this interval, the simulated population and per capita income growth rates were 0.8% per year and 2.1% per year, respectively. Under these conditions, the projected increase in agricultural demand for the 2006–2051 period due to variations in population growth and per capita income are, respectively, 35.4% and 149% higher compared to the 1961–2006 period. The decomposition of these forces for the 2006 – 2051 horizon revealed that population would explain 55% of the growth in agricultural demand, while per capita income would explain the remaining 45%. In the period extending to 2050, most of the global population variation will be concentrated in Africa (59.2%) and Asia (33.5%) (United Nations, 2019). Income, as an explanatory factor for the increase in agriculture demand, has Asia as the main region, followed by Africa (64.6% and 21.1% of the total variation, respectively) (Hertel; Baldos, 2016).

These forces, combined with changes in eating habits resulting from the expansion of the world middle class (Kharas, 2017) and the growing rate of urbanization (Warr, 2019), will sustain the demand for agricultural products in the next decade. The demand growth rates for products with higher income elasticity, such as animal protein, should be higher in emerging countries (Godfray et al., 2018), whose population seeks to reach consumption levels close to those observed in rich countries. The aging of the population (United Nations, 2019) and possible changes in diet preferences motivated by claims for healthier food and environmental issues (Godfray et al., 2018) should reduce the rate of increase in the demand for animal protein in the future. However, eventual structural and large-scale transformations are complex processes that take decades to be carried out (Elzen et al., 2012).

Agricultural demand and covid-19

The pandemic has caused a unique situation in human history, both in characteristics and proportions, as its deleterious consequences have simultaneously affected health and economy in different regions of the world.

Expectations of economic growth in Brazil and in the world have rapidly deteriorated over the first semester of 2020. The world's Gross Domestic Product (GDP) for 2020 showed a negative variation of -3.1% (International Monetary Fund, 2021), representing a 6.1 p.p decrease compared to the pre-crisis scenario (variation from +3% to -3.1%) (European Commission, 2020a, 2020b, International Monetary Fund, 2021). The Brazilian economy shrank 4.1% in 2020 (International Monetary Fund, 2021), which represents a drop of 6.4 p.p compared to pre-covid-19 growth projections.

The economic recovery has been slow and quite uneven between countries, between regions of a given country, and between different sectors of the economy. In 2021, the world economy grew 5.9% and the world GDP is projected to expand 4.4% in 2022 (International Monetary Fund, 2022). However, the Brazilian economy has not performed at the same pace. It grew 4.7% in 2021, and it is projected to stay close to stagnation in 2022 (e.g. GDP variation of only 0.3%) (International Monetary Fund, 2022).

There is still a great deal of expectation about the effectiveness of the economic emergency measures implemented by governments and their future impacts. In the Brazilian case, the latest Focus Report available[2] from February 21, 2022, projected an official inflation rate (e.g. IPCA) of 5.56% for 2022. That inflationary pressure is forcing Brazil's Central Bank to raise the basic interest rate (e.g. SELIC) in the economy, that is already projected to reach 12.25% in 2022 according to the latest Focus Report. The situation could worsen given the uncertainties regarding the possibility of new Covid-19 variants cycles over the next few months.

Agriculture has shown to be resilient and one of the few sectors capable of sustaining a positive variation in sectoral GDP, having registered a 2% growth rate in 2020 (IBGE, 2021). However, the impacts of the pandemic on the food sector demand in the short term and immediately after the pandemic, are likely to be asymmetric among its subsectors. This asymmetry reflects, among others, restrictions due to social distance, the deterioration of per capita income in the coming months, the availability of substitute products and the income-elasticity of demand for products.

The "food at home" group has been the least impacted by the Covid-19 crisis. Even so, according to income stratum, a slowdown in demand may reach -3.7% over the next year, compared to pre-crisis expectations[3]. Food groups with greater income-elasticity, such as meals away from home, organic, and animal protein segments are expected to be more heavily impacted. Depending on the case, the drop in demand over the next year may exceed 9.0% compared to the pre-crisis scenario (Martha Júnior, 2020).

Globalization and international trade remain essential for economic development. However, it is expected that there will be growing support to nationalize the production of a greater variety of inputs in order to support the production process within national borders. At the same time, there is a trend towards more demanding consumers asking for more qualified information on agricultural products, services, and environmental and social variables related to production (and eventual externalities).

Within this scenario, it is plausible that the growth of exports may face a more intense and competitive environment, particularly in view of the sharp per capita income drop caused by the covid-19 pandemic. Despite these uncertainties, for the time being, there are positive expectations for the expansion of Brazil's participation in global agricultural markets over the next decade. However, meeting the increasingly diverse and rigorous demands and capturing opportunities is neither negligible nor free of challenges.

Some digital technologies are expected to significantly grow in this context. The previously discussed trends point to a rapid expansion in the demand for digital solutions in monitoring, traceability of

[2] The Focus Report, by the Central Bank of Brazil (BCB), "[...] summarizes the statistics calculated considering market expectations collected up to the Friday prior to its release. It is released every Monday. The report presents the graphic evolution and weekly behavior of projections for price indices, economic activity, exchange rates, Selic rate, among other indicators. Projections are from the market, not from the Brazilian Central Bank[...]" (Banco Central do Brasil, 2020). It is worth of mentioning that Selic, an acronym originating from "Sistema Especial de Liquidação e de Custódia", is the basic interest rate in the Brazilian economy. It is the main monetary policy instrument used by Brazil's Central Bank to control inflation, and it thus influences interest rates on loans, financing and financial investments, etc.

[3] Constant prices are assumed, therefore not incorporating the dynamics when prices adjust. A drop of 6.50% in GDP per capita was considered, with reference to information in the Focus Report of June 19, 2020.

products, and the behavior of system's variables (sensors, geotechnologies, Big Data, etc.). The objective in expanding these digital solutions is to increase the amount of information about origin, safety, food quality, and agricultural production models, as well as their potential impacts in the environmental and social dimensions.

Key driving forces on the supply side

The supply side reflects the quantity of goods and services that producers choose to place on the market in a given period. Nevertheless, agricultural production is not a quick process. Decisions on what to produce, and with which technological packages take months in advance before harvesting grains, such as corn, or oilseeds, such as soybeans. In livestock and fruit production, and in the forestry sector, the delay between production decision-making and final results takes years. Thus, expectations regarding prices (products, inputs), production volume and associated risk variables (e.g., price and productivity) are extremely relevant in the decision-making process.

In the very short term, the producer has limited capacity to change production level and the technological package. In the short term, fluctuations in weather conditions and price volatility resulting from uncertainties on supply and demand, and speculative movements in the market, increase the risks in agricultural production and its market. Periods of great uncertainty, such the new coronavirus pandemic, introduce a greater degree of risk to the agricultural business, determine additional challenges to agricultural policies (producer's income and supply to consumers), and influence decision-making at different levels.

Decisions regarding production and the adopted technological package are adjusted by the farmer as time increases. The longer the time, the greater the possibility of implementing adjustments. In the longest term, all factors may vary. In the long run, generating and incorporating technological innovations into agricultural production systems have been the main and most successful strategy to ensure greater food supply and food security for the Brazilian population, which is mostly urban nowadays, ensuring viable economic conditions for farmers (Martha Júnior; Alves, 2018).

The technological packages developed for Brazilian agriculture, among others, include improved genetics, fertilizers, agrochemicals, cultural practices and conservationist production systems, such as no-tillage and crop-livestock-forest integration. More recently, technological possibilities have expanded and incorporated digital solutions into the universe of viable technologies. With the adoption of these different forms of technology, the goal is to effectively increase productivity gains in the use of resources and inputs. These are strategic aspects for the sustainability and competitiveness of national agriculture.

The role of technology in Brazilian agriculture

The overall style of development in Brazilian agriculture has been predominantly based on productivity gains, reflecting the increasing incorporation of technologies into the production system. Alves et al. (2013) worked with data from the 1995/96 and 2006 Agricultural Censuses and investigated the determinants of income in Brazilian agriculture. Within a decade, technology, drawing on the entrepreneurship of producers, public policies (such as rural credit), and the available stock of knowledge and technologies in tropical agriculture began to explain 68% of the variation in gross income in agriculture. This represented a 33% increase compared to 1995/1996. In the comparison between the Agricultural Censuses, the contribution of land and labor to the variation of income in agriculture was reduced by about 50% (from about 18% to 9%) and 30% (i.e., from approximately 31% to 22 %), respectively.

Therefore, land and labor, as factors of production, lost their relevance in explaining the development of a science-based agriculture, like the Brazilian one. Technology is the major driver behind the sustained development of Brazilian agriculture in recent decades, as well as for future projections. The positive response to investment in agricultural research, with returns higher than 10% (Hurley et al., 2014), occurs over long periods, usually over 20 years, depending on the technology (Alston, 2010; Baldos et al., 2018). Given this long period of maturation of agricultural research, the listed results were only possible due to a persistent and focused work in agricultural research and development (R&D), with a focus on innovation, which has been actively developed in Brazil since the 1970s.

It is anticipated that digital solutions will gain relevance in the coming decades, reinforcing the role of technology as a major driver explaining income in Brazilian agriculture. This happens because of the wide set of databases, technologies, and resources that make up digital agriculture, crosscutting to traditional technological aspects. These advances begin to gradually play a more important role in the management and efficiency of the technological component of national agriculture.

A brief reflection on risks in agriculture

Farmers carry out their activities in a dynamic and uncertain environment. Expectations (and volatility) for the prices of products and inputs guide their decision-making. According to the perception of business risk, farmers can opt for lower risk activities, even if this implies compromising the average income of the farm (Barry et al., 2000; Chavas, 2008; Moss, 2010). The level of risk aversion of individuals can vary over time according to wealth and previous experience, among other factors (Barry et al., 2000).

Variations in weather conditions can compromise the expected production results depending on the intensity, duration, and moment in which they occur in the production cycle. This weather production risk is a random effect, beyond the farmer's control, except in irrigated areas, which in Brazil make up less than 10% of the total cultivated area. Furthermore, Brazilian agriculture operates in soils with low chemical fertility and, throughout the entire production cycle, is pressured by the possible incidence of pests, diseases, and weeds. This makes it dependent on the continued use of modern inputs to remain productive, which nevertheless can accommodate substantial gains in agronomic and economic efficiency.

With unfavorable relative input-product prices, efficiency in the use of these inputs needs to increase to alleviate pressure on farmer's income. Digital technologies can act directly to increase efficiency gains and reduce production costs. The available precision farming instruments allow for improved adjustment in quantity requirements for a range of inputs, such as fertilizers, seeds, agrochemicals, fuels. Advanced models that support decision-making allow, within certain limits, to reduce the negative impacts of some forms of risk. Thus, intelligent warning systems enable adjustments in management, making it more effective in the face of pressures from diseases and pests or the effects of seasonality in forage production in pastoral systems.

Agricultural commodity prices are more volatile, e.g., they fluctuate more over time in relation to price volatility of other non-food goods and services (Tomek; Robinson, 2003). Among agricultural commodities, prices for beef cattle are generally less volatile (e.g., lower price risk for economic agents) compared to the prices of other products, such as soybeans or corn. Wedekin (2017) illustrated this fact by analyzing the price volatility of several agricultural commodities on the stock exchange, from January 2010 to February 2017. The volatility, expressed in percentage per year, for fed cattle, calf, soybeans, corn, coffee, and sugar was 11.0%, 17.0%, 23.1%, 24.6%, 32.6%, and 34.1%, respectively[4].

[4] Volatility is generally estimated based on daily price variation, however it is expressed as a percentage rate per year (Wedekin, 2017).

Within the scope of price risks, it is also necessary to consider the exchange risk in the sale of these products when purchasing inputs. Another important factor is the behavior of relative input-product prices, which are important in determining the level of production. In general, output and input prices are positively correlated (Tomek; Robinson, 2003), but the adjustments to changes in relative output-factor prices can take some time.

Price expectations in Brazilian agriculture gain complexity as a growing portion of national production is directed to exports and, therefore, the international market is a strong component of the price reference. It should be noted, however, that this growth in exports has not compromised the domestic supply and food security of Brazilians (Martha Júnior, 2020).

There are two pertinent additional considerations. First, digital solutions, as inputs to agricultural production, have their adoption subject to the perception of benefits and relative input-product prices. The second consideration is that the biggest opportunities for the group of digital solutions whose value depends on the information they provide for decision-making lies in their use to reduce production and market risks, the so-called business risk. This is due to the increased capacity to observe and capture (sensors, satellites, drones), record, and store data (Big Data, cloud storage), which once transferred and gradually used at greater speed (IoT), through advanced algorithms and models, allow basic and applied solutions (artificial intelligence, analytics) to be applied for the "real world". This increases the efficiency of management and operations and improves the decision-making process.

The choice-process of technology on rural properties

Two groups of technologies can be identified. The first one, the embodied knowledge type of technologies, is represented by technologies that incorporate large amounts of knowledge, and their use do not require much expertise as the technology value is already incorporated into it. Some examples are hybrid seeds, fertilizers and most agrochemicals. This type of technology is the most successful in twentieth-century agriculture (Miller et al., 2018; Lowenberg-Deboer, 2019). In digital agriculture, an example is Global Navigation Satellite Systems (GNSS) guidance (Lowenberg-Deboer, 2019).

The second group is information-intensive technologies. These technologies generate substantial amounts of data and information, which can then be used in the decision-making process. However, to explore the potential offered by these technologies, it is necessary to analyze and interpret data and information, which requires training and some degree of specialization (Miller et al., 2018; Lowenberg-Deboer, 2019). In digital agriculture, an example is variable rate technology of input application, such as fertilizers (Lowenberg-Deboer, 2019).

Regardless of the type, there are steps that should be used in the process of choosing technologies on a farm. Based on Alves' suggestion (2001), and adapting it to the context of digital technologies, the following are the critical steps for evaluating these digital solutions:

a) Description of the digital technology and of the proposal for its insertion in the production system.

b) Analysis of the strengths and weaknesses of digital technology vis-à-vis the technology that will be replaced on the farm. What are the expected gains?

c) Identification of necessary changes in the current production system and farm management to enable the adoption of digital technology. For example, what is the demand for resources and specialized technical assistance for the proper functioning of the digital solution?

d) Identification of restrictions, in farm's context, that may limit the best performance of the digital solution. For example, what are the limitations of IT infrastructure in the region?

e) Analysis of financial demand and financing possibilities in the case of capital acquisition costs. How does the type of contracting the digital solution affect the farmer's decision-making (fixed versus variable cost)? It is also important to consider whether there is a necessary minimum scale, or minimum prices, to make digital technology viable.

f) Evaluation of risks related to digital technology and the training needs of the farmer/manager.

g) Analysis of impacts (positive or negative) perceived by digital technology on the environment.

In cases where the expected social benefits outweigh the private benefits due to the use of technology, opportunities are created to eventually design a public policy to equalize the two (Alves, 2001). When in society's interest, the adoption of these digital technologies can be induced through a set of appropriate incentives.

An illustrative example

The farmer's perspectives on opportunity costs and risks involved in decision making are unique in relation to a given farmer-farm combination. This is because the quantity and quality of resources (land, labor, physical and human capital) and available inputs, which are subject to relevant relative prices, vary from case to case. Brazilian agriculture is extremely exposed to market signals[5]. Therefore, evaluating the economic performance of the activity is a critical step in the decision-making process.

The cornerstone of costing is the opportunity cost. This represents the value of the best alternative sacrificed for another economic alternative to be realized, given the restrictions on production. Expenses (explicit) and remuneration (implicit) of capital (own or third-party) must be computed. Expenses denote cash disbursements, while rents and interest paid from the use of capital represent how much the firm is sacrificing, in monetary terms, in order to carry out one activity against another alternative.

Table 1 presents an exercise related to soybeans in the 2019/2020 crop. Indicator descriptors are included in the table's footnote. The net return (NR) in the baseline scenario was USD 51.38[6]. This means that, in the situation represented in Table 1, it was possible to fully remunerate all production factors, with a positive balance. The main digital technologies have sought to increase efficiency gains in mechanized operations and in the use of inputs, notably fertilizers and agrochemicals. These three expense items – operations with machinery, fertilizers, and agrochemicals – represented 47.2% of the total cost (TC) in the exercise shown in Table 1.

Some of the potential benefits of digital solutions that can be adopted in the production system already have their cost represented in these expenses, as they refer to embodied-type technologies, which do not require additional remuneration. Alternatively, consider the feasibility of expanding such benefits through information-intensive technologies is being studied. These would allow, for example, to better use productivity and soil fertility maps, and identify crop growth deviations in support to

[5] According to statistics from the Organization for Economic Co-operation and Development, the 1995–2018 level of incentives to Brazilian agriculture ("producer support estimate", PSE) was on average only 1.6% of gross farm receipts. Considering the same period, the average levels of incentives received by farmers in the United States, China and the European Union were 13.0%, 8.7% and 26.9%, respectively (Organization for Economic Co-operation and Development, 2019).

[6] Average dollar exchange rate for 2020: USD 1.00 = BRL 5.1575

decision-making. There is no rule for remunerating these services, but it is estimated that depending on the particularities of each farm, this cost can currently vary between USD 0.97 and USD 7.76 per hectare.

A quick look at the information in Table 1 can reveal that a digital solution costing, for example, USD 7.76/ha would easily be paid for the NR of USD 51.38. However, the producer already has this NR without using the technology under evaluation. Therefore, it is important to look at the difference between the expected economic gain from the use of the digital solution under evaluation and the one already earned, translated by the NR in the baseline scenario. The differences would be, therefore, of USD 7.39, USD 14.77, and USD 44.32 for efficiency gains from the adoption of a digital solution of 2.5%, 5.0%, and 15.0%, respectively (Table 1).

Table 1. Illustration of the potential of digital technologies in reducing expenses and increasing net return in soybean production.

Expense reduction from adoption of digital solutions (%)	Net return (baseline)	Productivity shock	Productivity shock+price shock
	Net return (USD/ha)		
0.0	51.38	0.60	-30.72
2.5	58.77	7.99	-23.33
5 .0	66.16	15.37	-15.94
15.0	95.70	44.92	13.60
Net return, SEM[(1)]	35.58 to 67.19	-15.21 to 16.41	-46.52 to -14.91

Note: Considers 18 Conab panels for the 2019/2020 harvest. Net return (NR) is given by NR = gross return (GR) – TC. The reduction in expenses refers to operations with machinery, fertilizers, agrochemicals. SEM refers to the standard error of the mean, it is a measure of dispersion. For the baseline scenario, there was an average productivity of 53.4 bags (60 kg) per hectare. The average price at the time the decision was taken (2019) was USD 12.67/bag. In the baseline scenario, gross return (GR) is USD 677.11; variable costs (VC) of USD 451.91; total operating cost, TOC (VC + depreciation and other fixed costs) of USD 541.16; total costs (TC; TC= TOC added to capital remuneration and rent expenses) of USD 625.73. The productivity shock considered a 7.5% drop in soybean yield and the price shock a 5% drop in soybean price.

[(1)]SEM = Standard error of the mean.

In this context, two qualifications are possibly relevant. First, in the presence of negative productivity and price shocks, which reduce the returns compared to the baseline scenario, the economic performance of the system deteriorates. However, the differential responses expected by efficiency gains in the use and application of inputs in Table 1 were maintained. To put in another way, if the payment for a digital solution is as a "fixed cost", independent of the productivity level, the decision-making does not change[7]. The optimal decision would remain in place, except when the additional "fixed" expense is of such magnitude as to compromise the positive result of the business in the short term.

For example, with a 7.5% drop in productivity ("productivity shock"), the NR was slightly positive, at USD 0.60. The dispersion of NR, given by the standard error of the mean, ranged from - USD 15.21 to USD 16.41 (Table 1). This condition may indicate that those farmers uncertain about the performance of the digital technology could choose not to use it. If the technology delivers what it promises – in the case illustrated in Table 1 a reduction in expenses with mechanized operations, fertilizers, and agrochemicals

[7] If the payment for a digital solution is variable, depending on the level of productivity, the decision, at the margin, will change. In these situations, each different level of production coinciding with a different pay for the technology, would be optimized at a different level.

varying from 2.5% to 15.0% –, its adoption will improve economic performance, whether by expanding the positive result, or by reducing the negative one, as illustrated in the last column (joint shock of productivity and prices).

A second qualification concerns the relationships between revenue and cost, at their different levels. If the unit cost of production is higher than the product price, the business, as it stands, is not sustainable. In the short run, if the expenses represented by the variable costs were covered, the farmer should remain in the activity if he/she had expectations that, in the future, he/she would be able to fully remunerate the factors of production (NR equal to zero, or even positive in the case of "abnormal profits")[8]. Such expectations could reflect improvement in the price of the product, but which is beyond the farmer's control. In the market close to perfect competition where agriculture operates, the farmer is a price taker. It could, alternatively, reflect a critical and well-done analysis of the system by the decision-maker, in which opportunities under his/her control would be identified in order to improve resources- and inputs-use efficiency, and productivity in a sustained manner. The correct diagnosis, together with the ability to manage and implement appropriate actions, would allow to reduce production costs and/or increase returns and, potentially, obtain a more favorable NR.

The possibilities presented by the new generation of digital solutions allow well-established management tools, but still little used by farmers in their daily decisions to become part of their routine decision-making process, many times in an automatically manner. With the expansion of a range of data to aid farm management and field operations, the situation in which the use of a given resource or input could be maximized can be estimated with more accuracy and precision. And, given the expectations of changes in relative prices or production levels, it would be possible to re-estimate new optimal prices that maximize returns to the farmer.

Digital agriculture in the context of the production chain

Digital solutions act transversally in the agricultural production chain. An interesting example of such applications was presented by U.S. Agency for International Developmen (2019). The applicability of digital solutions ranges from the planning process, advancing to the input industry (pre-gate), to the agricultural production, as widely discussed in the chapters of this book, reaching the manufacturing industry and consumers. Transverse to these links, there is the transport sector and the financial sector, which increasingly benefit from digital solutions. The activities in these different links do not take place in an economic vacuum. Thus, a series of forward and backward effects can be observed in the production chain, which ultimately has the potential to multiply sectorial gains.

Data from CEPEA – ESALQ/USP (Luiz de Queiroz Superior School of Agriculture, 2020) showed that the Brazilian agribusiness GDP in 2019 totaled USD 300.53 billion, equivalent to 21.4% of Brazil's GDP. Of this total, the contribution of the input-industry, agricultural (on-farm) production, transformation-industry, and services sectors was USD 15.34 billion, USD 68.06 billion, USD 90.47 billion and USD 127.24 billion, respectively. The key point made by these numbers is that a strong, competitive, and sustainable agriculture is able to supply the national industry with a flow of quality raw materials at declining real prices, potentially increasing its competitiveness in a sustained manner. However, consolidating the

[8] The "zero" economic cost, considering explicit and implicit expenses, is not the same as the "zero" considered in accounting.

competitiveness and sustainability of a strong science-based agriculture requires modern inputs with high technological content. These are provided by urban activities. Therefore, the country's industrial and service sector has a large market to explore, if they are able to deliver quality products to agriculture at competitive prices (Martha Júnior; Alves, 2018).

Sustained efficiency gains provided by digital solutions can contribute favorably to the expansion of agricultural demand. In a recent work, Takasago et al. (2017) estimated the type II multipliers of production, employment, and income for Brazilian agriculture. The values found were 3.42, 1.84 and 5.55, respectively. Thus, for every USD 1 million increase in the sector's final demand, a total of direct, indirect and induced effects of USD 10.81 million is expected. Additionally, for every USD 1 million shock in the final demand of agriculture, there is stimulus for the creation of 309 total jobs, 170 direct, 36 indirect, and 103 induced. These numbers illustrate the significant potential for the application of knowledge and technologies that improve management and decision-making in farms, with a consequent increase in efficiency gains and reduction in production costs in different production stages[9].

Final considerations

Digital technologies have effective potential to improve the competitiveness and sustainability of Brazilian agricultural production chains in a sustainable way. Observing this potential, however, is not a trivial task.

The first concern regards the phenomenon of sectoral concentration and consolidation. This phenomenon, present in the input segments such as seeds, agrochemicals, fertilizers, and agricultural machinery, is advancing rapidly in the dimension of digital agriculture. Such movements point to greater concentration and consolidation in the agricultural production chain, "on-farm" and in the industry and services segments (Miles, 2019; Mooney, 2019; Klerkx; Rose, 2020).

This context introduces a second aspect, related to barriers to the adoption of digital technologies, and to the policy dimension its capacity to promote a more competitive environment for agricultural production in this coming digital era. One of the main barriers to the broader adoption of modern technologies concerns market imperfections. These change relative prices and the return on investment in technologies, and can lead to a widening of productive inequality. Thus, reducing market imperfections is a necessary condition to expand production in a more inclusive way, and increase the effectiveness of policies focused on implementing technologies in agriculture (Martha Júnior; Alves, 2018).

The adoption of a new technology implies that farmers will consider the proposed technology when they have the perception it is superior (competitive) to other technological alternatives already in use at the farm, especially when the sector operates with low levels of incentives, such as in the Brazilian case. The process of choosing the technology on a farm involves the analysis of several factors, and it is conditioned by the farmer's capacity to assimilate and effectively adopt this set of knowledge and technologies in accordance with the recommendations. In some cases, research and rural extension have difficulty in translating and transferring the existing knowledge and respective recommendations into a language that can be easily absorbed by farmers.

[9] Heath (2018) estimated the potential effects of digital agriculture on the Australian economy. His simulations were based on productivity gains attributed to greater efficiency and to a greater value of production and marketing. It was found that the gains would be in the order of A$ 20.3 billion (about R$ 71.90 billion). About two-thirds of the gains were attributed to automation (labor savings), better understanding of genetics, and management to achieve plant and animal productivity gains with more efficient use of inputs, and improved access to markets.

In this regard, a broad and competent work of pertinent technical-economic analyses and subsequent dissemination of the results is necessary to stimulate their faster implementation. Digital solutions can help reducing some of the forms of market imperfection by allowing: a) faster dissemination of knowledge and recommendations (e.g. contributes to reducing the perception of risk); b) better monitoring of key variables (e.g., makes room for more favorable relative prices and efficiency gains in the management of system's components and their interactions); and potentially, c) greater economic attractiveness of the agricultural business (e.g., driven by traceability and transparency related to product and processes).

A third aspect concerns the behavior of digital solutions vis-à-vis other available innovations to be used in the system. The results observed by evaluating one given technology, in isolation, may not be transferable to situations in which other technologies, digital or not, are simultaneously implemented. The results from the interaction of different knowledge and technologies may be different when introduced simultaneously into the production system.

Furthermore, an increasing number of digital solutions are now becoming available for farmers' decision-making, such as different possibilities in robots/automation, artificial intelligence, and sensors. However, very little is still known about how farmers will make their choices, whether they will prioritize one digital technology or a set of them (Klerkx; Rose, 2020). The choices of inputs and technologies used in the production systems depend on their relative prices and, obviously, on the expected returns from their use. In the short term, substantial variations in factors' relative prices may make the adoption of technologies unfeasible, especially the more capital intensive ones.

The fourth aspect, compounded by the crisis of the new coronavirus, indicates digital technologies are being more eagerly adopted in people's daily lives, as well as in some productive and financial activities (teleworking, electronic commerce, etc.). These changes in the way the economy works can accelerate structural changes, impacting the socioeconomic dimension. Despite some sectors gaining momentum with the anticipation of a more intense digital age, the prolonged high unemployment rates in the strata of less qualified people can be observed. This portion of the population, which has already been harmed by the crisis and by the slow recovery of the economy, may see employment and income expectations further deteriorated as they may lack the skills or familiarity with the minimum tools to participate in this digital age.

With a vision on the future, strengthening the sustainability and competitiveness of Brazilian agriculture requires a solid base in agricultural research, without the slowdown in required spending levels. From a strategic perspective, the intensity of investments in public agricultural research requires equal conditions when facing the main international competitors. Developed countries invest around 3.12% of agricultural GDP in public research (Heisey; Fuglie 2018). Brazil invested approximately 1.8% of its agricultural GDP in research, mostly public, until 2013 (Martha Júnior; Alves, 2018). There was a reduction in this level of investment with the recession between 2015-2017, which added to the weak economic recovery in the following years. Two relevant consequences arise from this reflection: a) the need to increase investments in public agricultural research in the country in order to guarantee the continuity of the virtuous cycle of innovation in the agricultural sector; b) the need to encourage engagement of the private sector, as governmental contribution alone, even if increased, will not be enough to sustain the required levels of investment in research for a competitive agriculture in the coming decades.

Last, but not least, the implementation of successful strategies by the private and public sectors, and the design of public policies with greater impact on the competitiveness and sustainability of Brazilian agricultural production, require adequate assessment tools to provide credible and verifiable analyses of the technical-economic-environmental-social dimensions. Without a robust understanding of these dimensions, strategies, policies, programs and, ultimately, the associated decision-making may prove inappropriate and impractical, with unintended consequences (outcomes). In this context, digital agriculture approaches, which make use of large databases (Big Data), advanced models and modeling techniques (artificial intelligence, analytics), in different areas of knowledge, are of great relevance to support the decision-making process at its different levels.

References

ALSTON, J. M. **The benefits from agricultural research and development, innovation, and productivity growth**. Paris: OECD Publishing, 2010. (OECD food agriculture and fisheries papers, n. 31). DOI: 10.1787/5km91nfsnkwg-en.

ALVES, E. Que fazer antes de difundir a tecnologia? Tema para discussão. **Cadernos de Ciência & Tecnologia**, v. 18, n. 2, p.135-138, maio/ago. 2001. DOI: 10.35977/0104-1096.cct2001.v18.8846.

ALVES, E. R. A.; SOUZA, G. S.; ROCHA, D. P.; MARRA, R. Fatos marcantes da agricultura brasileira. In: ALVES, E. R. A.; SOUZA, G. S.; GOMES, E. G. (ed.). **Contribuições da Embrapa para o desenvolvimento da agricultura no Brasil**. Brasília, DF: Embrapa, 2013. p.13-45.

BALDOS, U. L. C.; VINES, F. G.; HERTEL, T. H.; FUGLIE, K. O. R&D spending, knowledge capital, and agricultural productivity growth: a Bayesian approach. **American Journal of Agricultural Economics**, v. 101, n. 1, p. 291-310, Jan. 2019. DOI: 10.1093/ajae/aay039.

BANCO CENTRAL DO BRASIL. **Focus**: relatório de mercado. Available at: https://www.bcb.gov.br/publicacoes/focus. Accessed on: 23 June 2020.

BARRY, P. J.; ELLINGER, P.N.; HOPKIN, J. A.; BAKER, C. B. **Financial management in agriculture**. 6th ed. Danville: Interstate, 2000. 678 p.

BOUMPHREY, S.; BREHMER, Z. **Megatrend analysis**: putting the consumer at the heart of business. London: Euromonitor International, 2017. 25 p.

CHAVAS, J.-P. On the economics of agricultural production. **The Australian Journal of Agricultural and Resource Economics**, v. 52, n. 4, p. 365-380, Dec. 2008. DOI: 10.1111/j.1467-8489.2008.00442.x.

ELZEN, B.; BARBIER, M.; CERF, M.; GRIN, J. Stimulating transitions towards sustainable farming systems. In: DARNHOFER, I.; GIBBON, D.; DEDIEU, B. (ed.). **Farming systems research into the 21st century**: the new dynamic. Dordrecht: Springer Netherlands, 2012. p. 431-455. DOI: 10.1007/978-94-007-4503-2_19.

EMBRAPA. **Visão 2014-2034**: o futuro do desenvolvimento tecnológico da agricultura brasileira. Brasília, DF, 2014. 194 p.

EMBRAPA. **Visão 2030**: o futuro da agricultura brasileira. Brasília, DF, 2018. 212 p.

ESCOLA SUPERIOR DE AGRICULTURA LUIZ DE QUEIROZ. Centro de Estudos Avançados em Economia Agricola. **PIB do agronegócio brasileiro**. Available at: https://www.cepea. esalq.usp.br/br/pib-do-agronegocio-brasileiro.aspx. Accessed on: 15 May 2020.

EUROPEAN COMMISSION. **European economic forecast**: Autumn 2019. Nov. 2019. (European economy: institutional paper, 115). Available at: https://ec.europa.eu/info/sites/info/files/economy-finance/ip115_en_0.pdf. Accessed on: 7 May 2020a.

EUROPEAN COMMISSION. **European economic forecast**: Spring 2020. May 2020. (European economy: institutional paper, 125). Available at: https://ec.europa.eu/info/sites/info/files/economy-finance/ip125_en.pdf. Accessed on: 7 May 2020b.

FAO. **The future of food and agriculture**: trends and challenges. Rome, 2017. 163 p.

GLOBAL megatrends and impacts on the food value chain. Apr. 2019. Available at: https://www.pwc.co.nz/pdfs/2019pdfs/global-megatrends-and-impacts-on-the-food-value-chain.pdf. Accessed on: 18 Aug. 2019.

GODFRAY, C. J.; AVEYARD, P.; GARNETT, T.; HALL, J. H.; KEY, T. J.; LORIMER, J. Meat consumption, health, and the environment. **Science**, v. 361, n. 6399, eaam5324, 2018. DOI: 10.1126/science.aam5324.

GUILLEMETTE, Y.; TURNER, D. The long view: scenarios for the world economy to 2060. **OECD Economic Policy Paper**, n. 22, July, 2018. DOI: 10.1787/2226583X.

HARARI, Y. N. **Sapiens**: a brief history of humankind. New York: HarperCollins Publishers, 2015. 443 p.

HAZELL, P.; WOOD, S. Drivers of change in global agriculture. **Philosophical Transactions of the Royal Society B**, v. 363, n. 1491, p. 495-515, July 2008. DOI: 10.1098/rstb.2007.2166.

HEATH, R. An analysis of the potential of digital agriculture for the Australian economy. **Farm Policy Journal**, v. 5, p. 9-23, 2018. Available at: http://103.37.8.116/~farminst/product/an-analysis-of -the-potential-of-digital-agriculture-for-the-australian-economy-by-richard-heath. Accessed on: 4 Mar. 2021.

HERTEL, T. W.; BALDOS, U. L. C. **Global change and the challenges of sustainably feeding a growing planet**. Cham: Springer International Publishing, 2016. 184 p. DOI: 10.1007/978-3-319-22662-0.

HURLEY, T. M.; RAO, X.; PARDEY, P. G. Re-examining the reported rates of return to food and agricultural research and development. **American Journal of Agricultural Economics**, v. 96, n. 5, p. 1492-1504, Oct. 2014. DOI: 10.1093/ajae/aau047.

IBGE. **PIB cai 4,1% em 2020 e fecha o ano em R$ 7,4 trilhões**. Available at: https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/30165-pib-cai-4-1-em-2020-e-fecha-o-ano-em-r-7-4-trilhoes. Accessed on: 4 Mar. 2021.

INTERNATIONAL MONETARY FUND. **World Economic Outlook update**, Jan. 2022. Available at: https://www.imf.org/en/Publications/WEO/Issues/2022/01/25/world-economic-outlook-update-january-2022. Accessed on: 21 Feb. 2022.

INTERNATIONAL MONETARY FUND. **World Economic Outlook**, October 2021. Available at: https://www. International Monetary Fund.org/en/Publications/WEO/Issues/2021/10/12/world-economic-outlook-october-2021. Accessed on: 3 Nov. 2021.

KHARAS, H. **The unprecedented expansion of the global middle class**: an update. Global Economy Development Working Paper, n. 100, Feb. 2017. 27 p. (Brookings Institution). Available at: https://think-asia.org/bitstream/handle/11540/7251/global_20170228_global-middle-class.pdf?sequence=1. Accessed on: 1 Jun. 2020.

KIROVA, M.; MONTANARI, F.; FERREIRA, I.; PESCE, M.; ALBUQUERQUE, J. D.; MONTFORT, C.; NEIRYNCK, R.; MORONI, J.; TRAON, D.; PERRIN, M.; ECHARRI, J.; ARCOS PUJADES, A.; LOPEZ MONTESINOS, E.; PELAYO, E. **Megatrends in the agri-food sector: global overview and possible policy response from an EU perspective**. Available at: https://ec.europa.eu/knowledge4policy/publication/megatrends-agri-food-sector-global-overview-possible-policy-response-eu-perspective_en. Accessed on: 1 Jun. 2020.

KLERKX, L.; ROSE, D. Dealing with the game-changing technologies of agriculture 4.0: How do we manage diversity and responsibility in food system transition pathways? **Global Food Security**, v. 24, article 100347, Mar 2020. DOI: 10.1016/j.gfs.2019.100347.

LOWENBERG-DEBOER, JDEBOER, J. L. The economics of precision agriculture. In: STAFFORD, J. (ed.). **Precision agriculture for sustainability**. Cambridge: Burleigh Dodds Science Publishing Limited, 2019. Disponível em: https://shop.bdspublishing.com/store/bds/detail/workgroup/3-190-56391. Accessed on: 28 Nov. 2019.

MARTHA JÚNIOR, G. B. Uma agropecuária forte amortece os impactos negativos da Covid-19. **Revista de Política Agrícola**, v. 29, n. 2, abr./maio/jun. 2020. Available at: https://ainfo.cnptia.embrapa. br/digital/bitstream/item/215297/1/Uma-agropecuaria-forte.pdf. Accessed on: 13 Oct. 2020.

MARTHA JÚNIOR, G. B.; ALVES, E. R. de A. Brazil's agriculture modernization and Embrapa. In: AMANN, E.; AZZONI, C.; BAER, W. (ed.). **The Oxford handbook of the Brazilian economy**. New York: Oxford University Press, 2018. p. 309-337. DOI: 10.1093/oxfordhb/9780190499983.013.15.

MILES, C. The combine will tell the truth: on precision agriculture and algorithmic rationality. **Big Data & Society**, p. 1-12, Jan./June 2019. DOI: 10.1177%2F2053951719849444.

MILLER, N. J.; GRIFFIN, T. W.; CIAMPITTI, I. A.; SHARDA, A. Farm adoption of embodied knowledge and information intensive precision agriculture technology bundles. **Precision Agriculture**, v. 20, p. 348-361, 2018. DOI: 10.1007/s11119-018-9611-4.

MOONEY, P. **Blocking the chain**: industrial food chain concentration, Big Data platforms and food sovereignty solutions. Available at: https://www.etcgroup.org/sites/www.etcgroup.org/files/files/blockingthechain_english_web.pdf. Acesso em: 23 Mar. 2019.

MOSS, C. B. **Risk, uncertainty and the agricultural firm**. New Jersey: World Scientific, 2010. 292 p. DOI: 10.1142/7469.

NELSON, G. Drivers of change in ecosystem condition and services. In: CARPENTER, S. R. (ed.). **Ecosystems and human well-being**: scenarios - findings of the scenarios working group. [Washington, D.C.]: Island Press, 2005. p. 173-222.

TAKASAGO, M.; CUNHA, C. A.; OLIVIER, A. K. G. Relevância da agropecuária brasileira: uma análise insumo-produto. **Revista Espacios**, v. 38, n. 36, p. 31, 2017. Available at: http://www.revistaespacios.com/a17v38n36/a17v38n36p31.pdf. Accessed on: 18 Aug. 2019.

THE ECONOMIST INTELLIGENCE UNIT. **Global food trends to 2030**: with a closer look at the GCC. Available at: https://eiuperspectives.economist.com/strategy-leadership/ global-food-trends-2030-closer-look-gcc. Accessed on: 18 Aug. 2019.

TOMEK, W. G.; ROBINSON, K. L. **Agricultural product prices**. 4th ed. Ithaca: Cornell University Press, 2003. 428 p.

USAID. **Digital tools in USAID agricultural programming toolkit**. [S.l.]: Usaid, 2018. 26 p. Available at: https://www.usaid.gov/digitalag/documents/ how-program-digital-tools-agriculture-programs. Accessed on: 28 Nov. 2019.

WARR, P. **Urbanisation and the demand for food**. Australian National University, Dec. 2019. (Working Papers in Trade and Development, n. 9). Available at: https://acde.crawford.anu.edu.au/sites/default/files/publication/acde_crawford_anu_edu_au/2019-12/adec_td_wp_2019_10_warr.pdf. Accessed on: 28 Nov. 2019.

WEDEKIN, I. **Economia da pecuária de corte**: fundamentos e o ciclo de preços. São Paulo: Wedekin Consultores, 2017. 180 p.

Photo: Digilife (AdobeStock)

16

Challenges, trends and opportunities in digital agriculture in Brazil

Édson Luis Bolfe | Jayme Garcia Arnal Barbedo | Silvia Maria Fonseca Silveira Massruhá | Kleber Xavier Sampaio de Souza | Eduardo Delgado Assad

Introduction

The combination of soil conditions, climate, relief, science, technology, public policies and the agricultural entrepreneurship made Brazil one of the world leaders in agricultural production and export. Recent forecasts by the Ministry of Agriculture, Livestock and Supply (Brazil, 2019b) indicate that grain production could surpass the current level of 250 million tons, reaching between 300 and 350 million tons in the 2028/2029 crop-year. As for meat production (chicken, pork and beef), projections indicate going from the current 26 million to 33 million tons by the end of the next decade. There is also a growing demand for cotton, cellulose, milk, sugar and fruits, especially mango, grape and apple. The domestic market and international demand are indicated as the main growth factors for most of these products.

The growth of this production should continue based on productivity. Total factor productivity (TFP) has grown on average 3.50% per year over the past few years, and is forecasted to grow at 2.92% per year for the next decade (Gasques et al., 2016). Embrapa (2018) also highlights the importance of Brazilian agricultural intensification in the coming years, with emphasis on multiple crops per year in the same area, recovery of degraded pastures, precision irrigation and more sustainable use of inputs and natural resources. In turn, population growth, continued urbanization, longer life expectancy, changes in dietary patterns and economic power are the driving factors of greater global demand for food, energy and water.

Digital technologies can help solve this complex equation with countless economic, social and environmental variables, which require producing more food, with quality and with less use of natural resources. Digital agriculture, also called “4.0”, comprises technologies, which are already operational or under development, such as robotics, nanotechnology, synthetic protein, cellular agriculture, gene editing technology, artificial intelligence, blockchain and machine learning. These technologies can have widespread transformative effects for future development of agriculture and agrifood systems (Klerkxa; Roseb, 2020).

Bolfe and Massruhá (2020) point out that the process of digital transformation in rural properties is no longer an option, it is an essential path to make Brazilian agriculture more competitive and with greater added value. This transformation is understood as interdisciplinary and transversal, not limited to regions, crops or social class. Its potential benefits amplify innovations and interaction between links in agricultural production chains, promoting new approaches and applications for input manufacturers, rural producers, processors, distributors and consumers (Figure 1).

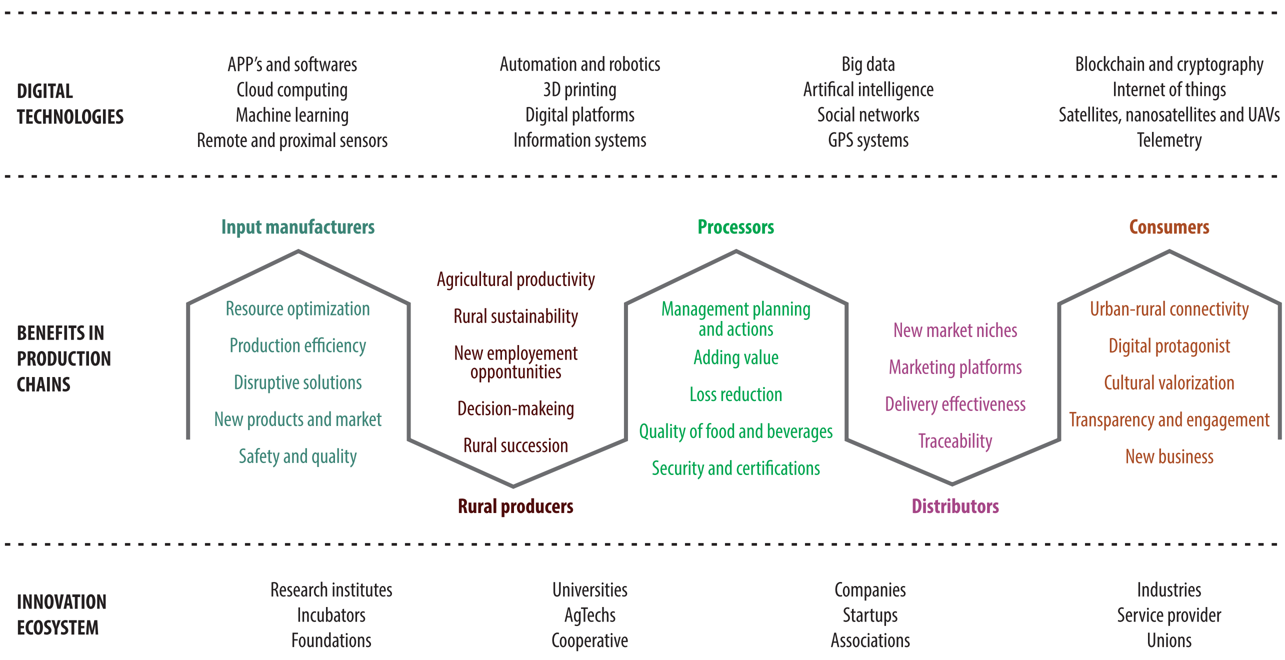


Figure 1. Potential benefits of digital transformation in agricultural production chains.

Source: Bolfe and Massruhá (2020).

This digital transformation environment also shapes development agendas at various scales. Internationally, it can be associated with the 2030 Agenda, which includes 17 Sustainable Development Goals (SDGs) (United Nations, 2015). In this context, the digital transformation in agriculture can also significantly contribute to achieving these goals, particularly in reducing hunger, health and well-being, decent employment and economic growth, reducing inequalities, responsible consumption and production, combating climate change, life on earth, peace, justice and strong institutions.

Estimates by the UNGC (United Nations Global Compact, 2017) indicate that the world market for digital agriculture, in 2021, will be worth 15 billion dollars, and that 80% of companies expect to have competitive advantages in this sector. However, recent international aspects involving trade and health issues between the United States and China, as regards the Covid-19 pandemic (United Nations, 2020), generate an environment of uncertainty to some degree, but with greater expectations of potential growth in the use of digital technologies in agriculture as of 2020.

In this context, this chapter lists some of the main scientific, technological, social and economic challenges and, subsequently, points out trends and opportunities for the future of Brazilian agriculture.

Scientific and technological challenges

Digital online services

The offer of online digital services to rural producers gained momentum in the early 2010s and has expanded since then. With the widespread use of smartphones, most of those services migrated to this platform (Duncombe, 2016). Many technology startups applied to agriculture (AgTechs) that have emerged in recent years have wagered on this type of technology. As it is still relatively new, how to best offer this type of product is still being defined, but this technology is already a reality, and most of the technologies mentioned in this section are or will be incorporated into portfolios of already existing digital services or under development. The challenges that still need to be overcome concern aspects that are not necessarily technological, such as ownership of the data generated by this type of tool, lack of synchronization between the needs of producers and the information generated by the tools, and data security (Rotz et al., 2019). Regardless of what solutions will be implemented for these problems, new digital service platforms will continue to be developed, many of them based on the technologies discussed as follows.

Management and monitoring of plant production

There are several factors that need to be constantly monitored in agricultural management, for instance production, productivity, presence of diseases, pests, weeds, nutritional deficiencies, water stress, among others. One of the main challenges regards monitoring stress, which can be divided into three stages: stress detection, determining the cause of the stress and solving the problem. Despite recent innovations in artificial intelligence, the process as a whole is still mostly manual. However, the degree of automation has increased, and several companies (startups in particular) already offer services in this regard (Wolfert et al., 2017).

A large number of mathematical models have been developed to process different variables and provide indications regarding crop susceptibility to stress-inducing events. For example, data on precipitation, moisture and leaf wetness can be used to calculate the probability of incidence of certain diseases. These models have been improved and fed with increasingly higher quality data, making them a fundamental part of the integrated management of rural properties. However, for effective stress management, it requires detecting them directly in the field. The challenge is to achieve this detection early enough to avoid significant damage. Although there are proximal methods for stress detection, the trend is to make use of these remotely obtained images much more. In the short and medium term, drones will likely dominate this activity due to the high temporal resolution of these images, enabling to detect problems even in individual leaves (Barbedo, 2019a). As more sophisticated sensors are embedded in satellites, they will likely gain attention, mainly due to their wide-ranging coverage.

Conventional cameras (RGB) have limited ability to detect early stresses, as they cannot go beyond human visual capacity. Therefore, it is vital to have sensors that can capture other spectrum bands besides the visible one, such as multispectral and hyperspectral cameras. Multispectral cameras, which typically include three to five spectral bands, are increasingly being used. However, the falling costs and miniaturization of hyperspectral cameras, which separately capture hundreds of spectral bands, will make them an especially attractive alternative in the near future. (Thomas et al., 2018).

Once stress is detected, it is necessary to determine its cause in order to take appropriate actions. Under certain conditions, current sensors can provide sufficient information so that models based on artificial intelligence can provide a reliable classification for the problem being observed (Barbedo, 2018, 2019b), but in most cases it requires that a specialist should do this identification, or it should be done through laboratory analyses. The problem is that different types of stress often produce similar visual signals (Barbedo, 2019b). The spectral profiles produced by different agents tend to differ to a greater degree, but even using sensitive hyperspectral sensors the confusion index is high (Thomas et al., 2018). In the future, the trend is to combine imaging with other sources of information (meteorological variables, management history of the property, soil characteristics, etc.) to increase the degree of automation of the process, although the complete elimination of manual activities is unlikely in the near future. In addition to identifying stress, in many cases it is also important to determine the severity of symptoms so as to deal with the problem. Although there are several algorithms for this purpose, many of the difficulties mentioned are also applicable in this case (Bock et al., 2020).

After locating and identifying the stress, it is necessary to act to eliminate the problem. In many cases, it is necessary to apply products such as pesticides and nutrients. Autonomous vehicles have been developed by several research groups so that this activity can be carried out not only without the need for permanent human supervision, but also at the location and in the necessary quantity, reducing costs and environmental impacts (Reina, 2016). In the near future, it will be possible to have one or more of these vehicles monitoring and operating within the property. In parallel, actuators can also be installed on agricultural machinery to carry out these same activities.

It is noteworthy that artificial intelligence algorithms have been used in other applications, such as crop prediction, location of fault lines, determination of production quality, determination of the degree of ripeness of fruits/grains, among others (Liakos et al., 2019). The trend that is being observed is that tools based on artificial intelligence and machine learning will continue to gain space and will be part of the routine of most properties in the near future.

Management and monitoring of animal production

Appropriate management of livestock farms has evolved considerably in recent years, especially in the case of dairy and beef cattle in the intensive system. However, farm management implementing the extensive production system still faces significant challenges (Barbedo; Koenigkan, 2018). More effective control of the variables involved in property management is essential to maximize profits and reduce the number and severity of problems. Two alternatives have been used, albeit in a limited way, for monitoring large animal husbandry properties: sensors attached to the animals and drones for remote monitoring (Barbedo; Koenigkan, 2018).

Sensors can be attached to animals through ear tags or collar tags, which can collect different information about these animals, including location, temperature and patterns of movement and chewing (Rahman et al., 2018). This information allows to detect potential problems, such as diseases, and infer various aspects of animal behavior, which is important, for example, to create effective mechanisms to accelerate fattening and define the optimal slaughter point (Miller et al., 2019). For the data to be collected at the required frequency, there needs to be effective communication between the individual sensors and the data processing center. Receivers can be installed on poles distributed throughout the property, or data can be collected by drones flying over the animals (Barbedo et al., 2019). Both options have limitations in the case of large properties: in the case of fixed receivers, it would require installing numerous receivers, which represents a high initial cost, in addition to the difficulty of maintaining equipment located in more distant regions. As for drones, in addition to their limited autonomy, the planning of flights must be careful so as to include all animals. There are already commercial solutions offering monitoring through individual animal sensors, as shown in simple internet searches. However, such solutions are not suitable for all types of properties, and the costs are still high. It is noteworthy that technology costs tend to fall as their use increases.

The use of drones for imaging animals is more recent, and the practical use of this type of technology still depends on more research efforts and on the development of new algorithms. There are several efforts in this direction (Barbedo et al., 2019, 2020), since, once made available, this type of technology has several advantages: it does not need specific infrastructure, it is a comparatively cheaper option, several types of sensors can be embedded in drones (RGB, thermal, multispectral, hyperspectral cameras), in addition to the potential of providing other types of information in addition to that which can be obtained with ear tags. There are ongoing studies that, based on animal measurements obtained using the captured images, are focused on estimating the weight of each animal without having to use a scale. Other information that may be obtained in the future using drones include the number of animals in a given area, and the detection of anomalous events such as disease and calf births. However, there are still some limitations that need to be overcome, such as the relatively short autonomy of current UAVs and the difficulty of identifying individual animals when they are grouped together (Barbedo et al., 2020). In the case of autonomy limitation, possible future solutions include the use of images captured at an angle to cover larger areas (Barbedo et al., 2020) and the development of new drones with greater autonomy, such as the "balloon drone". As new solutions emerge, the use of drones in livestock is expected to grow substantially in the near future.

It is also important to mention the use of satellites. Although the spatial resolution of images captured by satellite is not yet sufficient to allow its effective use in monitoring herds, advances in imaging technology and the multiplication of micro and nano-satellite constellations with specific purposes tend to make the use of this type of equipment viable in the future. This does not mean that this technology will replace the others, but it will be an additional alternative that will certainly be beneficial under certain conditions.

Databases in agriculture

The evolution observed over the last two decades in relation to machine learning techniques has made most detection, recognition and classification problems treatable, potentially leading to developing tools of great practical use. However, for such tools to be reliable and robust, the database used to produce the models should be representative of all the variability found in practice (Barbedo, 2018). In most cases, this involves collecting a large number of samples, such as images, measurements or analyses. In the case of images, for example, there are situations that require collecting hundreds of thousands of samples (Barbedo, 2018). The challenge is even greater when the samples need to be properly annotated, that is, information about what is represented in that sample, where the sample was collected, and additional information needs to be correctly generated so as to correctly infer the model. As a result, it is often impossible for a single research group to be able to build a truly representative database (Barbedo, 2019b).

There are two alternatives that have been applied in some circumstances and that will likely prevail in the future. The first is citizen science (Irwin, 2002). This approach, which is mentioned in the 2030 Agriculture Vision document (Embrapa, 2018a), makes use of non-professional volunteers to collect data as part of scientific research, particularly in ecology and environmental science (Silvertown, 2009). In the case of plant disease detection, for example, producers and rural workers could collect images of symptoms in the field and, after they are sent to a server, these images could be labeled by phytopathologists. As mobile devices with imaging capabilities become ubiquitous, the challenge will be to find mechanisms to promote volunteer participation (Barbedo, 2019b).

The second alternative that could prevail in the future is that of sharing the generated databases (Barbedo, 2018). Most technological and scientific challenges are addressed simultaneously by several research groups, each generating its own dataset. If such databases were made available and integrated, the resulting set would likely be much more representative and applicable to real world conditions. Embrapa has been contributing to this type of efforts through the availability of databases such as Digipathos (Embrapa, 2019), one of the first to be part of the Network of Scientific Data Repositories of the State of São Paulo (Fundação de Amparo à Pesquisa do Estado de São Paulo, 2019). A complementary step for adding value to databases is adherence to the findable, accessible, interoperable and reusable (FAIR) principles, which dictate the findability, accessibility, interoperability and reusability standards (Wilkinson et al., 2016).

Socioeconomic challenges

Connectivity in the field

Brazil is among the top ten world markets for mobile telecommunications and fixed broadband data (Agência Nacional de Telecomunicações, 2020). The 2017 agricultural census indicated that internet access grew by 1,900% compared to 2006, which is accessed by around 30% of rural producers (1.43 million in 2017), 659,000 via broadband and 909,000 via mobile internet (IBGE, 2017). Despite representing a relatively high increase, these data indicate that approximately 3.5 million rural establishments – that is, 70% – did not have access to the internet. A study with 750 Brazilian farmers indicates that 47% use at least one tool in precision agriculture, while 33% use two or more, and the young profile of Brazilian rural producers less than 45 years old for some regions and production systems, is one of the reasons for this receptivity to new technologies (Mckinsey Consultoria, 2020).

Even with recent investments by the public and private sectors, the lack of connectivity in rural areas is still one of the main challenges for the insertion of agriculture in the digital transformation process. Territorial dimensions, the low demographic density of a large part of the rural area and socioeconomic inequalities are some of the main obstacles to increasing the availability of internet access in the country. The National Bank for Economic and Social Development (BNDES) (Banco Nacional de Desenvolvimento Econômico e Social, 2017) estimates that greater connectivity in agriculture through the internet of things (IoT) could generate between 50 and 200 billion dollars of annual economic impact in 2025. It is also highlighted that standardization and interoperability of the components of IoT solutions should be sought in order to achieve a greater scale of adoption, faster development of new services and applications, thus fostering the capacity for innovation.

Connectivity is essential to improve technical assistance, distance education, access to market information, the use of management software and applications, and the integration of agricultural machinery and equipment, reducing production costs and improving farm productivity. The Research Foundation of the State of São Paulo (2020) emphasizes that connection infrastructure and data interoperability are the major obstacles to the inclusion of Brazilian agriculture in the 4.0 era, which should help the producer overcome the challenge of expanding the supply of food at affordable prices and in a sustainable way.

Private initiatives to increase internet access in rural properties via satellite, antenna network and bluetooth technologies are expanding in Brazil. An example is ConectarAgro (2020), which attempts to encourage and promote solutions for connectivity in rural areas through tower, radio and antenna technologies.

However, small and medium producers have greater difficulties due to implementation costs. Possible public resources to improve the internet infrastructure may originate from the proposed bill No. 172/2020 (Brazil, 2020a), which is being processed in the National Congress, aimed at modifying the General Telecommunications Law for access to the Fund for Universalization of Telecommunications Services (FUST). The proposed bill provides for the financing of infrastructure expansion in rural or urban regions with a low Human Development Index (HDI), encouraging the use and development of new connectivity technologies to promote economic and social development. Greater connectivity in rural areas is also highlighted in the National Internet of Things Plan (Brazil, 2019a) and in the discussions of the Agro 4.0 Chamber, which has a work group for the issue of Connectivity in the Field (Brazil, 2019a, 2020b).

Costs of digital technologies

Data from the IFAD (International Fund for Agricultural Development, 2020) indicate that around 63% of the poorest people in the world work in agriculture, with the vast majority in small rural properties. In Brazil, according to the Agricultural Census (IBGE, 2017), of the 5 million rural establishments, 4.5 million have an agricultural area of less than 100 ha, that is, they represent around 90% of rural producers. A survey carried out by Embrapa, Sebrae and Inpe (Bolfe et al., 2020) with 753 rural producers, companies and service providers in digital agriculture from all Brazilian regions detected that 67% of these farmers and 58% of service providers indicate that the main challenge for implementing and keeping digital transformation in the property is still investing in machines, equipment and/or applications. Thus, for a significant portion of rural producers, especially small and medium-sized ones, the digital transformation process is still perceived as difficult in view of the current perception of potential economic benefits.

On the other hand, a study estimated a relevant potential economic impact regarding the use of the main technologies in precision agriculture in Brazil for sugarcane, corn and soy products. It was observed that a scenario of 10% increase in the productivity of these crops, with or without a reduction or increase in fertilizers, could increase the GDP of the country's economy by around R$ 11 billion and generate more than 450 thousand jobs (Costa; Guilhoto, 2013). DeBoer (2019) analyzed precision agriculture for sustainability and emphasized that its applications increase the ability to identify the spatial variability within the field, using this information for a more targeted crop management, operating resources more efficiently, making agriculture more productive, sustainable, while reducing its environmental impact.

Another important trend facing the challenge of costs in digital agriculture is the free availability of public and private instruments and training. Some examples are the platforms and applications available that support the management of property and agricultural production, such as: WebAgritec, ZARC Plantio Certo, SatVeg, Agritempo, WebAmbiente, Roda da Reprodução, BioInsumos e AGro (Embrapa Informática Agropecuária, 2019a, 2019b, 2020a, 2020b); AFSoft, Siscob, Qualisolo (Embrapa Instrumentação, 2020); MapOrgânico, Geoweb Matopiba, GeoInfo (Embrapa Territorial, 2020); RenovaCalc, AgroTag e Aquisys (Embrapa Meio Ambiente, 2020).

Rural family succession

The 2017 Census of Agriculture indicated that of the total of five million agricultural producers, 15% declared that they had never attended school, 14% had literacy level, and 43%, had at most elementary education. Thus, 73% of all producers have, at most, elementary education as their education level – complete or partial. It is worth mentioning that 1.1 million producers (23%) declared not knowing how to read and write (IBGE, 2017).

Embrapa (2018), in its study on the future of Brazilian agriculture, points out that 91% of the population will be concentrated in urban areas in 2030. It also emphasizes that the development of technologies suited to different socioeconomic and environmental conditions is not enough to raise the Brazilian agricultural productivity and family income, since producers have a low level of education and lack access to technical assistance and rural extension, thus, incorporating technologies is difficult or even impossible. The study also highlights that the issues associated with income and the demographic depletion of the countryside are altering an important structural element of national agriculture, the hereditary succession in the command/management of properties.

Ownership and management are understood as two major dimensions in the family succession process, in which the digital technology factor is pointed out as one of the opportunities for future succession processes in Brazil (PWC Brasil, 2019). Thus, it is important to have actions to facilitate the succession decision-making process, especially for small and medium rural producers, regarding what should be produced and how this production will be carried out, generating information and assisting in the management of production adjusted to the reality of rural establishments.

Sustainable rural development

The great challenge of world agriculture is to raise its level of economic, social and environmental sustainability. Among the Brazilian goals proposed in the Sustainable Development Goals Agenda for 2030, "ending hunger, achieving food security and improving nutrition and promoting sustainable agriculture" is highlighted (Ipea, 2018). Specifically in agriculture, some of the challenges are to "ensure sustainable food production systems, through research policies, technical assistance and rural extension, among others, in order to implement resilient agricultural practices that increase production and productivity, while helping to protect, recover and conserve ecosystem services, strengthening the capacity to adapt to climate change, extreme weather conditions, droughts, floods and other disasters, progressively improving the quality of land, soil, water and air" and "increase agricultural productivity and income of small food producers, particularly women, family farmers, traditional peoples and communities, targeting both the production of self-consumption and guaranteeing the social reproduction of these populations, as well as their socioeconomic development."

According to the Ministry of Agriculture, Livestock and Supply (Brasil, 2019a), Brazilian agriculture has increased its total production in recent decades, and based on productivity, this growth should continue until 2030. One of the challenges is the need for greater integration of geotechnologies with new remote sensing image processing and fusion algorithms to raise the level and accuracy of real-time use of natural resources, further enhancing agricultural sustainability (Bolfe, 2019).

Among the innovation challenges, Embrapa (2019) points out the need for Brazilian agriculture to increase: a) the efficiency of water use in irrigated agricultural systems for grain, vegetables, fruit, pasture and sugarcane; b) the adaptive capacity and resilience of agricultural production systems with greater projected economic impact and relevance for food security based on climate change scenarios; c) the guidance of land use and occupation in land conversion and expansion areas of the agricultural frontier in the Cerrado, Caatinga and Amazon biomes. For Brazil to definitively assume the leading role in global sustainable agricultural production, it will need greater public and private investments in science, innovation, entrepreneurship, connectivity infrastructure, communication and professional training in digital agriculture (Bolfe; Massruhá, 2020).

Trends and opportunities

Disruptive digital technologies

The quantity and quality of new technologies available for use in agriculture have not only continuously increased over the last few decades but also intensified. Examples of disruptive technologies that are increasingly being used in agriculture for various purposes include nanosatellites (Houborg; Mccabe, 2016), remote and proximal sensors (Mahlein, 2016; Adão et al., 2017), artificial intelligence algorithms (Liakos et al., 2018), drones (Barbedo; Koenigkan, 2018), Big Data techniques (Wolfert et al., 2018), internet of things (Tzounist et al., 2017), cloud computing (Roopaei et al., 2017), blockchain and cryptography (Lin et al., 2017), genomic editing (Chen et al., 2019), 3D printing, robotics (Bechar; Vigneault, 2016), augmented reality (Huuskonen; Oksanen, 2018), and other technologies. Most of these technologies have been discussed in detail throughout this book and are part of the research portfolio being carried out in the context of Embrapa.

Although the offer of such technologies is evidently positive, their usefulness can only be maximized through mechanisms and systems that aggregate the vast amount of data generated by these technologies. More importantly, such tools must be able to generate information that can be immediately used in decision-making. These aggregating technologies will play an increasingly fundamental role in all productive sectors, which is demonstrated by the investments that have been made towards this end (Rose et al., 2016).

Increasing the impact of this type of integrated systems faces major challenges. In particular, the appropriate integration of data from different sources will still require substantial research efforts. Significant advances have been achieved in some areas: feedback between genotyping and phenotyping has been successfully applied in many genetic improvement efforts, and meteorological data have been integrated with information obtained through images to determine the phytosanitary status of crops (Mahlein, 2016). However, it is likely that there is a high degree of complementarity between different types of data that have not yet been explored, causing many technologies not reaching their full potential. Another important challenge regards creating mechanisms for integrating systems to deal with the heterogeneity of potential users. In addition to the level of education that varies considerably, it is important to take into account the different type and level of information each user expects to receive. While most users want to receive fully processed data in the form of information directly related to decision making, there are those who want a more detailed report of what happens on the property. Thus, greater flexibility in data visualization through a user-friendly interface is also an important objective in the near future.

It is important to note that emerging technologies may cause major changes that cannot yet be predicted, such as quantum computing, which can speed up calculation in systems that involve massive calculations, such as simulating scenarios about climate impacts in different areas, price volatility and market fluctuations (Preskill, 2018; Woerner; Egger, 2019), and swarm robotics, in which a large number of robots act in a coordinated manner to collect data (Bayindir, 2016). These are technologies that can significantly change the current scenario, giving way to new possibilities that are not yet viable at the current stage.

Training in digital agriculture

Numerous public and private initiatives seek to increase training in digital agriculture for rural producers, which also favor the process of young people remaining in the countryside. An innovative example is

SENAR (2019), which offers courses in precision agriculture, providing information on the state of the art in agricultural techniques for rural management, promoting rationality and efficiency in production. Another initiative comes from SEBRAE (2019), which provides support to small producers through technological service providers, such as in digital agriculture.

In addition to the availability of online platforms, applications and courses from various research centers, the actions of the Precision Agriculture Research Network (Embrapa, 2018b) are highlighted. These actions generate scientific knowledge, provide technical publications and train multipliers/extensionists on the variability of production and of soil and environmental parameters, of plants, pests and crops diseases such as soybean, corn, cotton, wheat, eucalyptus, sugarcane, orange, grape, apple and peach. At the Brazilian higher education level, BNDES (National Bank for Economic and Social Development, 2017) points out the need to incorporate new disciplines related to IoT and precision agriculture in the courses offered in rural areas, as well as expand the offer of extension courses and postgraduate courses to train technology specialists with agricultural knowledge.

Thus, there are opportunities to provide greater dynamism and integration between research, teaching, industry, commerce, technical assistance and rural extension; take advantage of the more connected rural world and improve the distance education process in the countryside. Digital training can attract more young people to generate more interdisciplinary solutions in the day-to-day life of rural properties, increasing productivity with less pressure on natural resources. An innovative, entrepreneurial and multiplier profile is essential for all who seek digital transformation in agriculture.

Consumer market in the digital age

The higher level of consumer information, made possible by social networks, allows raising awareness about the quality and origin of food and the socio-environmental responsibility of agricultural production systems. The various information and communication technologies benefit the rural-urban relationship by better understanding the role of each sector, enabling to value regional culture and local products, help valuing and maintaining biodiversity, and support rural tourism. Conventional businesses will be developed from the perspective of the digital market, in which the relationship between consumers and customers will be strengthened through business ecosystems, the intensive use of automation and the convergence of ICT in agriculture (Embrapa, 2018b).

The digital economy with cryptocurrencies also boosts virtual cooperatives, new businesses, and digital platforms with direct producer to consumer integration. Bolfe and Massruhá (2020) emphasize that in this technological revolution, people are the main protagonists who will increasingly have a decisive role in decision-making, because, through digital technologies, people will be more demanding and will require more information about the products consumed.

According to a study conducted by CEPEA (Luiz de Queiroz Superior School of Agriculture, 2020), the current covid-19 pandemic can potentially change society's habits even more, by increasing awareness and efforts to meet hygiene and health levels known to science, but not yet prioritized. It is highlighted that different countries must adopt more robust health protocols, and in addition, they need to raise the global discussion on the consistency of disease surveillance and control systems, which affect animals and humans, to ensure food supply and safety.

In this scenario, only the producer that incorporates new digital technologies will be able to provide more transparency in their production process and will respond to the demands of the national and international market. Thus, there are great opportunities for the development of technologies aligned with digital transformation, which generate information about origin, quality, production methods, environmental and social impacts of agricultural production, among others, such as animal well-being and adequate use of agricultural inputs.

Digital platforms

The growing digital transformation of agriculture drives the demand for solutions that integrate property, production and marketing management information, which are available to rural producers via computers or smartphones. In this context, research institutions, universities, large companies, startups, cooperatives and associations have invested in the development of digital platforms, providing innovative solutions with the integration and analysis of data via geostatistics, artificial intelligence, cloud processing and computer vision.

SigmaABC is an example of a platform that integrates user information (farms, fields, machines and implements, production costs) with data collected in the field, geophysical surveys, phytotechnical data, automatic meteorological stations, global models and regional weather forecasting, mathematical models (diseases, pests, weeds, soil water, potential yield) and remote sensing models (vegetation indices) at different spatial and temporal scales (Fundação ABC, 2020).

Another digital platform format in agriculture is AgroAPI, which offers agricultural information and generated models (Embrapa, 2019). (Embrapa, 2019). It provides opportunities to generate new products and businesses for companies, startups, public and private institutions to create software, web systems and mobile applications for the agricultural sector, with reduced costs and time. Access to information and models is performed virtually, through Application Programming Interface (APIs). These applications include a set of standards and programming languages that allow, in an automated way, agile and secure communication between different systems.

The platforms for the sale of beverages and foods are also consolidated realities in Brazil, and they serve countless consumer profiles. With the current pandemic associated with covid-19, e-commerce giants are leveraging their capabilities in logistics, supplies and technology to also supply urban centers, especially in Asia. RaboResearch (2020) highlights that these companies can further solidify their power of influence with consumers, linking rural producers and processors to distributors and retailers, effectively organizing agricultural production, processing, management of inventories and distribution channels.

Opportunities are also highlighted for the coming years concerning the development of integrated digital platforms, on topics such as: a) support for data analysis and decision-making on the property, with geospatial information on agriculture, vegetation, soil and water resources to support Environmental Regularization Programs (PRA), Environmental Reserve Quotas (CRA) and Payments for Environmental Services (PSA); b) connectivity between rural producers and consumers, supporting the traceability process and certification of the quality and origin of products such as milk, honey, eggs, meat, grains, fruits, sugar, biofuels, fibers, wood and cellulose; and c) support for decision-making and the management of agricultural public policies, based on mathematical, statistical and computational models, with the use of artificial intelligence, computer vision and remote sensing image processing (Embrapa, 2018a, 2018b, 2019).

Future risk projection systems

Climate change has always been one of the main factors in determining the risks to agricultural activities. Using the tools available today, it is necessary to compile, systematize and update information on the possible impacts related to the rising temperatures in Brazilian agriculture related to climate change. For the purpose of short-term planning, all the information currently made available by Embrapa for supporting property management and agricultural production, such as WebAgritec, ZARC Plantio Certo, SatVeg and Agritempo, are sufficient for decision-making.

However, for medium and long-term projections and analysis of future risks in agriculture, one of the challenges is incorporating regional climate models, which allow evaluating the future behavior of crops in terms of climate risk and productivity. Embrapa, on an experimental basis, is developing a new system called the Agricultural Scenario Simulator (ScenAgri) (Embrapa, 2020), which incorporates the aforementioned aspects and combines the foundation of Agricultural Zoning of Climate Risk (ZARC). The system is based on high-performance computing to support researchers investigating the impacts of climate change on Brazilian agriculture. Some studies have already shown the importance of this future projection in the medium and long term for diseases of plants, forages, eucalyptus, grains and sugarcane (Ghini et al., 2011a, 2011b; Marin; Nassif, 2013; Assad et al., 2016).

In the near future, rural producers may have, in their applications, systems that show the crop vulnerability to climate in the short (Plantio Certo), medium and long-term (SCenAgri) conditions. As pointed out above, one of the main challenges is to solve the problem of treating a large amount of data in the models, but technological advances will allow reducing this limitation as an impediment in the medium term. The main sectors that have looked for information about future climate impacts are pulp and paper, citrus and beef cattle.

With the increase in greenhouse gas emissions, resulting from anthropic actions, and its negative consequences for natural ecosystems and for the certification of Brazilian agricultural products, another great opportunity is related to technological solutions that incorporate the determination of the balance of greenhouse gas emissions by production systems. These technologies are based on the Greenhouse Gas Protocol (GHG) and are as the basis for Low Carbon Beef (LCB) or Carbon Neutral Beef (CNB) certifications (Alves et al., 2018).

Thus, in the future, in addition to the recommendations contained in the WebAgritec system, costs per production system and the calculation of the balance of emissions based on the GHG protocol will be incorporated. Thus, at each crop cycle or integrated systems, the rural producer will have the productivity and the carbon "footprint" on his rural property, which will help in the certification of his product. This certification takes place with the analysis of the balance of emissions, which will be carried out after the assimilation of emission factors originating from the Ministry of Science, Technology and Innovation´s National Inventory of Greenhouse Gases.

Traceability and certifications

Based on new national and international consumption patterns, food certification traceability processes have intensified in recent years. A study by Embrapa (2018) on the future of Brazilian agriculture for 2030 highlights that the traceability of products that contain information on their place of origin, inputs used, harvesting, slaughtering, processing, conservation, quality, storage and transport will become an essential condition for customer service, which will require transparency in relation to such characteristics.

Porpino and Bolfe (2020) emphasize that the search for certification of food products by Brazilian companies is an increasing pressure imposed by the consumer market, which demands guarantees about the nutritional, sanitary and hygienic characteristics of foods. There is a set of regulations and standards created by the National Health Surveillance Agency (ANVISA) and by the Ministry of Agriculture, Livestock and Supply for certain certifications, such as the "Good Manufacturing Practices" (GMP) and the Federal Inspection Service (SIF). In addition to the described general certifications for food safety, there are increasing opportunities for agriculture to reach more demanding markets and consumers for processes and products with specific certifications, in particular: Socioenvironmental certifications such as Fair Trade, Certified Humane, Rainforest or Organic (Brasil, 2003); Good Agricultural Practices (FAO, 2016); Animal Welfare (Brazil, 2017); Geographical Indications (National Institute of Industrial Property, 2019); International Organization for Standardization (2020); and Food Safety System Certification (2020). These certifications consider the complexity of agriculture and are based on nationally and/or internationally recognized metrics, criteria and protocols.

New opportunities are also envisioned for digital agriculture facing innovation challenges (Embrapa, 2019), which highlight the need to:

a) Provide digital and cyberphysical solutions to support the identification, traceability, sensing and certification of livestock and animal and vegetable products. A great deal of support for traceability can come from the use of blockchain technology, as it provides a large distributed database that can track what happened in the various links of the production chain.

b) Expand granting the certificates of geographical indication to agricultural products and processes, with intrinsic value and proper identity of the place of origin, such as soil, vegetation and climate.

c) Optimize traceability and certification in accordance with standards of control agencies and consumer demands in the animal protein, eggs, milk, fruits, vegetables and grains chains.

d) Expand the traceability and rapid diagnosis of pathogens, toxins and drug residues carried by food of animal origin, of economic and public health interest.

Society 5.0

Digital agriculture, also known here as "Agriculture 4.0", has been presented as an alternative to solve major agricultural challenges. Note that digital agriculture extends the idea of observing, measuring and connecting intelligent machines from precision agriculture to Big Data platforms and automated machine learning, sensors, satellites, drones and robots.

Digital technologies are facilitators that can optimize agricultural planning and production processes to achieve sustainability goals, enable better decision-making and remodel the functioning of agrifood markets, improve the quality of life of agricultural workers and the rural population, being able to attract a younger generation to agriculture and new rural businesses.

The robustness of Brazilian agribusiness favors the use of these new technologies, but the country is still having to overcome challenges related to training, telecommunications infrastructure, regulation, standard-setting and information security, in addition to high costs. Without a doubt, the covid-19 pandemic marked the end of the 20th century and, officially, the beginning of the 21st century, operating as an accelerator of futures, starting a new revolution in modern society.

After significant progress in mechanization, electrification, information and network technology, modern society has entered a new technology development era: the parallel era of dual virtual and augmented reality technology. Similarly, our society is shifting from a machine society (Society 1.0), electrical society (Society 2.0), information society (Society 3.0) and network society (Society 4.0) to its fifth paradigm: the parallel society or 5.0 Societies (World Economic Forum, 2019), in which there should be no separation between the physical and virtual world.

The basic research theory in Societies 5.0 is parallel intelligence, which is a new methodology that extends traditional theories of artificial intelligence to those emerging from cyber-physical-social systems (Cyber-Physical-Social Systems – CPSS) (Zhang, 2016). This concept of parallel intelligence can be presented as one of the enabling technologies for a more predictive and intelligent agriculture, which can contribute to meeting the new demands for increased sustainable production and productivity in three dimensions: economic, environmental and social. It is in this context that the concept of Agriculture 5.0 is inserted, which, in addition to the massive use of artificial intelligence and biotechnology in agricultural productive processes, it will ensure the production and distribution of food in a more economical and ecologically efficient manner than is currently practiced (Fraser; Campbell, 2019).

In addition to greater demand for food, there is another trend towards behavioral change in populations, which is due to growing urbanization, increased life expectancy, new work relationships and access to information. This combination brings to the environment of cities the concept of "urban agriculture" (FAO, 2011), which includes different aspects, such as indoor production, in controlled environments, combined organic production, raising bees and small animals, community gardens, production on roofs, etc.

In the context of urban development, the preservation and conservation of the environment must be taken into account, and other examples: promote collection, treatment and recycling of solid waste, using water rationally, using clean energy efficiently, develop and use digital technologies and innovative business models such as the internet of things (IoT) and wearable technologies, use autonomous vehicles, circular and shared economy, ensure zero net emission of greenhouse gases and propose new housing solutions, taking into account the principles of sustainable development.

Final considerations

Brazil has an innovative role in the global context of the digital transformation of agriculture. Mobile applications support decision making on numerous practices involving animal and plant production. The use of applications has increasingly supported the monitoring of phytosanitary conditions, the application of pesticides, biological control, animal welfare, soil management and irrigation management. Planning activities associated with the ZARC and the Rural Environmental Registry (CAR) are already part of the day-to-day activities of rural properties. These instruments support land use and occupation planning, the recovery of degraded areas, the implementation of more resilient and low-carbon agricultural systems, such as integrated crop-livestock-forestry (ICLF) and no-tillage.

However, there are still important scientific, technological, social and economic challenges to be overcome in order to integrate the digital transformation of Brazilian agriculture in the different agricultural classes and regions. There are also countless opportunities for research institutes, universities, companies, startups, cooperatives, associations and unions to generate more integrated digital solutions for planning, managing, harvesting and marketing products such as milk, honey, eggs, meat, grains, fruits, sugar, biofuels, fibers, wood and cellulose (Figure 2).

The pandemic linked to covid-19 is accelerating and shaping the digitalization of all links in agricultural production chains. The need for greater food security, with the possibility of using technologies that reduce physical contact, drives new applications from input suppliers to rural producers, from marketing to transport, and from distribution to final consumers.

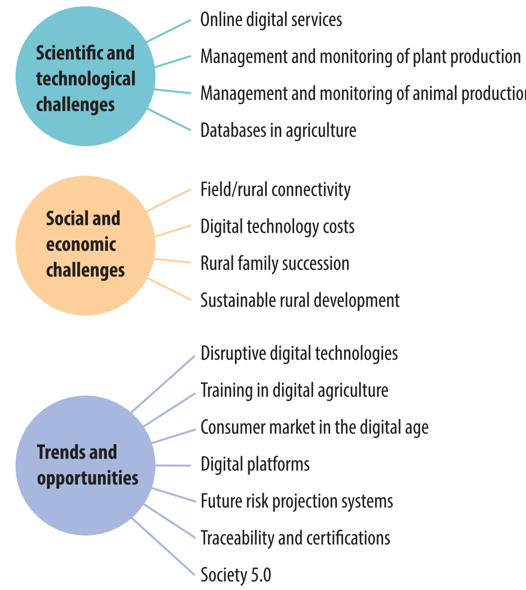


Figure 2. Main challenges and opportunities in the digital transformation of Brazilian agriculture.

In the "new normal" post-pandemic, digital connectivity and content services associated with links in the chains could expand with the growing concerns about the health of populations and the sanitary and nutritional safety of food. Public and private managers, entrepreneurs, service providers and rural producers need to consider, in their decisions, the aspects of digital transformation and its implications and interconnections with other links in the production chains and food security. E-commerce giants have taken advantage of the already installed capacity in logistics and distribution to sell food products in certain urban centers worldwide. However, digital gaps between the poorest and richest families, as well as between rural and urban populations, will probably persist.

These and other conditions are foreseeing the future of digitization of Brazilian agriculture, when research, innovation and business are expected to expand rapidly in infrastructure and services such as:

- Cognitive artificial intelligence for monitoring production.
- Multi-scale and multi-source analyses of agricultural risks.
- Real-time monitoring of properties by remote sensing.
- Machine and equipment maintenance of prediction systems.
- Processing of agricultural big data and small data in the cloud.
- Sales platforms via short circuits integrating producers and consumers.
- Distance learning and work applications with safe administrative procedures and social interaction of staff teams.
- Blockchain and digital encryption technologies for the security of commercial transactions and the traceability of products and food.
- Technical-financial management systems considering economic, environmental and social aspects of the property.
- Security and privacy of data and information generated in all digital processes.

Hence, the digital transformation of Brazilian agriculture will play a far more relevant role in the coming years regarding the production of food, fiber and energy in greater quantity, quality and sustainability.

References

ADÃO, T.; HRUŠKA, J.; PÁDUA, L.; BESSA, J.; PERES, E.; MORAIS, R.; SOUSA, J.J. Hyperspectral imaging: a review on UAV-based sensors, data processing and applications for agriculture and forestry. **Remote Sensing**, v. 9, article 1110, 2017. DOI: 10.3390/rs9111110.

AGÊNCIA NACIONAL DE TELECOMUNICAÇÕES (Brasil). **Relatório e comparação internacional**. Brasília, DF, 2020. 19 p. Available at: https://www.anatel.gov.br/ dados/relatorios-de-acompanhamento/2020. Accessed on: 14 Oct. 2020.

ALVES, F. V.; ALMEIDA, R. G. de; LAURA, V. A.; PORFIRIO-DA-SILVA, V.; GOMES, R. da C.; MACEDO, M. C. M.; PEREIRA, M. de A.; FERREIRA, A. D.; BUNGENSTAB, D. J. **50 perguntas, 50 respostas sobre a Carne Carbono Neutro (CCN)**. Brasília, DF: Embrapa, 2018. 29 p.

ASSAD, E. D.; OLIVEIRA, A. F.; NAKAI, A. M.; PAVÃO, E. J.; MONTEIRO, J. E.; PELLEGRINO, G. Q. Impactos e vulnerabilidades da agricultura brasileira as mudanças climáticas. In: TEIXEIRA, B. S.; ORSINI, J. A. M.; CRUZ, M. R. (ed.). **Modelagem climática e vulnerabilidades setoriais à mudança do clima no Brasil**. Brasília, DF: Ministério da Ciência Tecnologia e Inovação, 2016. 590 p. Available at: https://files.cercomp.ufg.br/weby/up/804/o/ModelagemClimticaeVulnerabilidadeSetoriaisMudanadoClimanoBrasil.pdf?1528299061. Accessed on: 14 Oct. 2020.

BANCO NACIONAL DE DESENVOLVIMENTO ECONÔMICO E SOCIAL. **Internet das Coisas**: um plano de ação para o Brasil. Brasília, DF: BNDES; MCTIC: McKinsey/Fundação CPqD: 2017. 65 p. Availabe at: https://www.bndes.gov.br/wps/portal/ site/home/onde-atuamos/inovacao/internet-das-coisas. Accessed on:14 Oct. 2020.

BARBEDO, J. G. A. A review on the use of unmanned aerial vehicles and imaging sensors for monitoring and assessing plant stresses. **Drones**, v. 3, n. 2, article 40, 2019a. DOI: 10.3390/drones3020040.

BARBEDO, J. G. A. Detection of nutrition deficiencies in plants using proximal images and machine learning: A review. **Computers and Electronics in Agriculture**, v. 162, p. 482-492, July 2019b. DOI: 10.1016/j.compag.2019.04.035.

BARBEDO, J. G. A.; KOENIGKAN, L. V. Perspectives on the use of unmanned aerial systems to monitor cattle. **Outlook on Agriculture**, v. 47, n. 3, p. 214-222, June 2018. DOI: 10.1177/0030727018781876.

BARBEDO, J. G. A.; KOENIGKAN, L. V.; SANTOS, P. M.; RIBEIRO, A. R. B. Counting Cattle in UAV Images-Dealing with Clustered Animals and Animal/Background Contrast Changes. **Sensors**, v. 20, n. 7, article 2126, 2020. DOI: 10.3390/s20072126.

BARBEDO, J. G. A.; KOENIGKAN, L. V.; SANTOS, T. T.; SANTOS, P. M. A study on the detection of cattle in UAV images using deep learning. **Sensors**, v. 19, article 5436, 2019. DOI: 10.3390/s19245436.

BAYINDIR, L. A review of swarm robotics tasks. **Neurocomputing**, v. 172, p. 292-321, 2016. DOI: 10.1016/j.neucom.2015.05.116.

BECHAR, A.; VIGNEAULT, C. Agricultural robots for field operations: Concepts and components. **Biosystems Engineering**, v. 149, p. 94-111, 2016. DOI: 10.1016/j.biosystemseng.2016.06.014.

BOCK, C. H.; BARBEDO, J. G. A.; DEL PONTE, E. M.; BOHNENKAMP, D.; MAHLEIN, A.-K. Fromvisual estimates to fully automated sensor-based measurements of plant disease severity. **Phytopathology Research**, v. 2, n. 9, Apr. 2020. DOI: 10.1186/s42483-020-00049-8.

BOLFE, E. Application of geotechnologies in the development of sustainable agriculture in Brazil. **International Journal of Advanced Engineering Research and Science**, v. 6, p. 458-463, 2019. DOI: 10.22161/ijaers.612.53.

BOLFE, E. L.; JORGE, L. A. C.; SANCHES, I.; COSTA, C. C. DA; LUCHIARI JR., A.; VICTÓRIA, D.; INAMASU, R.; GREGO, C.; FERREIRA, V.; RAMIREZ, A. **Agricultura digital no Brasil**: tendências, desafios e oportunidades: resultados de pesquisa online. Campinas: Embrapa, 2020. 44 p. Available at: https://www.embrapa.br/agropensa/produtos-agropensa. Accessed on: 14 Oct. 2020.

BOLFE, E.; MASSRUHÁ, S. A transformação digital e a sustentabilidade agrícola. **Agroanalysis**, v. 40, p. 32-34, Mar. 2020.

BRASIL. **Decreto Federal nº 9.013 de 2017**. Available at: http://www.planalto.gov.br/ ccivil_03/_ato2015-2018/2017/decreto/d9013.htm. Accessed on: 15 Mar. 2020.

BRASIL. **Decreto nº 9.854 de 2019a**. Institui o Plano Nacional de Internet das Coisas e dispõe sobre a Câmara 4.0. Available at: http://www.planalto.gov.br/ ccivil_03/_Ato2019-2022/2019/Decreto/D9854.htm. Accessed on: 14 May 2020.

BRASIL. **Lei nº 10.831 de 23 de dezembro de 2003**. Dispõe sobre a agricultura orgânica e dá outras providências. Available at: https://www.planalto.gov.br/ccivil_03/leis/2003/l10.831.htm. Accessed on: 15 Mar. 2020.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Projeções do agronegócio**: Brasil 2018/19 a 2028/29 projeções de longo prazo. Brasília, DF: Mapa/ACE, 2019b. 126 p.

BRASIL. Ministério da Ciência, Tecnologia, Inovações e Comunicações - MCTIC; Ministério da Agricultura, Pecuária e Abastecimento - MAPA. **Câmara Agro 4.0**. Available at: https://www.embrapa.br/busca-de-noticias/-/noticia/47656442/mapa-e-mctic- promovem-primeira-reuniao-da-camara-do-agro-40. Acessed on: 5 maio 2020b.

BRASIL. **Projeto de Lei n° 172 de 2020**. Altera a Lei Geral de Telecomunicações e a Lei do Fust. Available at: https://www25.senado.leg.br/web/atividade/materias/-/materia/140555. Acessed on: 10 May 2020a.

ESCOLA SUPERIOR DE AGRICULTURA LUIZ DE QUEIROZ. Centro de Estudos. Avançados em Economia Agrícola. **Coronavírus e o agronegócio – covid-19 e agroalimentos**: recalibrando expectativas. Piracicaba, 2020. 16 p.

CHEN, K.; WANG, Y.; ZHANG, R.; ZHANG, H. E.; GAO, C. CRISPR/Cas Genome Editing and Precision Plant Breeding in Agriculture. **Annual Review of Plant Biology**, v. 70, p. 667-697, 2019. DOI: 10.1146/annurev-arplant-050718-100049.

CONECTARAGRO. Agricultura 4.0. Available at: https://conectaragro.com.br. Acessed on: 14 May 2020.

COSTA, C.; GUILHOTO, J. M. Impactos Potenciais da Agricultura de Precisão sobre a Economia Brasileira. Revista de Economia e Agronegócio, v. 10, p. 177-204, 2013. DOI: 10.25070/rea.v10i2.201.

DEBOER, J. **Precision agriculture for sustainability**. Cambridge, United Kingdom: Burleigh Dodds Science Publishing Limited, 2019. 514 p.

DUNCOMBE, R. Mobile phones for agricultural and rural development: a literature review and suggestions for future research. **European Journal of Development Research**, v. 28, p. 213-235, 2016. DOI: 10.1057/ejdr.2014.60.

EMBRAPA INFORMÁTICA AGROPECUÁRIA. **AgroAP**. Campinas, 2019a. Available at: https://www.agroapi.cnptia.embrapa.br/portal. Acessed on: 24 May 2020.

EMBRAPA INFORMÁTICA AGROPECUÁRIA. **Digipathosrep**. Campinas, 2019b. Available at: https://www.digipathos-rep.cnptia.embrapa.br. Acessed on: 10 May 2020.

EMBRAPA INFORMÁTICA AGROPECUÁRIA. **SCenAgri**. Campinas, 2020a. Available at: https://www.scenagri.cnptia.embrapa.br/scenagri/publico/login.xhtml. Acessed on: 13 May 2020.

EMBRAPA INFORMÁTICA AGROPECUÁRIA. **Soluções tecnológicas**. Available at: https://www. embrapa.br/informatica-agropecuaria/solucoes-tecnologicas. Acessed on: 14 May 2020b.

EMBRAPA INSTRUMENTAÇÃO. **Soluções tecnológicas**. Available at: https://www. embrapa.br/instrumentacao/solucoes-tecnologicas. Acessed on: 14 May 2020.

EMBRAPA MEIO AMBIENTE. **Soluções tecnológicas**. Available at: https://www. embrapa.br/meio-ambiente/solucoes-tecnologicas. Acessed on: 14 May 2020.

EMBRAPA TERRITORIAL. **Soluções tecnológicas**. Available at: https://www.embrapa.br/territorial/solucoes-tecnologicas. Acessed on: 14 May 2020.

EMBRAPA. **Portfólios de P&D da Embrapa**: desafios de inovação. Brasília, DF: SPD/Ideare, 2019.

EMBRAPA. **Rede AP**. 2018b. Available at: https://www.embrapa.br/tema-automacao-e-agricultura-de-precisao. Acessed on: 14 May 2020.

EMBRAPA. **Visão 2030**: o futuro da agricultura brasileira. Brasília, DF, 2018a. 212 p.

FAO. **Good Agricultural Practice – GAP**. 2016. Available at: http:// www.fao.org/3/a-i6677e.pdf. Acessed on: 10 Mar. 2020.

FAO. **The place of urban and peri-urban agriculture in national food security programmes**. 2011. 60 p. Available at: http://www.fao.org/3/i2177e/i2177e00.pdf. Acessed on: 14 May 2020.

FRASER, E.; CAMPBELL, M. Agriculture 5.0: reconciling production with planetary health. **One-earth**, v.1, n. 3, p. 278-280, Nov. 2019. DOI: 10.1016/j.oneear.2019.10.022.

FOOD SAFETY SYSTEM CERTIFICATION. **Food Safety System Certification 22000**. 2019. Available at: https://www.fssc22000.com. Acessed on: 10 Mar. 2020.

FUNDAÇÃO ABC. **SigmaABC**. 2020. Available at: https://www.sigmaabc.org. Acessed on: 24 May 2020.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. Agricultura 4.0. **Revista Fapesp**, ed. 287, jan. 2020. Available at: https://revistapesquisa.fapesp.br/2020/01/02/agricultura-4-0. Acessed on: 15 May 2020.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. **Rede de repositório de dados**. São Paulo, 2019. Available at: http://agencia.fapesp.br/fapesp-lanca-rede-de-repositorios-de-dados-cientificos-do-estado-de-sao-paulo/32251. Acessed on: 10 May 2020.

GASQUES, J. G.; BACCHI, M. P. R.; RODRIGUES, L.; BASTOS, E. T.; VALDEZ, C. Produtividade da agricultura brasileira: a hipótese da desaceleração. In: VIEIRA FILHO, J. E. R.; GASQUES, J. G. (org.). **Agricultura, transformação produtiva e sustentabilidade**. Brasília, DF: Ipea, 2016. p. 143-163.

GHINI, R.; BETTIOL, W.; HAMADA, E. Diseases in tropical and plantation crops as affected by climate changes: current knowledge and perspectives. **Plant Pathology**, v. 60, p. 122-132, 2011a. DOI: 10.1111/j.1365-3059.2010.02403.x.

GHINI, R.; HAMADA, E.; PEDRO JUNIOR, M.; GONÇALVES, R. Incubation period of Hemileia vastatrix in coffee plants in Brazil simulated under climate change. **Summa Phytopathologica**, v. 37, n. 2, p. 85-93, Apr./June 2011b. DOI: 10.1590/S0100-54052011000200001.

HOUBORG, R.; MCCABE, M. F. High-resolution NDVI from planet's constellation of earth observing nano-satellites: a new data source for precision agriculture. **Remote Sensing**, v. 8, n. 9, article 768, 2016. DOI: 10.3390/rs8090768.

HUUSKONEN, J.; OKSANEN, T. Soil sampling with drones and augmented reality in precision agriculture.**Computers and Electronics in Agriculture**, v. 154, p. 25-35, 2018. DOI: 10.1016/j.compag.2018.08.039.

IBGE. **Censo agropecuário 2017**. Available at: https://censos. ibge.gov.br/agro/2017. Acessed on: 5 May 2020.

INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL (Brasil). **Legislações de indicação geográfica**. 2019. Available at: http://www.inpi.gov.br/menu-servicos/indicacao-geografica. Acessed on: 10 May 2020.

INTERNATIONAL FUND FOR AGRICULTURAL DEVELOPMENT. **Covid-19**. 2020. Available at: https://www.ifad.org/en/covid19. Acessed on: 10 May 2020.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. 2019. Available at: https://www.iso.org/home.html. Acessed on: 10 Mar. 2020.

IRWIN, A. **Citizen science**: a study of people, expertise and sustainable development. London: Routledge, 2002. 216 p.

KLERKXA, L.; ROSEB, D. Dealing with the game-changing technologies of Agriculture 4.0. **Global Food Security**, v. 24, n.3, p. 1-7, Mar. 2020. DOI: 10.1016/j.gfs.2019.100347.

LIAKOS, K. G.; BUSATO, P.; MOSHOU, D.; PEARSON, S.; BOCHTIS, D. Machine learning in agriculture: a review. **Sensors**, v. 18, n. 8, article 2674, 2018. DOI: 10.3390/s18082674.

LIN, Y. P.; PETWAY, J. R.; ANTHONY, J.; MUKHTAR, H.; LIAO, S. W.; CHOU, C. F.; HO, Y.F. Blockchain: the evolutionary next step for ICT e-agriculture. **Environments**, v. 4, article 50, 2017. DOI: 10.3390/environments4030050.

MAHLEIN, A.-K. 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, Jan. 2016. DOI: 10.1094/PDIS-03-15-0340-FE.

MARIN, F. R.; NASSIF, D. S. P. Mudanças climáticas e a cana-de-açúcar no Brasil: fisiologia, conjuntura e cenário futuro. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 17, p. 232-239, fev. 2013. DOI: 10.1590/S1415-43662013000200015.

MCKINSEY CONSULTORIA. **Nossas publicações**. 2020. Available at: https://www.mckinsey.com.br. Acessed on: 14 May 2020.

MILLER, G. A.; HYSLOP, J. J.; BARCLAY, D.; EDWARDS, A.; THOMSON, W.; DUTHIE, C.-A. Using 3Dimaging and machine learning to predict liveweight and carcass characteristics of live finishing beef cattle. **Frontiers in Sustainable Food Systems**, v. 3, May 2019. DOI: 10.3389/fsufs.2019.00030.

NAÇÕES UNIDAS **The 2030 Agenda for Sustainable Development in the new global and regional context**: scenarios and projections in the current crisis. Santiago: Economic Commission for Latin America and the Caribbean, 2020. 61 p.

NACÕES UNIDAS. **Objetivos de Desenvolvimento Sustentável**. 2015. Available at: https://nacoesunidas.org/pos2015. Acessed on: 5 maio 2020.

PRESKILL, J. Quantum computing in the NISQ era and beyond. **Quantum**, v. 2, n. 79, 2018. DOI: 10.22331/q-2018-08-06-79.

PWC BRASIL. **Empresas familiares e plano de sucessão**. 2019. Available at: https://www.pwc.com.br/ pt/sala-de-imprensa/artigos/empresas-familiares-e-plano-de-sucessao.html. Acessed on: 6 May 2020.

RAHMAN, A.; SMITH, D. V.; LITTLE, B.; INGHAM, A. B.; GREENWOOD, P. L.; BISHOP-HURLEY, G. J.Cattle behaviour classification from collar, halter, and ear tag sensors. **Information Processing in Agriculture**, v. 5, n. 1, p. 124-133, Mar. 2018. DOI: 10.1016/j.inpa.2017.10.001.

REINA, G.; MILELLA, A.; ROUVEURE, R.; NIELSEN, M.; WORST, R.; BLAS, M. R. Ambientawareness for agricultural robotic vehicles. **Biosystems Engineering**, v. 146, p. 114-132, June 2016. DOI: 10.1016/j.biosystemseng.2015.12.010.

ROOPAEI, M.; RAD, P.; CHOO, K. R. Cloud of Things in Smart Agriculture: Intelligent Irrigation Monitoring by Thermal Imaging. **IEEE Cloud Computing**, v. 4, n. 1, p. 10-15, Mar. 2017. DOI: 10.1109/MCC.2017.5.

ROSE, D. C.; SUTHERLAND, W. J.; PARKER, C.; LOBLEY, M.; WINTER, M.; MORRIS, C.; TWINING, S.; FFOULKES, C.; AMANO, T.; DICKS, L. V. Decision support tools for agriculture: Towards effective design and delivery. **Agricultural Systems**, v. 149, p. 165-174, Nov. 2016. DOI: 10.1111/soru.12233.

ROTZ, S.; DUNCAN, E.; SMALL, M.; BOTSCHNER, J.; DARA, R.; MOSBY, I.; REED, M.; FRASER, E. D. The Politics of Digital Agricultural Technologies: A Preliminary Review. **Sociologia Ruralis**, v. 59, p. 203-229, Feb. 2019. DOI: 10.1111/soru.12233.

SEBRAE. **SebraeTec**. 2019. Available at: http://www.sistemasebraetec.sebrae.com.br. Acessed on: 14 May 2020.

SENAR. **EAD**. 2019. Available at: http://ead.senar.org.br/cursos/ agricultura-de-precisao. Acessed on: 25 May 2020.

SILVERTOWN, J. A new dawn for citizen science. **Trends in Ecology & Evolution**. v. 24, n. 9, p. 467-471, 2009. DOI: 10.1016/j.tree.2009.03.017.

THOMAS, S.; KUSKA, M. T.; BOHNENKAMP, D.; BRUGGER, A.; ALISAAC, E.; WAHABZADA, M.; BEHMANN, J.; MAHLEIN, A.-K. Benefits of hyperspectral imaging for plant disease detection and plant protection: a technical perspective. **Journal of Plant Diseases and Protection**, v. 125, p. 5-20, 2018. DOI: 10.1007/s41348-017-0124-6.

TZOUNIS, A.; KATSOULAS, N.; BARTZANAS, T.; KITTAS, C. Internet of Things in agriculture, recent advances and future challenges. **Biosystems Engineering**, v. 164, p. 31-48, Dec. 2017. DOI: 10.1016/j.biosystemseng.2017.09.007.

UNITED NATIONS GLOBAL COMPACT - UNGC. **Digital agriculture**. 2017. http://breakthrough.unglobalcompact.org/disruptive-technologies/digital-agriculture. Acessed on: 19 May 2020.

WOERNER, S; EGGER, D. Quantum risk analysis. **Quantum Information**, n. 5, article 15, Feb. 2019.

WOLFERT, S.; GE, L.; VERDOUW, C.; BOGAARDT, M. J. Big data in smart farming – a review. **Agricultural Systems**, v. 153, p. 69-80, May 2017. DOI: 10.1016/j.agsy.2017.01.023.

WORLD ECONOMIC FORUM. **Modern society has reached its limits. Society 5.0 will liberate us.** 2019. Available at: https://www.weforum.org/agenda/2019/01/modern- society-has-reached-its-limits-society-5-0-will-liberate-us. Acessed on: 19 May 2020.

ZHANG, Y. Grorec: a group-centric intelligent recommender system integrating social, mobile and big data technologies. **IEEE Transactions on Services Computing**, v. 9, n. 5, p. 786-795, Sept./Oct. 2016. DOI: 10.1109/TSC.2016.2592520.

Embrapa

Digital Agriculture

MINISTRY OF
AGRICULTURE AND
LIVESTOCK

BRAZILIAN GOVERNMENT
BRASIL
UNITING AND REBUILDING

ISBN 978-65-89957-72-0

CGPE 018055