WESTERN PARANÁ STATE UNIVERSITY – UNIOESTE
CENTER FOR EXACT AND TECHNOLOGICAL SCIENCES – CAMPUS CASCAVEL
GRADUATE PROGRAM IN AGRICULTURAL ENGINEERING

RICARDO SOBJAK

INCORPORATION OF COMPUTATIONAL MODULES AS MICROSERVICES IN THE AGDATABOX PLATFORM AND DEVELOPMENT OF AGDATABOX-MAP APPLICATION

CASCAVEL
2021

RICARDO SOBJAK

INCORPORATION OF COMPUTATIONAL MODULES AS MICROSERVICES IN THE AGDATABOX PLATFORM AND DEVELOPMENT OF AGDATABOX-MAP APPLICATION

Thesis presented to the Graduate Program in Agricultural Engineering (PGEAGRI) in compliance with the requirements for obtaining the title of Ph.D. at Agricultural Engineering, concentration area in Biological and Agroindustrial Systems Engineering.

Advisor: Ph.D. Eduardo Godoy de Souza.

Co-advisor: Ph.D. Claudio Leones Bazzi.

CASCAVEL – PARANÁ – BRASIL

2021

CATALOG RECORD

Ficha de identificação da obra elaborada através do Formulário de Geração Automática do Sistema de Bibliotecas da UNIOESTE.

Sobjak, Ricardo
Incorporation of computational modules as microservices in the AgDataBox platform and development of the AgDataBox-Map application / Ricardo Sobjak; orientador Eduardo Godoy de Souza; coorientador Claudio Leones Bazzi. -- Cascavel, 2021.
256 p.

Tese (Doutorado Campus de Cascavel) -- Universidade Estadual do Oeste do Paraná, Centro de Ciências Exatas e Tecnológicas, Programa de Pós-Graduação em Engenharia Agrícola, 2021.

1. Precision agriculture. 2. Digital agriculture. 3. Agriculture software. 4. Microservices architecture. I. Souza, Eduardo Godoy de, orient. II. Bazzi, Claudio Leones, coorient. III. Título.

Portuguese, English, and standards reviewed by Ana Maria Vasconcelos on November 5, 2021.

RICARDO SOBJAK

INCORPORATION OF COMPUTATIONAL MODULES AS MICROSERVICES IN THE AGDATABOX PLATFORM AND DEVELOPMENT OF AGDATABOX-MAP APPLICATION

Thesis presented to the Graduate Program in Agricultural Engineering in compliance with the requirements for obtaining Doctor title at Agricultural Engineering, concentration area in Biological and Agroindustrial Systems, research line Geoprocessing, Spatial Statistics, and Precision Agriculture, **APPROVED** by the following examining board:

Advisor - Eduardo Godoy de Souza
Western Paraná State University (UNIOESTE) – Cascavel Campus

Fabio Henrique Rojo Baio
Federal University of Mato Grosso do Sul (UFMS)

José Paulo Molin
University of São Paulo (USP)

Luis Henrique Bassoi
Brazilian Agricultural Research Corporation (EMBRAPA)

Miguel Angel Uribe Opazo
Western Paraná State University (UNIOESTE) – Cascavel Campus

Erivelto Mercante
Western Paraná State University (UNIOESTE) – Cascavel Campus

Cascavel, August 11, 2021.

BIOGRAPHY

Ricardo Sobjak, Brazilian, married, from Medianeira/PR, son of Sérgio Luis Sobjak and Alice Pugslei Sobjak. I graduated in Computer Technology at Federal Technological University of Paraná (UTFPR), in 2008; I concluded Design and Development of Object-Based Systems for the Internet Environment specialization course at UTFPR in 2009; I finalized my master in Agricultural Engineering at the Western Paraná State University (UNIOESTE) in 2012. After that, I worked as an IT technician at UTFPR - Campus Medianeira (2007-2011). Then, I worked as a computer science professor at the Federal Institute of Santa Catarina (IFSC) - Campus Chapecó (2011-2012). Currently, I work as a computer science professor at UTFPR - Campus Medianeira (since December 2012). I joined the *Stricto Sensu* Graduate Course in Agricultural Engineering at UNIOESTE, Ph.D. level, in March 2017, in the Biological and Agroindustrial Systems area, whose topic is related to precision agriculture and digital agriculture.

DEDICATION

To God,

to my wife,

to my children,

to my parents and my brothers,

I dedicate this work with all my love.

ACKNOWLEDGEMENTS

I thank God for life and the opportunities I've received. All honor and glory to Him.

To my wife, Emanuela, for loving, understanding, supporting me with patience during this moment.

To my children, Eduardo and Isabel, for simply being precious in my life.

To my parents, Sérgio and Alice, for love, teachings, and dedication they had along my whole life.

To my brothers, Anderson and Rodrigo, and other familiar members for their support, growth, and learning together.

To Professor Ph.D. Eduardo Godoy de Souza for his guidance, trust, advice, friendship, and support for this work.

To Professor Ph.D. Claudio Leones Bazzi for the opportunities offered, partnership, friendship, and for helping to coordinate this work.

To the Western Paraná State University (UNIOESTE), for the opportunity to participate in the Doctoral Program, as well as to use its infrastructure.

To the Federal Technological University of Paraná (UTFPR), for the opportunity to hold a Doctorate and for the infrastructure provided in researching in partnership with UNIOESTE.

To friends and colleagues from PGEAGRI, from whom we were able to study, work and learn together.

To all other professors, colleagues, and employees who, directly or indirectly, participated in carrying out this research.

INCORPORATION OF COMPUTATIONAL MODULES AS MICROSERVICES IN THE AGDATABOX PLATFORM AND DEVELOPMENT OF AGDATABOX-MAP APPLICATION

ABSTRACT

Digital technologies can provide farmers, specialists, and researchers in the agricultural area the in-deep knowledge of their cropping field, in addition to effective farm management, aiming at increasing profitability and reducing environmental impact. Among the available technologies, there are others, such as precision agriculture, Big Data, Internet of Things (IoT), Unmanned Aerial Vehicles (UAV), robotics, and automation. Digital agriculture requires specific portals and platforms for its adoption. The free web platform AgDataBox (ADB) contributes to farmers' inclusion in the digital agriculture phase. ADB-Map web application and ADB microservices architecture (ADB-MSA) take part on of thematic maps creation (TMs) and the delineation of management zones (MZs) in a friendly and integrated way. Thus, this study aims to integrate the functionalities for creating TMs and MZs through microservices in ADB-MSA and incorporating them into ADB-Map application. ADB-MSA provided eight microservices, six of them (statistics, spatial, interpolation, clustering, rectification, and lime/nutrient recommendation) execute procedures based on JavaScript, R, and Python programming languages, while the others are used to store data. ADB-Map was rewritten in JavaScript language and Angular framework, based on a software architecture that decoupled the front-end from the back-end and now uses the resources of ADB-MSA. In the case study, the procedures to create TMs and delineate MZs were carried out satisfactorily with data from a commercial area. Thus, the MZs were generated and evaluated to apply fertilizer according to phosphorus and potassium requirement. In order to improve the interpolator selection process, the new semivariogram model selection criteria were adopted (i) effective spatial dependence index (%ESDI) > 25%, (ii) first semivariance significance index ($\%\gamma(1)$) < 50% and (iii) slope of the model ends index (%SMEI) > 20%, which were applied according to three methods: 1) only with the interpolator selection index (ISI) without application regarding the proposed criteria; 2) the criteria applied after the interpolator selection analysis + ISI, and 3) the criteria applied during the interpolator selection analysis + ISI. Thus, it was observed that, usually, the three methods selected different models and Method 3 was considered the best.

Keywords: digital agriculture, precision agriculture, management zones.

INCORPORAÇÃO DE MÓDULOS COMPUTACIONAIS COMO MICROSERVICES NA PLATAFORMA AGDATABOX E DESENVOLVIMENTO DA APLICAÇÃO AGDATABOX-MAP

RESUMO

As tecnologias digitais podem proporcionar ao agricultor, especialistas e pesquisadores, o conhecimento aprofundado da área de cultivo, além de uma gestão eficaz da propriedade agrícola, objetivando o aumento da lucratividade e a diminuição do impacto ambiental. Dentre as tecnologias disponíveis, destacam-se agricultura de precisão, Big Data, Internet das coisas (IoT), veículos aéreos não tripulados (VANT), robótica e automação. A agricultura digital necessita da disponibilização de portais e plataformas específicos para sua adoção. A plataforma web AgDataBox (ADB) é gratuita e contribui para a inclusão dos agricultores na fase da agricultura digital. A aplicação ADB-Map e a arquitetura de microserviços da ADB (ADB-MSA) tornam amigáveis e integrados aos processos de criação de mapas temáticos (MTs) e de delineamento de zonas de manejo (ZMs). Assim, o objetivo deste trabalho foi integrar as funcionalidades para a criação de MTs e delineamento de ZMs por meio de microserviços web na ADB-MSA e incorporá-las na aplicação web ADB-Map. ADB-MSA fornece oito microsserviços, seis dos quais (estatísticas, espaciais, interpolação, agrupamento, retificação e recomendação de calcário/nutrientes) que executam procedimentos baseados nas linguagens de programação JavaScript, R e Python, enquanto os outros dois são usados para armazenar dados. A ADB-Map foi totalmente reescrita em linguagem JavaScript e framework Angular, baseada em uma arquitetura de software que desacoplou o front-end do back-end e agora consome os recursos da ADB-MSA. No estudo de caso, os procedimentos para criar MTs e delinear ZMs foram realizados satisfatoriamente com dados de uma área comercial. Assim, as ZMs para a aplicação do fertilizante foram delineadas e avaliadas de acordo com a necessidade de fósforo e potássio. Para melhorar o processo de seleção de interpoladores, novos critérios de seleção de modelos de semivariograma foram adotados: (i) índice de dependência espacial efetiva (%ESDI) > 25%, (ii) índice de importância da primeira semivariância ($\%\gamma(1)$) < 50% e (iii) índice de inclinação das extremidades do modelo (%SMEI) > 20%, os quais foram aplicados de acordo com três métodos: 1) apenas com o índice de seleção de interpolador (ISI) sem aplicação dos critérios propostos; 2) os critérios aplicados após a análise de seleção de interpolador + ISI, e 3) os critérios aplicados durante a análise de seleção de interpolador + ISI. Observou-se que geralmente os três métodos selecionaram modelos diferentes e que o Método 3 foi considerado o melhor.

Palavras-chave: agricultura digital, agricultura de precisão, zonas de manejo.

SUMMARY

LIST OF TABLES

PAPER 3 Process improvement to select the best interpolator and its parameters and create thematic maps

LIST OF FIGURES

LITERATURE REVIEW

PAPER 1 Incorporation of computational routines in a microservice architecture in AgDataBox platform

PAPER 2 AgDataBox-Map: a web application for creating thematic maps and management zones in precision agriculture and digital agriculture

PAPER 3 Process improvement of selecting the best interpolator and its parameters to create thematic maps

LIST OF ABBREVIATIONS AND ACRONYMS

ADB	AgDataBox
ADB-CLU-API	AgDataBox Clustering API
ADB-Data-API	AgDataBox Data API
ADB-INT-API	ADB Interpolation API
ADB-RCT-API	AgDataBox Rectification API
ADB-REC-API	ADB Recommendation API
ADB-RS	AgDataBox Remote Sensing
ADB-SP-API	AgDataBox Spatial API
ADB-ST-API	AgDataBox Statistics API
Al	Aluminum
ANOVA	Analysis of variance
API	Application Programming Interface
ASC	Average silhouette coefficient
B	Boron
C	Carbon
C0	Nugget Effect
C1	Partial Sill
Ca	Calcium
Cl	Chlorine
CRD	Coefficient of relative deviation
CSV	Comma-Separated Values
CTC	Cation exchange capacity
Cu	Copper
CV	Coefficient of variation
DA	Digital Agriculture
DSS	Decision Support System
ECI	Error comparison index
FCM	Fuzzy C-means
Fe	Iron
FI	Fragmentation index
FMIS	Farm Management Information Systems
FPI	Fuzziness performance index
GA	Global accuracy
GIS	Geographic information system
GNSS	Global Navigation Satellite Systems
GQI	Global Quality Index
GPS	Global Positioning System
H	Hydrogen
ha	Hectare
HTTP	Hypertext Transfer Protocol
ICVI	Improved Cluster Validation Index
IDW	Inverse distance weighting
IoT	Internet of Things
IoTAg	IoT-enabled Agricultural
INPI	National Institute of Industrial Property
ISI	Interpolator Selection Index
JSON	JavaScript Object Notation
K	Potassium
Kp	Kappa
MA	Moving Average
MAD	Mean absolute difference
ME	Mean error

Mg	Magnesium
MGQI	Modified Global Quality Index
Mn	Manganese
Mo	Molybdenum
MPE	Modified partition entropy
MSA	Microservices architecture
MULTISPATI-PCA	Multivariate spatial analysis based on the Moran's index and PCA
MZ	Management zone
N	Nitrogen
Na	Sodium
NCE	Normalized classification entropy
NDVI	Normalized difference vegetation index
NDRE	Normalized difference red edge index
NIR	Near-infrared spectroscopy
NN	Nearest neighbor
O	Oxygen
OK	Ordinary kriging
OM	Organic matter
P	Phosphorus
PA	Precision Agriculture
PCA	Principal Component Analysis
Ra	Range
RE	Relative efficiency
REST	Representational State Transfer
RGB	Red, green, and blue
RME	Reduced mean error
RPC	Remote Procedures Call
S	Sulfur
SD	Spatial dependence
SDI	Spatial dependence index
SDUM	Software for defining management zones
SDME	Standard deviation of mean error
SDRME	Standard deviation of the reduced mean error
SI	Smoothness Index
SOAP	Simple Object Access Protocol
SRP	Soil resistance to penetration
TM	Thematic map
UAS	Unmanned aerial systems
UAV	Unmanned aerial vehicles
UGV	Unmanned ground vehicle
UN	United Nations Organization
URI	Uniform Resource Identifier
UTM	Universal Transverse Mercator
VRA	Variable-rate application
VR	Variance reduction
Vis-NIR	Visible-Near Infrared
WSDL	Web Services Description Language
WSN	Wireless sensor networks
XB	Xie and Beni Index
XML	Extensible Markup Language
Zn	Zinc

1 INTRODUCTION

Among the factors that stimulate the development of technologies applied to agriculture, food production stands out. According to the United Nations (UN) projections, the world population will be 8.5, 9.7, and 10.9 billion people in 2030, 2050, and 2100, respectively (United Nations, 2019). The challenge for agriculture will be to produce more with greater profitability. In this context, precision agriculture (PA) and, more recently, digital agriculture (DA) are inserted. DA uses PA technology, smart grids and data management tools. DA aims at using all available information and knowledge to enable the automation of sustainable processes in agriculture. DA made available, cheaper and more powerful sensors, actuators and microprocessors, high-bandwidth cellular communication, cloud communication, and Big Data. As a result, the information flow is no longer coming only from the used agricultural resource but also from new services offered with new algorithms that transform data into valuable intelligence (CEMA, 2017). In this new DA paradigm, large amounts of data are made available, and the challenge is to add value to them with the insertion of data portals and work platforms. At the portals, the end-user can view their data but no longer insert them manually. By platforms, this user can transform data into new and more powerful information.

The significant expansion of digital technologies in agriculture, not only for PA, led to the emergence of the new era in agriculture, called “agriculture 4.0”, driven by the movement of industry 4.0 (Zambon et al., 2019). The transition, therefore, from Agriculture 3.0 stage, characterized using PA, to Agriculture 4.0, which evolves from PA to DA, requires the provision of specific portals and platforms (CEMA, 2017). Therefore, Brazil needs to be aware of this technological evolution and provide free web platforms for integrating data, software, procedures, and methodologies for PA. The availability of the free digital platform AgDataBox (ADB) for the web contributes to the inclusion of Brazil in this phase of agriculture. The ADB-Map application creates thematic maps (TMs) and delineates friendly and integrated management zones (MZs) (Michelon et al., 2019; Borges et al., 2020; Dall'agnol et al., 2020). This platform is a continuation of the software to define management zones (SDUM – Bazzi et al. 2019b) project, which has already been registered with the National Institute of Industrial Property (INPI) (registration BR 51 2014 000720 D) and made available free to be applied.

The web microservices architecture (MSA) allows the scalability of the ADB Digital platform, allowing modules to be built with different software development technologies and later integrated. The main content of this trial is structured in papers, presented after the contextualization chapters (chapters 1 to 4):

- Paper 1 (chapter 5): MSA of ADB digital platform is presented. It provides computational routines to create TMs and delineate MZs as web services;
- Paper 2 (chapter 6): ADB-Map web application is presented. It makes it possible to create TMs and delineate MZs in a friendly way and integrated with the MSA of ADB digital platform;

- Paper 3 (chapter 7): it presents the improvements made during the interpolator selection process, a service of ADB platform. This service considers three new criteria for the semivariogram analysis: the minimum effective spatial dependence, the importance of the first semivariance, and the non-tendency of the purest nugget effect.

2 OBJECTIVES

2.1 General objectives

Integrate functionalities for thematic maps (TMs) creation and management zones (MZs) design in some microservice architecture (MSA) for the digital platform AgDataBox (ADB) and incorporate them in the ADB-Map web application.

2.2 Specific objectives

- To create an MSA web to integrate, incorporate and make available the functionalities for TMs creation and MZs delineation.
- To develop a new ADB-Map web application to create TMs and delineate MZs in a friendly, automated, and integrated way with MSA functionalities.
- To improve procedures for data interpolation and selection of the best interpolator.
- To evaluate the ADB-Map from the user's perspective.

3 LITERATURE REVIEW

3.1 Soil attributes in agriculture

Agriculture undergoes constant transformations, seeking for efficiency in soil management, during the appropriate application of inputs, in mechanized operations aiming at increasing crop yield and investment income. The agricultural ecosystem is complex and requires knowledge on different factors that influence crop yield. Some of them are controllable, such as soil fertility, cultivars adapted to local conditions, irrigation, pest, disease control, among others. On the other hand, climate actions are examples of uncontrollable factors.

Proper management of soil fertility depends on understanding the availability of nutrients in soil for plant absorption. Soil nutrients are divided into macro and micronutrients. Macronutrients are required in greater amounts by plants and may be of organic origin such as carbon (C), oxygen (O), and hydrogen (H), or of chemical origins, such as nitrogen (N), phosphorus (P), potassium (K), sulfur (S), calcium (Ca), and magnesium (Mg). Plants consume micronutrients in smaller amounts, but they also play an essential role along crop growth and development. They are boron (B), chlorine (Cl), copper (Cu), iron (Fe), manganese (Mn), molybdenum (Mo), and zinc (Zn) (Mendes, 2007). Although, macronutrients generally become deficient in soil before the others due to higher consumption made by plants (POTAFOS, 1998).

Nitrogen and potassium are more extracted by soybeans. For example, part of the first one is provided by the soil (25 to 35%) and part by symbiotic fixation of atmospheric N_2 (65 to 85%) (Borkert et al., 1994). Although phosphorus is the least extracted among the three primary macronutrients, it is usually the nutrient used in greater amount, either because of its low content in soil or because of its dynamics in tropical soils (fixation) (Vitti and Trevisan, 2000). The frequent limitation of corn yield is related, in part, to the low availability of calcium and phosphorus in most Brazilian soils (Coutinho et al., 1991). Therefore, Liming and phosphate fertilization are requirements to improve both cropping development and production (Santos et al., 2006).

The evaluation of soil chemical fertility is helpful to define the quantities and types of fertilizers, correctives, and general management that should be applied to the soil to keep or recover its yield (Ronquim, 2010). Fertilization begins with soil analysis, continues with the correction of acidity, and ends with the correct application of fertilizer (Malavolta, 1992). Nutrients' uptake is done with chemical mineral fertilizers, organic matter, minerals taken from deposits or the air (in biological nitrogen fixation). Organic matter provides an increase in soil fertility, as it practically contains macro and micronutrients and provides a better structure to soil. Mineral fertilizers (which are different from organic matter) have nutrients in high concentrations that are highly soluble and can be quickly absorbed by plants and/or leached more easily (Ronquim, 2010).

Liming is the best known and most used practice for correcting arable soil acidity and supplying Ca and Mg nutrients (Souza and Lobato, 1986). Soil acidity is a limiting factor for agricultural production due to toxicity caused by Aluminum (Al), limiting plant roots' growth in naturally acidic soils (Coleman and Thomas, 1967). It is mainly caused by the intense weathering that occurs over time, a phenomenon that causes leaching and removal of basic cations from cation exchange capacity (CTC) of soil, mainly Ca, Mg, K, and sodium (Na), which give place in CTC to exchangeable aluminum and undissociated hydrogen (Van Raij, 2011). In acidic soils, or even those with low aluminum saturation, adequate root development does not occur in most cropping plants, making it difficult to absorb water and nutrients, and causing yield decrease (Souza and Ritchey, 1986). Toxicity caused by Al and Mn is reduced with liming by increasing Ca and Mg contents. In addition, liming also promotes an increase in CTC and N, P, S, and Mo availability (Bernardi et al., 2003).

The physical attributes of a soil, on the other hand, are directly related to infiltration capacity, retention, and availability of water for crops, as well as on making easy air circulation and root development capacity (Beuttler et al., 2012; Carvalho et al., 2012). Furthermore, studies related to soil compaction, which uses its density as an indicator attribute, have pointed out that its increase decreases agricultural yield (Abreu, et al., 2003; Faraco et al., 2008; Reichert et al., 2009).

Another essential feature of soil is its texture, which refers to the mineral particles sizes that compose it (Oliveira Junior et al., 2010). Soil texture is determined by percetage presence of sand, silt, and clay on it. Clay textured soils is characterized due to their smaller particle size, while sandy textured soils present larger particle size (POTAFOS, 1998). Therefore, there are several interrelationships among physical, chemical, and biological soil attributes (Carneiro et al., 2009). The practices applied to manage soil can cause changes in its physical, chemical, and biological attributes (Niero et al., 2010). Any change in soil management can directly change its structure and biological activity, consequently, its fertility, damaging its quality and crop yield (Carneiro et al., 2009).

3.2 Precision Agriculture

Precision agriculture (PA) is a cropping management system based on crop and soil characteristics' spatial and temporal variability within a field (Stafford, 2000). The International Society of Precision Agriculture (ISPA, 2019) defines PA as “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 some estimated variability to improve resource use efficiency, productivity, quality, profitability and sustainability of agricultural production”. However, it is not a new concept, as farmers have realized it since the early days of agriculture. The farmers divided the property into smaller areas to seek for the most suitable conditions for growing crops, as they knew the soil characteristics. For them, precision was about ensuring enough food to sustain the family (Oliver, 2010). What we know as current PA practices have been in use since the 1980s (Zhang et al., 2002), has attracted worldwide interest commercially (CAMBOURIS et al., 2014), and available since the early 1990s (McBride and Daberkow, 2003).

PA only gained more prominence with the spread of global positioning system technologies (GPS), currently called global navigation satellite systems (GNSS) (standard generic term for satellite navigation systems), geographic information system (GIS), remote sensing technologies, and different sensors to evaluate location variability and crop characteristics. These technologies have provided information to assist producers in a more accurate agricultural system management (Pathak et al., 2019). Thus, PA is an agricultural management system that includes information technologies and attempts to provide amounts and types of inputs based on actual cropping needs in small fields inside a large farm (e.g., variable-rate application (VRA), yield monitors, remote sensing) (Bongiovanni and Lowenberg-Deboer, 2004).

The PA approach can be summarized in a three-step process (Fig. 1). The first step is to identify the variability present in a field (Cambouris et al., 2014); spatial and temporal variabilities are found out in many attributes that influence cropping yield (Schepers et al., 2004); for example, spatial variability of P and K can influence corn yield (Molin et al., 2007). Bottega et al. (2013) study has recorded some spatial and temporal dependence on soybean yield.

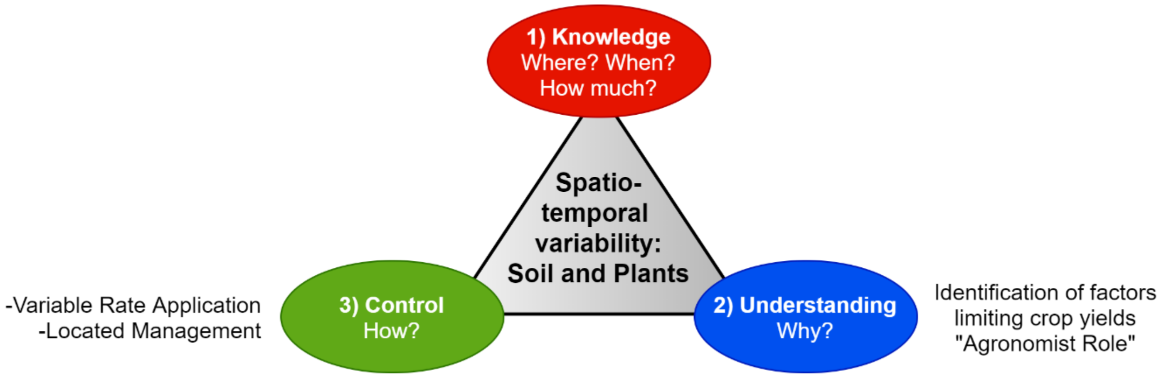


Fig. 1 Approach of precision agriculture in three steps, which involves knowledge, understanding and control of variability.
Source: Cambouris et al. (2014).

The second step is to analyze variability within the field to make a management decision. In this stage, the specialist has a fundamental role, working with the farmer, who knows well field conditions (Cambouris et al., 2014). This decision-making involves inputs and how to use them at variable rates (Salehi and Rezaei-Moghaddam, 2008). The third step involves managing variability within the field in the most effective way to increase yield and profitability while minimizing environmental risks (Cambouris et al., 2014; Mondal and Basu, 2009). Some of the expected benefits are cost reduction, such as applying inputs where they are needed and improving water resources management (Mintert et al., 2016). VRA is a technique used to manage field variability, which requires specialized agricultural equipment, and another technique for managing field variability is by MZs.

Barnes et al. (2019) observed in their research that technologies applied in PA are categorized as guidance, recording, and reacting (Fig. 2). Guidance technologies provide precise guidance and control to carry out field operations. These technologies are dependent on GNSS for their applicability. Driver assistance technologies aim at relieving farmers from the physical workload or time spent in the field as well as simplify and optimize processes (e.g., automatic steering systems) (Groher et al., 2020).

Data acquisition in PA can be performed in different ways, such as sample collection in a georeferenced sampling grid and subsequent analyses in laboratory (Molin and Tavares, 2019), use of field sensors according to the proximal soil sensing concept (PSS, defined by Viscarra Rossel et al. (2011), and remote sensing technologies that are alternatives to using sensors in agriculture (Groher et al., 2020). Sensors can be more cost-effective than conventional laboratory chemistry analyses, compact, faster, more accurate and energy-efficient, wireless, and smarter (Viscarra Rossel et al., 2011).

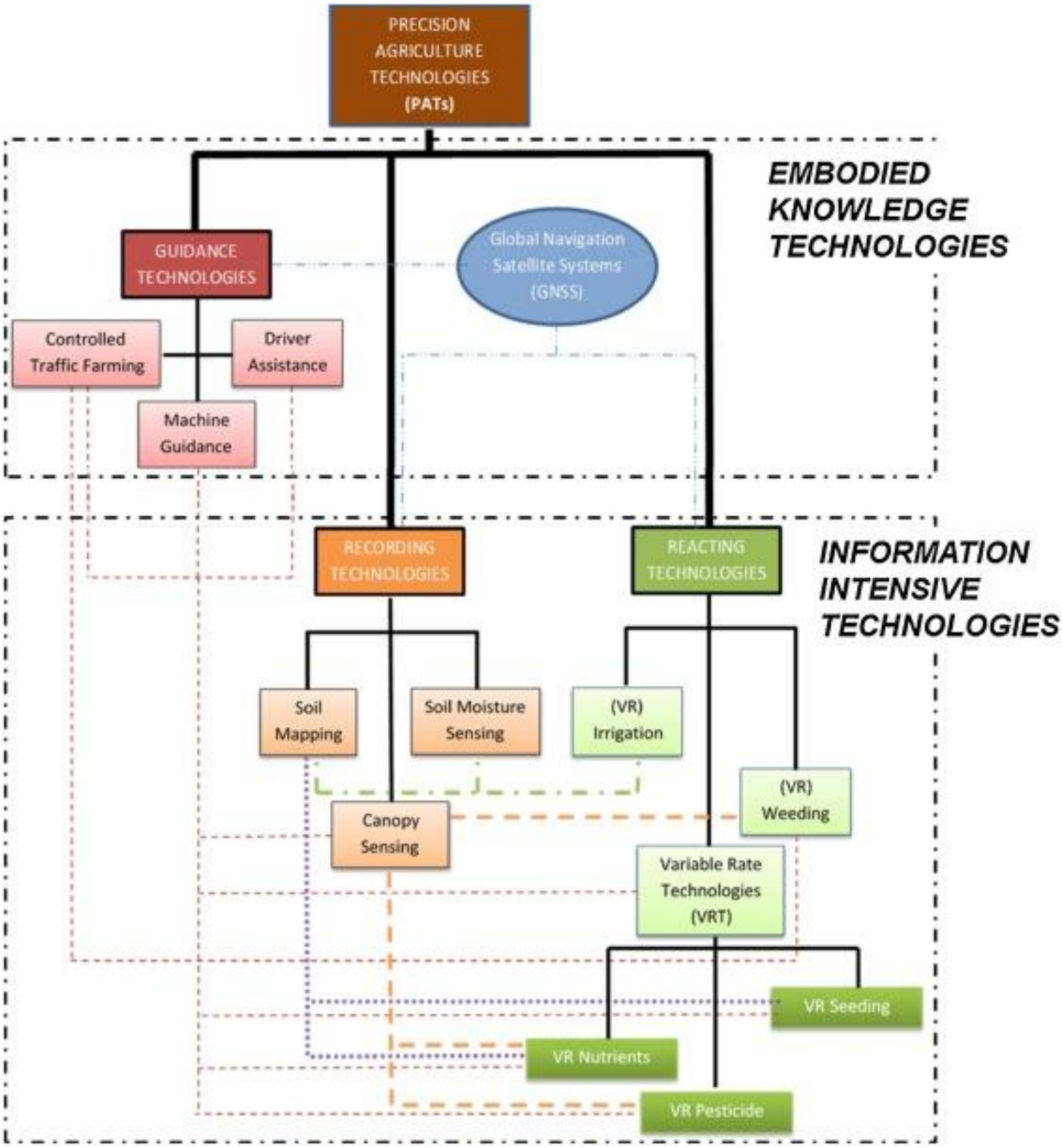


Fig. 2 Hierarchy of Precision Agricultural technologies.
Source: Barnes et al. (2019).

3.3 Digital agriculture

Agricultural field knowledgement is an essential factor for business success. Farmers who want to keep up with the competitive market and increase their profitability need suitable digital technological tools to provide the correct information for decision making that positively impacts on business. The challenge is to produce more in less space and with available use of natural resources. DA is also referred to as an intelligent farming and agriculture 4.0, and it is the fulfillment of "digital land" concept proposed in the 1990s, which is characterized as an advance in PA with emphasis on agricultural production procedures (Shen et al., 2010). The introduction of advanced biotechnologies and digital tools seeks to offer new strategies to feed the growing world population (Fraser and Campbell, 2019).

Based on this concept, digital technologies such as artificial intelligence, robotics, Big Data, and IoT (Alm et al., 2016) are adopted to manage processes within the farm. Broson (2018) points out that Big Data, PA, and automation are considered the primary digital artifacts responsible for technological changes in agriculture. The new revolution in agriculture, called “agriculture 5.0”, has as principle associating production with planet's health (Fraser and Campbell, 2019), thus, DA is an essential ally in supplying this demand.

Data collection and use have influenced decision-making in agriculture to become increasingly important (Pham and Stack, 2018). Big Data concept is related to a large volume of collected data, usually from digital sources. At DA, large volumes of data can be generated from satellites and unmanned aerial vehicles equipped with multispectral cameras, IoT with different types of sensors, mobile device applications, software computers, among others. Data is stored in computer databases and subsequently analyzed by algorithms specialized in extracting information, and agriculture is designed to be the next big data industry. Agricultural machines are equipped with sensors and cameras to capture detailed data at the field level, such as soil moisture, plants, temperature, seeding, fertilizer and herbicide spray rate, yield, fuel usage, and machine performance (Pham and Stack, 2018). Despite the ability to obtain different types of data and large volumes, it is necessary to go through validation. Molin et al. (2020) point out that the challenge is establishing a consistent set of variables that guarantee robust and satisfactory results for all techniques.

Digital technologies are essential for several actors in the agricultural ecosystem, such as scientists, to ease interaction among analysts and farmers. Farmers have the potential to provide large amounts of valuable data about their activities and experiences (Eitzinger et al., 2019). Much of PA data can be acquired passively, directly from agricultural equipment instrumented with sensors for data collection and GNSS to identify geographic positioning. However, as Big Data and PA are linked to automation, machines can do the job “intelligently,” performing semi-autonomous and learning actions over time (Broson, 2018). Decision-making is based on computational algorithms that interpret large sets of agricultural data and generate useful information and insights for PA equipment to perform pre-programmed actions (Wolfert et al., 2017).

The fast expansion of Internet-enabled devices has led to IoT occurence. IoT is among the top strategic technology trends for 2021, within the context of carrying out operations anywhere (Burke, 2020). IoT-enabled Agricultural (IoTAg) is growing quickly. IoT technology allows monitoring and controlling crop parameters by sensors and devices to obtain food quality in quantity (Uddin et al., 2017). IoTAg segment is projected to reach $4.5 billion by 2025, according to PwC (Columbus, 2021). There are several researches developed with IoT application in agriculture (Akhter and Sofi (2021); Nóbrega et al., 2018; Taneja et al., 2018; Mohanraj et al., 2016; Ryu et al., 2015).

In this growing environment of devices connected to Internet and integrated with intelligent systems that can share, process, store, and analyze data with each other, the result is an extensive network of cooperation, in which a large volume of data is generated. The challenge for DA is to create systems for data management in the agricultural sector, which range from environmental, soil, plant, production, and market conditions. Data analysis techniques and intelligence algorithms can provide insights into business and assist farmers with their decision-making, whether for managing production or marketing the product. Studies are reporting the software development with different applicability within DA context:

- Precision Agriculture Methodologies for Cost-Benefit Analysis (PAMCoBA; Medici et al., 2021) is a web tool designed to provide guidelines for farmers over their decisions to invest in selected precision agricultural Technologies. This tool explores data regarding existing PA technologies, crops, and agricultural operations, guiding farmers on selecting the most appropriate technologies for farm-specific context.
- GeoFarmer (Eitzinger et al., 2019) is a platform for sharing experiences among farmers and specialists. It aims at sharing knowledge for better management of crops and farms.
- CLUeFARM (Colezea et al., 2018) is a hybrid platform with local and cloud computing that provides resources for farmers to manage their farms, such as data analysis, IoT integration, and external data sources.
- Smart platform to help farmers manage their greenhouses and interact with other farmers (Musat et al., 2017). The platform is based on cloud computing and manages data models, data analysis, and provision of services.
- AEGIS (Shen et al., 2010) is a Canadian DA system for agricultural risk management, which can help assessing risks due to climate change, developing a revenue protection plan for producers, and generating a soil quality management plan.
- Farm management information systems (FMIS; Kaloxylos et al., 2014) assist producers with an efficient management, enabling decision-making based on data collection and processing. There are similar trials in this area (Paraforos et al., 2017; Ampatzidis et al., 2016; Kaloxylos et al., 2014).

3.4 Management zones

Delineating management zones (MZs) in an agricultural field is one of the ways to practice PA. An MZ is a subregion of a field that expresses a functionally homogeneous combination of yield-limiting factors for which a single rate of a specific crop input is appropriate (Doerge, 2000; Moral et al., 2010; Moshia et al., 2014; Bobryk et al., 2016). Although MZs may be used with localized application machinery, their use is more common with conventional machinery. After MZs are delineated, they may also be used in a smart sampling, in which the number of samples needed to delineate soil variability in the field may be reduced to a sample composed per zone. This approach (smart sampling) can reduce laboratory costs and, at the same time, keep reliability level (Ferguson and Hergert, 2009; Mallarino and Wittry, 2004) and improve nutrient use efficiency, maintaining or increasing yield and potentially reducing the nutrients overload on the environment (Moshia et al., 2014; Khosla et al., 2002).

Despite the original concept that an MZ is a subregion of a field that expresses a functionally homogeneous combination of limiting production factors, the target agricultural variable may be other than yields, such as infestation by pests and diseases, water content, Brix value, and soil resistance to penetration (SRP). An MZ may be used for one or several years, usually from three to five years. This fact is significant when choosing variables. If the plan is to use it only once, which might be the case of a weed infestation, the farmer may use variables that are not constant over time (such as weed infestation) to delineate MZs. However, in most cases, someone may wish to use the MZs for several years and, thus, must use variables relatively constant over time such as soil properties (Carvalho et al., 2016), soil apparent electrical conductivity (Castrignanò et al., 2018; Martínez-Casasnova et al., 2018), farmer experience (Martínez-Casasnova, 2018; Schenatto et al., 2017a), crop yield (Blackmore, 2000), topographic attributes and soil electrical conductivity (Peralta et al. 2015; Fraisse et al., 2001), satellite images (Damian et al., 2020; Breunig et al., 2020; Zhang et al. 2010) or a combination of multiple layers of data (Schepers et al. 2004; Betzek et al., 2018; Betzek et al., 2019; Gavioli et al., 2019).

Among the approaches presented in literature to delineate MZs using yield maps, two stand out (Xiang et al., 2007): 1) the empirical method, which uses the frequency distribution of yield and specialized knowledge to divide the field usually into three or four zones (Blackmore, 2000; Molin, 2002); and 2) cluster analysis, such as K-means and Fuzzy C-Means (FCM) (Taylor et al., 2003; Taylor et al., 2007; Yan et al., 2007). Although empirical classification methods are simpler, cluster analysis allows a greater degree of differentiation among MZs. Empirical methods are used primarily when the target variable (usually yield) is used in MZs delineation. Molin (2002) used corn, soybean, and wheat yield to delineate MZ by empirical methods. When using attributes correlated to the target variable to create MZs, it is possible to generally use clustering methods.

Some protocols have already been proposed (Santos and Saraiva, 2015; Cordoba et al., 2016; Souza et al., 2018) to delineate MZs properly. In the protocol organized by Souza et al. (2018), the process to delineate MZs follows these phases: (i) data processing, (ii) data normalization, (iii) selection of variables to delineate MZs, (iv) data interpolation, (v) application of methods to delineate MZs, (vi) MZs rectification and (vii) MZs evaluation (Fig. 3).

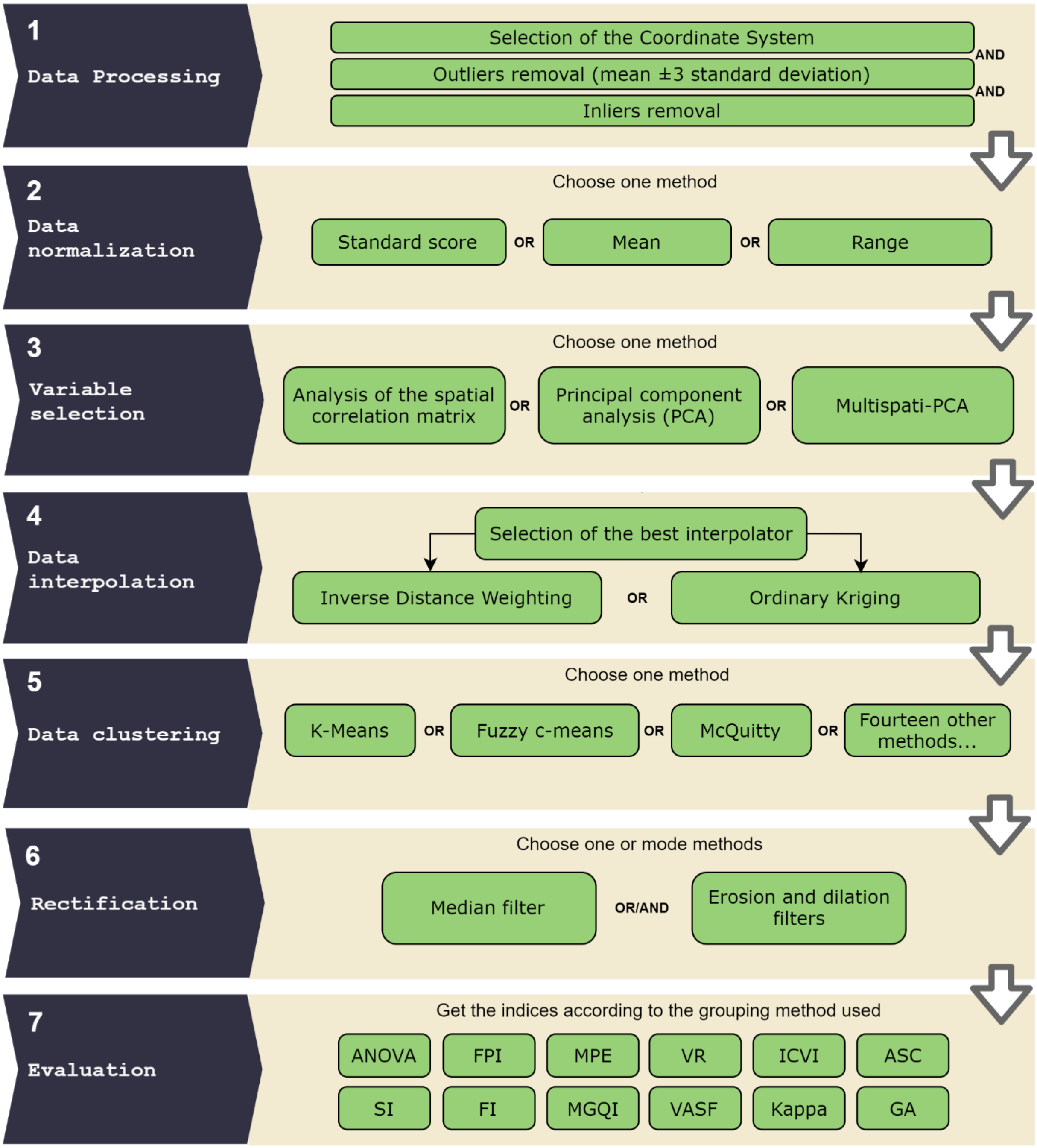


Fig. 3 Protocol steps for management zones Delineation.
Source: Adapted from Souza et al. (2018).

3. 4. 1 Data processing

A GIS is a system that creates, manages, analyzes, and maps all kinds of data. A GIS software and a file with at least three columns are used to build 2D TMs, representing X (longitude) and Y (latitude) coordinates and the measured attribute value. For 3D TMs, one more coordinate is needed, Z (altitude). These coordinates are associated with a coordinate system, and the most typical coordinate systems are the geographic and universal transverse Mercator (UTM). After data are uploaded to GIS, it is necessary an exploratory analysis. The exploratory analysis employs various techniques (mainly graphical) to maximize data set perception as well as to detect and remove outliers. For example, according to Córdoba et al. (2016), values outside the mean ± three standard deviations are identified as outliers and must be removed.

Data that differ significantly from their neighborhood but are within the general variation range of data set are called inliers (Cordoba et al., 2016) and should be removed using, for example, the local Moran's index (I_i, Anselin, 1995; Levine, 2004, Equation 1). While the global Moran's index quantifies a spatial autocorrelation as a whole, local indicators of spatial association (LISA) measure the spatial autocorrelation degree at each specific location (Anselin, 1995).

$$I_i = \frac{z_i - \bar{z}}{\sigma^2} \sum_{j=1, j \neq i}^{n} \left[W_{ij} \left(z_j - \bar{z} \right) \right], \tag{1}$$

where, $\bar{z}$ is the mean value of z with the number of samples of n; z_i is the value of the variable at location i; z_j is the value at other locations ($j \neq i$); σ^2 is the variance of z; and W_{ij} is the weighted distance between z_i and z_j, which can be defined as the inverse of the distance.

Spatial weights must be defined to calculate Moran's index, and spatial relationships of data are defined by the spatial proximity matrix ($\mathbf{W}$). This is an *n* x *n*-dimensional symmetric matrix. The connection degree among regions *i* and *j*, represented by weights w_{ij}, is defined by some proximity criterion that indicates the influence of one region over another (Almeida, 2012). Greater weight is attributed to geographically closer regions, and lesser weight is given to more distant regions. Thus, the most common criteria for assigning values to each matrix element are distance, contiguity, and neighborhood.

3. 4. 2 Normalization methods

The data clustering techniques using FCM algorithm are the most broadly employed processes to define MZs. In this process, it is necessary to choose a similarity measure, and the most used is the Euclidean distance. However, with this distance, the algorithm is sensitive to the interval of input variables, requiring their normalization, which may be done by dividing the value of each variable by the maximum value, mean, or sum of observations.

Schenatto et al. (2017b) analyzed the influence of data normalization methods for defining MZs. The tests were conducted in three experimental fields with 9.9, 15.0, and 19.8 ha, in Southern Brazil. The variables (attributes) used to define MZs were selected using Moran's bivariate spatial autocorrelation statistic, and data were normalized using range (Equation 2), mean (Equation 3), and standard score (Equation 4) methods. MZs were defined using FCM algorithm, which generated clusters with two, three, and four classes. It was proven that normalization is necessary when MZs definition uses more than one variable during the clustering process, and similarity measure is the Euclidean distance. The range method was considered the best normalization method.

- Range (Equation 2): it is based on dataset range it is necessary to normalize. According to Anderberg (1973) and Milligan and Cooper (1988), it is not indicated when there are outliers in the data. Some changes to this method involve the numerator, with Milligan and Cooper (1988) presenting, on the numerator, only the value of datum P_i or P_i-$P_{i_{minimum}}$, in which case, normalized data will be among values from zero to one.

$$Z_{iN} = \frac{X_i - Median}{Max(X) - Min(X)}, \tag{2}$$

  where, Z_{iN} – normalized point i; X_i – original data value i; $Min(X)$ – minimum value of dataset; $Max(X)$ – maximum value of dataset.

- Mean (Equation 3 – Swindel, 1997): it is well known and employed, hoping that the means represent dataset well. However, for Anderberg (1973), mean value is sensitive and may be altered by adding any constant, thus, easily modifying the normalized data distribution.

$$Z_{iN} = \frac{X_i}{\bar{X}}, \tag{3}$$

  where, Z_{iN} – normalized observation i; X_i – original data value i; $\bar{X}$ – arithmetic mean of all map pixels or of sample set to be normalized.

- Standard Score or Z-Score (Equation 4 – Larscheid and Blackmore, 1996): is used for transforming normal variables to standard score where the transformed variable will have a mean of 0.0 and a variance of 1.0.

$$Z_{iN} = \frac{X_i - \bar{X}}{s}, \tag{4}$$

  where Z_{iN} – normalized observation i; X_i – original data value i; $\bar{X}$ – arithmetic mean of all map pixels or of sample set to be normalized; s - standard deviation.

- Min-Max (Equation 5 – Milligan and Cooper, 1988): it is a variation of the range method containing changes in the numerator, and they present in the numerator, in which case the normalization result will be among values 0 and 1.

$$Z_{iN} = \frac{X_i - Min(X)}{Max(X) - Min(X)}, \tag{5}$$

  where, Z_{iN} – normalized observation i; X_i – original data value i; $Min(X)$ – minimum value of dataset; $Max(X)$ – maximum value of dataset.

3. 4. 3 Variable selection

The weighting and selection of variables are complex tasks in cluster analysis. The capacity of clustering software to process many variables encourages users to employ many variables in this process. However, we must be aware that the choice of variables and weights attributed to them often influences clusters' determination (Gnanadesikan et al., 1995). Three variable selection techniques that may be applied in combination with FCM algorithm are (i) spatial correlation analysis (Reich, 2008; Schepers et al., 2004), (ii) principal component analysis (PCA) (Hotelling, 1933; Jolliffe, 2011), used by Fraisse et al. (2001), Li et al. (2007), Moral et al. (2010), and Cohen et al. (2013), and (iii) multivariate spatial analysis based on Moran’s index and PCA (MULTISPATI-PCA) (Dray et al., 2008), applied by Córdoba et al. (2013, 2016) and Peralta et al. (2015).

Gavioli et al. (2016) studied the efficiency of each of these three techniques and proposed a new method, named MPCA-SC, based on the combined use of Moran's bivariate spatial autocorrelation statistic and MULTISPATI-PCA. This evaluation was carried out from data collected from 2010 to 2014 in three agricultural areas of Paraná state, Brazil, with corn and soybean, generating two, three, and four classes. The MZs delineated were different according to the method used, with MPCA-SC method providing the best performance.

3. 4. 4 Data interpolation

Sample data is generally interpolated in a dense and regular grid to generate TMs and MZs that are continuous and smooth. This task is carried out with the aid of interpolation methods. The interpolation methods most used in precision agriculture (PA) are Ordinary Kriging (OK) and inverse distance weighted (IDW), which are differentiated by how weights are attributed to the different samples, which may influence the estimated values (Reza et al., 2010).

3. 4. 4. 1 Geostatistical Analysis

The semivariogram chart (Fig. 4) is determined from observed values in two stages (Oliver and Webster, 2015). The first stage is to calculate the empirical semivariogram, which summarizes spatial relations in data. Semivariances are calculated from an estimator, as the classic proposed by Matheron (1963). Each calculated semivariance for a particular lag (h) is only an estimate of a mean semivariance $\hat{\gamma}(h)$ for that lag. As such, it is subject to error.

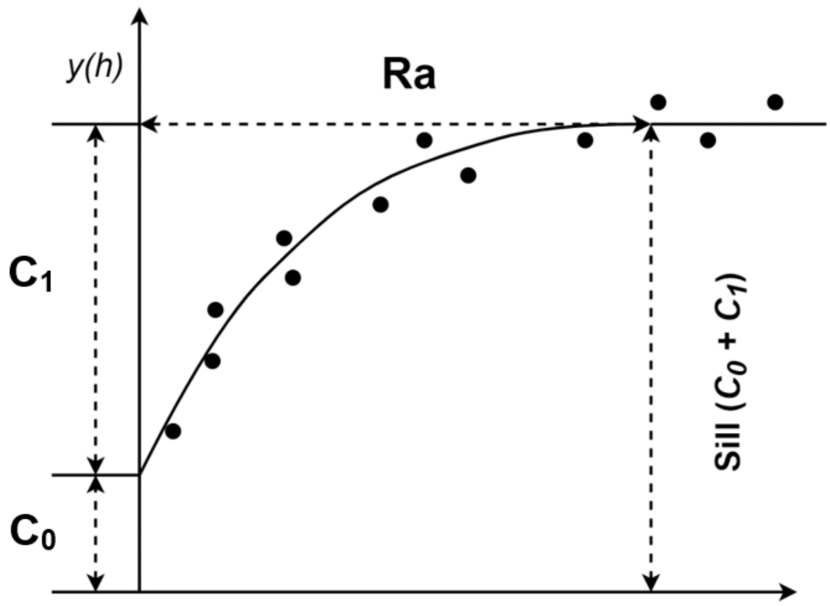


Fig. 4 Semivariogram chart with the four main elements: nugget effect (C_0), partial sill (C_1), sill ($C_0 + C_1$), and the range of spatial autocorrelation (Ra).

The second stage is adjusting a mathematical model that best represents the distribution of semivariances in each lag distance. This mathematical model should describe the spatial variation to estimate or predict values at unsampled places by kriging optimally. Therefore, the geostatistical analysis is carried out in two moments: the semivariogram analysis and interpolation by Kriging (Oliver and Webster, 2015). Only some mathematical functions are suitable for this purpose and choosing and fitting a model must be done carefully (Lark, 2000). Once we calculate an experimental variogram, we can fit it using several variogram models, such as spherical, exponential, Gaussian, and Matérn (Isaaks and Srivastava, 1989, Gamero et al., 2020).

It consists of main parameters: the nugget effect (C_0), the partial sill (C_1), the sill (C_0 + C_1), and the range (Ra), adjusted according to the spatial data variation characteristics (Matheron, 1963):

- Nugget effect (C_0): is the semivariance value for zero distance (Webster, 1985) and represents the component of random variation, i.e., variability for scales smaller than the distance among sample points. According to Cressie (1993), C_0 represents local small-scale variations, such as measurement errors. It corresponds to the point where the semivariogram touches the ordinate axis. This point reveals the semivariogram discontinuity for distances that are closer than the closest distance among the sampling points.
- Partial sill (C_1): represents the spatial differences among C_0 values and plateau, an interval in which the semivariogram grows, representing spatial dependence (Cressie, 1993). Thus, it is also known as dispersion variance.

- Range (Ra): is the distance where the variogram reaches sill, and from this distance, samples are not correlated. Places further apart than this are spatially independent (Oliver and Webster, 2015).

Semivariance tends to increase with the distance among sample locations, or lag distance (h) to a possible constant value (sill) at a given separation distance, called spatial dependence range. Samples distanced greater than the range are not spatially related (Webster, 1985; Cambardela et al., 1994).

The estimator based on method-of-moments proposed by Matheron (1963) is defined in Equation 6 and is unbiased for actual theoretical values. It is the average of the squared differences among observations separated by the distance h.

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2, \tag{6}$$

where $\hat{\gamma}(h)$ is the value of semivariance estimate; $Z(x_i)$ is the value of variable Z at point x_i; $Z(x_i + h)$ is the value of variable Z at point $x_i + h$; $N(h)$ is the number of pairs separated by a determined distance h.

Journel and Huijbregts (1978) suggest the definition of lag increment and the number of lags as at least 30 pairs of points, and the range Ra was limited to half of the maximum distance among points (cutoff = 50%). The semivariances' calculation should not exceed distances among points greater than half of maximum distance (Clark, 1979). Points located beyond cutoff are considered non-influential (Isaaks and Srivastava, 1989).

The spatial dependence index (%SDI – Biondi et al., 1994 – Equation 7) can be used to evaluate SD degree of a variable using semivariograms. The adopted %SDI classification (Konopatzki et al., 2012) was: very low for %SDI < 20%; low for 20 ≤ %SDI < 40%; medium for 40 ≤ %SDI < 60%; high for 60 ≤ %SDI < 80%; and very high for %SDI > 80%. This classification has the advantage of having five interpretation levels instead of three proposed by Cambardela et al. (1994) and is proportional to the spatial variability (the highest %SDI, the highest SD).

$$\%SDI = \frac{C_1}{C_0 + C_1} * 100 = \frac{C_1}{C} * 100, \tag{7}$$

where C_0 is the nugget effect, C_1 is the partial sill, and C ($C_0 + C_1$) is the sill.

3.4.4.1.1 Semivariogram models

Variogram modeling is a crucial stage of geostatistical analysis. Knowing semivariances values for any distances (vector h) is necessary for further interpolation using the kriging method (Borkowski and Kwiatkowska-Malina, 2017). After estimating the experimental semivariogram, it is required to fit a smooth curve to the experimental values to describe the sequence's principal features, which is made by a mathematical expression that can describe variances of random processes (Oliver and Webster, 2015). Among the existing theoretical semivariogram models, the most used to explain most spatial phenomena are spherical, exponential, Gaussian, and Matérn (Gamero et al., 2020) (Fig. 5).

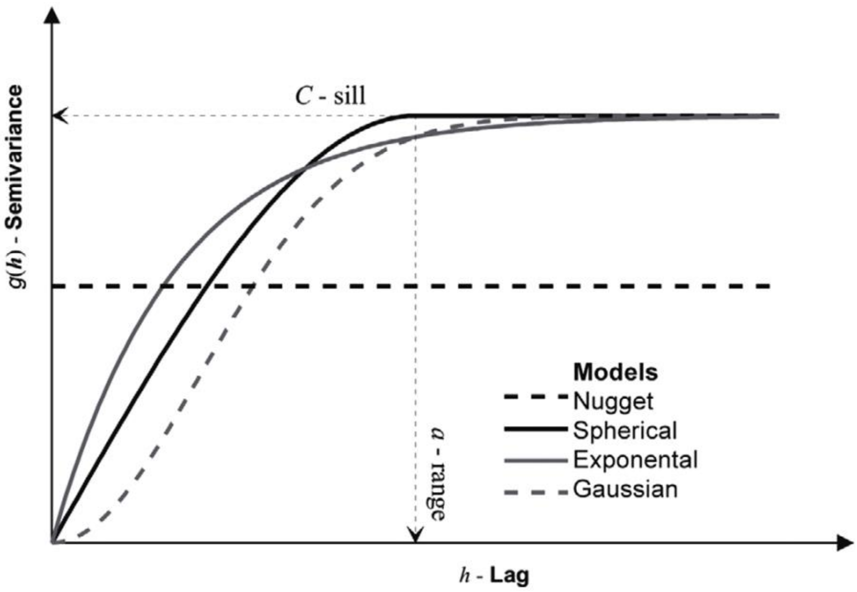


Fig. 5 The most common variogram models: nugget, spherical, exponential, and Gaussian. Source: Stach (2007).

a) Spherical model (Equation 8): This model behaves linearly at small separation distances near the origin, but lies flat at longer distances and sills at a (Isaaks and Srivastava, 1989).

$$\gamma(h) = \begin{cases} C_0 + C_1\left[1.5\frac{h}{a} - 0.5\left(\frac{h}{a}\right)^3\right], & if\ h < a \\ C_0 + C_1, & if\ h \geq a. \end{cases} \tag{8}$$

where a is range; h is distance; C_0 is the nugget effect; and $C_1$is the partial sill.

b) Exponential model (Equation 9): This function approaches its sill asymptotically and does not have a finite interval (Oliver and Webster, 2015). For practical purposes, with practical range defined as that distance, sill's variogram value is 95%. The model is linear at very short distances near the origin. However, it rises more steeply and then flattens out more gradually (Isaaks and Srivastava, 1989).

$$\gamma(h) = \begin{cases} 0, & if\ h = 0 \\ C_0 + C_1\left[1 - exp\left(-\frac{h}{a}\right)\right], & if\ h > 0. \end{cases} \tag{9}$$

where a is range; h is distance; C_0 is the nugget effect; and $C_1$is the partial sill.

c) Gaussian model (Equation 10): reaches its sill asymptotically, and parameter a is defined as the practice range or distance at which the variogram value is 95% of sill (Isaaks and Srivastava, 1989). The Gaussian model represents a slow increase of variogram values near the origin, for example, a parabolic behavior (Azevedo and Soares, 2017). The range (a) is the distance from where the model reaches 95% of sill.

$$\gamma(h) = \begin{cases} 0, & if\ h = 0 \\ C_0 + C_1\left[1 - exp\left(-\left(\frac{h}{a}\right)^2\right)\right], & if\ h > 0. \end{cases} \tag{10}$$

where a is range; h is distance; C_0 is the nugget effect; and C_1 is the partial sill.

d) Matérn model (Fig. 6, Equation 11): Matérn's family of models is a generalization of other theoretical models, and its fundamental characteristic is the inclusion of a parameter (κ) that determines smoothing (Minasny and McBratney, 2005). For κ = 0.5, the Matérn model is equivalent to the exponential model and, for κ tending to infinity, it is equivalent to the Gaussian model (Uribe-Opazo et al., 2012). This model has excellent flexibility for modeling spatial covariance and can model many local spatial processes to balance both extremes. Thus, it can be used as a general model of soil variation (Minasny and McBratney, 2005).

$$\gamma(h) = \begin{cases} 0, & if\ h = 0 \\ C_0 + C_1\left[1 - \frac{2}{\Gamma(\kappa)}\left(\frac{h\sqrt{\kappa}}{a}\right)^{\kappa} Bk\left(\frac{2h\sqrt{\kappa}}{a}\right)\right], & if\ h > 0. \end{cases} \tag{11}$$

where a is range; h is distance; C_0 is the nugget effect; C_1 is the partial sill; Bk is the Bessel function of order κ, $\Gamma(\kappa)$ is the Gamma function and κ is the smoothness parameter.

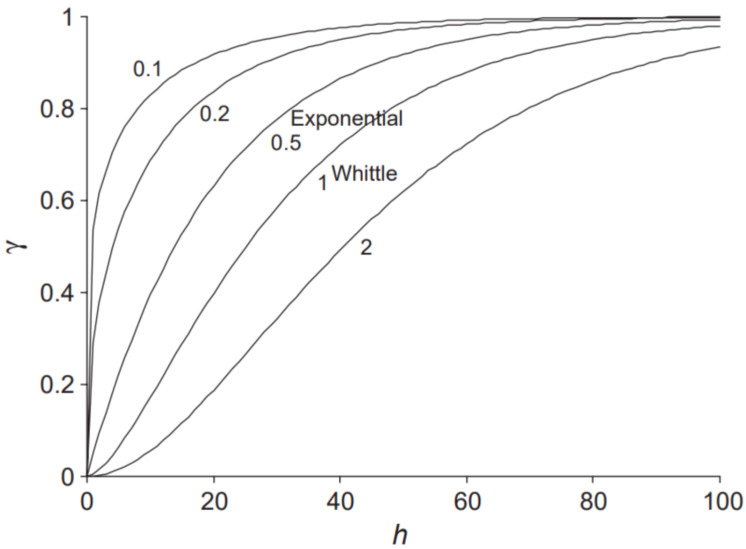


Fig. 6 Plots of the Matérn variogram with varying smoothness parameters.
Source: Minasny and McBratney (2005).

e) Pure Nugget effect model (Equation 12): The lack of spatial continuity in data set is demonstrated to increase the nugget effect, making the estimation procedure as a simple averaging of the available data (Isaaks and Srivastava, 1989).

$$\gamma(h) = \begin{cases} 0, & if\ h = 0 \\ C_0, & if\ h > 0. \end{cases} \tag{12}$$

where a is range; h is distance; and C_0 is the nugget effect.

3. 4. 4. 2 Ordinary Kriging

Interpolation by OK (Cressie, 1993) is done after adjusting the semivariogram model, and the value to be estimated at the point of interest can be calculated by Equation 13. Kriging makes estimation based on a continuous model of stochastic spatial variation. Thus, it makes the best use of existing knowledge by considering how a property varies in space by the variogram model (Oliver and Webster, 2015).

$$\hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i * \mathrm{Z}\,(x_i), \qquad (13)$$

where $\hat{Z}(x_0)$ – estimated value at a given location; λ_i – weight attributed to the sample values; $Z(x_i)$ – sampled attribute value; n – number of neighboring locations employed for interpolating the point, where $\sum_{i=1}^{n} \lambda_i = 1$.

Interpolation by kriging has weights determined from spatial analysis, based on experimental semivariogram (Cressie, 1993). Weights are determined by statistical dependence (i.e., covariances) among sampled locations yet they respect some measurement uncertainty. In general, the greater the covariability, the greater the weight (Wikle et al., 2019).

Kriging has been identified as an interpolator with better performance over other interpolators, as it is based on the estimator's unbiasedness and the interpolator's minimum variance (Diggle and Ribeiro, 2007, Vieira, 2000). However, to have the correct performance and proper use in creating the TM, it is necessary to meet the requirements of spatial dependence modeling (Oliver and Webster, 2015). The procedure performance can be influenced by variability and spatial structure of data, the semivariogram model, the search radius, and the number of the closest neighboring points used (Isaaks and Sriivastava, 1989; Reza et al., 2010).

3. 4. 4. 3 Inverse distance weighting

IDW interpolator (Equation 14) is a deterministic estimator that considers sample point weights evaluated during the interpolation process. Thus, the influence of each sampled point is inversely proportional to the distance raised to the power of the point to be estimated (Isaaks and Sriivastava, 1989).

$$\hat{\mathrm{Z}}_i = \frac{\sum_{i=1}^{n}\left(\frac{1}{d_i^p} * Z_i\right)}{\sum_{i=1}^{n}\left(\frac{1}{d_i^p}\right)}, \qquad (14)$$

where $\hat{\mathrm{Z}}_i$ – interpolated value; Z_i – sampled attribute value; d_i^p – Euclidean distance between the ith neighborhood point and the sampled point, elevated to the power of p > 0.

It is a purely mathematical process in which data is weighted so that the influence among them decreases as distance increases. The chosen power value predetermines the weight factor; that is, the higher this value, the lesser the influence of the farthest points. This method is fast and requires little computational cost (Mazzini and Schettini, 2009).

3. 4. 5 Determination of the best semivariogram model and its parameters

Bier and Souza (2017) proposed the interpolation selection index (ISI – Equation 15) to automatize selecting the best interpolation method, which assumes a lower value as better the interpolator is. Betzek et al. (2019) developed computational routines in geoR to determine the best semivariogram model (and its parameters) and the best power to be used in IDW interpolator using ISI. In geostatistics module, six semivariogram models are tested (spherical, Gaussian, exponential, Matérn 1.0, Matérn 1.5, and Matérn 2.0), as well as two statistical methods to optimize the semivariogram adjustment, ordinary least squares (OLS) and weighted least squares (WLS – Cressie, 1985), thus totalize twelve different models. For each model, 25 different parameter sets (five initial values for the contribution parameter and five for range) are used, totalizing 300 different adjustments analyzed to find the best one. By cross-validation (Isaaks and Sriivastava, 1989), mean error (ME, Equation 16) and standard deviation of mean error (SDME, Equation 17) are calculated. ME and SDME values calculated for each parameter set are stored and used to determine ISI, thus, identifying the best adjustment for each model analyzed.

$$ISI = \left\{ \frac{abs(ME)}{max\left|_{i=1}^{j}[abs(ME)]\right.} + \frac{\left[SDME - min\left|_{i=1}^{j} SDME\right.\right]}{max\left|_{i=1}^{j}[abs(SDME)]\right.} \right\}, \quad (15)$$

where ME is the mean error and $SDME$ is the standard deviation of mean error of the crossed validation; n is the number of data; abs is the module value; $min|_{i=1}^{j}$ is the lowest value found among the compared j models; $max|_{i=1}^{j}$ is the highest value found among the compared j models.

$$ME = \frac{1}{n}\sum_{i=1}^{n} Z(s_i) - \hat{Z}(s_{(i)}), \quad (16)$$

$$SDME = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Z(s_i) - \hat{Z}(s_{(i)})\right)^2}, \quad (17)$$

where n is the number of data; $Z(s_i)$ is the value observed at the point $s_{(i)}$; $\hat{Z}(s_{(i)})$ is the predicted value at the point $s_{(i)}$.

The statistic called error comparison index (ECI – Equation 18 – Souza et al., 2016) was used to determine the best semivariogram fit in each analyzed j model, which assumes a lower value for the model is better stochastic methods of interpolation. The best semivariogram of each j model was used in ISI analysis.

$$ECI_i = \frac{|RME_i|}{10^{-10} + max\Big|_{i=1}^{j} |RME|} + \frac{|SDRME_i - 1|}{10^{-10} + max\Big|_{i=1}^{j} |SDRME - 1|}, \tag{18}$$

where ECI_i is the error comparison index for model i; and $max\Big|_{i=1}^{j}$ is the highest value among the compared j semivariograms. The arbitrary constant 10^{-10} was included to avoid division by zero.

The reduced mean error (RME – Equation 19) and standard deviation of the reduced mean error (SDRME – Equation 20) was determinate by ordinary kriging cross-validation.

$$RME = \frac{1}{n}\sum_{i=1}^{n}\frac{Z(s_i) - \hat{Z}(s_i)}{\hat{\sigma}\left(\hat{Z}(s_i)\right)}, \tag{19}$$

$$SDRME = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{\left|Z(s_i) - \hat{Z}(s_i)\right|}{\hat{\sigma}\left(\hat{Z}(s_i)\right)}}, \tag{20}$$

where $Z(s_i) - \hat{Z}(s_i)$ is the prediction error associated with estimating yield at spatial location s_i; $Z(s_i)$ is the observed value; $\hat{Z}(s_i)$ is the estimated value obtained from the ordinary kriging cross-validation; $\hat{\sigma}\left(\hat{Z}(s_i)\right)$ is the estimated standard deviation associated with the estimated value, and n is the sample size.

3. 4. 6 Clustering Methods

The purpose of cluster analysis methods is to divide the data points of an agricultural area into classes, which are also denominated clusters, employing a similarity evaluation function for this division. Clustering methods are considered more complex than empirical methods, yet they allow a more significant differentiation among classes by less subjective criteria. They employ several variables in the process of defining MZs. Among the several clustering algorithm options described in literature, two algorithms have been often applied in studies regarding MZs generation : K-Means (MacQueen, 1967) and FCM (Bezdek, 1981). Examples of software that are specific for MZ delineation using FCM are Management Zone Analyst (MZA, Fridgen et al., 2004), FuzME (Minasny and McBratney, 2002), Software for Defining Management zones (SDUM – Bazzi et al. 2019b), ZoneMAP (Zhang et al., 2010), and the friendly interface software proposed by Albornoz et al. (2017).

Gavioli et al. (2019) evaluated twenty clustering algorithms using data obtained from 2010 to 2015 at three commercial agriculture areas cropped with soybean and corn in Paraná state, Brazil. From the variables of elevation, clay, sand, silt, SRP, declivity, and bulk density, a method based on the main component analysis (PCA) was applied to generate new variables used as input variables for the clustering algorithms. The methods by McQuitty and Fanny were considered the best ones because they produced the most significant reductions in yield variance (target variable) for the studied three areas. Such methods generated zones with high internal homogeneity and less fragmentation (appropriate for field operations). The classic FCM and K-Means algorithms created subareas that were significantly different in only two areas, for which the results obtained were similar to those ones of the algorithms by McQuitty and Fanny.

3. 4. 7 Management zone rectification

Data rectification allows changing the format of qualitative data layers minimally. Qualitative layers data are a resultfrom the grouping process or discrete TMs. Regardless of the method used to delimit these zones, isolated spots or pixels usually appear. The rectification methods of layers are based on morphological filters used in digital processing of image: median, opening, closure, and with the combination of opening and closure. These indices have already been used to reduce MZs fragmentation (Betzek et al., 2018; Albornoz et al., 2017; Córdoba et al., 2016; Gonzalez and Woods, 2008).

Morphological filters act out as a non-linear mathematical function used to segment useful information and object description, such as shape, edges, and skeletons (Gonzalez and Woods, 2008). In a combined morphological function (opening and erosion), small objects are removed, and the subsequent dilation tries to restore the remaining objects shape (Gonzalez and Woods, 2008). For example, the image opening filter preserves unsegmented parts of objects by firstly image dilation by merging an object's neighboring pixels into the object and then image erosion by removing the object's boundary pixels. Instead, image closure is erosion followed by dilation to eliminate non-segmented parts of the background (He et al., 2016).

In the median filter, pixel intensity values are examined in a small region of the filter size and the median intensity value is selected for the central pixel (He et al., 2016). Thus, the median acts out similarly to an open-closing. However, the open-closing has advantages over the median, since it requires less computation and decomposes the noise suppression task into two independent steps, i.e., suppressing positive spikes via the opening and negative spikes via the closing (Maragos, 2009).

3. 4. 8 Distance measures

Clustering techniques separate data objects into clusters that have some similarity relationships among data. A distance measure determines similarity among data objects. The similarity among objects within a cluster is the most relevant factor during the clustering process. A good cluster finds the maximum similarity among its members (Saxena et al., 2017).

Distance measures are classified according to the similarity, which defines the similarity degree among data objects (Couso et al., 2013):

- Diagonal distance (Equation 21) is a distance used to standardize measurements at the moment when the variance equality is detected during the clustering process (Odeh et al., 1992).

$$Dist(E_i, E_j) = \sqrt{(x_{il} - x_{jl}).A_D.(x_{il} - x_{jl})}, \tag{21}$$

where $Dist$ is the distance among the points; E is the attribute values for each point; AD is the diagonal matrix.

- Euclidean distance among points x_{il} and x_{jl} is the length of the line segment connecting them (Equation 22). It is highly sensitive to noise and usually not applied to data with many attributes (Yahyaoui and Own, 2018). Thus, it is recommended to normalize data beforehand when data are not on the same measurement scale (Liu et al., 2014).

$$Dist(E_i, E_j) = \sqrt{\sum_{l=1}^{M} (x_{il} - x_{jl})^2}, \tag{22}$$

where x_{il} and x_{jl} are the attribute values for each point.

- Mahalanobis Distance (Equation 23 – Mahalanobis, 1936) is a measure that considers the covariance among the analyzed data. Moreover, it corrects some Euclidean distance restrictions since it automatically considers the scale of the coordinates axes and the covariance among characteristics (Liu et al., 2014).

$$Dist(E_i, E_j) = \sqrt{(x_{il} - x_{jl})^T.S^{-1}.(x_{il} - x_{jl})}, \tag{23}$$

where $Dist$ is the distance among the points; x_{il} and x_{jl} are the attribute values for each point; S is the covariance matrix.

3. 4. 9 Management zones evaluation

After MZs delineation, it is necessary to evaluate whether it is effective. Some indices help in evaluation and decision-making for clustering adoption and others to compare MZs or TMs.

i. Evaluation of the management zones quality

a) **Variance reduction** (VR – Equation 24 – Xiang et al., 2007, Schenatto et al., 2017b) is calculated for the mean yield, expecting that the sum of data variances for each MZ is smaller than the total variance of the field.

$$VR = \left(1 - \frac{\sum_{i=1}^{c} W_i * V_{MZ_i}}{V_{field}}\right) * 100, \tag{24}$$

where c is the MZs number, W_i is the field rate of i-th MZ to the total field, V_{MZ_i} is data variance of the i-th MZ, and V_{field} is the field data variance.

b) **Fuzziness performance index** (FPI – Equation 25 – McBratney and Moore, 1985; Fridgen et al., 2004): measures the degree of separation among the fuzzy c groups generated from a data set. FPI varies from 0 to 1.

$$FPI = 1 - \frac{c}{(c-1)}\left[1 - \sum_{j=1}^{n}\sum_{i=1}^{c}(m_{ij})^2/n\right], \quad (25)$$

where c is the number of groups; n is the number of elements in the data set; and m_{ij} is the element of fuzzy membership matrix M.

c) **Modified partition entropy** (MPE – Equation 26 – McBratney and Moore, 1985; Fridgen et al., 2004): estimates the difficulty level of the *c* groups organization.

$$MPE = \frac{-\sum_{j=1}^{n}\sum_{i=1}^{c} m_{ij}\, log(m_{ij})/n}{log\, c}, \quad (26)$$

where c is the number of groups; n is the number of elements in data set; and m_{ij} is the element of the fuzzy membership matrix M.

d) **Improved cluster validation index** (ICVI – Equation 27 – Gavioli et al., 2016): is a composition of FPI, MPE, and VR% indices.

$$ICVI_i = \frac{1}{3} * \left(\frac{FPI_i}{Max\{FPI\}} + \frac{MPE_i}{Max\{MPE\}} + \left(1 - \frac{VR\%_i}{Max\{VR\%\}}\right)\right), \quad (27)$$

where FPI_i is FPI value for *i*-th MZ; MPE_i is the MPE value for *i*-th MZ; $VR\%_i$ is VR value for *i*-th MZ; and $Max\{Index_X\}$ represents the maximum value of $Index_X$ among n MZs.

e) **Analysis of variance** (ANOVA): The mean values of a target variable can be compared among classes in an MZ delineated using Tukey's test to identify whether the sub-regions presented significant differences among classes.

f) **Smoothness index** (SI – Equation 28 – Gavioli et al., 2016): The evaluation of the best methods to define clustering must also include the visual aspect of the clustering created and, therefore, it must be taken into consideration the contour curves smoothness since it facilitates visual interpretation and application of agricultural inputs in varied rates. SI calculates the frequency of class changes in TM in horizontal, vertical, and diagonal directions, pixel by pixel. On the hypothesis that the map is a single completely homogeneous area, a smoothness index of 100% will be obtained due to the absence of class changes. Likewise, if the map was generated with random values, SI would value close to zero.

$$SI = 100 - \left(\frac{\sum_{i=1}^{k} C_{H_i}}{4P_H} + \frac{\sum_{j=1}^{k} C_{V_j}}{4P_V} + \frac{\sum_{l=1}^{k} C_{DR_l}}{4P_{DR}} + \frac{\sum_{m=1}^{k} C_{DL_m}}{4P_{DL}}\right) * 100, \quad (28)$$

where C_{H_i} is the number of changes on line *i* (horizontal); C_{V_j} is the number of changes in column *j* (vertical); C_{DR_l} is the number of changes on diagonal *l* (diagonal right – DR); C_{DL_m} is the number of changes on diagonal *m* (diagonal left – DL); *k* is the maximum number of pixels on a line, column, or diagonal; P_H is the possibility of a pixel change on horizontal; P_V is the possibility of a pixel change on vertical; P_{DR} is the possibility of a pixel change on DR; and P_{DL} is the possibility of a pixel change on the diagonal left DL.

g) **Average Silhouette Coefficient** (ASC – Equation 29 – Rousseeuw, 1987) is obtained from the silhouette coefficient (SC), an evaluation index that measures satisfactory internal formation and external separation among clusters. SC value for point *p*, which is denoted by sc_p, is calculated using the mean of the intragroup distances a_p and the mean of intergroup distances b_p. ASC values vary from -1 to 1: -1 indicates an incorrect clustering, whereas 1 indicates groups with the best intra-group formation and intergroup separation possible. Kaufman and Rousseeuw (1990) classify the structure of the formed groups as: very robust (0.71 < ASC ≤ 1.00); reasonable unit (0.51 < ASC ≤ 0.70; weak (0.26 < ASC ≤ 0.50); none (ASC < 0.26). When the structure is weak, it is recommended to use another grouping.

$$sc_p = \frac{b_p - a_p}{Max(a_p, b_p)}, \tag{29}$$

where a_p - Mean of distances among point p and all other points in the same group; b_p - Mean of distances among point p and all points in the closest group that contains p.

h) **Fragmentation Index** (FI% – Equation 30 – Souza et al., 2021): it takes into account how higher is the number of zones (NMZ) in comparison with the number of classes (NC). The higher FI%, the higher fragmentation.

$$FI\% = 100\frac{NMZ - Nc}{Nc}, \tag{30}$$

i) **Global Quality Index** (GQI – Equation 31 – Beneduzzi, 2020): it looks for finding the best number of classes during MZs delineation, taking into account the values of ICVI, SIr%, and FIr%. The SIr% and Fir% are SI% and FI% index values after MZ rectification.

$$GQI_i = \frac{ICVI_i * (100 + FIr\%_i)}{SIr\%_i}, \tag{31}$$

j) **Modified Global Quality Index** (MGQI – Equation 32) this coefficient, proposed in this work, is an adaptation of GQI to include the ASC coefficient.

$$MGQI = \frac{ICVI * (100 + FIr\%)}{SIr\% * ASC}, \tag{32}$$

ii. Comparison between thematic maps and between management zones

a) **Coefficient of relative deviation** (CRD – Equation 33 – Coelho et al., 2009): it calculates the mean difference in modulus of the interpolated values on a thematic map compared to a map taken as a reference.

$$CRD = \sum_{i=1}^{n} ABS(\frac{Zi_B - Zi_A}{Zi_A}), \tag{33}$$

where Zi_A is the estimated value at the location i on the reference map, Zi_B is the value at location i on the map to be compared, and n is the total number of interpolated locations on the maps.

b) **Mean absolute difference** (MAD – Equation 34): it computes the mean absolute difference among values on both maps.

$$MAD = \frac{\sum_{i=1}^{n} ABS(Zi_B - Zi_A)}{n}, \tag{34}$$

where Zi_A is the value of location (pixel) i on the reference map, Zi_B is the value at location (pixel) i on the map to be compared, and n is the total number of observations on the maps.

c) **Kappa index** (Kp – Equation 35 – Cohen, 1960): In comparing thematic maps, one of the ways used to determine the accuracy of a thematic classification is the Kappa's index, which adopts a reference for comparison with the maps produced. The accuracy analysis of mappings is obtained by the confusion matrices, and, subsequently, Kappa's index of agreement is calculated (Congalton, 1991). In PA, Kappa index has been used to compare thematic maps generated by different interpolators (Betzek et al., 2018; Schenatto et al., 2017b; Xiang et al., 2007). Kappa value ranges from 0 to 1, with values close to 0 representing no agreement beyond chance and 1 representing full agreement. Landis and Koch (1977) proposed the following classification: 0 < Kp ≤ 0.2 indicates no agreement, 0.2 < Kp ≤ 0.4 weak agreement, 0.4 < Kp ≤ 0.6 moderate agreement, 0.6 < Kp ≤ 0.8 strong agreement, and 0.8 < Kp ≤ 1 very strong agreement.

$$K = \frac{n \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} (x_{i+} * x_{+i})}{n^2 - \sum_{i-1}^{r} (x_{i+} * x_{+i})}, \tag{35}$$

where n is the number of observations (sample points); r is the number of classes in the error matrix; x_{ii} is the number of combinations of line i and column i; x_{i+} is the total number of observations of line i; and x_{+i} is the total number of observations of column i.

d) **Global accuracy** (GA – Equation 36 – Foody, 2002): like *Kp*, GA measures the degree of agreement among maps (MZs) and corresponds to the simple percent agreement.

$$GA = \frac{\sum_{i=1}^{c} x_{ii}}{n}, \tag{36}$$

where, $\sum_{i=1}^{c} x_{ii}$ is the sum of the main diagonal of the error matrix with c classes and a total of N samples collected (number of points interpolated).

3.5 Thematic maps

Maps representing the ground and a topic associated with it are called TMs and aim at informing by graphical symbols where a specific geographical phenomenon occurs. TMs development is associated to data collection, analysis, interpretation, and information representation on a map facilitates the similarities identification and enables the spatial correlations visualization. However, it is first necessary to interpolate data onto a dense and regular grid to provide values for locations that are not sampled. This task is carried out with interpolation methods and geostatistical analysis, and it is the most used interpolation method. One specific case of TMs is contour maps built by connecting points of the same value and applying them to geographical phenomena that show continuity in geographic space. Another case is choropleth maps that use color to show specific variable's values within a defined geographic area. Contour and choropleth maps can be built from categorical data (elevation, temperature, precipitation, humidity, and atmospheric pressure) or relative data (density, percentages, and indices).

Usually, soil samples are analyzed to determine the levels of nutrients on soil. Therefore, the sampling must be dense enough to determine nutrients variability on soil so that fertilizers may be used in a profitable and environmentally sustainable manner (Ferguson and Hergert, 2009; Franzen et al., 2002). Sampling with minimum densities from 1 sample ha^{-1} (Ferguson and Hergert, 2009) to 2.5 samples ha^{-1} (Journel and Huijbregts, 1978; Doerge, 2000) is suggested and must be composed of at least eight individual samples (Wollenhaupt et al., 1994). For building TMs, it is necessary to follow a protocol such as the one proposed by Souza et al. (2018), presented in Fig. 7.

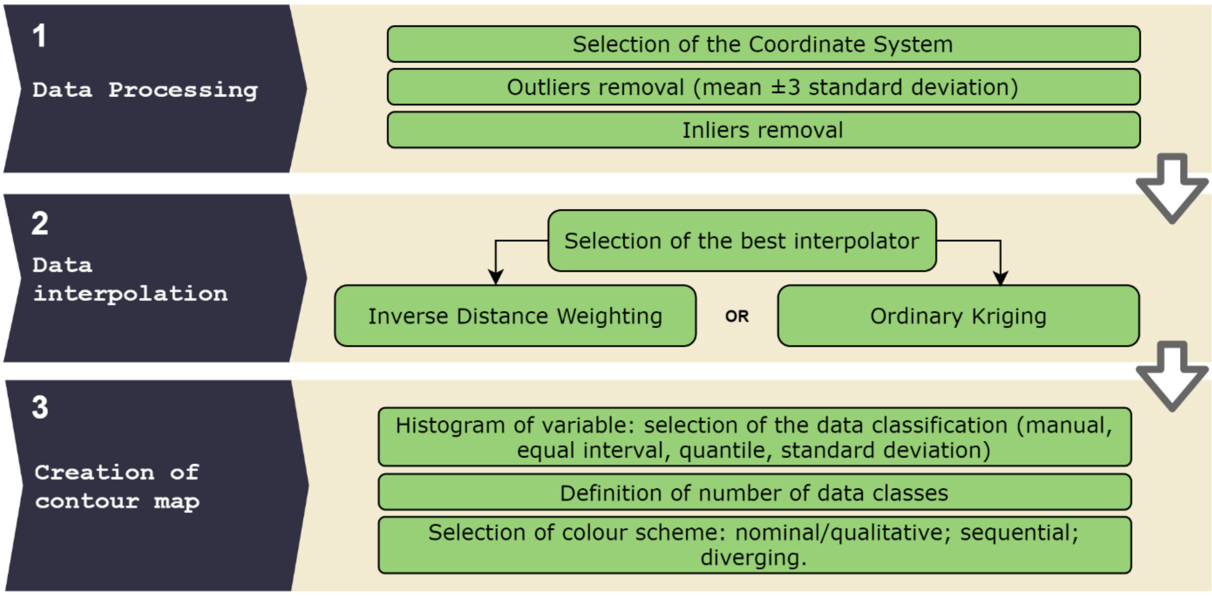
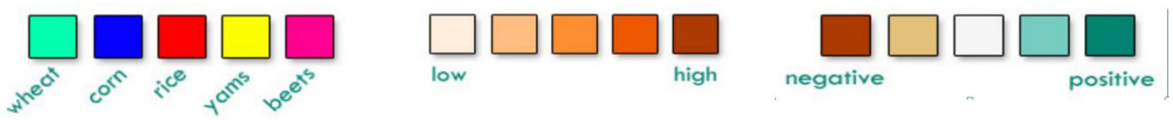


Fig. 7 Flowchart of the typical protocol to create a thematic map.
Source: Souza et al. (2018).

After data interpolation, It must be decided the number of classes and the method used to divide data into intervals to build TMs with such data. The goal is to group similar observations and split apart substantially different observations (Indiemapper, 2016). The most common grouping forms are 1) Manual interval; 2) Equal interval (equal-sized classes); 3) Quantile (classes with the same number of elements); and 4) Standard deviation (the class size is a multiple of the standard deviation). The number of data classes is also an essential part of the contour map design. Increasing the number of data classes results in a more revealing map but requires more colors. Generally, it is advised not to exceed seven classes.

After selecting the way to classify data, it is also essential to choose an effective color scheme for TM and, thus, define the three-color dimensions (hue, lightness, and saturation) for each class. There are three types of color schemes (Fig. 8): 1) Nominal/qualitative (unordered data such as land use): different hues with the same lightness and saturation; 2) Sequential (ordered data such as numerical data): single or multihued with different lightness/saturation; and 3) Diverging (when there is a mid-point, such as zero, or if someone wants to compare with an average such as profit): two different hues with different lightness/saturation starting from a central neutral color, usually white.



a) Nominal/qualitative b) Sequential c) Diverging

Fig. 8 Three types of color schemes: nominal/qualitative (a), sequential (b), and diverging (c).

3.6 Software for delineation of Management Zones

The process for delineating MZ can be made more accessible when using specialized software for this purpose. There are few options available to perform all phases of MZs delineation but use several software to accomplish the task in many cases. For example, some software tools are SAS (Statistical Analysis System), SPSS statistical software, Statistic (StatSoft Inc., currently maintained by TIBCO Software Inc), GS+, ArcGis (Environmental Systems Research Institute, Redlands, CA), Software R (R Core Team, 2014), FuzMe, MZA, Matlab and GRASS GIS (Damian et al., 2020; Méndez-Vázquez et al., 2019; Oldoni et al., 2019; Behera et al., 2018; Peralta et al., 2015; Chang et al., 2014)

FuzME (Minasny and McBratney, 2002) is a Fortran software capable of delineating MZs (Arshad et al., 2019; Behera et al., 2018) by FCM algorithm. It provides FPI and MPE indices, which are used to detect the class ideal number. This software is not specialized in performing all the tasks of a protocol for generating MZs; therefore, the pre-and post-processing steps must be performed with other software.

MZA (Fridgen et al., 2004) is a precursor to offer specialized tools for MZs delineation. Several researches involving MZs have already been carried out using MZA (Breunig et al., 2020; Damian et al., 2020; Peralta et al., 2015). It was developed with Microsoft Visual Basic 6.0 and operated in a Windows environment. Software calculates descriptive statistics and delineates MZs using FCM algorithm. The evaluative measures of the generated clusters can be made by FPI indices and normalized classification entropy (NCE).

The R software has been used for the development of studies concerning MZ delineation. Frequent use is given by the autonomy offered to the researcher to install libraries (packages) or develop their functions in execution scripts. It can be used to perform the entire MZ delineation process (Gavioli et al., 2019, Betzek et al., 2019) or in part of the process (Méndez-Vázquez et al., 2019; Oldoni et al., 2019; Gili et al., 2017). In addition, some practical applications for MZs design have been developed using the R software as part of the execution processes (Dall'agnol et al., 2020; Paccioretti et al., 2020).

ZoneMAP (Zhang et al., 2010) is an easy-to-use tool that uses FCM to delineate MZs with several data sources, including remote sensing images and field data collected by users. It can access a remote sensing database, and pre-processes data automatically, including format conversion and projection adjustment. The optimal number of MZ is determined based on the variance reduction.

EZZone (Lowrance et al., 2016) is web software to delineate MZ with FCM algorithm, composed of back-end and front-end. The back-end was developed in Python programming language and used SciPy and NumPy scientific computing libraries. The front-end uses HTML and JavaScript technologies, along with OpenLayers and jQuery libraries. The tool has a friendly interface in which MZs are delineated by a semi-supervised process interactively with the user. There is no need to create an account to use it. The software presents the optimal amount of MZs for the analyzed dataset. Once outlined, MZs can be rectified by the user.

The software by Albornoz et al. (2017) was developed in C++ and used FCM to delineate MZs. The authors characterize it as a user-friendly software focused on end-users, without the need-to-know advanced GIS and statistical skills. It generates files in ESRI Shapefile format. In determining the optimal number of MZs, the Euclidean distance of FPI cluster evaluation indices, normalized classification entropy (NCE), and Xie and Beni (XB) were used. In addition, the software allows rectifying delineated MZs with automatic post-processing, including the application of fashion, erosion and dilation filters, and the merging of smaller areas.

GeoFis (Leroux et al., 2018) is an open-source software focused on PA, allowing delineating MZs. It was developed in Java and can run on Windows, Linux, or Docker container systems. It uses R software for executing routines. It interpolates data by IDW and kriging. MZs delineation is done by segmentation algorithm (Pedroso et al., 2010).

AgDataBox (ADB; Michelon et al., 2019; Borges et al., 2020; Dall'agnol et al., 2020) is a digital platform that provides free computational tools for farmers, researchers, and service providers focused on PA by integrating data, software, procedures, and methodologies, in an attempt to enable the agricultural sector with free technologies.

FastMapping (Paccioretti et al., 2020) emerged as web software that can accomplish several protocol steps to delineate MZs. It was developed entirely in R software language. It allows cleaning, normalization, interpolation, data grouping, and evaluation indices. FCM from MULTISPATI-PCA (KM-sPC) algorithm is used (Córdoba et al., 2013) and its purpose is to reinforce that the Fast-track module, more available it is in ADB-Map software.

Other software applied in trials that design MZs are MATLAB (MathWorks Inc.) with fuzzy clusters (Fu et al., 2010), Statistica software with k-means algorithm (Whetton et al., 2018), and ArcGis with a neighborhood function via point statistics tool (Spatial Analyst toolset, ArcGIS PRO) (Ohana-Levi et al., 2019).

3.7 Web services

Modern software tools are, in most cases, built to be accessible by a computer network, with a web interface, stored data in relational databases or based on NoSQL (Not Only SQL) philosophy and can be integrated with other software, consuming or providing data. However, one of the challenges for integrating distributed software components is heterogeneity of programming languages, communication protocols, operating systems, data communication means, and hardware architecture of the devices.

Web service is a technology that allows interoperability of software on a computer network, in which information exchange (communication) is done over the hypertext transfer protocol (HTTP). Thus, an application programming interface (API) of a web service defines a standard for representing messages exchanged, usually in JavaScript Object Notation (JSON) or Extensible Markup Language (XML) (Grahl et al., 2017). The main advantage of using web services is the possibility of creating systems using reusable and loosely coupled software components

The most common classifications for web services are two and differ in how they are structured: tunneling style and Representational State Transfer (REST) architectural style (Fielding, 2000).

Tunneling is a style based on the principle of remote procedures call (RPC) for exchanging messages in simple object access protocol (SOAP) patterns and uses a language to describe the entire functioning of the web services description language (WSDL) (Grahl et al., 2017). Since there are many protocols and standards, this community is also known as “big web services” or WS-* stack (Richardson and Ruby, 2007).

On the other hand, the REST architectural style allows developing web services more simply, aiming to fully explore HTTP protocol for communication among applications without additional protocols. Furthermore, HTTP protocol is considered robust and sufficient to create web services (Fielding, 2000). Therefore, web services can be built on a lightweight architecture without encapsulating a protocol on HTTP protocol, as with SOAP. Fielding (2000) states that it is possible to integrate heterogeneous systems, to exchange messages and information without losing semantics of data among the parties involved, as well as guarantee security, integrity, and consistency in the distributed data using REST.

The application of web services to the agricultural sector is not new. For example, Spilke and Zürnstein (2005) highlight the potential of web services for data transfer among partners in agriculture and application integration, including outsourced service. However, in new software solutions, it has been preferred to use REST-based services. For example, research involving IoT in agriculture used REST architecture, such as animal monitoring systems (Nóbrega et al., 2018), for monitoring and forecasting data in a rose nursery (Rodríguez et al., 2017), platform for integrating sensors in an educational environment (Gunasekera et al., 2018) and in connected farms project (Ryu et al., 2015).

Therefore, during the decision support systems (DSS), it is also possible to find studies using the REST architecture, such as AgroDSS, which allows the use of data mining services (Rupnik et al., 2018), and e-Agriculture, which helps farmers in different stages of crop development (Mohanraj et al., 2016).

3.8 Microservices

Distributed computing systems have evolved by homogeneous cluster architectures in the 1990s, grid computing in the 2000s, cloud computing in the late 2000s, and, recently, ubiquitous computing and IoT (García-Valls et al., 2018). Software paradigms have also evolved; they have ceased to be isolated applications and have become part of collaborative platforms in which different partners can exchange information. One of the challenges of distributed applications is scalability, related to the software's ability to grow and expand due to business demands (Soldani et al., 2018). New software and technical architectures are studied and proposed to improve this situation and one of them is microservices architecture (MSA). In web software architecture, two approaches have been shown to the organization of its components, the traditional (monolithic) and the microservices-based (Fig. 9).

Software developed on monolithic approach is seen as a single artifact (Lewis and Fowler, 2014), a single web product deployed on the application server, containing the main logical parts, view (graphical user interface), data model (data representation and storage), and control functionalities. Therefore, the application can be difficult to understand and modify in monolithic approach, especially when it becomes large. In addition, a large codebase decreases the development team productivity. Some of the challenges found by the development team are regarding the division of activities for implementation, replacement of team members, difficulty in working independently, coordination of all development efforts, and code redistribution (Richardson, 2018).

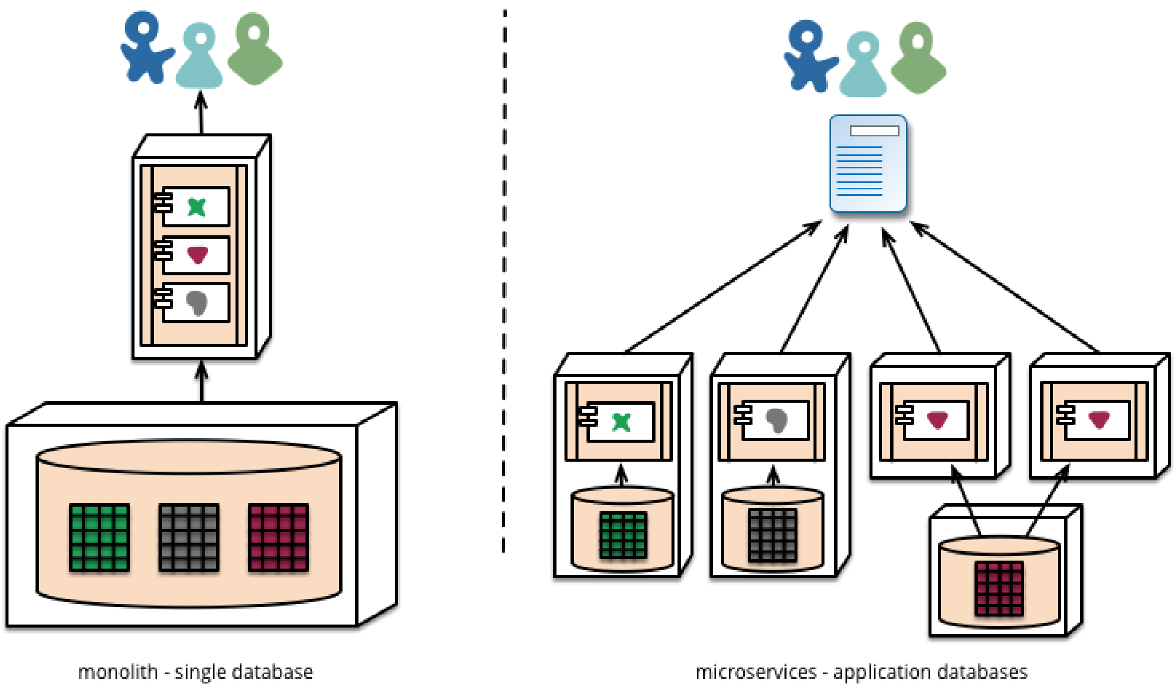


Fig. 9 Comparison of software architecture in traditional (monolithic), with a single artifact, and microservice approaches, where the application comprises several artifacts.
Source: Lewis and Fowler (2014).

On the other hand, the MSA aims to organize, logically and structurally, an application in small cohesive components. The application is seen as a set of small services, modular and loosely coupled, each one of them dedicated to a single activity (Ciavotta et al., 2017). Each microservice can be its database. An essential feature of microservices is to support the application's continuous delivery/deployment, which provides agile software provisioning

(García-Valls et al., 2018). Furthermore, an application can grow in different independent parts. As a result, the microservice approach provides scalability to applications, which is an essential factor in supporting several platforms and devices.

This approach has aroused the researchers' interest and software developers to create solutions in the most different industries, such as HazMate (Cherradi et al., 2017), which is real-time software for environmental information to assist on the logistics of transporting dangerous cargo in urban areas, as it provides creation of safer routes. MSA allowed the system to be distributed and business-oriented, as each microservice has business implementation. In dairy farming, a system based on "fog computing" was developed to analyze and monitor animal's health and behavior (Taneja et al., 2018). However, no research publications were found out with reports of trials using MSA to develop software focused on TMs creation and MZs delineation.

3.9 AgDataBox digital platform

This digital platform ADB has several application programming interfaces (API) in MSA (ADB-MSA), which consists of a set of resources accessible by the hypertext transfer protocol (HTTP) to transfer request and response messages expressed in JSON format. ADB-MSA, where data and processing routines are centered, enables interoperability of several applications. The following applications that consume ADB-MSA resources are under the development phase: 1-ADB-Mobile (Schenatto et al., 2017c); 2-ADB-Map (Borges et al., 2020; Michelon et al., 2019); 3-ADB-Admin; 4-ADB-IoT; and 5-ADB-SR. Fig. 10 presents the architecture of Digital platform ADB.

This platform originated from SDUM, developed on a desktop environment, which required its installation on computers with high processing capacity and memory availability due to the complexity of the implemented functionalities. Despite the acceptance of SDUM by researchers and producers, we opted for migrating to a web platform, including new modules and functionalities, but maintaining its gratuity.

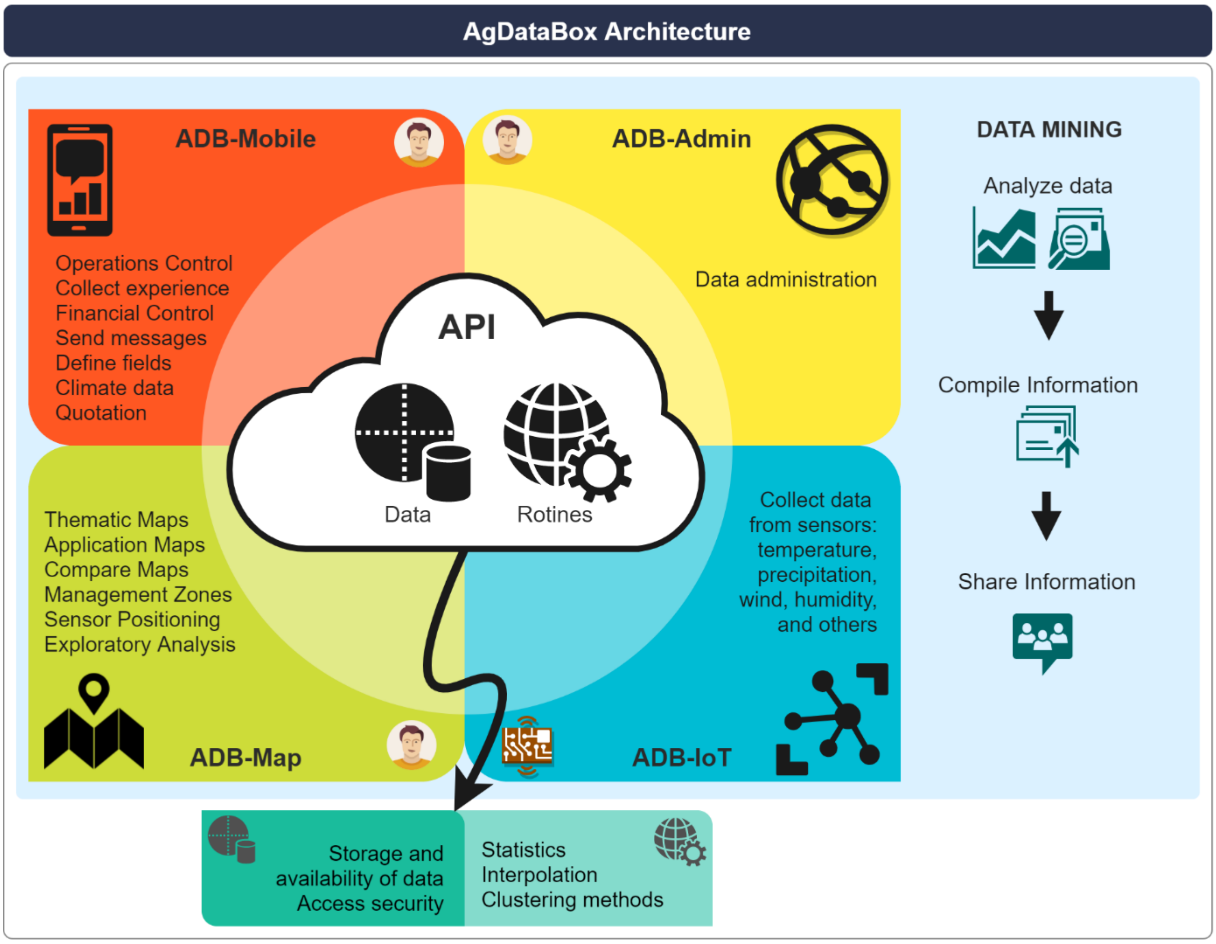


Fig. 10 AgDataBox digital platform architecture with its ADB-Data-API, ADB-Mobile, ADB-Admin, ADB-Map, and ADB-IoT applications.

3. 9. 1 AgDataBox Map (ADB-Map)

ADB-Map is the application that works with spatial data aiming to create TM and MZ to subsidize PA and DA. ADB-Map functionalities are divided into different layers, composed of (i) a back-end, which contains algorithms and rules of business operation, and (ii) a front-end, which is the interface of interaction with the user. This approach is a trend of use in modern software (Eitzinger et al., 2019).

The available features in ADB-Map are:

- Data importing/exporting;
- Statistical analysis and data cleaning;
- Spatial operations with grids;
- Data normalization;
- Data interpolation and TMs creation by different interpolators, including selecting the best interpolator between OK and IDW;
- Definition and evaluation of MZs, involving variable selection methods, data clustering methods, rectification, and quality evaluation;
- Fertilizers and lime recommendation.

ADB-Map preliminary version (Borges et al., 2020; Michelon et al., 2019) was implemented in a monolithic approach following the Model-View-Controller (MVC) standard, based on Java language with web technologies, such as VRaptor framework, and deployed on Apache Tomcat application server, PostgreSQL database, and many of functionalities implemented in PL/pgSQL procedural language. This application was restructured, thus, it generates a new application in new application architecture and with different technologies. This application is discussed as part of this work (Chapter 6 – Paper 2).

3. 9. 2 AgDataBox Data API (ADB-Data-API)

According to the scenario with several technologies that make up ADB digital platform, there was a need to integrate the generated data and implemented procedures among different software, portals, and devices that create the platform. ADB-Data-API (Bazzi et al., 2019a) was created to centralize and share data on Web, and is registered with the Brazilian National Institute of Industrial Property (INPI, BR 51 2018 000899-2). This API manages agricultural data (acting as a database abstraction layer), making data persistence and availability for client applications, and operating in data security management. The types of data currently managed in ADB-Data-API refer to map resources, area characteristics, climate, and agricultural area management (Fig. 11).

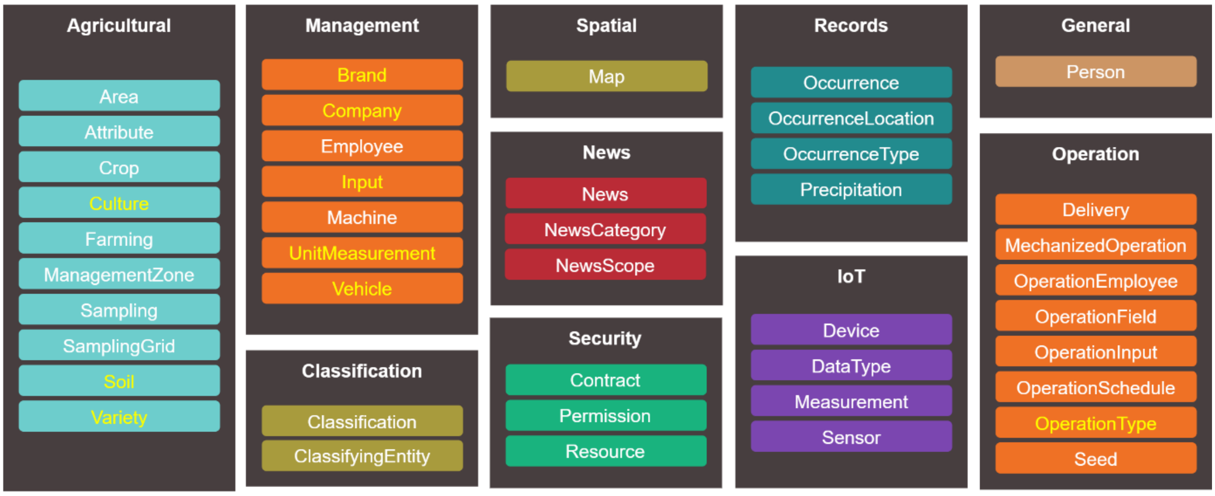


Fig. 11 Data model managed by ADB-Data-API.

ADB-Data-API is based on the REST architectural style. Communication among applications and ADB-Data-API is done over HTTP protocol, which is fundamental for the web functioning, thus, it provides greater ease for integration among applications and ADB-Data-API. JSON is the format adopted to represent data when transferring among applications and ADB-Data-API.

All data stored on ADB-Data-API is associated with its owner; however, ADB-Data-API allows sharing it with other registered users. Access permissions are granted by data type and by access level, such as viewing, creating, editing, and deleting. Access authentication is handled with login and password, the phone number. The e-mail address is considered as the user's login. When authentication is performed, the client receives a token, a key encrypted by ADB-Data-API, and must be used to carry out operations on available resources. All data storage is done in PostgreSQL databases and managed by a web interface, which is not intended to the users' access, but is consumed by applications, so it is called an API.

When communicating with ADB-Data-API, HTTP methods (get, post, put or delete), a uniform resource identifier (URI), and a data representation in a standardized format are used, in this case, JSON (Fig. 12).

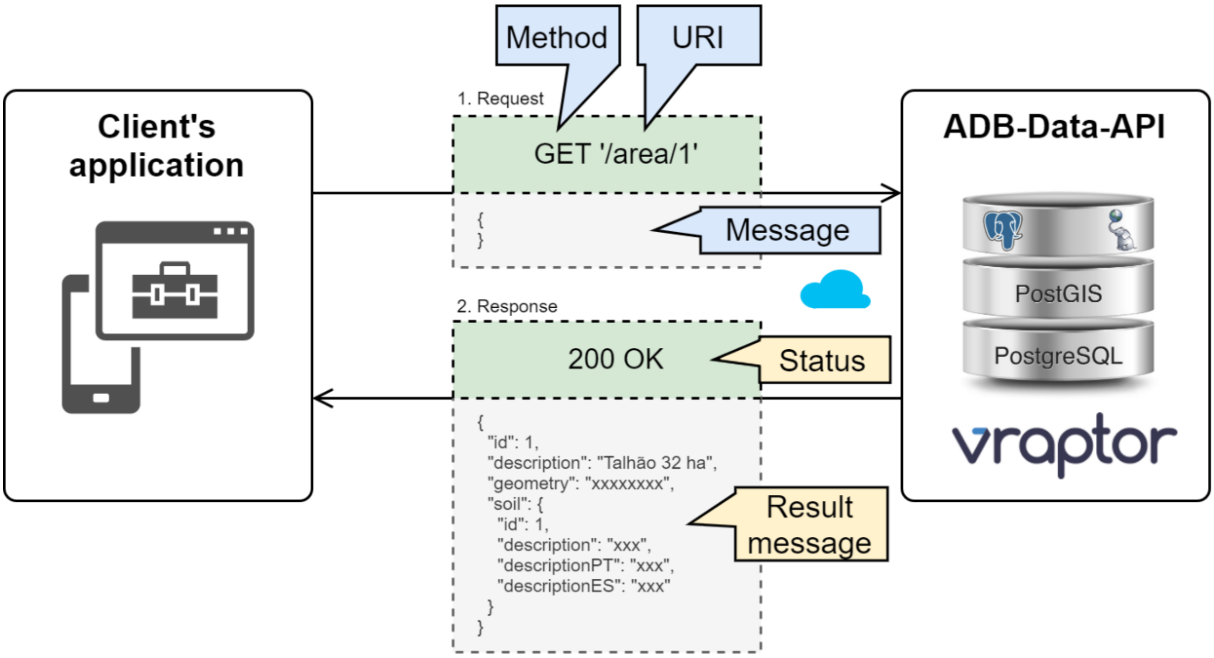


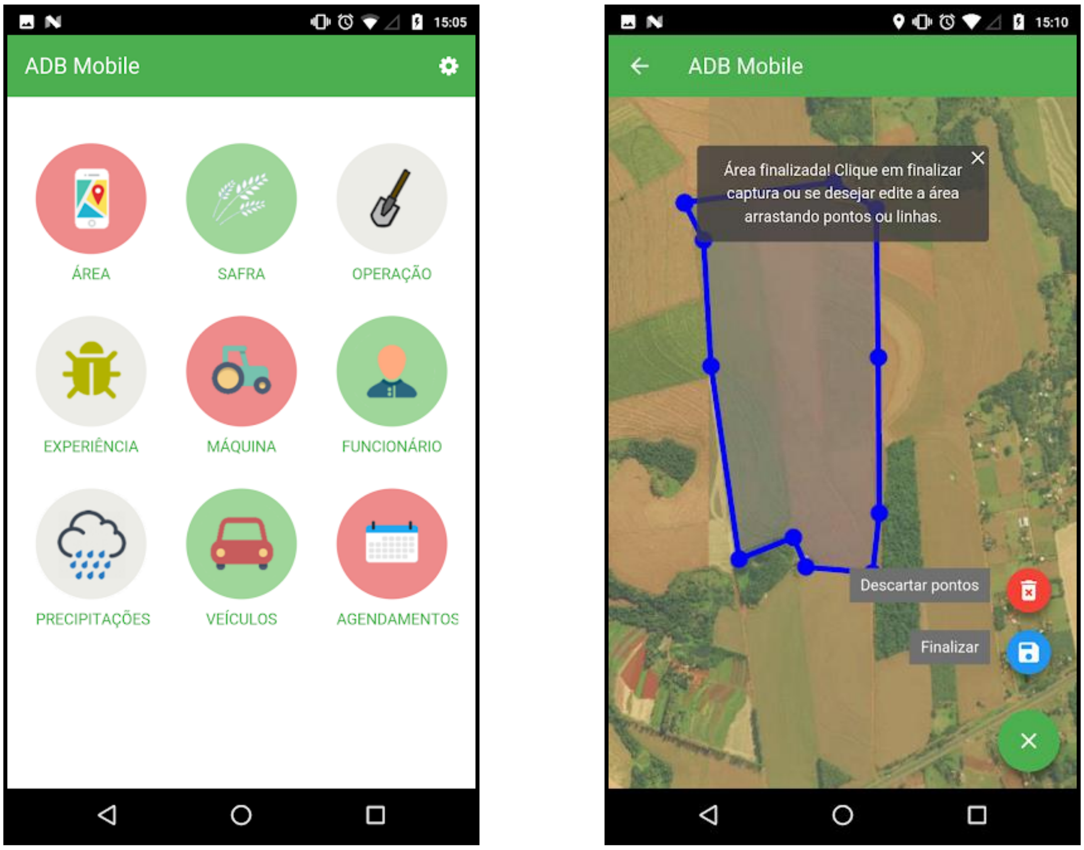
Fig. 12 Data request representation and response process in communication among a client's application and API

3. 9. 3 AgDataBox Mobile (ADB-Mobile)

ADB-Mobile application (Fig. 13, Schenatto et al., 2017c) operates on the Android operating system. It has two main objectives: (i) to be a valuable tool for farmers to record facts and organize operations on their property, able to keep the record of all operations and occurrences of a harvest; these data are stored locally on a mobile device and a data server, (ii) allow the registration of the producer's variable experience regarding the division of areas in MZs. ADB-Mobile allows you to perform operations in offline mode and later synchronize data with ADB-Data-API in online mode.

3. 9. 4 AgDataBox Admin (ADB-Admin)

ADB-Admin is a web application (Fig. 14) whose main objective is to manage the resources offered in API (ADB-Data-API) for storing platform data. It was developed in PHP programming language. The resources administration in the software is done to make it possible to view, create, edit, and delete the user's data and other users who have granted access permission.



(a) Home screen. (b) Boundaries demarcation's screen.

Fig. 13 Home screen (a) and boundaries demarcation screen (b) in ADB-Mobile application.

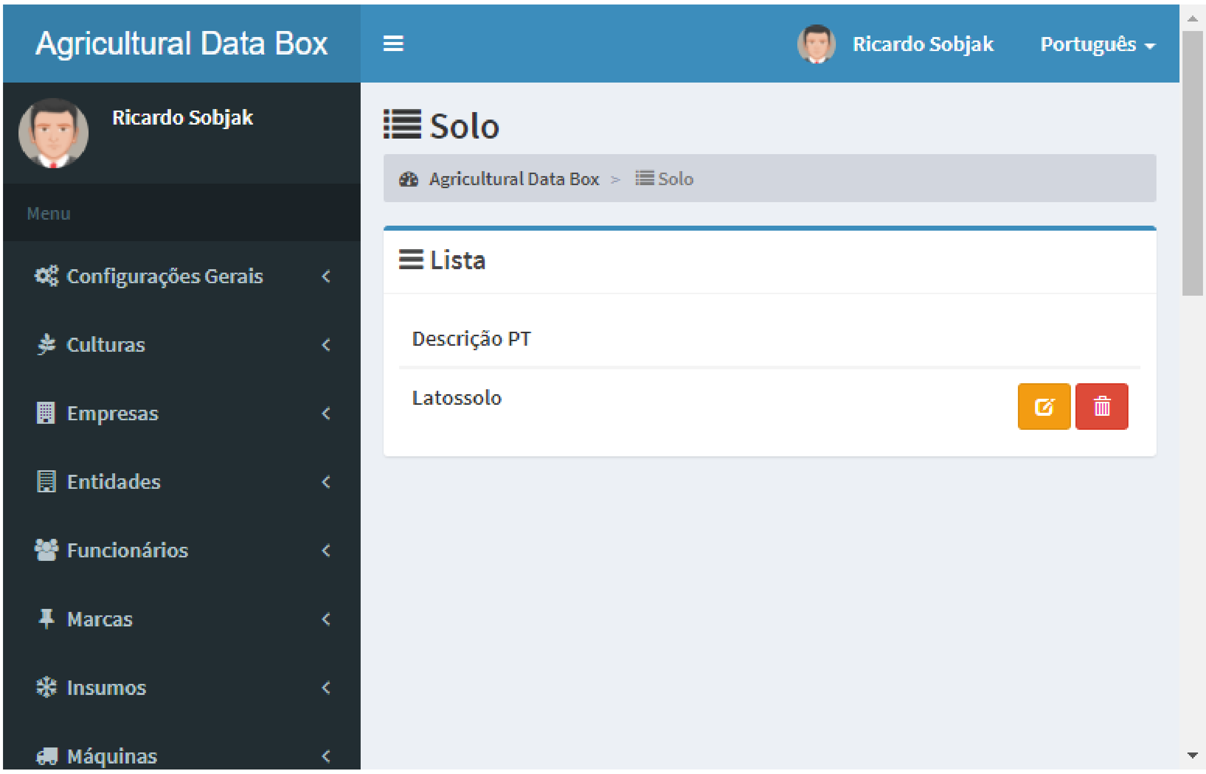


Fig. 14 An example of AgDataBox-Admin screen.

3. 9. 5 AgDataBox Remote Sensing (ADB-RS)

ADB-RS (Conti, 2021) is an application that aims at acquiring and processing images obtained by remote sensing. Data is extracted from a multispectral image, transformed into information, and exported for external applications. Application is integrated with ADB-MSA and allows the extracted data to be used, for example, in ADB-Map MZs delineation.

4 REFERENCES

Abreu, S. L., Reichert, J. M., Silva, V. R. da, Reinert, D. J., Blume, E. (2003). Variabilidade espacial de propriedades físico hídricas do solo, da produtividade e da qualidade de grãos de trigo em argissolo franco arenoso sob plantio direto. **Ciência Rural**, 33 (2), pp. 275-282.

Albornoz, E. M., Kemerer, A. C., Galarza, R., Mastaglia, N., Melchiori, R., Martínez, C. E. (2018). Development and evaluation of an automatic software for management zone delineation. **Precision Agriculture**, 19 (3), pp. 463-476.

Almeida, E. (2012). **Econometria espacial aplicada**. Editora Alínea, Campinas.

Akhter, R., Sofi, S. A. (2021). Precision agriculture using IoT data analytics and machine learning. **Journal of King Saud University - Computer and Information Sciences**, (in press).

Alm, E., Colliander, N., Lind, F., Stohne, V., Sundström, O., Wilms, M., Smits, M. (2016). Digitizing the Netherlands: How the Netherlands Can Drive and Benefit from an Accelerated Digitized Economy in Europe. Boston Consulting Group, accessed 18 December 2020, <https://www.bcg.com/en-nl/digitizing-the-netherlands>.

Ampatzidis, Y., Tan, L., Haley, R., Whiting, M. D. (2016). Cloud-based harvest management information system for hand-harvested specialty crops. **Computers and Electronics in Agriculture**, 122 (1), pp. 161–167.

Anderberg, M. R. (1973). **Cluster Analysis for Applications**. Academic Press, New York.

Anselin, L. (1995). Local indicators of spatial association-LISA. **Geographical Analysis**, 27 (2), pp. 93-115.

Arshad, M., Li, N., Zhao, D., Sefton, M., Triantafilis, J. (2019). Comparing management zone maps to address infertility and sodicity in sugarcane fields. **Soil and Tillage Research**, 193, pp.122-132.

Azevedo, L., Soares, A. (2017). **Geostatistical Methods for Reservoir Geophysics. Advances in Oil and Gas Exploration & Production Series**. Springer International Publishing, Cham. 159 p.

Barnes, A. P., Soto, I., Eory, V., Beck, B., Balafoutis, A., Sánchez, B., Vangeyte, J., Fountas, S., Van der Wal, T. Gómez-Barbero, M. (2019). Exploring the adoption of precision agricultural technologies: A cross regional study of EU farmers. **Land Use Policy**, 80, pp. 163-174.

Bazzi, C. L., Jasse, E. P., Magalhães, P. S. G., Michelon, G. K., Souza, E. G., Schenatto, K., Sobjak, R. (2019a). AgDataBox API – Integration of data and software in precision agriculture, **SoftwareX**, 10.

Bazzi, C. L., Souza, E. G., Schenatto, K., Betzek, N. M., Gavioli, A. (2019b). A software for the delineation of crop management zones (SDUM). **Australian Journal of Crop Science. Southern Cross Journals**, 13, pp. 26-34.

Behera, S. K., Mathur, R. K., Shukla, A. K., Suresh, K., Prakash, C. (2018). Spatial variability of soil properties and delineation of soil management zones of oil palm plantations grown in a hot and humid tropical region of southern India. **Catena**, 165, pp. 251-259.

Beneduzzi, H. M. (2020). **Módulo computacional para cálculo da necessidade de nitrogênio, fósforo e potássio a partir de suas disponibilidades no solo**. PhD thesis, Western Paraná State University (UNIOESTE), Cascavel.

Bernardi, A. C. de C., Machado, P. L. O. de, Freitas, P. L. de, Coelho, M. R., Leandro, W. M., Oliveira Júnior, J. P. de, Oliveira R. P. de, Santos, H. G. dos, Madari, B. E., Carvalho, M. da C. S. (2003). **Correção do solo e adubação no sistema de plantio direto nos cerrados,** 46. Embrapa Solos, Rio de Janeiro. 22p.

Betzek, N. M., Souza, E. G., Bazzi, C. L., Schenatto, K., Gavioli, A. (2018). Rectification methods for optimization of management zones. **Computers and Electronics in Agriculture**, 146 (1), pp. 1–11.

Betzek, N. M., Souza, E. G., Bazzi, C. L., Schenatto K., Gavioli, A., Magalhães, P. S. G. (2019). Computational routines for the automatic selection of the best parameters used by interpolation methods to create thematic maps. **Computers and Electronics in Agriculture**, 157, pp. 49-62.

Beutler, A. N., Munareto, J. D., Ramão, C. J., Galon, L., Dias, N. P., Pozzebon, B. C. (2012). Propriedades físicas do solo e produtividade de arroz irrigado em diferentes sistemas de manejo. **Revista Brasileira de Ciência do Solo**, 36, pp. 1601-1607.

Bezdek, J. C. (1981). **Pattern Recognition with Fuzzy Objective Function Algorithms**. Plenum Press, New York.

Bier, V. A., Souza, E. G. (2017). Interpolation selection index for delineation of thematic maps. **Computers and Electronics in Agriculture**, 136 (1), pp. 202-209.

Biondi, F., Myers, D. E., Avery, C. C. (1994). Geostatistically modeling stem size and increment in an old-growth forest. **Canadian Journal of Forest Research**, 24 (7), pp. 1354-1368.

Blackmore, S. (2000). The interpretation of trends from multiple yield maps. **Computers and Electronics in Agriculture**, 26 (1), pp. 37-51.

Bobryk, C. W., Myers, D. B., Kitchen, N. R., Shanahan, J. F., Sudduth, K. A., Drummond, S. T., Gunzenhauser, B., Gomez Raboteaux, N. N. (2016). Validating a digital soil map with corn yield data for precision agriculture decision support. **Agronomy Journal**, 108 (3), pp. 957-965.

Bongiovanni, R., Lowenberg-Deboer, J. (2004). Precision Agriculture and Sustainability. **Precision Agriculture**, 5, pp. 359–387.

Borges, L. G., Bazzi, C. L., Souza, E. G., Magalhães, P. S. G., Michelon, G. K. (2020). Web software to create thematic maps for precision agriculture. **Pesq. agropec. bras**., 55.

Borkert, C. M., Yorinori, J. T., Correa-Ferreira, B. S., Almeida, A. M. R., Ferreira, L. P., Sfredo, G. J. (1994). Seja o doutor da sua soja. **Informações Agronômicas**, 66, pp 1-6.

Borkowski, A., Kwiatkowska-Malina, J. (2017). Geostatistical modelling as an assessment tool of soil pollution based on deposition from atmospheric air. **Geosciences Journal**, 21, pp. 645-653.

Bottega, E. L., Queiroz, D. M. de, Pinto, F. de A. de C., Souza, C. M. A. de. (2013). Variabilidade espacial de atributos do solo em sistema de semeadura direta com rotação de culturas no cerrado brasileiro. **Revista Ciência Agronômica**, 44 (1), pp. 1-9.

Breunig, F. M., Galvão, L. S., Dalagnol, R., Santi, A. L., Shuisen Chen, D. P. D. F. (2020). Assessing the effect of spatial resolution on the delineation of management zones for smallholder farming in southern Brazil. **Remote Sensing Applications: Society and Environment**, 19.

Bronson, K. (2019). Digitization and Big Data in Food Security and Sustainability. In: Ferranti, P., Berry, E. M., Anderson, J. R. (Eds.), **Encyclopedia of Food Security and Sustainability**. Elsevier, Oxford, UK, pp. 582–587.

Burke, B. (2020). Top Strategic Technology Trends for 2021. Gartner, accessed 20 April 2021, <https://www.gartner.com/en/publications/top-tech-trends-2021>.

Cambardella, C. A., Mooman, T. B., Novak, J. M., Parkin, T. B., Karlen, D. L., Turv, R. F., Konopka, A. E. (1994). Field-scale variability of soil properties in central Iowa soil. **Soil Science Society of America Journal**, 58 (5), pp. 1501-1511.

Cambouris, A. N., Zebarth, B. J., Ziadi, N., Perron, I. (2014). Precision Agriculture in Potato Production. **Potato Research**, 57, pp. 249–262.

Carneiro, M. A. C., Souza, E. D. de, Reis, E. F. dos, Pereira, H. S., Azevedo, W. R. de. (2009). Atributos físicos, químicos e biológicos de solo de Cerrado sob diferentes sistemas de uso e manejo. **Revista Brasileira de Ciência do Solo**, 33, pp. 147-157.

Carvalho, L. A., Meurer, I., Silva Junior, C. A., Centurion, J. F. (2012). Spatial variability of soil physical properties in two management systems in sugarcane crop. **Engenharia Agrícola**, 32 (1), pp. 60-68.

Carvalho, P. S. M., Franco, L. B., Silva, S. A., Sodré, G. A., Queiroz, D. M., Lima, J. S. S. (2016). Cacao Crop Management Zones Determination Based on Soil Properties and Crop Yield. **Rev. Bras. Ciênc. Solo**, 40, pp. 1-17.

Castrignanò, A., Buttafuoco, G., Quarto, R., Parisi, D., Viscarra Rossel, R. A., Terribile, F., Langella, G., Venezia, A. (2018). A geostatistical sensor data fusion approach for delineating homogeneous management zones in Precision Agriculture. **Catena**, 167 (1), pp. 293–304.

CEMA - European Agricultural Machinery. (2017). **Digital farming: what does it really mean?** CEMA, Brussels. 9p.

Chang, D., Zhang, J., Zhu, L., Ge, S. H., Li, P. Y., Liu, G. S. (2014). Delineation of management zones using an active canopy sensor for a tobacco field. **Computers and Electronics in Agriculture**, 109, pp. 172-178.

Cherradi, G., Bouziri, A. E., Boulmakoul, A., Zeitouni, K. (2017). Real-Time HazMat Environmental Information System: A micro-service based architecture. **Procedia Computer Science**, 109 (1), pp. 982-987.

Ciavotta, M., Alge, M., Menato, S., Rovere, D., Pedrazzoli, P. (2017). A Microservice-based Middleware for the Digital Factory. **Procedia Manufacturing**, 11 (1), pp. 931-938.

Clark, I. 1979. **Practical geostatistics**. Applied Science Publishers, London.

Coelho, E. C., Souza, E. G., Uribe-Opazo, M. A., Pinheiro Neto, R. (2009). Influência da densidade amostral e do tipo de interpolador na elaboração de mapas temáticos. **Acta Scientiarum**, 31 (1), pp. 165-174.

Cohen, J. A. (1960). Coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, 20 (1), pp. 37-46.

Coleman, N. T., Thomas, G. W. (1967). The Basic Chemistry of Soil Acidity. In: Pearson, R. W., Adams, F. (Eds.), **Soil Acidity and Liming**. American Soc. of Agronomy Inc. Publishers, Madison, pp. 1-41.

Colezea, M., Musat, G., Pop, F., Negru C., Dumitrascu, A., Mocanu, M. (2018). CLUeFARM: Integrated web-service platform for smart farms. **Computers and Electronics in Agriculture**, 154 (1), pp. 134–154.

Columbus, L. (2021). **10 Ways AI Has The Potential To Improve Agriculture In 2021**. Forbes, accessed 21 April 2021, <https://www.forbes.com/sites/louiscolumbus/2021/02/17/10-ways-ai-has-the-potential-to-improve-agriculture-in-2021/?sh=4a2d99da7f3b>.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing of Environment**, 37 (1), p. 35-46.

Conti, G. (2021). **Aplicação computacional AgDataBox-RS: Gerenciamento de dados de sensoriamento remoto**. PhD thesis, Western Paraná State University (UNIOESTE), Cascavel.

Córdoba, M., Bruno, C., Costa, J. L., Balzarini, M. (2013). Subfield management class delineation using cluster analysis from spatial principal components of soil variables. **Computers and Electronics in Agriculture**, 97 (1), pp. 6-14.

Córdoba, M. A., Bruno, C. I., Costa, J. L., Peralta, N. R., Balzarini, M. G. (2016). Protocol for multivariate homogeneous zone delineation in precision agriculture. **Biosystems Engineering**, 143, pp. 95-107.

Couso, I., Garrido, L., Sánchez, L. (2013). Similarity and dissimilarity measures between fuzzy sets: A formal relational study. **Information Sciences**, 229 (1), pp. 122-141.

Coutinho, E. L. M., Natale, W., Stupiello, J. J., Carnier, P. E. (1991). Avaliação da eficiência agronômica de fertilizantes fosfatados para a cultura do milho. **Científica**, 19, pp. 93-104.

Cressie, N. (1985). Fitting variogram models by weighted least squares. **Mathematical Geology**, 17 (4), pp. 563-586.

Cressie, N. A. C. (1993). **Statistics for Spatial Data**. Wiley-Interscience Publication, New York.

Dall'agnol, R. W., Michelon, G. K., Bazzi, C. L., Magalhães, P. S. G., Souza, E. G., Betzek, N. M., Sobjak, R. (2020). Web applications for spatial analyses and thematic map generation. **Computers and Electronics in Agriculture**, 172.

Damian, J. M., Pias, O. H. C., Cherubin, M. R., Fonseca, A. Z., Fornari, E. Z., Santi, A. L. (2020). Applying the NDVI from satellite images in delimiting management zones for annual crops**. Sci. agric.**, 77 (1), pp. 1-11.

Dave, R. N. (1992). Generalized fuzzy c-shells clustering and detection of circular and elliptical boundaries. **Pattern Recognition**, 25 (7), pp. 713-721.

Diggle, P. J., Ribeiro Jr., P. J. (2007). **Model-based geostatistics**. Springer, New York.

Doerge, T. A. (2000). Management Zone Concepts. **Site-Specific Management Guidelines**. Potash and Phosphate Institute. University South Dakota, Brokings.

Dray, S., Saïd, S., Débias, F. (2008). Spatial ordination of vegetation data using a generalization of Wartenberg's multivariate spatial correlation. **Journal of Vegetation Science**, 19 (1), pp. 45-56.

Eitzinger, A., Cock, J., Atzmanstorfer, K., Binder, C. R., Läderach, P., Bonilla-Findji, O., Bartling, M. Mwongera, C., Zurita, L., Jarvis, A. (2019). GeoFarmer: A monitoring and feedback system for agricultural development projects. **Computers and Electronics in Agriculture**, 158, pp. 109-121.

Faraco, M. A., Uribe-Opazo, M. A., Silva, E. A. A., Johann, J. A., Borssoi, J. (2008). Selection criteria of spatial variability models used in thematical maps of soil physical attributes and soybean yield. **Revista Brasileira de Ciência do Solo**, 32 (2), pp. 463-476.

Ferguson, R. B., Hergert, G. W. (2009). Soil Sampling for Precision Agriculture. **Precision Agriculture**, pp. 1-4.

Fielding, R. T. (2000). **Architectural Styles and the Design of Network-Based Software Architectures**. Ph.D. thesis, University of California, Irvine.

Foody, G. M. (2002). Status of land cover classification accuracy assessment. **Remote Sensing of Environment**, 80 (1), pp. 185–201.

Fraisse, C. W., Sudduth, K. A., Kitchen, J. R. (2001). Delineation of site-specific management zones by unsupervised classification of topographic attributes and soil electrical conductivity. **International Journal of the American Society of Agricultural and Biological Engineers**, 1 (44), pp. 155-166.

Franzen, D. W., Hopkins, D. H., Sweeney, M. D., Ulmer, M. K., Halvorson, A. D. (2002). Evaluation of Soil Survey Scale for Zone Development of Site-Specific Nitrogen Management. **Agronomy Journal**, 94 (2), pp. 381-389.

Fraser, E. D. G., Campbell, M. (2019). Agriculture 5.0: Reconciling Production with Planetary Health. **One Earth**, 1 (3), pp. 278-280.

Fridgen, J. J., Kitchen, N. R., Sudduth, K. A., Drummond, S. T., Wiebold, W. J., Fraisse, C. W. (2004). Management zone analyst (MZA): software for subfield management zone delineation. **Agronomy Journal**, 96 (1), pp. 100-108.

Fu, Q., Wang, Z., Jiang, Q. (2010). Delineating soil nutrient management zones based on fuzzy clustering optimized by PSO. **Mathematical and Computer Modelling**, 51 (11–12), pp. 1299-1305.

Gamero, P., Uribe-Opazo, M. A., De Bastiani, F., Johann, J. A., Guedes, L. P. C. (2020). Variabilidade espacial da precipitação no cultivo de milho segunda safra no paraná utilizando o modelo WAVE. **Irriga**, 25 (3), pp. 521–536.

García-Valls, M., Dubey, A., Botti, V. (2018). Introducing the new paradigm of Social Dispersed Computing: Applications, Technologies, and Challenges. **Journal of Systems Architecture**, 91, pp. 83-102.

Gavioli, A., Souza, E. G., Bazzi, C. L., Guedes, L. P. C., Schenatto, K. (2016). Optimization of management zone delineation by using spatial principal components. **Computers and Electronics in Agriculture**, 127 (1), pp. 302-310.

Gavioli, A., Souza, E. G., Bazzi, C. L., Schenatto, K., Betzek, N. M. (2019). Identification of management zones in precision agriculture: An evaluation of alternative cluster analysis methods. **Biosystems Engineering**, 181, pp. 86-102.

Gili, A., Álvarez, C., Bagnato, R., Noellemeyer, E. (2017). Comparison of three methods for delineating management zones for site-specific crop management. **Computers and Electronics in Agriculture**, 139, pp. 213-223.

Gnanadesikan, R., Kettenring, J., Tsao, S. (1995). Weighting and selection of variables for cluster analysis. **Journal of Classification**, 12 (1), pp. 113-136.

Gonzalez, R. C., Woods, R. (2008). **Digital image processing**, 3. Pearson Prentice Hall, New Jersey.

Grahl, M., Bluhm, T., Grün, M., Hennig, C., Holtz, A., Krom, J. G., Kühner, G., Laqua, H., Lewerentz, M., Riemann, H., Spring, A., Werner, A. (2017). Archive WEB API: A web service for the experiment data archive of Wendelstein 7-X. **Fusion Engineering and Design**, 123 (1), pp. 1015-1019.

Groher, T., Heitkämper, K., Walter, A., Liebisch, F., Umstätter, C. (2020). Status quo of adoption of precision agriculture enabling technologies in Swiss plant production. **Precision Agriculture**, 21, pp. 1327–1350.

Gunasekera, K., Borrero, A. N., Vasuian, F., Bryceson, K. P. (2018). Experiences in building an IoT infrastructure for agriculture education. **Procedia Computer Science**, 135 (1), pp. 155-162.

Hawkins, E., Singh, M. (2019). **The art and science of variable rate seeding**. Michigan State University Extension, accessed 20 April 2021, <https://www.canr.msu.edu/news/the-art-and-science-of-variable-rate-seeding>.

He, H. J., Zheng, C., Sun, D. W. (2016). Chapter 2 - Image Segmentation Techniques. In: Sun, D.-W. (Eds.), **Computer Vision Technology for Food Quality Evaluation**, 2. Academic Press, pp. 45-63.

He, X., Ding, Y., Zhang, D., Yang, L., Cui, T., Zhong, X. (2019). Development of a variable-rate seeding control system for corn planters Part I: Design and laboratory experiment. **Computers and Electronics in Agriculture**, 162, pp. 318-327.

Holland, K. H., Scheppers, J. S. (2010). Derivation of a Variable Rate Nitrogen Application Model for In-Season Fertilization of Corn. **Agronomy Journal**, 102 (5), pp. 1415-1424.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. **Journal of Educational Psychology**, 24 (6), pp. 417-441.

Indiemapper. (2016). **The basics of data classification**. Accessed 15 March 2019, <http://indiemapper.com/app/learnmore.php?l=classification>.

Isaaks, E. H., Srivastava, R. M. (1989). **Applied geostatistics**. Oxford University Press, New York.

ISPA. (2019). **ISPA Newsletter 7(7) July 2019: Official Definition of PA, The Impact Factor, Workshop DAP for Turkey, 14th ICPA recap, Upcoming Events, Jobs**. International Society of Precision Agriculture. Accessed 02 September 2021, <https://ispag.org/site/newsletter/?id=90>.

Jolliffe I. (2011). Principal Component Analysis. In: Lovric M. (eds) **International Encyclopedia of Statistical Science**. Springer, Berlin, Heidelberg.

Journel, A. G., Huijbregts, C. J. (1978). **Mining Geostatistics**. Academic Press, London-New York-San Francisco.

Kaloxylos, A., Groumas, A., Sarris, V., Katsikas, L., Magdalinos, P., Antoniou, E., Maestre Terol, C. (2014). A cloud-based Farm Management System: Architecture and implementation. **Computers and Electronics in Agriculture**, 100 (1), pp. 168–179.

Kaufman, L., Rousseeuw, P. J. (1990). **Finding groups in data.** John Wiley & Sons, Hoboken.

Khosla, R., Fleming, K., Delgado, J. A., Shaver, T., Westfall D. G. (2002). Use of site-specific management zones to improve nitrogen management for precision agriculture. **J. Soil Water Conserv.**, 57 (6), pp. 513–518.

Konopatzki, M. R., Souza, E. G., Nóbrega, L. H., Uribe-Opazo, M. A., Suszek, G. (2012). Spatial variability of yield and other parameters associated with pear trees. **Engenharia Agrícola**, 32 (2), pp. 381-392.

Landis, J. R., Koch, G. G. (1977). The measurement of observer agreement for categorical data. **Biometrics**, 33 (1), pp. 159-174.

Lark, R. M. (2000). Estimating variograms of soil properties by the method-of-moments and maximum likelihood. **European Journal of Soil Science**, 51, pp. 717-728.

Larscheid G., Blackmore, B. S. (1996). Interactions between farm managers and information systems with respect to yield mapping. In: **International Conference on Precision Agriculture**, 3. Springer, Minneapolis, pp.1153-1163.

Leroux, C., Jones, H., Pichon, L., Guillaume, S., Lamour, J., Taylor, J., Naud, O., Crestey, T., Lablee, J., Tisseyre, B. (2018). Geofis: an open source, decision-support tool for precision agriculture data. **Agriculture**, 8 (6), pp. 14-21.

Levine N. (2004). **CrimeStat III**: a spatial statistics program for the analysis of crime incident locations. Ned Levine & Associates, Houston, National Institute of Justice, Washington.

Lewis, J., Fowler, M. (2014). **Microservices**. martinFowler.com, accessed 20 November 2020, <http://martinfowler.com/articles/microservices.html>.

Li, Y., Shi, Z., Li, F., Li, H. Y. (2007). Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. **Computers and Electronics in Agriculture**, 56 (2), pp. 174-186.

Liu, Q., Chu, X., Xiao, J., Zhu, H. (2014). Optimizing Non-orthogonal Space Distance Using PSO in Software Cost Estimation. In: **IEEE 38th Annual Computer Software and Applications Conference, 38.** IEEE, Vasteras, pp. 21-26.

Lowrance, C., Fountas, S., Liakos, V., Vellidis, G. (2016). EZZone – An Online Tool for Delineating Management Zones. In: **Proceedings of the 13th International Conference on Precision Agriculture**, St. Louis, pp. 1-7.

Luzardo, A. J. R., Castañeda Filho, R. M., Rubim, I. B. (2017). Análise espacial exploratória com o emprego do índice de Moran. **GEOgraphia**, 19 (40), pp. 161-179.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In: **Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability**, 1. University of California Press, Berkeley, pp. 281–297.

Mahalanobis, P. C. (1936). On the Generalized Distance in Statistics. In: **Proceedings of the National Institute of Science of India**, 2, pp. 49-55.

Malavolta, E. (1992). **ABC da análise de solos e folhas**: amostragem, interpretação e sugestões de adubação. Ceres, São Paulo. 124p.

Mallarino, A. P., Wittry, D. J. (2004). Efficacy of grid and zone soil sampling approaches for site-specific assessment of phosphorus, potassium, pH, and organic matter. **Precision Agriculture**, 5 (2), pp. 131-144.

Maragos, P. (2009). Chapter 13 - Morphological Filtering. Bovik, A. (Eds.), **The Essential Guide to Image Processing**. Academic Press, pp. 293-321.

Martínez-Casasnovas, J. A., Escolà, A., Arnó, J. (2018). Use of Farmer Knowledge in the Delineation of Potential Management Zones in Precision Agriculture: A Case Study in Maize (Zea mays L.). **Agriculture**, 8 (6), p. 1-18.

Matese, A., Toscano, P., Di Gennaro, S. F., Genesio, L., Vaccari, F. P., Primicerio, J., Belli, C., Zaldei, A., Bianconi, R., Gioli, B. (2015). Intercomparison of UAV, Aircraft, and Satellite Remote Sensing Platforms for Precision Viticulture. **Remote Sensing**, 7 (3), pp. 2971-2990.

Matheron, G. (1963). Principles of geostatistics. **Economic Geology**, 58 (8), p. 1246-1266.

Mazzini, P. L. F., Schettini, C. A. F. (2009). Avaliação de metodologias de interpolação espacial aplicadas a dados hidrográficos costeiros quase sinópticos. **Brazilian Journal of Aquatic Science and Technology**, 13 (1), pp. 53-64.

McBratney, A. B., Moore, A. W. (1985). Application of fuzzy sets to climatic classification. Agricultural and Forest Meteorology. **Goettingen**, 35 (1-4), pp. 165-185.

McBride, W. D., Daberkow, S. G. (2003). Information and the adoption of precision farming technologies. **Journal of Agribusiness**, 21 (1), pp. 21–38

Mendes, A. M. S. (2007). **Introdução a fertilidade do solo**. Superintendência Federal de Agricultura, Pecuária e Abastecimento do Estado da Bahia – SFA - BA/SDC/MAPA. p. 64.

Medici, M., Pedersen, S. M., Canavari, M., Anken, T., Stamatelopoulos, P., Tsiropoulos, Z., Zotos, A., Tohidloo, G. (2021). A web-tool for calculating the economic performance of precision agriculture technology. **Computers and Electronics in Agriculture**, 181.

Méndez-Vázquez, L. J., Lira-Noriega, A., Lasa-Covarrubias, R., Cerdeira-Estrada, S. (2019). Delineation of site-specific management zones for pest control purposes: Exploring precision agriculture and species distribution modeling approaches. **Computers and Electronics in Agriculture**, 167, pp. 1-15.

Michelon, G. K., Bazzi, C. L., Upadhyaya, S., Souza, E. G., Magalhães, P. S. G., Borges, L. F., Schenatto, K., Sobjak, R., Gavioli, A., Betzek, N. M. (2019). Software AgDataBox-Map to precision agriculture management. **SoftwareX**, 10.

Milligan, G. W., Cooper, M. C. (1988). A study of standardization of variables in cluster analysis. **Journal of Classification**, 5 (2), pp. 181–204.

Minasny, B., McBratney, A. B. (2002). **FuzME version 3**. Australian Centre for Precision Agriculture, The University of Sydney.

Minasny, B., McBratney, A. B. (2005). The Matern function as a general model for soil variograms. **Geoderma**, 128 (3–4), pp. 192-207.

Mintert, J., Widmar, D., Langemeier, M., Boehlje, M., Erickson, B. (2016). The challenges of precision agriculture: is big data the answer. In: **Southern Agricultural Economics Association Annual Meeting**, San Antonio, Texas, pp. 1–9.

Mohanraj, I., Ashokumar, K., Naren, J. (2016). Field Monitoring and Automation Using IOT in Agriculture Domain. **Procedia Computer Science**, 93 (1), pp. 931-939.

Molin, J. P. (2002). Definição de unidades de manejo a partir de mapas de produtividade. **Engenharia Agrícola**, 22 (1), pp. 83-92.

Molin, J. P., Bazame, H. C., Maldaner, L., Corredo, L. P., Martello, M., Canata, T. F. (2020). Precision agriculture and the digital contributions for site-specific management of the fields. **Revista Ciência Agronômica**, 51 (e20207720).

Molin, J. P., Tavares, T. R. (2019). Sensor systems for mapping soil fertility attributes: challenges, advances, and perspectives in Brazilian tropical soils. **Engenharia Agrícola**, 39 (special issue), pp. 126-147.

Molin, J. P., Vieira Junior, P. A., Dourado Neto, D., Faulin, G. D. C., Mascarin, I. (2007). Variação espacial na produtividade de milho safrinha devido aos macronutrientes e à população de plantas. **Revista Brasileira De Milho e Sorgo**, 6 (3), pp. 309–324.

Mondal, P., Basu, M. (2009). Adoption of PA technologies in India and Some Developing Countries: scope, present status and strategies. **Pro Natu Sci.**, 19, pp. 659-666.

Moral, F. J., Terrón, J. M., Silva, J. R. M. (2010). Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. **Soil Tillage Research**, 106 (2), pp. 335-343.

Moshia, M. E., Khosla, R., Longchamps, L., Reich, R., Davis, J. G., Westfall, D. G. (2014). Precision manure management across site-specific management zones: grain yield and economic analysis. **Agronomy Journal**, 106 (6), pp. 2146-2156.

Musat, G. A., Colezea, M., Pop, F., Negru C., Mocanu, M., Esposito, C., Castiglione, A. (2018). Advanced services for efficient management of smart farms. **Journal of Parallel and Distributed Computing**, 116, pp. 3–17.

Nóbrega, L., Tavares, A., Cardoso, A., Gonçalves, P. (2018). Animal monitoring based on IoT technologies. In: **IoT Vertical and Topical Summit on Agriculture.** IEEE, Tuscany, pp.1–5.

Niero, L. A. C., Dechen, S. C. F., Coelho, R. M., Maria, I. C. de. (2010). Avaliações visuais como índice de qualidade do solo e sua validação por análises físicas e químicas em um Latossolo Vermelho distroférrico com usos e manejos distintos. **Revista Brasileira de Ciência do Solo Viçosa**, 34 (4), pp. 1271-1282.

Odeh, I. O. A., McBratney, A. B., Chittleborough, D. J. (1992). Soil pattern recognition with fuzzy-c-means: application to classification and soil–landform interrelationships. **Soil Science Society of America Journal**, 56 (1), pp. 505-516.

Ohana-Levi, N., Bahat, I., Peeters, A., Shtein, A., Netzer, Y., Cohen, Y., Ben-Gal, A. (2019). A weighted multivariate spatial clustering model to determine irrigation management zones. **Computers and Electronics in Agriculture**, 162, pp. 719-731.

Oldoni, H., Terra, V. S. S., Timm, L. C., Reisser Júnior, C., Monteiro, A. B. (2019). Delineation of management zones in a peach orchard using multivariate and geostatistical analyses. **Soil and Tillage Research**, 191, p. 1-10.

Oliveira Junior, J. C. de, Souza, L. C. P., Melo, V. F. (2010). Variabilidade de atributos físicos e químicos de solos da formação guabirotuba em diferentes unidades de amostragem. **Revista Brasileira de Ciência do Solo**, 34 (5), pp. 1491-1502.
Oliver, M. A. (2010). An overview of geostatistics and precision agriculture. In: Oliver, M. A. (Eds.), **Geostatistical Applications for Precision Agriculture**. Springer, Dordrecht, pp. 1–32.

Oliver, M. A., Webster, R. (2015). **Basic steps in geostatistics**: the variogram and Kriging. Springer-Verlag, London.

Paccioretti, P., Córdoba, M., Balzarini, M. (2020). FastMapping: Software to create field maps and identify management zones in precision agriculture. **Computers and Electronics in Agriculture**, 175, p. 1-7.

Pathak, H. S., Brown, P., Best, T. (2019). A systematic literature review of the factors affecting the precision agriculture adoption process. **Precision Agriculture**, 20, pp. 1292–1316.

Paraforos, D. S., Vassiliadis, V., Kortenbruck, D., Stamkopoulos, K., Ziogas, V., Sapounas, A. A., Griepentrog, H. W. (2017). Multi-level automation of farm management information systems. **Computers and Electronics in Agriculture**, 142 (1), pp. 504–514.

Pham, X., Stack, M. (2018). How data analytics is transforming agriculture. **Business Horizons**, 61 (1), pp. 125-133.

Pedroso, M., Taylor, J., Tisseyre, B., Charnomordic, B., Guillaume, S. (2010). A segmentation algorithm for the delineation of management zones. **Computer and electronics in agriculture**, 70 (1), pp. 199-208.

Peralta, N. R., Costa, J. L., Balzarini, M., Franco, M. C., Córdoba, M., Bullock, D. (2015). Delineation of management zones to improve nitrogen management of wheat. **Computers and Electronics in Agriculture**, 110 (1), pp. 103–113.

POTAFOS, Instituto da Potassa & Fosfato. (1998). **Manual internacional de fertilidade do solo**, 2. POTAFOS, Piracicaba. 177p.

R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.

Reich, R. M. (2008). **Spatial Statistical Modeling of Natural Resources**. Colorado State University, Fort Collins.

Reichert, J. M., Suzuki, L. E. A. S., Reinert, D. J., Horn, R., Håkansson, I. (2009). Reference bulk density and critical degree of compactness for no-till crop production in subtropical highly weathered soils. **Soil & Tillage Research**, 102, pp. 242-254.

Reza, S. K., Sarkar, D., Daruah, U., Das, T. H. (2010). Evaluation and comparison of ordinary Kriging and inverse distance weighting methods for prediction of spatial variability of some chemical parameters of Dhalai district, Tripura. **Agropedology**, 20 (1), pp. 38-48.

Richardson, C. (2018). **Microservices Patterns**. Manning Publications, 520 p.

Richardson, L., Ruby, S. (2007). **RESTful Web Services**, 1. O'Reilly. 419 p.

Rodríguez, S., Gualotuña, T., Grilo, C. (2017). A System for the Monitoring and Predicting of Data in Precision Agriculture in a Rose Greenhouse Based on Wireless Sensor Networks. **Procedia Computer Science**, 121 (1), pp. 306-313.

Ronquim, C. C. (2010). **Conceitos de fertilidade do solo e manejo adequado para as regiões tropicais**. Embrapa, Boletim de Pesquisa e Desenvolvimento, 8.

Rousseeuw, P. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. **Journal of Computational and Applied Mathematics**, 20 (1), pp. 53-65.

Rupnik, R., Kukar, M., Vračar, P., Košir, D., Pevec, D., Bosnić, Z. (2019). AgroDSS: A decision support system for agriculture and farming. **Computers and Electronics in Agriculture**, 161, pp. 260-271.

Ryu, M., Yun, J., Miao, T., Ahn, I. Y., Choi, S. C., Kim, J. (2015). Design and implementation of a connected farm for smart farming system. In: **IEEE Sensors**, Busan, South Korea, pp. 1-4.

Salehi, S., Rezaei-Moghaddam, K., Ajili, A. (2008). Application of yield monitoring technologies: a model for sustainable agriculture. **Iran Agric Exten Edu**, 4 (1), pp. 15-32.

Santos, J. R., Bicudo, S. J., Nakagawa, J., Albuquerque, A. W. de, Cardoso, C. L. (2006). Atributos químicos do solo e produtividade do milho afetados por corretivos e manejo do solo. **Rev. bras. eng. agríc. ambient.**, 10 (2), pp. 323-330.

Santos, R. T., Saraiva, A. M. (2015). A Reference Process for Management Zones Delineation in Precision Agriculture. **IEEE Latin America Transactions**, 13 (3), pp. 727-738.

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., Er, M. J., Ding, W., Lin, C. T. (2017). A review of clustering techniques and developments. **Neurocomputing**, 267 (1), pp. 664-681.

Schenatto, K., Souza, E. G., Bazzi, C. L., Betzek, N. M., Gavioli, A., Beneduzzi, H. M. (2017a). Use of the farmer's experience variable in the generation of management zones. **Semina: Ciências Agrárias**, 38 (4), pp. 2305-2321.

Schenatto, K., Souza, E. G., Bazzi, C. L., Gavioli, A., Betzek, N. M., Beneduzzi, H. B. (2017b). Normalization of data for delineating management zones. **Computers and Electronics in Agriculture**, 143 (1), pp. 238-248.

Schenatto, K., Souza, E. G., Bazzi, C. L., Gavioli, A, Michelon, G. K. (2017c). Software de gerenciamento de dados agrícola: AGDATAFIELD_MOBILE. In: Rosalen, D. L., erbato, C., Turco, J. E. P (Eds.), **A importância da Engenharia Agrícola para a segurança alimentar,** 1. Sociedade Brasileira de Engenharia Agrícola, pp. 1-10.

Schepers, A. R., Shanahan, F. J., Liebig, M. A., Schepers, J. S., Johnson, S. H., Luchiari, J. A. (2004). Appropriateness of management zones for characterizing spatial variability of soil properties and irrigated corn yields across years. **Agronomy Journal**, 96 (1), pp. 195–203.

Shen, S., Basist, A., Howard, A. (2010). Structure of a Digital Agriculture System and Agricultural Risks Due to Climate Changes. **Agriculture and Agricultural Science Procedia**, 1 (1), pp. 42-51.

Soldani, J., Tamburri, D. A., Van Den Heuvel, W. J. (2018). The pains and gains of microservices: A Systematic grey literature review. **Journal of Systems and Software**, 146 (1), pp. 215–232.

Sousa, D. M. G., Lobato, E. (1986). Adubação fosfatada. In: **Savana: alimento e energia**, EMBRAPA-CPAC, Planaltina, pp. 33-60.

Souza, D. M. G., Ritchey, K. D. (1986). Uso do gesso no solo do Cerrado. In: **Seminário sobre uso do fosfogesso na agricultura**, 1. EMBRAPA, Brasília, pp. 119-144.

Souza, E. G., et al. Comparison of yield maps of three fields. **Computers and Electronics in Agriculture,** 2021. (in analysis).

Souza, E. G., Bazzi, C. L., Khosla, R., Uribe-Opazo, M. A., Reich, R. M. (2016). Interpolation type and data computation of crop yield maps is important for precision crop production. **Journal of Plant Nutrition**, 39 (4), pp. 531-538.

Souza, E. G., Schenatto, K., Bazzi, C. L. (2018). Creating thematic maps and management zones for agriculture fields. In: **Proceedings of the 14th International Conference On Precision Agriculture** (IPCA).

Spilke, J., Zürnstein, K. (2005). Webservices - Beschreibung eines Ansatzes zur Anwendungskopplung und von Nutzensmöglichkeiten im Agrarbereich. **Zeitschrift für Agrarinformatik**, 13 (2), pp. 33–40.

Stach, A. (2007). Temporal variability of the spatial structure of maximum daily precipitation totals. **Monitoring Środowiska Przyrodniczego**, 8, pp. 73–90.

Stafford, J. V. (2000). Implementing precision agriculture in the 21th century. **J. Agric. Eng. Res.,** 76, pp. 267–275.

Swindell, J. (1997). Mapping the spatial variability in the yield potential of arable land through GIS analysis of sequential yield maps. In 1st European Conference on Precision Agriculture (pp. 827- 834). Warwick.

Taneja, M., Byabazaire, J., Davy, A., Olariu, C. (2018). Fog assisted application support for animal behaviour analysis and health monitoring in dairy farming. In: **IEEE World Forum on Internet of Things** (WF-IoT), 4. Singapore, pp. 819-824.

Taylor, J. A., McBratney A. B., Whelan B. M. (2007). Establishing management classes for broadacre agricultural production. **Agronomy Journal**, 99 (5), pp. 1366-1376.

Taylor, J. C., Wood, G. A., Earl, R., Godwin, R. J. (2003). Soil Factors and their Influence on Within-field Crop Variability, Part II: Spatial Analysis and Determination of Management Zones. **Biosystems Engineering**, 84 (4), pp. 441-453.

Uddin, M. A., Mansour, A., Le Jeune, D., Aggoune, E. H. M. (2017). Agriculture Internet of things: AG-IoT. In: **International Telecommunication Networks and Applications Conference** (ITNAC), 27. pp. 1-6.

United Nations. (2019). **World Population Prospects 2019:** Highlights. Department of Economic and Social Affairs – Population Division – ST/ESA/SER.A/423, New York.

Uribe-Opazo, M. A., Borssoi, J. A., Galea, M. (2012). Influence diagnostics in Gaussian spatial linear models. **Journal of Applied Statistics** 39 (3), pp. 615–630.

Van Raij, B. (2011). **Fertilidade do solo e manejo de nutrientes**. International Plant Nutrition Institute. 420 p.

Vieira, S. R. (2000). Geoestatística em estudos de variabilidade espacial do solo. In: Novais, R. F. de, Alvarez V., V. H., Schaefer, C. E. G. R. (Eds.), **Tópicos em ciência do solo**. Sociedade Brasileira de Ciência do Solo, Viçosa.

Viscarra Rossel, R. A., Adamchuk, V. I., Sudduth, K. A., McKenzie, N. J., Lobsey, C. (2011). Proximal soil sensing: an effective approach for soil measurements in space and time. **Adv. Agron.**, 113, pp. 243-291.

Vitti, G. C., Trevisan, W. (2000). Manejo de macro e micronutrientes para alta produtividade da soja. **Informações Agronômicas**, 90, pp. 1-16.

Webster, R. (1985). Quantitative spatial analysis of soil in the field. In: Stewart, B. A. (Ed.), **Advance in soil science**, 3. Spriger-Verlag, New York.

Whetton, R. L., Waine, T. W., Mouazen, A. M. (2018). Evaluating management zone maps for variable rate fungicide application and selective harvest. **Computers and Electronics in Agriculture**, 153, pp. 202-212.

Wikle, C. K., Zammit-Mangion, A., Cressie, N. (2019). **Spatio-Temporal Statistics with R**. Chapman & Hall/CRC, Boca Raton, FL.

Wolfert, S., Ge, L., Verdouw, C., Bogaardt, M. (2017). Big data in smart farming: a review. **Agricultural Systems**, 153 (1), pp. 69–80.

Wollenhaupt, N. C., Wolkowski, R. P., Clayton, M. K. (1994). Mapping soil test phosphorus and potassium for variable-rate fertilizer application. **Journal of Production Agriculture**, 7 (4), pp. 441-448.

Xiang, L., Yu-Chun, P., Zhong-Qiang, G., Chun-Jiang, Z. (2007). Delineation and Scale Effect of Precision Agriculture Management Zones Using Yield Monitor Data Over Four Years. **Agricultural Sciences In China**, 6 (2), pp. 180-188.

Yahyaoui, H., Own, H. S. (2018). Unsupervised clustering of service performance behaviors. **Information Sciences**, 422 (1), pp. 558-571.

Yan, L., Zhou, S., Feng, L., Hong-Yi, L. (2007). Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. **Computers and Electronics in Agriculture**, 56 (2), pp. 174-186.

Zambon I., Cecchini M., Egidi G., Saporito M. G., Colantoni A. (2019). Revolution 4.0: Industry vs. Agriculture in a Future Development for SMEs. **Processes**, 7 (36), pp. 1-16.

Zhang, N., Wang, M., Wang, N. (2002). Precision agriculture - a worldwide overview. **Computers and Electronics in Agriculture**, 36 (2-3), pp. 113-132.

Zhang, X., Shi, L., Jia, X., Seielstad, G., Helgason, C. (2010). Zone mapping application for precision-farming: a decision support tool for variable rate application. **Precision Agriculture**, 11 (2), pp. 103-114.

5 PAPER 1 – INCORPORATION OF COMPUTATIONAL ROUTINES IN A MICROSERVICE ARCHITECTURE IN AGDATABOX PLATFORM

ABSTRACT: Agriculture has been undergoing a digital process that aims to apply digital technologies to make the sector more productive, profitable, and environmentally responsible. This trend has been adopted since applying precision agriculture (PA) techniques and, more recently, with digital agriculture (DA). DA aims to use all available information and knowledge to enable automation of sustainable processes in agriculture, applying data analysis methods and techniques by specific software and platforms to collect and transform data into meaningful information for agriculture. Platform AgDataBox (ADB) offers a set of tools to allow agriculture specialists to obtain, process, and visualize data for the right decision-making. However, its structure needed to be readjusted in new software architecture to allow aggregation of new functionalities and expand ADB platform. This study aimed to develop a web microservices architecture (ADB-MSA) to incorporate the required functionalities to create thematic maps (TMs) and delineate management zones (MZs). ADB-MSA provided eight microservices, six of which (statistics, spatial, interpolation, clustering, rectification, and lime/nutrient recommendation) execute procedures based on JavaScript, R, and Python programming languages, while the other two are used to store data. In the case study, the procedures to create TMs and delineate MZs were performed with data from one commercial area. Thus, the services provided in the architecture meet the steps of creating TMs and delineating MZs, as MZs for fertilizer application were generated and evaluated according to phosphorus and potassium requirements.

KEYWORDS: precision agriculture, digital agriculture, variable-rate application, ADB.

5.1 Introduction

Computational technologies and global positioning systems have been applied in food production for decades. Currently, agriculture is expected to undergo more fulsome digitalization with sensors to collect data and intelligent machines to mine them (Lajoie-O'Malley et al., 2020). Precision technologies for agriculture have been changing modern agriculture, with important implications for debates on environmental sustainability (Clapp and Ruder, 2020). Advances in digital technologies in recent years have included remote sensing, sophisticated upgrades to variable-rate technologies, robotics, and automated steering machines, unmanned aerial vehicles, wireless communication, and data analytics (Kamilaris et al., 2017). Applying these resources in agriculture is based on the premise that they offer more precision in decision-making and practice (Clapp and Ruder, 2020).

A digital agriculture (DA) scenario integrates multiple components, which play an essential role in collecting, storing, and analyzing agronomic data. Software infrastructure needs to be enabled to allow business demands and fulfill them as fast as possible. These components on DA ecosystem can be developed in different programming languages intended to different operating systems and using different communication protocols, data communication means, and device hardware architecture.

AgDataBox (ADB) is one of such technology that is focused on DA. This platform provides free computational tools for producers, researchers, and service providers, mainly addressing precision agriculture practices. This platform also integrates data, software, procedures, and methodologies to enable the development of agricultural management in Brazil with free technologies.

This web platform has an API for data storage, called ADB Data API (ADB-DATA-API), accessible by the hypertext transfer protocol (HTTP). ADB-DATA-API allows interoperability of different applications in which data is centralized. Some applications under testing or development consume ADB-DATA-API resources: (i) AgDataBox-Mobile (Schenatto et al., 2017b), (ii) AgDataBox-Map (Borges et al., 2020; Michelon et al., 2019), (iii) AgDataBox-Admin, and (iv) AgDataBox-IoT. Fig. 1 shows the ADB web platform architecture.

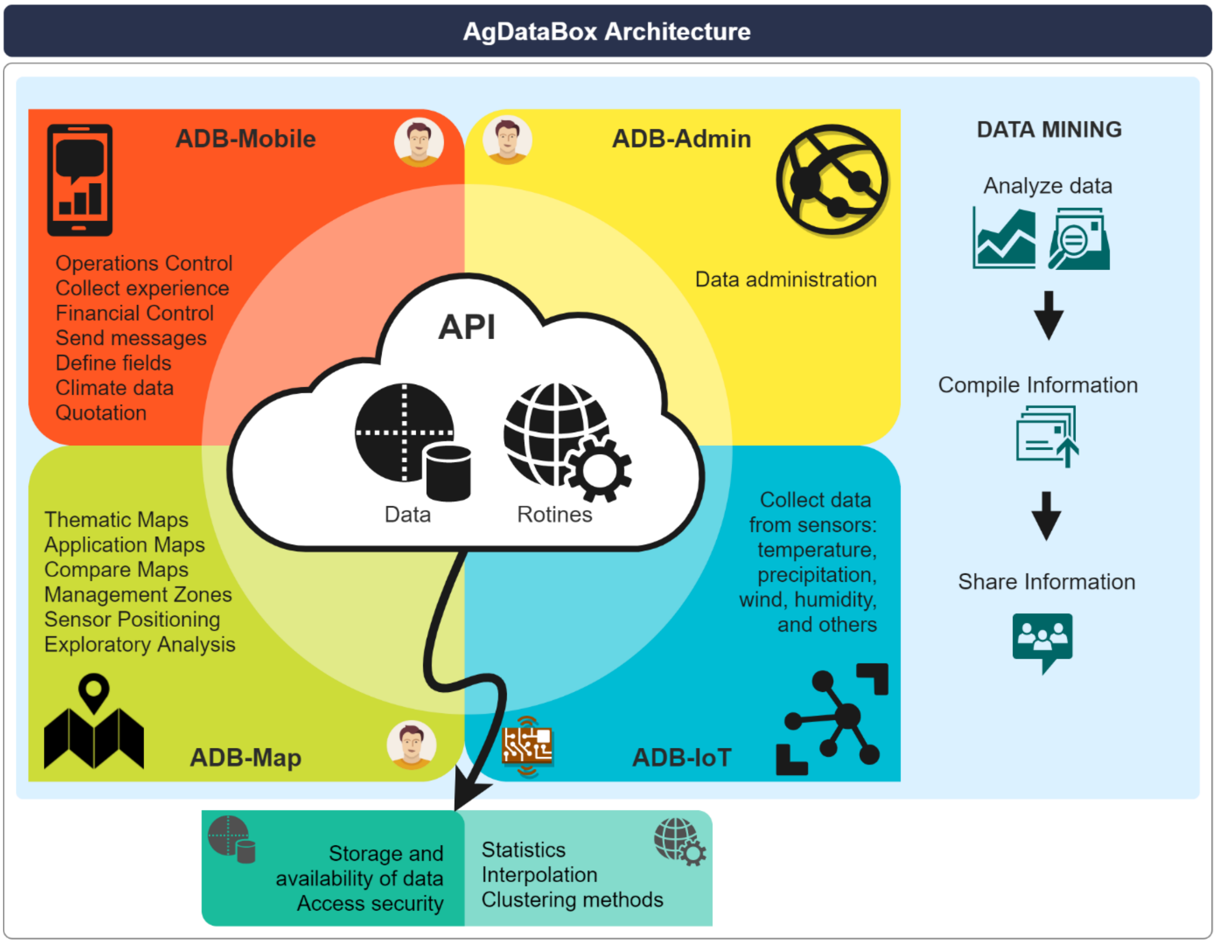


Fig. 1 AgDataBox web platform architecture with its AgDataBox Data API, AgDataBox Mobile, AgDataBox Admin, AgDataBox Map, and AgDataBox IoT applications.

This platform is originated from software to delineate crop management zones (SDUM) (Bazzi et al., 2019b). SDUM was developed on a desktop environment, and its installation requires computers with high processing capacity and memory availability due to the complexity of implemented functionalities. Despite SDUM acceptance by researchers and producers, we opted for migrating to a web platform to include new modules and functionalities while maintaining its free nature.

Web services are software integration technologies that allow software interoperability in a computer network, in which information is exchanged (communication) over HTTP. Thus, this application programming interface (API) defines a standard to represent the exchanged messages, usually in JavaScript object notation (JSON) or extensible markup language (XML) (Grahl et al., 2017). The main advantage of using web services is creating real-time systems using reusable and loosely-coupled software components.

One of the challenges of distributed applications is scalability, related to the software's capacity for growth and expansion due to business demands (Soldani et al., 2018). New software and technical architectures have been studied and proposed to improve the web services scenario. One of them is microservices architecture (MSA). MSA advocates decomposing an application into a set of small services and making them communicate with each other by lightweight mechanisms, such as the RESTful API (Lewis and Fowler, 2014). The application is logically and structurally organized into small cohesive components, and it is seen as a set of small, modular, and loosely coupled services, each one dedicated to a single activity (Ciavotta et al., 2017). Each microservice can be its database and are independent of each other.

Several studies have addressed the migration of applications to MSA (Balalaie et al., 2018; Hassan and Bahsoon, 2016; Lin et al., 2016), focusing on architectural principles, patterns, and practices common to this approach. Migration to microservices has become very popular in recent years. Companies migrate for different reasons, such as improving the quality of software or facilitating its maintenance (Taibi et al., 2017). Software migration from a monolithic architecture to MSA can generate a more significant initial effort, but the complexity of maintaining the code is reduced, and speed is increased in the long run (Lenarduzzi et al., 2020).

This approach has aroused the researchers' interest and software developers to create solutions in the most different fields of activity, such as HazMate (Cherradi et al., 2017), which is a real-time environmental information software to assist in logistics of transporting hazardous cargo in urban areas and the analysis of animals' health, behavior and their monitoring (Taneja et al., 2018). In this sense, ADB structuring in a microservices architecture is fundamental to expand, as data and procedures common to its applications can be shared harmonically. Therefore, this study aimed to develop a microservices architecture integrated into ADB platform and demonstrate its main functionalities to create thematic maps (TMs) and delineate management zones (MZs) available as web services.

5.2 Material and methods

5. 2. 1 Microservices architecture

An MSA was structured to integrate functionalities to generate thematic maps and delineate MZs in ADB platform (Fig. 2). Thus, there is the possibility to provide services to different applications without the need to rewrite much code.

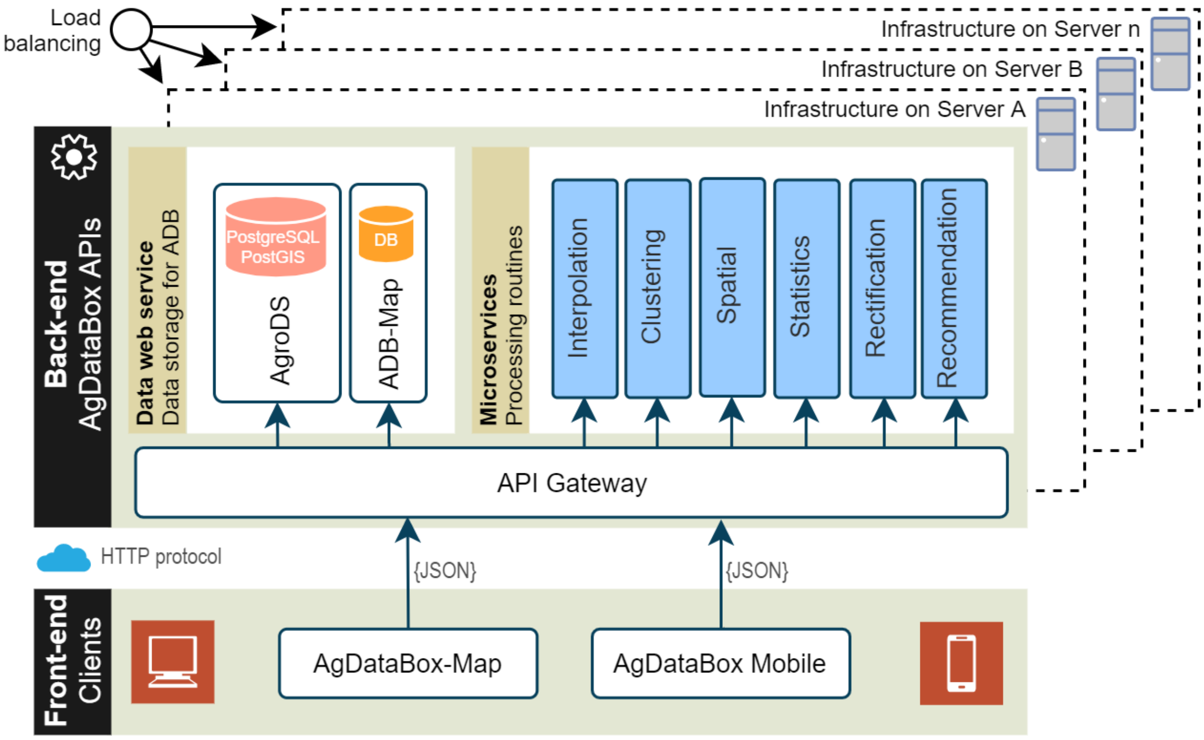


Fig. 2 Representation of the concepts involving the microservices architecture of the AgDataBox platform, where the back-end layer contains the microservices, with routines and data, and the front-end are the applications that consume the microservices.

There is a logical division of the parts that compose the platform. The user interacts with software artifacts called applications, such as ADB-Map and ADB-Mobile, in the front-end layer. On the other hand, other artifacts provide services for executing routines in a remote environment located in the back-end layer. For example, microservices were implemented on the platform's back-end to perform statistical analysis of data, cleaning, and data preparation (ADB Statistics API), analysis and operations with spatial data (ADB Spatial API), data grouping, and cluster evaluation metrics (ADB Clustering API), MZ rectification (ADB Rectification API), fertilizer and lime recommendation (ADB Recommendation API), agricultural data storage (ADB Data API), and data storage for ADB-Map application (ADB-Map API).

An API gateway was implemented to centralize clients' requests and dispatch them to the appropriate microservices. So, NGINX server was used to allow loading balance among servers where services are deployed.

The following technical aspects were considered in microservices architecture:

- Architecture: microservices are available as web APIs in the architectural style of representational state transfer (REST).
- Format for exchanging messages: JSON format represents data among clients' applications and microservices.
- Security: The execution of routines must be restricted only for users registered in ADB-DATA-API. Each microservice can store execution results in its repository and make data available to the user who generated them. Microservices must receive the authentication token generated by ADB-DATA-API to authorize access. The token must be passed on the clients' application at each request.

ADB platform is deployed on a virtualized network server in a private cloud, with a Linux operating system with 16 processor cores and 32 GB RAM. Docker tool is used to streamline services that make up ADB platform and allows a fast deployment of developed software artifacts. Each component of this structure is deployed in a separate container. The advantage was to allow the operating systems independence, programming languages, and libraries within ADB architecture.

Microservices general workflow is based on their interaction with a clients' application to execute routines; the client makes a request and waits for a response from microservice. In this interaction, microservice receives customer's data, standardizes it, executes the requested functionality routine, and delivers the results to the clients' application (Fig. 3).

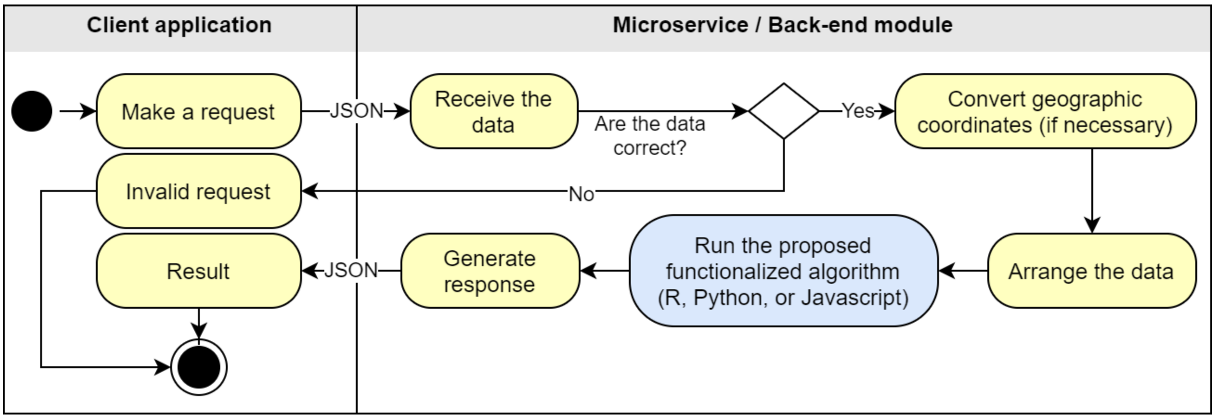


Fig. 3 Unified modeling language (UML) activity diagram demonstrates the general execution flow in the back-end modules in ADB-API.

All ADB-MSA microservices implement APIs in REST architecture. Except for ADB-DATA-API, the other microservices were developed in JavaScript, using Node.js platform and

Express framework. Thus, the features offered in microservices are executed in scripts implemented in statistical software languages R, Python, or JavaScript.

5. 2. 2 API for agricultural data management – AgDataBox Data API

The first version of ADB-DATA-API (Bazzi et al., 2019a) was updated and made available as a microservice in ADB platform. The technologies for software development used in API construction consisted of Java programming language, version 8, with resources of Java Enterprise Edition (JEE) platform; Maven for project management; VRaptor framework to build REST resources and organize flow of requests and responses in API; and Apache Tomcat as the application server. In addition, data are stored in database management system (DBMS) PostgreSQL, with PostGIS extension, for storage of data managed by ADB-DATA-API. This DBMS was also used by Rupnik et al. (2019).

Requests for data manipulation in ADB-API are made by HTTP, using a request method (get, post, put, or delete), a uniform resource identifier (URI), and a data representation in JSON standardized format. The response delivered by the server that hosts ADB-DATA-API is a message containing some information in its headers, such as the response status and the main content requested, also in JSON format.

5. 2. 3 Microservice for statistical analysis

Statistical analyses are available in ADB-Statistics-API microservice (ADB-ST-API) (Fig. 4). The used R statistical software libraries consisted of nortest, ade4, and spdep.

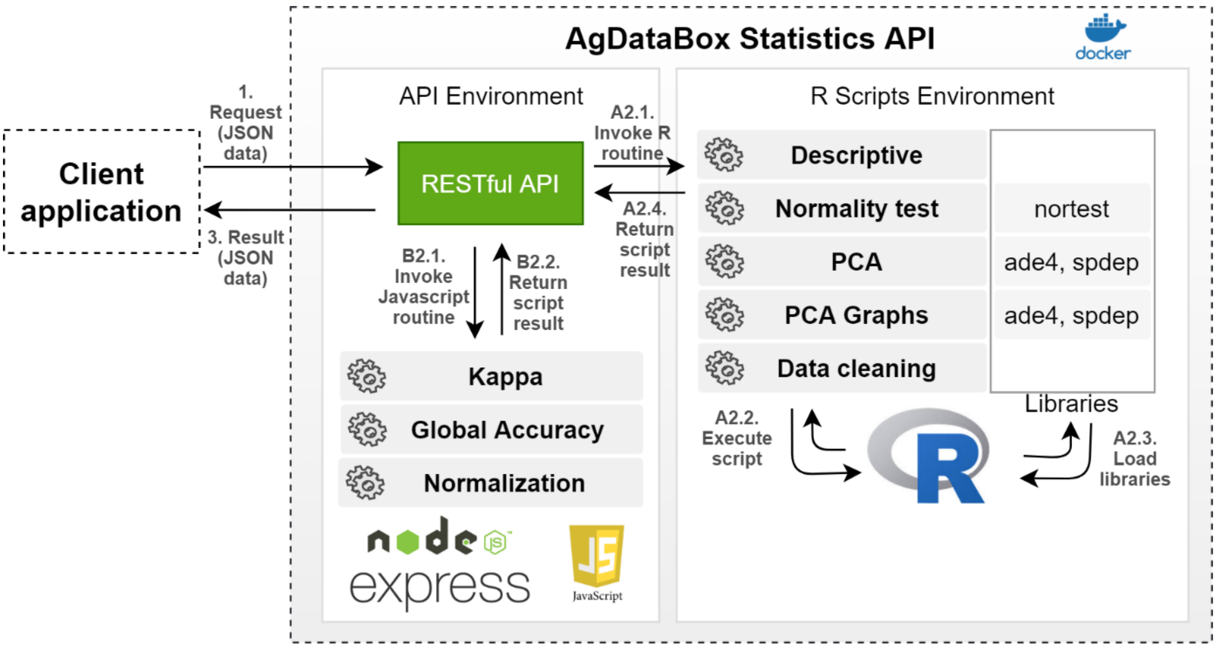


Fig. 4 AgDataBox Statistics API microservice components and workflow.

The following ADB-ST-API services run in R software environment:

- Descriptive statistics: measures of position (mean, median, mode, and quartiles) and dispersion (variance, standard deviation, and coefficient of variation – CV) stand out among the available statistics.
- Normality tests: data normality can be tested by Kolmogorov-Smirnov, Lilliefors, Cramer-Von Mises, Shapiro-Wilk, Shapiro-Francia, and Anderson-Darling tests.
- Data cleaning: it removes values lower than or equal to zero (useful for data collected by harvesters), duplicate data, outliers, and inliers.
- Principal component analysis (PCA) (Hotelling, 1933): it selects variables to be used in MZ delineation (Fraisse et al., 2001; Li et al., 2007; Moral et al., 2010; Cohen et al., 2013).

The following services run in JavaScript environment:

- Data normalization: The (i) range (Equation A1), (ii) mean (Equation A2), (iii) standard score (Equation A3), and (iv) min-max methods (Equation A4) are available.
- Agreement indices:
    - Kappa index (Equation A5) (Cohen, 1960; Congalton, 1991).
    - Global accuracy (GA) (Equation A6).
- Quantitative agreement indices:
    - The coefficient of relative deviation (CRD) (Coelho et al., 2009) Equation A7) calculates the mean difference in modulus of interpolated values on a thematic map compared to a map taken as a reference.
    - The mean absolute difference (MAD) (Equation A8) computes the mean absolute difference among values on both maps.

5. 2. 4 Microservice for analysis with spatial data

Statistical analysis procedures and spatial data preparation were made available in ADB Spatial API microservice (ADB-SP-API) (Fig. 5). The used R software libraries consisted of sp, rgdal, spdep, ade4, and adespatial.

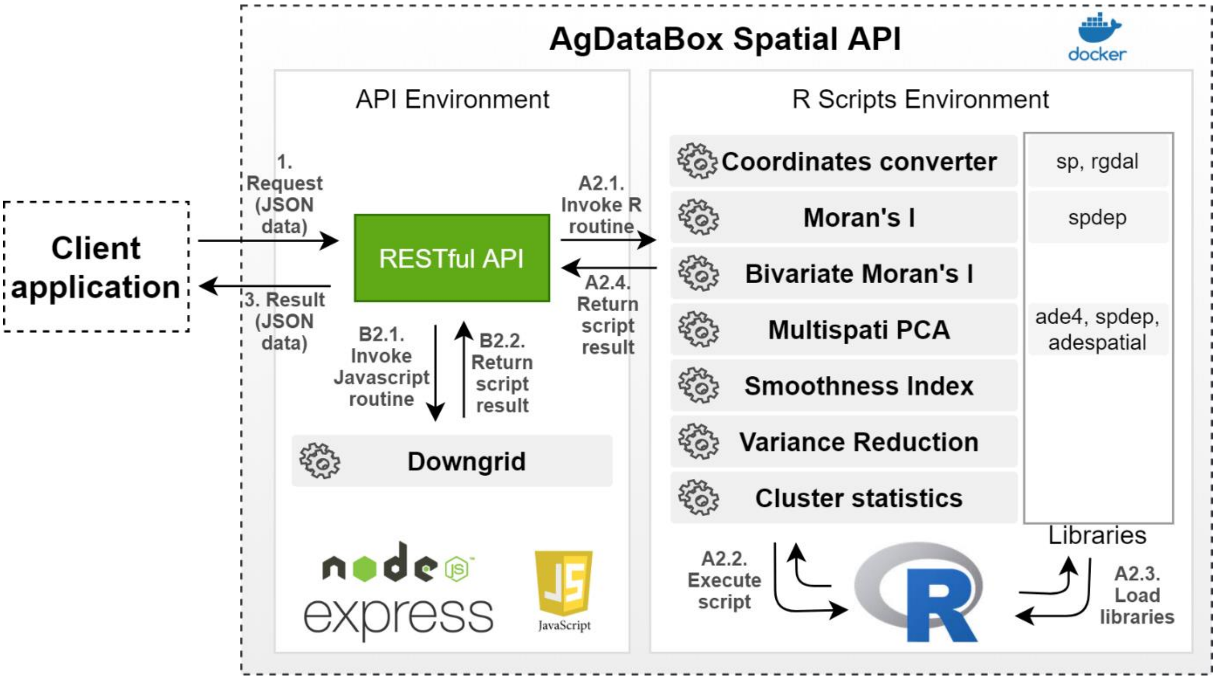


Fig. 5 AgDataBox Spatial API microservice components and workflow.

The following ADB-SP-API services run in R software environment:

- Coordinate's converter: it converts geographic coordinates into another system.
- Bivariate Moran's I (Reich, 2008; Schepers et al., 2004) (Equation A18): it calculates autocorrelation and bivariate correlation between two variables.
- Multivariate spatial analysis based on Moran's index and PCA (MULTISPATI-PCA; MPCA) (Dray et al., 2008): it calculates the spatial principal components (SPCs) of all stable variables.
- Smoothness index (SI%) (Equation A14) (Gavioli et al., 2016): it determines TMs smoothness. An SI closer to 100% indicates higher homogeneity of classes, and an SI closer to 0% indicates a higher presence of random values.
- Variance reduction (VR%) (Equation A9) (Xiang et al., 2007; Schenatto et al., 2017a): It is calculated for a variable, with the expectation that sum of data variances for each MZ is smaller than the total variance of the field.
- Cluster statistics: It calculates descriptive statistical measures of a variable of interest within each MZ class. The number of observations, mean, median, mode, minimum and maximum values, quartiles, standard deviation, variance, CV, skewness, and kurtosis are determined. The analysis of variance (ANOVA) by the Tukey test is used to identify whether sub-regions of design in MZs present significant differences in the mean value of variable of interest.
- Downgrade service: It reduces dataset density using only JavaScript environment.

5. 2. 5 Microservice for data interpolation

ADB Interpolation API (ADB-INT-API) microservice interpolates data by inverse distance weighting (IDW) (Equation A22), ordinary kriging (OK) (Equation A23 – Cressie, 1993), moving average (MA) (Equation A24), and nearest neighbor (NN).

Furthermore, it is possible to perform geostatistical analysis, select the best interpolation method between OK and IDW and determine its interpolation parameters (Fig. 6).

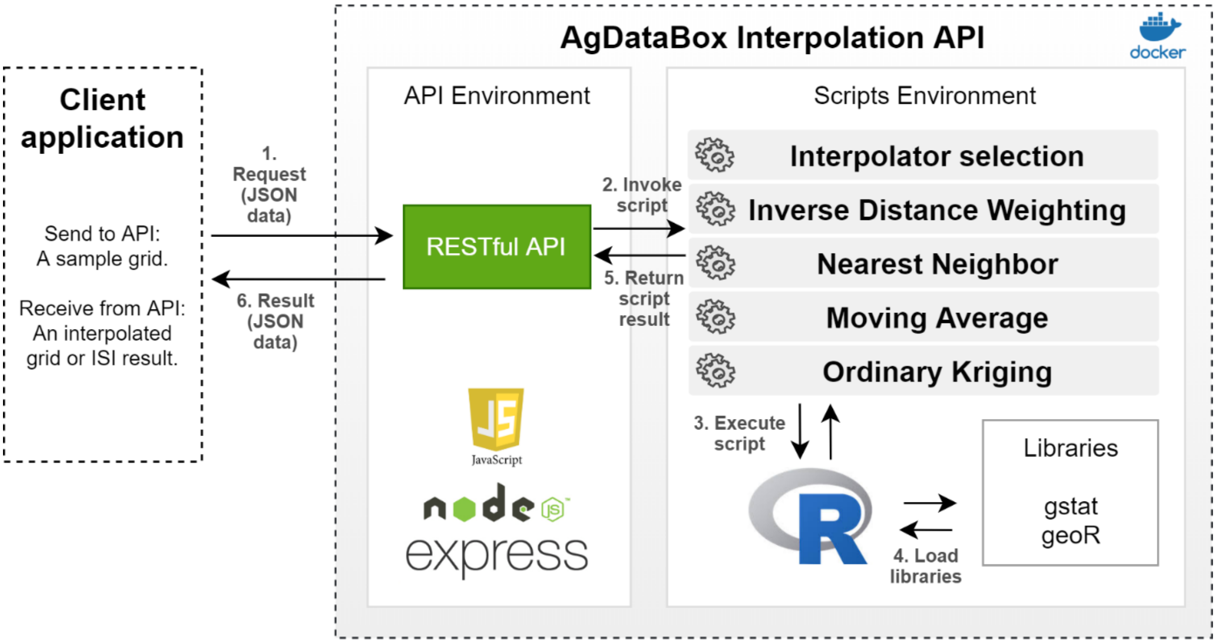


Fig. 6 AgDataBox Interpolation API microservice components and workflow.

Microservice aggregated scripts studied and implemented by Betzek et al. (2019) and improved by us and adding the newly implemented features. Algorithms that make interpolations were developed in R software, using the packages geoR (Ribeiro and Diggle, 2018) for OK, and as functions implemented directly in the PostgreSQL database by the procedural language PL/pgSQL for IDW. The algorithm to perform IDW interpolation based on PL/pgSQL was replaced by an R software script, using gstat package.

The interpolator selection is performed from the interpolator selection index (ISI) (Bier and Souza, 2017) (Equation A19), which allows determining the best semivariogram model and its parameters, as well as the best exponent and number of neighbors to be used in the IDW interpolator. The mean error (ME) (Equation A20) and the standard deviation of the mean error (SDME) (Equation A21) are calculated by cross-validation (Isaaks and Srivastava, 1989). In geostatistical analysis, computational routine tests fourteen different models (i) seven different semivariogram models (spherical, gaussian, exponential, Matérn 0.5, Matérn 1.0, Matérn 1.5, and Matérn 2.0), and (ii) two statistical methods to optimize semivariogram fit (ordinary least squares (OLS) and weighted least squares (WLS – Cressie, 1985)). In IDW analysis, the routine analyzes, by default, 88 analyses (i) twelve different values for the exponent (0.5, 1.0, 1.5, …, and 6.0), and (ii) the number of neighbors from 4 to 12. In geostatistics, the lag size is defined from the calculated amount of lags, relationship between cutoff, and the shortest distance among pairs of points. A significant limitation to address in this ADB-INT-API version is that anisotropy's eventual presence is not considered. Computational routines to calculate ISI were improved from Betzek et al. (2019) study.

5. 2. 6 Microservice for data clustering

The seventeen data clustering methods (Table 1) evaluated by Gavioli et al. (2019) were made available in ADB Clustering API (ADB-CLU-API) (Fig. 7). These methods showed the best results in the carried out research. Data clustering algorithms use R software libraries cclust, cluster, e1071, fastcluster, fclust, hybridHclust, optpart, and skmeans.

Table 1 Clustering methods available in the AgDataBox Clustering API

Methods	References
average linkage[a]	Jain and Dubes (1988)
centroid linkage[a]	Jain and Dubes (1988)
complete linkage[a]	Jain and Dubes (1988)
divisive analysis (diana)[a]	Kaufman and Rousseeuw (1990)
hybrid hierarchical clustering[a]	Chipman and Tibshirani (2006)
median linkage[a]	Jain and Dubes (1988)
McQuitty's method (mcquitty)[a]	McQuitty (1966)
Ward's method (ward)[a]	Ward (1963)
single linkage[a]	Jain and Dubes (1988)
bagged clustering[b]	Leisch (1999)
clustering large applications (clara)[b]	Kaufman and Rousseeuw (1990)
fuzzy analysis clustering (fanny)[b]	Kaufman and Rousseeuw (1990)
fuzzy c-means[b]	Bezdek (1981)
fuzzy c-shells[b]	Dave (1992)
hard competitive learning[b]	Xu and Wunsch (2009)
k-means[b]	MacQueen (1967)
neural gas[b]	Martinetz et al. (1993)
partitioning around medoids[b]	Kaufman and Rousseeuw (1990)
spherical k-means[b]	Dhillon and Modha (2001)
unsupervised fuzzy competitive learning[b]	Pal et al. (1996)

[a]: hierarchical method; [b]: partitioning method.

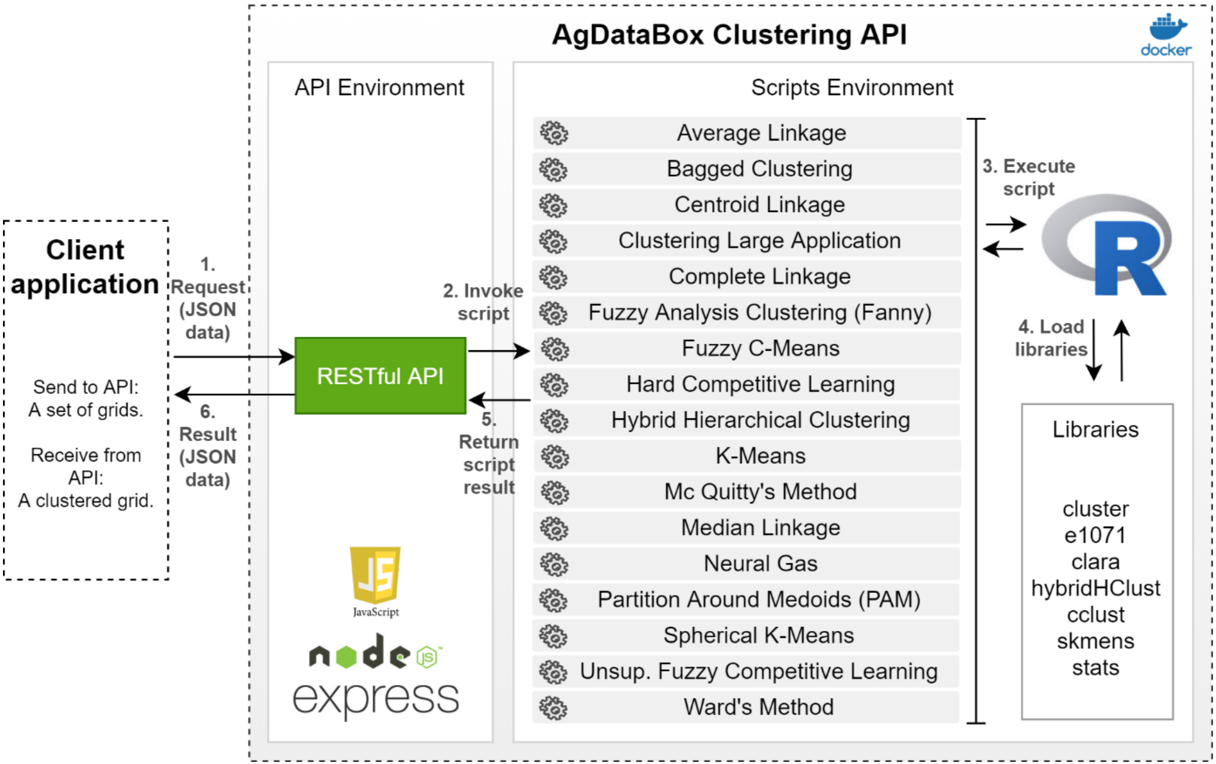


Fig. 7 AgDataBox Clustering API microservice components and workflow.

The indices for MZ quality evaluation obtained in this microservice are:

- Fuzziness performance index (FPI) (Equation A10) (McBratney and Moore, 1985; Fridgen et al., 2004): It measures the degree of separation among the fuzzy c groups generated from a data set. FPI varies from 0 to 1.
- Modified partition entropy (MPE) (Equation A11) (McBratney and Moore, 1985; Fridgen et al., 2004): It estimates the level of difficulty of c groups organization.
- Average silhouette coefficient (ASC) (Equation A12) (Rousseeuw, 1987): It measures the level of satisfactory internal formation and external separation among clusters.
- Improved cluster validation index (ICVI) (Equation A13) (Gavioli et al., 2016): It is a composition of FPI, MPE, and VR% indices.

5. 2. 7 Microservice for rectification of management zones

ADB Rectification API (ADB-RCT-API) microservice allows MZs rectification. The implemented rectification methods are based on median, erosion, and dilation morphological filters. Rectification algorithms were implemented in Python, using the OpenCV digital image processing library (Fig. 8).

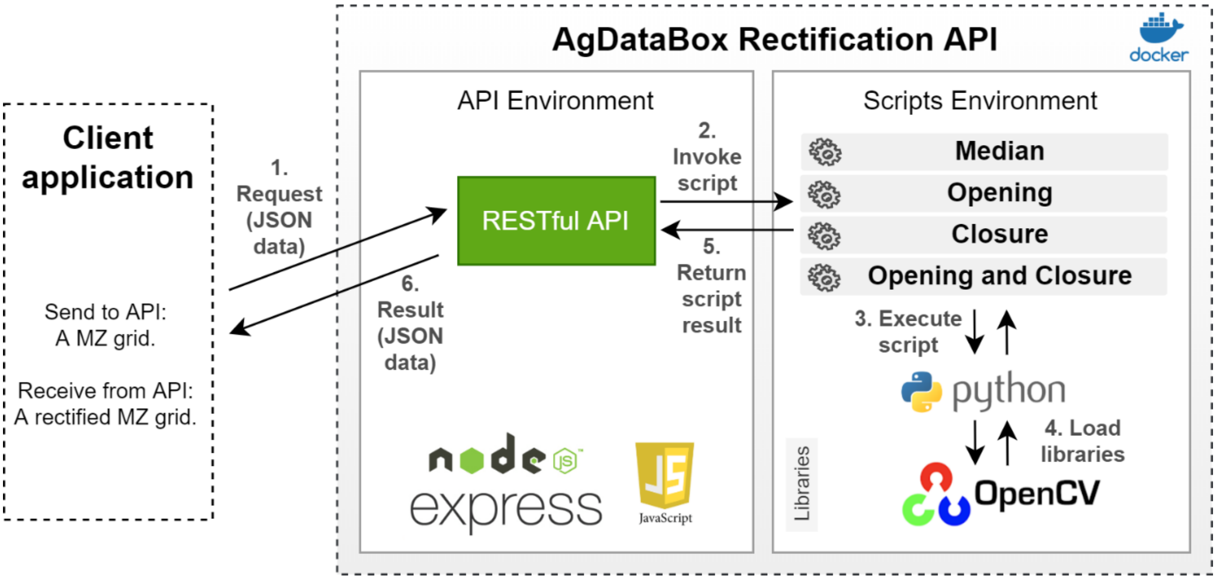


Fig. 8 AgDataBox Rectification API microservice components and workflow.

5. 2. 8 Microservice for calculation of nutrient and lime requirement

ADB Recommendation API (ADB-REC-API, Fig. 9) microservice calculates nutrient (N, P, and K) and lime requirements.

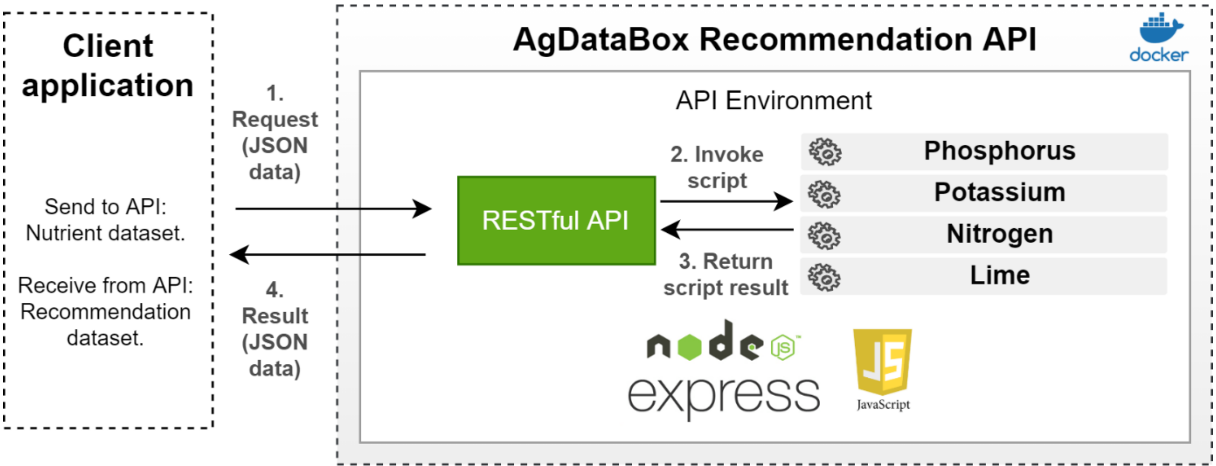


Fig. 9 AgDataBox Recommendation API microservice components and workflow.

The nutrient recommendation is based on Beneduzzi (2020) study, in which N recommendation is performed by yield expectation model for corn cropping, considering soil organic matter content (OM%). Two methods were implemented for P and K: soil nutrient availability and yield expectation, with the recommendation calculation for soybean and corn (Table 2). The recommendation for each nutrient is based on available fertilizers.

Table 2 Fertilizers used in the recommendation of nutrients, crops, and recommendation methods available in AgDataBox Recommendation API

Nutrient	Culture	Method	Fertilizer
N	Corn	YEOM	Urea (UR) Ammonium sulfate (AM) Ammonium nitrate (AN) Ammonium chloride (AC)
P	Corn and Soybean	A and YE	Simple Superphosphate (SS) Triple superphosphate (ST) Monoammonium phosphate (MAP) Daimonic phosphate (DAP) Araxa phosphate (ARAD)
K	Corn and Soybean	A and YE	Potassium chloride (KCL) Potassium sulfate (PS) Potassium and magnesium sulfate (PMS) Potassium double saltpeter (PDS)

N: nitrogen; P: phosphorus; K: potassium; YEOM: yield expectation considering the content of organic matter of the soil; A: availability; YE: yield expectation.

Lime recommendation is performed based on Moreira (2019) and uses the following methods: (i) exchangeable aluminum neutralization, (ii) exchangeable Al^{3+} neutralization, and increase in base cations (Ca^{2+} and Mg^{2+}), and (iii) base saturation.

5. 2. 9 Case study

Data from a 20-ha commercial agricultural field located in the municipality of Serranópolis do Iguaçu, Paraná state, Southern Brazil, with central coordinates −54.01232307º (longitude) and −25.39526307º (latitude) in Datum WGS 1984, were used to demonstrate some of MSA functionalities of ADB platform (Fig. 10). The sampling points with irregular distances were located along an imaginary line among the level curves following the terrain topography. The sample density of 2.6 ha^{-1} attends the suggestion of a minimum density from 1 sample ha^{-1} (Ferguson and Hergert, 2009) to 2.5 sample ha^{-1} (Journel and Huijbregts, 1978; Doerge, 2000).

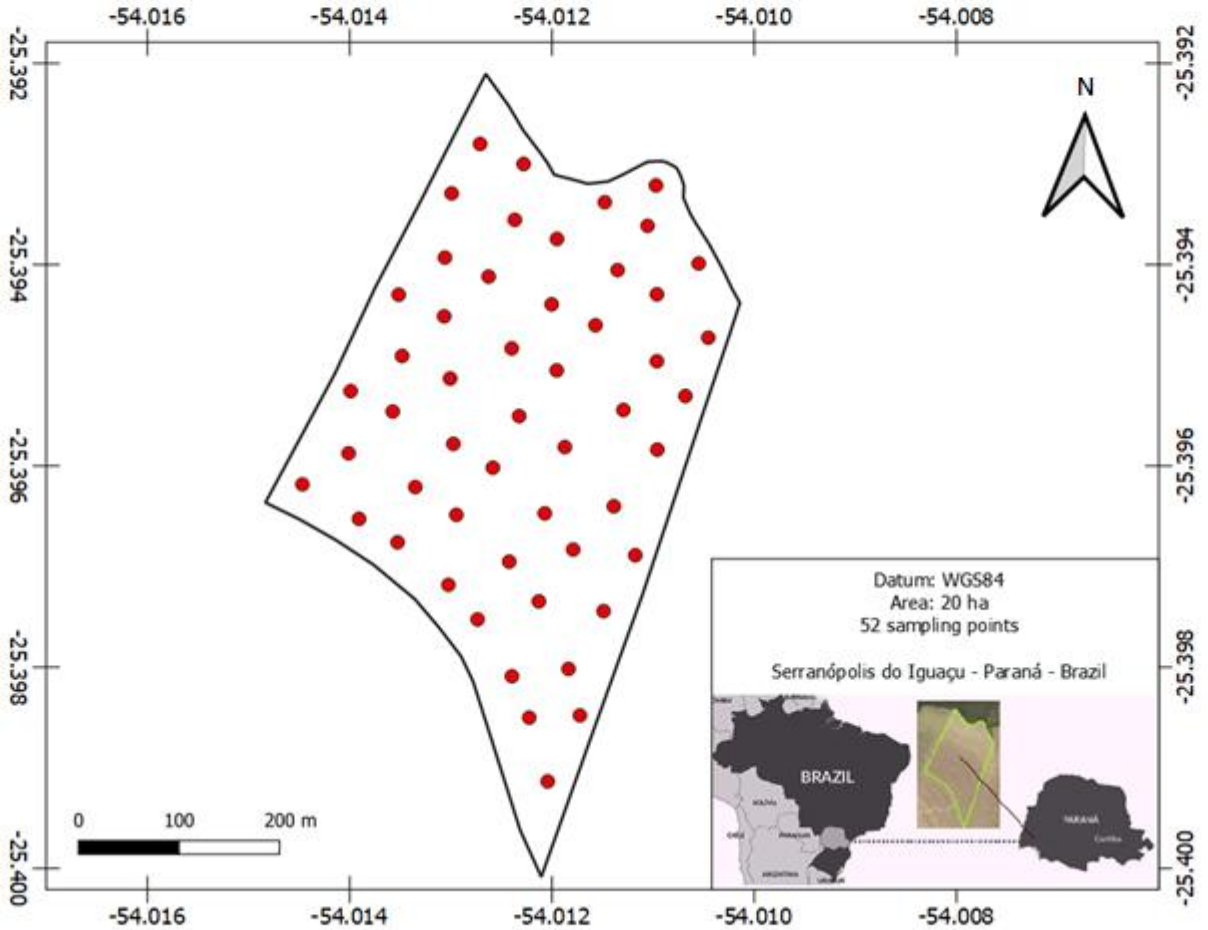


Fig. 10 Location of experiment and 52 sampling points in an experimental field in the municipality of Serranópolis do Iguaçu, Paraná state, Southern Brazil. Black contour delineates the 20 ha area used. Coordinates are in degrees (WGS 1984). The minimum and maximum distances among the sampling points are 45 and 706 m.

The case study demonstrates the steps for delineating temporary MZs for fertilizer application (Fig. 11).

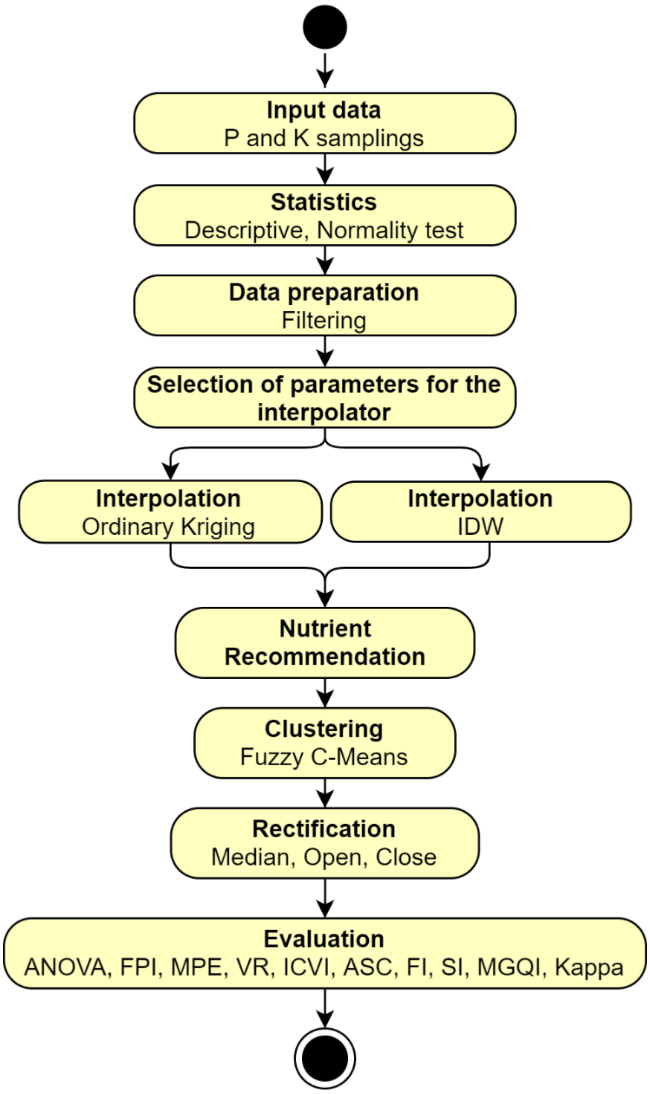


Fig. 11 Workflow for demonstrating activities involved in delineating management zones for nutrient application.

The used data consisted of P (mg dm^{-3}) and K (cmol$_c$ dm^{-3}) contents collected in 2019 in a 52-point sampling grid collected before soybean sowing during 2019/2020 growing season. The descriptive statistics of data were calculated using ADB-ST-API. CV was classified as proposed by Pimentel-Gomes (2009): low, when CV ≤ 10%; medium, when 10% < CV ≤ 20%; high, when 20% < CV ≤ 30%; and very high, when CV > 30%. Data normality was verified by Kolmogorov-Smirnov test. Data were normally distributed if at least one of the tests presented normality with 5% significance. Outliers were removed by the data cleaning procedure. Values outside the mean ± three standard deviations were identified as outliers and removed (Córdoba et al., 2016).

The sample data of P and K were interpolated by OK (if spatially dependent) or by IDW (if not), with interpolation parameters selected by ISI. In geostatistics, the Matheron (1963) classic estimator was used to calculate semivariances with at least 30 pairs of points (Journel and Huijbregts, 1978), and range (Ra) was limited to half of the maximum distance (MD) among points (cutoff = 0.5*MD). The lag size h was defined as the 44 meters, calculated from the number of lags (relation between cutoff and the shortest distance among pairs of points). The minimum and maximum amounts of 53 and 180 pairs were obtained with this lag distance to calculate semivariances. The maps with interpolated data were generated in ADB-Map application. Fertilizer recommendation was carried out for P and K by ADB-REC-API microservice.

Fertilizer recommendation MZs, called application zones (AZs), were delineated by fuzzy c-means method for data clustering and divided into 2, 3, and 4 classes. MZs were rectified until small spots were eliminated, by median, open, and close methods, applied until five times, as necessary.

The following procedures were used to evaluate MZs:

1. Statistical analysis by MZ class, including Tukey's mean difference test.
2. The following MZ quality indices were calculated for each: (i) FPI; (ii) MPE; (iii) ASC; (iv) SI; and (v) fragmentation index (FI) (Souza et al., 2021) (Equation A15). The VR (%) was calculated for each MZ, considering the variable of interest (P recommendation or K recommendation).
3. The following overall quality indices were calculated: (i) ICVI; and (ii) modified global quality index (MGQI) (Equation A17).
4. Agreement between original and Kappa rectified MZs was verified.

The following procedure was used to define the ideal number of classes: (1) comparison of means by the Tukey's test, as it is only interesting to divide the total area into classes that have a statistically distinct target variable (P or K recommendation) (Souza et al., 2018); and (2) choice of the number of classes that have the lowest MGQI, as it corresponds to a composition of VR%, FPI, MPE, SIr%, FIr%, and ASC indices.

5.3 Results and discussion

5. 3. 1 AgDataBox platform in a microservice architecture

Restructuring ADB platform in an MSA allows reusing procedures implemented in a PostgreSQL procedural database or directly in the preliminary version of ADB-Map application (Borges et al., 2020; Michelon et al., 2019). The preliminary ADB-Map version evolved from SDUM, which had most functionality implemented in PL/pgSQL procedural language. It was implemented in a monolithic approach following the model-view-controller (MVC) pattern, based on Java language with web technologies, such as VRaptor framework, and deployed in an Apache Tomcat application server. This application structuring approach added new features and reused them by other applications such as ADB-Mobile and ADB-IoT. This difficulty occurs because all functionalities in a monolithic architecture are encapsulated in a single application, not allowing its modules to run independently (Ponce et al., 2019).

ADB started to have a distributed network environment with the MSA creation and availability of web APIs, offering independence from application, programming language, operating system, and hardware on client-side. Therefore, time and effort are saved in developing clients' applications, as these features do not need to be re-implemented.

Furthermore, ADB platform microservices can be replicated among several servers, providing execution scaling according to the demand. Each component of the platform, microservices, and applications are created as containers of Docker tool, allowing the quick provisioning of services on server with Docker installed. As identified in their research, Li et al. (2020) pointed out that containerization is an emerging tactic for MSA, which allows for better performance than virtual machines.

Before MSA, efforts were concentrated on extracting functionalities aggregated in PostgreSQL database, creating a portfolio of scripts implemented in R statistical software language. Routines for data interpolation (Betzek et al., 2019), selection of best parameters for ordinary kriging (OK) (Betzek et al., 2019), data clustering (Gavioli et al., 2019), and selection of variables for MZ delineation (Gavioli et al., 2016) stand out. Furthermore, R statistical software has been used as an engine for executing specialized functionalities not only in ADB (Paccioretti et al., 2020; Leroux et al., 2018; Rupnik et al., 2019).

This study showed that the functionalities implemented in R statistical software scripts could be incorporated into microservices architecture and made available to any application that needs to consume them. Moreover, there was a need to refactorize other features that were in PL/pgSQL procedural language, such as selection of the best parameters for IDW interpolator (Betzek et al., 2019), bivariate Moran's I correlation, data normalization methods (Schenatto et al., 2017a), and smoothness index, which were reimplemented in R software, leading to a performance gain when executing the functionalities.

The structuring of microservices was performed from the functional decomposition, using Model-Driven approach. Ponce et al. (2019) found out most MSA migration studies following this approach. Scripts in R needed to be consumed from JavaScript language used in microservices development in Node.JS platform. For this purpose, the processes execution of the operating system, which executes "Rscript" utility of R, is used. Similarly, Dall'agnol et al. (2020) developed a web application in Java language with VRaptor framework in MVC standard, which executes scripts by the invocation of R statistical software.

5. 3. 2 Statistical analysis and data preparation

Statistical analysis and data preparation can be performed by ADB-ST-API microservice. Resources were made available following the same pattern as ADB microservices platform, based on REST architecture and data represented in JSON format. Features consisted of descriptive statistical analysis, data normality tests, PCA, PCA charting, Kappa and GA agreement statistics, data normalization, and data cleaning (Appendix B, Table 9). Outliers and null and duplicate data are removed in data cleaning. Data cleaning is the only resource that needs to receive geographic coordinates and dataset to carry out duplicate data removal procedure.

Resources for PCA are made available in this microservice, being useful for selecting variables for delineating MZs. This analysis has been used in several studies involving MZ delineation (Betzek et al., 2019; Gavioli et al., 2019; Gavioli et al., 2016). PCA request and response example is shown in Fig. 12.


```json
{
  "datasetType": "multi",
  "neighborLowerDistance": 0,
  "neighborUpperDistance": 240,
  "layers": [
    {
      "name": "Soybean",
      "dataset": [5.061]
    },
    {
      "name": "P",
      "dataset": [17.4, ...]
    },
    {
      "name": "K",
      "dataset": [0.28, ...]
    }
  ]
}
```


a) Request


```json
[
  {
    "component": 1,
    "eigenvalue": 1.20827139669403,
    "eigenvector": [
      {
        "name": "Soybean",
        "value": 0.214773828467946
      },
      ...
    ],
    "observations": [
      {
        "order": 1,
        "value": -0.396482307458985
      },
      ...
    ],
    "proportion": 40.2757132231342
  },
  ...
]
```


b) Response

Fig. 12 Example of JSON objects of request (a) and response (b) of the principal component analysis.

JSON objects of request for resources of descriptive statistical analysis and data normality test are exemplified in Fig. 13.

```
{
  "count": 51,
  "average": 0.356,
  "mode": 0.28,
  "standardDeviation": 0.169,
  "sampleStandardDeviation": 0.171,
  "variance": 0.028,
  "sampleVariance": 0.029,
  "medianAbsoluteDeviation": 0.090,
  "interquartileRange": 0.19,
  "quantile0": 0.09,
  "quantile25": 0.25,
  "quantile50": 0.33,
  "quantile75": 0.44,
  "quantile100": 0.8
}
```


a) Descriptive statistics response sample


```
[
  {
    "test": "ad",
    "description": "Anderson-Darling",
    "statistic": 0.732,
    "pValue": 0.052
  },
  ...
]
```


b) Normality test response sample

Fig. 13 Example of JSON objects resulting from the descriptive statistics analysis (a) and data normality test (b) in AgDataBox Statistics API.

P and K data did not show normality by Kolmogorov-Smirnov test at 5% significance. However, variables presented normality after data cleaning procedure was performed. Only one observation was removed, as an outlier, for each variable (Table 3).

Table 3 Descriptive statistics of data

	N	Minimum	1Q	Mean	Median	3Q	Maximum	S	CV%
P*	52	4.4	10.1	19.3	16.3	24.8	80.7	13.7	71 (VH)
P (cleaned)	51	4.4	10.1	18.1	15.8	24.1	53.0	10.7	59 (VH)
K*	52	0.09	0.25	0.37	0.34	0.45	0.94	0.19	51 (VH)
K (cleaned)	51	0.09	0.25	0.36	0.33	0.44	0.80	0.17	48 (VH)

CV: coefficient of variation: low (L) when CV ≤ 10%, medium (M) when 10% < CV ≤ 20%, high (H) when 20% < CV ≤ 30%, and very high (VH) when CV > 30%; P: phosphorus; K: potassium; N: number of observations; S: standard deviation; 1Q: 1st quartile; 3Q: 3st quartile.
* No normality at 5% significance level.

Phosphorus (P) presented a mean of 19.3 mg dm^{-3} with the original sampling and 18.1 mg dm^{-3} after data cleaning, which is considered very high (Pauletti and Motta, 2019) for annual crops productions in Paraná state, considering a clay content in the experimental area above 400 g kg^{-1}. Potassium (K) had a mean of 0.37 cmol$_c$ dm^{-3} with the original sampling data and 0.36 cmol$_c$ dm^{-3} after data cleaning, which are considered high values (Pauletti and Motta, 2019). The coefficient of variation (CV) for the variables was considered very high (higher than 30%) (Pimentel-Gomes, 2009).

Datasets need to be normalized to delineate MZs (Schenatto et al., 2017a). Thus, ADB-ST-API offers the possibility to normalize data using range, average, z-score, and min-max methods. The request for ADB-ST-API is carried out by sending a dataset without geographic coordinates, and, as a response, there is the normalized dataset. For example, variables P and K were normalized by range, average, z-score, and min-max methods (Table 4), demonstrating these features using ADB-ST-API.

Table 4 Descriptive statistics of variables P and K after normalization by range, average, z-score, and min-max methods

Nutrients	Method	Minimum	Mean	Median	Maximum	S	CV%
Phosphorus	Range	-0.23	0.05	0.00	0.77	0.22	454 (VL)
	Average	0.25	1.00	0.87	2.92	0.59	59 (VL)
	Z-score	-1.29	0.00	-0.22	3.29	1.01	*
	Min-Max	0.00	0.28	0.23	1.00	0.22	78 (VL)
Potassium	Range	-0.34	0.04	0.00	0.66	0.24	651 (VL)
	Average	0.25	1.00	0.93	2.25	0.48	48 (VL)
	Z-score	-1.57	0.00	-0.16	2.62	1.01	*
	Min-Max	0.00	0.38	0.34	1.00	0.24	64 (VL)

CV%: coefficient of variation: low (L) when CV ≤ 10%, medium (M) when 10% < CV ≤ 20%, high (H) when 20% < CV ≤ 30%, and very high (VH) when CV > 30%; * The CV% could not be calculated. S: standard deviation.

5. 3. 3 Spatial operations

ADB-SP-API microservice provides spatial data preparation and analysis procedures (Appendix B, Table 10). In addition, it features grid preparation functionalities, such as the conversion of geographic coordinates and grid degradation (down grid). The conversion of geographic coordinates is necessary for ADB, as resources are implemented in different microservices that work with Lat/Long coordinate system and others with Universal Transverse Mercator (UTM).

Grid degradation allows adjusting grids with different amounts of data and geographic positions (Fig. 14). ADB has procedures that show the same number of observations per grid and geographic coordinates correspondence in both grids, such as data clustering and variable selection procedures (PCA, MULTISPATI-PCA, and spatial correlation matrix). The common cases for grid degradation are datasets originating from harvesters instrumented with a harvest monitor and images from multispectral cameras, which need to be adjusted to be used to select variables for MZ delineation.

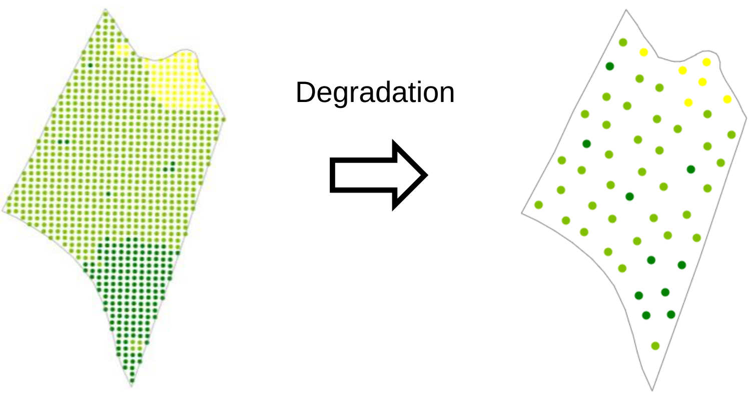


Fig. 14 Example of grid degradation procedure where the dense grid (interpolated grid, harvester yield map, digital imaging, and others) is reduced to a 52-point grid.

The spatial correlation matrix, used to select variables for MZ delineation (Bazzi et al., 2013), is constructed based on Moran's bivariate spatial autocorrelation statistic (Reich, 2008; Schepers et al., 2004). Statistic that determines the correlation degree between two variables and the test significance is calculated by ADB-SP-API.

Smoothness indices, variance reduction, and cluster statistics are used to evaluate MZs.

5. 3. 4 Data interpolation

Data interpolation resources implemented in ADB-INT-API microservice allow selecting the best parameters for IDW and OK interpolators and identifying which method is the best for a given dataset from ISI. Interpolation was also available from IDW, OK, MA, and NN methods (Appendix B, Table 11).

The resources ISI for IDW and ISI for geostatistics execute independent scripts to calculate ISI by identifying the best fit for IDW or OK interpolator, with results merged to determine the best interpolator among them. Lists of objects that contain ISI value are generated for each chosen combination, that is, the exponent range and range of neighbors to be tested in ISI for IDW resource, as well as the combinations of semivariogram models, fitting methods, partial sill intervals, and range intervals in ISI resource for Geostatistics. The best settings for IDW and OK interpolators are chosen from the lowest ISI value. The request and response JSON objects are exemplified in Fig. 15.


```
{
  "models": ["exp","gaus","sph","matern"],
  "methods": ["ols","wls"],
  "kappas": ["0.5","1","1.5","2"],
  "lambda": 1,
  "autoLags": true,
  "amountLags": 10,
  "estimator": "classical",
  "cutoff": 50,
  "pairs": 30,
  "amountRangeIntervals": 5,
  "amountPartialSillIntervals": 5,
  "dataset": [{
      "coordinates": [196955, 7187436],
      "data": 3.1
    }, ...
  ]
}
```


a) ISI request JSON


```
[
  {
    "model": "matern",
    "method": "ols",
    "kappa": 2,
    "nuggetEffect": 0.0228,
    "partialSill": 0.07,
    "range": 352.8649,
    "averageError": -0.0011,
    "stdDevAverageError": 0.1634,
    "ice": 0.8286,
    "isi": 0.7782,
    "sdi": 0.7545,
  },
  ...
]
```


b) ISI response JSON

Fig. 15 Example of (a) request and (b) service invocation response of JSON objects for selecting the interpolator.

Resources for data interpolation (IDW, OK, MA, and NN) receive a sample dataset and return an interpolated dataset, considering the informed parameterization. The semivariogram graph can be generated in the ADB-INT-API, and the request JSON object is represented in Fig. 16.

```json
{
  "variogram": {
    "estimator": "classical",
    "model": "matern",
    "method": "ols",
    "partialSill": 0.07,
    "range": 352.8649,
    "nuggetEffect": 0.0228,
    "kappa": "2",
    "lambda": 1,
    "autoLags": "true",
    "amountLags": false,
    "cutoff": 50,
    "cutoffInMeter": false,
    "pairs": 30
  },
  "width": 500,
  "height": 400,
  "lineColor": "#001188",
  "labelMain": "Semivariogram",
  "labelX": "Distance",
  "labelY": "Semivariance",
  "dataset": [
    {
      "coordinates": [
        196955.676165181,
        7187436.24639107
      ],
      "data": 0.28
    },
    ...
  ]
}
```

Fig. 16 Example of a JSON object required to generate the semivariogram graph in ADB Interpolation API, where semivariogram parameters, styles, and dataset are passed.

The scripts that determine ISI values for IDW and geostatistics were optimized for faster execution. The main change was in the adoption of parallel processing execution. As a result, all these analyses were executed sequentially in a single thread, both in R and in PostgreSQL database. The parallelism allows allocating several threads to be executed concurrently, managed by the “parallel” library from R. The number of threads is automatically determined based on the number of cores provided by the server in which microservice is hosted.

The best-fit analysis algorithm for IDW was initially implemented directly in the PostgreSQL database by PL/pgSQL procedural language, which analyzes twelve different values for exponent (0.5, 1.0, …, and 6.0), considering a fixed number of eight neighboring points and using ISI (Equation A19) to identify the best value for the exponent, as during the best semivariogram model selection (Betzek et al., 2019). However, this algorithm was rewritten in R language and had its logic modified to analyze a larger range of exponents and consider a range of neighbors in the analysis.

The algorithm ISI for geostatistics tests seven different semivariogram models (spherical, Gaussian, exponential, Matérn 0.5, Matérn 1.0, Matérn 1.5, and Matérn 2.0) and two statistical methods of semivariogram fit optimization (OLS and WLS), totaling fourteen different models (Betzek et al., 2019). ISI selects the best model.

Microservice resources allowed the construction of thematic maps of P and K contents of the experimental area (Fig. 17). Maps were generated by IDW and OK interpolators, considering the original variables, with no evidence of data normality, and the cleaned data, with removed outliers.

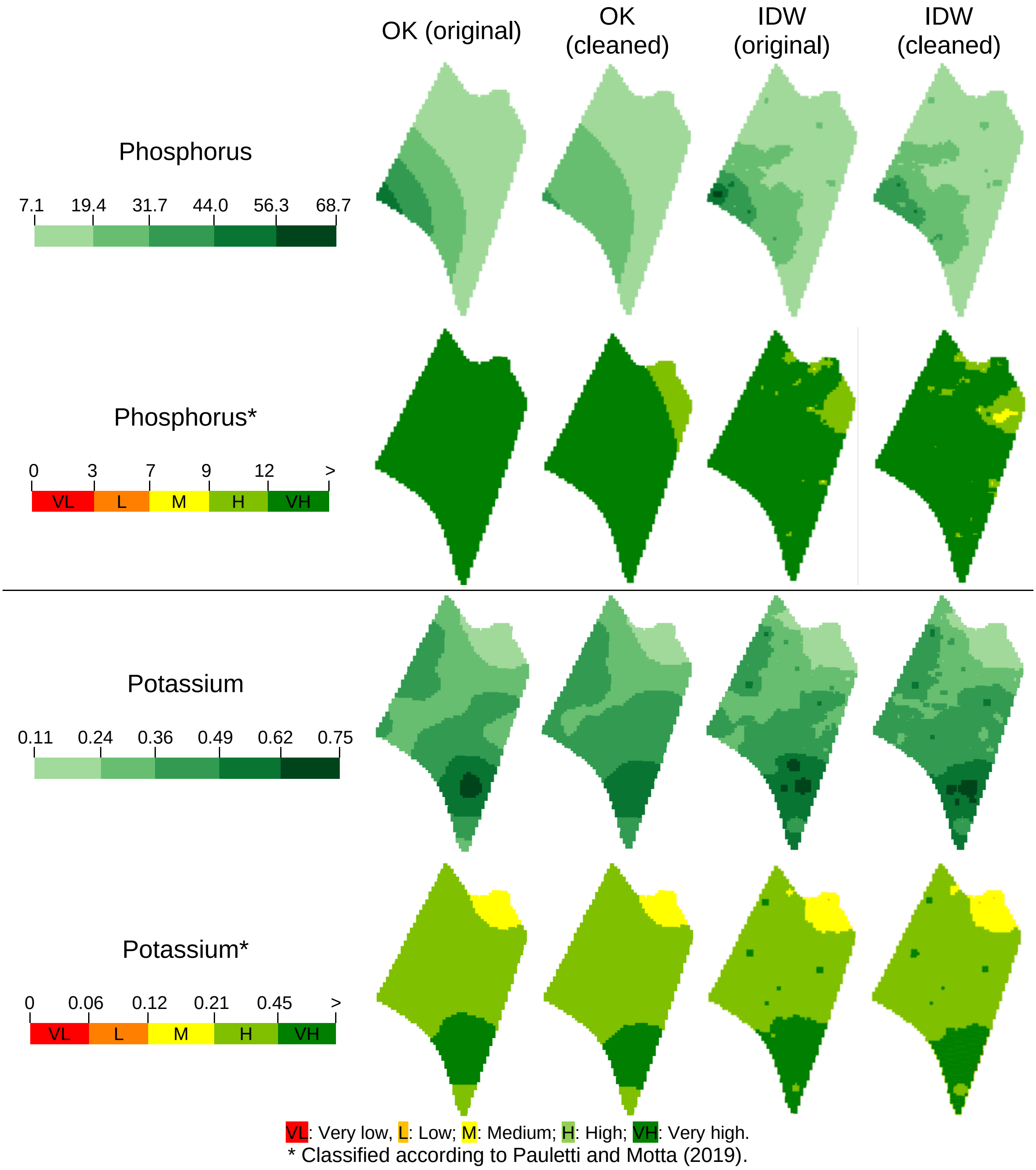
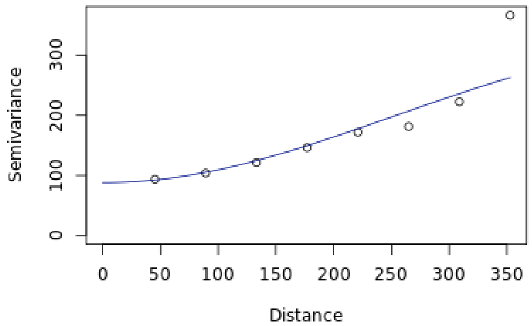
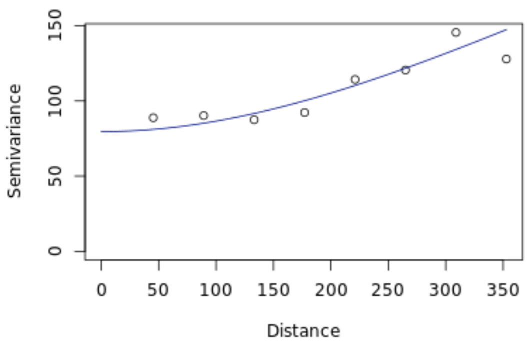
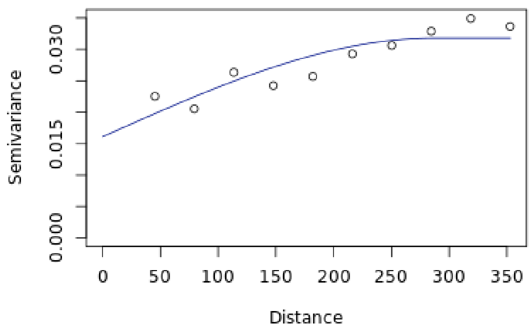
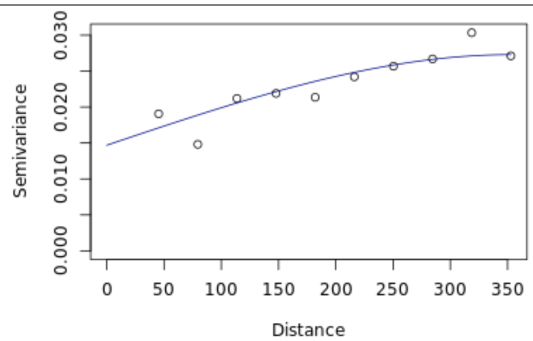


Fig. 17 Phosphorus and potassium availability maps interpolated by ordinary kriging (OK) and inverse distance weighting (IDW), using original and cleaned sample grids.

The maps are classified according to the interpretation of P (mg dm^{-3}) and K (cmol$_c$ dm^{-3}) available on soil (Pauletti and Motta, 2019). Clay content in the experimental area higher than 400 g kg^{-1} is considered for P. The best exponent determined by ISI for interpolation by IDW was 1. The number of neighbors varied according to the datasets, reaching 9 (original data) and 6 (cleaned data) for P and 12 (original data) and 11 (cleaned data) for K. The initial parameters of the interpolator selection analysis were set with the exponent range from 1 to 6, considering a variation of 0.5 (1, 1.5, 2, 2.5, …, 6) and the number of neighbors from 4 to 12, totaling 88 analyses for IDW.

ISI for geostatistics analysis showed that the parameters selected for dataset were those presented in the semivariograms (Table 5). The analysis parameters considered exponential, spherical, Gaussian, and Matérn models, the OLS and WLS fitting methods, kappa values for Matérn 0.5, 1.0, 1.5, and 2.0 models, and ten sill intervals and range, totaling 1400 geostatistical analyses. The semivariogram model of variable K changed between the original and cleaned dataset and from Gaussian to Matérn, but it remained spherical for the variable K.

Table 5 Semivariograms of the variables P and K with original and cleaned datasets

	Original	**Cleaned**
P	 Model: Gaussian, Method: WLS, NE: 87.85, PS: 277.18, R: 353	 Model: Matérn, Model: OLS, Kappa: 2, NE: 79.57, PS: 352.86, R: 348
K	 Model: Spherical, Model: OLS, NE=0.0161, PS: 0.0157, R: 287	 Model: Spherical, Model: OLS, NE: 0.0147, PS: 0.0126, R: 352

P: phosphorus; K: potassium; NE: nugget effect; PS: partial sill; R: range (m); OLS: ordinary least squares; WLS: weighted least squares.

5. 3. 5 Nutrient recommendation

ADB-REC-API microservice calculates P, K, nitrogen (N), and lime recommendations using the parameters passed to API (Appendix B, Table 12). The fertilizer recommendation considers the parameters differentiation according to the desired nutrient (P, K, or N) and the calculation method (availability of nutrients in the soil or expected crop yield). The request and response JSON objects are exemplified in Fig. 18.


```
{
  "culture": "soy",
  "soil": "clay",
  "source": [
    "ss", "st", "map", "dap", "arad"
  ],
  "dataset": [17.4, ...]
}
```


a) Recommendation request JSON

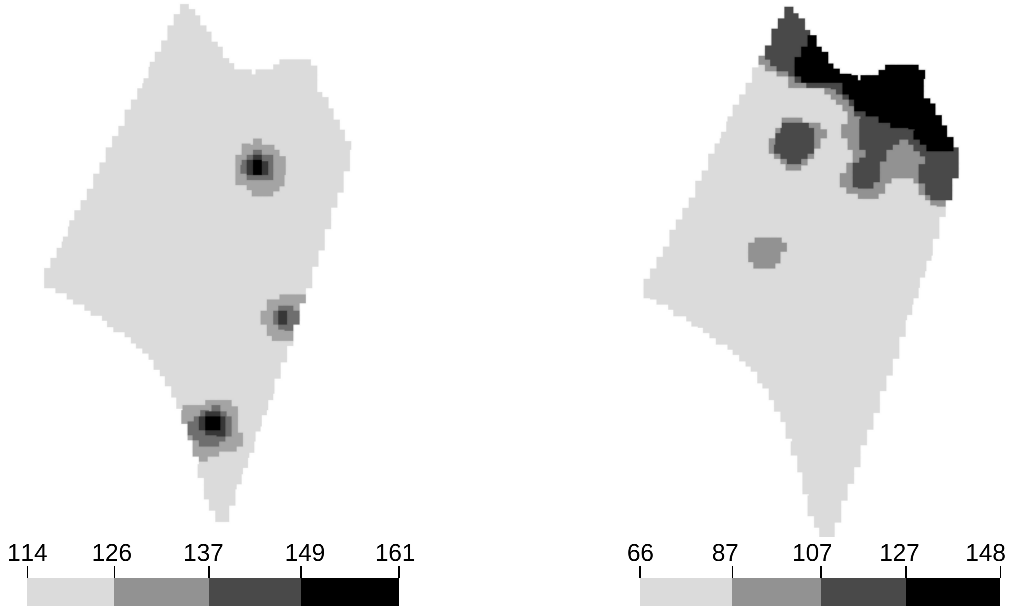

```
{
  "source": [
    {"name": "ss", "dataset": [275, ...]},
    {"name": "st", "dataset": [122.10, ...]},
    {"name": "map", "dataset": [114.4, ...]},
    {"name": "dap", "dataset": [140.8, ...]},
    {"name": "arad", "dataset": [916.85, ...]}
  ]
}
```


b) Recommendation response JSON

Fig. 18 Example of JSON objects for the request (a) and response received (b) from the microservice AgDataBox Recommendation API.

Each recommendation method must send a dataset of available nutrient content on soil (P, K, or organic matter). Soybean and corn are considered in the P and K recommendation and corn in the N recommendation. Fertilizer recommendation is based on the product to be applied. For example, P recommendation is calculated based on single superphosphate (SS), triple superphosphate (TS), monoammonium phosphate (MAP), diammonium phosphate (DAP), and Araxá phosphate (ARAD). The products available to calculate K recommendation are potassium oxide, potassium sulfate, potassium and magnesium sulfate, and potassium double saltpeter. N recommendation can be made from urea, ammonium sulfate, ammonium nitrate, and ammonium chloride. As a response, ADB-REC-API returns a dataset containing the amount of fertilizer to be applied in each observation of the nutrient availability dataset, according to the geographic position.

The fertilizer application recommendations were generated with the study area's P and K experimental data (Fig. 19). The P (Fig. 19a) and K (Fig. 19b) recommendation maps for soybean were interpolated by IDW with exponent 1.5 and 4.5, respectively, 9 and 11 neighbors, divided into four classes by equal distance classification. The P recommendations were calculated from the soil nutrient availability for SS, TS, MAP, DAP, and ADAD and K for KCL, SP, SPD, and SPM. The recommendation for most of the area is the minimum rate because P contents are very high (Table 4) (Pauletti and Motta, 2019). There is a recommendation to apply higher K rates at the top of the map.



IDW_E1.5_N9
a) P recommendation by monoammonium phosphate (MAP) fertilizer (t ha^{-1})

IDW_E4.5_N11
b) K recommendation by potassium chloride (KCL) fertilizer (t ha^{-1})

Fig. 19 Fertilizer recommendation maps (t ha^{-1}) to supply the nutrient requirements (a) phosphorus (P) and (b) potassium (K) for soybean crop in the experimental field.

The recommendations variability for each product was considered low (CV < 10%) for P and high (20% < CV ≤ 30%) for K (Table 6).

Table 6 Descriptive statistics for P and K recommendations

Nutrient	Product	Min.	Mean	Median	Max.	S	CV (%)
P	SS	275.00	281.57	275.00	398.59	24.52	9 (l)
	ST	122.10	125.02	122.10	176.97	10.88	9 (l)
	MAP	114.40	117.13	114.40	165.81	10.20	9 (l)
	DAP	140.80	144.17	140.80	204.08	12.55	9 (l)
	ARAD	916.85	938.77	916.85	1328.89	81.73	9 (l)
K	KCL	66.80	66.80	81.70	147.57	24.76	30 (h)
	PS	80.00	80.00	97.85	176.73	29.65	30 (h)
	PMS	154.00	154.00	188.36	340.21	57.08	30 (h)
	PDS	285.60	285.60	349.32	630.94	105.85	30 (h)

P: phosphorus; K: potassium; Min.: minimum; Max.: maximum; S: standard deviation; CV: coefficient of variation – low (l), high (h); SS: simple superphosphate; TS: triple superphosphate; MAP: monoammonium phosphate; DAP: diamonic phosphate; ARAD: araxa phosphate; KCL: potassium chloride; PS: potassium sulphate; PMS: potassium and magnesium sulfate; PDS: potassium double saltpeter.

5. 3. 6 Data clustering

MZ delineation is performed by data clustering algorithms available in the ADB-CLU-API microservice (Appendix B, Table 13). Each clustering method can be executed independently, requiring different parameters in the request. The request parameters common to all methods consist of datasets and class numbers to perform clustering. Only one request to the microservice allows obtaining several clustered datasets, such as in 2, 3, and 4 classes. ADB-CLU-API microservice request object is exemplified in Fig. 20a. Distance metrics were provided in each method according to the availability of libraries made available for data clustering in R statistical software.


```json
{
  "datasetType": "condensed",
  "datasetResponseFormat": "condensed",
  "classes": [ 2, 3, 4],
  "normalize": "amplitude",
  "metric": "euclidean",
  "weightingExponent": 1.3,
  "iterations": 500,
  "coordinates": [
    [196946.6854, 7187348.9489],
    ...
  ],
  "layers": [
    {
      "name": "K-2019",
      "dataset": [
        0.5018, ...
      ]
    },
    ...
  ]
}
```

```json
[
  {
    "classes": 2,
    "dataset": [ ... ],
    "asc": 0.588845195058417,
    "pc": 0.972124804306403,
    "fpi": 0.0557503913871948,
    "pe": 0.0504424525942423,
    "mpe": 0.07277307620799,
    "xb": 5.42925584588642e-05
  },
  ...
]
```

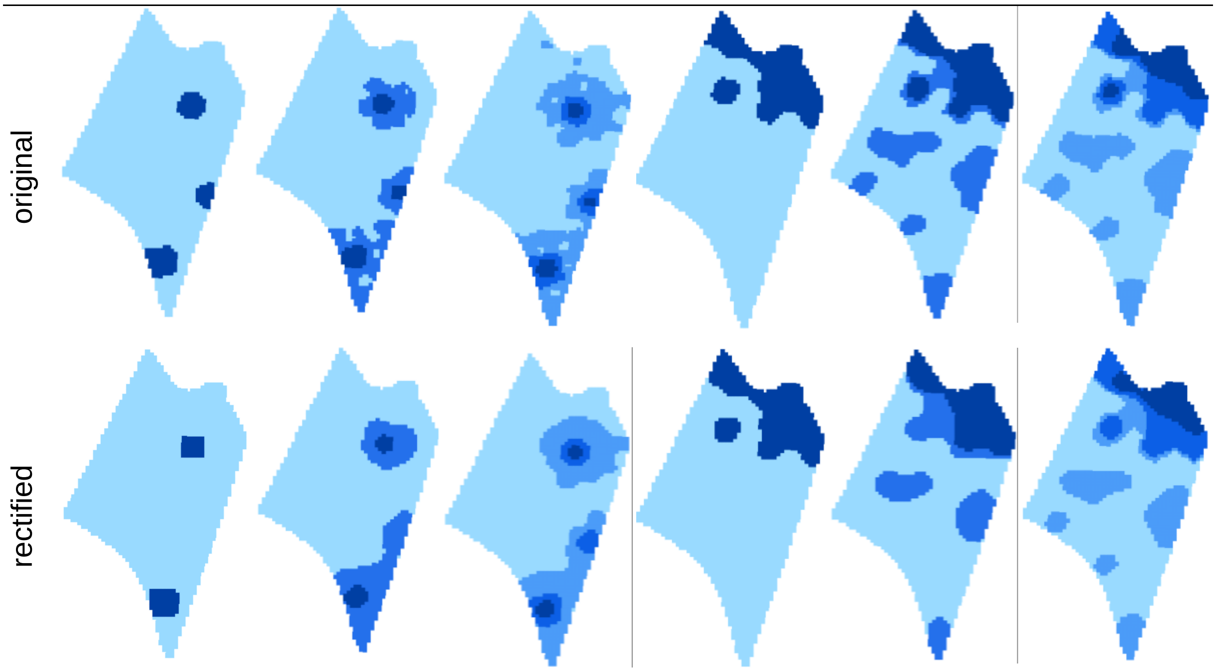

a) Request b) Response

Fig. 20 Example of JSON object to perform the request (a) and the received response (b) in the AgDataBox Clustering API microservice.

As a result, microservice returns a JSON object (Fig. 20b) containing the clustered datasets and evaluation indices. ASC is calculated in all methods. FCM and UFCL return PC, FPI, PE, MPE, and XB. The k-means method returns the sum of squares (SS) and within-clusters sum of squares (WCSS) values.

P and K recommendation maps were clustered by the FCM method, divided into 2, 3, and 4 classes (Table 7), named AZ_P and AZ_K.

Table 7 Application zones for phosphorus and potassium before and after map rectification.



	Phosphorus (AZ_P)			Potassium (AZ_K)		
	2 classes	**3 classes**	**4 classes**	**2 classes**	**3 classes**	**4 classes**
Kp	0.77(S)	0.84 (VS)	0.90 (VS)	0.99 (VS)	0.75(S)	0.90 (VS)
GA	0.98	0.94	0.95	1.00	0.86	0.94
Met	Open	Median and Close	Median and Close	Median and Close	Median and Close	Median

AZ_P: Application zones for phosphorus fertilizer; AZ_K: Application zones for potassium fertilizer; Met: Rectification method; GA: Global Accuracy. Kp: Kappa; Kappa agreement: N = no agreement (Kp ≤ 0.2); W = weak (0.2 < Kp ≤ 0.4); M = moderate (0.4 < Kp ≤ 0.6); S = strong (0.6 < Kp ≤ 0.8); VS = very strong (0.8 < Kp ≤ 1); Met: Method of rectification. The darker the classes are painted, the highest are the effectively recommended amount.

5. 3. 7 Data rectification

Isolated spots or pixels usually appear regardless of the method used to delimit these zones. Therefore, differentiated applications of nutrients in these spots may be operationally infeasible. In this sense, the rectification methods allow adjusting MZs/AZs and thematic maps to smooth edges and eliminate small spots. Gonzalez and Woods (2008), Córdoba et al. (2016), Albornoz et al. (2017), and Betzek et al. (2018) used median, dilation, and erosion filters to reduce MZ fragmentation.

The resources implemented in ADB-REC-API microservice are divided according to the implemented rectification method: median, erosion, dilation, and the combination of erosion and dilation (Appendix B, Table 14). The median filter modifies the central pixel of a small region, called filter size, by the median value calculated by the intensity of pixels in this region. The image opening effect reserves the unsegmented parts of objects using the first image dilation by merging the neighboring pixels of an object into the object and then image erosion by removing the boundary pixels from the object. On the contrary, image closure is erosion followed by dilation to eliminate non-segmented parts of the background (He et al., 2016).

Visual observations show that small spots were removed after applying median digital image processing filters and opening and closing effect to rectify AZs AZ_P and AZ_K (Table 7). The comparison among rectified AZs and original AZs by Kappa index shows a strong agreement in AZ_P in 2 classes and AP_K in 3 classes. The agreement was very strong in the others. Kappa index ranged from 0.77 to 0.99 among the compared AZs. GA ranged from 0.86 to 1.00, indicating that, even after rectification, AZs did not lose the visual essence compared to the original AZs.

5. 3. 8 Evaluation of management zones

The ideal number of classes is defined by considering (1) means from Tukey test since the total area should only be divided into classes that have a target variable (P or K recommended) statistically different (Souza et al., 2018), and (2) the number of classes with the lowest MGQI, as it corresponds to a composition of indices such as VR, FPI, MPE, ASC, SIr, and FIr.

The resources to evaluate MZs from ADB-MSA showed that the mean amount of fertilizers recommended for P and K correction is different in all classes of AZ_P and AZ_K (Table 8). The quantities required are based on MAP and KCL formulations to correct P and K application, respectively. The best grouping for both cases (P and K) used two classes, showing the lowest MGQI. The darker the classes are painted, the higher are the effectively recommended amount. Therefore, the highest percentage of an area requires a smaller amount of P and K fertilizers. On the other hand, AZ_P divided into two classes shows that the area is almost homogeneous because only 4% of it needs a larger amount of fertilizer.

Table 8 Quality indices of management zones for fertilizer application

MZ	Cl.	A%	Mean	CV%	VR%	FIr%	SIr%	ASC	FPI	MPE	ICVI	MGQI
AZ_P 2 classes	1	96	114.7 a	1	80	50	99	0.83	0.019	0.023	0.30	0.56
	2	4	157.0 b	6								
AZ_P 3 classes	1	78	115.3 a	1	75	67	96	0.71	0.047	0.047	0.66	1.62
	2	19	121.9 b	4								
	3	2	140.7 c	5								
AZ_P 4 classes	1	68	115.0 a	1	88	100	94	0.68	0.054	0.048	0.67	2.11
	2	26	119.1 b	2								
	3	5	130.2 c	3								
	4	1	145.1 d	5								
AZ_K 2 classes	1	78	70.0 a	9	85	50	98	0.77	0.028	0.035	0.50	1.00
	2	22	125.6 b	13								
AZ_K 3 classes	1	60	67.2 a	2	89	100	94	0.71	0.045	0.046	0.69	2.15
	2	20	84.2 b	19								
	3	20	125.7 c	14								
AZ_K 4 classes	1	59	66.8 a	1	96	150	90	0.72	0.044	0.038	0.60	2.30
	2	20	79.6 b	8								
	3	11	112.4 c	6								
	4	10	141.4 d	4								

Cl.: class; A%: percentage of class in the area; CV%: coefficient of variation: low (L) when CV ≤ 10%, medium (M) when 10% < CV ≤ 20%; SIr%: smoothness index of the rectified management zone; ASC: average silhouette coefficient; FPI: fuzziness performance index; MPE: modified partition entropy; VR%: variance reduction; ICVI: improved cluster validation index; FIr%: fragmentation index of the rectified management zone; MGQI: modified global quality index; a, b, c, d (Mean): difference of means by the Tukey's test. Best grouping highlighted in blue.

CV% was considered low in 15 of 18 classes and medium in three of them. As it was expected, reduction in variance (VR%) and fragmentation of zones (FR%) increased with the number of classes, while smoothness index (SIr%) decreased. The highest values of ASC and the lowest values of FPI, MPE, and ICVI were obtained in the divisions into two classes.

5.4 Conclusions

ADB-MSA enabled to provide the features to create TMs and delineate MZs in microservices accessible via web APIs. Thus, ADB-MSA became (i) a flexible platform to absorb new business demands, (ii) an integrative and reusable platform by different applications, and (iii) a scalable platform, with the ability to allocate services in different operating environments.

The case study demonstrated the possibility of generating TMs and delineating MZs consuming ADB-MSA resources. The ideal number of MZs classes for P (AZ_P) and K (AZ_K) applications were considered two, using MGQI index.

5.5 Acknowledgments

The authors would like to thank the Western Paraná State University (UNIOESTE), the Federal University of Technology of Paraná (UTFPR), the Coordination for the Upgrading of Higher Education Personnel (CAPES), the National Council for Scientific and Technological Development (CNPq), the Itaipu Technological Park Foundation (FPTI), and the Ministry of Agriculture, Livestock and Food Supply (MAPA) for funding this project.

5.6 References

Albornoz, E. M., Kemerer, A. C., Galarza, R., Mastaglia, N., Melchiori, R., Martínez, C. E. (2018). Development and evaluation of an automatic software for management zone delineation. **Precision Agriculture**,19 (3), pp. 463-476.

Anderberg, M. R. (1973). **Cluster Analysis for Applications**. Academic Press, New York.

Balalaie, A., Heydarnoori, A., Jamshidi, P., Tamburri, D. A., Lynn, T. (2018). Microservices migration patterns. **Software: Practice and Experience**, 48 (11), pp. 2019-2042.

Bazzi, C. L., Jasse, E. P., Magalhães, P. S. G., Michelon, G. K., Souza, E. G., Schenatto, K., Sobjak, R. (2019a). AgDataBox API – Integration of data and software in precision agriculture, **SoftwareX**, 10. p. 100327.

Bazzi, C. L., Souza, E. G., Betzek, N. M. (2015). **Software para Definição de Unidades de Manejo**: Teoria e prática. 1. UNIOESTE, Cascavel.

Bazzi, C. L., Souza, E. G., Schenatto, K., Betzek, N. M., Gavioli, A. (2019b). A software for the delineation of crop management zones (SDUM). **Australian Journal of Crop Science. Southern Cross Journals**, 13, pp. 26-34.

Bazzi, C. L., Souza, E. G., Uribe-Opazo, M. A., Nóbrega, L. H. P., Rocha, D. M. (2013). Management zones definition using soil chemical and physical attributes in a soybean area. **Engenharia Agrícola**, 33 (5), pp. 952–964.

Beneduzzi, H. M. (2020). **Módulo computacional para cálculo da necessidade de nitrogênio, fósforo e potássio a partir de suas disponibilidades no solo**. PhD thesis, Western Paraná State University (UNIOESTE), Cascavel.

Bezdek, J. C. (1981). **Pattern Recognition with Fuzzy Objective Function Algorithms**. Plenum Press, New York.

Betzek, N. M., Souza, E. G., Bazzi, C. L., Schenatto K., Gavioli, A., Magalhães, P. S. G. (2019). Computational routines for the automatic selection of the best parameters used by interpolation methods to create thematic maps. **Computers and Electronics in Agriculture**, 157, pp. 49-62.

Betzek, N. M., Souza, E. G., Bazzi, C. L., Schenatto, K., Gavioli, A. (2018). Rectification methods for optimization of management zones. **Computers and Electronics in Agriculture**, 146 (1), pp. 1–11.

Bier, V. A., Souza, E. G. (2017). Interpolation selection index for delineation of thematic maps. **Computers and Electronics in Agriculture**, 136 (1), pp. 202-209.

Borges, L. G., Bazzi, C. L., Souza, E. G., Magalhães, P. S. G., Michelon, G. K. (2020). Web software to create thematic maps for precision agriculture. **Pesq. agropec. bras**., 55.

Cherradi, G., Bouziri, A. E., Boulmakoul, A., Zeitouni, K. (2017). Real-Time HazMat Environmental Information System: A micro-service based architecture. **Procedia Computer Science**, 109 (1), pp. 982-987.

Chipman, H., Tibshirani, R. (2006). Hybrid Hierarchical Clustering with Applications to Microarray Data. **Biostatistics**, 7, pp. 302-317.

Ciavotta, M., Alge, M., Menato, S., Rovere, D., Pedrazzoli, P. (2017). A Microservice-based Middleware for the Digital Factory. **Procedia Manufacturing**, 11 (1), pp. 931-938.

Clapp, J., Ruder, S. L. (2020). Precision Technologies for Agriculture: Digital Farming, Gene-Edited Crops, and the Politics of Sustainability. **Global Environmental Politics**, 20 (3), pp. 49-69.

Coelho, E. C., Souza, E. G., Uribe-Opazo, M. A., Pinheiro Neto, R. (2009). Influência da densidade amostral e do tipo de interpolador na elaboração de mapas temáticos. **Acta Scientiarum**, 31 (1), pp. 165-174.

Cohen, J. A. (1960). Coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, 20 (1), pp. 37-46.

Cohen, S., Cohen, Y., Alchanatis, V., Levi, O. (2013). Combining spectral and spatial information from aerial hyperspectral images for delineating homogenous management zones. **Biosystems Engineering**, 114 (4), pp. 435-443.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing of Environment**, 37 (1), p. 35-46.

Córdoba, M. A., Bruno, C. I., Costa, J. L., Peralta, N. R., Balzarini, M. G. (2016). Protocol for multivariate homogeneous zone delineation in precision agriculture. **Biosystems Engineering**, 143, pp. 95-107.

Dall'agnol, R. W., Michelon, G. K., Bazzi, C. L., Magalhães, P. S. G., Souza, E. G., Betzek, N. M., Sobjak, R. (2020). Web applications for spatial analyses and thematic map generation. **Computers and Electronics in Agriculture**, 172.

Dave, R. N. (1992). Generalized fuzzy c-shells clustering and detection of circular and elliptical boundaries. **Pattern Recognition**, 25 (7), pp. 713-721.

Dhillon, I. S., Modha, D. S. (2001). Concept decompositions for large sparse text data using clustering. **Machine Learning**, 42, pp. 143-175.

Doerge, T. A. (2000). **Management Zone Concepts**. Site-Specific Management Guidelines. Potash and Phosphate Institute. University South Dakota, Brokings.

Dray, S., Saïd, S., Débias, F. (2008). Spatial ordination of vegetation data using a generalization of Wartenberg’s multivariate spatial correlation. **Journal of Vegetation Science**, 19 (1), pp. 45-56.

Ferguson, R. B.; Hergert, G. W. 2009. Soil Sampling for Precision Agriculture. **Precision Agriculture**, pp. 1-4.

Foody, G. M. (2002). Status of land cover classification accuracy assessment. **Remote Sensing of Environment**, 80 (1), pp. 185–201.

Fraisse, C. W., Sudduth, K. A., Kitchen, J. R. (2001). Delineation of site-specific management zones by unsupervised classification of topographic attributes and soil electrical conductivity. **International Journal of the American Society of Agricultural and Biological Engineers**, 1 (44), pp. 155-166.

Fridgen, J. J., Kitchen, N. R., Sudduth, K. A., Drummond, S. T., Wiebold, W. J., Fraisse, C. W. (2004). Management zone analyst (MZA): software for subfield management zone delineation. **Agronomy Journal**, 96 (1), pp. 100-108.

Gavioli, A., Souza, E. G., Bazzi, C. L., Guedes, L. P. C., Schenatto, K. (2016). Optimization of management zone delineation by using spatial principal components. **Computers and Electronics in Agriculture**, 127 (1), pp. 302-310.

Gavioli, A., Souza, E. G., Bazzi, C. L., Schenatto, K., Betzek, N. M. (2019). Identification of management zones in precision agriculture: An evaluation of alternative cluster analysis methods. **Biosystems Engineering**, 181, pp. 86-102.

Gonzalez, R. C., Woods, R. (2008). **Digital image processing**. 3. Pearson Prentice Hall, New Jersey.

Grahl, M., Bluhm, T., Grün, M., Hennig, C., Holtz, A., Krom, J. G., Kühner, G., Laqua, H., Lewerentz, M., Riemann, H., Spring, A., Werner, A. (2017). Archive WEB API: A web service for the experiment data archive of Wendelstein 7-X. **Fusion Engineering and Design**, 123 (1), pp. 1015-1019.

Hassan, S., Bahsoon, R. (2016). Microservices and their design trade-offs: A self-adaptive roadmap. In: **Proceedings of the 2016 IEEE International Conference on Services Computing (SCC)**. IEEE Press, pp. 813-818.

He, H. J., Zheng, C., Sun, D. W. (2016). Chapter 2 - Image Segmentation Techniques. In: Sun, D.-W. (Eds.), **Computer Vision Technology for Food Quality Evaluation**, 2. Academic Press, pp. 45-63.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. **Journal of Educational Psychology**, 24 (6), pp. 417-441.

Isaaks, E. H., Srivastava, R. M. (1989). **Applied geostatistics**. Oxford University Press, New York.

Jain, A. K., Dubes, R. (1988). **Algorithms for clustering data**. Prentice-Hall, Englewood Cliffs.

Johnson, R. A., Wichern, D. W. (2007). **Applied Multivariate Statistical Analysis**. 6. Pearson Prentice Hall, Upper Saddle River.

Journel, A. G., Huijbregts, C. J. (1978). **Mining Geostatistics**. Academic Press, London-New York-San Francisco.

Kamilaris, A., Kartakoullis, A., Prenafeta-Boldú, F. X. (2017). A Review on the Practice of Big Data Analysis in Agriculture. **Computers and Electronics in Agriculture**, 143, pp. 23–37.

Kaufman, L., Rousseeuw, P. J. (1990). **Finding groups in data.** John Wiley & Sons, Hoboken.

Lajoie-O'malley, A.; Bronson, K.; Van Der Burg, S.; Klerkx, L. (2020). The future(s) of digital agriculture and sustainable food systems: An analysis of high-level policy documents. **Ecosystem Services**, 45.

Landis, J. R., Koch, G. G. (1977). The measurement of observer agreement for categorical data. **Biometrics**, 33 (1), pp. 159-174.

Larscheid G., Blackmore, B. S. (1996). Interactions between farm managers and information systems with respect to yield mapping. In: **International Conference on Precision Agriculture**, 3. Springer, Minneapolis, pp.1153-1163.

Leisch, F. (1999). Bagged clustering. In: **SFB adaptive information systems and modelling in economics and management science**, 51. Vienna University of Economics and Business, Vienna, pp. 1-11.

Lenarduzzi, V., Lomio, F., Saarimäki, N., Taibi, D. (2020). Does migrating a monolithic system to microservices decrease the technical debt? **Journal of Systems and Software**, 169.

Leroux, C., Jones, H., Pichon, L., Guillaume, S., Lamour, J., Taylor, J., Naud, O., Crestey, T., Lablee, J., Tisseyre, B. (2018). Geofis: an open source, decision-support tool for precision agriculture data. **Agriculture**, 8 (6), pp. 14-21.

Lewis, J., Fowler, M. (2014). **Microservices**. martinFowler.com, accessed 20 November 2020, <http://martinfowler.com/articles/microservices.html>.

Li, S., Zhang, H., Jia, Z., Zhong, C., Zhang, C., Shan, Z., Shen, J., Babar, M. A. (2020). Understanding and Addressing Quality Attributes of Microservices Architecture: A Systematic Literature Review. **Information and Software Technology**, 131 (106449), pp. 1-23.

Li, Y., Shi, Z., Li, F., Li, H. Y. (2007). Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. **Computers and Electronics in Agriculture**, 56 (2), pp. 174-186.

Lin, J., Lin, L. C., Huang, S. (2016). Migrating web applications to clouds with microservice architectures. In: **Proceedings of the 2016 International Conference on Applied System Innovation (ICASI)**. IEEE Press, pp. 1-4.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In: **Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability**, 1. University of California Press, Berkeley, pp. 281–297.

Martinetz, T. M., Berkovich, S. G., Schulten, K. J. (1993). "Neural-gas" network for vector quantization and its application to time-series prediction. **IEEE Transactions on Neural Networks**, 4 (4), pp. 558-569.

Matheron, G. 1963. Principles of geostatistics. **Economic Geology**, 58 (8), p. 1246-1266.

McBratney, A. B., Moore, A. W. (1985). Application of fuzzy sets to climatic classification. Agricultural and Forest Meteorology. **Goettingen**, 35 (1-4), pp. 165-185.

McQuitty, L. L. (1966). Similarity Analysis by Reciprocal Pairs for Discrete and Continuous Data. **Educational and Psychological Measurement**, 26, pp. 825-831.

Michelon, G. K., Bazzi, C. L., Upadhyaya, S., Souza, E. G., Magalhães, P. S. G., Borges, L. F., Schenatto, K., Sobjak, R., Gavioli, A., Betzek, N. M. (2019). Software AgDataBox-Map to precision agriculture management. **SoftwareX**, 10.

Milligan, G. W., Cooper, M. C. (1988). A study of standardization of variables in cluster analysis. **Journal of Classification**, 5 (2), pp. 181–204.

Moral, F. J., Terrón, J. M., Silva, J. R. M. (2010). Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. **Soil Tillage Research**, 106 (2), pp. 335-343.

Moreira, W. K. O. (2019). **Módulo computacional para delineamento de mapas de aplicação de calcário a partir dos atributos químicos do solo**. Master's dissertation, Western Paraná State University (UNIOESTE), Cascavel.

Paccioretti, P., Córdoba, M., Balzarini, M. (2020). FastMapping: Software to create field maps and identify management zones in precision agriculture. **Computers and Electronics in Agriculture**, 175, p. 1-7.

Pal, N. R., Bezdek, J. C., Hathaway, R. J. (1996). Sequential competitive learning and the fuzzy c-means clustering algorithm. **Neural Networks**, 9 (5), pp. 787-796.

Pauletti, V., Motta, A. C. V. (2019). **Manual de adubação e calagem para o estado do Paraná**, 2. Núcleo Estadual Paraná da Sociedade Brasileira de Ciência do Solo - NEPAR-SBCS, Curitiba. 289 p.

Peralta, N. R., Costa, J. L., Balzarini, M., Franco, M. C., Córdoba, M., Bullock, D. (2015). Delineation of management zones to improve nitrogen management of wheat. **Computers and Electronics in Agriculture**, 110 (1), pp. 103–113.

Pimentel-Gomes, F. (2009). **Curso de estatística experimental**, 15. FEALQ, Piracicaba.

Ponce, F., Márquez, G., Astudillo, H. (2019). Migrating from monolithic architecture to microservices: A Rapid Review. In: **International Conference of the Chilean Computer Science Society (SCCC)**, 38. Concepcion, Chile, pp. 1-7.

Reich, R. M. (2008). **Spatial Statistical Modeling of Natural Resources**. Colorado State University, Fort Collins.

Ribeiro, P. J., Diggle, P. J. (2018). **geoR**: Analysis of Geostatistics Data.

Rousseeuw, P. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. **Journal of Computational and Applied Mathematics**, 20 (1), pp. 53-65.

Rupnik, R., Kukar, M., Vračar, P., Košir, D., Pevec, D., Bosnić, Z. (2019). AgroDSS: A decision support system for agriculture and farming. **Computers and Electronics in Agriculture**, 161, pp. 260-271.

Schenatto, K., Souza, E. G., Bazzi, C. L., Bier, V. A., Betzek, N. M., Gavioli, A. (2016). Data interpolation in the definition of management zones. **Acta Scientiarum**, 38 (1), pp. 31-40.

Schenatto, K., Souza, E. G., Bazzi, C. L., Gavioli, A., Betzek, N. M., Beneduzzi, H. B. (2017a). Normalization of data for delineating management zones. **Computers and Electronics in Agriculture**, 143 (1), pp. 238-248.

Schenatto, K., Souza, E. G., Bazzi, C. L., Gavioli, A, Michelon, G. K. (2017b). Software de gerenciamento de dados agrícola: AGDATAFIELD_MOBILE. In: Rosalen, D. L., erbato, C., Turco, J. E. P (Eds.), **A importância da Engenharia Agrícola para a segurança alimentar,** 1. Sociedade Brasileira de Engenharia Agrícola, pp. 1-10.

Schepers, A. R., Shanahan, F. J., Liebig, M. A., Schepers, J. S., Johnson, S. H., Luchiari, J. A. (2004). Appropriateness of management zones for characterizing spatial variability of soil properties and irrigated corn yields across years. **Agronomy Journal**, 96 (1), pp. 195–203.

Soldani, J., Tamburri, D. A., Van Den Heuvel, W. J. (2018). The pains and gains of microservices: A Systematic grey literature review. **Journal of Systems and Software**, 146 (1), pp. 215–232.

Souza, E. G., Bazzi, C. L., Khosla, R., Uribe-Opazo, M. A., Reich, R. M. (2016). Interpolation type and data computation of crop yield maps is important for precision crop production. **Journal of Plant Nutrition**, 39 (4), pp. 531-538.

Souza, E. G., Schenatto, K., Bazzi, C. L. (2018). Creating thematic maps and management zones for agriculture fields. In: **Proceedings of the 14th International Conference On Precision Agriculture** (IPCA).

Swindell, J. (1997). Mapping the spatial variability in the yield potential of arable land through GIS analysis of sequential yield maps. In 1st European Conference on Precision Agriculture (pp. 827- 834). Warwick.

Taibi D., Lenarduzzi V., Pahl C. (2017). Processes, motivations, and issues for migrating to microservices architectures: An empirical investigation. **IEEE Cloud Comput**., 4 (5), pp. 22-32.

Taneja, M., Byabazaire, J., Davy, A., Olariu, C. (2018). Fog assisted application support for animal behaviour analysis and health monitoring in dairy farming. In: **IEEE World Forum on Internet of Things** (WF-IoT), 4. Singapore, pp. 819-824.

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. **Journal of the American Statistical Association**, 58 (301), pp. 236-244.

Xiang, L., Yu-Chun, P., Zhong-Qiang, G., Chun-Jiang, Z. (2007). Delineation and Scale Effect of Precision Agriculture Management Zones Using Yield Monitor Data Over Four Years. **Agricultural Sciences In China**, 6 (2), pp. 180-188.

Xu, R., Wunsch, D. C. (2009). **Clustering. Piscataway**: IEEE Press.

Appendix A

Normalization methods:

- **Range** (Equation A1): it is based on the dataset range that is wanted to be normalized. According to Anderberg (1973) and Milligan and Cooper (1988), it is not indicated when there are outliers in the data.

$$Z_{iN} = \frac{X_i - Median}{Max(X) - Min(X)}, \tag{A1}$$

where, Z_{iN} – normalized observation i; X_i – original data value i; $Min(X)$ – minimum value of dataset; $Max(X)$ – maximum value of the dataset.

- **Mean** (Swindel, 1997 – Equation A2): it is well known and employed, hoping that the means represent dataset well. However, for Anderberg (1973), the mean value is sensitive and may be changed by adding any constant, thus, easily modifying the normalized data distribution.

$$Z_{iN} = \frac{X_i}{\bar{X}}, \tag{A2}$$

where, Z_{iN} – normalized observation i; X_i – original data value i; $\bar{X}$ – sample mean of dataset.

- **Standard score or Z-score** (Larscheid and Blackmore, 1996 – Equation A3): is used to transform normal variables to standard score where the transformed variable will have a mean of 0.0 and a variance of 1.0.

$$Z_{iN} = \frac{X_i - \bar{X}}{s}, \tag{A3}$$

where, Z_{iN} – normalized observation i; X_i – original data value i; $\bar{X}$ – sample mean of dataset; s – standard deviation of dataset.

- **Min-Max method** (Milligan and Cooper, 1988 – Equation A4): it is a variation of the range method containing changes in the numerator, and they present in the numerator, in which case the normalized data will vary from 0 to 1.

$$Z_{iN} = \frac{X_i - Min(X)}{Max(X) - Min(X)}, \tag{A4}$$

where, Z_{iN} – normalized observation i; X_i – original data value i; $Min(X)$ – minimum value of dataset; $Max(X)$ – maximum value of dataset.

Comparison among thematic maps and among management zones

- **Kappa coefficient** (Kp – Equation A5 – Cohen, 1960; Congalton, 1991): measures the degree of agreement among MZ maps generated by clustering algorithms. Landis and Koch (1977) proposed the following classification: 0 < Kp ≤ 0.2 indicates no agreement, 0.2 < Kp ≤ 0.4 weak agreement, 0.4 < Kp ≤ 0.6 moderate agreement, 0.6 < Kp ≤ 0.8 strong agreement, and 0.8 < Kp ≤ 1 very strong agreement.

$$K_p = \frac{\{n\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}(x_{i+} * x_{+i})\}}{\{n^2 - \sum_{i=1}^{r}(x_{i+} * x_{+i})\}}, \quad \text{(A5)}$$

where X_{ii} is the value in row i and column i, $X_{i+}$is the sum of line i, and X_{+i} is the sum of column i of the error matrix, N is the total number of points interpolated and sorted by the matrix, and c is the number of classes of the error matrix.

- **Global accuracy** (GA – Equation A6 – Foody, 2002): like *Kp*, GA measures the degree of agreement among maps (MZs) and corresponds to the simple percent agreement.

$$GA = \frac{\sum_{i=1}^{c} x_{ii}}{n}, \quad \text{(A6)}$$

where, $\sum_{i=1}^{c} x_{ii}$ is the sum of the main diagonal of the error matrix with c classes and a total of N collected samples (number of interpolated points).

- **Coefficient of relative deviation** (CRD – Equation A7 – Coelho et al., 2009): it calculates the mean difference in modulus of the interpolated values on a thematic map compared to a map taken as reference.

$$CRD = \sum_{i=1}^{n} ABS(\frac{Zi_B - Zi_A}{Zi_A}), \quad \text{(A7)}$$

where Zi_A is the estimated value at location i on the reference map, Zi_B is the value at location i on the map to be compared, and n is the total number of interpolated locations on the maps.

- **Mean absolute difference** (MAD – Equation A8): it computes the mean absolute difference among values on both maps.

$$MAD = \frac{\sum_{i=1}^{n} ABS(Zi_B - Zi_A)}{n}, \quad \text{(A8)}$$

where Zi_A is the value of the location (pixel) i on the reference map, Zi_B is the value at location (pixel) i on the map to be compared, and n is the total number of observations on the maps.

Indices of MZs quality:

- **Variance reduction** (VR% – Equation A9 – Xiang et al., 2007; Schenatto et al., 2017): is calculated for a variable, with the expectation that the sum of data variances for each MZ is smaller than the total variance of the field.

$$VR\% = \left(1 - \frac{\sum_{i=1}^{c} W_i * V_{MZ_i}}{V_{field}}\right) * 100, \quad \text{(A9)}$$

where *c* is the number of MZs; W_i is the ratio of *i-th* MZ to the entire field; V_{mzi} is the data variance of *i-th* MZ; V_{field} is the field data variance.

- **Fuzziness performance index** (FPI – Equation A10 – McBratney and Moore, 1985; Fridgen et al., 2004): measures the degree of separation among fuzzy c groups generated from a data set. FPI varies from 0 to 1.

$$FPI = 1 - \frac{c}{(c-1)}\left[1 - \sum_{j=1}^{n}\sum_{i=1}^{c}(m_{ij})^2/n\right], \tag{A10}$$

  where c is the number of groups; n is the number of elements in the data set; m_{ij} is the element of fuzzy membership matrix M.

- **Modified partition entropy** (MPE – Equation A11 – McBratney and Moore, 1985; Fridgen et al., 2004): estimates the difficulty level to organize c groups.

$$MPE = \frac{-\sum_{j=1}^{n}\sum_{i=1}^{c} m_{ij}\, log(m_{ij})/n}{log\, c}, \tag{A11}$$

  where c is the number of groups; n is the number of elements in the data set; m_{ij} is the element of the fuzzy membership matrix M.

- **Average silhouette coefficient** (ASC – Equation A12 – Rousseeuw, 1987): the ASC coefficient is obtained from the silhouette coefficient (SC), an evaluation index that measures both levels of satisfactory internal formation and external separation of groups. SC value for point p, denoted by sc_p, is calculated using the mean of the intra-group distances ap and the mean of inter-group distances b_p. Kaufman and Rousseeuw (1990) proposed the following classification: 0.71 < ASC ≤ 1.00, a strong structure has been obtained; 0.51 < ASC ≤ 0.70, a reasonable structure has been obtained; 0.26 < ASC ≤ 0.50 the structure is weak and could be artificial (try additional methods); ASC ≤ 0.26 no substantial structure has been obtained.

$$sc_p = \frac{b_p - a_p}{Max(a_p, b_p)}, \tag{A12}$$

  where a_p is the mean of distances from point p to all other points in the same group; b_p is the mean of distances from point p to all points in the closest group containing p.

- **Improved cluster validation index** (ICVI – Equation A13 – Gavioli et al., 2016): is a composition of FPI, MPE, and VR% indices.

$$ICVI_i = \frac{1}{3} * \left(\frac{FPI_i}{Max\{FPI\}} + \frac{MPE_i}{Max\{MPE\}} + \left(1 - \frac{VR\%_i}{Max\{VR\%\}}\right)\right), \tag{A13}$$

  where FPI_i is FPI value of *i-th* variable selection method; MPE_i is MPE value of *i-th* variable selection method; $VR\%_i$ is VR% value of *i-th* variable selection method; *Max{Index_X}* represents the maximum value of *Index_X* among the n variable selection methods.

- **Smoothness index** (SI – Gavioli et al., 2016 – Equation A14): it gives pixel-by-pixel frequency of change of classes in a thematic map in horizontal and vertical directions and along the diagonal. It also characterizes smoothness of MZs boundary curves. If a map has an entirely homogeneous area, SI is equal to 100% due to the lack of class changes. On the other hand, if the map is entirely generated with random values, SI% would have a value close to 0.

$$SI = 100 - \left(\frac{\sum_{i=1}^{k} NM_{H_i}}{4P_H} + \frac{\sum_{j=1}^{k} NM_{V_j}}{4P_V} + \frac{\sum_{l=1}^{k} NM_{DD_l}}{4P_{DD}} + \frac{\sum_{m=1}^{k} NM_{DE_m}}{4P_{DE}} \right) * 100, \quad \text{(A14)}$$

where NM_{H_i} is the number of changes in row *i* (horizontal); NM_{V_j} is the number of changes in column *j* (vertical); NM_{DD_l} is the number of changes in diagonal *l* (right diagonal DD); NM_{DE_m} is the number of changes in diagonal *m* (left diagonal DE); k is the maximum number of pixels in a row, column, or diagonal; P_H is the possibility of changes in horizontal pixels; P_V is the possibility of changes in vertical pixels; P_{DD} is the possibility of changes in the right diagonal DD; P_{DE} is the possibility of changes in the left diagonal DE.

- **Fragmentation index** (FI% – Equation A15): it considers how higher is the number of zones (NMZ) in comparison with the number of classes (NC). The higher FI%, the higher fragmentation.

$$FI\% = 100 \frac{MZ - c}{c}, \quad \text{(A15)}$$

where MZ is the number of zones; c is the number of established classes.

- **Global quality index** (GQI – Equation A16 – Beneduzzi, 2020): it looks for finding the best number of classes during ZMs delineation, considering the values of ICVI, SIr, and FIr:

$$GQI_i = \frac{ICVI_i * (100 + FIr_i)}{SIr_i}, \quad \text{(A16)}$$

- **Modified global quality index** (MGQI – Equation A17): this coefficient is an adaptation of GQI to include ASC coefficient. MGQI will be better (close to zero) the higher ASC and smoothness of the rectified map (measured by SIr), the smaller is fragmentation of rectified map (measured by FIr) the smaller ICVI will be, and it will correspond to the best amount of classes for an MZ design.

$$MGQI_i = \frac{ICVI_i * (100 + FIr_i)}{SIr_i * ASC} \quad \text{(A17)}$$

Variables' selection to delineate management zones

- Spatial correlation matrix: variable selection by SCM uses Bivariate Moran's I (Reich, 2008; Schepers et al., 2004 - Equation A18), and applied in three steps, as described by Bazzi et al. (2013) and Schenatto et al. (2016): (i) variables without spatial dependence are eliminated, (ii) variables without correlation with the target variable are eliminated, and (iii) redundant variables are eliminated.

$$I_{YZ} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} y_i z_j}{W\sqrt{m_Y^2 m_Z^2}} \tag{A18}$$

where I_{YZ} is the degree of spatial association between Y and Z variables, ranging from -1 to 1, as it is followed: positive correlation I_{YZ} > 0 and negative correlation I_{YZ} < 0; w_{ij} is the ij element of spatial association matrix, calculated by $w_{ij} = (1/(1 + D_{ij}))$, so that D_{ij} is the distance among i and j points; y_i and z_i: transformed y and z values, respectively, at point i ($i = 1, 2, \ldots, n$), to get a zero average by the formulas $y_i = (y_i - \bar{Y})$ and $z_j = (z_j - \bar{Z})$, where $\bar{Y}$ and $\bar{Z}$ are the sample means of Y and Z variables; W is the sum of spatial association degrees obtained by w_{ij} matrix, for $i \neq j$; m_Y^2 and m_Z^2: sample variance of Y and Z variables, respectively.

- PCA: PCs calculation from all stable variables, such that the number of PCs selected is based on the criterion of representation of at least 80% of the total variability of data associated with the original variables (Johnson and Wichern, 2007);
- MULTISPATI-PCA: calculation of spatial principal components (SPCs) from all stable variables that were significantly correlated with the target variable, such that the amount of SPCs selected was also based on the criterion of representation of at least 80% of the total variability of the original data (Córdoba et al., 2016; Peralta et al., 2015; Gavioli et al., 2016).

Selection of interpolators

The selection of interpolators determined the best model among the ordinary kriging and inverse distance weighting by the interpolator selection index (ISI - Bier and Souza, 2017 - Equation A19). ISI is determined from cross-validation (Isaaks and Srivastava, 1989), which calculates the mean error (ME - Equation A20) and the standard deviation of the mean error (SDME, Equation A21). ME and SDME values calculated for each parameter set are stored and used to determine ISI, which compares deterministic and stochastic interpolation methods, thus, identifying the best adjustment for each model analyzed. The best interpolator registered the lowest ISI value.

$$ISI = \left\{ \frac{abs(ME)}{max\Big|_{i=1}^{j}[abs(ME)]} + \frac{\left[SDME - min\Big|_{i=1}^{j} SDME\right]}{max\Big|_{i=1}^{j}[abs(SDME)]} \right\}, \tag{A19}$$

where ME is the mean error; $SDME$ is the standard deviation of mean error of the crossed validation; n is the number of data; abs is the module value; $min|_{i=1}^{j}$ is the lowest value among the compared j models; $max|_{i=1}^{j}$ is the highest value among the compared j models.

$$ME = \frac{1}{n}\sum_{i=1}^{n} Z(s_i) - \hat{Z}(s_i), \tag{A20}$$

$$SDME = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Z(s_i) - \hat{Z}(s_i)\right)^2}, \tag{A21}$$

where n is the number of data; $Z(s_i)$ is the value observed at the point s_i; $\hat{Z}(s_i)$ is the predicted value at the point s_i.

The statistic called error comparison index (ECI – Souza et al., 2016 – Equation A22) was used to determine the best semivariogram fit in each j model analyzed, which assumes that the lowest value for the model is the best stochastic methods of interpolation. The best semivariogram of each j model was used in ISI analysis.

$$ECI_i = \frac{|RME_i|}{10^{-10} + max\left|_{i\,=\,1}^{\,j}|RME|\right.} + \frac{|SDRME_i - 1|}{10^{-10} + max\left|_{i\,=\,1}^{\,j}|SDRME - 1|\right.}, \tag{A22}$$

where ECI_i is the error comparison index for model i; and $max\left|_{i\,=\,1}^{\,j}\right.$ is the highest value among the compared j semivariograms. The arbitrary constant 10^{-10} was included to avoid division by zero.

The reduced mean error (RME – Equation A23) and the standard deviation of the reduced mean error (SDRME – Equation A24) were determined by ordinary kriging cross-validation.

$$RME = \frac{1}{n}\sum_{i=1}^{n}\frac{Z(s_i) - \hat{Z}(s_i)}{\hat{\sigma}(\hat{Z}(s_i))}, \tag{A23}$$

$$SDRME = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{\left|Z(s_i) - \hat{Z}(s_i)\right|}{\hat{\sigma}(\hat{Z}(s_i))}}, \tag{A24}$$

where $Z(s_i) - \hat{Z}(s_i)$ is the prediction error associated by estimating yield at spatial location s_i; $Z(s_i)$ is the observed value; $\hat{Z}(s_i)$ is the estimated value obtained from the ordinary kriging cross-validation; $\hat{\sigma}(\hat{Z}(s_i))$ is the estimated standard deviation associated with the estimated value, and n is the sample size.

Methods of interpolation

- **Inverse distance weighting** (IDW – Equation A25): the interpolation considers a weight for the observed samples. Weights for the points are considered by the inverse of distance increased to power so that the greater the power, the lesser the influence of the most distant points.

$$\hat{Z}_i = \frac{\sum_{i=1}^{n}\left(\frac{1}{d_i^p} * Z_i\right)}{\sum_{i=1}^{n}\left(\frac{1}{d_i^p}\right)}, \tag{A25}$$

where, $\hat{Z}_i$ – interpolated value; Z_i – sampled attribute value; d_i^p – Euclidean distance among the i[th] neighborhood point and the sampled point, elevated to the power of p > 0.

- **Ordinary Kriging** (OK – Equation A26 – Cressie, 1993): is made after adjusting the semivariogram model, and the value to be estimated at the point of interest.

$$\hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i * \mathrm{Z}\,(x_i), \tag{A26}$$

where $\hat{Z}(x_0)$ – estimated value at a given location; λ_i – weight attributed to the sample values; $Z(x_i)$ – sampled attribute value; n – number of neighboring locations employed for interpolating the point, where summation of λ_i weights must be equal to one.

- **Moving average** (MA – Equation A27): estimates non-sampled point values based on the mean of sampled points within a predefined radial distance as given i. The points within predefined radial distance are equally weighted (i.e, weight is 1/n) and the resulting value is the arithmetic average of the identified neighboring data. All points within the predefined radial distance are equally weighted (i.e, weight is 1/n) (Bazzi et al., 2015).

$$\hat{Z}_i = \frac{\sum_{i=1}^{n} Z_i}{\mathrm{n}}, \tag{A27}$$

where: $\hat{Z}_i$ is the interpolated value of the non-sampled point; Z_i is the value of the neighboring sample point; n is the number of neighboring sample points used to interpolate the non-sampled point.

- **Nearest neighbor** (NN): assigns the value of the nearest point to each.

Appendix B

Table 9 Resources available in the AgDataBox Statistics API

Resources	Method	Request Parameters	Responses
Descriptive	POST	• Dataset	• Count, mean, mode, median, standard deviation, variance, median absolute deviation, interquartile range, minimum, maximum, and quartiles.
Normality	POST	• Dataset	• One-sample Kolmogorov-Smirnov test. • Lilliefors (Kolmogorov-Smirnov) normality test. • Cramer-von Mises normality test. • Shapiro-Wilk normality test. • Shapiro-Francia normality test. • Anderson-Darling normality test.
PCA	POST	• Multiple datasets	• Eigenvalue, eigenvector, and observation scores.
PCA graphs	POST	• Multiple datasets • Graph type: Scree-plot, Biplot	• Graph.
Kappa	POST	• Two datasets	• Kappa statistics.
Global Accuracy	POST	• Two datasets	• Global Accuracy statistics.
Normalization	POST	• Dataset	• Dataset normalized.
Data cleaning	POST	• Dataset	• Dataset cleaned (outlier, null, and duplicated values).

Table 10 Resources available in AgDataBox Spatial API (ADB-SP-API)

Resources	Script languages	Coordinate systems	Request parameters	Responses
Coordinate's converter	R	UTM and Lat/Long	• Coordinate set • Input datum • Output datum	• Coordinate set in the output datum
Moran's i	R	UTM	• Dataset	• Moran's I statistic • p-value
Bivariate Moran's i	R	UTM	• Two datasets	• Bivariate Moran's I statistic • p-value
Multispati PCA	R	UTM	• Multiple datasets	• Eigenvalue • Eigenvector • Observation scores
Downgrid	Javascript (Turf.js)	Lat/Long	• Dataset • Destination grid • Method: Moving average or Nearest Neighbors	• The dataset in the destination grid
Smoothness Index	R	UTM	• Dataset	• Smoothness index
Variance reduction	R	UTM	• Clustering dataset • Dataset of the target variable	• Percentage of variance reduction
Descriptive statistics by cluster	R	UTM	• Clustering dataset • Dataset of the target variable	• Descriptive statistical measures by cluster • Tukey's test

UTM: universal transverse Mercator; Lat: latitude; Long: longitude.

Table 11 Resources of the AgDataBox Interpolation API

Resources	Method	Request parameters	Responses
Ordinary Kriging	POST	• Semivariogram parameters (Model, nugget effect, partial sill, range, cutoff) • Sampling dataset • Area boundaries	• Interpolated dataset
IDW	POST	• Exponent • Neighbors • Sampling dataset • Area boundaries	• Interpolated dataset
Moving Average	POST	• Neighbors • Sampling dataset • Area boundaries	• Interpolated dataset
Nearest neighbors	POST	• Sampling dataset • Area boundaries	• Interpolated dataset
ISI for IDW	POST	• Minimum exponent • Maximum exponent • Step for exponent • Minimum neighbors • Maximum neighbors • Sampling dataset	• Result list of the IDW analyses.
ISI for Geostatistics	POST	• Models • Methods • Kappas • Lambda • Lags • Estimator • Maximum distance • Pairs • Amount of range intervals • Amount of partial sill intervals • Sampling dataset	• Result list of the Geostatistical analyses.
Semivariogram	POST	• Model • Kappa • Nugget effect • Partial Sill • Range • Maximum distance • Pairs • Lags • Lambda • Sampling dataset	• Semivariogram image

IDW: inverse distance weighting; ISI: interpolator selection index.

Table 12 Resources available according to the nutrient recommendation microservice (AgDataBox Recommendation API)

Resource	Methods	Request parameters	Responses
P	Nutrient available	• Culture: soybean or corn. • Soil texture: clay or sand. • Source: SS, TS, MAP, DAP, and ARAD. • Dataset (P available) with geographic coordinates.	Recommendation dataset per pixel/observation
	Yield expectation	• Culture: soybean or corn. • Clay content (g kg^{-1}). • Yield expectation (t ha^{-1}). • Source: SS, TS, MAP, DAP, and ARAD. • Dataset (P available) with geographic coordinates.	Recommendation dataset per pixel/observation
K	Nutrient available	• Culture: soybean or corn. • Soil texture: clay or sand. • Source: KCL, PS, PMS, and PDS. • Dataset (K available) with geographic coordinates.	Recommendation dataset per pixel/observation
	Yield expectation	• Culture: soybean or corn. • Yield expectation (t ha^{-1}). • Source: KCL, PS, PMS, and PDS. • Dataset (K available) with geographic coordinates.	Recommendation dataset per pixel/observation
N	Yield expectation	• Culture: only corn. • Previous culture: grass or legume. • Yield expectation (t ha^{-1}). • Efficiency coefficient. • Source: UR, AM, AN, and AC. • Dataset (OM available) with geographic coordinates.	Recommendation dataset per pixel/observation

P: phosphorus; K: potassium; N: nitrogen; OM: organic matter; SS: simple superphosphate; TS: triple superphosphate; MAP: monoammonium phosphate; DAP: diamonic phosphate; ARAD: Araxa phosphate; KCL: potassium chloride; PS: potassium sulphate; PMS: potassium and magnesium sulfate; PDS: potassium double saltpeter; UR: urea; AM: ammonium sulfate; AN: ammonium nitrate; AC: ammonium chloride.

Table 13 Resources available in AgDataBox Clustering API

Resources	Request parameters	Metric	Responses
average linkage	• Multiple datasets with GC. • The number of classes.	-	• Clustering dataset. • Index: ASC.
centroid linkage	• Multiple datasets with GC. • The number of classes.	-	• Clustering dataset. • Index: ASC.
complete linkage	• Multiple datasets with GC. • The number of classes.	-	• Clustering dataset. • Index: ASC.
hybrid hierarchical clustering	• Multiple datasets with GC. • The number of classes.	-	• Clustering dataset. • Index: ASC.
median linkage	• Multiple datasets with GC. • The number of classes.	-	• Clustering dataset. • Index: ASC.
McQuitty's method	• Multiple datasets with GC. • The number of classes.	-	• Clustering dataset. • Index: ASC.
Ward's method	• Multiple datasets with GC. • The number of classes.	-	• Clustering dataset. • Index: ASC.
bagged clustering	• Multiple datasets with GC. • The number of classes. • Centers. • Hierarchical method. • Iterations.	B, C, E, MN, MX, MI	• Clustering dataset. • Index: ASC.
clustering large applications	• Multiple datasets with GC. • The number of classes.	B, E, MN	• Clustering dataset. • Index: ASC.
fuzzy analysis clustering	• Multiple datasets with GC. • The number of classes. • Iterations. • Weighting exponent.	E, MN	• Clustering dataset. • Index: ASC.
fuzzy c-means	• Multiple datasets with GC. • The number of classes. • Iterations. • Weighting exponent.	E, MN	• Clustering dataset. • Indices: ASC, PC, FPI, PE, MPE, XB
hard competitive learning	• Multiple datasets with GC. • The number of classes. • Iterations.	E, MN	• Clustering dataset. • Index: ASC.
k-means	• Multiple datasets with GC. • The number of classes. • Iterations.	E, MN	• Clustering dataset. • Indices: ASC, SS, WCSS.
neural gas	• Multiple datasets with GC. • The number of classes. • Iterations.	E, MN	• Clustering dataset. • Index: ASC.
partitioning around medoids	• Multiple datasets with GC. • The number of classes.	E, MN	• Clustering dataset. • Index: ASC.
spherical k-means	• Multiple datasets with GC. • The number of classes. • Internal method. • Weighting exponent.	-	• Clustering dataset. • Index: ASC.
unsupervised fuzzy	• Multiple datasets with GC. • The number of classes.	E, MN	• Clustering dataset.

competitive learning	• Iterations. • Weighting exponent.		• Indices: ASC, PC, FPI, PE, MPE, XB.

GC: geographic coordinates; B: binary. E: Euclidean; MN: Manhattan; MX: maximum; MI: Minkowski; ASC: average silhouette coefficient; PC: partition coefficient; FPI: fuzziness performance index; PE: partition entropy; MPE: modified partition entropy; XB: Xie and Beni index; SS: sum of squares; WCSS: within-clusters sum of squares.

Table 14 Resources available in AgDataBox Rectification API

Resources	HTTP Method	Request parameters	Response
Median	POST	• Dataset with GC. • Size of Kernel. • Kernel format. • Iterations.	• Rectified dataset
Close	POST	• Dataset with GC. • Size of Kernel. • Kernel format. • Iterations.	• Rectified dataset
Open	POST	• Dataset with GC. • Size of Kernel. • Kernel format. • Iterations.	• Rectified dataset
Open and Close	POST	• Dataset with GC. • Size of Kernel. • Kernel format. • Iterations.	• Rectified dataset

GC: Geographic coordinates.

6 PAPER 2 – AGDATABOX-MAP: A WEB APPLICATION FOR CREATING THEMATIC MAPS AND MANAGEMENT ZONES IN PRECISION AGRICULTURE AND DIGITAL AGRICULTURE

ABSTRACT: Agriculture has been challenged to produce more, with greater profitability and less environmental impact. Therefore, digital technologies have become valuable tools for data collection, information analysis, and decision-making. Thus, this work aimed to develop the AgDataBox-Map (ADB-Map) application, whose purpose is to generate thematic maps (TMs) and delineate management zones (MZs) in an easy, friendly, and efficient way. The application was developed for the web, consuming resources from ADB microservices architecture (ADB-MSA), and provides graphical interfaces to prepare data, perform statistical analysis, select variables, interpolate data, delineate MZs, and calculate lime and fertilizer recommendations. In the case study, it was possible to select the best MZ design (elevation variable selected by the spatial correlation matrix, divided into two classes by the Fuzzy C-Means method) using the Modified Global Quality Index (MGQI), defined and applied in this work, which is composed of other indices available in ADB-Map. The application was considered acceptable by the community with a quality evaluation made by the users.

KEYWORDS: data interpolation, variable-rate application, site-specific management.

6.1 Introduction

The new era of agricultural production has converged towards the massive use of digital technologies to improve the production process. The continuous need for food production has stimulated the development of technologies applied to agriculture. The challenge for agriculture is to produce more, with greater profitability and less environmental impact. In this context, precision agriculture (PA) and digital agriculture (DA) are included. PA is not a recent practice, as studied since the 1980s (Zhang et al., 2002). However, this technology only gained greater prominence with the diffusion of Global Position System (GPS) technologies, currently called Global Navigation Satellite System (GNSS), Geographic Information System (GIS), and remote sensing technologies.

Fertilizer distribution varies according to soil conditions in an agricultural field was one of the first practices in PA. Its evolution allowed adopting practices such as automatic guidance of agricultural vehicles and implements, autonomous machinery and processes, product traceability, on-farm research, and software for the overall management of agricultural production systems (Gebbers and Adamchuk, 2010).

Industry 4.0, or "Fourth Industrial Revolution," has rapidly transformed several sectors from "disruptive" digital technologies such as Blockchain, Internet of Things, Artificial Intelligence, and Immersed Reality (Trendov et al., 2019). In DA, available information and knowledge are used to enable automation of sustainable processes in agriculture. Large amounts of data are generated, and the challenge is to add value to them with the insertion of data portals and work platforms. In portals, the user can view their data without having to enter them manually. By platforms, this user can transform data into new and more robust information.

More than 570 million small farms worldwide (Lowder et al., 2016), and agriculture and food production account for 28% of the entire global workforce (Ilostat, 2019 as cited in Trendov et al., 2019). From the evolution of wireless communications means, IoT hardware supply, and cost reduction, it is expected that DA collaborates to democratize the digital technologies application in agriculture, not only for large rural producers but also to reach the small ones. Furthermore, Digital technologies are creating opportunities to integrate smallholders into a digital agri-food system (Gray et al., 2018).

The digitization of agricultural sector depends on some primary conditions that, according to Trendov et al. (2019), include infrastructure availability and connectivity (mobile subscriptions, network coverage, Internet access, and electricity supply), accessibility, educational achievement (literacy, education in information and communication technologies (ICT) and institutional support. Most of these factors depend on government intervention. However, non-governmental initiatives can contribute to agriculture's digitization. Furthermore, it is essential to make specific portals and platforms available to keep up with the growing demand for agricultural data processing. Therefore, Brazil needs to be aware of this technological evolution and make available free web platforms to integrate data, software, procedures, and methodologies for PA.

The web platform AgDataBox (ADB) has collaborated with this phase of agriculture, enabling PA methodologies. Digital platforms can be seen as "software and web applications that act as mediators among service providers and service recipients" (Hanafizadeh et al., 2020). In this sense, the ADB is a platform that provides free computational tools for farmers, researchers, and service providers, focused on PA by the integration of data, software, procedures, and methodologies to contribute to agriculture development in the country using free technologies. This web platform has a microservices architecture (MSA), called ADB-MSA, which consists of a set of resources accessible remotely, through the hypertext transfer protocol (HTTP), to process and store data from agricultural environment. ADB-MSA allows interoperability of several applications in which data and processing routines are centralized. The following applications, underdevelopment, consume ADB-MSA resources: 1) ADB-Mobile; 2) ADB-Map; 3) ADB-Admin; 4) ADB-IoT; 5) ADB-Remote Sensing.

Applications software are essential resources for PA development. Cisternas et al. (2020) identified in literature that the most cited terms in PA publications are GIS, multispectral images, soil mapping, variable rate applications (VRA), variable rate fertilization (VRF), variable rate irrigation (VRI), yield maps and yield monitors, and GIS is the most researched one of them. Furthermore, GIS is often combined with yield maps and other technologies.

Some protocols have already been proposed (Santos and Saraiva, 2015; Cordoba et al., 2016; Souza et al., 2018) to delineate MZs properly. For example, in the protocol by Souza et al. (2018), the process to delineate MZs follows phases of (i) data processing, (ii) data normalization, (iii) selection of variables to design MZs, (iv) data interpolation, (v) application of methods to design MZs, (vi) MZs rectification and (vii) MZs evaluation (Fig. 1).

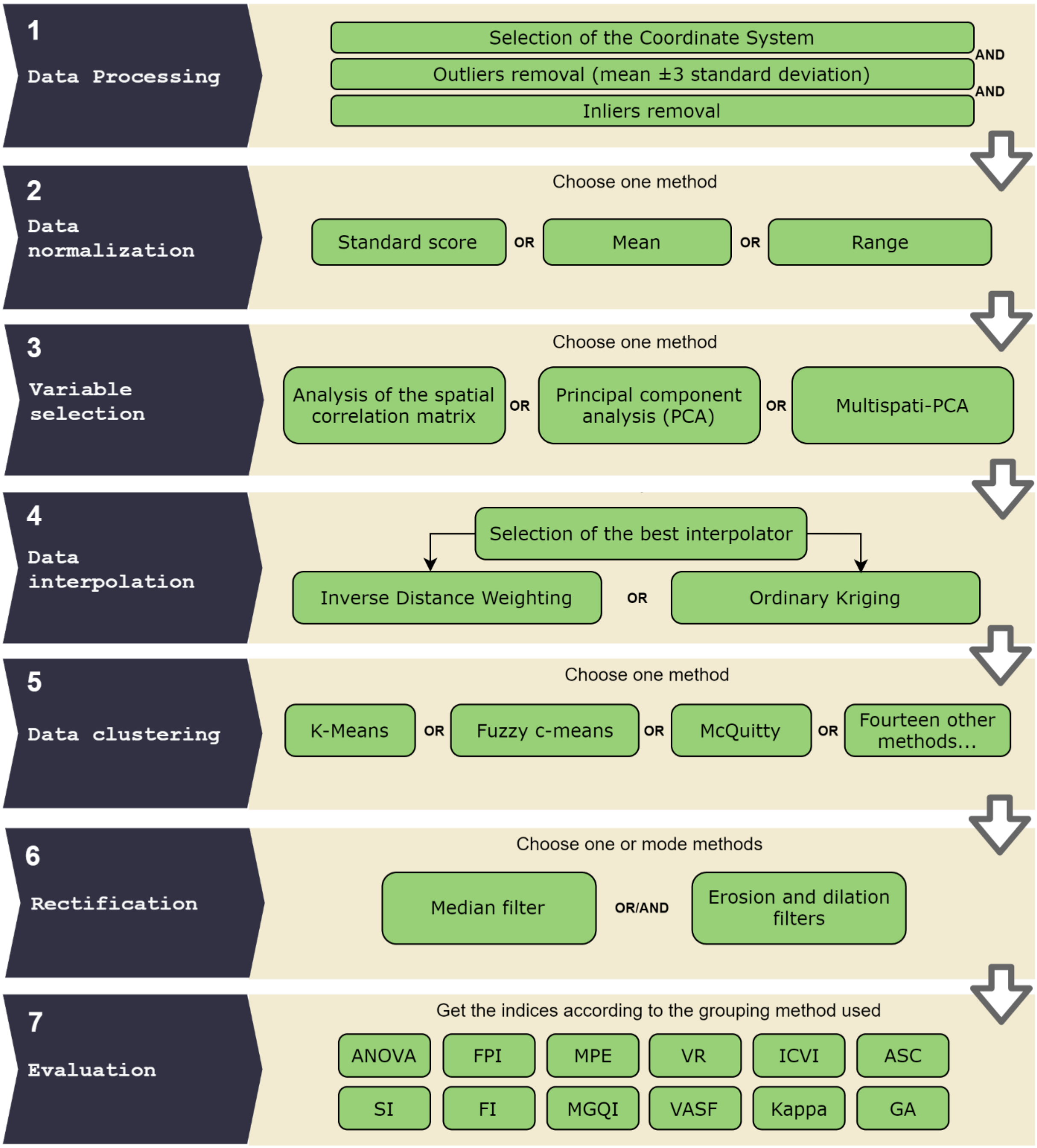


Fig. 1 Protocol steps for the design of management zones.
Source: Adapted from Souza et al. (2018).

In literature, it is not easy to find an application that offers all the necessary features to create TMs and delineate MZs. Examples of software that are specific for MZ delineation are Management Zone Analyst (MZA, Fridgen et al., 2004), FuzME (Minasny and McBratney, 2002), Software for Defining Management Zones (SDUM; Bazzi et al., 2013; Bazzi et al., 2019), ZoneMAP (Zhang et al., 2010), a friendly interface software proposed by Albornoz et al. (2017), and FastMapping (Paccioretti et al., 2020). MZA is a precursor and very popular software in MZ development (Breunig et al., 2020; Damian et al., 2020; Peralta et al., 2015). However, in many cases, a software combination is made to complete the steps involved in MZs delineation. Among them, there are: Statistical Analysis System (SAS), SPSS statistical software, Statistic (StatSoft Inc., currently maintained by TIBCO Software Inc), GS+, ArcGis (Environmental Systems Research Institute, Redlands, CA), Software R (R Core Team, 2014), FuzMe, MZA, Matlab and GRASS GIS (Damian et al., 2020; Méndez-Vázquez et al., 2019; Oldoni et al., 2019; Behera et al., 2018; Peralta et al., 2015; Chang et al., 2014).

In this sense, ADB-Map application aims at mitigating the problem of using different software to create TMs and delineate MZs and provide user-friendly interfaces and procedures. This proposal converges to digitize agriculture. ADB-Map application continues SDUM, registered with the National Institute of Industrial Property (INPI, abbreviation in Portuguese) (registry BR 51 2014 000720 D), and freely available for use. SDUM was developed in a desktop environment, requiring installation on computers, and, despite SDUM acceptance by researchers and producers, it migrated to a web platform with the inclusion of new modules and features but remaining its gratuity. Among the limiting factors for the adoption of PA management by farmers, such as the high investment cost, small operational field, lack of technical training, compatibility problems between hardware and software, there are the complexity to use some computational tools and time needed to learn how to use them (Nicol and Nicol, 2021; Rotz et al., 2019; Bambini et al., 2013; Reichart and Jürgens, 2009). Therefore, ADB-Map also aims to reduce complexity for PA user learning and usage.

Some studies have resulted in a preliminary version of this web application (Borges et al., 2020; Michelon et al., 2019). On the other hand, with the fast software development technologies transformation, which involves standards, languages, and methodologies, it was necessary to modify the entire application structure so that it could be prepared to receive new features and be able to respond quickly to the implementation demands.

As a GIS, ADB-Map works with spatial data and offers the necessary tools to create TMs and delineate MZs, which are subsidies for PA techniques. ADB-Map's functionalities are divided into conceptual modules (Fig. 2), composed of the back-end, which contains algorithms and rules of business operation, and the front-end, user interaction interface.

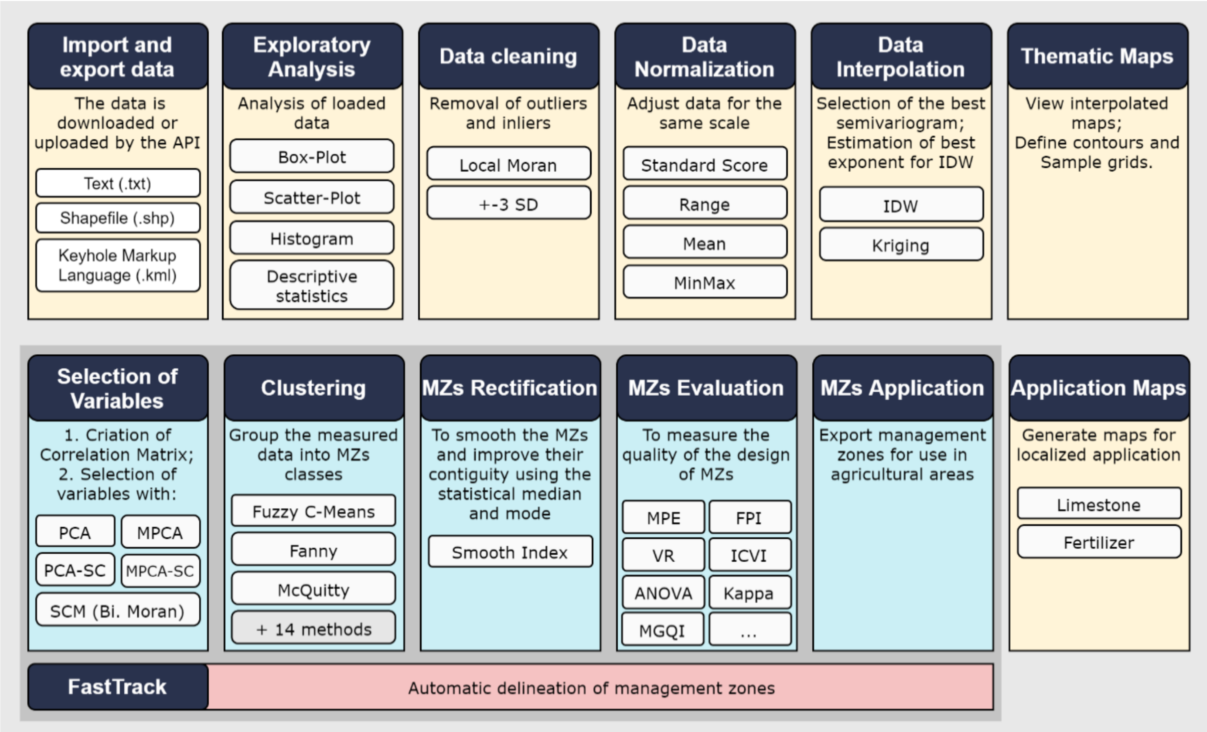


Fig. 2 Overview of modules that make up AgDataBox-Map application.

The modules that integrate ADB-Map are:

a) Import and export: import and export data from/to files;

b) Descriptive and exploratory statistical analysis: calculation of measures of central tendency (mean, median, and mode), measures of dispersion (total amplitude, variance, standard deviation, and coefficient of variation (CV)), measures of distribution shape (asymmetry and kurtosis coefficients), and data normality tests;

c) Data cleaning: removal of duplicate, null, outliers, and inliers observations;

d) Data normalization: variables standardization on the same numerical scale using the methods of amplitude, mean, standard score, and min-max;

e) Data interpolation: selection of the best interpolator for a data set and application of statistical techniques to estimate values in non-sampled locations;

f) Thematic maps: TMs creation using interpolated spatial data and definition of layer styles depending on the type of data classification (equal distances, quantiles, standard deviation, and manual range), number of classes, color palettes, type of marker, and pixel size;

g) Variables selection: use of techniques to select variables for MZs design using these techniques: (i) principal component analysis (PCA), (ii) multivariate spatial analysis based on Moran and PCA index (MULTISPATI-PCA), and (iii) spatial correlation matrix;

h) Clustering: delineate MZs by seventeen data clustering methods (Fuzzy C-Means, K-Means, and others);

i) MZs rectification: apply digital image processing filters (median, opening, closure, and with the combination of opening and closure) to remove small spots in MZs;
j) MZs evaluation: use of descriptive statistics and indices to evaluate the quality of the delineated MZs;
k) MZs application: of MZs export to a field operation;
l) Application maps: create application maps for lime and fertilizer from soil attribute availability maps.

This work aimed to develop the new ADB-Map web application to make TM creation and MZ design processes user-friendly and integrated, making it scalable and integrated with ADB-MSA. ADB-Map was divided into two parts, the front-end, which contains the user's interfaces, and the back-end, which performs functionalities and stores the user's data.

6.2 Material and methods

i. AgDataBox-Map architecture

Compared to the preliminary version, ADB-Map application has been completely rewritten (Borges et al., 2020; Michelon et al., 2019). The new application structure aims to allow scalability and make it more flexible for agile implementation of new demands. In this approach, the application's horizontal separation into two layers, front-end and back-end (Fig. 3), is different from the preliminary version, which was based on the monolithic conception. There is no separation between front-end and back-end according to the monolithic view, but application is treated as a single artifact in Model-View-Controller (MVC) architecture, in which these layers are dependent and difficult to reuse for other applications.

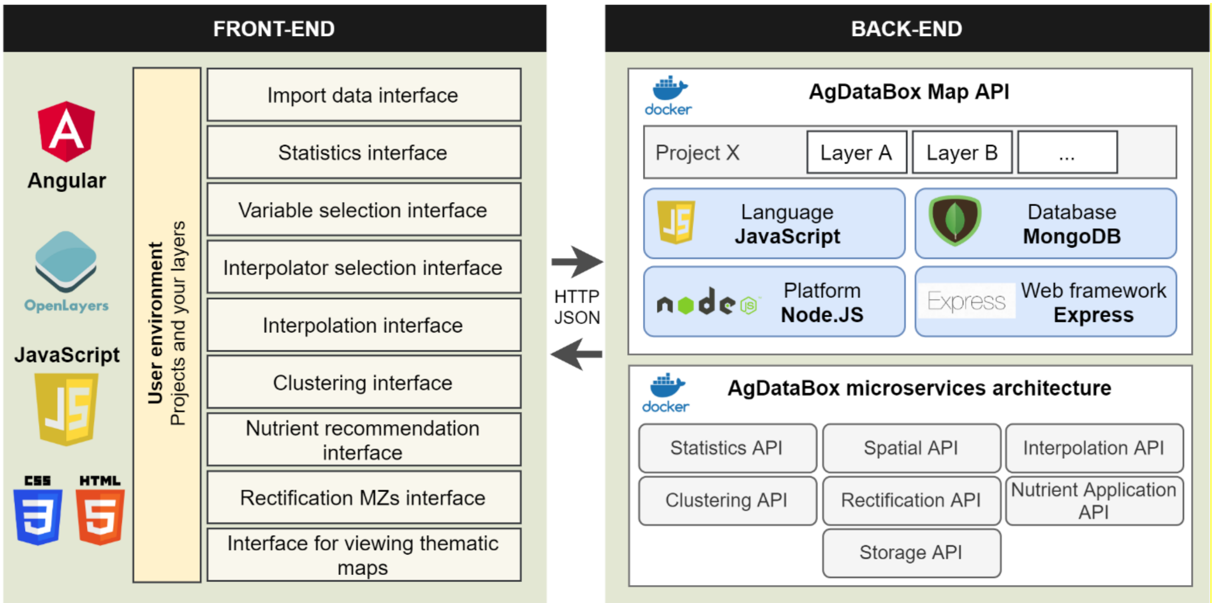


Fig. 3 Architecture of the AgDataBox-Map application containing an overview of the front-end and back-end functionalities.

Each part of the application (front-end and back-end) is deployed in a Docker tool container and the other services of ADB platform. Back-end application comprises several microservices that implement web APIs in the REST architectural style and are deployed in ADB Microservices Architecture (ADB-MSA). Requests to the back-end are made by HTTP, which uses a method (get, post, put or delete), a Uniform Resource Identifier (URI), and a representation of data in a standardized format, in this case, JavaScript Object Notation (JSON). APIs respond to requests with messages in JSON format.

The multilayer architecture approach is also used by other software, like GeoFarmer (Eitzinger et al., 2019), a system of modular components (functionalities and interfaces) that communicate with a central cloud application, which includes the central database where all information is compiled. Cloud applications' backend also communicates with external components and services.

ADB-Map is under the testing phase and is accessible by URL https://adb.md.utfpr.edu.br/map. In addition, there are demonstration videos on URL https://adb.md.utfpr.edu.br/help/map to help people with their first contact with ADB-Map.

ii. AgDataBox-Map application development

ADB-Map front-end (Fig. 3) was implemented to be accessible via Web, with the development of technologies based on JavaScript language, style formatting in Cascading Style Sheets (CSS), and content structuring in Hypertext Markup Language (HTML). The developed platform Node.js was used by the web framework Angular (version 9), TypeScript, the language used by Angular, and OpenLayers (version 6) to present the spatial objects. All technologies used are free to be used.

A microservice was implemented to store ADB-Map data and deployed to ADB-MSA in an operating environment (container) separate from the front-end. This microservice contains a web API implemented with technologies for software development: Node.js platform, Express web framework, and Mongoose library. Data are stored in MongoDB non-relational database (often called NoSQL database), unlike the preliminary version (Borges et al., 2020; Michelon et al., 2019) that uses PostgreSQL. The objective of changing database was to offer flexibility to manage the application's data model and store a large volume of data from the projects' layers (data grids). In this database, data are not stored in tables but in document collections.

Data are organized into two collections, one for the project and the other for layers. Thus, it is possible to represent data models that the application manages (Fig. 4). The project represents a set of layers. A layer represents a set of spatial data, such as crop yield, soil chemical attributes, and satellite image spectral bands.

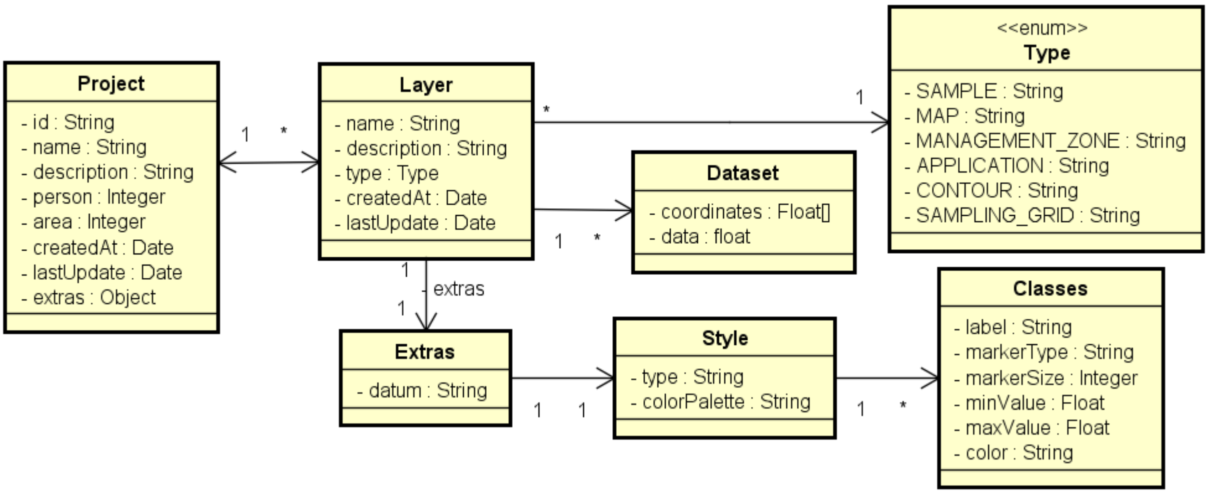


Fig. 4 Class diagram of data managed by AgDataBox-Map application.

Data exchanged between the front-end and back-end is represented in JSON format (Frame 1).

```
{
    "project": {
        "person": number,
        "area": number,
        "name": string,
        "description": string,
        "createdAt": date,
        "lastUpdate": date,
        "extras": {
            "viewMap": boolean
        },
        "id": string
    },
    "type": string,
    "name": string,
    "description": string,
    "data set": [ {"coordinates": [number, number], "data": number }, ...]
}
```

Frame 1 Example of data represented in JSON format.

Layers are categorized according to the representation of the data type:

- Sampling grid: sample data set, usually with low data density, as soil chemical analysis received from a laboratory.
- Defining sampling grid: definition of a sampling grid with the respective georeferenced sampling points to carry out sample collections.
- Interpolated grid: data set, usually with a higher sample density, arranged evenly in the spacing among observations.
- Application grid: calculation result for lime and fertilizer recommendation;
- Management zone: resulting from clustering processes.
- Boundaries: a set of geographic coordinates with the field boundaries.

iii. Data importation and exportation

Data import and export in ADB-Map are done from text files separated by columns (.txt or .csv). The imported data are stored in ADB-MSA (Fig. 3) and consulted by ADB-Map or another application. The file must have rows and columns, where the columns contain geographic coordinates (X and Y) and data variable are separated by a special character, such as tab, space, semicolon, or comma. Each line in the file represents an observation of the variables.

The import process is done following the steps below:

- Data file pre-processing: this step occurs before using the software, in which the user prepares data to be read by ADB-Map. During this step, identifying geographic coordinates, variables numerical formatting, standardization of column separators, and removing orphan lines must be observed.

- The user selects the input file that contains geographic coordinates and variables' data.
- During the import screen, the user defines the file import parameters to recognize the correct way in which data are arranged.
- Select the DATUM of geographic coordinates.
- Indicate the columns of file that represent X and Y geographic coordinates.
- Select which variables will be imported.
- After user intervention, a computational process reads and stores data in ADB-MSA.

iv. Statistical analysis

In ADB-Map, graphical interfaces were implemented to perform descriptive statistical analysis for the layers. The analyses are obtained by computational routines available in microservices of ADB-MSA (Fig. 3). The graphical interfaces to perform statistical procedures are (i) screen for statistical analysis of quantitative layers (Sampling grids, interpolation, and nutrient/liming recommendation), (ii) screen for statistical analysis in MZs, and (iii) screen for data normalization.

During the exploratory data analysis, position measures (mean and median), dispersion measures (variance, standard deviation, and coefficient of variation (CV)), and distribution shape measures (asymmetry and kurtosis coefficient) can be calculated) to identify and evaluate whether data are homogeneous and normally distributed.

Data normality can be tested by the Kolmogorov-Smirnov, Lilliefors, Cramer-Von Mises, Shapiro Wilk, Shapiro Francia, and Anderson Darling tests. The provided normalization methods are (i) Range (Anderberg, 1973; Milligan and Cooper, 1988 – Equation A1), (i) Mean (Swindel, 1997 – Equation A2), (iii) Standard Score or Z-Score (Larscheid and Blackmore, 1996 – Equation A3), and (iv) Min-Max method (Milligan and Cooper, 1988 – Equation A4).

v. Variables' selection to delineate management zones

ADB-Map uses computational routines developed in statistical R software (Gavioli et al., 2016) implemented in ADB-MSA microservices (Fig. 3) to select variables to be used in MZs delineation. The methods implemented in ADB-Map to select variables are spatial correlation matrix (SCM - Reich, 2008; Schepers et al., 2004), principal component analysis (PCA - Hotelling, 1933), and multivariate spatial analysis based on Moran's index and PCA (MULTISPATI-PCA - Dray et al., 2008):

- Spatial correlation matrix: Variable selection by SCM uses the Bivariate Moran's I (Reich, 2008; Schepers et al., 2004 - Equation A5), and applied in three steps, as described by Bazzi et al. (2013) and Schenatto et al. (2016): (i) variables without spatial dependence are eliminated, (ii) variables without correlation with the target variable are eliminated, and (iii) redundant variables are eliminated.

- PCA: PCs calculation from all stable variables, such the number of selected PCs is based on the criterion of representation of at least 80% of total variability of data associated with the original variables (Johnson and Wichern, 2007);
- MULTISPATI-PCA: calculation of spatial principal components (SPCs) from all stable variables that were significantly correlated with the target variable, such the amount of selected SPCs, was also based on the criterion of representation of at least 80% of the total variability of the original data (Córdoba et al., 2016; Peralta et al., 2015; Gavioli et al., 2016).

vi. Selection of interpolators

The selection of interpolators was obtained by using computational routines developed in statistical R software (Betzek et al., 2019), and geoR package, which are available in ADB-MSA (Fig. 3). Routines determine whether the best interpolator for a data set is Ordinary Kriging (OK) or Inverse Distance Weighting (IDW). For the OK interpolator (Equation A16 – Cressie, 1993), the theoretical model best fits the experimental semivariogram, and its parameters (nugget effect, partial sill, and range) are determined. Also, the exponent value and number of neighbors are determined by IDW interpolator (Equation A15).

In geostatistical analysis, the computational routine tests seven different semivariogram models (spherical, gaussian, exponential, Matérn 0.5, Matérn 1.0, Matérn 1.5 and Matérn 2.0) and two statistical methods to optimize semivariogram fit (ordinary least squares (OLS) and weighted least squares (WLS – Cressie, 1985)), totaling fourteen different models. During IDW analysis, the routine analyzes by default twelve different values for the exponent (0.5; 1.0; 1.5; ...; and 6.0) and the number of neighbors from 4 to 12, totaling 88 analyses. The best model is selected by the interpolator selection index (ISI – Bier and Souza, 2017 – Equation A6). ISI is determined from cross-validation (Isaaks and Srivastava, 1989), which calculates the mean error (ME – Equation A7) and the standard deviation of the mean error (SDME – Equation A8). The ME and SDME values calculated for each parameter set are stored and used to determine ISI, which compares the deterministic and stochastic interpolation methods, thus, identifying the best adjustment for each model analyzed. The best interpolator is the one with the lowest ISI value.

The statistic called error comparison index (ECI – Souza et al., 2016 – Equation A9) was used To determine the best semivariogram fit in each j model analyzed, which assumes that a lower value for the model is better stochastic methods of interpolation. The best semivariogram of each j model was used in ISI analysis. The reduced mean error (RME –

Equation A10) and the standard deviation of the reduced mean error (SDRME – Equation A11) were determinated by ordinary kriging cross-validation.

The selection of the best semivariogram model considers three selection criteria when performing the best interpolator analysis: (i) a minimum of 25% of effective spatial dependence (%ESD – Equation A12), (ii) the selected semivariogram model should contemplate a fraction of SD due to only the first semivariance significance index ($\%\gamma(1)$, Equation A13) lower than 50%, and (iii) the degree of inclination between the nugget effect and the last adjusted semivariance, estimated by slope of the model ends index (%SMEI, Equation A14) should be greater than 20%. Otherwise, there is an indication of a pure nugget effect.

Thus, the selection of the best interpolator model should not depend only on the ISI but on the criteria presented in Table 1:

Table 1 Criteria to select the best interpolation method

Criterion 1 Minimum of effective spatial dependence		**Criterion 2** Spatial dependence due only to the first semivariance		**Criterion 3** The model needs to express spatial dependence	**The best interpolation method**
If $\%ESDI$ > 25%	and	If $\%\gamma(1)$ < 50%	and	If %SMEI > 20%	IDW or OK with the lowest ISI
If $\%ESDI$ ≤ 25%	or	If $\%\gamma(1)$ ≥ 50%	or	If %SMEI ≤ 20%	IDW with the lowest ISI

%ESDI: Effective spatial dependence index; $\%\gamma(1)$: First semivariance significance index; IDW: Inverse distance weighting; OK: Ordinary Kriging; ISI: Interpolator selection index.

In ADB-Map, the interpolator selection routines are used in three screens:

- Screen for interpolator selection: determines the best interpolator (OK or IDW) and determines the interpolation parameters.
- Screen for IDW interpolation: determines the best exponent value and the number of neighbors to be used in interpolation.
- Screen for OK interpolation: selects a theoretical model that best fits the experimental semivariogram and determines its parameters (nugget effect, partial sill, and range).

vii. Data interpolation

Graphical interfaces were implemented for each interpolation method available for better interaction with the user in ADB-MSA. The interpolation is performed by computational routines developed in statistics R software and made available in a microservice at the ADB-MSA (Fig. 3).

The data interpolation methods available in ADB-Map are (i) IDW (Equation A15), (ii) OK (Equation A16), (iii) Moving Average (MA – Equation A17), and (iv) Nearest Neighbor (NN).

viii. Data clustering

A specific interface performs the data clustering process in ADB-Map. It is possible to choose one of the seventeen clustering methods (Table 2) available for grouping layers by a computational process, which aims to find groups of similar data, delimiting different groups within a data set. The protocol for designing MZs (Souza et al., 2018) uses these clustering methods and has already been evaluated in studies involving MZs (Gavioli et al., 2019).

Table 2 Clustering methods available on AgDataBox-Map

Methods	References
average linkage[a]	Jain and Dubes (1988)
centroid linkage[a]	Jain and Dubes (1988)
complete linkage[a]	Jain and Dubes (1988)
divisive analysis (diana)[a]	Kaufman and Rousseeuw (1990)
hybrid hierarchical clustering[a]	Chipman and Tibshirani (2006)
median linkage[a]	Jain and Dubes (1988)
McQuitty's method (mcquitty)[a]	McQuitty (1966)
Ward's method (ward)[a]	Ward (1963)
single linkage[a]	Jain and Dubes (1988)
bagged clustering[b]	Leisch (1999)
clustering large applications (clara)[b]	Kaufman and Rousseeuw (1990)
fuzzy analysis clustering (fanny)[b]	Kaufman and Rousseeuw (1990)
fuzzy c-means[b]	Bezdek (1981)
fuzzy c-shells[b]	Dave (1992)
hard competitive learning[b]	Xu and Wunsch (2009)
k-means[b]	MacQueen (1967)
neural gas[b]	Martinetz et al. (1993)
partitioning around medoids[b]	Kaufman and Rousseeuw (1990)
spherical k-means[b]	Dhillon and Modha (2001)
unsupervised fuzzy competitive learning[b]	Pal et al. (1996)

a: hierarchical method; b: partitioning method.

In ADB-Map, a graphical interface was implemented to make the data clustering. The procedures for clustering data are performed in computational routines developed in statistical R software and made available in ADB-MSA (Fig. 3).

ix. Management zones rectification

The data rectification interface was implemented to select one or more layers to choose a rectification method and its parameters. The rectification process is performed in ADB-MSA, which has a microservice with rectification routines implemented in Python language and uses OpenCV library. The rectification methods of layers are based on morphological filters used in digital processing of image: median, opening, closure, and with the combination of opening and closure. These indices have already been used to reduce MZs fragmentation (Betzek et al., 2018; Albornoz et al., 2017; Córdoba et al., 2016; Gonzalez and Woods, 2008).

x. Evaluation of management zones quality

Indices to evaluate MZs quality are obtained and presented in some ADB-Map's graphical interfaces. The indices calculation is done in computational routines developed in the statistical R software available in ADB-MSA's microservices (Fig. 3). They are (i) variance reduction (VR% – Equation A18 – Xiang et al., 2007; Schenatto et al., 2017), (ii) Fuzziness Performance Index (FPI – Equation A19 – McBratney and Moore, 1985; Fridgen et al., 2004), (iii) Modified Partition Entropy (MPE – Equation A20 – McBratney and Moore, 1985; Fridgen et al., 2004), (iv) Improved Cluster Validation Index (ICVI – Equation A21 – Gavioli et al., 2016), (v) Tukey test (ANOVA), (vi) Smoothness Index (SI% – Equation A22 – Gavioli et al., 2016), (vii) Average Silhouette Coefficient (ASC – Equation A23 – Rousseeuw, 1987), (viii) Fragmentation index (FI% – Equation A24), (ix) Global Quality Index (GQI – Equation A25 – Beneduzzi, 2020), and (x) Modified Global Quality Index (MGQI – Equation A26).

xi. Comparison between thematic maps and between management zones

Also, ADB-Map's graphical interfaces calculate indices to compare thematic maps with management zones. They are: (i) Coefficient of relative deviation (CRD – Equation A27 – Coelho et al., 2009), (ii) Mean absolute difference (MAD – Equation A28), (iii) Kappa coefficient (Kp – Equation A29 – Cohen, 1960; Congalton, 1991), and (iv) Global accuracy (GA – Equation A30 – Foody, 2002).

xii. Nutrient and liming recommendation

The fertilizer recommendation is based on the study by Beneduzzi (2020), in which the recommendation of N is made by yield expectation model, considering soil's organic matter content (OM%) for corn cropping. Two methods were implemented for P and K: soil nutrient availability and yield expectation, and the recommendation calculation can be done for soybean and corn crops (Table 3). For each nutrient, the recommendation is based on available fertilizers.

Table 3 Fertilizers used in nutrients' recommendation, crops, and recommendation methods available in AgDataBox Recommendation API

Nutrients	Cultures	Methods	Fertilizing
N	Corn	YEOM	Urea (UR) Ammonium sulfate (AM) Ammonium nitrate (AN) Ammonium chloride (AC)
P	Corn and Soybean	SNA and YE	Simple Superphosphate (SS) Triple superphosphate (ST) Monoammonium phosphate (MAP) Diammonium phosphate (DAP) Araxa phosphate (ARAD)
K	Corn and Soybean	SNA and YE	Potassium oxide (KCL) Potassium sulfate (PS) Potassium and magnesium sulfate (PMS) Potassium double saltpeter (PDS)

N: Nitrogen; P: Phosphorus; K: Potassium; YEOM: Yield expectation considering the content of organic matter on soil; SNA: soil nutrient availability; YE: Yield expectation.

Lime recommendation is made based on the study of Moreira (2019). Thus, the following methods to calculate the required lime were selected: Method 1: neutralization of exchangeable aluminum; Method 2: exchangeable Al^{3+} neutralization and elevation of base cations (Ca^{2+} and Mg^{2+}); and Method 3: base saturation.

xiii. Definition of the case study

Data from a 20-ha agricultural field located in the Serranópolis do Iguaçu – PR city were used to demonstrate ADB-Map's functionalities, with central coordinates −54.01232307º (longitude) and −25.39526307º (latitude) in Datum WGS 1984 (Fig. 5). The sampling points with irregular distances were located along an imaginary line among the level curves following the terrain topography. The sample density of 2.6 ha^{-1} attends the suggest of a minimum density from 1 sample ha^{-1} (Ferguson and Hergert, 2009) to 2.5 sample ha^{-1} (Journel and Huijbregts, 1978; Doerge, 2000). The minimum and maximum distances between the sampling points are 45 and 706 m.

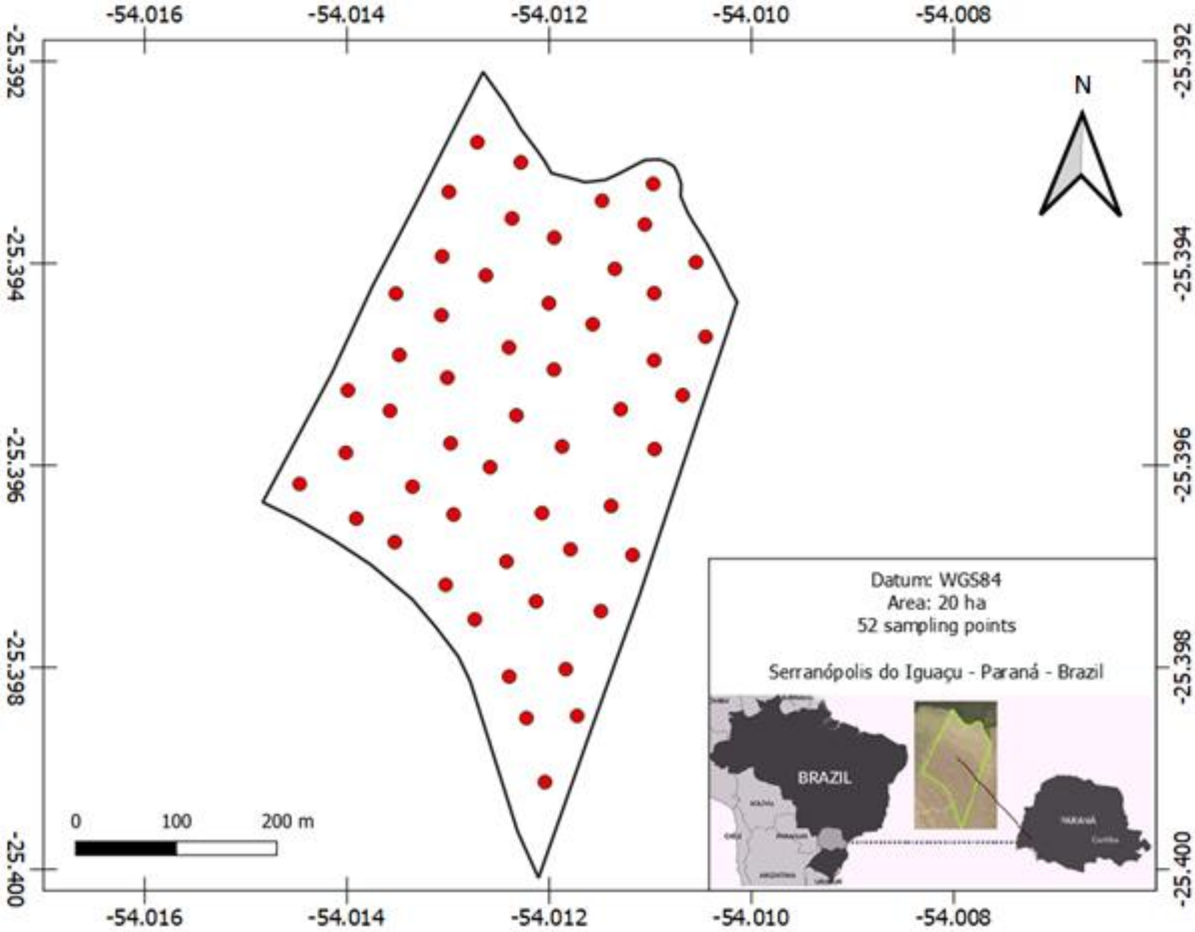


Fig. 5 Location of experiment and 52 sampling points in an experimental field in the municipality of Serranópolis do Iguaçu, Paraná state, Southern Brazil. Black contour delineates the 20 ha area used. Coordinates are in degrees (WGS 1984). The minimum and maximum distances between the sampling points are 45 and 706 m.

According to the workflow (Fig. 6) used in this case study, the following modules were used:

- Input data: The field data used were elevation, soil texture (sand, clay, and silt), soil resistance penetration from 0 to 20 m depth in 2018 (SPR_0-20), and soybeans yields in 2018/2019 (SY_18/19) and 2019/2020 (SY_19/20) crops. Soil samples were taken from 0 to 0.20 m depth and analyzed in a commercial laboratory. Around each sampling point (using a GNSS Juno SB Trimble Navigation Limited, Westminster, CO, USA) and using a 3-m radius, eight subsamples were randomly collected, two per quadrant, within a symmetrical circle divided into four quadrants.
- Statistics: descriptive statistics measures and data normality.
- Data preparation:
  - Cleaning: removal of outliers and inliers.
  - Grid cutting out: points that are outside of field boundaries are eliminated.

- Adjust grid: process that aims to interpolate data in missing points so that two grids have the same number of points and the same geographic coordinates. Data interpolation was done by IDW method.
- Normalization: All variables were normalized before further analysis. Yield values were normalized (Y_N) using the mean method (Equation A2 – Swindel, 1997) to stabilize these data, which are generally heavily influenced by variations in climate and precipitation. This method showed better performance with variables that vary from year to year. Then, since there is more than one year of yield data, the arithmetic mean of normalized yield was calculated, generating a single variable corresponding to the mean normalized yield ($\overline{Y_N}$).

  The remaining variables were used to delineate MZs using FCM with Euclidean distance. However, this method requires normalizing data since this distance is sensitive to the variables' amplitude (Bezdek 1981; Fridgen et al. 2004, Schenatto et al. 2017). The normalization used the range method (Equation A1 – Mielke and Berry, 2007), as Schenatto et al. (2017) suggested. They recommended it as the most suitable method to normalize data before the clustering using Euclidean distance similarity since this similarity measure is sensitive to the variables' variance. With this, the same amplitude was kept for data, regardless of the variable used.

- Variable selection: the variable selection to delineate MZs by SCM, PCA, and MULTISPATI-PCA methods.
- Interpolator selection: Selection of the best interpolator using IDW and OK methods. In geostatistics, the Matheron (1963) classic estimator was used to calculate semivariances with at least 30 pairs of points (Journel and Huijbregts, 1978), and the range (Ra) was limited to half of the maximum distance (MD) among points (cutoff = 0.5*MD). The lag size h was defined as the 44 meters, calculated from the number of lags (relation between cutoff and the shortest distance among pairs of points). The minimum and maximum amounts of 53 and 180 pairs were obtained with this lag distance to calculate semivariances. The significant limitation to address in this ADB-Map version is that anisotropy's eventual presence is not considered.
- Interpolation: Since selected variables showed spatial dependence, ordinary kriging interpolation was performed after normalizing variables. A 9x9 m (1/100 of the larger dimension in the Eastern/Western direction) grid was created to

increase the density of points to allow smoother delineation and more continuous MZs.

- MZs definition: MZs delineation using FCM method.
- MZs Rectification: MZs rectification with median, opening, closure, opening/closure methods, square kernel format, and 3x3, 5x5, and 7x7 kernel sizes applied as needed in each MZ for stain removal.
- MZs Evaluation: determination of statistics for MZs evaluation.
- Nutrient/lime recommendation: calculation of lime and nutrient recommendation.

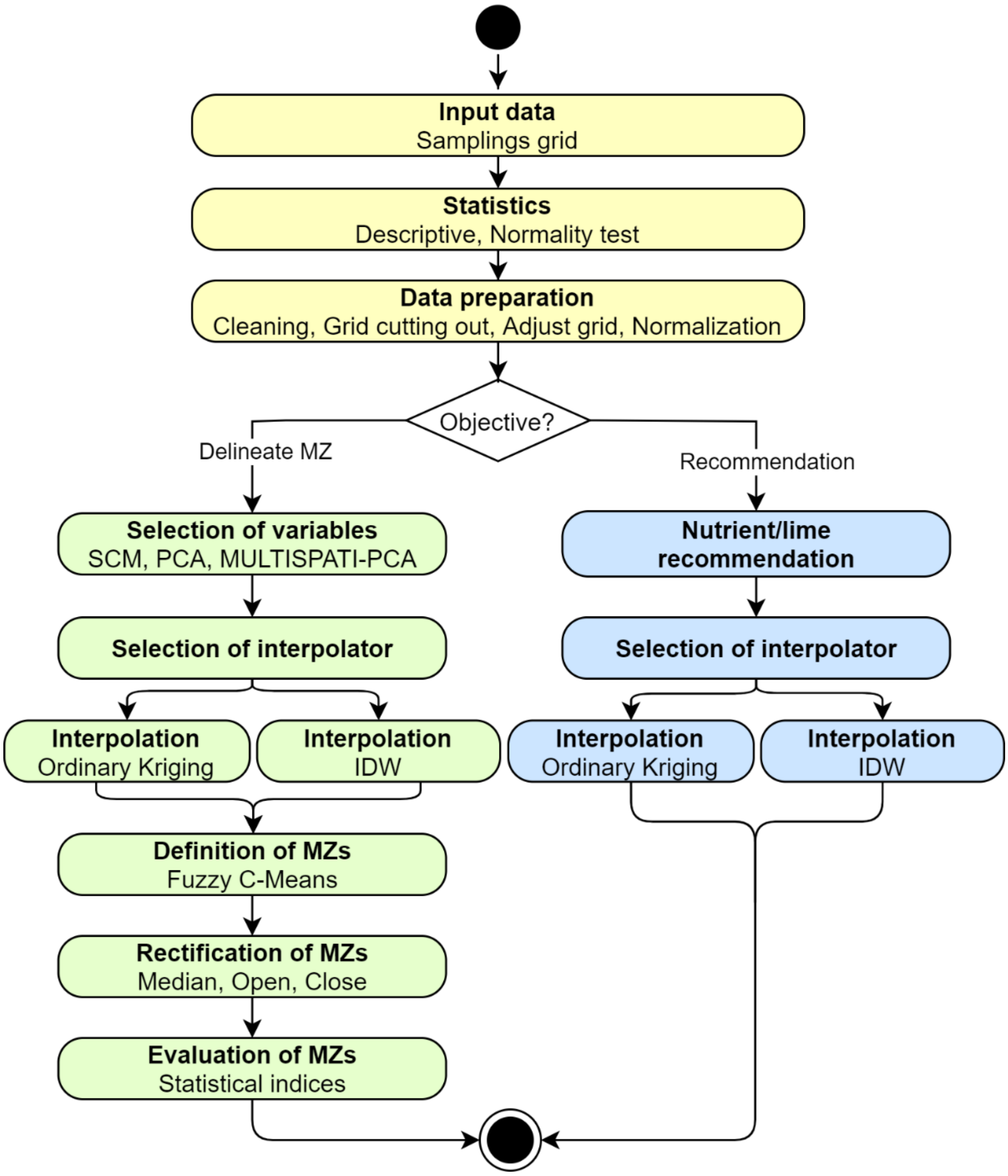


Fig. 6 Workflow to demonstrate the activities of (i) designing management zones and (ii) recommending nutrients and lime.

xiv. Software evaluation

The software evaluation captured the user's perception of ADB-Map, based on a software quality model. The product quality model was based on the ISO/IEC 9126 and 14598 standards, which define a general-purpose quality model, quality characteristics and provide examples of metrics. These standards were revised, expanded, and restructured, receiving the names ISO/IEC 25030 (ISO/IEC, 2019) and ISO/IEC 25040 (ISO/IEC, 2011), constituting part of the set of standards ISO/IEC 25000 (ISO/IEC, 2014).

The evaluation methodology (Fig. 7) consisted of the following activities:

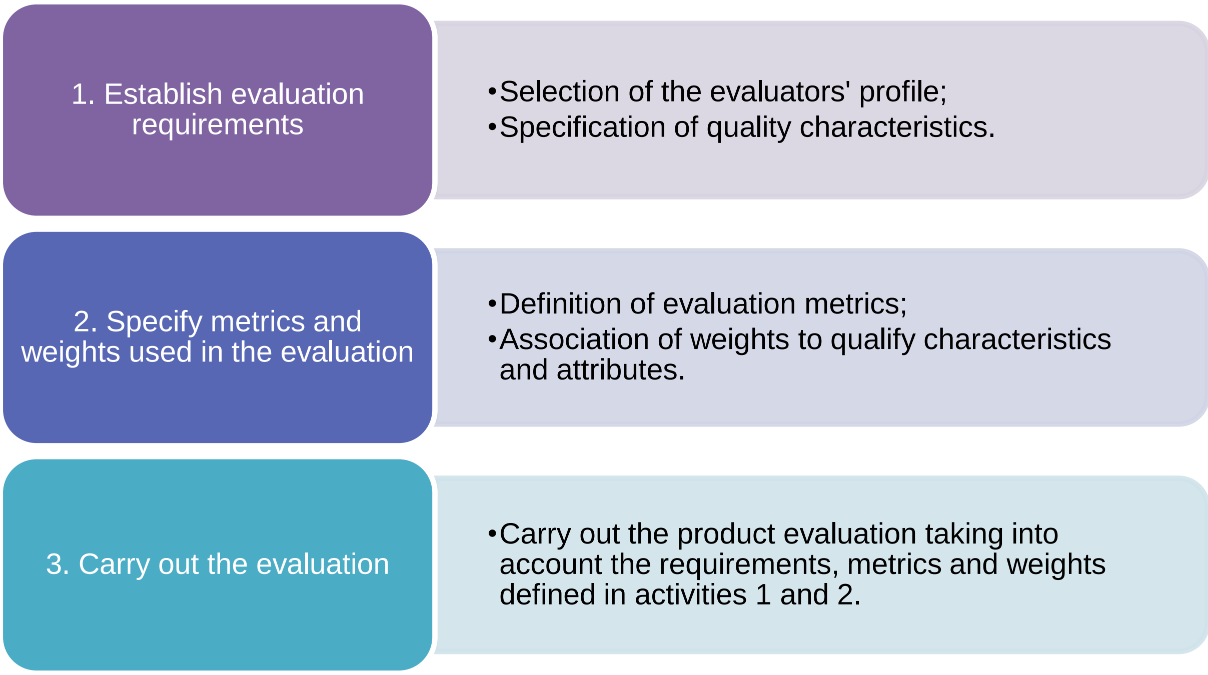


Fig. 7 Activity plan defined for the user's ADB-Map quality evaluation.

The first step was to establish the evaluation requirements and define the evaluators' profile and evaluation criteria. Next, the target audience of interest for the research was defined as users already registered in ADB-Map application who have already used several application features. In this case, they are undergraduate and graduate students, professors, and professionals in PA area. For the sample to represent the intended target audience, software evaluation must be carried out with at least eight evaluators (NBR ISO/IEC 14598-6; ABNT, 2004). Then, quality characteristics were evaluated using key-questions (Appendix B) related to the perception of the six characteristics specified in the adopted quality model (Table 4).

Table 4 Software quality characteristics of ISO/IEC 9126-1

Characteristics	Meaning	Key-question
Functionality	Shows the set of functions that meet explicit and implicit needs for the purpose for which the product is intended	Does it satisfy the needs?
Reliability	Shows product's capability to keep its performance over time and under established conditions	Is it immune to failure?
Usability	Shows how easy is to use the product	Is it easy to be used?
Efficiency	Shows the relationship among the product's performance level and the number of resources used under established conditions	Is it fast and 'lean'?
Maintainability	Shows the effort needed to make modifications to the product	Is it easy to modify?
Portability	Shows product's ability to be transferred from one environment to another	Is it easy to use in another environment?

Source: ABNT (2003).

In the second step of evaluation, scoring levels and judgment criteria were defined. The interviewers assigned scores from 1 to 5 for each question of the evaluation instrument, referring to a qualitative scale (Table 5). The interviewers answered the questions regarding agreement or satisfaction, depending on the context, with 1 being the lowest and 5 the highest. Each level has its respective quantitative score. Thus, it was possible to determine the quantitative score for each question, the average score for each characteristic of the quality model, and global score (average of all characteristics).

Table 5 Scoring levels for ADB-Map user's evaluation instrument

User rating note	Meaning	Quantitative score
5	Total	1
4	Great	0.75
3	Moderate	0.5
2	Lowest	0.25
1	Disagreement/Unsatisfied	0

The criteria to judge the obtained results were based on the evaluation scale for the characteristics proposed in NBR ISO/IEC 14598-6 standard – Annex C (ABNT, 2004). The acceptance criterion of ADB-Map by the community was adopted to be larger than 0.70 (70%) in each characteristic of the quality model evaluated by the user and globally. Data tabulation and characteristics calculation were performed using the Microsoft Excel® software.

In the third step, the evaluation plan was elaborated. The interviewers were invited to participate in the study using an invitation letter, sent by digital means, containing instructions to carry out the evaluation, together with a link to access ADB-Map and the evaluation instrument. As the interviewers were the application's users, they already had the credentials to access the software. The evaluation instrument was created using the online questionnaire tool Google Forms. The interviewers' responses were automatically obtained by this data collection tool. The data collection period was from September 2020 to November 2020. This period was necessary to obtain the minimum number of eight evaluators for each category, as recommended by NBR ISO/IEC 14598-6 standard.

After receiving the invitation, fourteen ADB-Map users, aged from 24 to 63, made the proposed evaluation. Most interviewers (86%) have master degree. The others have higher education (7%) and doctorate (7%). As for the relationship with AP, most interviewers (93%) are undergraduate or graduate students. The others (7%) are researchers in the area. Despite being invited, no professional from the PA area participated of the evaluation. As for the evaluator's contact time with the AP, most work between 1 and 2 years (50%), followed by 0 and 1 year (21%), 2 and 3 years (7%), 4 and 5 years (7 %), over 10 years (7%), and over 25 years (7%). Among other software (Fig. 8) used by interviewers to develop AP activities, R (57%), ArcGis/ArcMap (43%), QGis (43%), and Surfer (36%) stand out.

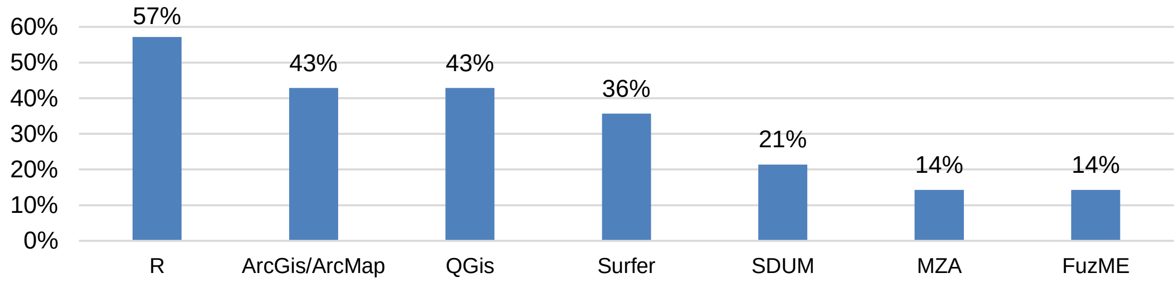


Fig. 8 Software already used by the evaluator to perform work with precision agriculture.

6.3 Results and discussion

i. Running AgDataBox-Map and importing data

Access to ADB-Map is restricted to registered and authenticated users. The user must accept ADB disclaimer. User management and authentication are done centrally at ADB-MSA and shared with other applications that make up ADB platform. The user has a particular work context, which is not visible to other users of application. ADB-Map's primary Graphical User Interface (GUI) was designed so that the user has three important access points: (1) list of projects and layers; (2) tools menu; (3) component to visualize spatial data (Fig. 9).

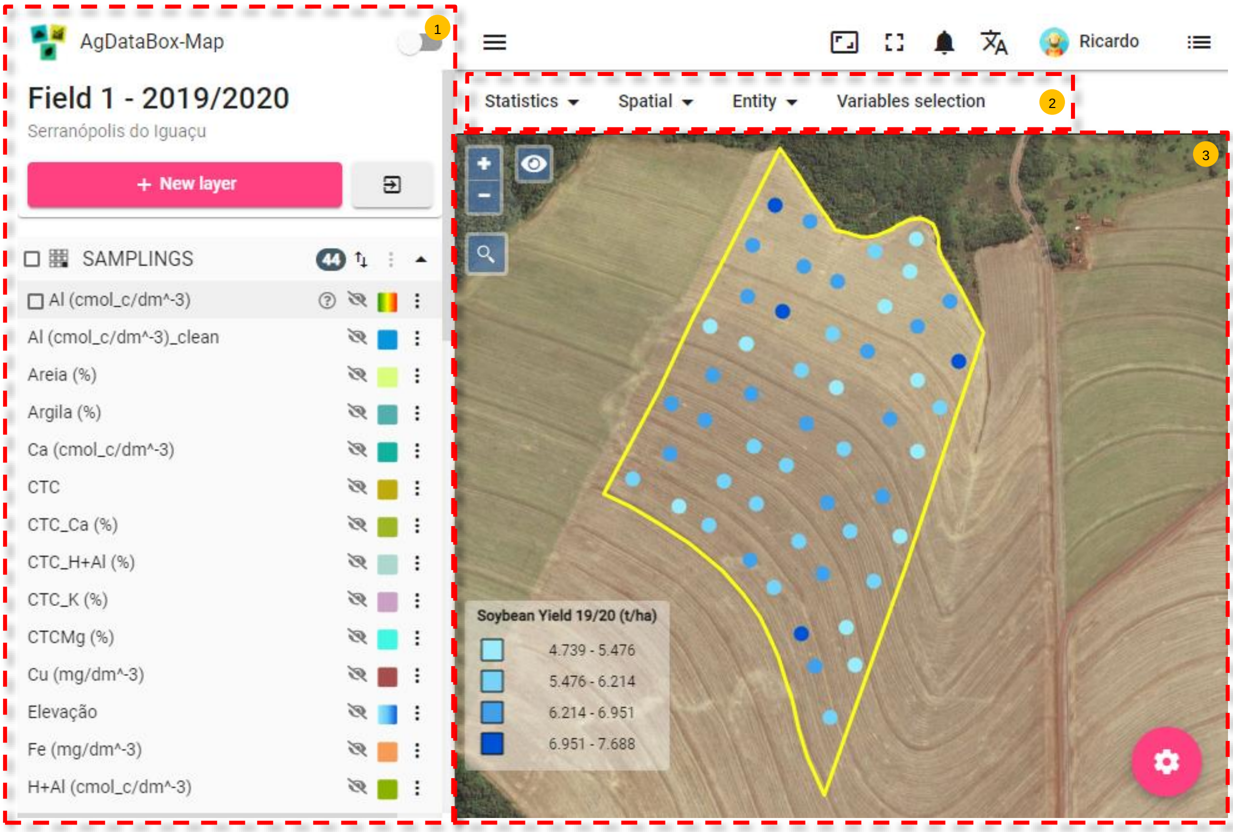


Fig. 9 Graphical User Interface of AgDataBox-Map application, presenting (1) the list of project layers, (2) menu, and (3) the spatial data visualization component.

It is necessary to create a project to use the application. In the workspace project, it is possible to create the following types of layers: Sampling grid, Boundaries, Data interpolation, Management Zone, Defining sampling grid, and Nutrient/Lime recommendation (Fig. 10).

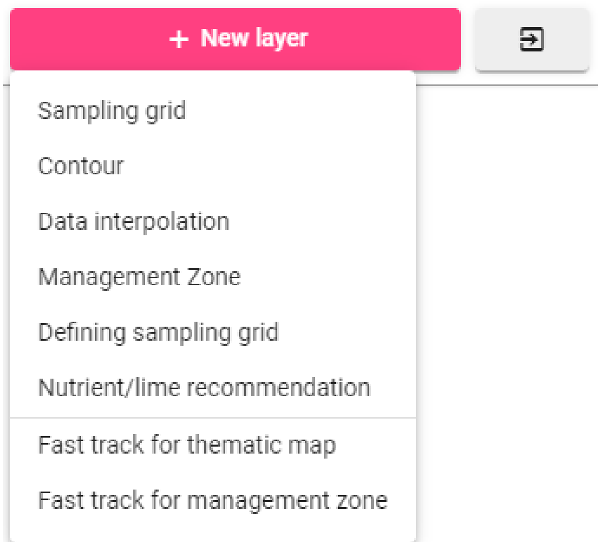


Fig. 10 Component screen to create a new layer showing the layers types available in ADB-Map application.

The features for creating TMs and defining MZs in ADB-Map depend on external data in the user's domain but need to be imported into ADB-Map's database. Some of the data sets required by ADB-Map are from a sample collection of attributes, thematic maps of attributes, field demarcations, satellite, and aerial images, among others, which can be stored in text files, spreadsheets, ESRI Shapefile, or other formats standard for data set representation.

The data import screen (Fig. 11) allows you to import multiple columns (variables) from the text file. For example, it is possible to specify the columns that represent X and Y geographic coordinates, in addition to the reference datum. Imported data is converted to the same project Datum. The file columns selected for import are stored in ADB-MSA.

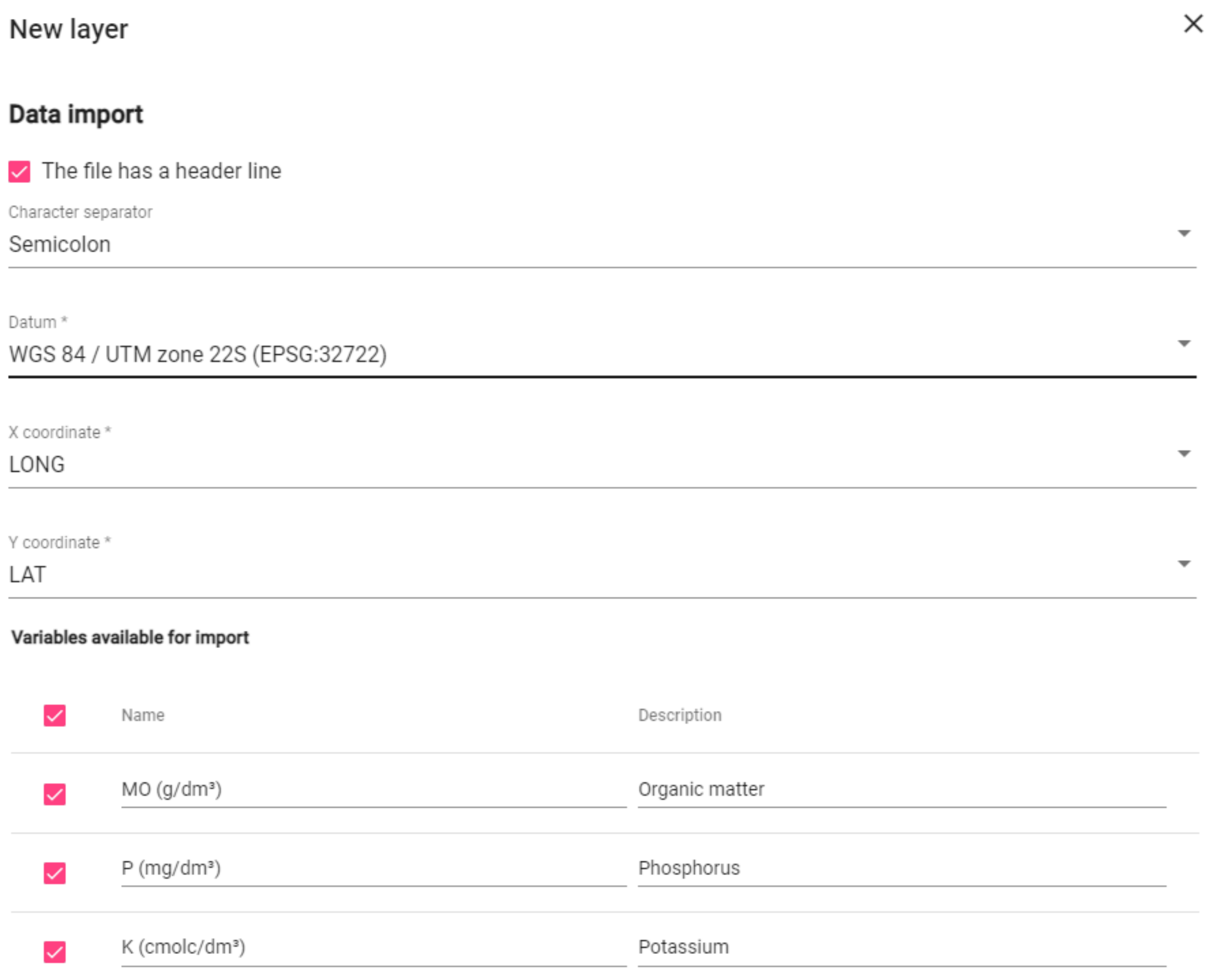
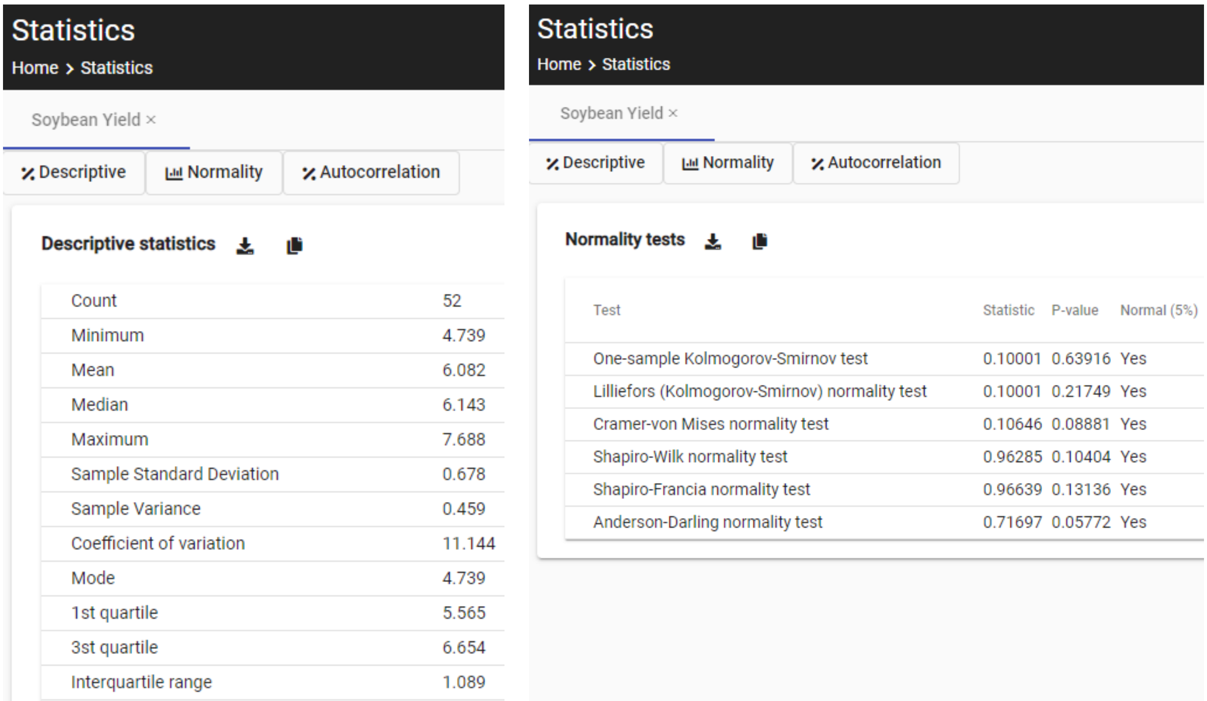


Fig. 11 Import layers interface to read a text file containing multiple variables.

ii. Statistical analysis and data preparation

ADB-Map's graphical statistical analysis interface allows performing descriptive statistical analysis of data (Fig. 12a) and testing normality (Fig. 12b).



(a) Screen of descriptive statistics analysis

(b) Screen of data normality tests

Fig. 12 Graphical user interfaces for (a) descriptive statistical analysis and (b) data normality tests.

There are data cleaning procedures to remove data with errors and atypical data from data set. This procedure may remove duplicate data (in the same geographic position), negative, null, or missing data, outliers, and inliers (Fig. 13).

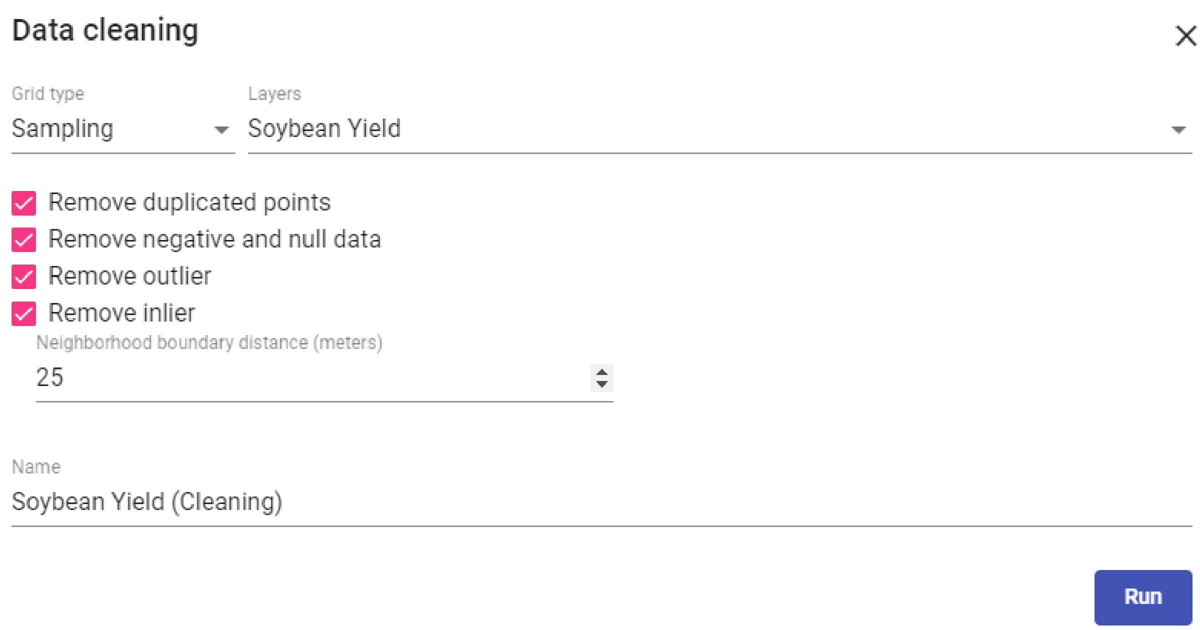


Fig. 13 Screen for data cleaning in data preparation process: remove duplicate, negative and null data, outlier, and inlier.

There are work situations that need grid cutting out. This procedure aims to eliminate points that are outside field boundaries. One of the most common cases is removing points obtained by the harvest monitor outside the field's limits due to bedside maneuvers. This situation can be solved with the functionality called “Grid cutting out” (Fig. 14), which aims to cut a grid eliminating the points that are outside the field boundaries.

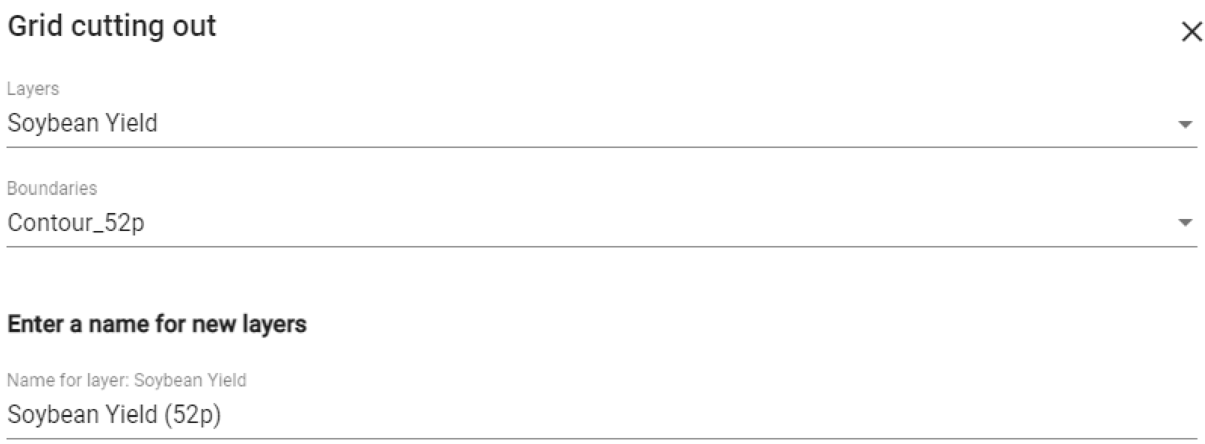
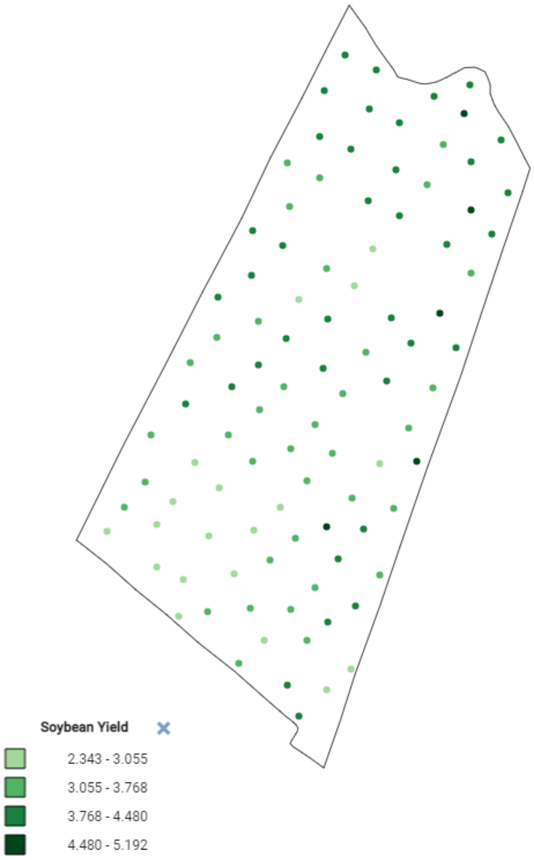
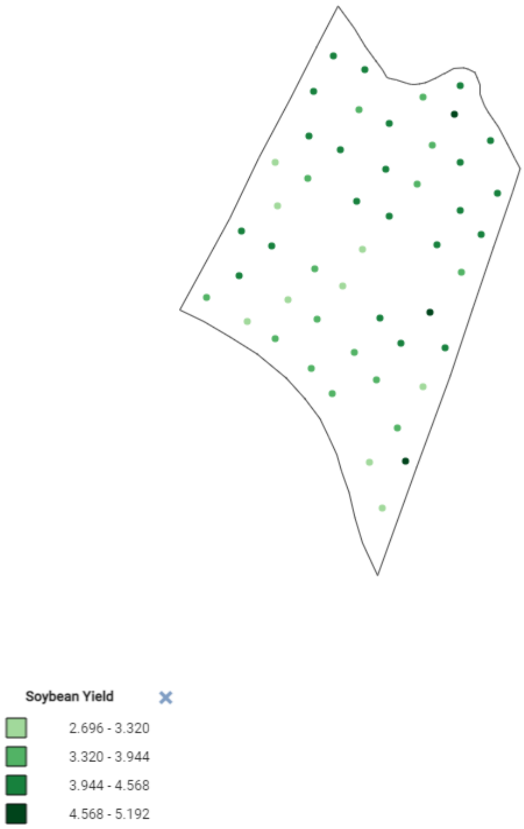


Fig. 14 Screen for grid cutting out during data preparation process to eliminate map pixels that are outside the field boundaries.

A situation applied in the experimental field was to reduce the work field, which, until the 2018/2019 soybean crop (SY_18/19), was managed with 36.6 ha^{-1} (Fig. 15a) and, after that, was resized to 20.0 ha^{-1} (Fig. 15b). From 100 sampling points, the field now has 52. For studies that depend on the crossing of data collected at these different times, there is a need to cut data grids, adjusting all grids from the 2018/2019 crop to the new format.



(a) Grid with 100 sampling points



(b) Grid with 52 sampling points

Fig. 15 Soybean yield sample data (t ha^{-1}) from the 2018/2019 crop (SY_18/19) with the number of (a) original points after data cleaning and (b) applied to the grid cutting out procedure.

The descriptive statistics of data (Table 6) show that there was an increase of 58% in soybean yield from 2018/2019 to 2019/2020 crops. The coefficient of variation (CV) for yield in both years was considered medium (Pimentel-Gomes, 2009).

Table 6 Descriptive data statistics

Variables	Samples	Minimum	Mean	Median	Maximum	SD	CV%
Elevation (m)	52	316	334	335	348	9.5	3 (L)
Clay (%)	52	61.0	73.8	74.0	84.0	3.9	5 (L)
Sand* (%)	52	0.7	2.7	2.6	7.0	1.1	43 (VH)
Silt* (%)	52	14.3	23.6	23.3	36.0	3.8	16 (M)
SPR_0-20* (kPa)	52	867	1478	1470	2440	355	24 (H)
SY_18/19* (t ha^{-1})	49	2.7	3.9	4.0	5.2	0.5	14 (M)
SY_19/20* (t ha^{-1})	52	3.9	5.0	5.0	6.3	0.6	11 (M)

SD: Standard deviation; CV: Coefficient of variation: low (L) when CV ≤ 10%, medium (M) when 10% < CV ≤ 20%, high (H) when 20% < CV ≤ 30%, and very high (VH) when CV > 30%.
* No normality at 5% significance level (Pimentel-Gomes, 2009).
SY_18/19: soybean yield in 2018/2019; SY_19/20: soybean yield in 2019/2020; SPR_0-20: Soil penetration resistance at a depth of 0–20 cm.

Experimental variables were normalized according to two methods: (i) the yield values with the mean method (Equation A2), and (ii) the remaining variables with the range method

(Equation A1), as suggested by Schenatto et al. (2017). The mean normalized soybean yield (mnSY) was obtained by the arithmetic average of nSY_18/19 and nSY_19/20. This operation was performed by the “Grid math” procedure (Fig. 16), which allows performing mathematical operations among the layers of the project.

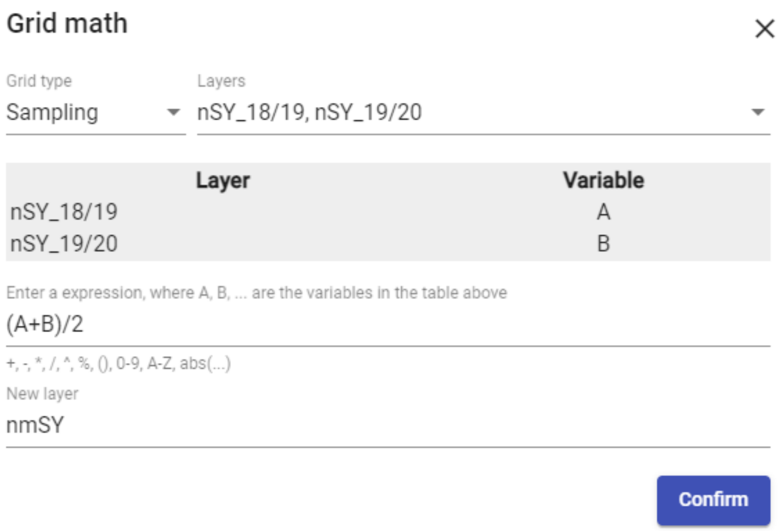
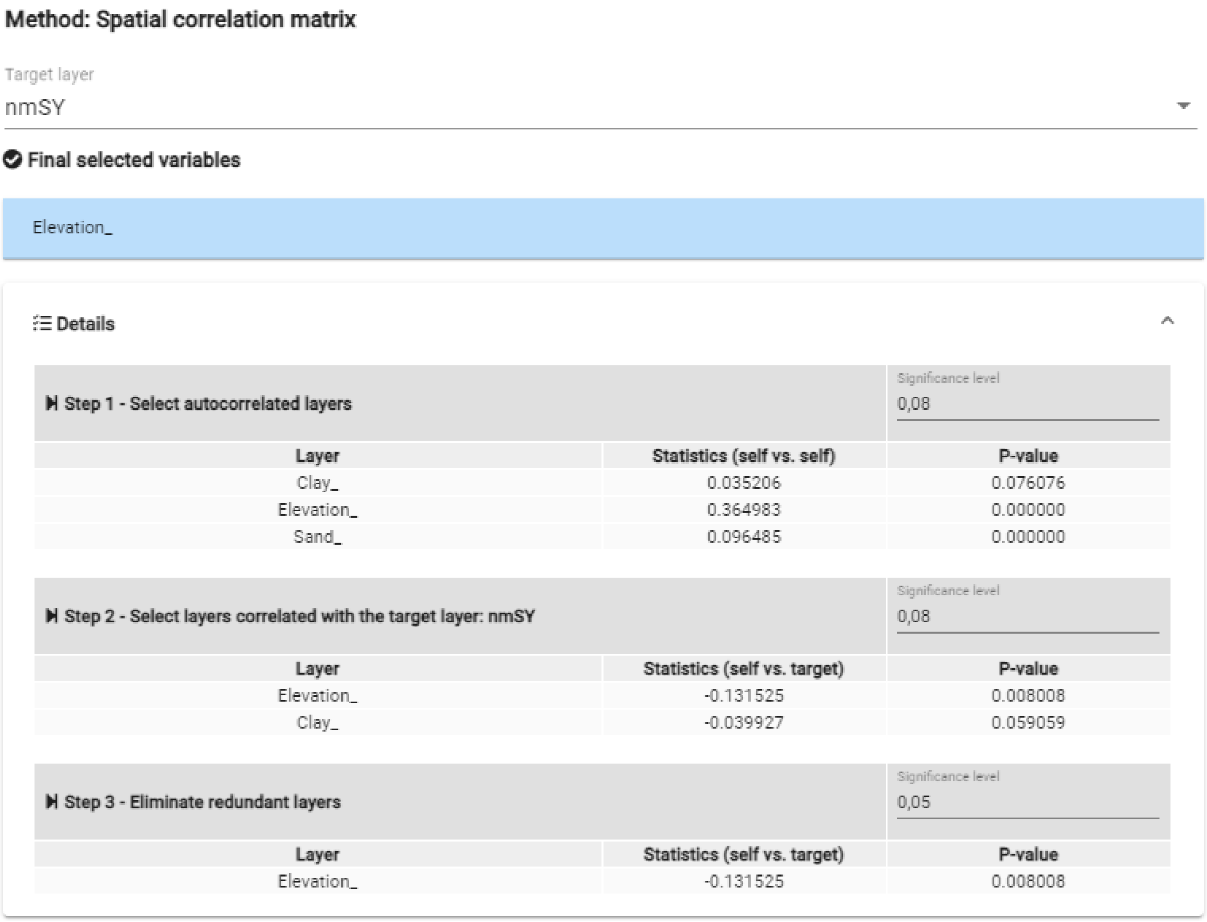


Fig. 16 Screen for grid math where a grid is generated by applying a mathematical expression over multiple existing grids.

iii. Variable Selection

The graphical interface implemented in ADB-Map allows selecting variables to delineate MZs. For each method, the result is presented on a different screen:

(a) Spatial correlation matrix method (SCM; Bazzi et al., 2013; Fig. 17): the process was performed with variables mnSY (mean of nSY_18/19 and nSY_19/20), elevation, clay, sand, and silt. The variables selected in each step were: (i) spatially dependent: elevation, clay, and sand; (ii) with spatial correlation with the target variable: clay and elevation; and (iii) after eliminating redundant variables: elevation.



Method: Spatial correlation matrix

Target layer

nmSY

Final selected variables

Elevation_

Details

Step 1 - Select autocorrelated layers — Significance level 0,08

Layer	Statistics (self vs. self)	P-value
Clay_	0.035206	0.076076
Elevation_	0.364983	0.000000
Sand_	0.096485	0.000000

Step 2 - Select layers correlated with the target layer: nmSY — Significance level 0,08

Layer	Statistics (self vs. target)	P-value
Elevation_	-0.131525	0.008008
Clay_	-0.039927	0.059059

Step 3 - Eliminate redundant layers — Significance level 0,05

Layer	Statistics (self vs. target)	P-value
Elevation_	-0.131525	0.008008

Fig. 17 AgDataBox-Map screen showing the result of variable selection by the spatial correlation matrix method.

Autocorrelation (spatial dependence) and cross-correlation (bivariate) between variables can be observed in the spatial correlation matrix (SCM) (Fig. 18), provided by ADB-Map application. However, the values of Bivariate Moran's I indices are low because they are not normalized.

	Sand	**Clay**	**Elevation**	**SPR_0-20**	**Silt**	**nmSY**
Sand	0.09648 **pv: 0**					
Clay	-0.04113 pv: 0.1	0.03521 pv: 0.08				
Elevation	0.06314 pv: 0.01	0.0817 pv: 0	0.36498 **pv: 0**			
SPR_0-20	-0.00717 pv: 0.78	0.02569 pv: 0.2	0.05548 pv: 0.33	-0.00189 pv: 0.92		
Silt	0.00929 pv: 0.73	-0.02412 pv: 0.23	-0.12423 pv: 0.02	-0.02738 pv: 0.12	0.02259 pv: 0.27	
nmSY	-0.02387 pv: 0.27	-0.03993 **pv: 0.06**	-0.13152 **pv: 0.01**	-0.02831 pv: 0.17	0.0548 pv: 0.01	0.02480 pv: 0.24

pv: p-value (significative when p-value < 0.05).

Fig. 18 Spatial correlation matrix, in which the cells (i) in blue indicate the significantly auto-correlated variables, (ii) in green the variables correlated with the average soybean yield, and (iii) in salmon the redundant variables, which should be eliminated.

(b) Principal component analysis (PCA; Fig. 19): After PCA's execution, the variable selection is made after applying one of the defined selection criteria. Using the selection criteria of Johnson and Wichern (2007), the principal components (PCs) that together accumulated at least 80% of variance of data set were selected, which in this case were PCs from 1 to 3, which accumulated 86% of variance.

Method: Principal Component Analysis

Selection criteria

Johnson and Wichern (2007) - Accumulated variance > 80%

	Variable	Eigenvalue	Proportion (%)	Accumulated (%)
☑	PC1	2.111	42.21	42.21
☑	PC2	1.181	23.61	65.83
☑	PC3	0.997	19.95	85.77
☐	PC4	0.696	13.92	99.69
☐	PC5	0.015	0.31	100

Fig. 19 Screen showing the variable selection by the principal component analysis method.

(c) MULTISPATI-PCA (Fig. 20): the variable selection by MULTISPATI-PCA method is similar to PCA analysis has selected the three spatial PCs by the selection criterion of Johnson and Wichern (2007).

Method: Multivariate spatial analysis based on Moran's index and PCA

Selection criteria

Johnson and Wichern (2007) - Accumulated variance > 80%

	Variable	Eigenvalue	Proportion (%)	Accumulated (%)
☑	PC1	0.836	36.05	36.05
☑	PC2	0.244	37.27	73.32
☑	PC3	0.005	26.68	100

Fig. 20 Screen that presents an example of variable selection results by the MULTISPATI-PCA method.

iv. Data interpolation

ADB-Map's graphical interfaces for data interpolation allow using the interpolation methods IDW, OK, MA, and NN. The interface to select the interpolator and determine its parameters was implemented (Fig. 21) to determine the best method of interpolation between OK (if there is spatial dependence in the data), or IDW (otherwise), which uses ECI and ISI indices.

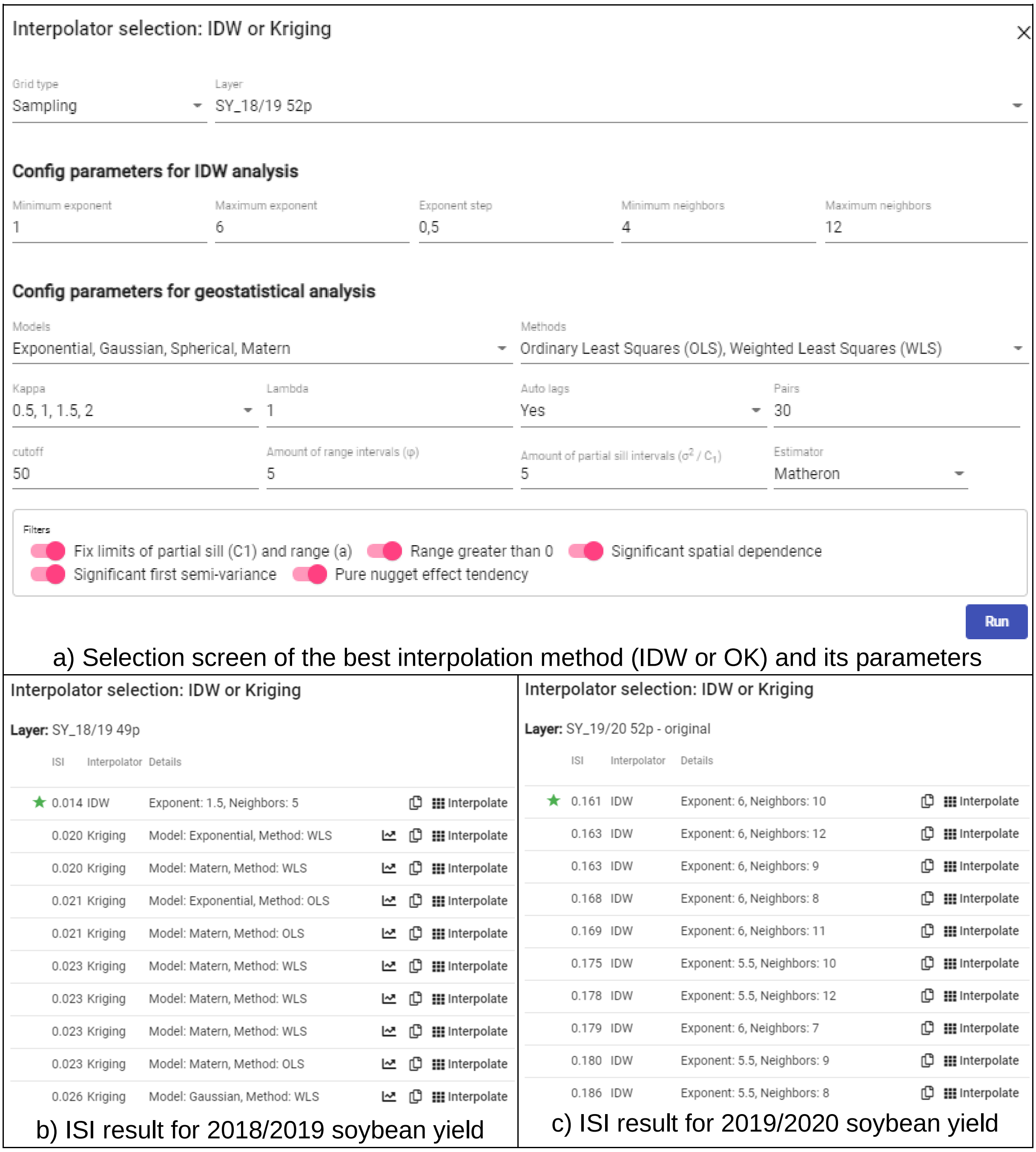
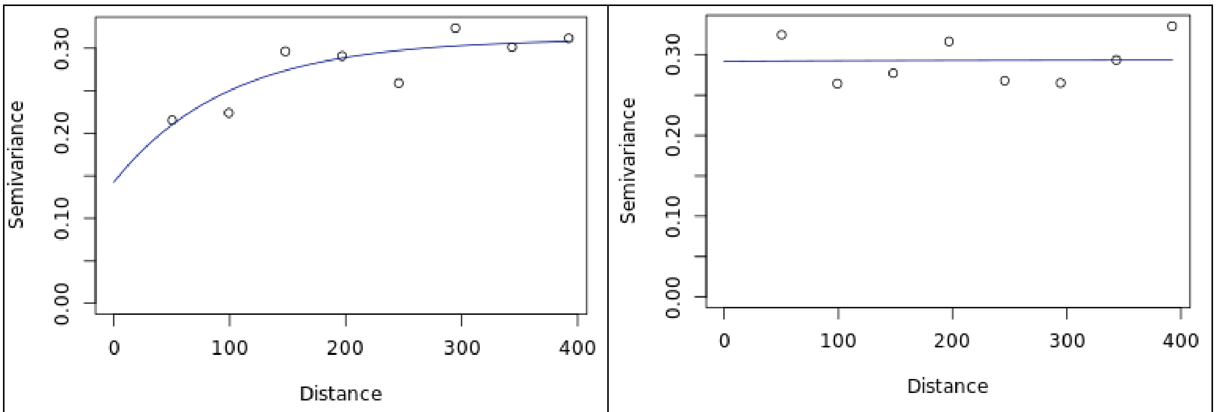


Fig. 21. Graphical interfaces (a) to select the best interpolation method, Inverse Distance Weighting (IDW) or Ordinary Kriging (OK); Ranking of the ten best results for soybean yield during the (b) 2018/2019 and (c) 2019/2020 crops.

The best interpolation method selection between IDW and OK is performed using a graphical interface (Fig. 21a) in which the parameters for IDW and OK interpolators are defined. For IDW, exponents' range and the neighbors' range to be analyzed are defined. For OK, the semivariogram models, adjustment methods, Lambda, number of lags, pairs, cutoff, amount of partial sill intervals, and range intervals are defined. The ranking of the best models (Fig. 21b and Fig. 21c) is done in descending order of the best performance. Ten of them are presented, but the list is long and varies according to the number of parameters selected in the analysis. From this list, someone can select the method/model to interpolate. According to the harvests, the interpolator selection result changed (SY_18/19, Fig. 21b; SY_19/20; Fig. 21c). IDW was the best interpolator for the two years of soybean yield, but it varied the exponent value and the optimal number of neighbors. The second to the tenth position of ISI in variable SY_18/19 indicates the OK interpolator, but in variable SY_19/20, they indicate IDW. All geostatistical models were eliminated during the interpolator selection for SY_19/20, as this variable has a pure nugget effect.

It was possible to obtain the semivariogram graph with the adjusted theoretical model for each resulting geostatistical model presented on an ADB-Map screen. Considering only the result of OK interpolator, we generate the semivariogram for variable SY_18/19 (Fig. 22a), adjusted with the exponential model (Nugget effect = 0.14, partial sill = 0.17, and range = 98 m). The semivariogram of variable SY_19/20 (Fig. 22b) was generated considering pure nugget effect (Nugget effect = 0.29, partial sill = 0 and range = 0).



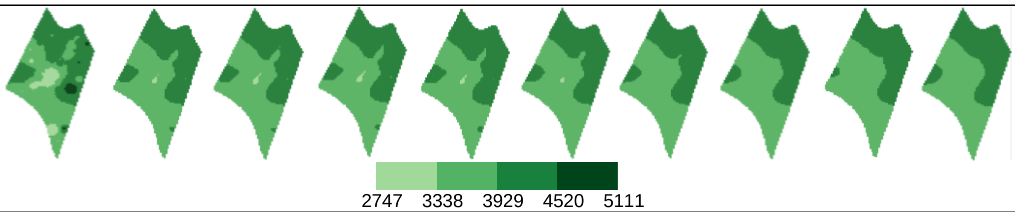
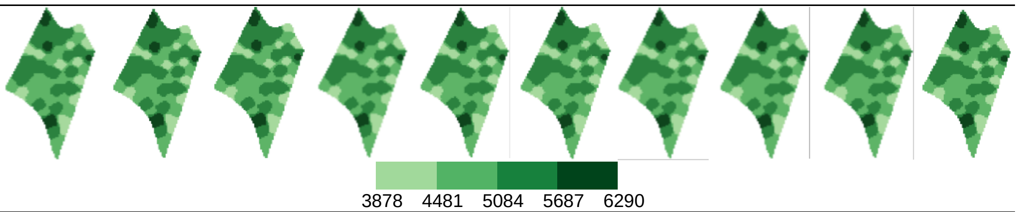
(a) SY_18/19 semivariogram adjusted by the exponential model

(b) SY_19/20 semivariogram adjusted by pure nugget effect

Fig. 22 Semivariogram of soybean yield variables in crops (a) 2018/2019 and (b) 2019/2020 considering the best geostatistical model selected in the interpolator selection process.

SY_18/19 and SY_19/20 sample grids were interpolated according to the parameters obtained during the interpolator selection (Fig. 22), from the first to the tenth position of ISI (Table 7). The maps were divided into four classes by equal distances, considering the minimum and maximum values among the ten maps. The maps differed from the first map (considered as reference) from 3.20% to 4.24% (by CRD) in SY_18/19 variable, from 0.0% to 0.15% (by CRD) in SY_19/20 variable. This variation represents on average 119 to 158 (kg ha^{-1}), in SY_18/19 variable, and 0.07 to 7.38 (kg ha^{-1}), in SY_19/20 variable.

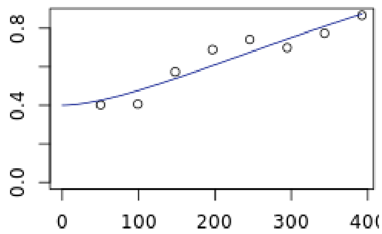
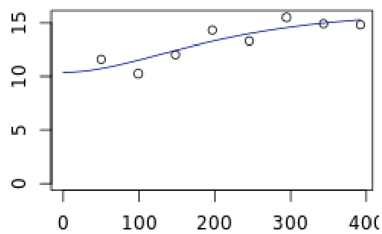
Table 7 Soybean yield sampling grids interpolated according to interpolator selection and ISI ranking from first to the tenth position

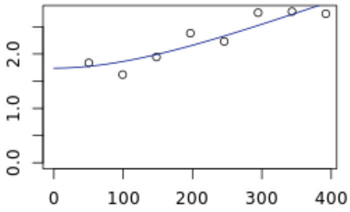
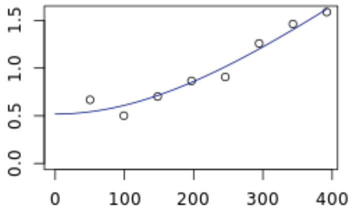
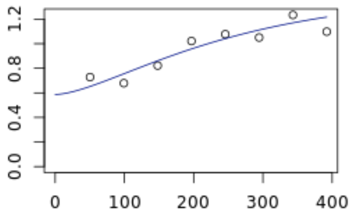
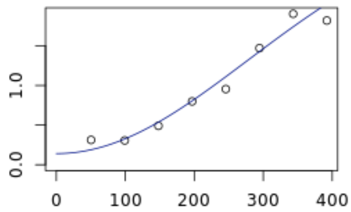
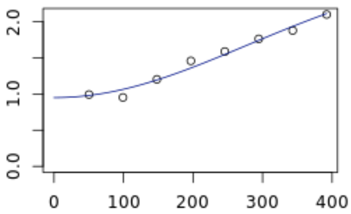
Variables	Interpolator selection index (ISI) ranking									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
SY_18/19										
CRD% MAD (kg ha^{-1})	Ref.	3.26% 121	3.26% 121	3.20% 119	3.20% 119	3.48% 129	3.65% 136	3.75% 139	4.24% 158	3.94% 147
SY_19/20										
CRD% MAD (kg ha^{-1})	Ref.	0.002% 0.101	0.002% 0.0998	0.004% 0.186	0.001% 0.0674	0.15% 7.30	0.15% 7.38	0.006% 0.318	0.15% 7.23	0.15% 7.17

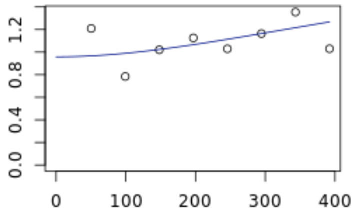
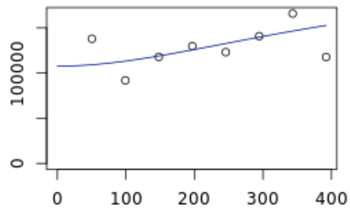
Ref.: Reference; ISI: Interpolator selection index; SY_18/19: Soybean yield in 2018/2019 crop; SY_19/20: Soybean yield in the 2019/2020 crop.

IDW was considered the best interpolator for the mean of SY_18/19 and SY_19/20 (mSY), elevation, and silt variables. On the other hand, OK was considered the best for clay, sand, PCA 1 to 3, MPCA 1 to 3, and SPR_0-20 variables (Table 8).

Table 8 Interpolator selection analysis and thematic maps of the studied variables

Varia-bles	Selected interpolator	Interpolator parameters	Semivariogram (Interpolation by Ordinary kriging)	Thematic Map
Mean soybean yield (mSY)	IDW	Exponent: 2 Neighbors: 9		
Elevation	IDW	Exponent: 6 Neighbors: 4		
Sand	Ordinary Kriging	Model: Matérn 1 Method: OLS Nugget Effect: 0.400 Partial Sill: 1.192 Range: 392 m		
Clay	Ordinary Kriging	Model: Matérn 2 Method: OLS Nugget Effect: 10.4 Partial Sill: 5.61 Range: 94 m		
Silt	IDW	Exponent: 1 Neighbors: 9		

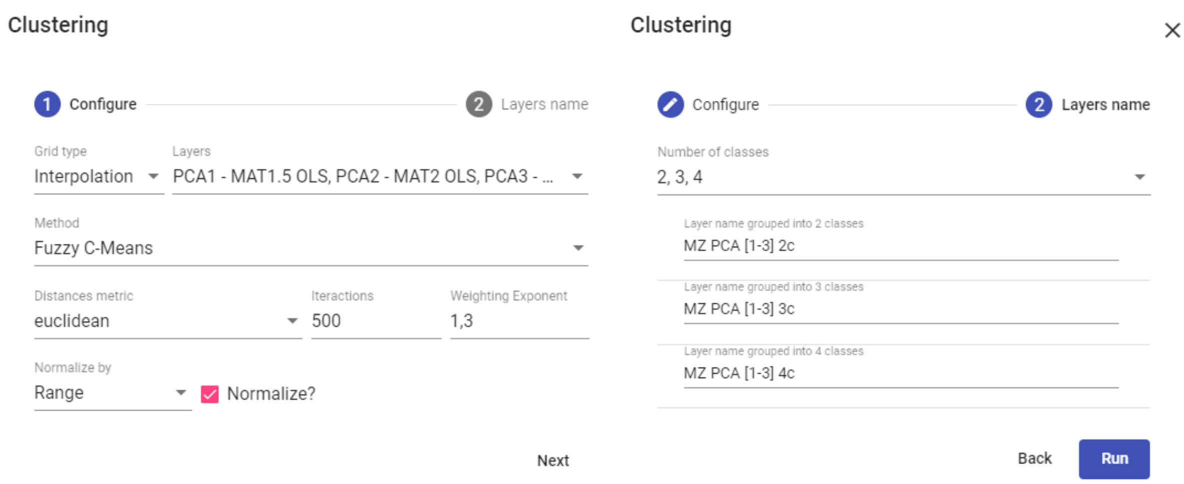
Varia-bles	Selected interpolator	Interpolator parameters	Semivariogram (Interpolation by Ordinary kriging)	Thematic Map
PC1	Ordinary Kriging	Model: Matérn 1.5 Method: OLS Nugget Effect: 1.74 Partial Sill: 4.53 Range: 392 m		
PC2	Ordinary Kriging	Model: Matérn 2 Method: OLS Nugget Effect: 0.52 Partial Sill: 5.87 Range: 392 m		
PC3	Ordinary Kriging	Model: Matérn 1 Method: WLS Nugget Effect: 0.59 Partial Sill: 0.81 Range: 172 m		
MPCA1	Ordinary Kriging	Model: Gaussian Method: WLS Nugget Effect: 0.141 Partial Sill: 2.99 Range: 392 m		
MPCA2	Ordinary Kriging	Model: Gaussian Method: OLS Nugget Effect: 0.950 Partial Sill: 1.84 Range: 392 m		

Varia-bles	Selected interpolator	Interpolator parameters	Semivariogram (Interpolation by Ordinary kriging)	Thematic Map
MPCA3	Ordinary Kriging	Model: Matérn 2 Method: WLS Nugget Effect: 0.958 Partial Sill: 0.826 Range: 245 m		
SPR_0-20	Ordinary Kriging	Model: Matérn 2 Method: WLS Nugget Effect: 107717 Partial Sill: 83622 Range: 181 m		

IDW: Inverse distance weighting; OLS: Ordinary least squares; WLS: Weighted least squares.

v. Clustering data

In the graphical interface for data clustering (Fig. 23), variables (layers), the clustering method, its parameters, and the number of classes are chosen to carry out the clustering. A new grid is generated in ADB-Map for each number of chosen classes.



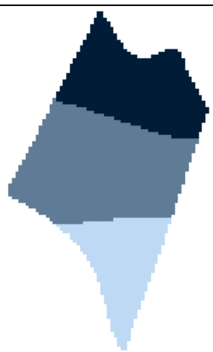
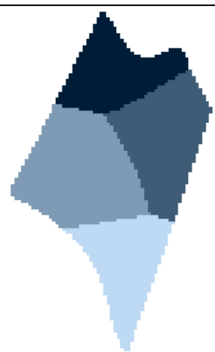
a) Configure analysis b) Number of classes and layers name

Fig. 23 Graphical user interfaces for data clustering: configure analysis (a) and definition of class numbers and the name of the new layers.

The grids interpolated from the variables selected in the study were clustered into 2, 3, and 4 classes To demonstrate the data clustering process by Fuzzy C-means (FCM; Table 9) method, as it follows:

- SCM_EL: Elevation variable clustering, selected by SCM method;
- SCM_EL-MSY: Clustering of the elevation variable, selected by SCM method, with the addition of the nmSY variable;
- PCA_1-3: principal components (1 to 3) selected by Johnson and Wichern (2007) and Ferreira (1996) criteria;
- PCA_1-2: principal components (1 to 2) selected by Kaiser (1960) criterion;
- MPCA_1-3: spatial principal components (1 to 3) selected by Johnson and Wichern (2007) criterion.

Table 9 Management zones delineated with the variables selected by SCM, PCA, and MULTISPATI-PCA selection methods, clustered into 2, 3, and 4 classes

Selection method	2 classes	3 classes	4 classes
SCM _EL			
SCM_EL-MSY			
PCA_1-3			
PCA_1-2			
MPCA_1-3			

Kappa index showed that the agreement among the designs divided into two classes was, for the most part, very strong (Table 10). Moderate agreement was obtained to compare PCA_1-3 x PCA_1-2. In the division into three classes, comparisons among MZs varied from no agreement to very strong agreement, which was obtained only in PCA_1-2 x MPCA_1-3 and SCM_EL x SCM_EL-MSY. In the division into four classes, the comparisons varied from no agreement to strong agreement, and in most cases, they were weak or no agreement.

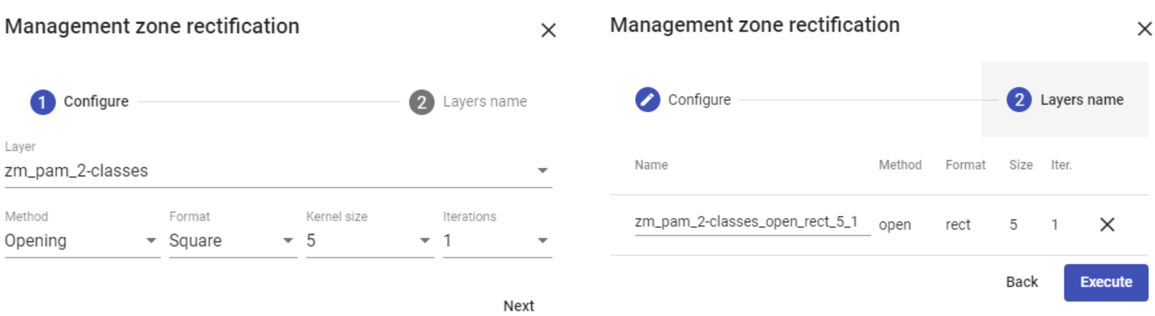
Table 10 Agreement between management zones (MZs) measured by Kappa coefficient (Kp) and Global accuracy (GA)

NC	VSM	Variable selection method			
		SCM_EL	SCM_EL-MSY	PCA_1-3	PCA_1-2
2	MPCA_1-3	0.96 (VS) – 0.98	0.82 (VS) – 0.91	0.89 (VS) – 0.94	0.94 (VS) – 0.97
	PCA_1-2	0.93 (VS) – 0.97	0.88 (VS) – 0.94	0.46 (M) – 0.64	
	PCA_1-3	0.91 (VS) – 0.96	0.72 (S) – 0.86		
	SCM_EL-MSY	0.81 (VS) – 0.91			
3	MPCA_1-3	0.21 (W) – 0.47	0.27 (W) – 0.52	0.63 (S) – 0.76	0.80 (VS) – 0.87
	PCA_1-2	0.27 (W) – 0.51	0.34 (W) – 0.57	0.46 (M) – 0.67	
	PCA_1-3	0.04 (N) – 0.36	0.04 (N) – 0.36		
	SCM_EL-MSY	0.87 (VS) – 0.91			
4	MPCA_1-3	0.49 (M) – 0.62	0.32 (W) – 0.50	0.26 (W) – 0.45	0.19 (N) – 0.45
	PCA_1-2	0.22 (W) – 0.41	-0.09 (N) – 0.17	0.78 (S) – 0.84	
	PCA_1-3	0.23 (W) – 0.42	-0.11 (N) – 0.17		
	SCM_EL-MSY	0.04 (N) – 0.29			

NC: number of classes; VSM: Variable selection method; Kappa agreement: N = no agreement ($Kp \leq 0.2$); W = weak ($0.2 < Kp \leq 0.4$); M = moderate ($0.4 < Kp \leq 0.6$); S = strong ($0.6 < Kp \leq 0.8$); VS = very strong ($0.8 < Kp \leq 1$).

vi. Rectification of management zones

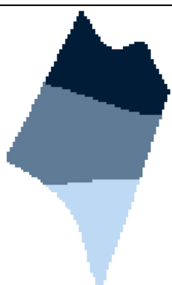
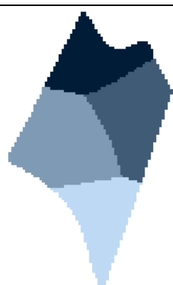
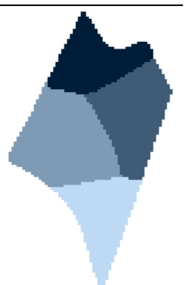
MZs rectification (Betzek et al., 2018) eliminates spots or isolated pixels in data clustering process. In ADB-Map rectification module, there is a graphical interface configuration screen (Fig. 24), where someone chooses the rectification method, the brush (kernel) format and size, and the number of interactions over MZ. The small spots observed (three and four classes) were removed with the rectification process (Table 11).



a) Screen to configure the process b) Screen to defining new layers' name

Fig. 24 Graphical user interfaces for management zone rectification: (a) selecting the layers to be rectified and the methods to be used (b) and defining the name of the rectified layers.

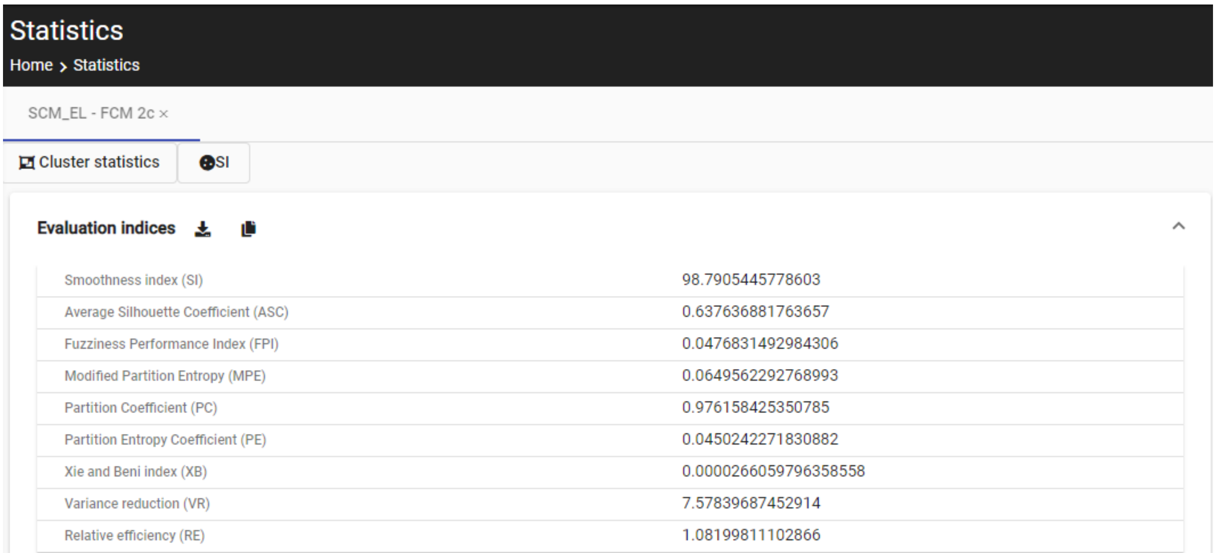
Table 11 Management zones, original and rectified, delineated with the variables selected by SCM, PCA, and MULTISPATI-PCA selection methods, clustered into 2, 3, and 4 classes

	2 classes		**3 classes**		**4 classes**	
	MZo	**MZr**	**MZo**	**MZr**	**MZo**	**MZr**
SCM_EL						
Kp/GA	0.99 – 1.00		0.99 – 0.99		0.97 – 0.97	
SCM_EL-MSY						
Kp/GA	1.00 – 1.00		0.96 – 0.97		0.96 – 0.97	
PCA_1-3						
Kp/GA	1.00 – 1.00		1.00 – 1.00		1.00 – 1.00	
PCA_1-2						
Kp/GA	1.00 – 1.00		1.00 – 1.00		1.00 – 1.00	
MPCA_1-3						
Kp/GA	1.00 – 1.00		1.00 – 1.00		1.00 – 1.00	

Kp: Kappa; GA: Global Agreement; MZo: Original management zone: MZr: Rectified management zone.

vii. Evaluation of the management zones

The delineated MZs were evaluated in the “Statistics” graphic interface by calculating the quality indices of each cluster (Fig. 25a), descriptive statistical analysis of the target variable (Fig. 25b), and the calculation of agreement indices among the maps (Fig. 26). The overall cluster quality indices (Fig. 25a) are (i) the smoothness index (SI), (ii) average silhouette coefficient (ASC), (iii) fuzzy performance index (FPI), (iv) modified entropy partition (MPE), (v) partition coefficient (PC), (vi) partition entropy coefficient (PE), (vii) Xie and Beni index (XB), (viii) variance reduction (VR%), and (iv) relative efficiency (RE). Statistical measures of position (mean, median, and quartiles), dispersion (standard deviation, variance, and coefficient of variation), skewness, kurtosis, and Tukey's mean difference test are obtained by clustering class (Fig. 25b).


Statistics
Home > Statistics
SCM_EL - FCM 2c ×
Cluster statistics
SI
Evaluation indices

Smoothness index (SI)	98.7905445778603
Average Silhouette Coefficient (ASC)	0.637636881763657
Fuzziness Performance Index (FPI)	0.0476831492984306
Modified Partition Entropy (MPE)	0.0649562292768993
Partition Coefficient (PC)	0.976158425350785
Partition Entropy Coefficient (PE)	0.0450242271830882
Xie and Beni index (XB)	0.0000266059796358558
Variance reduction (VR)	7.57839687452914
Relative efficiency (RE)	1.08199811102866

a) Quality indices of each grouping

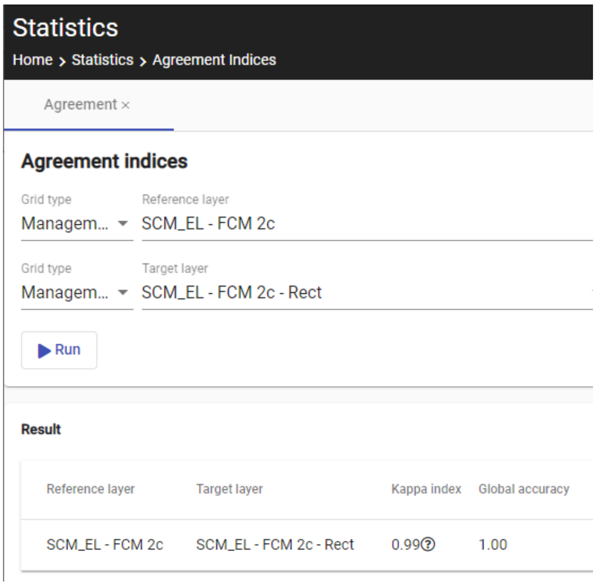

Layer statistics by group (Yield Soybean Mean (t/ha) into SCM_EL - FCM 2c)

Class	Count	Avg	Tk	SD	Var	CV	Min	Q1	Me	Q3	Max	Skew	Kurt
1	28	4.30	b	0.27	0.07	6.29	3.81	4.12	4.28	4.50	4.85	0.27	-0.48
2	24	4.49	a	0.34	0.12	7.67	3.83	4.35	4.52	4.77	5.06	-0.47	-0.67

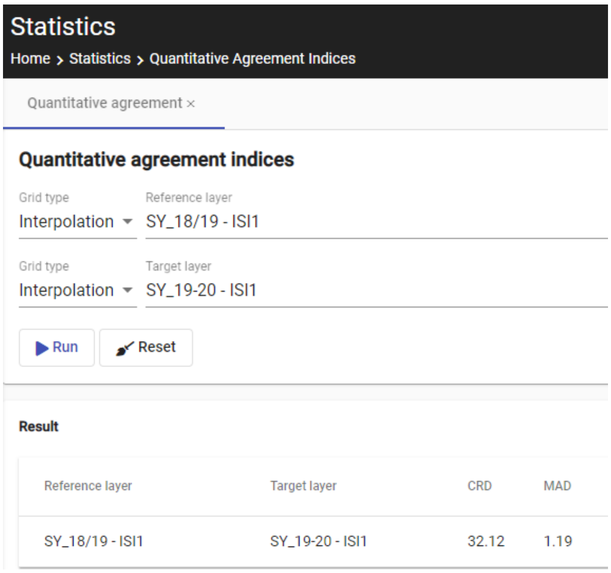
b) Descriptive statistical analysis of the target variable

Fig. 25 Graphical user interface for descriptive statistical analysis by class and obtaining grouping quality indices: (a) quality indices for each cluster and (b) descriptive statistical analysis of the target variable.

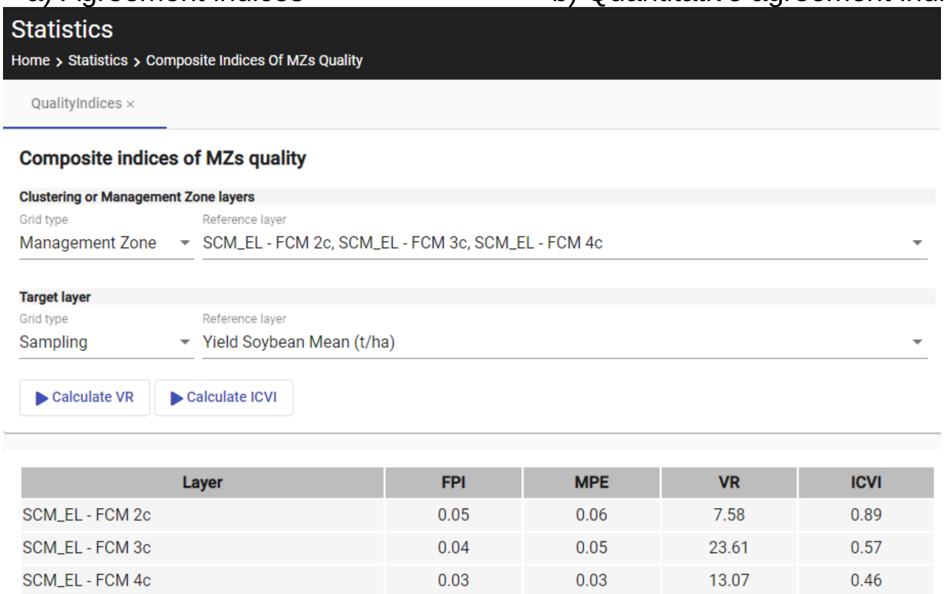
In the statistical analysis module, it is possible to calculate the Kappa and Global Accuracy indices (Fig. 26a) ammong TMs (as long as discretized) and MZs, the CRD and MAD quantitative concordance indices (Fig. 26b), and the composite indices for the MZs quality evaluation (Fig. 26c).



a) Agreement indices



b) Quantitative agreement indices



c) Composite indices of MZs quality

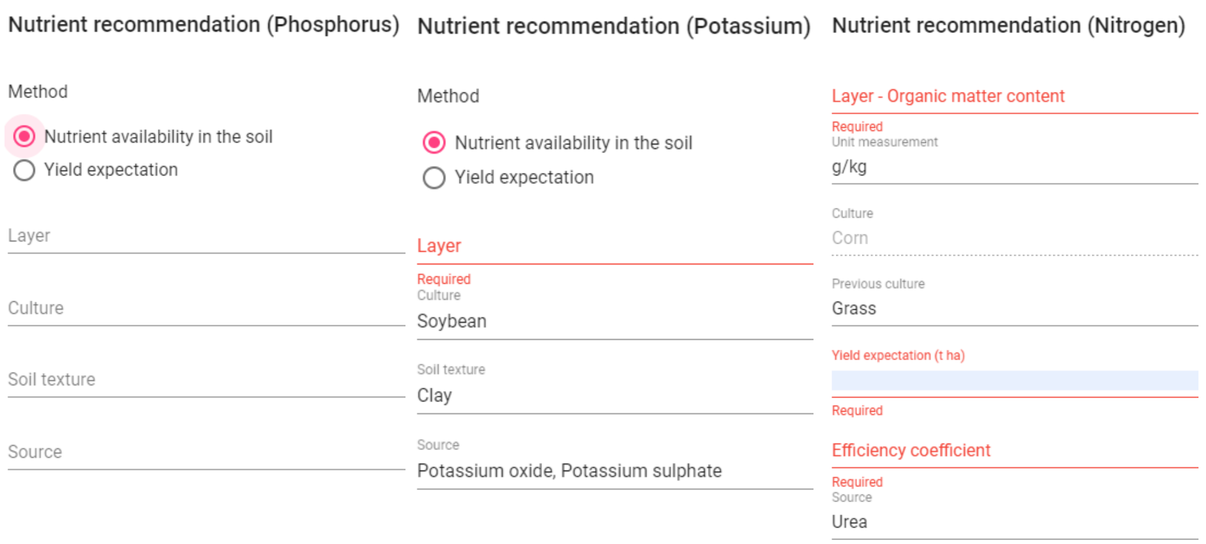
Fig. 26 Tabs of the statistical analysis interface to calculate (a) the Kappa and Global Accuracy indices, (b) the quantitative agreement indices, and (c) the composite indices of management zone quality.

MZs presented smoothness (SI) above 96%, with the highest smoothness being 99% (Table 12). As observed by other authors (Betzek et al., 2018), SI decreased with classes. The best ASC was obtained in SCM_EL design in four classes (0.65) and the worst in MPCA_1-3 in two classes (0.35). SCM_EL design, divided into three and four classes, obtained the best ICVI (0.33) and the MPCA_1-3 design into two classes, the worst (0.96). The VR% varied from 2% (PCA_1-3; 2 classes) to 50% (SCM_EL-MSY; 4 classes). When divided into two classes, the mean soybean yield was different among all classes only in SCM_EL, SCM_EL-MSY, PCA_1-2, and MPCA_1-3 designs. The lowest MGQI value was found in SCM_EL design, in four classes (0.53), and the highest in MPCA_1-3 design, in two classes (2.08).

The procedure to select the best method combination to MZs delineation was carried out in two steps: (i) the Tukey's test was applied to identify whether the generated classes showed significant differences in terms of the normalized average yield values; (ii) it was chosen SCM_EL_2-classes as the best clustering method combination by the best MGQI (the lowest, 0.71).

viii. Nutrient recommendation

ADB-Map has modules for nutrient recommendation (Beneduzzi, 2020) and lime recommendation (Moreira, 2019). With this, graphical interfaces are available to calculate the amount of fertilizer needed to correct the nutrients phosphorus (Fig. 27a), potassium (Fig. 27b), and nitrogen (Fig. 27c) on soil, as well as the lime recommendation graphical interface (Fig. 28). The interfaces are similar, but some fields differ due to the parameters needed to perform the calculations.



(a) Phosphorus recommendation (b) Potassium (c) recommendation (c) Nitrogen recommendation

Fig. 27 Screens to calculate phosphorus, potassium, and nitrogen recommendation.

Table 12 Evaluation statistics of management zones delineated with the elevation variable, selected by SCM method (SCM_EL)

VSM	NC	Tukey's Test C1	C2	C3	C4	SIr%	ASC	FPI	MPE	VR%	ICVI	FIr%	MGQI
SCM_EL	2	4.30a	4.49b	-	-	99	0.64	0.05	0.07	7.6	0.45	0	0.71
SCM_EL-MSY		4.26a	4.62b	-	-	99	0.53	0.08	0.10	28.6	0.42	0	0.81
PCA_1-3		4.32a	4.45a	-	-	99	0.39	0.15	0.19	2.2	0.80	0	2.07
PCA_1-2		4.27a	4.55b	-	-	99	0.51	0.08	0.10	18.2	0.49	0	0.96
MPCA_1-3		4.30a	4.49b	-	-	99	0.35	0.21	0.26	7.6	0.96	0	2.08
SCM_EL	3	4.27a	4.27a	4.61b	-	97	0.62	0.04	0.05	23.6	0.33	0	0.55
SCM_EL-MSY		4.14a	4.36a	4.61b	-	97	0.45	0.13	0.13	31.3	0.45	0	1.05
PCA_1-3		4.26a	4.37a	4.63b	-	98	0.44	0.11	0.12	16.6	0.46	0	0.98
PCA_1-2		4.21a	4.28a	4.58b	-	98	0.48	0.09	0.10	23.0	0.46	0	0.98
MPCA_1-3		4.24a	4.29a	4.63b	-	98	0.39	0.15	0.16	23.6	0.64	0	1.67
SCM_EL	4	4.27a	4.33ab	4.36ab	4.63b	96	0.65	0.03	0.03	13.1	0.33	0	0.53
SCM_EL-MSY		4.14a	4.35b	4.40b	4.81c	96	0.41	0.13	0.12	50.1	0.36	25	1.13
PCA_1-3		4.19a	4.31a	4.40ab	4.63b	97	0.45	0.10	0.10	17.6	0.55	0	1.30
PCA_1-2		4.21a	4.24a	4.40ab	4.63b	97	0.47	0.10	0.10	22.8	0.49	0	1.08
MPCA_1-3		4.24a	4.30a	4.30a	4.72b	97	0.44	0.13	0.12	24.3	0.55	0	1.29

VSM: variable selection method (SCM - Spatial Correlation Matrix, PCA - Principal Component Analysis, and MPCA - MULTISPATI-PCA); NC: Number of classes; SIr%: Smoothness index of the rectified management zone; ASC: Average silhouette coefficient; FPI: Fuzziness performance index; MPE: Modified partition entropy; VR%: Variance reduction; ICVI: Improved cluster validation index; Fir%: Fragmentation index of the rectified management zone; MGQI: Modified global quality index; Mean followed by the same letters did not differ by Tukey test at 5% probability. Clusterings that presented all statistically different classes are highlighted in gray. The best clustering for each field is highlighted in blue.

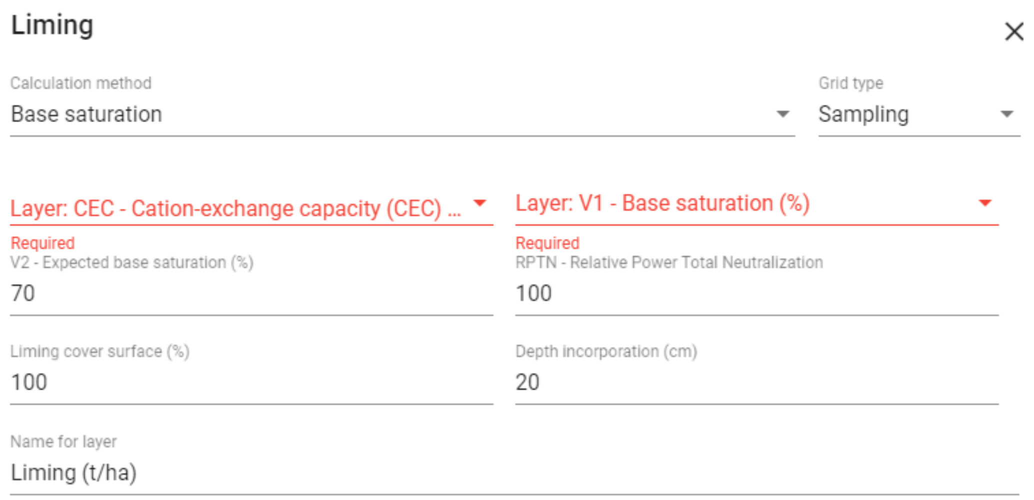


Fig. 28 Screen to calculate lime recommendation.

ix. User's evaluation of the software

As for the general functionality of the ADB-Map: (i) it was consensus that the functionalities offered by the ADB-Map are appropriate for the AP (Fig. 29a), (ii) on the execution of requests made to ADB-Map, the agreement was higher (50%) and total (50%) that they are correctly executed (Fig. 29b), (iii) in the TMs creation and MZs defining, 21% (Fig. 29c) and 7% (Fig. 29d) of respondents moderately agree that they are generated correctly; however, most of them have a great or total agreement (Fig. 29c and Fig. 29d), and (iv) most interviewers (93%) agree that the ADB-Map prevents unauthorized access to software and data (Fig. 29e).

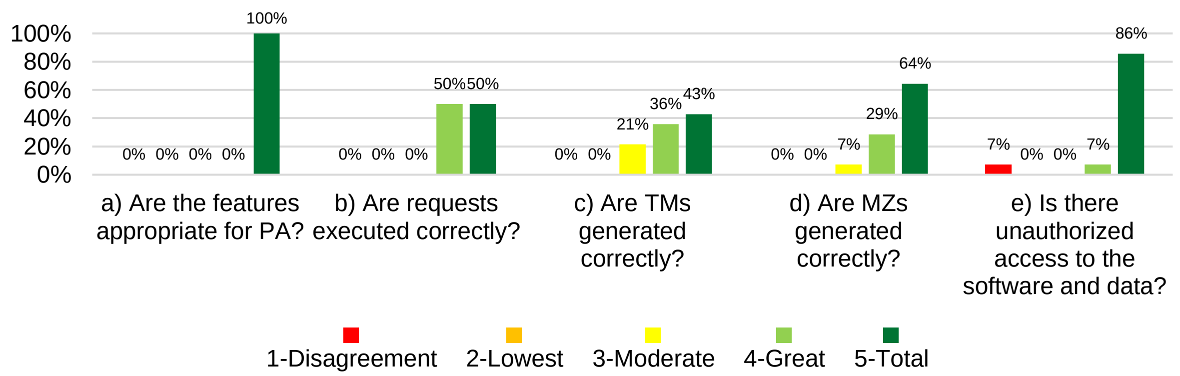


Fig. 29 Evaluation of AgDataBox-Map's general functionality. PA: precision agriculture, TMs: thematic maps, and ZMs: management zones.

Most of the features offered by the ADB-Map (Fig. 30) were considered of great or total importance. However, there are some answers that users do not know to answer (Interpolation by MA, Interpolation by NN, Spatial Correlation Matrix, Grid cutting out, and Agreement indices).

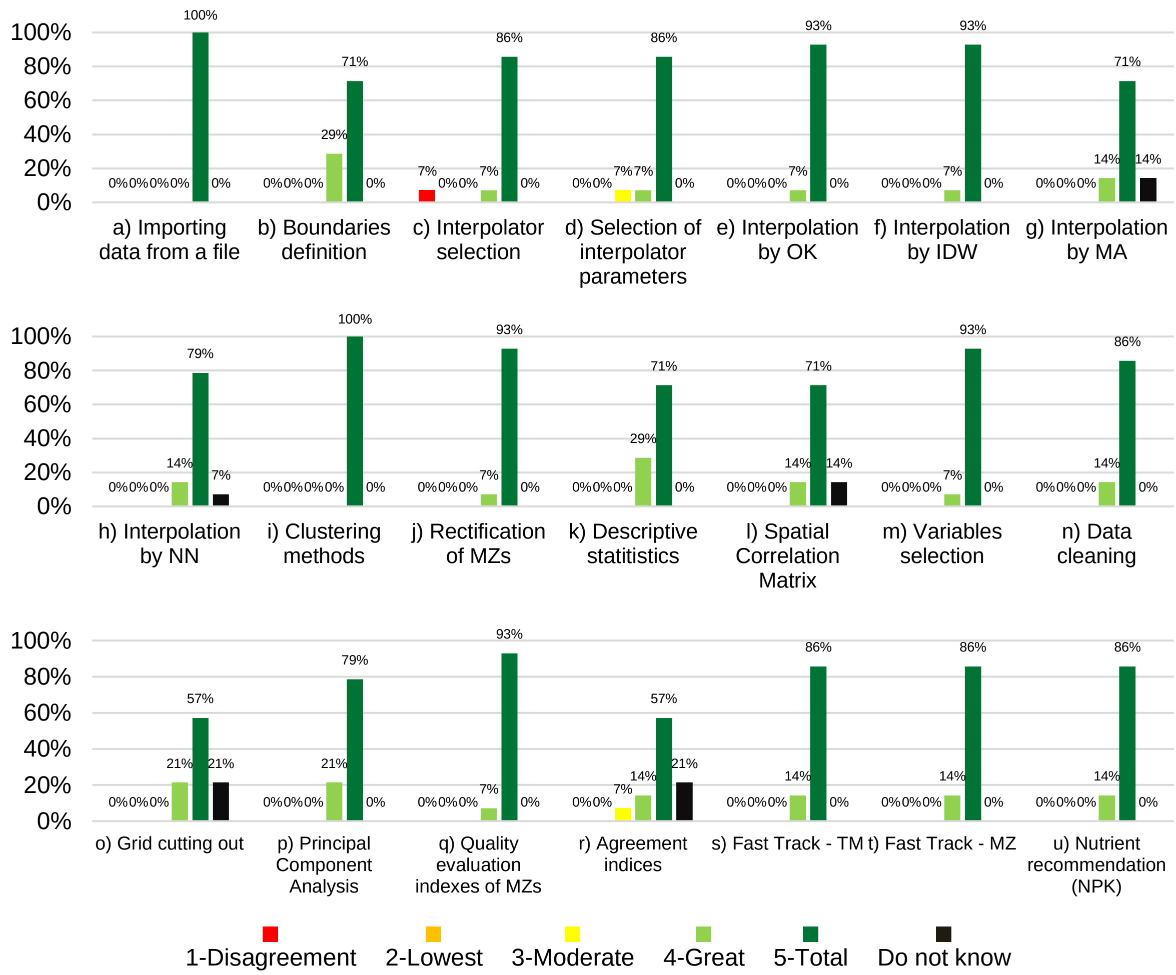


Fig. 30 AgDataBox-Map evaluation of user perception of the importance of ADB-Map features.

In the reliability aspect, most interviewers agree that the ADB-Map does not frequently fail, but 14% moderately agree (Fig. 31a). It is important that it reacts to software failure events, gets back up and running, and makes users' data available for use. In this aspect, most respondents indicated great or total agreement that the ADB-Map returns to work in case of failure (Fig. 31b). In addition, there was a total agreement regarding the availability of user data in case of application failure (Fig. 31c).

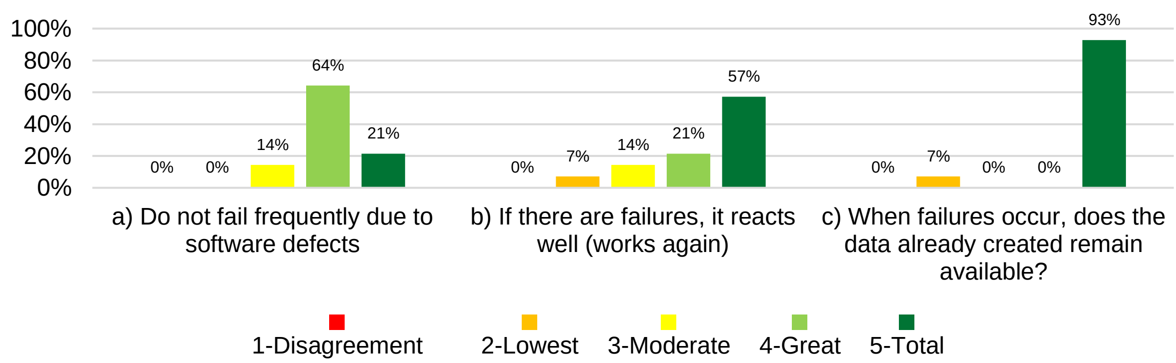


Fig. 31 AgDataBox-Map user evaluation of application reliability.

Regarding the software usability evaluation: a total agreement of users prevails in aspects (i) ease of access to the ADB-Map (Fig. 32a, 93%), (ii) ease of learning to use the software (Fig. 32c, 93%), (iii) adequacy of layer categories (Fig. 32d, 71%), and (iv) adequacy of font type, text size, page colors, and shadows (Fig. 32e, 64%). Furthermore, it prevails great and total agreement regarding the ease-of-use aspect (Fig. 32b, 93%).

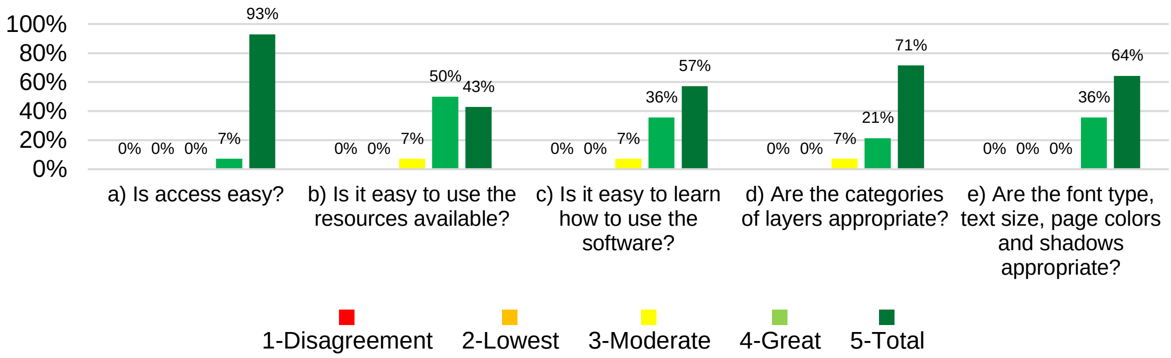


Fig. 32 AgDataBox-Map user evaluation of usability aspects.

When investigating the ease-of-use aspect by features available in the ADB-Map (Fig. 33), it is clear that the predominance of responses is in total agreement that it is easy to use. However, some of the features of the ADB-Map were not used by a small part of the interviewers (Boundaries definition, OK Interpolation, MA interpolation, NN interpolation, MZ rectification, SCM, Grid cutting out, PCA, Composite indices of MZs quality, agreement indices, Fast track – TM, and Nutrient/lime recommendation).

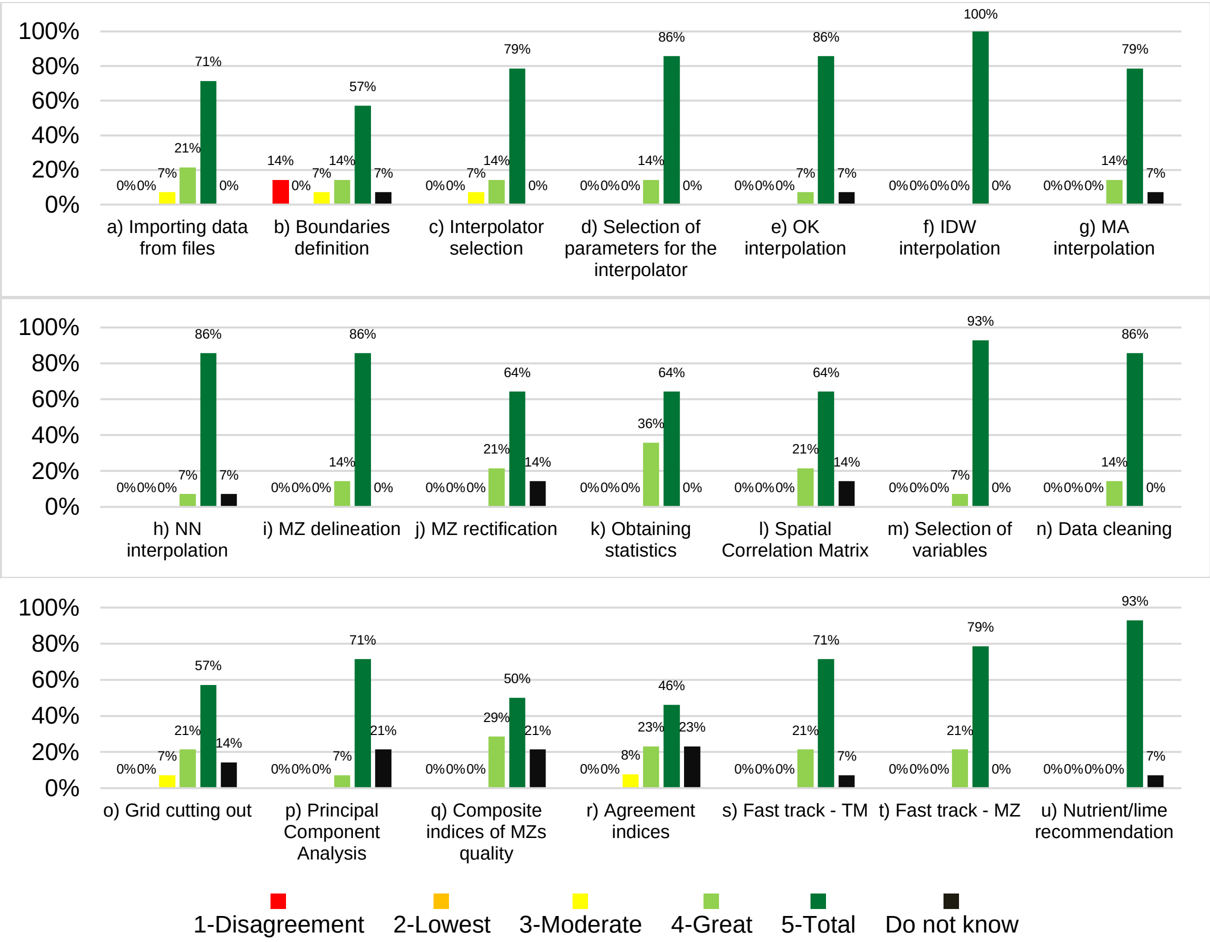


Fig. 33 AgDataBox-Map user evaluation of the features ease of use.

In the application efficiency aspect, it is clear that the interviewers were completely satisfied with the time to execute the seven evaluated features (Fig. 34), even with the interpolator selection process (Fig. 34b), which consumes the most processing resources and server memory, was not a cause for dissatisfaction by users.

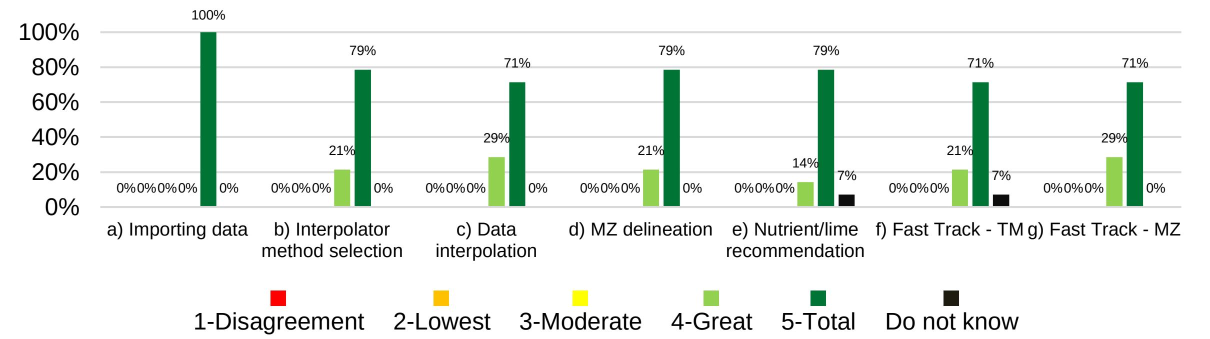


Fig. 34 AgDataBox-Map user evaluation of efficiency regarding the execution time of available features.

Among the interviewers who observed problems and failures due to software defects, 69% had a total and 31% great satisfaction with the time taken to solve problems (Fig. 35).

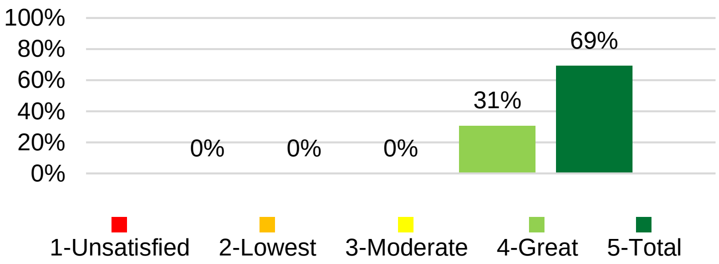


Fig. 35 AgDataBox-Map user evaluation regarding the application maintainability.

In terms of application portability, note that the application was run on Google Chrome, Mozilla Firefox, Microsoft Edge, and Safari browsers (Fig. 36). Most of the interviewers ran ADB-Map on Google Chrome (86%) and were completely satisfied with the quality of running on this browser. Interviewers did not use Opera, Internet Explorer, and other browsers.

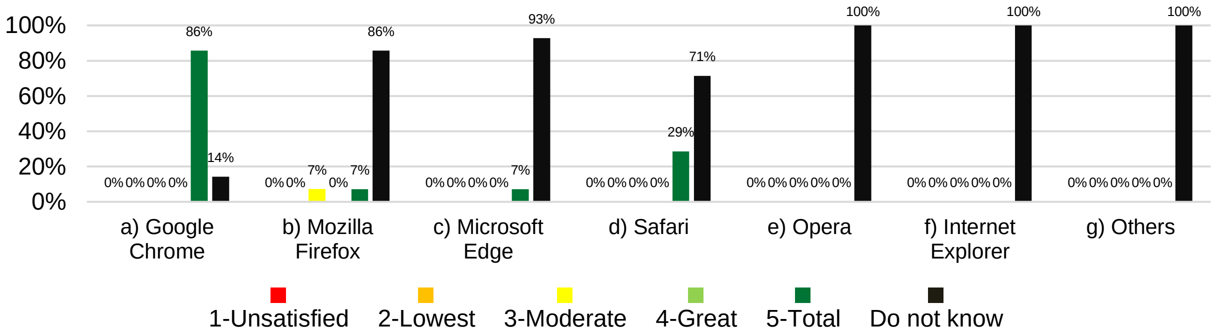


Fig. 36 Browsers used by interviewers to run the AgDataBox-Map.

The Windows operating system (OS) was used by 79% of interviewers who were totally satisfied with running the ADB-Map (Fig. 37). Linux and iOS were used by a small part of the interviewers, who showed great and total satisfaction when the ADB-Map was run on these OSs.

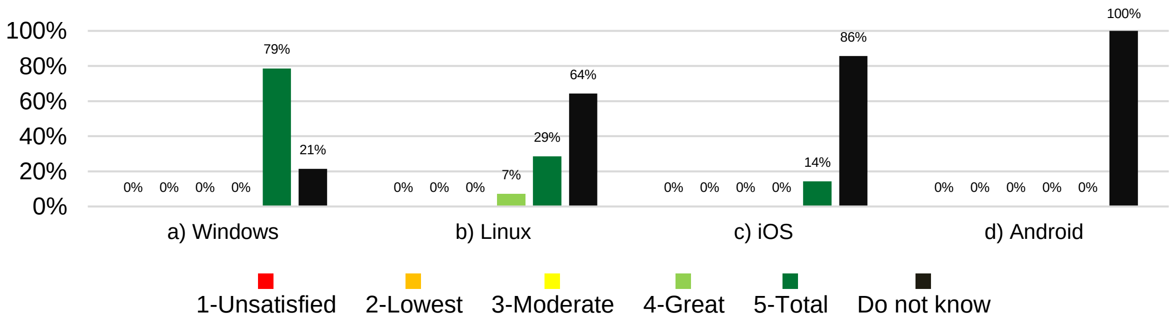


Fig. 37 Operational systems used by interviewers to run the AgDataBox-Map.

All interviewers used the ADB-Map on computers or notebooks, and only 7% also used it on mobile devices (Fig. 38). On both types of devices, the satisfaction of use was total.

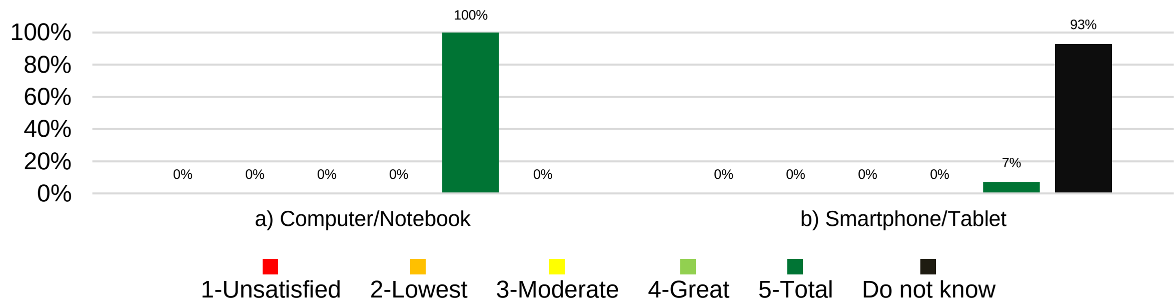


Fig. 38 Electronic devices used by interviewers to run the AgDataBox-Map.

In all categories of the quality model that the ADB-Map was evaluated, a score great than 70% was obtained (Fig. 39). Furthermore, the global score for the user's quality evaluation was 92%. Therefore, we consider the software well accepted by users.

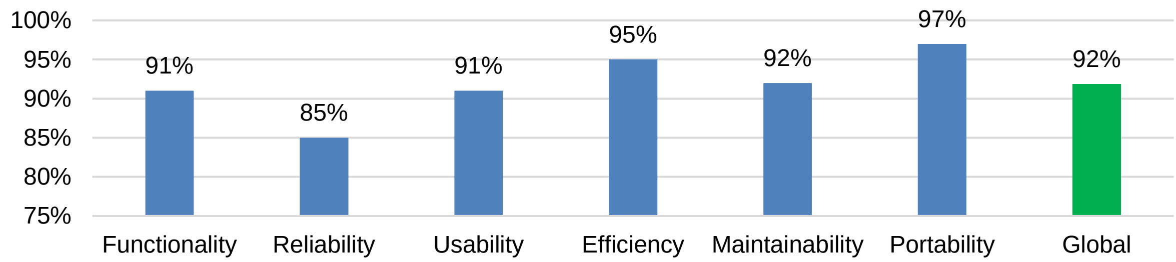


Fig. 39 Score obtained in the AgDataBox-Map evaluation in each category of the software quality model.

6.4 Conclusions

The application developed in this research allowed quickly generating TMs and delineating MZs, as it provides user-friendly graphical interfaces. Moreover, the features could be integrated and consumed from ADB-MSA.

In the quality evaluation performed by the user, the ADB-Map application received a score of 92%, considering the criteria defined by the researchers. Thus, we consider the application accepted by the community and generate TMs, and delineate MZs.

The SCM_EL (elevation variable selected by the spatial correlation matrix) design, divided into two classes, was considered the best in the case study.

The new MGQI index evaluated the MZs quality, considering the composition of several indices provided by the ADB-Map.

6.5 Acknowledgments

The authors would like to thank the Western Paraná State University (UNIOESTE), the Federal University of Technology of Paraná (UTFPR), the Coordination for the Upgrading of Higher Education Personnel (CAPES), the National Council for Scientific and Technological

Development (CNPq), the Itaipu Technological Park Foundation (FPTI), and the Ministry of Agriculture, Livestock and Food Supply (MAPA) for funding this project.

6.6 References

ABNT. (2003). **NBR ISO/IEC 9126-1:2003**: engenharia de software: qualidade de produto. Parte 1: modelo de qualidade. Associação Brasileira de Normas Técnicas, Rio de Janeiro.

ABNT. (2004). **NBR ISO/IEC 14598-6:2004**: engenharia de software: avaliação de produto. Parte 6: documentação de módulos de avaliação. Associação Brasileira de Normas Técnicas, Rio de Janeiro.

Albornoz, E. M., Kemerer, A. C., Galarza, R., Mastaglia, N., Melchiori, R., Martínez, C. E. 2018. Development and evaluation of an automatic software for management zone delineation. **Precision Agriculture**,19 (3), pp. 463-476.

Anderberg, M. R. (1973). **Cluster Analysis for Applications**. Academic Press, New York.

Bambini, M. D., Mendes, C. I. C., Moura, M. F., Oliveira, S. R. M. (2013). Software para agropecuária: panorama do mercado brasileiro. **Parc. Estrat.**, 18 (36), pp. 175-198.

Bazzi, C. L., Souza, E. G., Schenatto, K., Betzek, N. M., Gavioli, A. (2019). A software for the delineation of crop management zones (SDUM). **Australian Journal of Crop Science. Southern Cross Journals**, 13, pp. 26-34.

Bazzi, C. L., Souza, E. G. de, Betzek, N. M. (2015). **Software para Definição de Unidades de Manejo**: Teoria e prática. 1. UNIOESTE, Cascavel.

Bazzi, C. L., Souza, E. G., Uribe-Opazo, M. A., Nóbrega, L. H. P., Rocha, D. M. (2013). Management zones definition using soil chemical and physical attributes in a soybean area. **Engenharia Agrícola**, 33 (5), pp. 952–964.

Behera, S. K., Mathur, R. K., Shukla, A. K., Suresh, K., Prakash, C. (2018). Spatial variability of soil properties and delineation of soil management zones of oil palm plantations grown in a hot and humid tropical region of southern India. **CATENA**, 165, pp. 251-259.

Beneduzzi, H. M. (2020). **Módulo computacional para cálculo da necessidade de nitrogênio, fósforo e potássio a partir de suas disponibilidades no solo**. PhD thesis, Western Paraná State University (UNIOESTE), Cascavel.

Betzek, N. M., Souza, E. G., Bazzi, C. L., Schenatto, K., Gavioli, A. (2018). Rectification methods for optimization of management zones. **Computers and Electronics in Agriculture**, 146 (1), pp. 1–11.

Betzek, N. M., Souza, E. G., Bazzi, C. L., Schenatto K., Gavioli, A., Magalhães, P. S. G. (2019). Computational routines for the automatic selection of the best parameters used by interpolation methods to create thematic maps. **Computers and Electronics in Agriculture**, 157, pp. 49-62.

Bezdek, J. C. (1981). **Pattern Recognition with Fuzzy Objective Function Algorithms**. Plenum Press, New York.

Bier, V. A., Souza, E. G. (2017). Interpolation selection index for delineation of thematic maps. **Computers and Electronics in Agriculture**, 136 (1), pp. 202-209.

Breunig, F. M., Galvão, L. S., Dalagnol, R., Santi, A. L., Shuisen Chen, D. P. D. F. (2020). Assessing the effect of spatial resolution on the delineation of management zones for smallholder farming in southern Brazil. **Remote Sensing Applications: Society and Environment**, 19.

Borges, L. G., Bazzi, C. L., Souza, E. G., Magalhães, P. S. G., Michelon, G. K. (2020). Web software to create thematic maps for precision agriculture. **Pesq. agropec. bras**., 55.

Chang, D., Zhang, J., Zhu, L., Ge, S. H., Li, P. Y., Liu, G. S. (2014). Delineation of management zones using an active canopy sensor for a tobacco field. **Computers and Electronics in Agriculture**, 109, pp. 172-178.

Chipman, H., Tibshirani, R. (2006). Hybrid Hierarchical Clustering with Applications to Microarray Data. **Biostatistics**, 7, pp. 302-317.

Cisternas, I., Velásquez, I., Caro, A., Rodríguez, A. (2020). Systematic literature review of implementations of precision agriculture, **Computers and Electronics in Agriculture**, 176, pp. 1-11.

Coelho, E. C., Souza, E. G., Uribe-Opazo, M. A., Pinheiro Neto, R. (2009). Influência da densidade amostral e do tipo de interpolador na elaboração de mapas temáticos. **Acta Scientiarum**, 31 (1), pp. 165-174.

Cohen, J. A. (1960). Coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, 20 (1), pp. 37-46.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing of Environment**, 37 (1), p. 35-46.

Córdoba, M. A., Bruno, C. I., Costa, J. L., Peralta, N. R., Balzarini, M. G. (2016). Protocol for multivariate homogeneous zone delineation in precision agriculture. **Biosystems Engineering**, 143, pp. 95-107.

Damian, J. M., Pias, O. H. C., Cherubin, M. R., Fonseca, A. Z., Fornari, E. Z., Santi, A. L. (2020). Applying the NDVI from satellite images in delimiting management zones for annual crops**. Sci. agric.**, 77 (1), pp. 1-11.

Dave, R. N. (1992). Generalized fuzzy c-shells clustering and detection of circular and elliptical boundaries. **Pattern Recognition**, 25 (7), pp. 713-721.

Dhillon, I. S., Modha, D. S. (2001). Concept decompositions for large sparse text data using clustering. **Machine Learning**, 42, pp. 143-175.

Doerge, T. A. (2000). **Management Zone Concepts**. Site-Specific Management Guidelines. Potash and Phosphate Institute. University South Dakota, Brokings.

Dray, S., Saïd, S., Débias, F. (2008). Spatial ordination of vegetation data using a generalization of Wartenberg’s multivariate spatial correlation. **Journal of Vegetation Science**, 19 (1), pp. 45-56.

Eitzinger, A., Cock, J., Atzmanstorfer, K., Binder, C. R., Läderach, P., Bonilla-Findji, O., Bartling, M. Mwongera, C., Zurita, L., Jarvis, A. (2019). GeoFarmer: A monitoring and feedback system for agricultural development projects. **Computers and Electronics in Agriculture**, 158, pp. 109-121.

Ferguson, R. B.; Hergert, G. W. 2009. Soil Sampling for Precision Agriculture. **Precision Agriculture**, pp. 1-4.

Foody, G. M. (2002). Status of land cover classification accuracy assessment. **Remote Sensing of Environment**, 80 (1), pp. 185–201.

Fridgen, J. J., Kitchen, N. R., Sudduth, K. A., Drummond, S. T., Wiebold, W. J., Fraisse, C. W. (2004). Management zone analyst (MZA): software for subfield management zone delineation. **Agronomy Journal**, 96 (1), pp. 100-108.

Gavioli, A., Souza, E. G., Bazzi, C. L., Guedes, L. P. C., Schenatto, K. (2016). Optimization of management zone delineation by using spatial principal components. **Computers and Electronics in Agriculture**, 127 (1), pp. 302-310.

Gavioli, A., Souza, E. G., Bazzi, C. L., Schenatto, K., Betzek, N. M. (2019). Identification of management zones in precision agriculture: An evaluation of alternative cluster analysis methods. **Biosystems Engineering**, 181, pp. 86-102.

Gebbers, R., Adamchuk, V. I. (2010). Precision Agriculture and Food Security. **Science**, 327 (5967), pp. 828–831.

Gonzalez, R. C., Woods, R. (2008). **Digital image processing**. 3. Pearson Prentice Hall, New Jersey.

Gray, B., Babcock, L., Tobias, L., McCord, M., Herrera, A., Osei, C., Cadavid, R. (2018). **Digital farmer profile**: Reimagining Smallholder Agriculture. United States Agency for International Development (USAID), Washington.

Hanafizadeh, P., Khosravi, B., Tabatabaeian, S. H. (2020). Rethinking dominant theories used in information systems field in the digital platform era. **Digital Policy, Regulation and Governance**, 22 (4), pp. 363-384.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. **Journal of Educational Psychology**, 24 (6), pp. 417-441.

ILOSTAT, 2019. **Employment database**. Geneva: International Labour Organization. [Data retrieved May 2019].

Isaaks, E. H., Srivastava, R. M. (1989). **Applied geostatistics**. Oxford University Press, New York.

ISO/IEC. (2014). **ISO/IEC 25000:2014** - Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Guide to SQuaRE. Int'l Organization for Standardization.

ISO/IEC. (2019). **ISO/IEC 25030:2019** - Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Quality requirement. Int'l Organization for Standardization.

ISO/IEC. (2011). **ISO/IEC 25040:2011** - Software engineering–System and software Quality Requirements and Evaluation (SQuaRE) – Evaluation process. Int'l Organization for Standardization.

Jain, A. K., Dubes, R. (1988). **Algorithms for clustering data**. Prentice-Hall, Englewood Cliffs.

Johnson, R. A., Wichern, D. W. (2007). Applied Multivariate Statistical Analysis. 6. Pearson Prentice Hall, Upper Saddle River.

Journel, A. G., Huijbregts, C. J. (1978). **Mining Geostatistics**. Academic Press, London-New York-San Francisco.

Kaufman, L., Rousseeuw, P. J. (1990). **Finding groups in data.** John Wiley & Sons, Hoboken.

Landis, J. R., Koch, G. G. (1977). The measurement of observer agreement for categorical data. **Biometrics**, 33 (1), pp. 159-174.

Larscheid G., Blackmore, B. S. (1996). Interactions between farm managers and information systems with respect to yield mapping. In: **International Conference on Precision Agriculture**, 3. Springer, Minneapolis, pp.1153-1163.

Leisch, F. (1999). Bagged clustering. In: **SFB adaptive information systems and modelling in economics and management science**, 51. Vienna University of Economics and Business, Vienna, pp. 1-11.

Lowder, S. K., Skoet, J., Raney, T. (2016). The number, size and distribution of farms, smallholder farms, and family farms worldwide. **World Development**, 86, pp. 16–29.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In: **Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability**, 1. University of California Press, Berkeley, pp. 281–297.

Martinetz, T. M., Berkovich, S. G., Schulten, K. J. (1993). "Neural-gas" network for vector quantization and its application to time-series prediction. **IEEE Transactions on Neural Networks**, 4 (4), pp. 558-569.

Matheron, G. 1963. Principles of geostatistics. **Economic Geology**, 58 (8), p. 1246-1266.

Méndez-Vázquez, L. J., Lira-Noriega, A., Lasa-Covarrubias, R., Cerdeira-Estrada, S. (2019). Delineation of site-specific management zones for pest control purposes: Exploring precision agriculture and species distribution modeling approaches. **Computers and Electronics in Agriculture**, 167, pp. 1-15.

McBratney, A. B., Moore, A. W. (1985). Application of fuzzy sets to climatic classification. Agricultural and Forest Meteorology. **Goettingen**, 35 (1-4), pp. 165-185.

McQuitty, L. L. (1966). Similarity Analysis by Reciprocal Pairs for Discrete and Continuous Data. **Educational and Psychological Measurement**, 26, pp. 825-831.

Michelon, G. K., Bazzi, C. L., Upadhyaya, S., Souza, E. G., Magalhães, P. S. G., Borges, L. F., Schenatto, K., Sobjak, R., Gavioli, A., Betzek, N. M. (2019). Software AgDataBox-Map to precision agriculture management. **SoftwareX**, 10.

Mielke, P. W., Berry. K. J. (2007). **Permutation methods: a distance function approach**. Springer, New York.

Milligan, G. W., Cooper, M. C. (1988). A study of standardization of variables in cluster analysis. **Journal of Classification**, 5 (2), pp. 181–204.

Minasny, B., McBratney, A. B. (2002). **FuzME version 3**. Australian Centre for Precision Agriculture, The University of Sydney.

Moreira, W. K. O. (2019). **Módulo computacional para delineamento de mapas de aplicação de calcário a partir dos atributos químicos do solo**. Master's dissertation, Western Paraná State University (UNIOESTE), Cascavel.

Nicol, L. A., Nicol, C. J. (2021). Adoption of precision agriculture in Alberta irrigation districts with implications for sustainability. **The Journal of Rural and Community Development**, 16 (1), pp. 152–174.

Oldoni, H., Terra, V. S. S., Timm, L. C., Reisser Júnior, C., Monteiro, A. B. (2019). Delineation of management zones in a peach orchard using multivariate and geostatistical analyses. **Soil and Tillage Research**, 191, p. 1-10.

Paccioretti, P., Córdoba, M., Balzarini, M. (2020). FastMapping: Software to create field maps and identify management zones in precision agriculture. **Computers and Electronics in Agriculture**, 175, p. 1-7.

Pal, N. R., Bezdek, J. C., Hathaway, R. J. (1996). Sequential competitive learning and the fuzzy c-means clustering algorithm. **Neural Networks**, 9 (5), pp. 787-796.

Peralta, N. R., Costa, J. L., Balzarini, M., Franco, M. C., Córdoba, M., Bullock, D. (2015). Delineation of management zones to improve nitrogen management of wheat. **Computers and Electronics in Agriculture**, 110 (1), pp. 103–113.

Pimentel-Gomes, F. 2009. **Curso de estatística experimental**, 15. FEALQ, Piracicaba.

R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org>.

Reich, R. M. (2008). **Spatial Statistical Modeling of Natural Resources**. Colorado State University, Fort Collins.

Reichardt, M., Jürgens, C. (2009). Adoption and future perspective of precision farming in Germany: Results of several surveys among different agricultural target groups. **Precision Agriculture**, 10 (1), pp. 73–94.

Rotz, S., Duncan, E., Small, M., Botschner, J., Dara, R., Mosby, I., Reed, M., Fraser, E. D. G. (2019). The politics of digital agricultural technologies: A preliminary review. **Sociologia Ruralis**, 59 (2), pp. 203–229.

Rousseeuw, P. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. **Journal of Computational and Applied Mathematics**, 20 (1), pp. 53-65.

Santos, R. T., Saraiva, A. M. (2015). A Reference Process for Management Zones Delineation in Precision Agriculture. **IEEE Latin America Transactions**, 13 (3), pp. 727-738.

Schenatto, K., Souza, E. G., Bazzi, C. L., Bier, V. A., Betzek, N. M., Gavioli, A. (2016). Data interpolation in the definition of management zones. **Acta Scientiarum**, 38 (1), pp. 31-40.

Schenatto, K., Souza, E. G., Bazzi, C. L., Gavioli, A., Betzek, N. M., Beneduzzi, H. B. (2017). Normalization of data for delineating management zones. **Computers and Electronics in Agriculture**, 143 (1), pp. 238-248.

Schepers, A. R., Shanahan, F. J., Liebig, M. A., Schepers, J. S., Johnson, S. H., Luchiari, J. A. (2004). Appropriateness of management zones for characterizing spatial variability of soil properties and irrigated corn yields across years. **Agronomy Journal**, 96 (1), pp. 195–203.

Souza, E. G., Bazzi, C. L., Khosla, R., Uribe-Opazo, M. A., Reich, R. M. 2016. Interpolation type and data computation of crop yield maps is important for precision crop production. **Journal of Plant Nutrition**, 39 (4), pp. 531-538.

Souza, E. G., Schenatto, K., Bazzi, C. L. 2018. Creating thematic maps and management zones for agriculture fields. In: **Proceedings of the 14th International Conference On Precision Agriculture** (IPCA).

Swindell, J. (1997). Mapping the spatial variability in the yield potential of arable land through GIS analysis of sequential yield maps. In 1st European Conference on Precision Agriculture (pp. 827- 834). Warwick.

Trendov, N. M., Varas, S., Zeng, M. (2019). **Digital technologies**: In agriculture and rural areas. Food and Agriculture Organization of the United Nations (FAO), Rome. 26 p.

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. **Journal of the American Statistical Association**, 58 (301), pp. 236-244.

Xiang, L., Yu-Chun, P., Zhong-Qiang, G., Chun-Jiang, Z. (2007). Delineation and Scale Effect of Precision Agriculture Management Zones Using Yield Monitor Data Over Four Years. **Agricultural Sciences In China**, 6 (2), pp. 180-188.

Xu, R., Wunsch, D. C. (2009). **Clustering. Piscataway**: IEEE Press.

Zhang, N., Wang, M., Wang, N. 2002. Precision agriculture - a worldwide overview. **Computers and Electronics in Agriculture**, 36 (2-3), pp. 113-132.

Zhang, X., Shi, L., Jia, X., Seielstad, G., Helgason, C. (2010). Zone mapping application for precision-farming: a decision support tool for variable rate application. **Precision Agriculture**, 11 (2), pp. 103-114.

Appendix A

Data normalization methods

- Range (Anderberg, 1973; Milligan and Cooper, 1988 – Equation A1):

$$Z_{iN} = \frac{X_i - Median}{Max(X) - Min(X)}, \tag{A1}$$

where, Z_{iN} – normalized observation i; X_i – original data value i; $Min(X)$ – minimum value of data set; $Max(X)$ – maximum value of the data set.

- Mean (Swindel, 1997 – Equation A2):

$$Z_{iN} = \frac{X_i}{\overline{X}}, \tag{A2}$$

where, Z_{iN} – normalized observation i; X_i – original data value i; $\bar{X}$ – sample mean of data set.

- Standard Score or Z-Score (Larscheid and Blackmore, 1996 – Equation A3):

$$Z_{iN} = \frac{X_i - \bar{X}}{s}, \tag{A3}$$

where, Z_{iN} – normalized observation i; X_i – original data value i; $\bar{X}$ – sample mean of data set; s – standard deviation of data set.

- Min-Max method (Milligan and Cooper, 1988 – Equation A4):

$$Z_{iN} = \frac{X_i - Min(X)}{Max(X) - Min(X)}, \tag{A4}$$

where, Z_{iN} – normalized observation i; X_i – original data value i; $Min(X)$ – minimum value of data set; $Max(X)$ – maximum value of data set.

The Bivariate Moran's I (Reich, 2008; Schepers et al., 2004 – Equation A5):

$$I_{YZ} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} y_i z_j}{W\sqrt{m_Y^2 m_Z^2}}, \tag{A5}$$

where I_{YZ}: Degree of spatial association between Y and Z variables, ranging from -1 to 1, as it is followed: positive correlation $I_{YZ} > 0$ and negative correlation $I_{YZ} < 0$; w_{ij}: corresponds to the ij element of spatial association matrix, calculated by $w_{ij} = (1/(1 + D_{ij}))$, so that D_{ij} is the distance between i e j points; y_i and z_i: transformed y and z values, respectively, at point i ($i = 1, 2, \dots, n$), to get a zero average by the formulas $y_i = (y_i - \bar{Y})$ and $z_j = (z_j - \bar{Z})$, where $\bar{Y}$ and $\bar{Z}$ are the sample means of Y and Z variables; W: it is the sum of spatial association degrees obtained by w_{ij} matrix, for $i \neq j$; m_Y^2 and m_Z^2: sample variance of Y and Z variables, respectively.

The interpolator selection index (ISI – Bier and Souza, 2017 – Equation A6):

$$ISI = \left\{ \frac{abs(ME)}{max\left|_{i\,=\,1}^{\,j} [abs(ME)]\right.} + \frac{\left[SDME - min\left|_{i\,=\,1}^{\,j} SDME\right.\right]}{max\left|_{i\,=\,1}^{\,j} [abs(SDME)]\right.} \right\}, \tag{A6}$$

where ME (Equation A7) is the mean error; $SDME$ (Equation A8) is the standard deviation of mean error of crossed validation; n is the number of data; abs is the module value; $min|_{i=1}^{j}$ is the lowest value obtained among the compared j models; $max|_{i=1}^{j}$ is the highest value obtained among the compared j models.

$$ME = \frac{1}{n}\sum_{i=1}^{n} Z(s_i) - \hat{Z}(s_i), \tag{A7}$$

$$SDME = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Z(s_i) - \hat{Z}(s_i)\right)^2}, \tag{A8}$$

where n is the number of data; $Z(s_i)$ is the value observed at the point s_i; $\hat{Z}(s_i)$ is the predicted value at the point s_i.

The statistic called error comparison index (ECI – Souza et al., 2016 – Equation A9) was used to determine the best semivariogram fit in each j model analyzed, which assumes that a lower value for the model is better stochastic methods of interpolation. The best semivariogram of each j model was used in ISI analysis.

$$ECI_i = \frac{|RME_i|}{10^{-10} + max\Big|_{i\,=\,1}^{j}|RME|} + \frac{|SDRME_i - 1|}{10^{-10} + max\Big|_{i\,=\,1}^{j}|SDRME - 1|}, \tag{A9}$$

where ECI_i is the error comparison index for model i; and $max\Big|_{i\,=\,1}^{j}$ is the highest value among the compared j semivariograms. The arbitrary constant 10^{-10} was included to avoid division by zero.

The reduced mean error (RME – Equation A10) and the standard deviation of the reduced mean error (SDRME – Equation A11) were determined by ordinary kriging cross-validation.

$$RME = \frac{1}{n}\sum_{i=1}^{n}\frac{Z(s_i) - \hat{Z}(s_i)}{\hat{\sigma}(\hat{Z}(s_i))}, \tag{A10}$$

$$SDRME = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{\left|Z(s_i) - \hat{Z}(s_i)\right|}{\hat{\sigma}(\hat{Z}(s_i))}}, \tag{A11}$$

where $Z(s_i) - \hat{Z}(s_i)$ is the prediction error associated wbith estimating yield at spatial location s_i; $Z(s_i)$ is the observed value; $\hat{Z}(s_i)$ is the estimated value obtained from the ordinary kriging cross-validation; $\hat{\sigma}(\hat{Z}(s_i))$ is the estimated standard deviation associated with the estimated value, and n is the sample size.

The selection of the best semivariogram model

It considers three selection criteria when performing the best interpolator analysis:

1. A minimum of effective spatial dependence (%ESD) should be observed. The effective spatial dependence index (%ESDI – Equation A12), a new measure of the degree of spatial dependence, must be greater than 25%. This index considers the semivariance ($\gamma(1)$) in the first lag distance (h(1)).

$$\%ESDI = \frac{C - \gamma(1)}{C} * 100, \tag{A12}$$

where C is the sill (nugget effect + partial sill) and $\gamma(1)$ is the first semivariance of the semivariogram. The %ESDI was classified as %SDI.

2. The selected semivariogram model should contemplate a fraction of SD due only to the first semivariance significance index ($\%\gamma(1)$, Equation A13) lower than 50%.

$$\%\gamma(1) = \frac{\gamma(1) - C_0}{C_1} * 100, \tag{A13}$$

where C_0 is the nugget effect, C_1 is the partial sill, and $\gamma(1)$ is the first semivariance of the semivariogram.

3. The degree of inclination between the nugget effect and the last adjusted semivariance, estimated by the slope of the model ends index (%SMEI, Equation A14) should be greater than 20%. Otherwise, there is an indication of a pure nugget effect.

$$\%SMEI = \left(1 - \frac{\gamma_Z(0)}{10^{-10} + \gamma_Z(n)}\right) * 100 = \left(1 - \frac{C_0}{10^{-10} + \gamma_Z(n)}\right) * 100, \tag{A14}$$

where γ_Z is the adjusted theoric semivariance, $\gamma_Z(0) = C_0$ is the nugget effect, and $\gamma_Z(n)$ is the last adjusted theoretic semivariance, correspondent to the cutoff. The arbitrary constant 10^{-10} was included to avoid division by zero.

Data interpolation

The data interpolation methods available in ADB-Map are:

a) **Inverse Distance Weighting** (IDW – Equation A15): is calculated by:

$$\hat{Z}_i = \frac{\sum_{i=1}^{n}\left(\frac{1}{d_i^p} * Z_i\right)}{\sum_{i=1}^{n}\left(\frac{1}{d_i^p}\right)}, \tag{A15}$$

where, $\hat{Z}_i$ – interpolated value; Z_i – sampled attribute value; d_i^p – Euclidean distance between the ith neighborhood point and the sampled point, elevated to the power of p > 0.

b) **Ordinary Kriging** (OK – Equation A16) is made after adjusting the semivariogram model, and the value to be estimated at the point of interest.

$$\hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i * Z(x_i), \tag{A16}$$

where $\hat{Z}(x_0)$ – estimated value at a given location; λ_i – weight attributed to the sample values; $Z(x_i)$ – sampled attribute value; n – number of neighboring locations employed for interpolating the point, where the summation of the λ_i weights must be equal to one.

c) **Moving Average** (MA – Equation A17): estimates the non-sampled point values based on the mean of the sampled points within a predefined radial distance as given i. The points within the predefined radial distance are equally weighted (i.e., weight is 1/n) and the resulting value is the arithmetic average of the identified neighboring data (Bazzi et al., 2015).

$$\hat{Z}_i = \frac{\sum_{i=1}^{n} Z_i}{n}, \tag{A17}$$

where: $\hat{Z}_i$ is the interpolated value of the non-sampled point; Z_i is the neighboring sample point; n is the number of neighboring sample points used for interpolation of the non-sampled point.

Indices for evaluation of the management zones quality

a) **Variance reduction** (VR% – Xiang et al., 2007; Schenatto et al., 2017 – Equation A18): is calculated for a variable, with the expectation that the sum of data variances for each MZ is smaller than the total variance of the field.

$$VR\% = \left(1 - \frac{\sum_{i=1}^{c} W_i * V_{MZ_i}}{V_{field}}\right) * 100, \tag{A18}$$

where *c* is the number of MZs; W_i is the field rate of *i-th* MZ to the total field; V_{mzi} is the data variance of *i-th* MZ; V_{field} is the field data variance.

b) **Fuzziness Performance Index** (FPI – McBratney and Moore, 1985; Fridgen et al., 2004 – Equation A19): measures the degree of separation between the fuzzy c groups generated from a data set. FPI varies between 0 and 1.

$$FPI = 1 - \frac{c}{(c-1)}\left[1 - \sum_{j=1}^{n}\sum_{i=1}^{c}(\frac{m_{ij})^2}{n}\right], \tag{A19}$$

where *c* is the number of groups; *n* is the number of elements in data set; m_{ij} is the element of the fuzzy belongs to matrix *M*.

c) **Modified Partition Entropy** (MPE – McBratney and Moore, 1985; Fridgen et al., 2004 – Equation A20): estimates the difficulty level to organize *c* groups.

$$MPE = \frac{-\sum_{j=1}^{n}\sum_{i=1}^{c} m_{ij}\, log(m_{ij})/n}{log\, c}, \tag{A20}$$

where *c* is the number of groups; *n* is the number of elements in the data set; m_{ij} is the element of the fuzzy belongs to matrix *M*.

d) **Improved Cluster Validation Index** (ICVI – Gavioli et al., 2016 – Equation A21): is a composition of FPI, MPE, and VR% indices.

$$ICVI_i = \frac{1}{3} * \left(\frac{FPI_i}{Max\{FPI\}} + \frac{MPE_i}{Max\{MPE\}} + \left(1 - \frac{VR\%_i}{Max\{VR\%\}}\right)\right), \tag{A21}$$

where FPI_i is FPI value of the *i-th* variable selection method; MPE_i is the MPE value of the *i-th* variable selection method; $VR\%_i$ is the VR% value of the *i-th* variable selection method; *Max{Index_X}* represents the maximum value of the *Index_X* among the *n* variable selection methods.

e) **Analysis of Variance (ANOVA)**: Tukey test identified whether the sub-regions of design in MZs present significant differences on the average value of the target variable.

f) **Smoothness Index** (SI% – Gavioli et al., 2016 – Equation A22): gives pixel-by-pixel frequency of change of classes in a thematic map in horizontal and vertical directions and along the diagonal. It also characterizes the smoothness of MZs boundary curves. For example, if a map has an entirely homogeneous area, SI equals 100% due to the lack of class changes. On the other hand, if the map is entirely generated with random values, SI% would have a value close to 0.

$$SI = 100 - \left(\frac{\sum_{i=1}^{k} NM_{H_i}}{4P_H} + \frac{\sum_{j=1}^{k} NM_{V_j}}{4P_V} + \frac{\sum_{l=1}^{k} NM_{DD_l}}{4P_{DD}} + \frac{\sum_{m=1}^{k} NM_{DE_m}}{4P_{DE}} \right) * 100, \quad \text{(A22)}$$

where NM_{H_i} is the number of changes in row *i* (horizontal); NM_{V_j} is the number of changes in column *j* (vertical); NM_{DD_l} is the number of changes in diagonal *l* (right diagonal DD); NM_{DE_m} is the number of changes in diagonal *m* (left diagonal DE); k is the maximum number of pixels in a row, column, or diagonal; P_H is the possibility of changes in horizontal pixels; P_V is the possibility of changes in vertical pixels; P_{DD} is the possibility of changes in the right diagonal DD; P_{DE} is the possibility of changes in the left diagonal DE.

g) **Average Silhouette Coefficient** (ASC – Rousseeuw, 1987 – Equation A23): the ASC coefficient is obtained from the silhouette coefficient (SC), an evaluation index that measures both levels of satisfactory internal formation and external separation of groups. SC value for point *p*, which is denoted by sc_p, is calculated using the mean of intra-group distances *ap* and the mean of inter-group distances b_p:

$$sc_p = \frac{b_p - a_p}{Max(a_p, b_p)}, \quad \text{(A23)}$$

where a_p is the mean of distances among point *p* and all other points in the same group; b_p is the mean of distances among point *p* and all points in the closest group that contains *p*.

h) The **group silhouette coefficient** (GSC) is obtained by calculating the mean of SCs for the points of this group, and the value that corresponds to ASC coefficient of k groups is obtained by calculating the mean of GSC values of k groups. ASC values vary from -1 to 1; -1 indicates an incorrect clustering, whereas 1 indicates groups with the best intra-group formation and the best possible inter-group separation.

i) **Fragmentation index** (FI% – Equation A24): considers how higher is the number of zones (NMZ) in comparison with the number of classes (NC). The higher FI%, the higher is fragmentation.

$$FI\% = 100\frac{MZ - c}{c}, \tag{A24}$$

j) **Global Quality Index** (GQI – Beneduzzi, 2020 – Equation A25): aims at finding the best number of classes during ZMs delineation, considering the values of ICVI, SIr, and FIr:

$$GQI_i = \frac{ICVI_i * (100 + FIr_i)}{SIr_i}, \tag{A25}$$

k) **Modified Global Quality Index** (MGQI – Equation A26): this coefficient, proposed in this work, is an adaptation of GQI to include ASC coefficient.

$$MGQI_i = \frac{ICVI_i * (100 + FIr_i)}{SIr_i * ASC}, \tag{A26}$$

Indices for comparison among thematic maps and among management zones

a. **Coefficient of relative deviation** (CRD – Coelho et al., 2009 – Equation A27): calculates the mean difference in modulus of the interpolated values on a thematic map compared to a map taken as a reference.

$$CRD = \sum_{i=1}^{n} ABS(\frac{Zi_B - Zi_A}{Zi_A}), \tag{A27}$$

where Zi_A is the estimated value at the location i on the reference map, Zi_B is the value at location i on the map to be compared, and n is the total number of interpolated locations on the maps.

b. **Mean absolute difference** (MAD – Equation A28): computes the mean absolute difference among values on both maps.

$$MAD = \frac{\sum_{i=1}^{n} ABS(Zi_B - Zi_A)}{n}, \tag{A28}$$

where Zi_A is the value of location (pixel) i on the reference map, Zi_B is the value at location (pixel) i on the map to be compared, and n is the total number of observations on the maps.

c. **Kappa coefficient** (Kp – Cohen, 1960; Congalton, 1991 – Equation A29): measures the degree of agreement among MZ maps generated by the clustering algorithms. Landis and Koch (1977) proposed the following classification: 0 < Kp ≤ 0.2 indicates no agreement, 0.2 < Kp ≤ 0.4 weak agreement, 0.4 < Kp ≤ 0.6 moderate agreement, 0.6 < Kp ≤ 0.8 strong agreement, and 0.8 < Kp ≤ 1 very strong agreement.

$$K_p = \frac{\{n \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} (x_{i+} * x_{+i})\}}{\{n^2 - \sum_{i=1}^{r} (x_{i+} * x_{+i})\}}, \tag{A29}$$

where X_{ii} is the value in row i and column i, $X_{i+}$is the sum of line i, and X_{+i} is the sum of column i of the error matrix, N is the total number of points interpolated and sorted by the matrix, and c is the number of classes of the error matrix.

d. **Global accuracy** (GA – Foody, 2002 – Equation A30): like *Kp*, GA measures the degree of agreement among maps (MZs) and corresponds to the simple percent agreement.

$$GA = \frac{\sum_{i=1}^{c} x_{ii}}{n}, \tag{A30}$$

where, $\sum_{i=1}^{c} x_{ii}$ is the sum of the main diagonal of the error matrix with c classes and a total of N samples collected (number of points interpolated).

Appendix B - Definition of characteristics and sub-characteristics in each category of the quality model presented in the evaluation instrument

Category	Characteristic	Sub-characteristic
1 Functionality	1.1 Are the software's functionalities suitable for Precision Farming?	
	1.2 In general, what is proposed in each feature is done correctly?	
	1.3 Does the software generate maps properly?	
	1.4 Does the software generate Management Zones properly?	
	1.5 Prevent unauthorized access to software and data?	
	1.6 IMPORTANCE of each feature in ADB-Map	1.6.1 Importing data from a file (sample grids, contour, maps, ...)
		1.6.2 Delineate boundaries
		1.6.3 Select the best interpolation method (Ordinary Kriging versus Inverse Distance Weighting)
		1.6.4 Select interpolation parameter for each interpolation method
		1.6.5 Interpolate data by ordinary Kriging
		1.6.6 Interpolate data by Inverse Distance Weighting
		1.6.7 Interpolate data by moving average
		1.6.8 Interpolate data by nearest neighbor
		1.6.9 Delineate management zones
		1.6.10 Rectify management zones
		1.6.11 View statistical data
		1.6.12 Spatial correlation matrix
		1.6.13 Variable selection
		1.6.14 Data cleaning
		1.6.15 Grid cutting out
		1.6.16 Principal component analysis
		1.6.17 Quality evaluations indices of management zones (FPI, MPE, ICVI, SI)
		1.6.18 Map agreement index (GA and Kappa)

		1.6.19 Fast Track - Thematic map
		1.6.20 Fast Track - Management zones
		1.6.21 Nutrient recommendation (Nitrogen, Phosphorus, and Potassium)
2 Reliability	2.1 It does not frequently have software defects	
	2.2 When failures occur, it reacts well (it starts working again)	
	2.3 When failures occur, are data already created still available in ADB-Map?	
3 Usability	3.1 Is access to ADB-Map easy?	
	3.2 Is it easy to understand the logic of available resources and their applicability?	
	3.3 Is it easy to learn how to use software?	
	3.4 Are the layers categories adequate?	
	3.5 Are the font type, text size, page colors, and shadows adequate?	
	3.6 EASE of using the ADB-Map features	3.6.1 Importing data from a file (sample grids, contour, maps, ...)
		3.6.2 Delineate boundaries
		3.6.3 Select the best interpolation method (Ordinary Kriging versus Inverse Distance Weighting)
		3.6.4 Select interpolation parameter for each interpolation method
		3.6.5 Interpolate data by ordinary Kriging
		3.6.6 Interpolate data by Inverse Distance Weighting
		3.6.7 Interpolate data by moving average
		3.6.8 Interpolate data by nearest neighbor
		3.6.9 Delineate management zones
		3.6.10 Rectify management zones
		3.6.11 View statistical data
		3.6.12 Spatial correlation matrix
		3.6.13 Variable selection
		3.6.14 Data cleaning
		3.6.15 Grid cutting out
		3.6.16 Principal component analysis
		3.6.17 Quality evaluations indices of management zones (FPI, MPE, ICVI, SI)
		3.6.18 Map agreement index (GA and Kappa)
		3.6.19 Fast Track - Thematic Map
		3.6.20 Fast Track - Management zone
		3.6.21 Nutrient recommendation (Nitrogen, Phosphorus, and Potassium)
4 Efficiency	4.1 Time to run features	4.1.1 Data import
		4.1.2 Parameters selection for the interpolator
		4.1.3 Data Interpolation (Ordinary Kriging, inverse distance weighting, moving average, nearest neighbor)
		4.1.4 Delineate management zones
		4.1.5 Nutrient recommendation (Nitrogen, Phosphorus, and Potassium)
		4.1.6 Fast Track - Thematic Map
		4.1.7 Fast Track - Management zone

5 Maintainability	5.1 Are software failure(s) resolved quickly?	
6 Portability	6.1 ADB-Map execution quality in different browsers	6.1.1 Google Chrome
		6.1.2 Mozilla Firefox
		6.1.3 Microsoft Edge
		6.1.4 Safari
		6.1.5 Opera
		6.1.6 Internet Explorer
		6.1.7 Others
	6.2 ADB-Map execution quality on different operating systems	6.2.1 Windows
		6.2.2 Linux
		6.2.3 iOS
		6.2.5 Android
	6.3 Equipment that ran the ADB-Map	6.3.1 Computer/Notebook
		6.3.2 Smartphone/Tablet

7 PAPER 3 – PROCESS IMPROVEMENT OF SELECTING THE BEST INTERPOLATOR AND ITS PARAMETERS TO CREATE THEMATIC MAPS

ABSTRACT: Thematic maps (MTs) are essential tools to demonstrate the information of spatially distributed phenomena. In precision agriculture, they have the critical role of demonstrating the existing variability in factors that influence the crop's yield. This yield can be mapped, studied, and used in decision-making. A TM can be created from sampling data, a standard procedure for soil attributes, since the aggregate cost in the laboratory analysis makes it impossible to perform a meter-by-meter sampling. Statistical interpolation methods are used to estimate data in unknown locations, , such as inverse distance weighting (IDW) and ordinary Kriging (OK). For both interpolators, it is essential to use the appropriate parameters to estimate values in non-sampled locations, either the exponent value and number of neighbors for IDW or the theoretical model adjusted to the experimental semivariogram for OK. Thus, this trial aims at adopting additional criteria in selecting interpolators and evaluating their performance. The selection criteria were (i) effective spatial dependence index (%ESDI) > 25%, (ii) the first semivariance significance index ($\%\gamma(1)$) < 50% and (iii) slope of the model ends index (%SMEI) > 20%, which were applied according to three methods: 1) only with the interpolator selection index (ISI) without application of the proposed criteria; 2) the criteria applied after the interpolator selection analysis + ISI, and 3) the criteria applied during the interpolator selection analysis + ISI. The experimental data come from an experiment in two agricultural areas in Serranópolis do Iguaçu–PR, Brazil, using grids with good sampling density (2.7, 2.6, and 3.5 points per ha). It was observed that, usually, the three methods selected different models and that Method 3 was considered the best one. It is essential to highlight that the three criteria must be considered altogether in the semivariogram models' selection process. The coefficient of relative deviation (CRD) varied from 0.1 to 64% when comparing the maps generated by the three methods.

KEYWORDS: Kriging, Inverse distance weighting, precision agriculture.

7.1 Introduction

Maps representing a field and a topic associated with it are called thematic maps (TMs) and aim to inform, by graphical symbols, where a specific geographical phenomenon occurs. TMs have become an essential tool in geospatial science to understand spatial information (Fraser and Congalton, 2019), e.g., digital elevation model, slope map, soil map, aspect map, land use/land cover map, and contour map (Gojiya et al., 2018).

In agriculture, soil is the primary source of nutrients. Crop development is directly affected by nutrients availability in soil (Coutinho et al., 2019). Therefore, studying its properties and spatial variability patterns are expected to manipulate crop development to our ends (Mcbratney and Pringle, 2006). Nutrients, classified as macro and micronutrients, play an essential role in energy storage, electrode transport, plant’s enzyme activity and cannot be replaced by others (Mikula et al., 2020). Thus, fertilizer application in variable rate, according to the crop's spatial variability demand, has been widely discussed in Precision Agriculture context (Coutinho et al., 2019). However, crop yield is affected by other factors. Zhang et al. (2002) categorize these factors into six groups: (i) yield variability, (ii) field variability, (iii) soil variability, (iv) crop variability, (v) variability regarding anomalous factors, and (vi) management variability.

In Precision Agriculture, TM is an essential tool to assist analysts in decision making, as it allows to identify spatial variability within the field and manage the area in a localized way. TMs development is associated to data collection, analysis, interpretation, and information representation on a map, facilitating identifying similarities and enabling spatial correlations visualization. One specific case of TMs is contour maps built by connecting points of the same value and applying them to geographical phenomena that show continuity in a geographic space. Another is choropleth maps that use color to show ranges of values of a specific variable within a defined geographic area. Contour and choropleth maps can be built from categorical data (yield, elevation, temperature, precipitation, humidity, and atmospheric pressure) or relative data (density, percentages, and indexes) (Souza et al., 2021). Usually, both maps are called contour maps.

The advancement of computational technologies allows TMs' creation and analysis using different techniques, methodologies, and software. For example, the geographical information systems (GISs) can store, exhibit, recover and dissect spatial data in a friendly approach. GIS has been widely used in many studies for spatial and temporal data creation (Gojiya et al., 2018). In this sense, AgDataBox (ADB, http://adb.md.utfpr.edu.br; Michelon et al., 2019, Dall'agnol et al., 2020) web platform provides tools to create, store, recover, manage, exhibit, and analyze geographic and spatial data of TMs focused on agriculture.

Usually, the sampled data are interpolated in a dense and regular grid to generate continuous and smooth TMs. This task is carried out with the aid of interpolation methods. The most used methods in precision agriculture (PA) are inverse distance weighted interpolation (IDW) and ordinary Kriging (OK – Cressie, 1993), which are differentiated by how weights are attributed to different samples, which may influence the estimated values (Reza et al., 2010). IDW procedure has been used because it is quick and straightforward; Kriging has been used because it provides the best linear unbiased estimates. However, it is more complex and time-consuming (Mueller et al., 2004). IDW interpolator considers weights at the sample points, which are evaluated during the interpolation process. Each sampled point's influence is inversely proportional to the distance from the point to be estimated increased to a power (Isaaks and Srivastava, 1989). The value of the chosen power predetermines the weight factor; that is, the higher this value, the lower the most distant points' influence. IDW is a fast method and requires little computational cost (Mazzini and Schettini, 2009).

Kriging makes the estimation based on a continuous model of stochastic spatial variation. It makes the best use of existing knowledge by considering how a property varies in space by the variogram model (Oliver and Webster, 2015). In Kriging, weights are determined by statistical dependence (i.e., covariances) among sampled locations, yet respect the measurement uncertainty. In general, the greater the covariability, the greater the weight (Wikle et al., 2019).

Kriging has been identified as a BLUE interpolator: Best Linear Unbiased Estimator (Diggle and Ribeiro, 2007; Vieira, 2000; Isaaks and Srivastava, 1989). However, it must meet the spatial dependence (SD) modeling requests (Oliver and Webster, 2015; Cambardella et al., 1994) to have the correct performance and adequate use in creating a TM. The decision to use kriging depends on several factors, which influence the choice of a particular interpolation technique (Eldeiry and Garcia, 2012). The procedure's performance can be influenced by variability and spatial structure of data, semivariogram model, search radius, and the used number of the closest neighboring points (Reza et al., 2010; Isaaks and Srivastava, 1989). Therefore, interpolations' quality depends on variable's spatial structure under study (Amaral and Justina, 2019). The deterministic interpolator IDW does not consider SD and specific behavior of data, leading to less efficiency in mapping the spatial distribution of a given variable than Kriging (stochastic interpolator) (Betzek et al., 2019). However, when there is no SD (Rodrigues et al., 2018; Cambardella et al., 1994), the use of a deterministic interpolator can be more appropriate.

It is necessary to assess its accuracy to properly use a TM as a tool to support a decision-making. Precision assessment methods are often based on analyzing values' estimation error, comparing predicted values to known values (Fraser and Congalton, 2019). The best adjustment model for semivariogram and parameter estimation can be evaluated by examining the distribution of errors or residuals (Betzek et al., 2019) using the cross-validation technique. The cross-validation technique enables a comparison among estimated and sampled values using only the available information in data (Isaaks and Srivastava, 1989). Commonly used error measures include mean error, mean absolute error, mean squared error, and root mean squared error (Wackernargel, 2003, Carroll and Cressie, 1996).

The error comparison index (ECI – Souza et al., 2016) selects the best model from a set of semivariograms. Like ECI, the interpolator selection index (ISI – Bier and Souza, 2017) compares deterministic and stochastic interpolation methods. ISI enables selecting the best among several existing mathematical and geostatistical models in a simplified and less subjective manner, implemented in computerized systems (Bier and Souza, 2017). Betzek et al. (2019) developed computational routines in geoR to determine the best semivariogram model (and its parameters) based on ECI and ISI, developed in statistical software R, using the geoR library and functions implemented directly in the PostgreSQL database were developed by PL/pgSQL procedural. These computational routines were reimplemented, optimized, and made available on ADB platform in a microservice form and a graphical user interface (GUI) in ADB-Map application, which allows selecting the best interpolator and its parameters for a data set. The routines determine the best semivariogram model (and its parameters) for OK and the best power and number of neighbors used in IDW interpolator. In the geostatistics module, seven semivariogram models are tested (spherical, gaussian, exponential, Matérn 0.5, Matérn 1.0, Matérn 1.5, and Matérn 2.0), as well as two statistical methods to optimize the semivariogram adjustment, ordinary least squares (OLS) and weighted least squares (WLS – Cressie, 1985), thus totalizing fourteen different models. For each model, twenty-five different parameter sets (five initial values for the partial sill parameter and five for range) are used, totalizing 350 different adjustments being analyzed to find the best one.

There is no optimal number of neighbors and exponent to be used in IDW interpolation. Thus, this parameter should always be individually evaluated and optimized for each dataset before final interpolation (Amaral and Justina, 2019). Computational routines implemented in ADB-Map allow analyzing a range of values for the exponent (0.5, 1.0, ..., n) and a range of values for the number of neighbors (4, 5, ..., n). As in selecting the best semivariogram model, ISI is used to identify the best value for the exponent and number of neighbors.

In geostatistics, variograms (semivariograms) are not only used as an exploratory tool but allow estimating parameters (Diggle and Ribeiro Jr., 2007). After the experimental semivariogram construction, it is necessary to adjust a theoretical model representing data variability. The curve-fitting can be done “by eye” by trying different values for the model parameters and visually inspecting the fit to the sample variogram (Diggle and Ribeiro Jr., 2007). However, parametric covariance functions can be used to estimate semivariogram parameters. As a result, the variogram parameter estimates minimize the theoretical model's squared differences and experimental variogram (Li et al., 2018).

The SD degree among the variables can be verified by the spatial dependence index (%SDI - Biondi et al., 1994) associated with semivariograms; %SDI represents the percentage of the spatially correlated variance (the partial sill) to the total variation (the nugget effect + the partial sill). The %SDI classification, adapted from Cambardella et al. (1994), is considered: %SDI ≤ 25% - weak SD; 25% < %SDI ≤ 75% - moderate SD; and %SDI > 75% - strong SD. When SD is weak, the use of a deterministic interpolator can be more appropriate.

It was observed that, in some data sets, the routine implemented by Betzek et al. (2019) to select an interpolator, may mistakenly select a geostatistical model that does not have spatial dependence or consider a model with a lack of adjustment to the experimental semivariogram.

Therefore, this work aimed to adopt criteria to guarantee a minimum spatial dependence in the semivariograms applied to the interpolators' selection process. For that, the indices were proposed (i) effective spatial dependence index (%ESDI), (ii) the first semivariance significance index ($\%\gamma(1)$), and (iii) slope of the model ends index (%SMEI).

7.2 Material and methods

ADB-Map (http://adb.md.utfpr.edu.br/map; Michelon et al., 2019, Dall'agnol et al., 2020) application, which is included in ADB platform, was employed for (i) descriptive and exploratory analyses, (ii) data interpolation, (iii) selection of the best interpolation method, and (iv) TMs creation.

7. 2. 1 Thematic maps

It is necessary to construct TMs about attributes collected in agriculture fields to follow a protocol like the one presented in Fig. 1 (Souza et al., 2018):

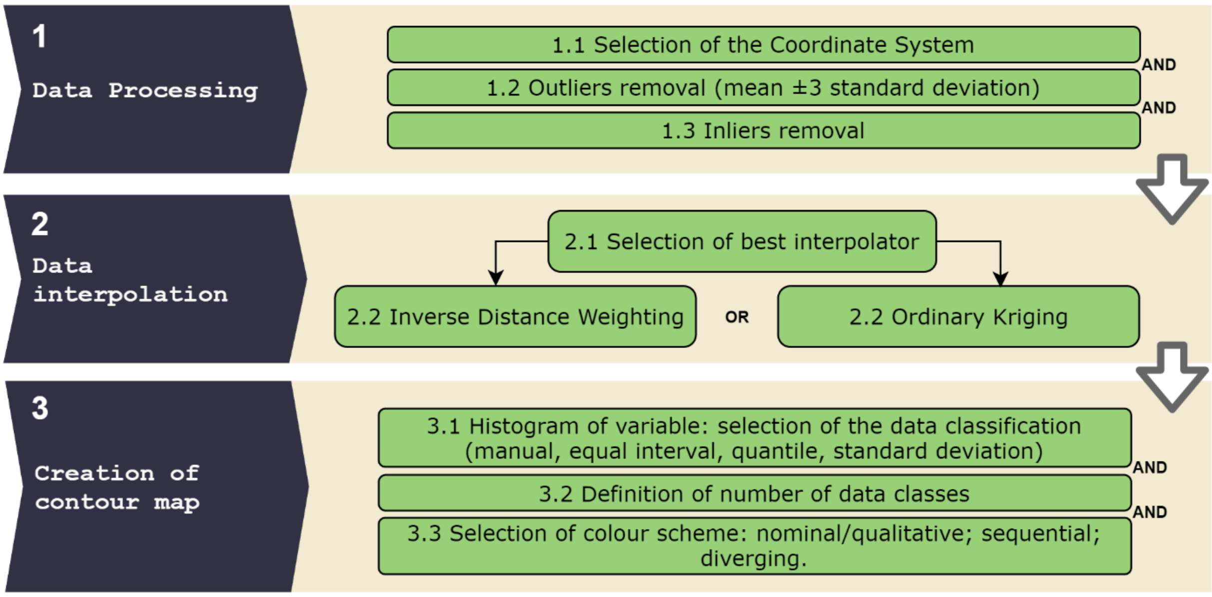


Fig. 1 Flowchart of the typical protocol to create a thematic map.
Source: Souza et al. (2018).

i. Location of the field, data collection, and selection of the coordinate system

Physical and chemical soil attributes were collected based on irregular sampling grids in two agricultural fields located in the municipality of Serranópolis do Iguaçu, western Paraná state, southern Brazil. The fields have been cultivated under a no-tillage system with a crop succession of soybean and corn. The coordinate systems were the geographic coordinate system (GCS) with WGS 1984 datum. The sampling points' locations were obtained by a GNSS receiver (Juno SB Trimble Navigation Limited, Westminster, CO, USA).

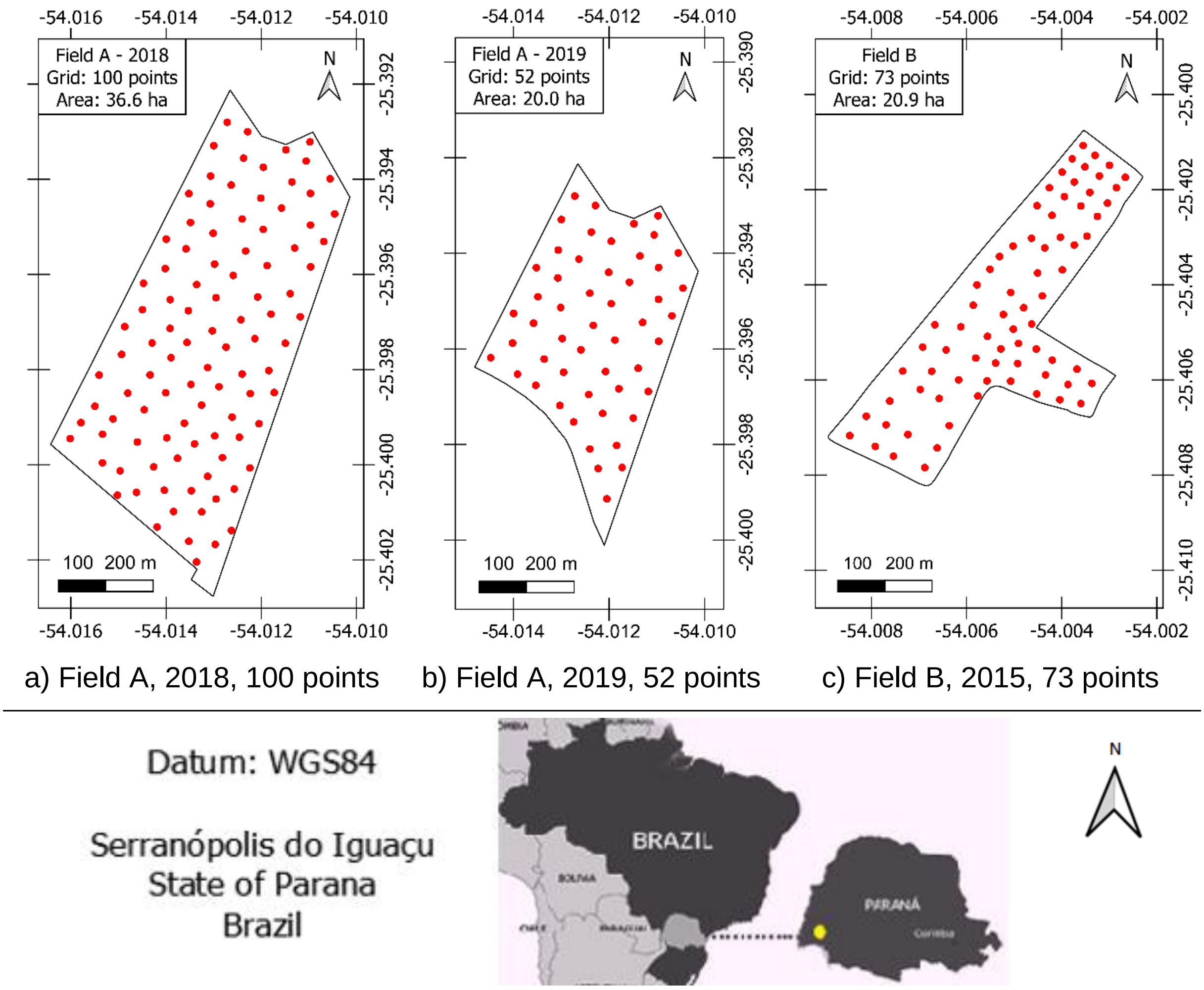


Fig. 2 Location of experimental fields and sampling grids of 100 points (Field A-2018; 36.6 ha), 52 points (Field A-2019; 20.0 ha), and 73 points (Field B-2015; 20.9 ha) in the municipality of Serranópolis do Iguaçu, Paraná state, Southern Brazil. Black contour delineates the fields used. Coordinates are in degrees (WGS 1984). The minimum and maximum distances among the sampling points are 41 and 1027 m in field A-2018, 45 and 706 m in field A-2019, and 31 and 838 m in Field B-2015.

Usually, soil samples are analyzed to determine the soil nutrient levels. Therefore, sampling must be dense enough to determine soil nutrients' variability so that fertilizers may be used in a profitable and environmentally sustainable way (Ferguson and Hergert, 2009; Franzen et al., 2002). Soil samples were taken from 0 to 0.20 m depth and analyzed in a commercial laboratory. Around each sampling point (using a GNSS Juno SB Trimble Navigation Limited, Westminster, CO, USA) and using a 3-m radius, eight subsamples were randomly collected, two per quadrant, within a symmetrical circle divided into four quadrants. Field A (Fig.s 2a and 2b) was sampled with 100 sampling points in 2018 (36.6 ha) and 52 in 2019 (20.0 ha) and Field B (Fig. 2c) was sampled with 73 sampling points (20.9 ha). The minimum and maximum distances among the sampling points are 41 and 1027 m in field A-2018, 45 and 706 m in field A-2019, and 31 and 838 m in Field B-2015. Thus, the sampled density corresponds, respectively, to 2.7, 2.6, and 3.5 points ha^{-1} (Table 1), which were considered enough to identify spatial variabilities of the variables of these fields given that they exceed the recommended minimum density of 1 sample ha^{-1} (Ferguson and Hergert, 2009) to 2.5 samples ha^{-1} (Doerge, 2000; Journel and Huijbregts, 1978). However, Oliver and Webster (2015) observed that at least between 100-150 samples are required for a reliable variogram, but Clark (1979) recommended at least 30-50 data points to use Kriging. Nevertheless, the threshold for a sufficient density in one case may not enough in another. Therefore, we decided to keep this sample density because we wanted to confirm the robustness of ADB's automated procedure and determine whether it can help be employed to determine when to use IDW and when to use OK (i.e., to determine whether the sample density is enough and/or if SD is detected; the pure nugget effect characterizes this case).

Table 1 Details of the study fields

Fields	**Areas (ha)**	**Geographical center coordinates (WGS84)**	**Elevation (m)**	**Sample points**	**Points (ha^{-1})**
A-2018	36.6	25º23′48″S 54º0′46″W	345	100	2.7
A-2019	20.0	25º23′43″S 54º0′44″W	334	52	2.6
B	20.9	25°24′28″S 54°00′17″W	355	73	3.5

Each point sample was composed of eight individual samples (Wollenhaupt et al., 1994). The sampling points were located along an imaginary line among intermediate contour lines with alternated distances provided a better fit at the smallest lag distances, essential in Kriging (Bier and Souza, 2017). The variables obtained from soil analysis were chemical attributes (organic matter (OM; g dm^{-3}), zinc (Zn; mg dm^{-3}), iron (Fe; mg dm^{-3}), manganese (Mn; mg dm^{-3}), phosphorus (P; mg dm^{-3}), potassium (K; $cmol_c$ dm^{-3}), copper (Cu; mg dm^{-3}), the potential of hydrogen (pH), calcium (Ca; $cmol_c$ dm^{-3}), magnesium (Mg; $cmol_c$ dm^{-3}), aluminum (Al; $cmol_c$ dm^{-3}), pH of buffer solution Shoemaker-McLean-Pratt (SMP) method, potential acidity (H+Al; $cmol_c$ dm^{-3}), the sum of bases (SB; $cmol_c$ dm^{-3}), base saturation (V%), aluminum saturation (m%), cation exchange capacity (CTC; $cmol_c$ dm^{-3}), the relationship between the content K/CTC (%), Mg/CTC (%), Ca/CTC (%) and H+Al/CTC (%)), and physical attributes (clay (%), sand (%), and silt (%)).

ii. Exploratory data analysis

Data were analyzed using descriptive and exploratory statistics and geostatistics. During the descriptive analysis of data, measures of central tendency (mean and median), of dispersion (standard deviation (SD) and coefficient of variation (CV)), and normality tests (Kolmogorov-Smirnov and Anderson-Darling tests at 0.05 significance level) were calculated.

Data were considered normal when, in at least one of the tests, they presented normality. The coefficient of variation (CV) was classified as low when CV ≤ 10%, medium when 10% < CV ≤ 20%, high when 20% < CV ≤ 30%, and very high when CV > 30% (Pimentel-Gomes, 2009). The exploratory data analysis (EDA) was used to detect and remove outliers and inliers. Using the module ADB-Map-Clean of platform AgDataBox, duplicate, negative or null points, outliers, and inliers were removed. The outliers were identified as values outside the mean ± 3 SD (Córdoba et al., 2016). The inliers were obtained by Moran's local spatial autocorrelation index (II) (Anselin, 1995).

iii. Analysis of spatial dependence

The semivariogram chart (Fig. 3) is determined from a set of observed values in two stages (Oliver and Webster, 2015) (i) the calculation of the empirical semivariogram that summarizes spatial relations in data, and (ii) the adjustment of a mathematical model that best represents semivariances' distribution in each lag distance. Each calculated semivariance for a particular lag (h) is only an estimate of a mean semivariance $\hat{\gamma}(h)$ for that lag. The four main elements are (i) nugget effect (C_0), (ii) partial sill (C_1), (iii) sill (C_0 + C_1), and (iv) the range of spatial autocorrelation (Ra).

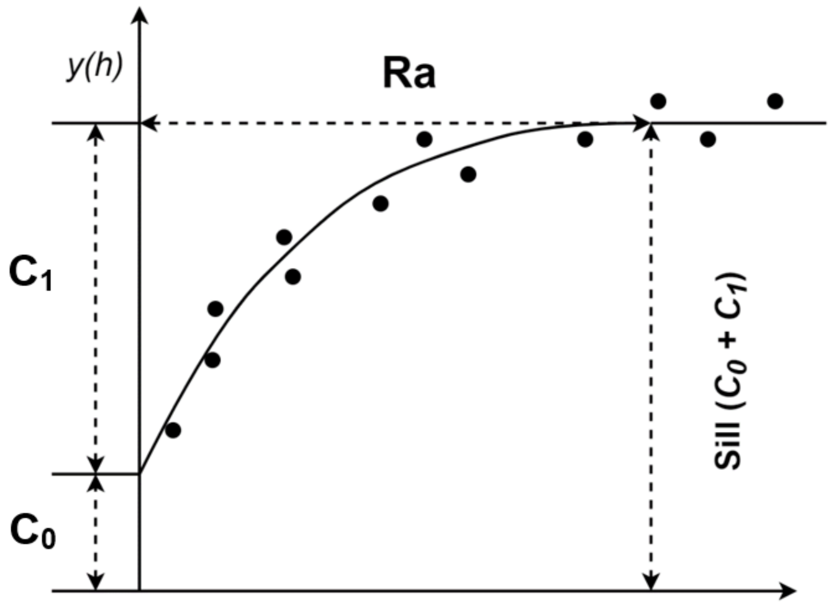


Fig. 3 Semivariogram chart with the four main elements: nugget effect (C_0), partial sill (C_1), sill (C_0 + C_1), and the range of spatial autocorrelation (Ra).

The nugget effect (C_0) is the semivariance value for zero distance (Webster, 1985) and represents the component of random variation, i.e., variability for scales is smaller than the distance between sample points. According to Cressie (1993), C_0 parameter represents small-scale local variations, such as measurement errors, and corresponds to where the semivariogram touches the ordinate axis. This point reveals semivariogram's discontinuity for distances shorter than the shortest distance between the sample points. The partial sill (C_1), also known as the dispersion variance, represents the spatial differences among C_0 values and sill, interval in which semivariogram develops and corresponds to SD (Cressie, 1993). The range (Ra) is the distance where variogram reaches sill, and from this distance, samples are not correlated (Oliver and Webster, 2015).

Semivariances are calculated from an estimator, as the classic proposed by Matheron (1963) (Equation 1). Unfortunately, outliers affect this estimator significantly, and even a single discrepant datum can distort the final variogram estimates. The alternative is to use one of the robust estimators, such as those of Cressie and Hawkins (1980), Dowd (1984), and Genton (1998) (Oliver and Webster, 2015).

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2, \tag{1}$$

where $\hat{\gamma}(h)$ is the value of the semivariance estimate; $Z(x_i)$ is the value of variable Z at point x_i; $Z(x_i + h)$ is the value of variable Z at point $x_i + h$; $N(h)$ is the number of pairs separated by a determined distance h (URIBE-OPAZO et al. 2012).

The mathematical model adjustment should describe the spatial variation to estimate or predict values at unsampled places optimally by Kriging (Oliver and Webster, 2015). Only certain mathematical functions are suitable for this purpose, so, choosing and fitting a model must be done with care (Lark, 2000). The theoretical models spherical, exponential, gaussian, and Matérn’s family are commonly used (Uribe-Opazo et al., 2012; Isaaks and Srivastava, 1989).

Semivariance tends to increase with the distance among sampled locations, or lag distance (h), to a more or less constant value (the sill or total semivariance) at a given separation distance, called SD range. Thus, samples separated by distances closer than the range are related spatially, and those ones separated by distances larger than the range are not spatially related (Webster, 1985).

The Matheron (1963) classic estimator was used to calculate semivariances with at least 30 pairs of points (Journel and Huijbregts, 1978), and the range Ra was limited to half of the maximum distance (MD) among points (cutoff = 0.5*MD). The semivariances' calculation should not exceed distances among points greater than half of the maximum distance (Clark, 1979). Points located beyond cutoff are considered non-influential (Isaaks and Srivastava, 1989). Lag size (h) was defined by calculating the number of lags, relationship between cutoff and the shortest distance among the pairs of points. Therefore, the lag h sizes were 43 m (Field A-2018), 44 m (Field A-2019), and 30 m (Field B-2015), while semivariances 102 and 438 in area A-2018, 53 and 180 in area A-2019, and 55 and 182 in area B-2015. A significant limitation to address in this ADB-Map version is that anisotropy's eventual presence is not considered.

To evaluate the degree of variable SD, we used the spatial dependence index (%SDI - Biondi et al., 1994 – Equation 2). The adopted %SDI classification (Konopatzki et al., 2012) was: very low for %SDI < 20%; low for 20 ≤ %SDI < 40%; medium for 40 ≤ %SDI < 60%; high for 60 ≤ %SDI < 80%; and very high for %SDI > 80%. This classification has the advantage of having five interpretation levels instead of three proposed by Cambardela et al. (1994) and is proportional to the spatial variability (the higher %SDI, the higher SD).

$$\%SDI = \frac{C_1}{C_0 + C_1} * 100 = \frac{C_1}{C} * 100, \tag{2}$$

where C_0 is the nugget effect, C_1 is the partial sill, and C is the sill.

Fig. 4 shows hypothetical sample points for which the spherical model was adjusted by routine in R. Considering that C_0 is 1 and C_1 is 9, the associated %SDI is 90%, corresponding to a strong SD. However, all semivariances are in the interval from 7 to 10. In this sense, this works presents a new index, the effective spatial dependence index (%ESDI – Equation 3), a new measure of SD degree. This index considers semivariance ($\gamma(1)$) in the first lag distance (h(1)).

$$\%ESDI = \frac{C - \gamma(1)}{C} * 100, \tag{3}$$

where C is the sill (nugget effect + partial sill) and $\gamma(1)$ is the first semivariance of the semivariogram. The %ESDI was classified as %SDI.

The second proposed index was the first semivariance significance index ($\%\gamma(1)$, Equation 4), SD fraction due only to ($\%\gamma(1)$).

$$\%\gamma(1) = \frac{\gamma(1) - C_0}{C_1} * 100, \tag{4}$$

where C_0 is the nugget effect, C_1 is the partial sill, and $\gamma(1)$ is the first semivariance of the semivariogram.

Furthermore, we also propose a slope of the model ends index (%SMEI, Equation 5), which aims to assess the inclination degree between the nugget effect and the last adjusted semivariance. When %SMEI is null, it is a pure nugget effect, characterizing a lack of SD.

$$\%SMEI = \left(1 - \frac{\gamma_Z(0)}{10^{-10} + \gamma_Z(n)}\right) * 100 = \left(1 - \frac{C_0}{10^{-10} + \gamma_Z(n)}\right) * 100, \tag{5}$$

where γ_Z is the adjusted theoric semivariance, $\gamma_Z(0) = C_0$ is the nugget effect, and $\gamma_Z(n)$ is the last adjusted theoretic semivariance, correspondent to the cutoff. The arbitrary constant 10^{-10} was included to avoid division by zero.

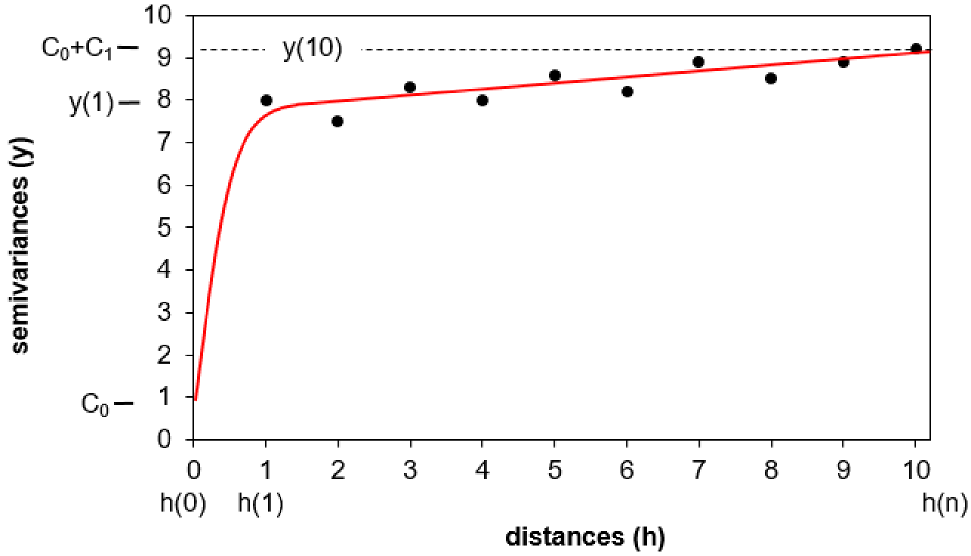


Fig. 4 Example of semivariogram chart adjusted with spherical semivariogram model, where $\gamma(1)$ is the first semivariance, γ_Z is the adjusted theoric semivariance, $\gamma_Z(0) = C_0$ is the nugget effect, and $\gamma_Z(n)$ is the last adjusted theoretic semivariance.

iv. Data Interpolation

The variables used to generate TM were interpolated using OK and IDW in a 9x9-m grid with pixels. ADB-Map application automatically sets the pixel size based on the area's size, with the value of 1 hundredth of the longest distance (horizontal or vertical). Computational routines were implemented in R language in ADB-Map application (Betzek et al., 2019).

a) Inverse distance weighting

IDW deterministic estimator (Equation 6) considers the closest points to the location to be estimated more representative than the most distant one according to the samples' linear distances. Twelve different values were used as IDW exponents (p) (0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, and 6.0).

$$\hat{Z}_i = \frac{\sum_{i=1}^{n}\left(\frac{1}{d_i^p} * Z_i\right)}{\sum_{i=1}^{n}\left(\frac{1}{d_i^p}\right)}, \quad (6)$$

where, $\hat{Z}_i$ – interpolated value; Z_i – sampled attribute value; d_i^p – Euclidean distance between the ith neighborhood point and the sampled point, elevated to the power of p > 0.

b) Ordinary Kriging

Variables' semivariograms were adjusted using theoretical models (spherical, gaussian, exponential, Matérn 0.5, Matérn 1.0, Matérn 1.5, and Matérn 2.0) by OLS and WLS methods. WLS weights were considered using the same number of pairs in each bin. Twenty-five different parameter sets (five initial values for the partial sill parameter and five for range) were used for each model, totalizing 350 adjustments.

c) Determination of the best semivariogram model and its parameters

Bier and Souza (2017) proposed the interpolation selection index (ISI – Equation 7) to automatize selecting the best interpolation method, which assumes a lower value as better the interpolator is. By cross-validation (Faraco et al., 2008; Isaaks and Srivastava, 1989), mean error (ME – Equation 8) and standard deviation of mean error (SDME – Equation 9) are calculated. ME and SDME values calculated for each parameter set are stored and used to determine ISI that compares the deterministic and stochastic interpolation methods, thus, identifying the best adjustment for each model analyzed.

$$ISI = \left\{\frac{abs(ME)}{max\Big|_{i\,=\,1}^{\,j}[abs(ME)]} + \frac{\left[SDME - min\Big|_{i\,=\,1}^{\,j} SDME\right]}{max\Big|_{i\,=\,1}^{\,j}[abs(SDME)]}\right\}, \tag{7}$$

where ME is the mean error; $SDME$ is the standard deviation of mean error of the crossed validation; n is the data number; abs is the module value; $min|_{i=1}^{j}$ is the lowest value among the compared j models; $max|_{i=1}^{j}$ is the highest value among the compared j models.

$$ME = \frac{1}{n}\sum_{i=1}^{n} Z(s_i) - \hat{Z}(s_i), \tag{8}$$

$$SDME = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Z(s_i) - \hat{Z}(s_i)\right)^2}, \tag{9}$$

where n is the data number; $Z(s_i)$ is the value observed at the point s_i; $\hat{Z}(s_i)$ is the predicted value at point s_i.

Statistic called error comparison index (ECI – Souza et al., 2016 – Equation 10) was used to determine the best semivariogram fit in each j model analyzed, which assumes a lower value for the model is better stochastic methods of interpolation. The best semivariogram of each j model was used in ISI analysis.

$$ECI_i = \frac{|RME_i|}{10^{-10} + max\Big|_{i\,=\,1}^{\,j}|RME|} + \frac{|SDRME_i - 1|}{10^{-10} + max\Big|_{i\,=\,1}^{\,j}|SDRME - 1|}, \tag{10}$$

where ECI_i is the error comparison index for model i; and $max\Big|_{i\,=\,1}^{\,j}$ is the highest value among the compared j semivariograms. The arbitrary constant 10^{-10} was included to avoid division by zero.

The reduced mean error (RME – Equation 11) and the standard deviation of the reduced mean error (SDRME – Equation 12) was determined by ordinary kriging cross-validation.

$$RME = \frac{1}{n}\sum_{i=1}^{n}\frac{Z(s_i) - \hat{Z}(s_i)}{\hat{\sigma}(\hat{Z}(s_i))}, \tag{11}$$

$$SDRME = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{\left|Z(s_i) - \hat{Z}(s_i)\right|}{\hat{\sigma}(\hat{Z}(s_i))}}, \tag{12}$$

where $Z(s_i) - \hat{Z}(s_i)$ is the prediction error associated with estimating yield at spatial location s_i; $Z(s_i)$ is the observed value; $\hat{Z}(s_i)$ is the estimated value obtained from the ordinary kriging cross-validation; $\hat{\sigma}(\hat{Z}(s_i))$ is the estimated standard deviation associated with the estimated value, and n is the sample size.

d) Improving models' selection using effective spatial dependence (%ESD)

Three problems should be addressed when selecting the best semivariogram:

1. A minimum of %ESD should be observed. We proposed that %ESDI must be greater than 25%.
2. The selected semivariogram model should contemplate a fraction of SD due only to (%γ(1)) lower than 50%.
3. The inclination degree of between the nugget effect and the last adjusted semivariance, estimated by %SMEI, should be greater than 20%. Otherwise, there is an indication of a pure nugget effect.

We proposed that the selection of the best interpolator model should not depend only on ISI but on the criteria presented on Table 2.

Table 2 Criteria to select the best interpolation method

Criterion 1 Minimum of effective spatial dependence		**Criterion 2** Spatial dependence due only to the first semivariance		**Criterion 3** The model needs to express spatial dependence	**The best interpolation method**
If $\%ESDI$ > 25%	and	If $\%\gamma(1)$ < 50%	and	If %SMEI > 20%	IDW or OK with the lowest ISI
If $\%ESDI$ ≤ 25%	or	If $\%\gamma(1)$ ≥ 50%	or	If %SMEI ≤ 20%	IDW with the lowest ISI

%ESDI: Effective spatial dependence index; $\%\gamma(1)$: First semivariance significance index; IDW: Inverse distance weighting; OK: Ordinary Kriging; ISI: Interpolator selection index.

The variable selection process was tested using three methods (Table 3): (i) method 1: best ISI, (ii) method 2 (Fig. 5): the three criteria (Table 2) are applied after geostatistics analysis, (iii) method 3 (Fig. 6): The three criteria are applied during geostatistics analysis.

Table 3 Methods used to select the best interpolation model

Methods	Selection of the best interpolation model
Method 1	Best ISI
Method 2	The three criteria are applied after geostatistics analysis + the best ISI
Method 3	The three criteria are applied during geostatistics analysis + the best ISI

The main difference between methods 2 and 3 is observed when the three criteria are applied. In method 2, the three criteria are applied to analyze geostatistical models after the ISI determination step and the best interpolator's indication (Fig. 5). For each semivariogram model and estimation method (Spherical OLS, Spherical WLS, Exponential OLS, Exponential WLS, etc.), all analyses to estimate semivariogram parameters are considered (5 partial sill intervals * 5 range intervals = 25 analysis). In method 3 (Fig. 6), a modification was proposed to filter out unsatisfactory geostatistical models before ECI has determined a semivariogram model's best fit. Therefore, when selecting the analyses by ECI, only the cleaned models (not discarded) by the new selection criteria are considered.

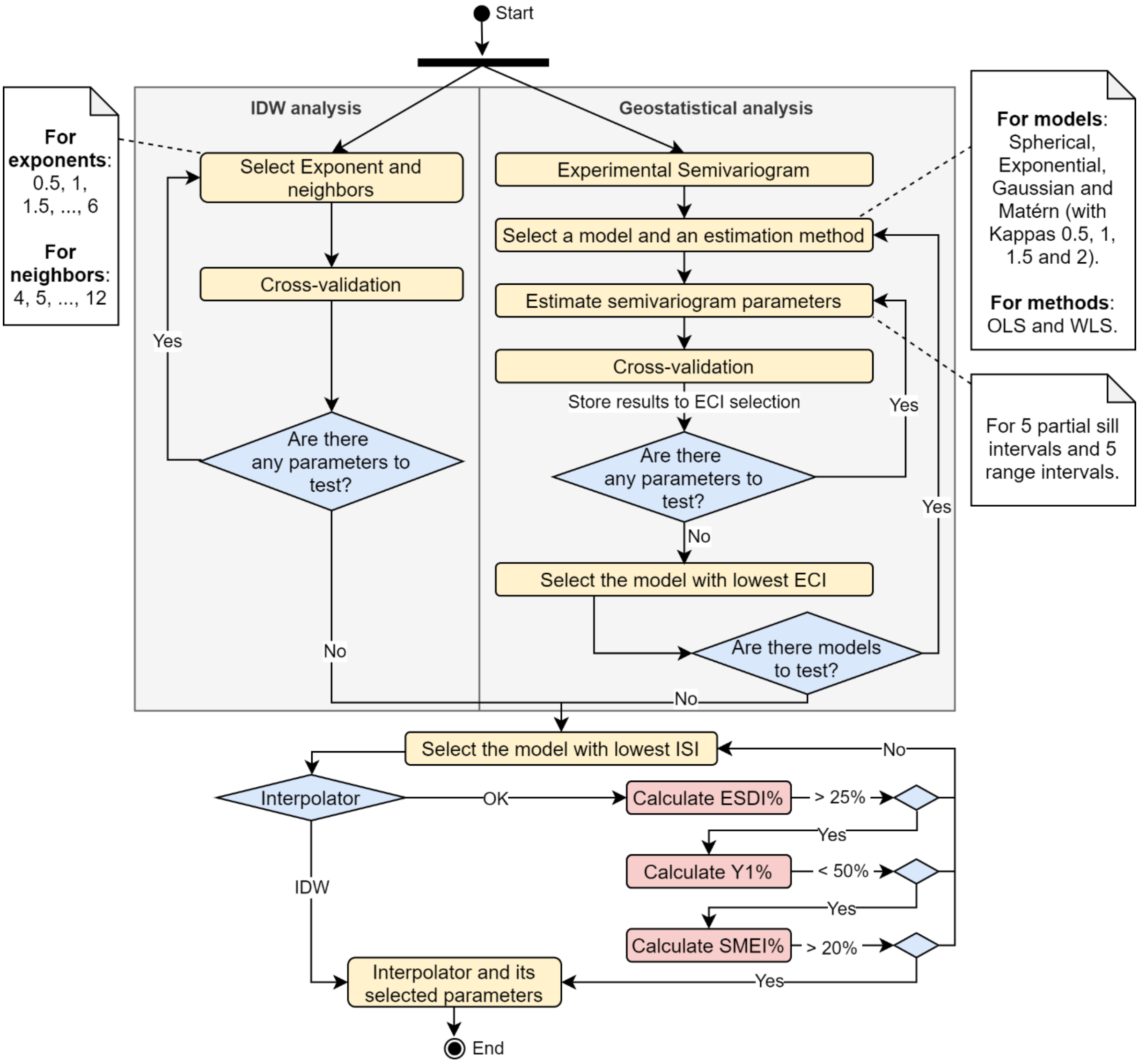


Fig. 5 Selection process of the best interpolator between inverse distance weighting and ordinary Kriging by method 2: the filters using %ESDI, $\%\gamma(1)$, and %SMEI were applied after geostatistics analysis.

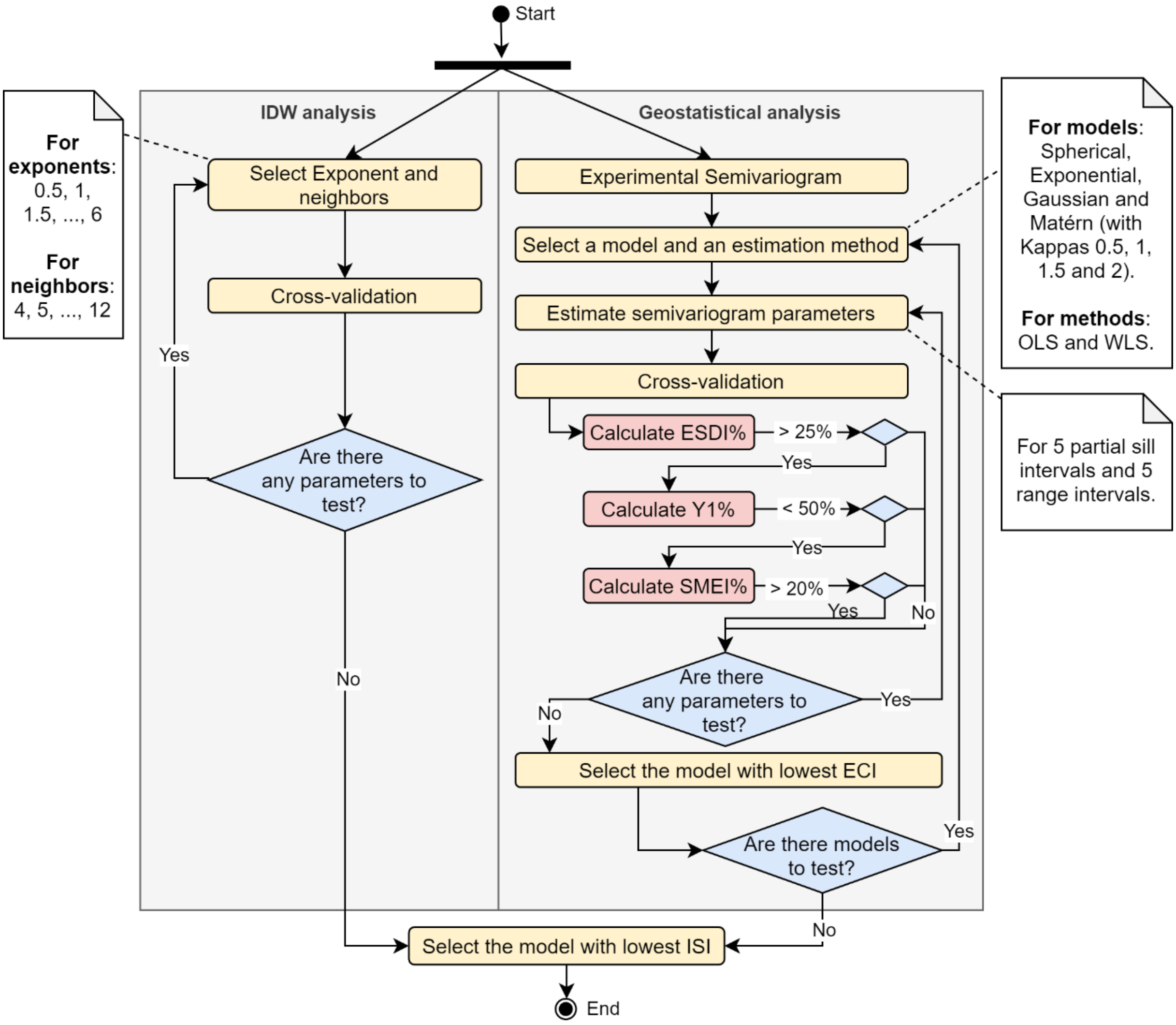


Fig. 6 Selection process of the best interpolator between inverse distance weighting and ordinary Kriging by method 3: the filters using %ESDI, $\%\gamma(1)$, and %SMEI were applied during geostatistics analysis.

Selection Methods 2 and 3 can lead to different results. The central aspect of method 3 is to allow another ‘fitted model’ to be selected in an interpolator selection analysis. In geostatistical analysis, for each combination of ‘geostatistical model’ (Spherical, Exponential, etc.) vs. ‘estimation method’ (OLS and WLS), 25 ‘fitted models’ (5 partial sill interval * 5 range intervals) are generated. When applying the selection criteria by Method 2, and eliminating the ‘fitted model’ that was considered the best, it is impossible to use another ‘fitted model’ from the same combination of ‘geostatistical model’ vs. ‘estimation method.’ In this case, the twenty five analyses were eliminated. On the other hand, selection by Method 3 makes it possible to use other ‘adjusted models’ within the combined analysis of ‘geostatistical model’ vs. ‘estimation method’.

v. Map's evaluation

The interpolated maps were compared using the coefficient of relative deviation (CRD – Equation 13) proposed by Coelho et al. (2009):

$$CRD = \sum_{i=1}^{n} \left| \frac{\hat{Z}_i - \hat{Z}_i^*}{\hat{Z}_i^*} \right| \frac{100}{n}, \quad (13)$$

where $\hat{Z}_i^*$ is the location value (pixel) i on the reference map, $\hat{Z}_i$ is the value at location (pixel) i on the map to be compared, and n is the total number of observations on the maps. The coefficient expresses the average absolute percent difference between both maps. The choice of a reference map used for comparison is arbitrary. For this study, the map generated by the best interpolator selected by Method 3 was considered the reference for each variable.

7.3 Results and discussion

i. Descriptive statistics

The descriptive analysis of variables (Table 4, 5, and 6) showed that CV varied from 5% (low, pH SMP) to 119% (very high, Al/CTC in Field A-2018), 5% (low, pH SMP, and Clay) to 123% (very high, aluminum saturation-m% in Field A-2019), and from 4% (low, pH SMP, Field B-2015) to 157% (very high, Al/CTC in Field B-2015). Variables Al, C, Ca, CTC, Al/CTC, Mg/CTC, K/CTC, Ca/CTC, Cu, Fe, K, Mg, OM, P, pH (CaCl2), pH SMP, V, m%, Clay, Sand, and Silt had points that were eliminated after eliminating outliers during EDA. Few outliers were found and eliminated in ten, nine, and twelve variables in Fields A-2018, A-2019, and B. In several cases, variables did not present normality at 5% significance level: i) Field A-2018: Al, Al/CTC, K/CTC, Cu, H+Al, K, and P; ii) Field A-2019: Al, m%, P, pH (CaCl2), Zn, and sand; and iii) Field B-2015: Al, C, H+Al, P and Al/CTC.

Table 4 Descriptive statistics of soil attributes in Field A-2018 (100 samples)

Variables	Samples remained	Mini-mum	Means	Medians	Maxi-mum	Standard deviations	CV%
Al* ($cmol_c/dm^{-3}$)	99	0.00	0.538	0.280	2.450	0.635	118 (VH)
C (g/kg)	98	15.5	21.30	21.43	26.45	2.15	10 (M)
Ca ($cmol_c/dm^{-3}$)	99	2.07	3.52	3.56	5.47	0.722	21 (H)
CTC ($cmol_c/dm^{-3}$)	100	10.8	13.32	13.20	16.50	1.10	8 (L)
Al/CTC* (%)	100	0.00	9.84	4.86	43.88	11.68	119 (VH)
Ca/CTC (%)	99	13.3	26.67	26.99	40.90	6.13	23 (H)
H/CTC (%)	99	42.8	56.6	57.0	69.1	5.38	10 (L)
K/CTC* (%)	100	1.18	3.34	3.04	5.79	1.13	34 (VH)
Mg/CTC (%)	100	4.73	9.49	9.85	14.71	2.30	24 (H)
Cu* (mg/dm^{-3})	98	1.86	4.02	3.65	8.66	1.37	34 (VH)
Fe (mg/dm^{-3})	99	4.88	15.60	15.26	29.04	4.38	28 (H)
H+Al* ($cmol_c/dm^{-3}$)	100	3.97	8.09	7.76	13.06	1.74	21 (H)
K* ($cmol_c/dm^{-3}$)	100	0.160	0.439	0.405	0.700	0.130	30 (H)
Mg ($cmol_c/dm^{-3}$)	100	0.690	1.245	1.270	1.840	0.247	20 (M)
Mn (mg/dm^{-3})	100	42.67	75.12	76.16	110.41	13.79	18 (M)
OM (g/dm^{-3})	98	26.72	36.73	36.95	45.60	3.70	10 (M)
P* (mg/dm^{-3})	98	2.60	9.65	8.40	23.50	4.48	46 (VH)
pH (CaCl2)	100	3.58	4.42	4.42	5.21	0.371	8 (L)
pH SMP	99	4.70	5.37	5.40	5.90	0.275	5 (L)
SB ($cmol_c/dm^{-3}$)	100	3.29	5.23	5.35	7.98	0.990	19 (M)
V%	99	20.85	39.47	39.73	57.23	8.62	22 (H)
Zn (mg/dm^{-3})	100	4.93	8.12	7.88	12.06	1.85	23 (H)

CV: Coefficient of variation: low (L) when CV ≤ 10%, medium (M) when 10% < CV ≤ 20%, high (H) when 20% < CV ≤ 30%, and very high (VH) when CV > 30%.

Al: aluminum; C: Carbon; Ca: calcium; CTC: cation exchange capacity; Al/CTC: aluminum adsorbed on CTC in %; Ca/CTC: calcium adsorbed on CTC in %; H/CTC: hydrogen adsorbed on CTC in %; K/CTC: potassium adsorbed on CTC in %; Mg/CTC: magnesium adsorbed on CTC in %; Cu: copper; Fe: iron; H+Al: potential acidity; K: potassium; Mg: magnesium; Mn: manganese; OM: organic matter; P: phosphorus; pH: the potential of hydrogen; pH SMP: pH of buffer solution Shoemaker-McLean-Pratt; SB: the sum of basis; V%: base saturation; Zn: zinc.

* No normality at 5% significance level.

Table 5 Descriptive statistics of soil attributes in Field A-2019 (52 samples)

Variables	Samples remained	Mini-mum	Means	Medians	Maxi-mum	Standard deviations	CV%
Al* ($cmol_c/dm^{-3}$)	51	0.00	0.27	0.15	1.15	0.30	112 (VH)
Ca ($cmol_c/dm^{-3}$)	52	1.50	4.08	4.20	6.90	1.21	30 (H)
CTC ($cmol_c/dm^{-3}$)	52	9.86	11.4	11.3	13.3	0.87	8 (L)
Ca/CTC (%)	52	14.2	35.4	37.6	51.8	8.59	24 (H)
H+Al/CTC (%)	52	20.6	46.5	42.9	79.3	11.4	24 (H)
K/CTC (%)	52	0.73	3.23	3.13	7.28	1.56	48 (VH)
Mg/CTC (%)	52	3.80	14.9	14.4	24.7	4.3	29 (H)
Cu (mg/dm^{-3})	52	4.3	9.2	8.5	14.1	2.2	24 (H)
Fe (mg/dm^{-3})	52	36	77	75	121	21	28 (H)
H+Al ($cmol_c/dm^{-3}$)	52	2.74	5.24	4.96	8.36	1.08	21 (H)
K ($cmol_c/dm^{-3}$)	51	0.090	0.356	0.330	0.800	0.171	48 (VH)
m%*	51	0.00	5.1	2.3	24.0	6.3	123 (VH)
Mg ($cmol_c/dm^{-3}$)	52	0.40	1.71	1.70	3.00	0.54	32 (VH)
Mn (mg/dm^{-3})	52	88	162	159	220	31	19 (M)
OM (g/dm^{-3})	52	14.7	25.8	26.8	41.6	5.3	21 (H)
P* (mg/dm^{-3})	51	4.4	18.1	15.8	53.0	11.0	59 (VH)
pH* (CaCl2)	51	3.80	4.50	4.50	5.30	0.35	8 (L)
pH SMP	52	5.30	5.96	6.00	6.80	0.29	5 (L)
SB ($cmol_c/dm^{-3}$)	52	2.2	6.2	6.3	10.6	1.6	27 (H)
V%	52	20.7	53.4	57.1	79.4	11.4	21 (H)
Zn* (mg/dm^{-3})	50	1.44	3.97	3.77	9.41	1.44	36 (VH)
Clay (%)	51	68.0	74.0	74.0	84.0	3.50	5 (L)
Sand* (%)	50	0.70	2.51	2.60	5.10	0.84	33 (VH)
Silt (%)	51	14.3	23.3	23.2	30.8	3.4	15 (M)

CV: Coefficient of variation: low (L) when CV ≤ 10%, medium (M) when 10% < CV ≤ 20%, high (H) when 20% < CV ≤ 30%, and very high (VH) when CV > 30%.

Al: aluminum; Ca: calcium; CTC: cation exchange capacity; Ca/CTC: calcium adsorbed on CTC in %; H+Al/CTC: aluminum more hydrogen adsorbed on CTC in %; K/CTC: potassium adsorbed on CTC in %; Mg/CTC: magnesium adsorbed on CTC in %; Cu: copper; Fe: iron; H+Al: potential acidity; K: potassium; m%: aluminum saturation; Mg: magnesium; Mn: manganese; OM: organic matter; P: phosphorus; pH: the potential of hydrogen; SB: the sum of basis; SMP: pH of buffer solution Shoemaker-McLean-Pratt; V%: base saturation; Zn: zinc; * No normality at 5% significance level.

Table 6 Descriptive statistics of soil attributes in Field B-2015 (73 samples)

Variables	Samples remained	Mini-mum	Means	Medians	Maxi-mum	Standard deviations	CV%
Al* ($cmol_c/dm^{-3}$)	72	0.000	0.065	0.020	0.390	0.095	146 (VH)
C* (g/kg)	72	16.9	21.7	21.4	27.7	2.4	11 (M)
Ca ($cmol_c/dm^{-3}$)	73	3.11	5.35	5.38	8.36	1.03	19 (M)
Cu (mg/dm^{-3})	72	11.6	14.8	14.8	20.3	1.6	11 (M)
Fe (mg/dm^{-3})	73	32.3	55.6	53.6	85.0	11.4	20 (H)
H+Al* ($cmol_c/dm^{-3}$)	73	3.18	5.87	5.76	9.00	1.05	18 (M)
K ($cmol_c/dm^{-3}$)	72	0.19	0.446	0.430	0.960	0.156	35 (VH)
Mg ($cmol_c/dm^{-3}$)	72	1.17	2.08	2.06	3.15	0.41	20 (M)
Mn (mg/dm^{-3})	73	224	316	313	400	49	15 (M)
P* (mg/dm^{-3})	72	4.8	12.4	11.1	29.9	5.4	43 (VH)
pH (CaCl2)	72	4.40	5.04	5.05	5.70	0.29	6 (L)
pH SMP	72	5.20	5.78	5.80	6.20	0.22	4 (L)
SB ($cmol_c/dm^{-3}$)	73	4.7	7.9	8.0	12.0	1.4	18 (M)
V%	73	34.3	57.2	58.2	79.1	8.4	15 (M)
Zn (mg/dm^{-3})	73	2.25	4.76	4.59	8.41	1.42	30 (H)
CTC ($cmol_c/dm^{-3}$)	72	12.5	13.8	13.6	15.8	0.7	5 (L)
Al/CTC* (%)	72	0.00	0.98	0.20	6.41	1.55	157 (VH)
Ca/CTC (%)	73	22.7	38.7	39.7	50.8	6,0	15 (M)
Mg/CTC (%)	72	8.60	15.1	15.4	22.2	2.8	18 (M)
K/CTC (%)	71	1.50	3.18	3.10	6.38	0.99	31 (VH)

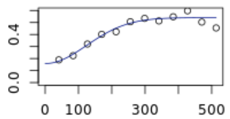
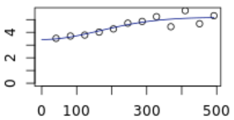
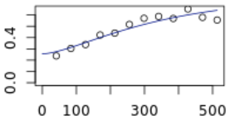
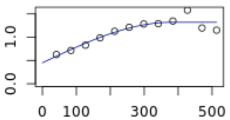
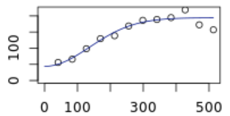
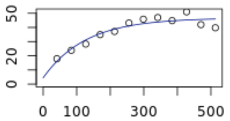
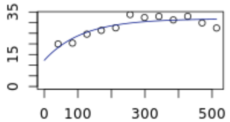
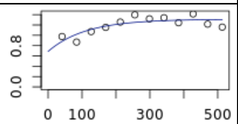
CV: Coefficient of variation: low (L) when CV ≤ 10%, medium (M) when 10% < CV ≤ 20%, high (H) when 20% < CV ≤ 30%, and very high (VH) when CV > 30%.
Al: aluminum; C: carbon; Ca: calcium; CTC: cation exchange capacity; Ca/CTC: calcium adsorbed on CTC in %; Al/CTC: aluminum adsorbed on CTC in %; K/CTC: potassium adsorbed on CTC in %; Mg/CTC: magnesium adsorbed on CTC in %; Cu: copper; Fe: iron; H+Al: potential acidity; K: potassium; Mg: magnesium; Mn: manganese; P: phosphorus; pH: the potential of hydrogen; pH SMP: pH of buffer solution Shoemaker-McLean-Pratt; SB: the sum of basis; V%: base saturation; Zn: zinc.
* No normality at 5% significance level.

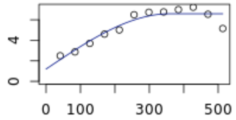
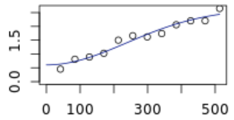
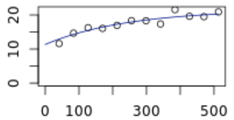
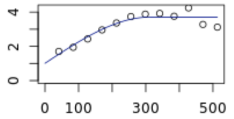
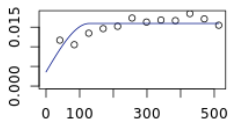
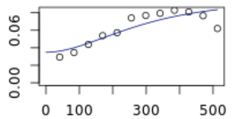
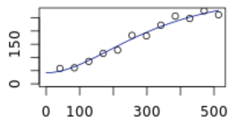
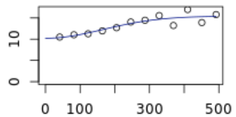
ii. Selection of the best interpolator model

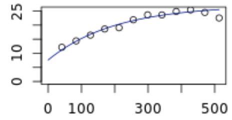
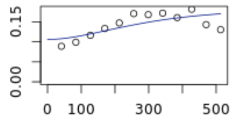
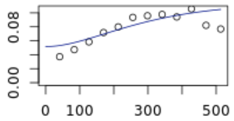
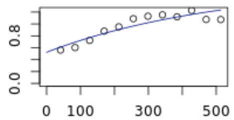
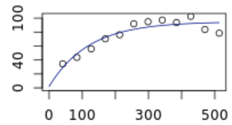
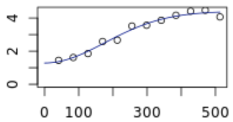
Method 1: The results of selecting the best interpolator model for IDW and OK using ISI for variables of Fields A-2018 (Table 7), A-2019 (Table 8), and B (Table 9) showed that the OK one is the best interpolator for 42 variables (10 in Field A-2018, 19 in Field A-2019, and 13 in Field B-2015) and IDW for 24 variables (12 in Field A-2018, 5 in Field A-2019, and 7 in Field B-2015).

During SD analysis, the 50%-cutoff limited range to 513 m (Field A-2018), 353 m (Field A-2019), and 419 m (Field B-2015). Therefore, the correspondent number of lags was twelve (Field A-2018), eight (Field A-2019), and fourteen (Field B-2015), always with a minimum of 30 pairs of points. The first semivariance corresponded to 41 m (Field A-2018), 45 m (Field A-2019), and 31 m (Field B-2015). ISI selected IDW as the best interpolator for i) Field A-2018: CTC, Ca/CTC, H/CTC, K/CTC, Mg/CTC, H+Al, K, Mn, pH CaCl2, pH SMP, V%, and Zn, ii) Field A-2019: Ca, Cu, K, m%, and SB, and iii) Field B-2015: Ca, Fe, Mg, Mn, Zn, Ca/CTC, and Mg/CTC. For the remained variables, OK was indicated as the best interpolator.

Table 7 Result of selecting the best interpolator model for inverse distance weighted interpolation (IDW) and ordinary Kriging (OK) using the interpolator selection index (ISI) for variables of Field A-2018, using Method 1: Selection using only ISI

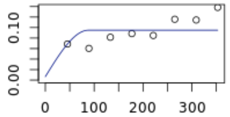
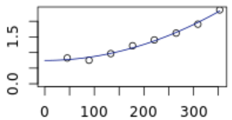
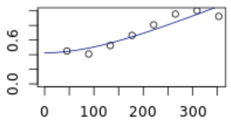
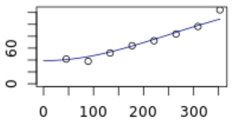
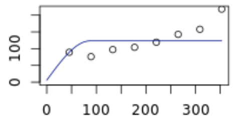
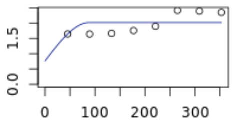
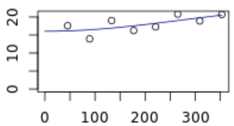
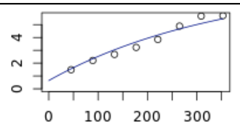
Variables	Geostatistics												IDW					Best Inter-polator
	Models	C_0	C_1	Ra	%SDI	%ESDI	$\%\gamma(1)$	%SMEI	ISI	ME*10^2	SDME	Semivariogram	Exp	Neig	ISI	ME*10^2	SDME	
Al	Gaussian WLS	0.158	0.382	176	71 (H)	65 (H)	8	71	0.00223	-0.00650	0.428		5	5	0.0679	0.013	0.458	OK
C	Gaussian OLS	3.44	1.75	234	34 (L)	32 (L)	5	33	0.0935	-0.817	2.031		1	7	0.419	-3.522	2.067	OK
Ca	Matérn 1 OLS	0.256	0.495	225	66 (H)	68 (H)	-4	60	0.0238	-0.037	0.529		4.5	5	0.0251	0.00876	0.536	OK
CTC	Spherical OLS	0.450	0.870	379	66 (H)	53 (M)	20	66	0.122	0.226	0.801		3	11	0.0302	0.0593	0.791	IDW
Al/CTC	Gaussian WLS	43.9	150	193	77 (H)	71 (H)	8	77	0.0436	2.266	7.507		5	5	0.0517	-0.0905	7.914	OK
Ca/CTC	Exponential OLS	4.42	42.2	127	91 (VH)	62 (H)	32	90	0.115	-2.598	4.506		4.5	4	0.0328	-0.519	4.534	IDW
H/CTC	Exponential OLS	12.3	19.5	113	61 (H)	37 (L)	39	61	0.138	1.671	4.741		1.5	7	0.0146	-0.0137	4.773	IDW
K/CTC	Exponential OLS	0.687	0.624	110	48 (M)	26 (L)	46	47	0.103	-0.221	0.960		3.5	12	0.0153	0.0332	0.949	IDW

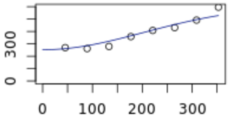
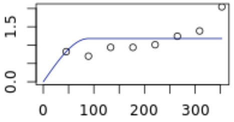
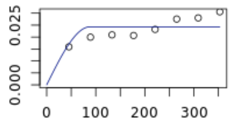
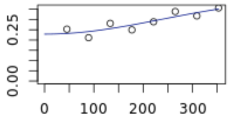
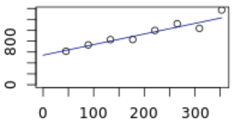
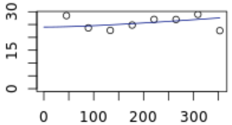
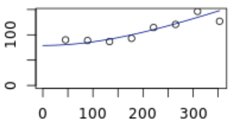
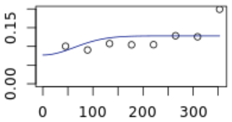
Variables	Geostatistics												IDW					Best Inter-polator
	Models	C_0	C_1	Ra	%SDI	%ESDI	$\%\gamma(1)$	%SMEI	ISI	ME*10^2	SDME	Semivariogram	Exp	Neig	ISI	ME*10^2	SDME	
Mg/CTC	Spherical OLS	1.197	5.39	359	82 (VH)	62 (H)	24	82	0.157	-0.817	1.638		1	6	0.0062	-0.0222	1.606	IDW
Cu	Gaussian OLS	0.605	2.03	336	77 (H)	83 (VH)	-8	75	0.123	0.591	0.736		6	7	0.540	4.505	0.702	OK
Fe	Exponential WLS	11.3	10.2	244	47 (M)	46 (M)	3	40	0.0243	0.129	4.024		2	11	0.0308	-0.292	3.989	OK
H+Al	Spherical OLS	1.01	2.70	313	73 (H)	54 (M)	26	73	0.131	0.672	1.281		5.5	10	0.0032	-0.00144	1.276	IDW
K	Spherical WLS	0.00361	0.0123	128	77 (H)	27 (L)	65	77	0.1156	-0.0292	0.107		3.5	7	0.00901	-0.000650	0.106	IDW
Mg	Matérn 2 OLS	0.0348	0.0565	128	62 (H)	68 (H)	-10	58	0.0575	-0.0268	0.184		5.5	5	0.108	-0.000691	0.203	OK
Mn	Matérn 2 OLS	42.06	284	138	87 (VH)	82 (VH)	6	85	0.0695	-1.401	7.579		1.5	10	0.0313	-0.161	7.665	IDW
OM	Gaussian OLS	10.22	5.19	234	34 (L)	32 (L)	5	33	0.0934	-1.408	3.501		1	7	0.419	-6.070	3.564	OK

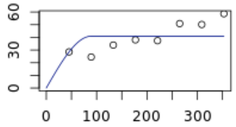
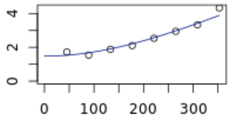
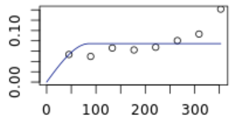
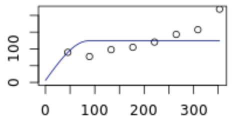
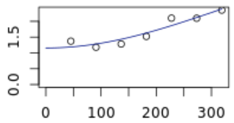
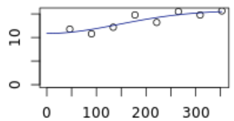
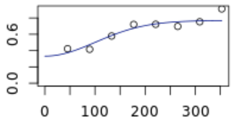
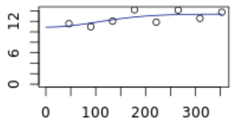
Variables	Models	C_0	C_1	Ra	%SDI	%ESDI	$\%\gamma(1)$	%SMEI	ISI	ME*10^2	SDME	Semivariogram	Exp	Neig	ISI	ME*10^2	SDME	Best Inter-polator
	Geostatistics												IDW					
P	Exponential WLS	7.62	19.4	200	72 (H)	55 (M)	23	70	0.0255	-0.175	3.735		3	9	0.0609	0.154	3.895	OK
pH CaCl2	Matérn 2 OLS	0.106	0.0744	128	41 (M)	51 (M)	-23	38	0.0598	-0.0177	0.308		1	8	0.0234	0.000577	0.303	IDW
pH SMP	Matérn 2 WLS	0.052	0.0618	128	54 (M)	67 (H)	-24	51	0.0745	-0.0323	0.205		4.5	9	0.00942	-0.00171	0.199	IDW
SB	Exponencial OLS	0.524	1.125	513	68 (H)	66 (H)	3	58	0.0159	-0.039	0.754		3.5	5	0.0172	-0.0226	0.759	OK
V%	Exponential OLS	1.94	92.9	120	98 (VH)	64 (H)	35	98	0.168	-4.941	6.104		2.5	8	0.0104	-0.0771	6.141	IDW
Zn	Gaussian OLS	1.29	3.12	250	71 (H)	67 (H)	5	71	0.140	-0.514	1.209		2.5	11	0.0159	0.0107	1.226	IDW

C_0: nugget effect; C_1: partial sill; Ra: range; %SDI: Spatial Dependence Index; %ESDI: Effective Spatial Dependence Index; $\%\gamma(1)$: First Semivariance Significance Index; ISI: Interpolator Selection Index; ME: Mean Error; SDME: Standard Deviation of Mean Error; IDW: Inverse Distance Weighting; OK: Ordinary Kriging; Exp: exponent; Neig: neighbors; OLS: Ordinary Least Squares; WLS: Weighted Least Squares. Classification of %SDI and ESDI: very low for %SDI < 20%; low for 20 ≤ %SDI < 40%; medium for 40 ≤ %SDI < 60%; high for 60 ≤ %SDI < 80%; and very high for %SDI > 80%. Values highlighted in Light Salmon do not agree with the criteria defined in Table 2 (%ESDI > 25%, %γ(1) < 50%, and %SMEI > 20).

Table 8 Result of selecting the best interpolator model for inverse distance weighted interpolation (IDW) and ordinary Kriging (OK) using the interpolator selection index (ISI) for variables of Field A-2019, using Method 1: Selection using only ISI

Varia-bles	Geostatistics												IDW					Best Inter-polator
	Models	C_0	C_1	Ra	%SDI	%ESDI	$\%\gamma(1)$	%SMEI	ISI	ME*10^2	SDME	Semivariogram	Exp	Neig	ISI	ME*10^2	SDME	
Al	Spherical WLS	0.00646	0.0885	88	93 (VH)	28 (L)	70	93	0.161	0.159	0.282		1	9	0.435	0.736	0.270	OK
Ca	Matérn 2 WLS	0.732	7.97	353	92 (VH)	91 (VH)	1	67	0.289	-1.295	0.925		1	9	0.223	-1.121	0.892	OK
CTC	Matérn 1.5 OLS	0.425	2.415	353	85 (VH)	84 (VH)	1	60	0.0242	-0.022	0.669		3	6	0.0488	-0.0259	0.687	OK
Ca/CTC	Gaussian WLS	38.9	111	353	74 (H)	72 (H)	2	64	0.251	-10.34	6.432		1	9	0.455	-20.34	6.353	OK
H+Al/CTC	Spherical WLS	6.1	117	87	95 (VH)	28 (L)	71	95	0.147	1.71	10.579		1	9	0.452	22.46	9.427	OK
K/CTC	Spherical OLS	0.764	1.25	88	62 (H)	18 (VL)	70	62	0.128	0.156	9.430		1	12	0.330	-2.674	1.316	OK
Mg/CTC	Matérn 1.5 WLS	15.9	13.12	287	45 (M)	39 (L)	13	22	0.025	0.416	1.420		1.5	6	0.0478	0.389	4.302	OK
Cu	Matérn 0.5 WLS	0.64	7.63	353	92 (VH)	82 (VH)	11	88	0.1170	1.898	1.416		4.5	5	0.0594	0.153	1.478	IDW

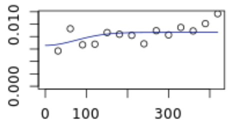
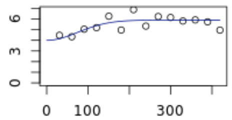
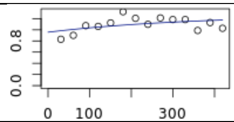
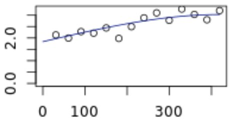
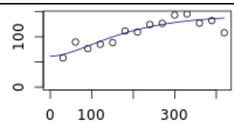
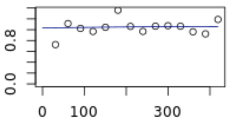
Varia-bles	Geostatistics												IDW					Best Inter-polator
	Models	C_0	C_1	Ra	%SDI	%ESDI	$\%\gamma(1)$	%SMEI	ISI	ME*10^2	SDME	Semivariogram	Exp	Neig	ISI	ME*10^2	SDME	
Fe	Gaussian OLS	253	353	286	58 (M)	56 (M)	5	52	0.0082	0.571	16.366		1.5	6	0.0837	-0.958	17.38	OK
H+Al	Spherical OLS	0.00	1.18	89.1	100 (VH)	30 (L)	70	100	0.111	0.161	16.471		6	4	0.551	2.328	1.105	OK
K	Spherical OLS	0.00	0.0243	88.2	100 (VH)	35 (L)	65	100	0.138	0.0392	1.012		1	10	0.0954	0.0173	0.145	IDW
Mg	Gaussian WLS	0.229	0.183	335	44 (M)	39 (L)	13	35	0.00139	0.000468	0.147		1	6	0.0310	0.0250	0.518	OK
Mn	Exponential WLS	542	506335	260037	100 (VH)	50 (M)	10	56	0.138	-15.778	26.064		6	9	0.221	-24.05	27.08	OK
OM	Matérn 2 WLS	24.0	6.97	170	23 (L)	-3 (VL)	126	13	0.024	-0.434	31.68		1	4	0.481	-7.212	5.814	OK
P	Matérn 2 WLS	79.0	340	337	81 (VH)	79 (H)	3	47	0.00303	-0.217	5.079		1.5	9	0.0645	-0.911	9.950	OK
pH CaCl2	Gaussian OLS	0.0768	0.0511	88.2	40 (L)	22 (L)	46	40	0.0392	-0.0642	9.342		6	8	0.321	-0.272	0.370	OK

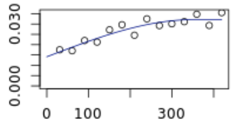
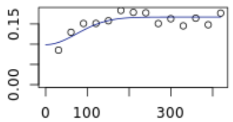
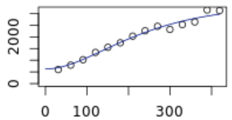
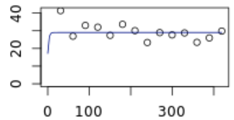
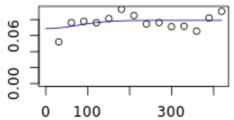
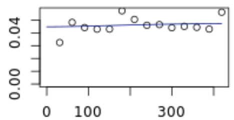
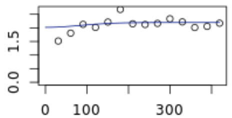
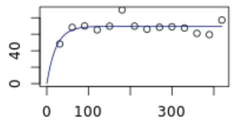
Varia-bles						Geostatistics									IDW			Best Inter-polator
	Models	C_0	C_1	Ra	%SDI	%ESDI	$\%\gamma(1)$	%SMEI	ISI	ME*10^2	SDME	Semivariogram	Exp	Neig	ISI	ME*10^2	SDME	
m%	Spherical WLS	0.0	40.9	89	100 (VH)	30 (L)	70	100	0.175	3.372	5.772		1	9	0.388	13.025	5.473	OK
SB	Matérn 2 WLS	1.48	12.8	353	90 (VH)	88 (VH)	2	62	0.277	-1.430	1.330		1	10	0.0876	-0.416	1.316	IDW
pH SMP	Spherical WLS	0.00	0.0746	86	100 (VH)	28 (L)	72	100	0.137	-0.0296	0.266		6	4	0.580	-0.550	0.281	OK
V%	Spherical WLS	5.9	118	87	95 (VH)	28 (L)	71	95	0.148	-1.66	10.662		1	9	0.443	-21.31	9.472	OK
Zn	Gaussian OLS	1.16	3.08	446	73 (H)	43 (M)	17	52	0.0229	0.0375	3.286		2.5	5	0.0476	0.0636	1.147	OK
Clay	Gaussian OLS	10.9	4.88	205	31 (L)	25 (L)	18	30	0.0396	-0.678	0.263		1	5	0.142	-0.996	3.639	OK
Sand	Gaussian WLS	0.332	0.435	142.9	57 (M)	45 (M)	21	57	0.0825	0.653	0.659		1	4	0.293	1.853	0.706	OK
Silt	Gaussian OLS	10.9	2.50	146	19 (VL)	14 (VL)	27	19	0.0243	0.277	1.126		1	10	0.0511	-0.407	3.351	OK

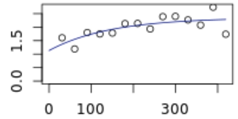
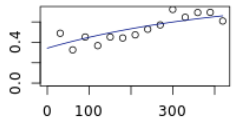
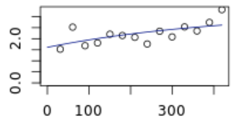
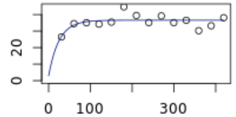
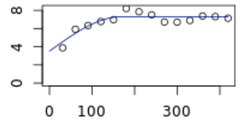
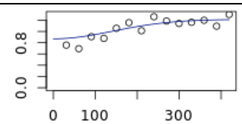
C_0: nugget effect; C_1: partial sill; Ra: range; %SDI: Spatial Dependence Index; %ESDI: Effective Spatial Dependence Index; $\%\gamma(1)$: the First Semivariance Significance Index; ISI: Interpolator Selection Index; ME: Mean Error; SDME: Standard Deviation of Mean Error; IDW: Inverse Distance

Weighting; OK: Ordinary Kriging; Exp: exponent; Neig: neighbors; OLS: Ordinary Least Squares; WLS: Weighted Least Squares. Classification of %SDI and ESDI: very low for %SDI < 20%; low for 20 ≤ %SDI < 40%; medium for 40 ≤ %SDI < 60%; high for 60 ≤ %SDI < 80%; and very high for %SDI > 80%. Values highlighted in Light Salmon do not agree with the criteria defined in Table 2 (%ESDI > 25%, %γ(1) < 50%, and %SMEI > 20).

Table 9 Result of selecting the best interpolator model for inverse distance weighted interpolation (IDW) and ordinary Kriging (OK) using the interpolator selection index (ISI) for variables of Field B-2015, using Method 1: Selection using only ISI

Varia-ble	Model / Method	Geostatistics											IDW					Best Inter-polator
		C_0	C_1	Ra	%SDI	%ESDI	$\%\gamma(1)$	%SMEI	ISI	ME*10^3	SDME	Semivariogram	Exp	Neig	ISI	ME*10^3	SDME	
Al	Gaussian WLS	0.00658	0.00209	105	24 (L)	34 (L)	-42	24	0.0456	0.138	0.0926		2.5	4	0.0783	-0.0378	0.0996	OK
C	Gaussian WLS	3.99	1.89	106	32 (L)	24 (L)	25	32	0.0681	-9.685	2.123		1	7	0.114	-12.84	2.176	OK
Ca	Exponential OLS	0.957	0.308	341	24 (L)	34 (L)	-41	19	0.0437	-1.157	1.015		1.5	7	0.0347	-0.800	1.010	IDW
Cu	Spherical OLS	1.836	1.19	419	39 (L)	30 (L)	25	39	0.116	9.89	1.536		3	10	0.646	57.66	1.531	OK
Fe	Matérn 1.5 WLS	61.4	81.0	94.5	57 (M)	59 (M)	-4	55	0.0689	17.95	9.579		1	5	0.0336	-1.994	9.632	IDW
H+Al	Matérn 1.5 WLS	1.04	0.02521	105	2 (VL)	32 (L)	-1232	2	0.0272	0.2684	1.051		1.5	7	0.0867	1.122	1.109	OK

Variable	Model / Method	Geostatistics C_0	C_1	Ra	%SDI	%ESDI	$\%\gamma(1)$	%SMEI	ISI	ME*10^3	SDME	Semivariogram	IDW Exp	Neig	ISI	ME*10^3	SDME	Best Interpolator
K	Spherical WLS	0.0141	0.0180	350	56 (M)	46 (M)	19	56	0.0026	0.0314	0.130		2	6	0.0344	-0.0992	0.134	OK
Mg	Gaussian OLS	0.0987	0.0682	105	41 (M)	49 (M)	-20	41	0.0972	-1.444	0.371		1.5	6	0.0533	0.677	0.370	IDW
Mn	Matérn 2 WLS	625	2706	103	81 (VH)	82 (VH)	-1	79	0.127	-159.4	30.04		1	4	0.00380	-11.04	27.75	IDW
P	Exponential WLS	17.15	11.86	2.4	41 (M)	-43 (VL)	204	41	0.0000 0003	- 0.00001	5.404		2	7	0.0691	-2.49	5.810	OK
pH CaCl2	Gaussian WLS	0.0687	0.0103	105	13 (VL)	34 (L)	-162	13	0.0801	-0.645	0.274		1	7	0.230	-2.695	0.273	OK
pH SMP	Matérn 2 WLS	0.0448	0.00297	105	6 (VL)	32 (L)	-415	5	0.0461	-0.145	0.219	3 	1.5	7	0.0625	0.0570	0.226	OK
SB	Gaussian OLS	2.019	0.178	118	8 (VL)	31 (L)	-281	8	0.0311	-1.457	1.421		1	7	0.0377	-1.620	1.429	OK
V%	Exponential WLS	0.00	69.7	22.2	100 (VH)	31 (L)	69	100	0.0759	-48.06	8.254		1	7	0.0768	20.69	8.681	OK

Varia-ble	Model / Method	Geostatistics											IDW					Best Inter-polator
		C_0	C_1	Ra	%SDI	%ESDI	$\%\gamma(1)$	%SMEI	ISI	ME*10^3	SDME	Semivariogram	Exp	Neig	ISI	ME*10^3	SDME	
Zn	Exponential OLS	1.132	1.22	145	52 (M)	32 (L)	39	50	0.0811	5.169	1.260		2.5	12	0.0642	1.654	1.311	IDW
CTC	Matérn 0.5 OLS	0.343	0.502	419	59 (M)	42 (M)	29	48	0.0748	-2.855	0.643		1	7	0.249	-9.229	0.654	OK
Al/CTC	Exponential OLS	1.61	1.59	419	50 (M)	52 (M)	-5	38	0.0512	2.396	1.530		2	9	0.0564	0.588	1.607	OK
Ca/CTC	Exponential WLS	2.93	33.6	25.0	92 (VH)	28 (L)	70	92	0.0745	-27.64	5.923		1	8	0.0337	-2.361	6.110	IDW
Mg/CTC	Spherical WLS	3.51	3.78	168	52 (M)	47 (M)	9	52	0.143	-14.43	2.544		2.5	6	0.0242	-0.944	2.541	IDW
K/CTC	Gaussian OLS	0.869	0.344	201	28 (L)	38 (L)	-33	28	0.0163	-0.521	0.894		1.5	8	0.0342	0.684	0.907	OK

C_0: nugget effect; C_1: partial sill; Ra: range; %SDI: Spatial Dependence Index; %ESDI: Effective Spatial Dependence Index; $\%\gamma(1)$: First Semivariance Significance Index; ISI: Interpolator Selection Index; ME: Mean Error; SDME: Standard Deviation of Mean Error; IDW: Inverse Distance Weighting; OK: Ordinary Kriging; Exp: exponent; Neig: neighbors; OLS: Ordinary Least Squares; WLS: Weighted Least Squares. Classification of %SDI and ESDI: very low for %SDI < 20%; low for 20 ≤ %SDI < 40%; medium for 40 ≤ %SDI < 60%; high for 60 ≤ %SDI < 80%; and very high for %SDI > 80%. Values highlighted in Light Salmon do not agree with the criteria defined in Table 2 (%ESDI > 25%, %γ(1) < 50%, and %SMEI > 20).

Some variables had their semivariogram models considered unsatisfactory, highlighted in Light Salmon (Tables 7, 8, and 9). They did not agree with the criteria defined in Table 2 (%ESDI > 25%, %γ(1) < 50%, and %SMEI > 20).

The variables' spatial dependences (SD, Fig. 7), measured by the traditional %SDI (Equation 2), were classified, on average, as medium (24%), as high (23%), and very high (30%). However, using %ESDI (Equation 3), SD was classified, on average, as medium (23%), as high (18%), and very high (11%). That means that the high and very high sum lowered from 53% to 29% and that %SDI masks the actual SD.

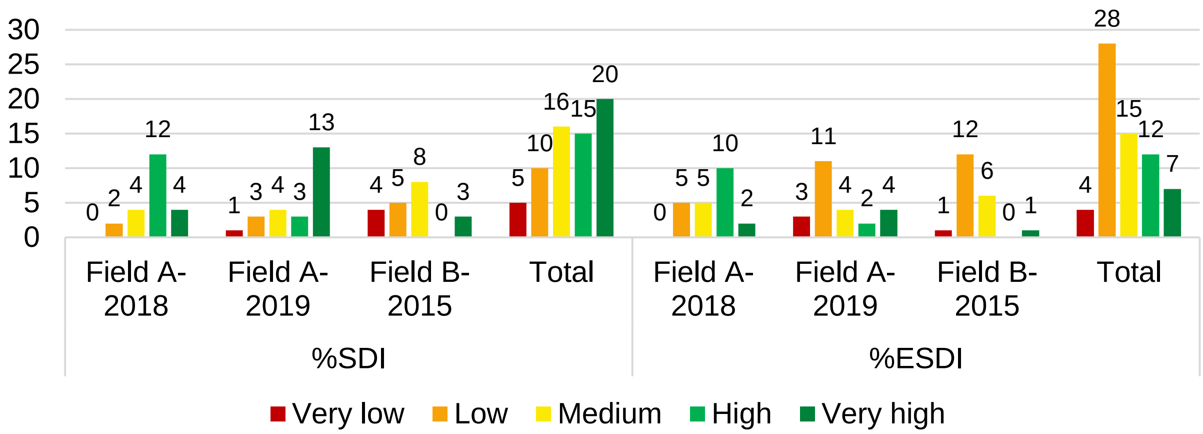


Fig. 7 Number of variables of each class for %SDI and %ESDI (very low, low, medium, high, and very high) for each Field (A-2018, A-2019, and B-2015).

According to the visual inspection of each variable semivariogram (Tables 7, 8, and 9), there seems to be a lack of adjustment of the model pointed out as the best for some variables in the Fields A-2018 (K), A-2019 (Al, H+Al/CTC, K/CTC, H+Al, K, m%, pH SMP, and V%) and B (Ca/CTC and V%). In other cases, there is an indication of pure nugget effect in Field A-2019 (OM, and pH CaCl2) and Field B-2015 (Al, Ca, H+Al, P, pH CaCl2, pH SMP, and SB). Clay and silt can also be included in this list (Field A-2019). Among the variables with "doubtful" or "pure nugget effect" adjustment, IDW interpolator was considered the best only for K, Fields A-2018, and A-2019, and Ca and Ca/CTC, in Field B-2015.

Another aspect observed was the fact that %SDI (Fig. 7) indicated wrongly the presence of strong spatial dependence (high or very high) in some variables in the following areas: (i) Field A-2018: K, (ii) Field A-2019: Al, K/CTC, H+Al/CTC, H+Al, K, m%, pH SMP, and V% and (iii) Field B-2015: V% and Ca/CTC. The first semivariance plotted in the semivariograms of these variables shows a high variance of data at the closest distances and that the model was adjusted incorrectly. In these cases, %SDI gives some false feeling of having an adequate model, which presents a strong spatial dependence.

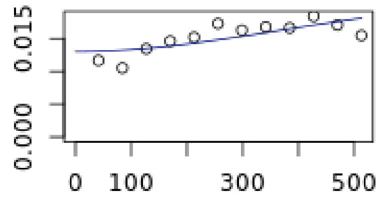
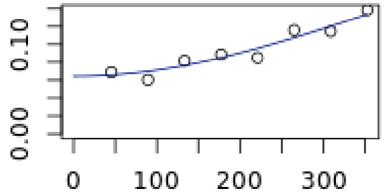
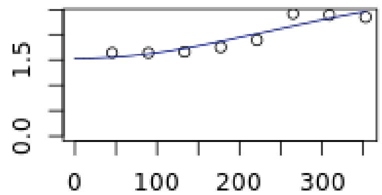
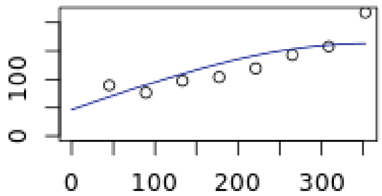
This kind of problem with semivariogram adjustments is due to the model's automatic adjustment to the semivariogram made by geoR package's routines. The automatic adjustment of models to semivariograms is pointed out in literature as a notoriously tricky task (WEBSTER; OLIVER, 1990; GOOVAERTS, 1997). As with any method for adjusting the variogram model, they all assume the model's basic structure in advance and then obtained the predefined model structure's optimal coefficients. Selecting the variogram model and its parameters is the most controversial aspect of geostatistics; shapes of valid variogram models are finite; sometimes, the model's optimal shape cannot be fitted, leading to reduced estimation accuracy (HAN; WANG; ZHENG, 2016). In this sense, it is proposed in this work criteria (using %ESDI, %γ(1), and %SMEI) to improve the semivariogram adjustment process, which are presented by Methods 2 and 3.

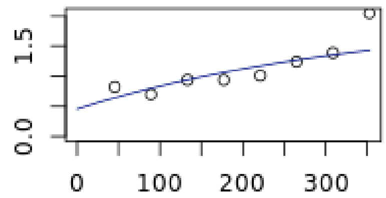
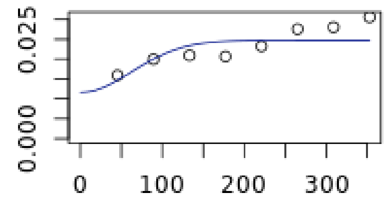
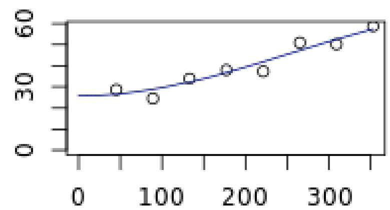
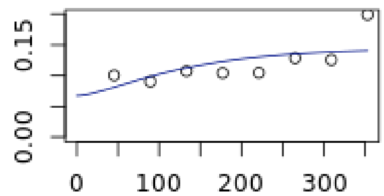
Method 2: This method was applied to variables with unsatisfactory semivariogram models (Tables 7, 8, and 9). As a result, other semivariogram models were selected for variables in Field A-2019 (Al, K/CTC, H+Al, m%, pH CaCl2, m%, and Clay). In another case, IDW interpolator was considered the best for variable SB (Field B-2015) (Table 10). It is noteworthy that variables OM and silt, from field A-2019, and C, H+Al, P, pH SMP, and V%, from Field B-2015, had all semivariogram models eliminated. In these cases, IDW interpolator was considered the best one.

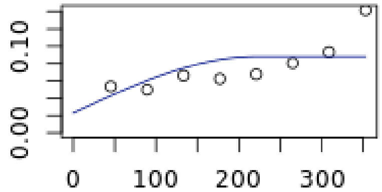
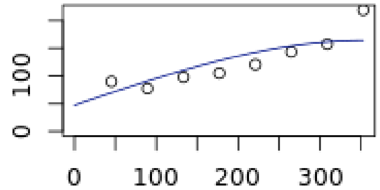
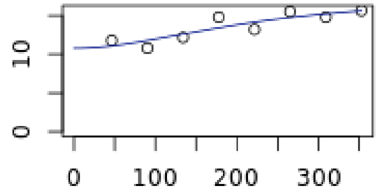
IDW interpolator had was considered using Method 1 as the best interpolator for variable K, in Fields A-2018 (Table 7) and A-2019 (Table 8), and for variables Ca and Ca/CTC, in Field B-2015 (Table 9). However, other semivariogram models' selection behavior was evaluated regardless of whether IDW was identified as the best. As a result, this allowed us to verify that the variable K, from Fields A-2018 and A-2019, and the variables Ca and Ca/CTC, from Field B-2015, could choose another semivariogram model (Table 10).

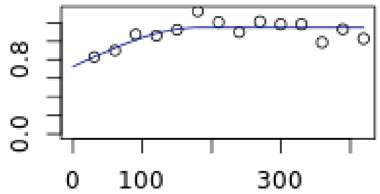
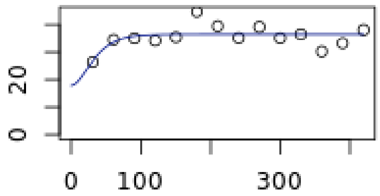
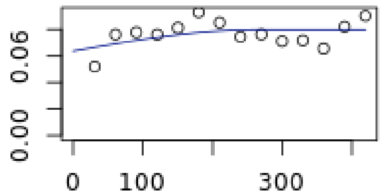
It is essential to highlight that the three criteria must be considered together in the semivariogram models' selection process. According to the semivariogram structure, a wrong model can be selected when it is not applied in association (see results in Table 11). This issue was the most important in Field A-2019 and the least important for Field A-2018.

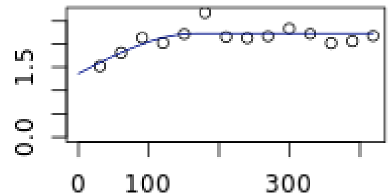
Table 10 Result of selecting the best interpolator model for ordinary Kriging (OK) using the interpolator selection index (ISI) for variables of Fields A-2018, A-2019, and B-2015 using Method 2: The three criteria (%ESDI > 25%, %γ(1) < 50%, and %SMEI > 20) are applied after geostatistics analysis

Variables/ Fields	Models	C_0	C_1	Ra	%SDI	%ESDI	$\%\gamma(1)$	%SMEI	ISI	ME*10^2	SDME	Semivariogram	Best Inter-polator
K Field A-2018	Gaussian OLS	0.0130	0.00808	513	38 (L)	45 (M)	-17	28	0.122	-0.015	0.113		IDW
Al Field A-2019	Matérn 2 WLS	0.0640	0.372	353	85 (VH)	84 (VH)	1	52	0.1919	0.279	0.272		OK
K/CTC Field A-2019	Gaussian OLS	1.54	1.47	353	49 (M)	33 (L)	12	38	0.1540	-1.010	1.324		OK
H+Al/CTC Field A-2019	Spherical OLS	46.5	116.6	353	71 (H)	45 (M)	37	71	0.2194	10.706	9.404		OK

Variables/ Fields	Models	C_0	C_1	Ra	%SDI	%ESDI	$\%\gamma(1)$	%SMEI	ISI	ME*10^2	SDME	Semivariogram	Best Inter-polator
H+Al Field A-2019	Exponential WLS	0.456	1.54	353	77 (H)	59 (M)	24	68	0.2251	1.357	0.916		OK
K Field A-2019	Gaussian WLS	0.0116	0.0131	88	53 (M)	36 (L)	33	53	0.1474	-0.126	0.136		IDW
m% Field A-2019	Gaussian OLS	25.73	49.89	353	66 (H)	62 (H)	6	55	0.2179	6.602	5.508		OK
OM Field A-2019	All geostatistical models were eliminated												IDW
pH CaCl2 Field A-2019	Matérn 1 OLS	0.0672	0.0773	88	53 (M)	31 (L)	43	52	0.1255	-0.205	0.313		OK

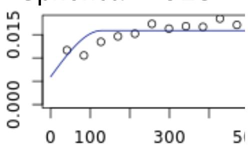
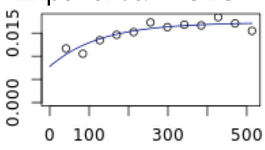
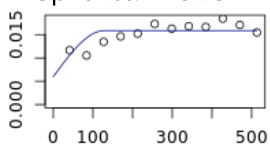
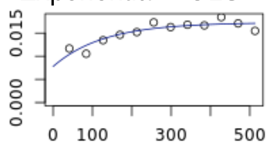
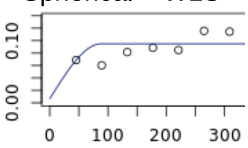
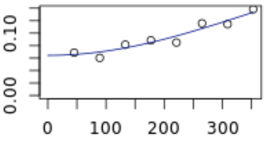
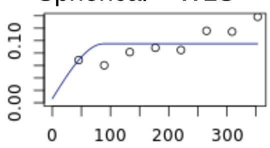
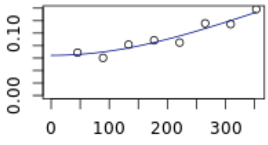
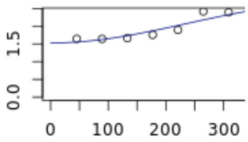
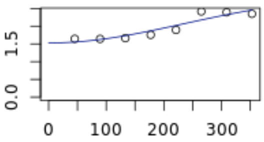
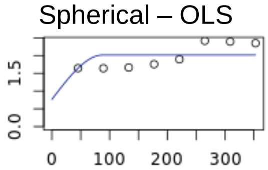
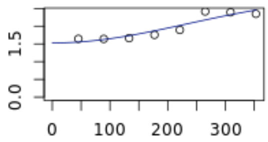
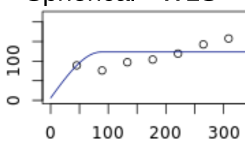
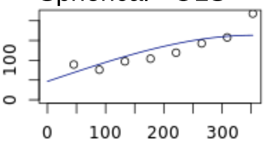
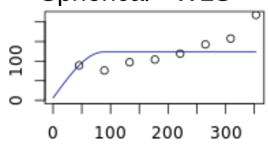
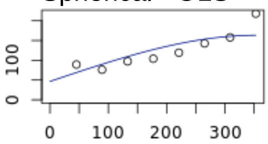
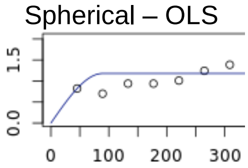
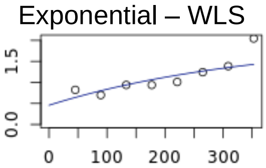
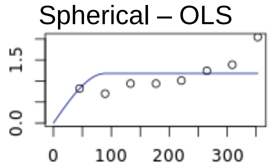
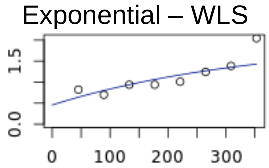
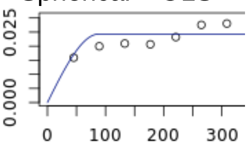
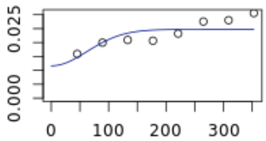
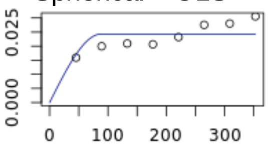
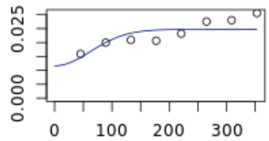
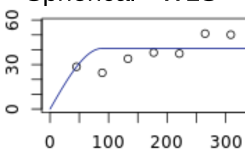
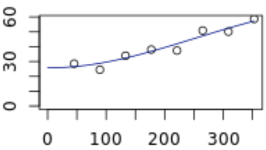
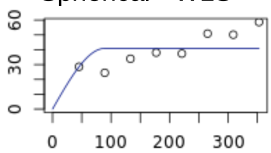
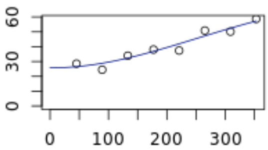
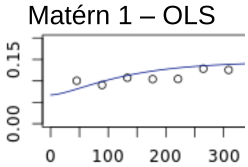
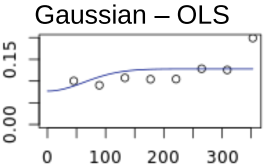
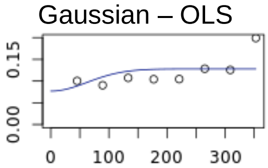
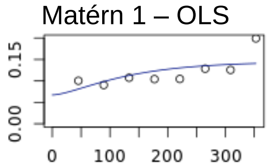
Variables/ Fields	Models	C_0	C_1	Ra	%SDI	%ESDI	$\%\gamma(1)$	%SMEI	ISI	ME*10^2	SDME	Semivariogram	Best Inter-polator
pH SMP Field A-2019	Spherical OLS	0.0230	0.0650	221	74 (H)	39 (L)	47	74	0.1827	-0.249	0.236		OK
V% Field A-2019	Spherical OLS	47.4	116.4	353	71 (H)	45 (M)	36	71	0.2247	-10.551	9.474		OK
Clay Field A-2019	Matérn 2 OLS	10.8	5.84	96	35 (L)	29 (L)	17	31	0.0452	-0.719	3.290		OK
Silt Field A-2019	All geostatistical models were removed												IDW
C Field B-2015	All geostatistical models were eliminated												IDW

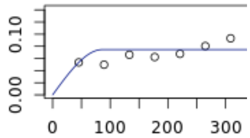
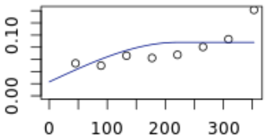
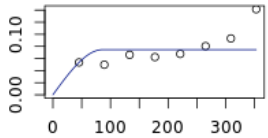
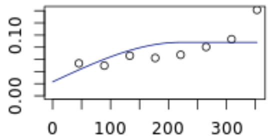
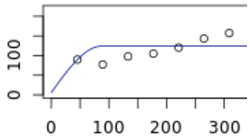
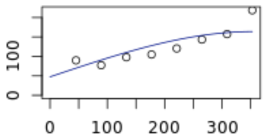
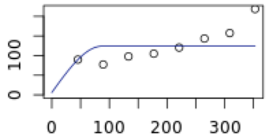
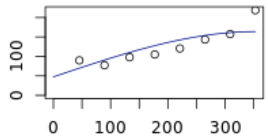
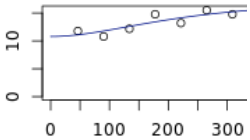
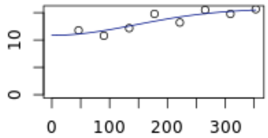
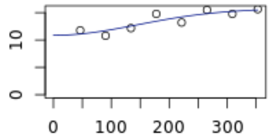
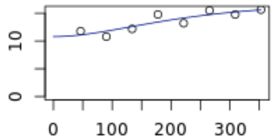
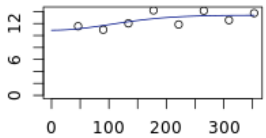
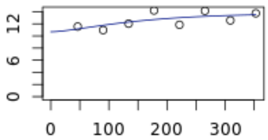
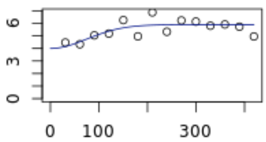
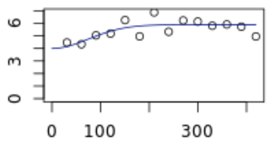
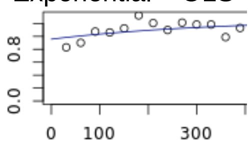
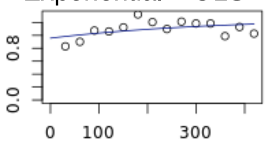
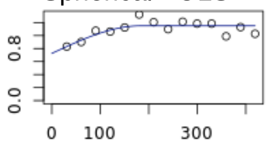
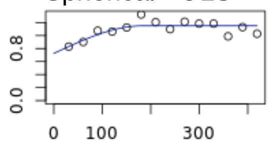
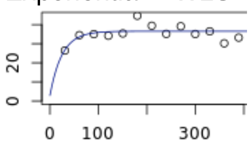
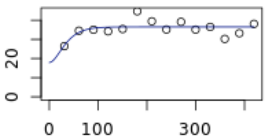
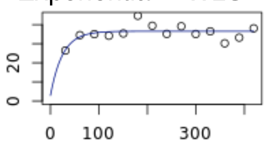
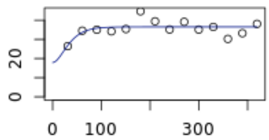
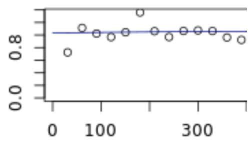
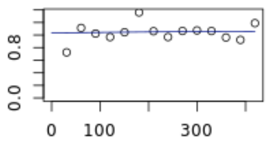
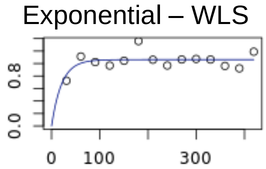
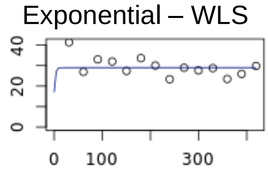
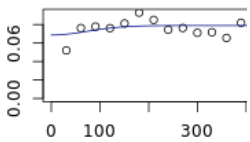
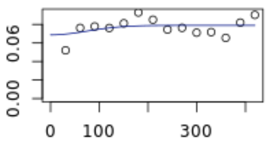
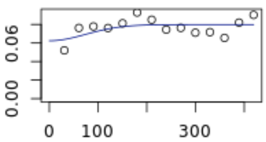
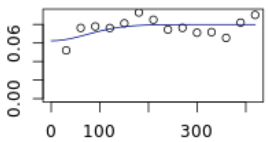
Variables/ Fields	Models	C_0	C_1	Ra	%SDI	%ESDI	$\%\gamma(1)$	%SMEI	ISI	ME*10^2	SDME	Semivariogram	Best Inter-polator
Ca Field B-2015	Spherical OLS	0.725	0.425	183	37 (L)	28 (L)	25	37	0.0445	-2.865	0.987		IDW
Ca/CTC Field B-2015	Matérn 2 WLS	17.9	18.6	16	51 (M)	28 (L)	46	51	0.0796	-29.649	5.921		IDW
H+Al Field B-2015	All geostatistical models were eliminated												IDW
P Field B-2015	All geostatistical models were eliminated												IDW
pH CaCl2 Field B-2015	Gaussian OLS	0.0624	0.0174	105	22 (L)	35 (L)	-59	22	0.0693	-0.891	0.271		OK

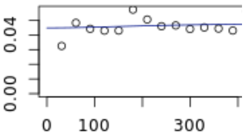
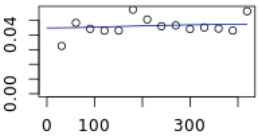
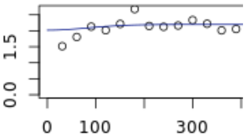
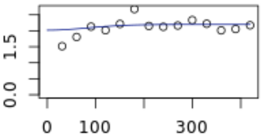
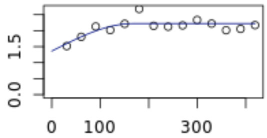
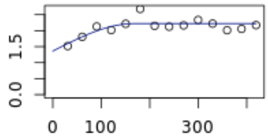
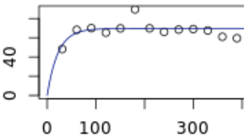
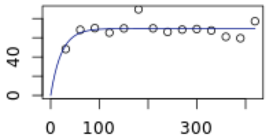
Variables/ Fields	Models	C_0	C_1	Ra	%SDI	%ESDI	$\%\gamma(1)$	%SMEI	ISI	ME*10^2	SDME	Semivariogram	Best Inter-polator
pH SMP Field B-2015	All geostatistical models were eliminated												IDW
SB Field B-2015	Spherical WLS	1.360	0.861	167	39 (L)	32 (L)	18	39	0.0527	-5.516	1.398		IDW
V% Field B-2015	All geostatistical models were removed												IDW

C_0: nugget effect; C_1: partial sill; Ra: range; %SDI: Spatial Dependence Index; %ESDI: Effective Spatial Dependence Index; $\%\gamma(1)$: First Semivariance Significance Index; %SMEI: Slope of the Model Ends Index; ISI: Interpolator Selection Index; ME: Mean Error; SDME: Standard Deviation of Mean Error; IDW: Inverse Distance Weighting; OK: Ordinary Kriging; OLS: Ordinary Least Squares; WLS: Weighted Least Squares. Classification of %SDI and ESDI: very low for %SDI/ESDI < 20%; low for 20 ≤ %SDI/ESDI < 40%; medium for 40 ≤ %SDI/ESDI < 60%; high for 60 ≤ %SDI/ESDI < 80%; and very high for %SDI/ESDI > 80%.

Table 11 Result of selecting the best interpolator model for ordinary Kriging (OK) with Method 2 using each criterion separately and all together for variables of Fields A-2018, A-2019, and B-2015

Variables/ Fields	Criterion 1 only %ESDI > 25%	Criterion 2 only $\%\gamma(1)$ < 50%	Criterion 3 only %SMEI > 20%	All criteria
K Field A-2018	Spherical – OLS* 	Exponential – OLS* 	Spherical – OLS* 	Exponential – OLS* 
Al Field A-2019	Spherical – WLS 	Matérn 2 – WLS 	Spherical – WLS 	Matérn 2 – WLS 
K/CTC Field A-2019	Gaussian – OLS 	Gaussian – OLS 	Spherical – OLS 	Gaussian – OLS 
H+Al/CTC Field A-2019	Spherical - WLS 	Spherical - OLS 	Spherical - WLS 	Spherical - OLS 
H+Al Field A-2019	Spherical – OLS 	Exponential – WLS 	Spherical – OLS 	Exponential – WLS 
K Field A-2019	Spherical – OLS* 	Gaussian – WLS* 	Spherical – OLS* 	Gaussian – WLS* 
m% Field A-2019	Spherical - WLS 	Gaussian - OLS 	Spherical - WLS 	Gaussian - OLS 
OM Field A-2019	All geostatistical models were eliminated	All geostatistical models were eliminated	All geostatistical models were eliminated	All geostatistical models were eliminated
pH CaCl2 Field A-2019	Matérn 1 – OLS 	Gaussian – OLS 	Gaussian – OLS 	Matérn 1 – OLS 

Variables/ Fields	Criterion 1 only %ESDI > 25%	Criterion 2 only $\%\gamma(1) < 50\%$	Criterion 3 only %SMEI > 20%	All criteria
pH SMP Field A-2019	Spherical - WLS 	Spherical - OLS 	Spherical - WLS 	Spherical - OLS 
V% Field A-2019	Spherical - WLS 	Spherical - WLS 	Spherical - WLS 	Spherical - WLS 
Clay Field A-2019	Matérn 2 – OLS 	Gaussian – OLS 	Gaussian – OLS 	Matérn 2 – OLS 
Silt Field A-2019	All geostatistical models were eliminated	Gaussian – OLS 	Matérn 1 – OLS 	All geostatistical models were eliminated
C Field B-2015	All geostatistical models were eliminated	Gaussian – WLS 	Gaussian – WLS 	All geostatistical models were eliminated
Ca Field B-2015	Exponential – OLS 	Exponential – OLS 	Spherical - OLS 	Spherical - OLS 
Ca/CTC Field B-2015	Exponential – WLS* 	Matérn 2 – WLS* 	Exponential – WLS* 	Matérn 2 – WLS* 
H+Al Field B-2015	Matérn 1.5 – WLS 	Matérn 1.5 – WLS 	Exponential – WLS 	All geostatistical models were eliminated
P Field B-2015	All geostatistical models were eliminated	All geostatistical models were eliminated	Exponential – WLS 	All geostatistical models were eliminated
pH CaCl2 Field B-2015	Gaussian – WLS 	Gaussian – WLS 	Gaussian – OLS 	Gaussian – OLS 

Variables/ Fields	Criterion 1 only %ESDI > 25%	Criterion 2 only $\%\gamma(1)$ < 50%	Criterion 3 only %SMEI > 20%	All criteria
pH SMP Field B-2015	Matérn 2 – WLS 	Matérn 2 – WLS 	All geostatistical models were eliminated	All geostatistical models were eliminated
SB Field B-2015	Gaussian - OLS 	Gaussian - OLS 	Spherical – WLS* 	Spherical – WLS* 
V% Field B-2015	Exponential – WLS 	All geostatistical models were eliminated	Exponential – WLS 	All geostatistical models were eliminated

OLS: Ordinary Least Squares; WLS: Weighted Least Squares; %ESDI: Effective Spatial Dependence Index; $\%\gamma(1)$: First Semivariance Significance Index; %SMEI: Slope of the Model Ends Index; Al: aluminum; C: carbon; Ca: calcium; Ca/CTC: calcium adsorbed on CTC in %; H+Al/CTC: aluminum more hydrogen adsorbed on CTC in %; H+Al: potential acidity; K: Potassium; K/CTC: potassium adsorbed on CTC in %; m%: aluminum saturation; OM: organic matter; P: phosphorus; pH: the potential of hydrogen; pH SMP: pH of buffer solution Shoemaker-McLean-Pratt; SB: sum of basis; V%: base saturation.
* The IDW interpolator was considered better than the model adjusted to the semivariogram.

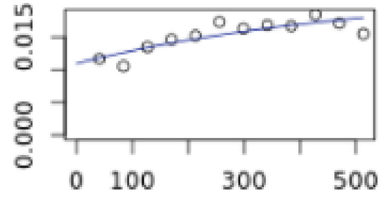
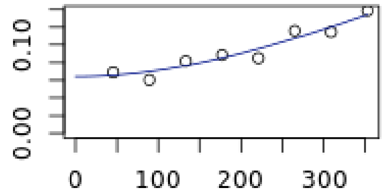
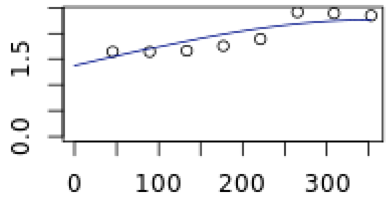
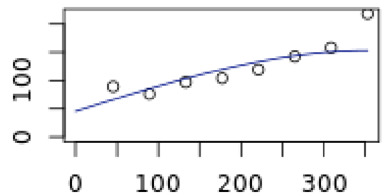
Method 3: This method, just as Method 2, was applied to the variables with unsatisfactory semivariogram models (Tables 7, 8, and 9). As a result, some models were eliminated in favor of others. In OM and Silt variables, from Field A-2019, and in C, H+Al, P, and V% variables, from Field B-2015, all geostatistical models were eliminated during the geostatistical analysis (Table 12). All other variables had changes in semivariogram parameters in comparison to Method 1.

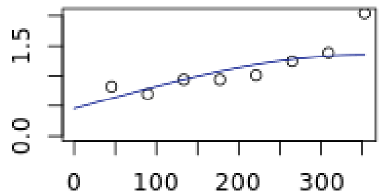
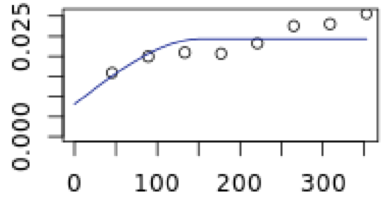
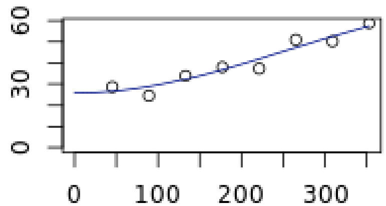
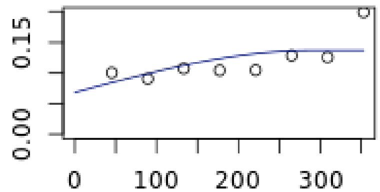
Other semivariogram models were selected for variables in Field A-2019 (Al, m%, pH CaCl2, and Clay) and Field B-2015 (pH CaCl2, pH SMP, and SB) (Table 12). In other cases, IDW interpolator was considered the best one: Field A-2019 (OM and silt) and Field B-2015 (C, H+Al, P, pH SMP, SB, and V%).

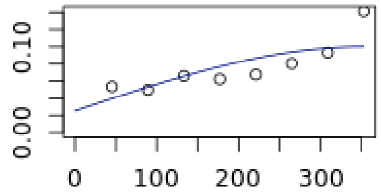
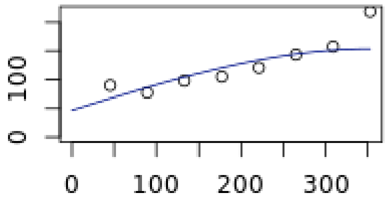
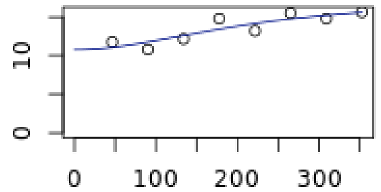
Variables as K/CTC, H+Al/CTC, and V% (Field A-2019) kept the model selected by Method 1 (Spherical – OLS or WLS) but with other semivariogram adjusting parameters. In variables H+Al, K, and pH SMP (Field A-2019) and Ca, the model selected by Method 1 (Spherical) remained; however, the method of adjusting the semivariogram changed between OLS and WLS. Variables Al, m%, and Clay (Field A-2019) and Ca and SB (Field B-2015) kept the model selected in Method 2. Despite maintaining the models, variables K (Field A-2018) and pH SMP (Field A-2019) changed the semivariogram adjustment parameters.

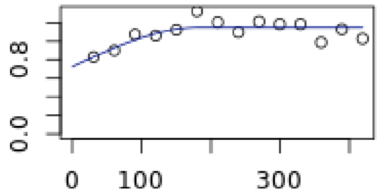
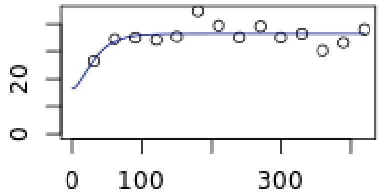
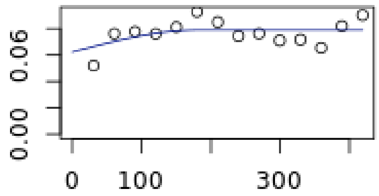
As it was expected, Methods 2 and 3 conducted different results. Method 3 allows another ‘fitted model’ to be selected in the geostatistical analysis, and as it was explained in section M&M, it is expected to lead to the best interpolator model (IDW or OK).

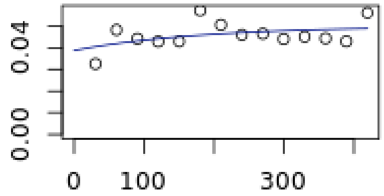
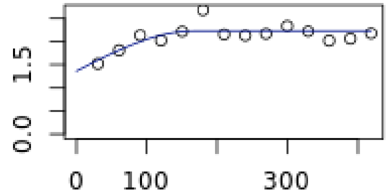
Table 12 Result of selecting the best interpolator model for ordinary Kriging (OK) using interpolator selection index (ISI) for variables of Fields A-2018, A-2019, and B-2015 using Method 3: The three criteria (%ESDI > 25%, %γ(1) < 50%, and %SMEI > 20) are applied during geostatistics analysis

Variables/ Fields	**Models**	**C_0**	**C_1**	**Ra**	**%SDI**	**%ESDI**	**$\%\gamma(1)$**	**%SMEI**	**ISI**	**ME*10^2**	**SDME**	**Semivariogram**	**Best Inter-polator**
K Field A-2018	Exponential OLS	0.0109	0.0111	513	50 (M)	47 (M)	7	39	0.126	-0.028	0.109		IDW
Al Field A-2019	Matérn 2 WLS	0.064	0.372	353	85 (VH)	84 (VH)	1	52	0.1919	0.279	0.272		OK
K/CTC (%) Field A-2019	Spherical OLS	1.37	0.894	353	39 (L)	27 (L)	31	39	0.1259	-0.968	1.290		OK
H+Al/CTC (%) Field A-2019	Spherical WLS	45.6	107.3	353	70 (H)	42 (M)	41	70	0.2150	10.507	9.399		OK

Variables/ Fields	Models	C_0	C_1	Ra	%SDI	%ESDI	$\%\gamma(1)$	%SMEI	ISI	ME*10^2	SDME	Semivariogram	Best Inter-polator
H+Al Field A-2019	Spherical WLS	0.454	0.899	353	66 (H)	39 (L)	41	66	0.1872	1.131	0.907		OK
K Field A-2019	Spherical WLS	0.00809	0.0163	154	67 (H)	35 (L)	48	67	0.1109	-0.091	0.137		IDW
m% Field A-2019	Gaussian OLS	25.73	49.89	353	66 (H)	62 (H)	6	55	0.2179	6.602	5.508		OK
OM Field A-2019	All geostatistical models were eliminated												IDW
pH CaCl2 Field A-2019	Spherical OLS	0.0682	0.0684	287	50 (M)	27 (L)	47	50	0.1248	-0.184	0.318		OK

Variables/ Fields	Models	C_0	C_1	Ra	%SDI	%ESDI	$\%\gamma(1)$	%SMEI	ISI	$ME*10^2$	SDME	Semivariogram	Best Inter-polator
pH SMP Field A-2019	Spherical OLS	0.0253	0.0749	353	75 (H)	47 (M)	38	75	0.2185	-0.319	0.233		OK
V% Field A-2019	Spherical WLS	46.7	106.7	353	70 (H)	41 (M)	40	70	0.2194	-10.318	9.467		OK
Clay Field A-2019	Matérn 2 OLS	10.78	5.83	96	35 (L)	29 (L)	17	31	0.0452	-0.719	3.290		OK
Silt Field A-2019	All geostatistical models were eliminated												IDW
C Field B-2015	All geostatistical models were eliminated												IDW

Variables/ Fields	Models	C_0	C_1	Ra	%SDI	%ESDI	$\%\gamma(1)$	%SMEI	ISI	$ME*10^2$	SDME	Semivariogram	Best Inter-polator
Ca Field B-2015	Spherical OLS	0.725	0.425	183	37 (L)	28 (L)	25	37	0.0445	-0.287	0.987		IDW
Ca/CTC Field B-2015	Matérn 1.5 WLS	16.55	19.20	18	55 (M)	28 (L)	50	55	0.0786	-2.898	5.921		IDW
H+Al Field B-2015	All geostatistical models were eliminated												IDW
P Field B-2015	All geostatistical models were eliminated												IDW
pH CaCl2 Field B-2015	Spherical WLS	0.062	0.017	183	21 (L)	34 (L)	-62	21	0.0673	-0.084	0.271		OK

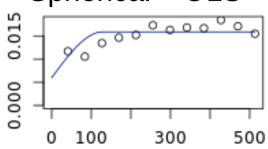
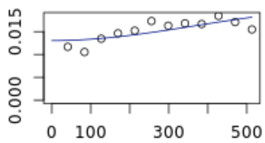
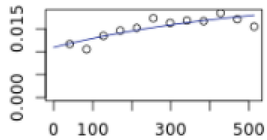
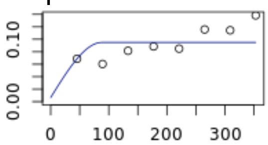
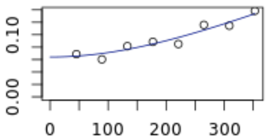
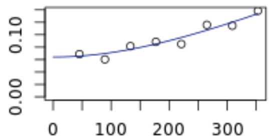
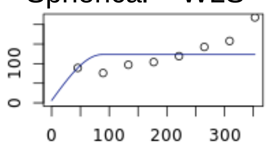
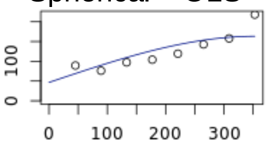
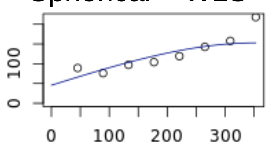
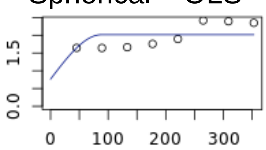
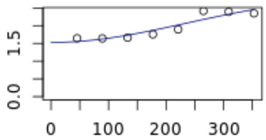
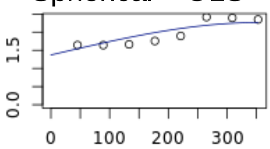
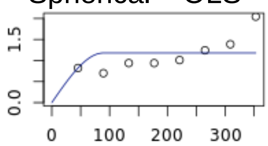
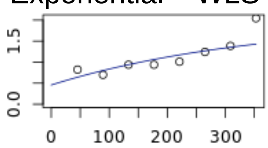
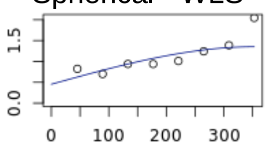
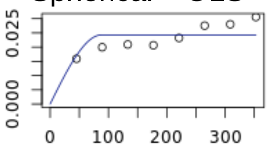
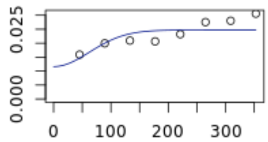
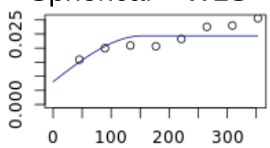
Variables/ Fields	Models	C_0	C_1	Ra	%SDI	%ESDI	$\%\gamma(1)$	%SMEI	ISI	ME*10^2	SDME	Semivariogram	Best Inter-polator
pH SMP Field B-2015	Exponential OLS	0.039	0.011	183	22 (L)	35 (L)	-57	21	0.0480	-0.035	0.217		IDW
SB Field B-2015	Spherical WLS	1.360	0.861	167	39 (L)	32 (L)	18	39	0.0527	-5.516	1.398		IDW
V% Field B-2015	All geostatistical models were eliminated												IDW

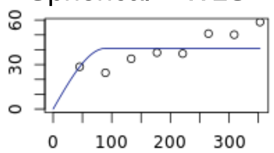
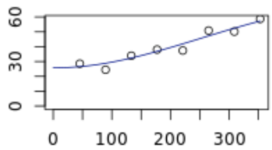
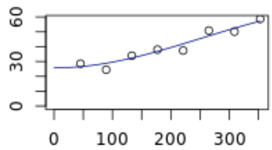
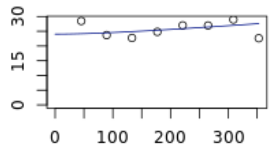
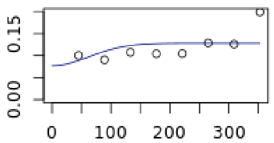
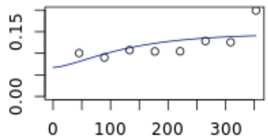
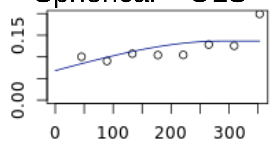
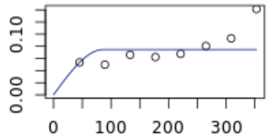
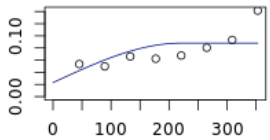
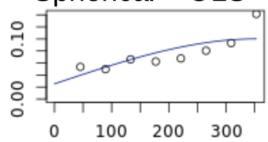
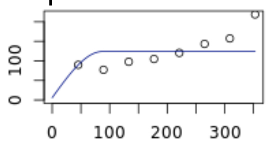
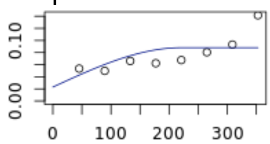
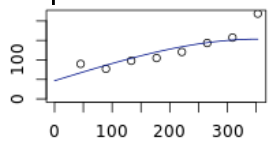
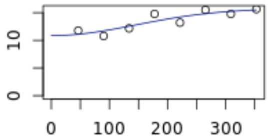
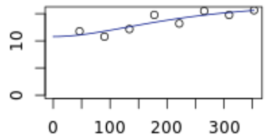
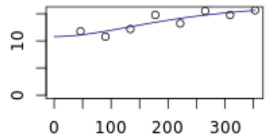
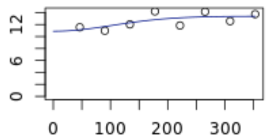
C_0: nugget effect; C_1: partial sill; Ra: range; %SDI: Spatial Dependence Index; %ESDI: Effective Spatial Dependence Index; First Semivariance Significance Index ($\%\gamma(1)$); %SMEI: Slope of the Model Ends Index; ISI: Interpolator Selection Index; ME: Mean Error; SDME: Standard Deviation of Mean Error; IDW: Inverse Distance Weighting; OK: Ordinary Kriging; OLS: Ordinary Least Squares; WLS: Weighted Least Squares. Classification of %SDI and ESDI: very low for %SDI/ESDI < 20%; low for 20 ≤ %SDI/ESDI < 40%; medium for 40 ≤ %SDI/ESDI < 60%; high for 60 ≤ %SDI/ESDI < 80%; and very high for %SDI/ESDI > 80%.

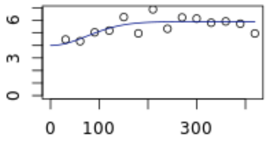
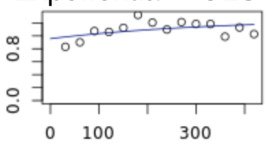
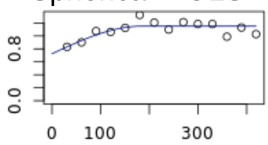
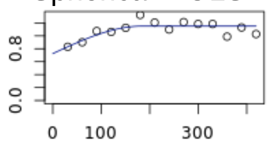
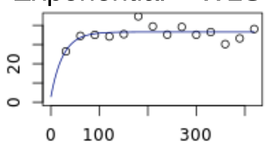
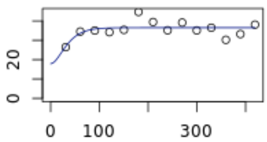
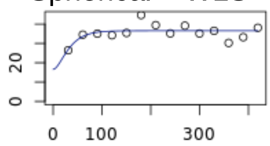
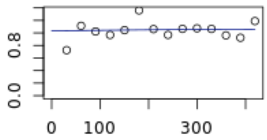
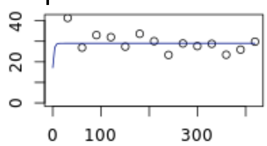
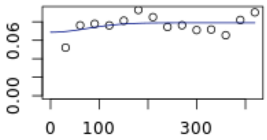
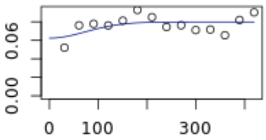
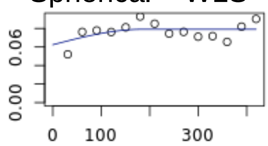
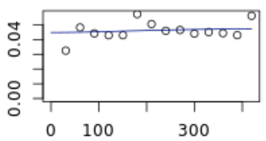
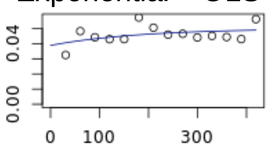
iii. Comparison of the three Methods

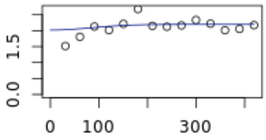
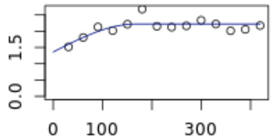
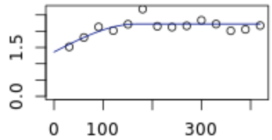
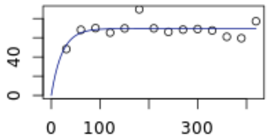
When comparing the interpolator selection result for the variables considered with inadequate geostatistical models, it was noticed that the selected interpolator might change according to the selection method (Table 13).

Table 13 The best interpolation models selected with each of the three methods

Variables/ Fields		**Method 1**	**Method 2**	**Method 3**
K Field A-2018	Best Semivari-ogram	Spherical – OLS 	Gaussian – OLS 	Exponential – OLS 
	Best Interpolator	IDWe3.5n7	IDWe3.5n7	IDWe3.5n7
Al Field A-2019	Best Semivari-ogram	Spherical – WLS 	Matérn 2 - WLS 	Matérn 2 - WLS 
	Best Interpolator	Spherical – WLS	Matérn 2 - WLS	Matérn 2 - WLS
H+Al/CTC Field A-2019	Best Semivari-ogram	Spherical – WLS 	Spherical – OLS 	Spherical – WLS 
	Best Interpolator	Spherical – WLS	Spherical – OLS	Spherical – WLS
K/CTC Field A-2019	Best Semivari-ogram	Spherical – OLS 	Gaussian – OLS 	Spherical – OLS 
	Best Interpolator	Spherical – OLS	Gaussian – OLS	Spherical – OLS
H+Al Field A-2019	Best Semivari-ogram	Spherical – OLS 	Exponential – WLS 	Spherical - WLS 
	Best Interpolator	Spherical – OLS	Exponential – WLS	Spherical – OLS
K Field A-2019	Best Semivari-ogram	Spherical – OLS 	Gaussian – WLS 	Spherical – WLS 

Variables/ Fields		Method 1	Method 2	Method 3
	Best Interpolator	IDWe1n10	IDWe1n10	IDWe1n10
m% Field A-2019	Best Semivari-ogram	Spherical – WLS 	Gaussian – OLS 	Gaussian – OLS 
	Best Interpolator	Spherical – WLS	Gaussian – OLS	Gaussian – OLS
OM Field A-2019	Best Semivari-ogram	Matérn 2 – WLS 	All geostatistical models were eliminated	All geostatistical models were eliminated
	Best Interpolator	Matérn 2 – WLS	IDWe1n4	IDWe1n4
pH CaCl2 Field A-2019	Best Semivari-ogram	Gaussian – OLS 	Matérn 1 – OLS 	Spherical – OLS 
	Best Interpolator	Gaussian – OLS	Matérn 1 – OLS	Spherical – OLS
pH SMP Field A-2019	Best Semivari-ogram	Spherical – WLS 	Spherical – OLS 	Spherical – OLS 
	Best Interpolator	Spherical – WLS	Spherical – OLS	Spherical – OLS
V% Field A-2019	Best Semivari-ogram	Spherical – WLS 	Spherical – OLS 	Spherical – WLS 
	Best Interpolator	Spherical – WLS	Spherical – OLS	Spherical – WLS
Clay Field A-2019	Best Semivari-ogram	Gaussian – OLS 	Matérn 2 – OLS 	Matérn 2 – OLS 
	Best Interpolator	Gaussian – OLS	Matérn 2 – OLS	Matérn 2 – OLS
Silt Field A-2019	Best Semivari-ogram	Gaussian – OLS 	All geostatistical models were eliminated	All geostatistical models were eliminated
	Best Interpolator	Gaussian – OLS	IDWe1n10	IDWe1n10

Variables/ Fields		Method 1	Method 2	Method 3
C Field B-2015	Best Semivari-ogram	Gaussian – WLS 	All geostatistical models were eliminated	All geostatistical models were eliminated
	Best Interpolator	Gaussian – WLS	IDWe1n7	IDWe1n7
Ca Field B-2015	Best Semivari-ogram	Exponential – OLS 	Spherical – OLS 	Spherical – OLS 
	Best Interpolator	IDWe1.5n7	IDWe1.5n7	IDWe1.5n7
Ca/CTC Field B-2015	Best Semivari-ogram	Exponential – WLS 	Matérn 2 – WLS 	Spherical – WLS 
	Best Interpolator	IDWe1n8	IDWe1n8	IDWe1n8
H+Al Field B-2015	Best Semivari-ogram	Matérn 1.5 – WLS 	All geostatistical models were eliminated	All geostatistical models were eliminated
	Best Interpolator	Matérn 1.5 – WLS	IDWe1.5n7	IDWe1.5n7
P Field B-2015	Best Semivari-ogram	Exponential – WLS 	All geostatistical models were eliminated	All geostatistical models were eliminated
	Best Interpolator	Exponential – WLS	IDWe2n7	IDWe2n7
pH CaCl2 Field B-2015	Best Semivari-ogram	Gaussian – WLS 	Gaussian – OLS 	Spherical – WLS 
	Best Interpolator	Gaussian – WLS	Gaussian – OLS	Spherical – WLS
pH SMP Field B-2015	Best Semivari-ogram	Matérn 2 – WLS 	All geostatistical models were eliminated	Exponential – OLS 
	Best Interpolator	Matérn 2 – WLS	IDWe1.5n7	IDWe1.5n7

Variables/ Fields		Method 1	Method 2	Method 3
SB Field B-2015	Best Semivari-ogram	Gaussian – OLS 	Spherical – WLS 	Spherical – WLS 
	Best Interpolator	Gaussian – OLS	IDWe1n7	IDWe1n7
V% Field B-2015	Best Semivari-ogram	Exponential – WLS 	All geostatistical models were eliminated	All geostatistical models were eliminated
	Best Interpolator	Exponential – WLS	IDWe1n7	IDWe1n7

Method 1: Only the best ISI; Method 2: the three criteria are applied after geostatistics analysis; Method 3: the three criteria are applied during geostatistics analysis; IDWe3.5n7 means: Inverse distance weighting with exponent 3.5 and 7 neighbors; OLS: Ordinary Least Squares; WLS: Weighted Least Squares; Al: aluminum; C: carbon; Ca: calcium; Ca/CTC: calcium adsorbed on CTC in %; H+Al/CTC: aluminum more hydrogen adsorbed on CTC in %; H+Al: potential acidity; K: Potassium; K/CTC: potassium adsorbed on CTC in %; m%: aluminum saturation; OM: organic matter; P: phosphorus; pH: the potential of hydrogen; pH SMP: pH of buffer solution Shoemaker-McLean-Pratt; SB: sum of basis; V%: base saturation.

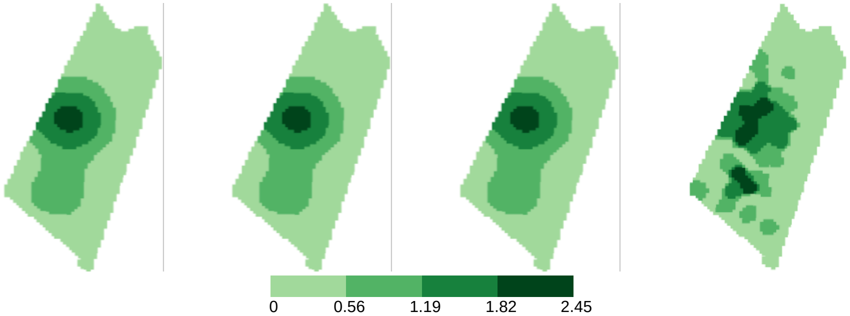
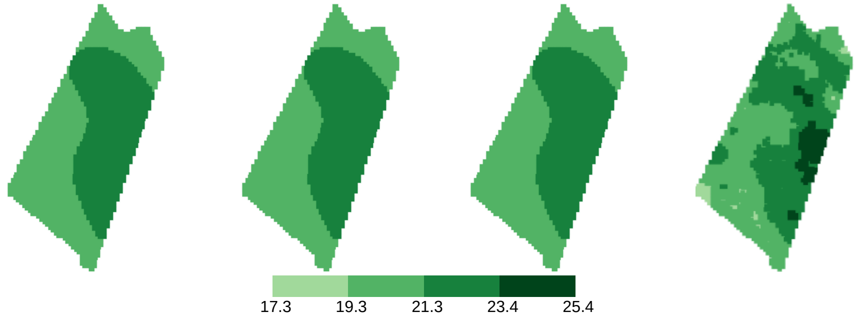
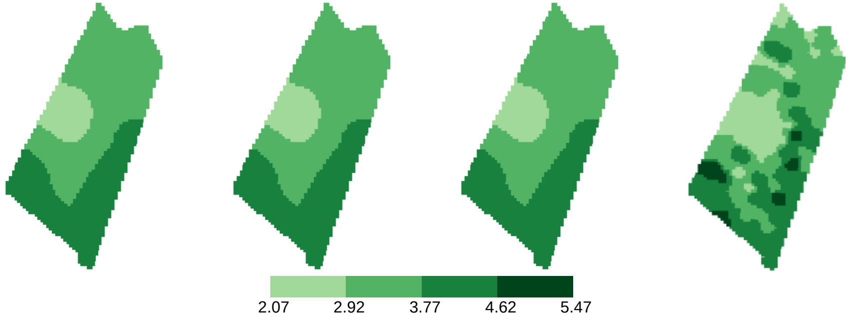
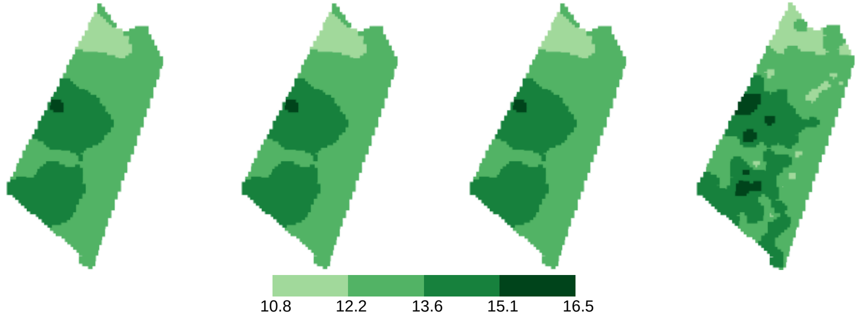
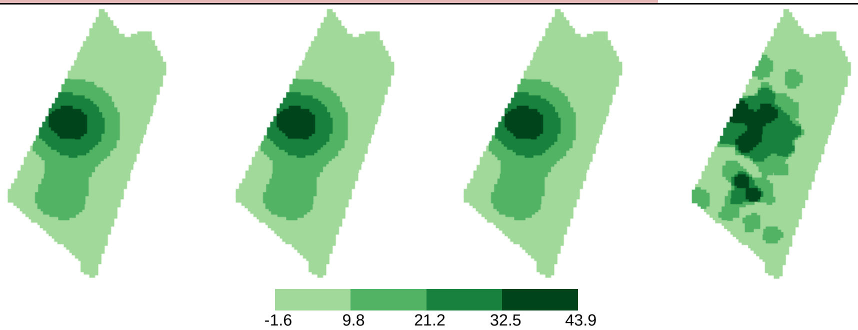
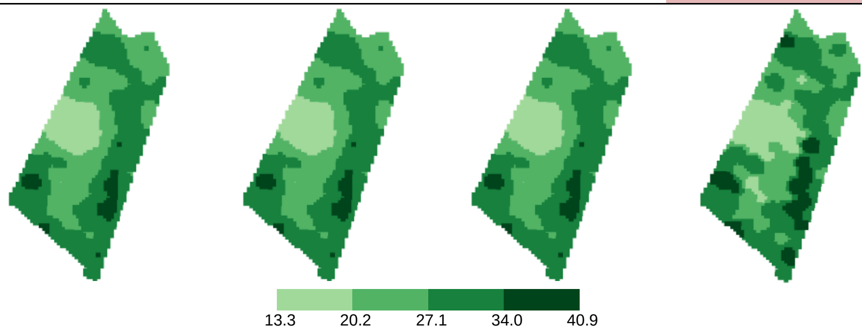
The variables OM and Silt, from Field A-2019, and C, H+Al, P, pH SMP, SB, and V%, from Field B-2015 registered that Method 1 had considered OK as the best interpolator, and, after applying the selection criteria by Methods 2 and 3, it started to consider IDW as the best interpolator. Most of these variables had all geostatistical models eliminated after applying the selection criteria, except for variables SB and pH SMP (by Method 3) from Field B-2015.

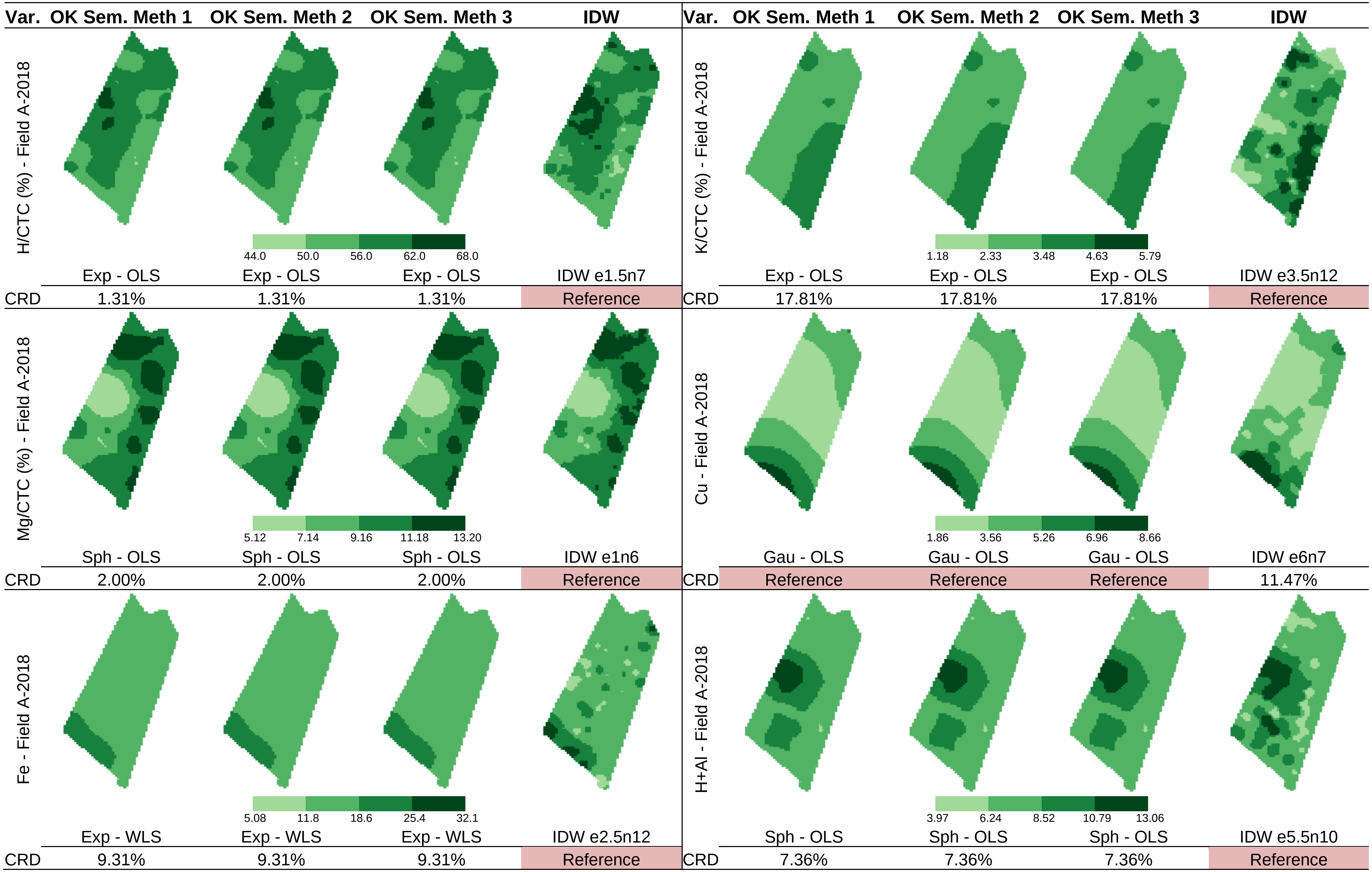
Even with eliminating inappropriate geostatistical models, K, from fields A-2018 and A-2019, and Ca and Ca/CTC, from Field B-2015, kept IDW as the best interpolator. The other variables, Al, K/CTC, H+Al/CTC, H+Al, m%, pH CaCl2, pH SMP, V%, and Clay, from field A-2019, and pH CaCl2, from Field B-2015, kept OK as the best interpolator, as selected by method 1. However, there was the selection of other geostatistical models after selection by Methods 2 and 3.

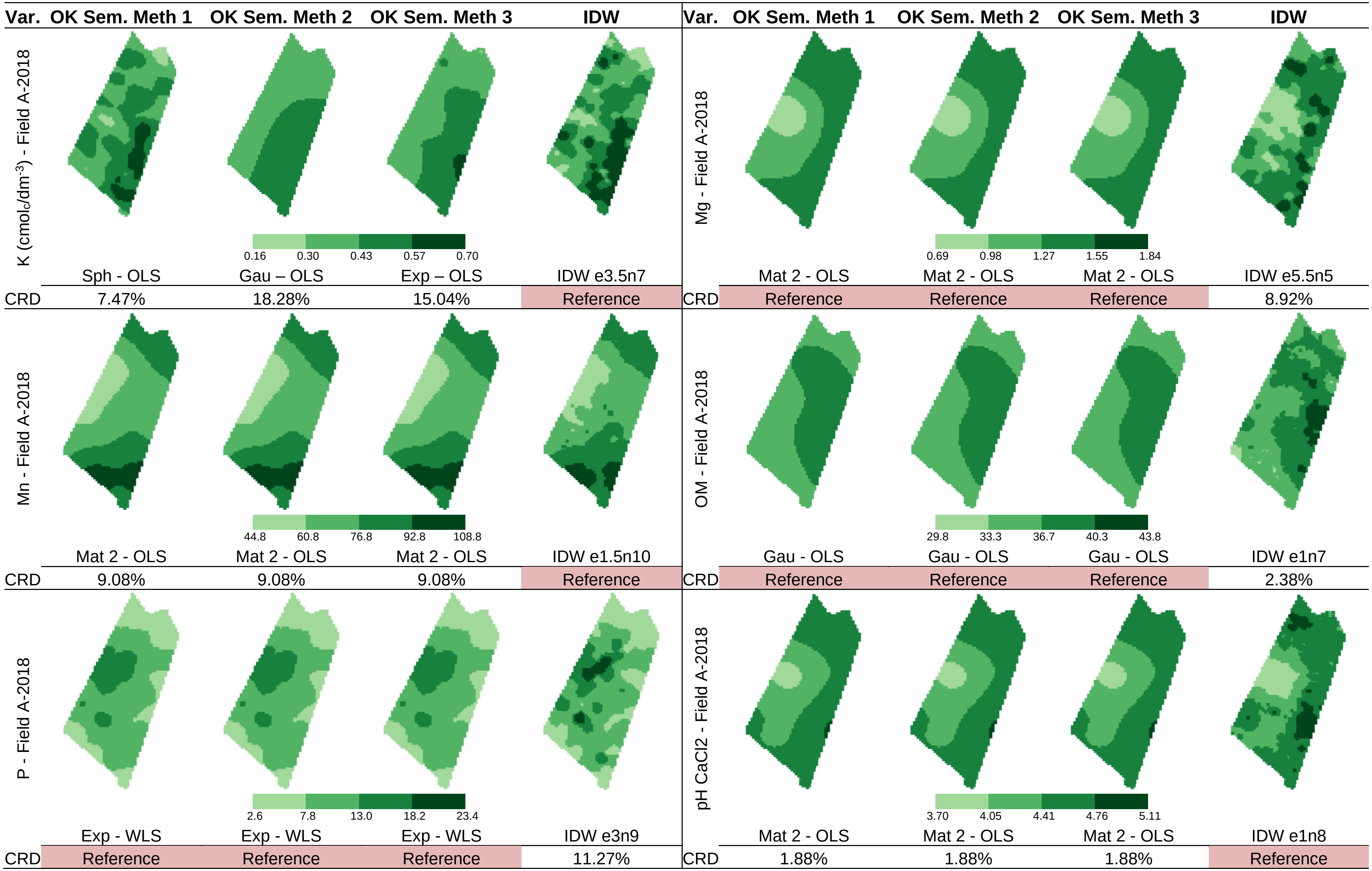
iv. Thematic maps

Thematic maps (TMs, Table 14) were generated by OK using the semivariogram selected by each of three methods and IDW with its best interpolator. The variables are the same as in Table 13. The best interpolator was considered the one selected with Method 3.

Table 14 Comparison of thematic maps created by OK using the semivariogram selected by each of three methods and IDW with its best interpolator

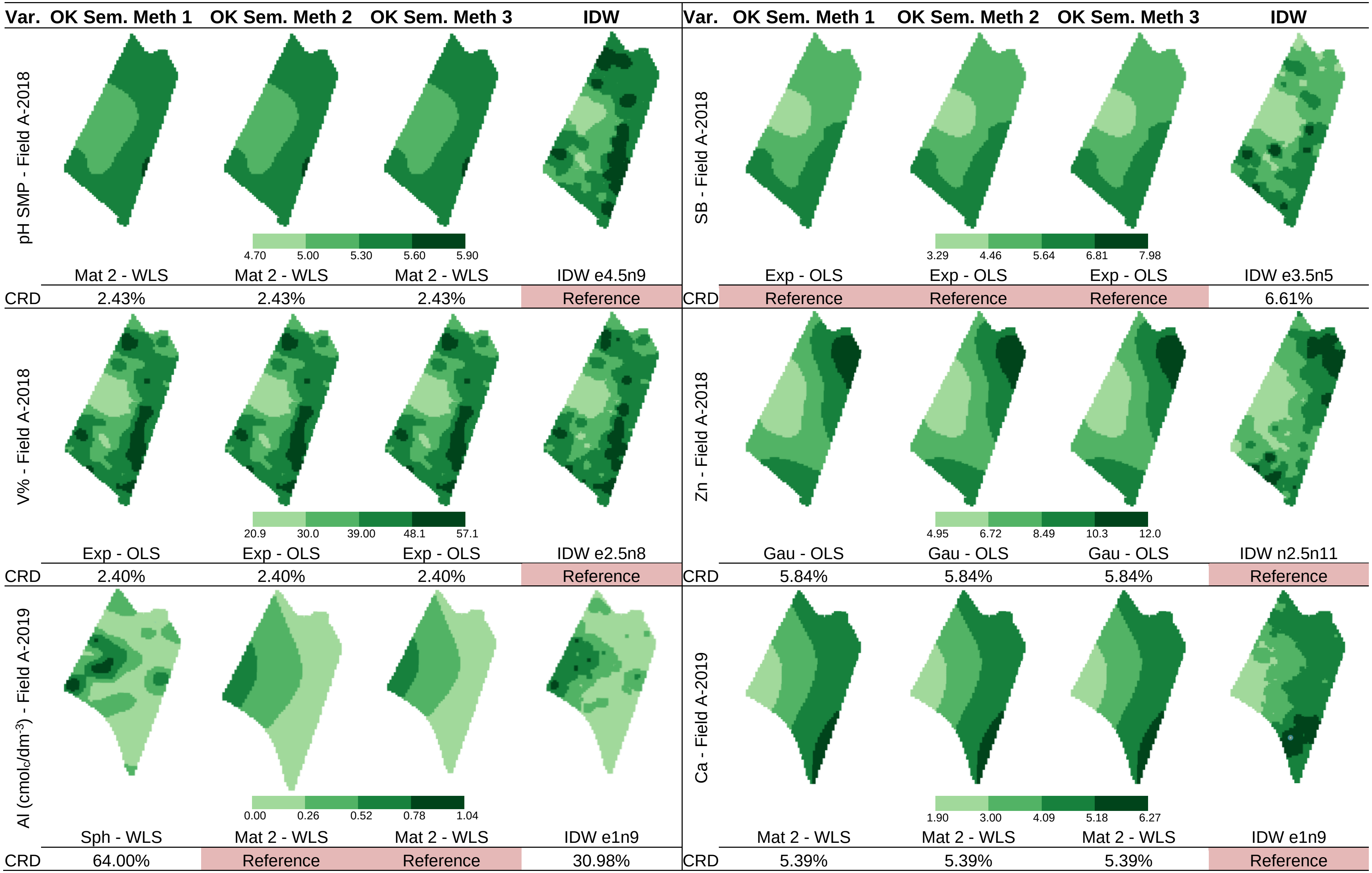
Var.	OK Sem. Meth 1	OK Sem. Meth 2	OK Sem. Meth 3	IDW	Var.	OK Sem. Meth 1	OK Sem. Meth 2	OK Sem. Meth 3	IDW
Al - Field A-2018					C - Field A-2018				
	Gau - WLS	Gau - WLS	Gau - WLS	IDW e5n5		Gau - OLS	Gau - OLS	Gau - OLS	IDW e1n7
CRD	Reference	Reference	Reference	59.72%	CRD	Reference	Reference	Reference	2.38%
Ca - Field A-2018					CTC - Field A-2018				
	Mat 1 - OLS	Mat 1 - OLS	Mat 1 - OLS	IDW e4.5n5		Sph - OLS	Sph - OLS	Sph - OLS	IDW e3n11
CRD	Reference	Reference	Reference	8.23%	CRD	2.25%	2.25%	2.25%	Reference
Al/CTC (%) - Field A-2018					Ca/CTC (%) - Field A-2018				
	Gau - WLS	Gau - WLS	Gau - WLS	IDW e5n5		Exp - OLS	Exp - OLS	Exp - OLS	IDW e4n4
CRD	Reference	Reference	Reference	74.95%	CRD	5.30%	5.30%	5.30%	Reference

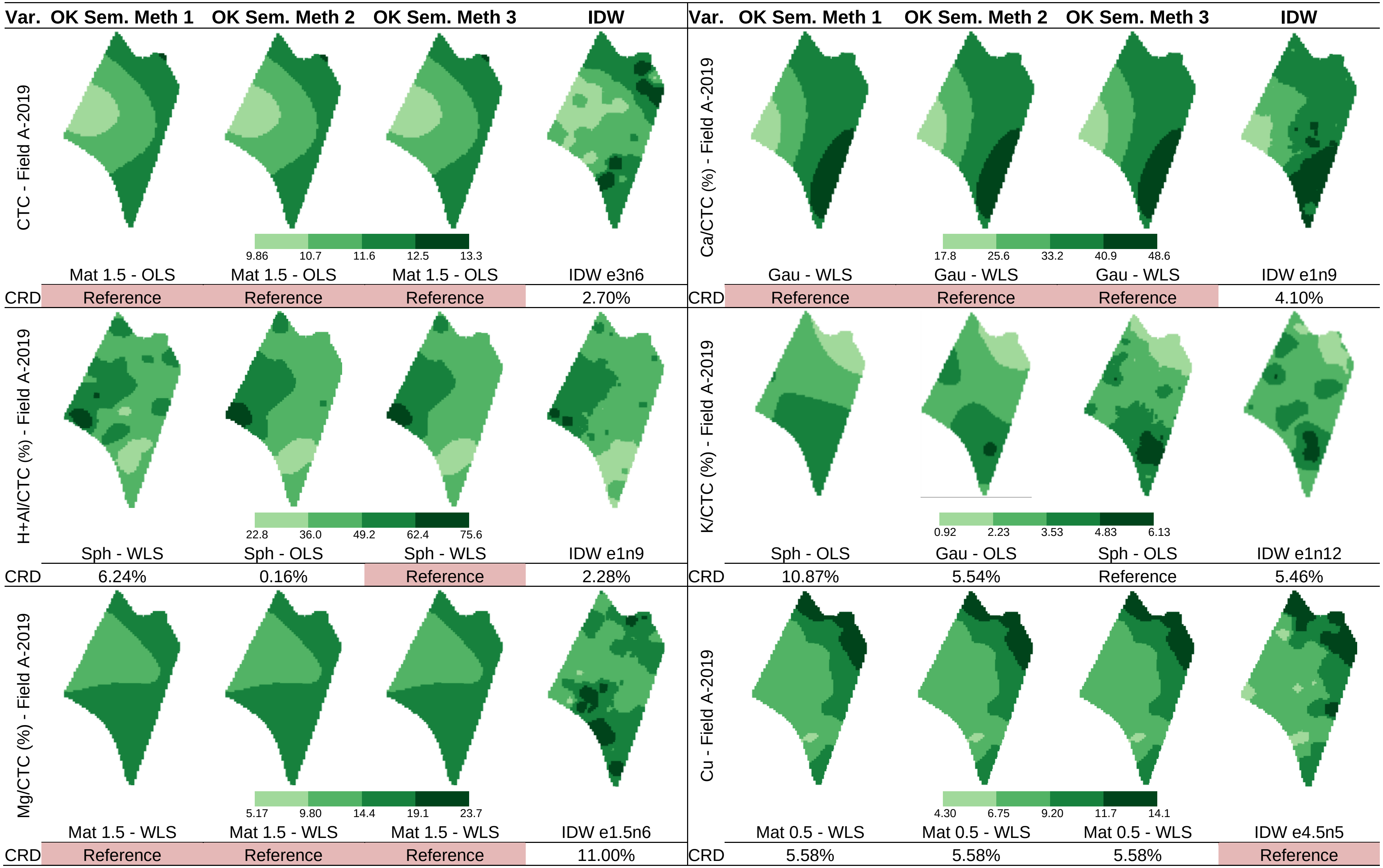

Var. OK Sem. Meth 1 OK Sem. Meth 2 OK Sem. Meth 3 IDW
H/CTC (%) - Field A-2018
44.0 50.0 56.0 62.0 68.0
Exp - OLS Exp - OLS Exp - OLS IDW e1.5n7
CRD 1.31% 1.31% 1.31% Reference
Mg/CTC (%) - Field A-2018
5.12 7.14 9.16 11.18 13.20
Sph - OLS Sph - OLS Sph - OLS IDW e1n6
CRD 2.00% 2.00% 2.00% Reference
Fe - Field A-2018
5.08 11.8 18.6 25.4 32.1
Exp - WLS Exp - WLS Exp - WLS IDW e2.5n12
CRD 9.31% 9.31% 9.31% Reference
Var. OK Sem. Meth 1 OK Sem. Meth 2 OK Sem. Meth 3 IDW
K/CTC (%) - Field A-2018
1.18 2.33 3.48 4.63 5.79
Exp - OLS Exp - OLS Exp - OLS IDW e3.5n12
CRD 17.81% 17.81% 17.81% Reference
Cu - Field A-2018
1.86 3.56 5.26 6.96 8.66
Gau - OLS Gau - OLS Gau - OLS IDW e6n7
CRD Reference Reference Reference 11.47%
H+Al - Field A-2018
3.97 6.24 8.52 10.79 13.06
Sph - OLS Sph - OLS Sph - OLS IDW e5.5n10
CRD 7.36% 7.36% 7.36% Reference



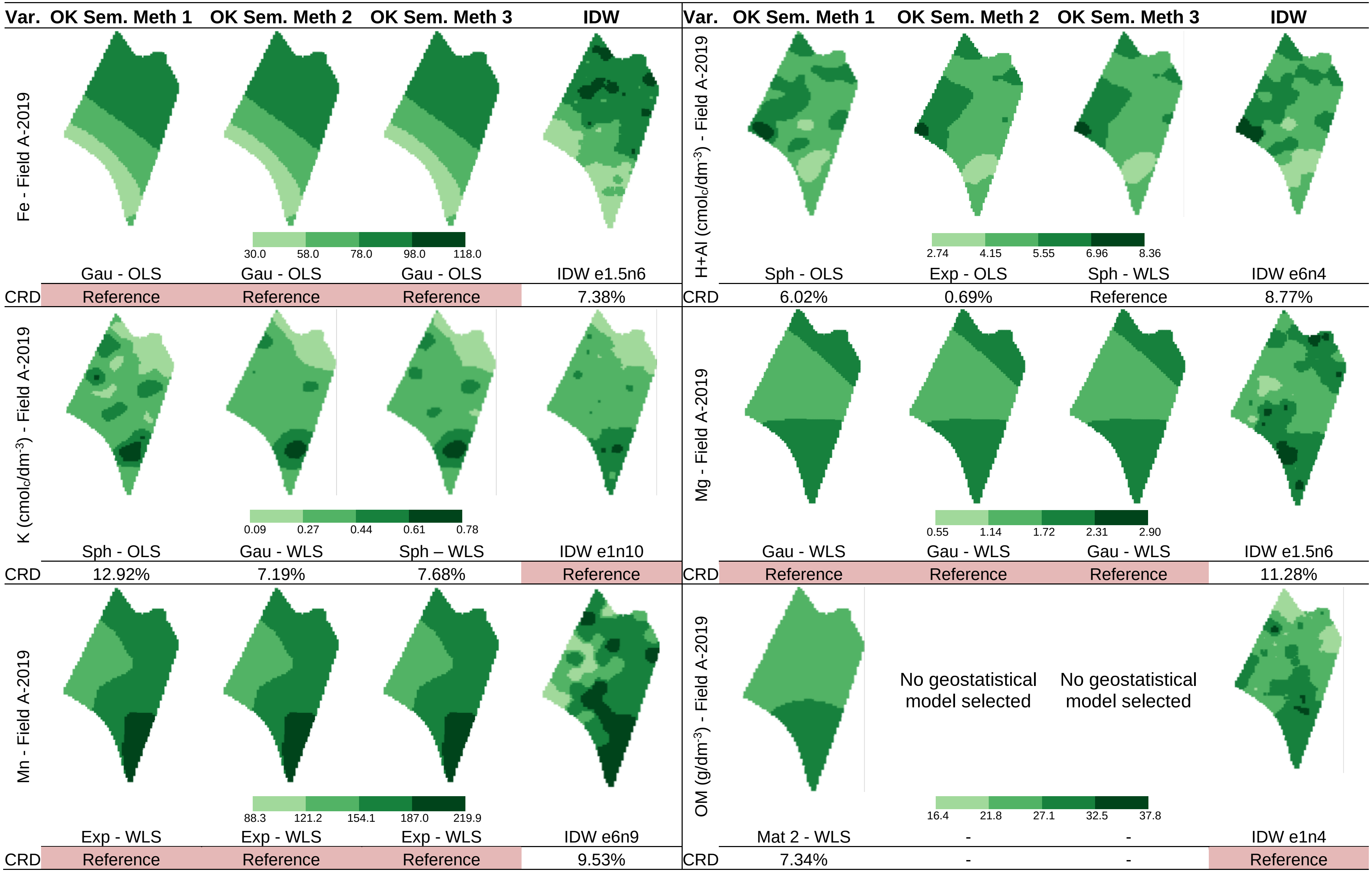
Var.	OK Sem. Meth 1	OK Sem. Meth 2	OK Sem. Meth 3	IDW
K ($cmol_c/dm^{-3}$) - Field A-2018	Sph - OLS	Gau – OLS	Exp – OLS	IDW e3.5n7
CRD	7.47%	18.28%	15.04%	Reference
Mn - Field A-2018	Mat 2 - OLS	Mat 2 - OLS	Mat 2 - OLS	IDW e1.5n10
CRD	9.08%	9.08%	9.08%	Reference
P - Field A-2018	Exp - WLS	Exp - WLS	Exp - WLS	IDW e3n9
CRD	Reference	Reference	Reference	11.27%

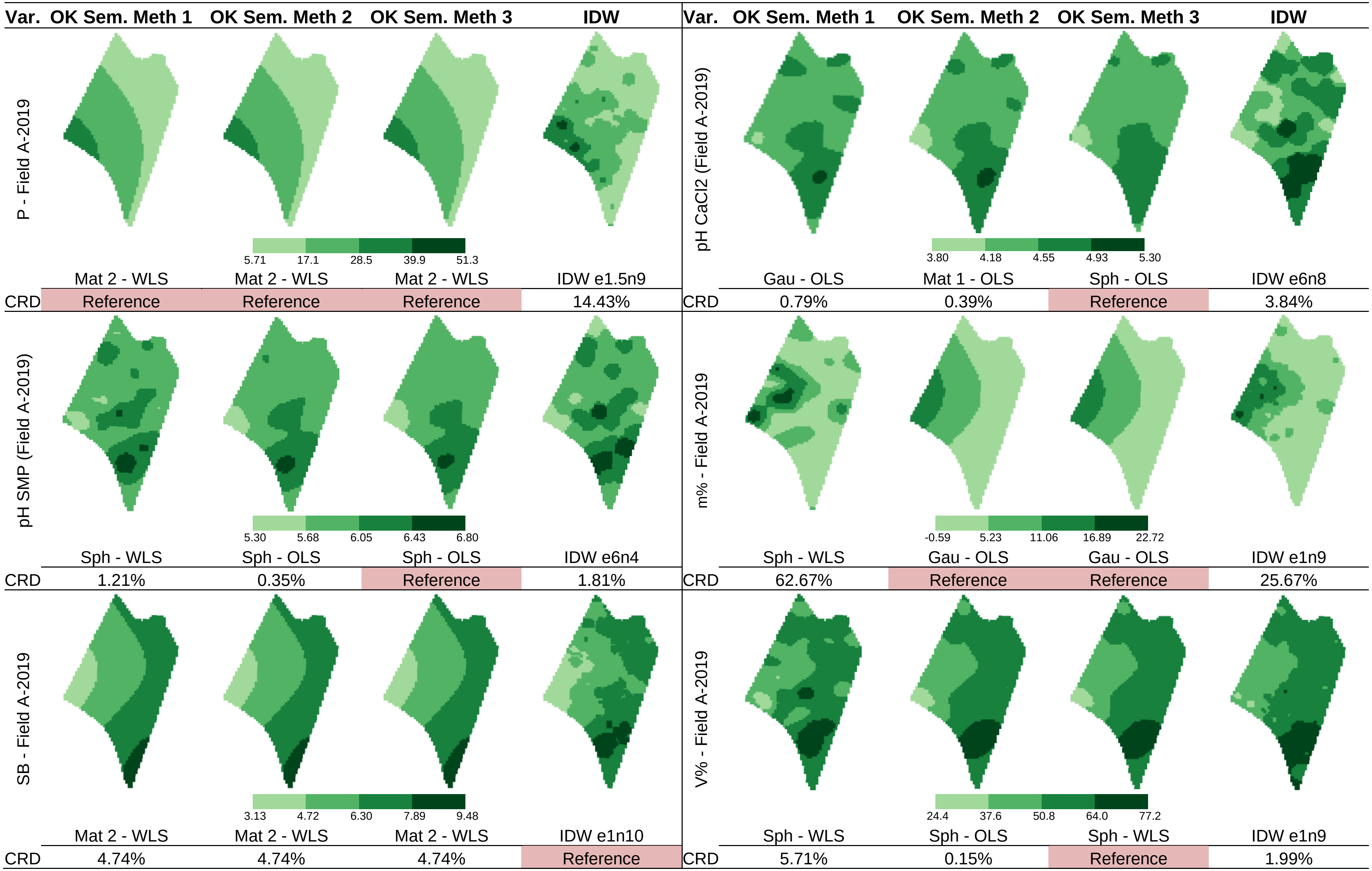
Var.	OK Sem. Meth 1	OK Sem. Meth 2	OK Sem. Meth 3	IDW
Mg - Field A-2018	Mat 2 - OLS	Mat 2 - OLS	Mat 2 - OLS	IDW e5.5n5
CRD	Reference	Reference	Reference	8.92%
OM - Field A-2018	Gau - OLS	Gau - OLS	Gau - OLS	IDW e1n7
CRD	Reference	Reference	Reference	2.38%
pH CaCl2 - Field A-2018	Mat 2 - OLS	Mat 2 - OLS	Mat 2 - OLS	IDW e1n8
CRD	1.88%	1.88%	1.88%	Reference

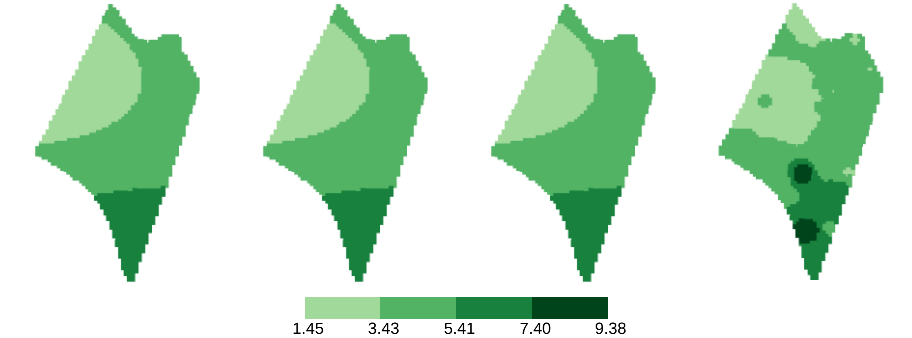

Var. OK Sem. Meth 1 OK Sem. Meth 2 OK Sem. Meth 3 IDW
pH SMP - Field A-2018
4.70 5.00 5.30 5.60 5.90
Mat 2 - WLS Mat 2 - WLS Mat 2 - WLS IDW e4.5n9
CRD 2.43% 2.43% 2.43% Reference
V% - Field A-2018
20.9 30.0 39.00 48.1 57.1
Exp - OLS Exp - OLS Exp - OLS IDW e2.5n8
CRD 2.40% 2.40% 2.40% Reference
Al (cmolc/dm-3) - Field A-2019
0.00 0.26 0.52 0.78 1.04
Sph - WLS Mat 2 - WLS Mat 2 - WLS IDW e1n9
CRD 64.00% Reference Reference 30.98%
Var. OK Sem. Meth 1 OK Sem. Meth 2 OK Sem. Meth 3 IDW
SB - Field A-2018
3.29 4.46 5.64 6.81 7.98
Exp - OLS Exp - OLS Exp - OLS IDW e3.5n5
CRD Reference Reference Reference 6.61%
Zn - Field A-2018
4.95 6.72 8.49 10.3 12.0
Gau - OLS Gau - OLS Gau - OLS IDW n2.5n11
CRD 5.84% 5.84% 5.84% Reference
Ca - Field A-2019
1.90 3.00 4.09 5.18 6.27
Mat 2 - WLS Mat 2 - WLS Mat 2 - WLS IDW e1n9
CRD 5.39% 5.39% 5.39% Reference

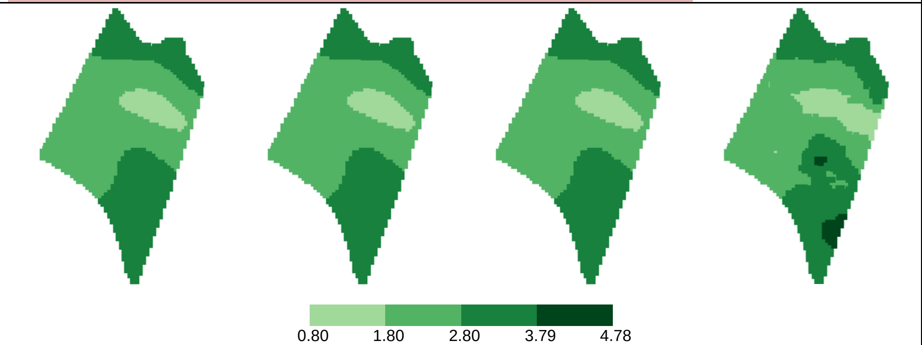


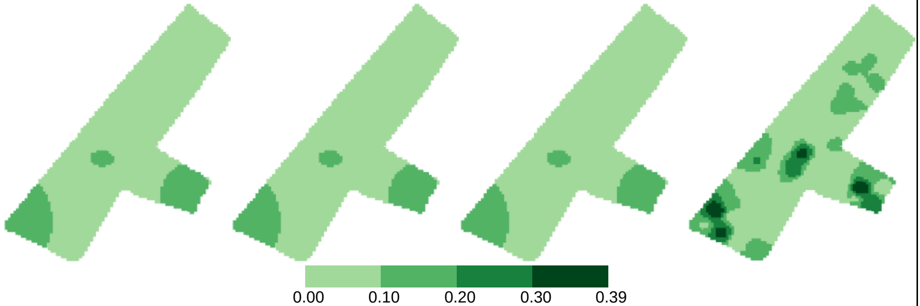
Var. OK Sem. Meth 1 OK Sem. Meth 2 OK Sem. Meth 3 IDW
CTC - Field A-2019
9.86 10.7 11.6 12.5 13.3
Mat 1.5 - OLS Mat 1.5 - OLS Mat 1.5 - OLS IDW e3n6
CRD Reference Reference Reference 2.70%
H+Al/CTC (%) - Field A-2019
22.8 36.0 49.2 62.4 75.6
Sph - WLS Sph - OLS Sph - WLS IDW e1n9
CRD 6.24% 0.16% Reference 2.28%
Mg/CTC (%) - Field A-2019
5.17 9.80 14.4 19.1 23.7
Mat 1.5 - WLS Mat 1.5 - WLS Mat 1.5 - WLS IDW e1.5n6
CRD Reference Reference Reference 11.00%
Var. OK Sem. Meth 1 OK Sem. Meth 2 OK Sem. Meth 3 IDW
Ca/CTC (%) - Field A-2019
17.8 25.6 33.2 40.9 48.6
Gau - WLS Gau - WLS Gau - WLS IDW e1n9
CRD Reference Reference Reference 4.10%
K/CTC (%) - Field A-2019
0.92 2.23 3.53 4.83 6.13
Sph - OLS Gau - OLS Sph - OLS IDW e1n12
CRD 10.87% 5.54% Reference 5.46%
Cu - Field A-2019
4.30 6.75 9.20 11.7 14.1
Mat 0.5 - WLS Mat 0.5 - WLS Mat 0.5 - WLS IDW e4.5n5
CRD 5.58% 5.58% 5.58% Reference


Var. OK Sem. Meth 1 OK Sem. Meth 2 OK Sem. Meth 3 IDW
Fe - Field A-2019
30.0 58.0 78.0 98.0 118.0
Gau - OLS Gau - OLS Gau - OLS IDW e1.5n6
CRD Reference Reference Reference 7.38%
H+Al (cmolc/dm-3) - Field A-2019
2.74 4.15 5.55 6.96 8.36
Sph - OLS Exp - OLS Sph - WLS IDW e6n4
CRD 6.02% 0.69% Reference 8.77%
K (cmolc/dm-3) - Field A-2019
0.09 0.27 0.44 0.61 0.78
Sph - OLS Gau - WLS Sph – WLS IDW e1n10
CRD 12.92% 7.19% 7.68% Reference
Mg - Field A-2019
0.55 1.14 1.72 2.31 2.90
Gau - WLS Gau - WLS Gau - WLS IDW e1.5n6
CRD Reference Reference Reference 11.28%
Mn - Field A-2019
88.3 121.2 154.1 187.0 219.9
Exp - WLS Exp - WLS Exp - WLS IDW e6n9
CRD Reference Reference Reference 9.53%
OM (g/dm-3) - Field A-2019
No geostatistical model selected
No geostatistical model selected
16.4 21.8 27.1 32.5 37.8
Mat 2 - WLS - - IDW e1n4
CRD 7.34% - - Reference

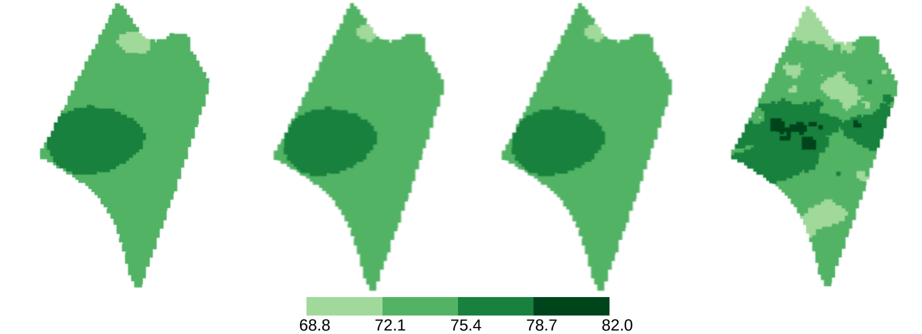

Var. OK Sem. Meth 1 OK Sem. Meth 2 OK Sem. Meth 3 IDW
Var. OK Sem. Meth 1 OK Sem. Meth 2 OK Sem. Meth 3 IDW
P - Field A-2019
5.71 17.1 28.5 39.9 51.3
Mat 2 - WLS Mat 2 - WLS Mat 2 - WLS IDW e1.5n9
CRD Reference Reference Reference 14.43%
pH CaCl2 (Field A-2019)
3.80 4.18 4.55 4.93 5.30
Gau - OLS Mat 1 - OLS Sph - OLS IDW e6n8
CRD 0.79% 0.39% Reference 3.84%
pH SMP (Field A-2019)
5.30 5.68 6.05 6.43 6.80
Sph - WLS Sph - OLS Sph - OLS IDW e6n4
CRD 1.21% 0.35% Reference 1.81%
m% - Field A-2019
-0.59 5.23 11.06 16.89 22.72
Sph - WLS Gau - OLS Gau - OLS IDW e1n9
CRD 62.67% Reference Reference 25.67%
SB - Field A-2019
3.13 4.72 6.30 7.89 9.48
Mat 2 - WLS Mat 2 - WLS Mat 2 - WLS IDW e1n10
CRD 4.74% 4.74% 4.74% Reference
V% - Field A-2019
24.4 37.6 50.8 64.0 77.2
Sph - WLS Sph - OLS Sph - WLS IDW e1n9
CRD 5.71% 0.15% Reference 1.99%

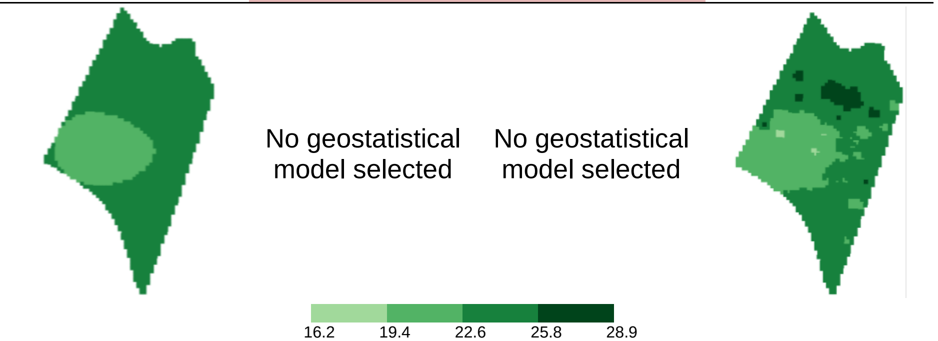


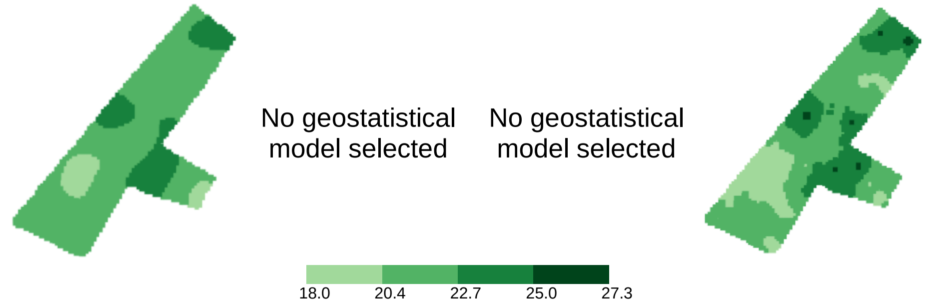




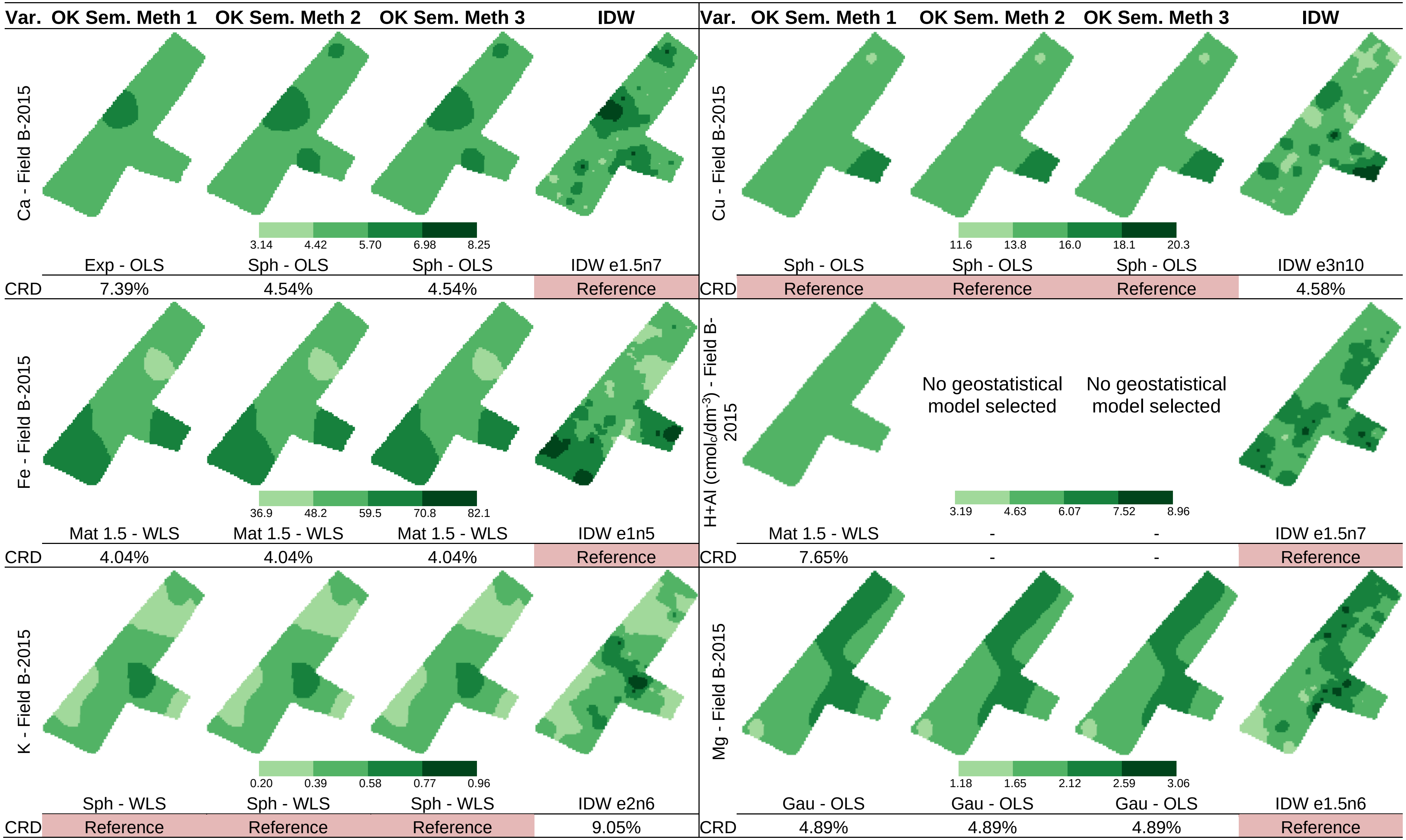
Var.	OK Sem. Meth 1	OK Sem. Meth 2	OK Sem. Meth 3	IDW
Zn - Field A-2019	Gau - OLS	Gau - OLS	Gau - OLS	IDW e2.5n5
CRD	Reference	Reference	Reference	12.16%
Sand- Field A-2019	Gau - WLS	Gau - WLS	Gau - WLS	IDW e1n4
CRD	Reference	Reference	Reference	6.24%
Al - Field B-2015	Gau - WLS	Gau - WLS	Gau - WLS	IDW e2.5n4
CRD	Reference	Reference	Reference	62.34%

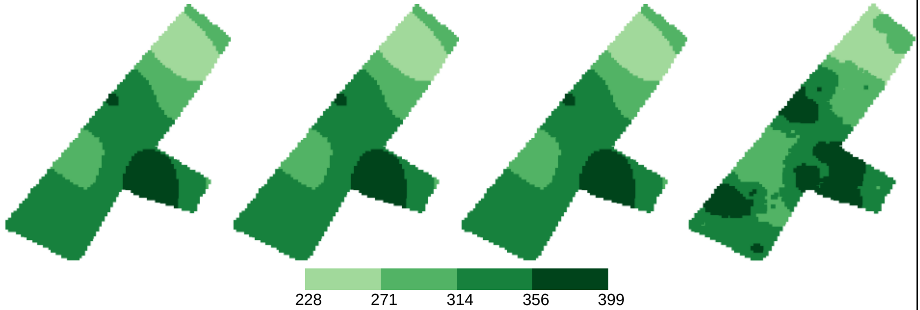
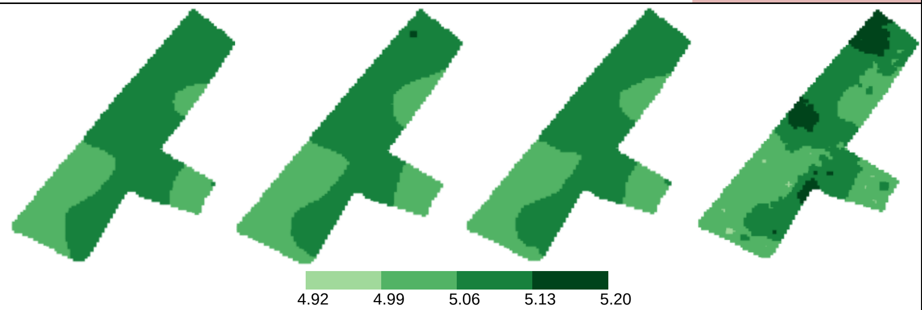
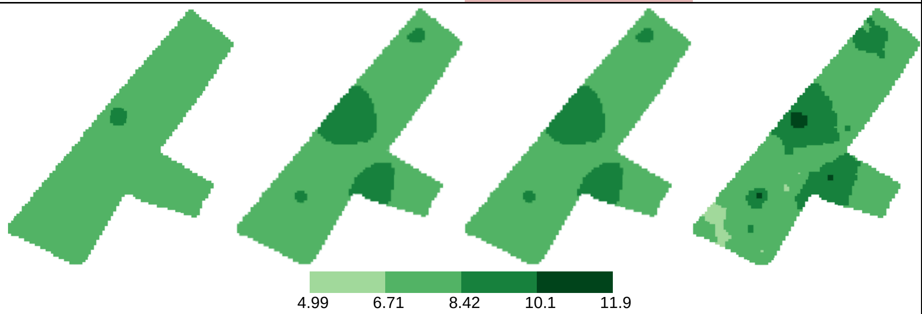


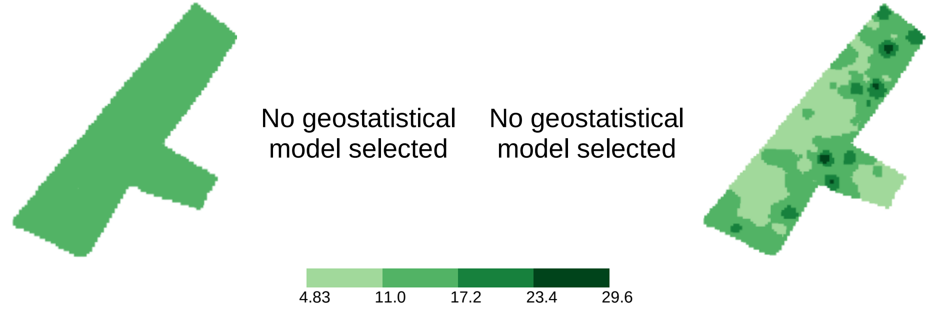
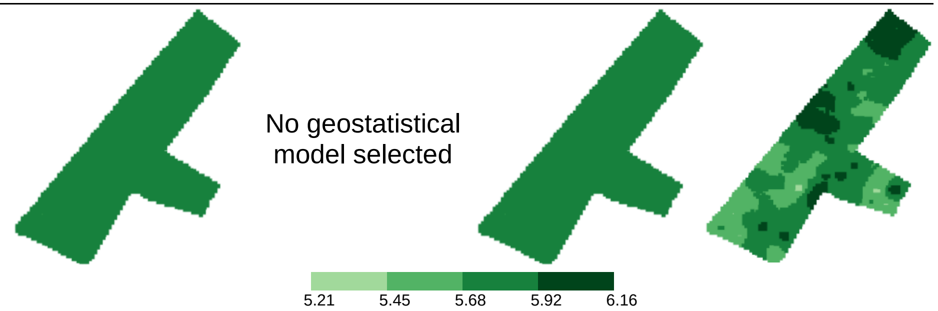
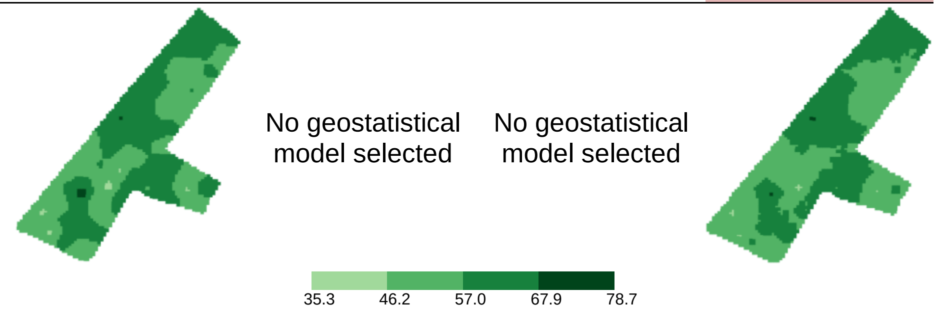


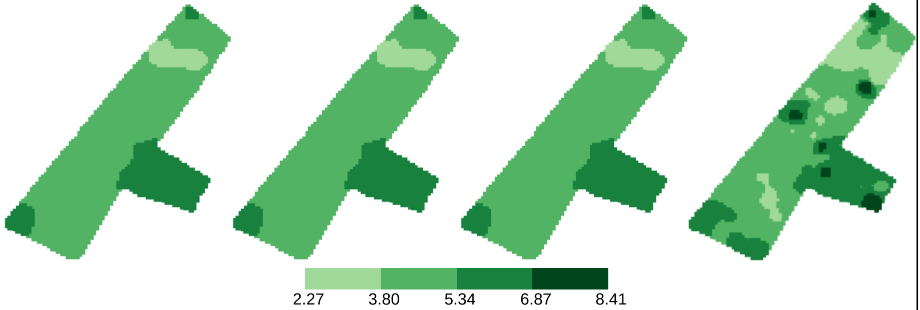
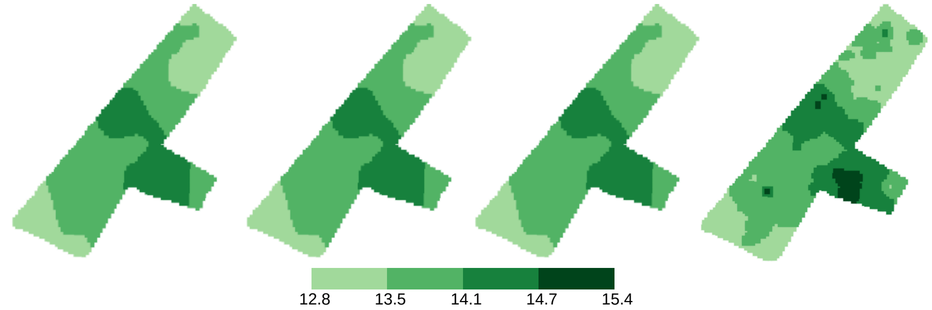
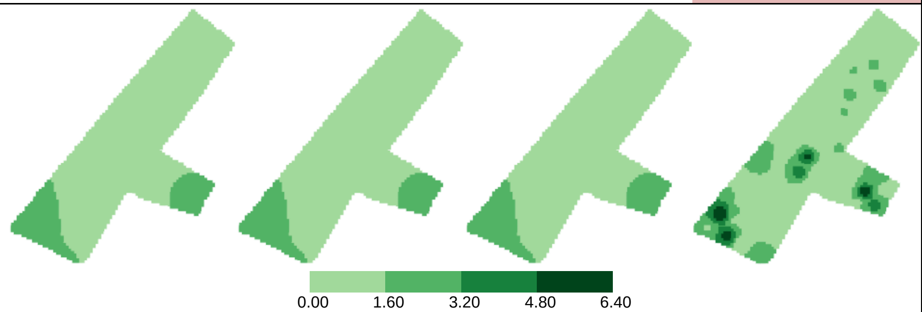
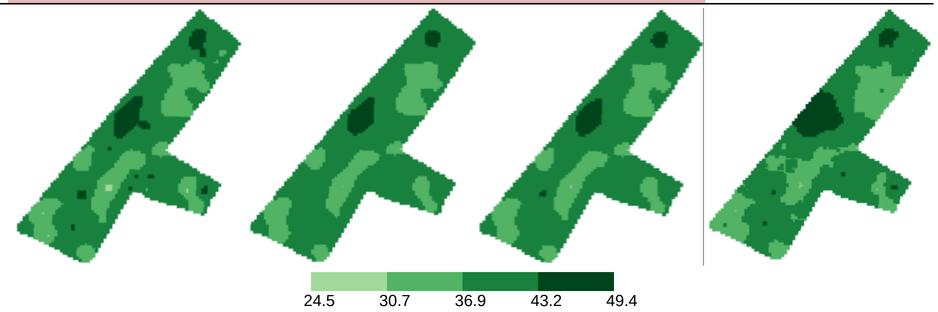
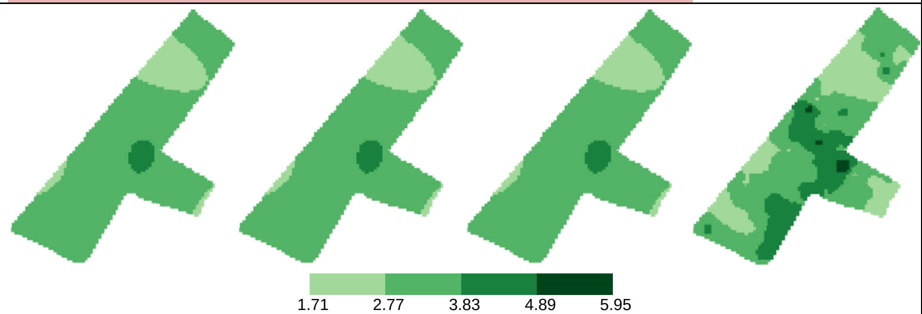
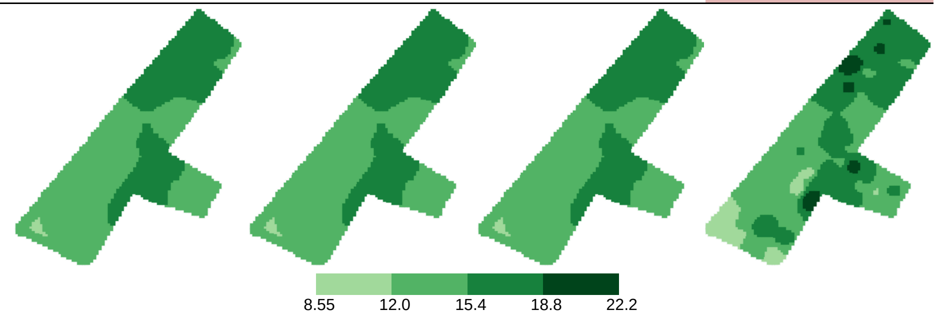


Var.	OK Sem. Meth 1	OK Sem. Meth 2	OK Sem. Meth 3	IDW
Clay (%) - Field A-2019	Gau - OLS	Mat 2 - OLS	Mat 2 - OLS	IDW e1n5
CRD	0.09%	Reference	Reference	1.16%
Silt (%) - Field A-2019	Gau - OLS	-	-	IDW e1n10
CRD	2.68%	-	-	Reference
C - Field B-2015	Gau - WLS	-	-	IDW e1n7
CRD	1.83%	-	-	Reference


Var.
OK Sem. Meth 1
OK Sem. Meth 2
OK Sem. Meth 3
IDW
Var.
OK Sem. Meth 1
OK Sem. Meth 2
OK Sem. Meth 3
IDW
Ca - Field B-2015
3.14 4.42 5.70 6.98 8.25
Exp - OLS
Sph - OLS
Sph - OLS
IDW e1.5n7
CRD
7.39%
4.54%
4.54%
Reference
Cu - Field B-2015
11.6 13.8 16.0 18.1 20.3
Sph - OLS
Sph - OLS
Sph - OLS
IDW e3n10
CRD
Reference
Reference
Reference
4.58%
Fe - Field B-2015
36.9 48.2 59.5 70.8 82.1
Mat 1.5 - WLS
Mat 1.5 - WLS
Mat 1.5 - WLS
IDW e1n5
CRD
4.04%
4.04%
4.04%
Reference
H+Al (cmolc/dm-3) - Field B-2015
No geostatistical model selected
No geostatistical model selected
3.19 4.63 6.07 7.52 8.96
Mat 1.5 - WLS
-
-
IDW e1.5n7
CRD
7.65%
-
-
Reference
K - Field B-2015
0.20 0.39 0.58 0.77 0.96
Sph - WLS
Sph - WLS
Sph - WLS
IDW e2n6
CRD
Reference
Reference
Reference
9.05%
Mg - Field B-2015
1.18 1.65 2.12 2.59 3.06
Gau - OLS
Gau - OLS
Gau - OLS
IDW e1.5n6
CRD
4.89%
4.89%
4.89%
Reference

Var.	OK Sem. Meth 1	OK Sem. Meth 2	OK Sem. Meth 3	IDW
Mn - Field B-2015				
	Mat 2 - OLS	Mat 2 - OLS	Mat 2 - OLS	IDW e1n4
CRD	3.04%	3.04%	3.04%	Reference
pH CaCl2 - Field B-2015				
	Gau - WLS	Gau - OLS	Sph - WLS	IDW e1n7
CRD	0.28%	0.23%	Reference	1.47%
SB - Field B-2015				
	Gau - OLS	Sph - WLS	Sph - WLS	IDW e1n7
CRD	6.06%	2.73%	2.73%	Reference

Var.	OK Sem. Meth 1	OK Sem. Meth 2	OK Sem. Meth 3	IDW
P (mg/dm^3) - Field B-2015				
	Exp - WLS	-	-	IDW e2n7
CRD	23.60%	-	-	Reference
pH SMP - Field B-2015				
	Mat 2 - WLS	-	Exp - OLS	IDW e1.5n7
CRD	1.65%	-	1.30%	Reference
V% - Field B-2015				
	Exp - WLS	-	-	IDW e1n7
CRD	2.47%	-	-	Reference

Var.	OK Sem. Meth 1	OK Sem. Meth 2	OK Sem. Meth 3	IDW	Var.	OK Sem. Meth 1	OK Sem. Meth 2	OK Sem. Meth 3	IDW
Zn - Field B-2015					CTC - Field B-2015				
	Exp - OLS	Exp - OLS	Exp - OLS	IDW e2.5n12		Mat 0.5 - OLS	Mat 0.5 - OLS	Mat 0.5 - OLS	IDW e1n7
CRD	8.59%	8.59%	8.59%	Reference	CRD	Reference	Reference	Reference	0.87%
Al/CTC - Field B-2015					Ca/CTC (%) - Field B-2015				
	Exp - OLS	Exp - OLS	Exp - OLS	IDW e2n9		Exp - WLS	Mat 2 - WLS	Mat 1.5 - WLS	IDW e1n8
CRD	Reference	Reference	Reference	48.99%	CRD	2.38%	2.27%	2.24%	Reference
K/CTC - Field B-2015					Mg/CTC - Field B-2015				
	Gau - OLS	Gau - OLS	Gau - OLS	IDW e1.5n8		Sph - WLS	Sph - WLS	Sph - WLS	IDW e2.5n6
CRD	Reference	Reference	Reference	9.96%	CRD	6.14%	6.14%	6.14%	Reference

Var.: Variable; OK Sem. Methods *(1, 2 or 3)* means: Ordinary Kriging using the semivariogram selected by method 1 (Only the best ISI), 2 (The three criteria are applied after geostatistics analysis), or 3 (The three criteria are applied during geostatistics analysis); Sph: Spherical; Exp: Exponential;

Gau: Gaussian: Mat: Matérn; IDW e3.5n7 means: Inverse distance weighting with exponent 3.5 and 7 neighbors; OLS: Ordinary Least Squares; WLS: Weighted Least Squares; CRD: Coefficient of Relative Deviation.
Al: aluminum; C: Carbon; Ca: calcium; CTC: cation exchange capacity; Al/CTC: aluminum adsorbed on CTC in %; Ca/CTC: calcium adsorbed on CTC in %; H/CTC: hydrogen adsorbed on CTC in %; H+Al/CTC: aluminum more hydrogen adsorbed on CTC in %; K/CTC: potassium adsorbed on CTC in %; Mg/CTC: magnesium adsorbed on CTC in %; Cu: copper; Fe: iron; H+Al: potential acidity; K: potassium; Mg: magnesium; Mn: manganese; OM: organic matter; P: phosphorus; pH: the potential of hydrogen; pH SMP: pH of buffer solution Shoemaker-McLean-Pratt; SB: the sum of basis; V%: base saturation; Zn: zinc.

Using CRD to compare the maps generated by the interpolator selected by Method 3 (IDW or OK) versus the best semivariogram model indicated by Method 1 (Fig. 8), it can be seen that:

- the selection of other interpolator parameters can result in large differences among the maps. In variable Al, from area A-2019, the best interpolator model, selected by Method 3 (Matérn 2 - OLS), deviated by 64% from the map selected by Method 1 (Spherical - WLS).
- the difference was below 5% in nine variables.
- the difference was from 5% to 10% in eight variables.
- over 10% in five variables.

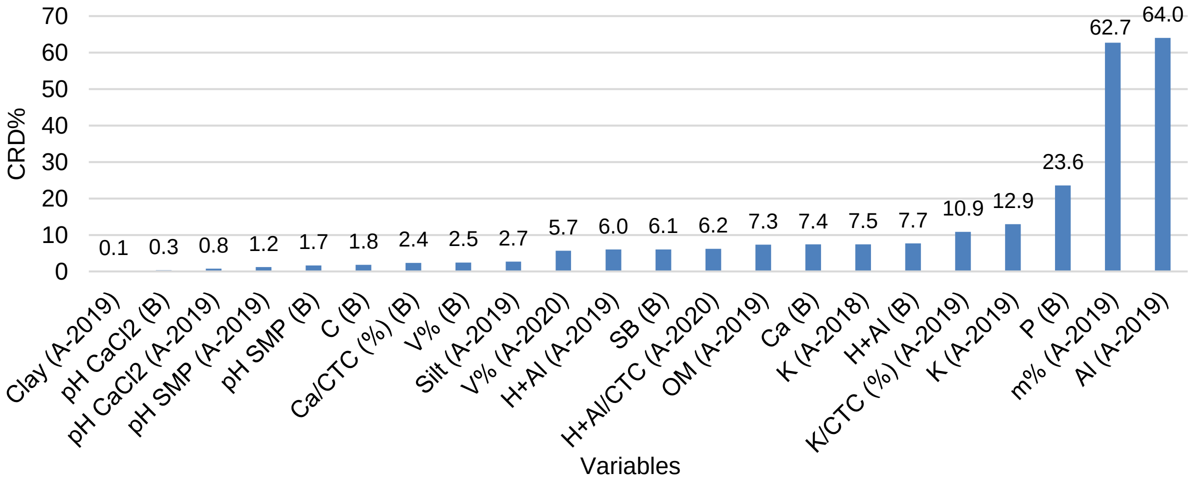


Fig. 8 The coefficient of relative deviation (CRD) between the interpolator selected by method 3 (IDW or OK) versus the best semivariogram model indicated by method 1.

When comparing the maps generated by the interpolator selected by method 3 (IDW or OK) versus the best semivariogram model indicated by method 2 (Fig. 9), it can be seen that:

- The most significant difference was observed in variable K (Field A-2018; 18%).
- The difference was below 5% in twelve variables.
- The difference between 5% and 10% in three variables.

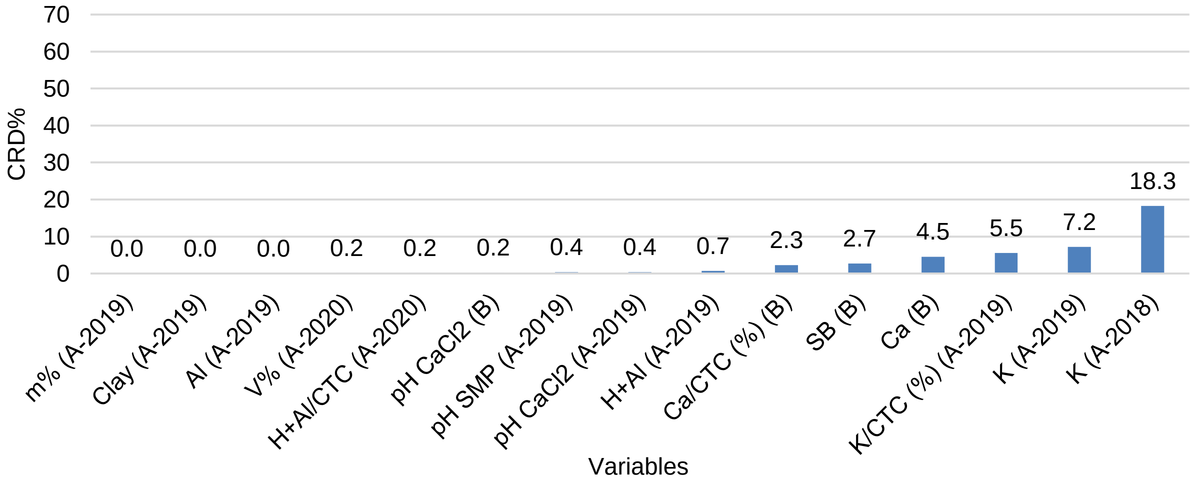


Fig. 9 The coefficient of relative deviation (CRD) between the interpolator selected by method 3 (IDW or OK) versus the best semivariogram model indicated by method 2.

Our study analyzed 66 cases, and in 31 of them, IDW outperformed OK. Consequently, in 35 cases, Kriging was better than IDW. These results confirm the ones recorded by Mueller et al. (2004) that for sample datasets with semivariograms that did not indicate spatial structure, IDW was a better choice than OK with a nugget model.

7.4 Conclusions

The inclusion of the three criteria (i) effective spatial dependence index (%ESDI) > 25%, (ii) the first semivariance significance index ($\%\gamma(1)$) < 50% and (iii) slope of the model ends index (%SMEI) > 20% improved the selection of the best interpolator using only the interpolator selection index (ISI – Bier and Souza, 2017).

The comparison carried out the methodology influence on selecting the best interpolator among the studied thematic maps using three Methods: (i) Method 1 - best ISI; (ii) Method 2 - the three criteria were applied after geostatistics analysis; Method 3 - the three criteria are applied during geostatistics analysis. Method 3 showed as the best approach. The coefficient of relative deviation (CRD) varied from 0.1 to 64% when comparing the maps generated by the three methods.

The newly proposed measurement of effective spatial dependence index (ESDI) of a semivariogram showed better performance than the usual spatial dependence index (%SDI) widely adopted in literature.

7.5 Acknowledgments

The authors would like to thank the Western Paraná State University (UNIOESTE), the Federal University of Technology of Paraná (UTFPR), the Coordination for the Upgrading of

Higher Education Personnel (CAPES), the National Council for Scientific and Technological Development (CNPq), the Itaipu Technological Park Foundation (FPTI), and the Ministry of Agriculture, Livestock and Food Supply (MAPA) for funding this project.

7.6 References

Amaral, L. R.; Justina, D. D. D. 2019. Spatial dependence degree and sampling neighborhood influence on interpolation process for fertilizer prescription maps. **Engenharia Agrícola,** 39 (special issue), pp. 85-95.

Anselin, L. 1995. Local indicators of spatial association-LISA. **Geographical Analysis**, 27 (2), pp. 93-115.

Betzek, N. M.; Souza, E. G.; Bazzi, C. L.; Schenatto K.; Gavioli, A.; Magalhães, P. S. G. 2019. Computational routines for the automatic selection of the best parameters used by interpolation methods to create thematic maps. **Computers and Electronics in Agriculture**, 157, pp. 49-62.

Bier, V. A.; Souza, E. G. 2017. Interpolation selection index for delineation of thematic maps. **Computers and Electronics in Agriculture**, 136 (1), pp. 202-209.

Biondi, F.; Myers, D. E.; Avery, C. C. 1994. Geostatistically modeling stem size and increment in an old-growth forest. **Canadian Journal of Forest Research**, 24 (7), pp. 1354-1368.

Cambardella, C. A.; Mooman, T. B.; Novak, J. M.; Parkin, T. B.; Karlen, D. L.; Turv, R. F.; Konopka, A. E. 1994. Field-scale variability of soil properties in central Iowa soil. **Soil Science Society of America Journal**, 58 (5), pp. 1501-1511.

Coelho, E. C.; Souza, E. G.; Uribe-Opazo, M. A.; Pinheiro Neto, R. 2009. Influência da densidade amostral e do tipo de interpolador na elaboração de mapas temáticos. **Acta Scientiarum**, 31 (1), pp. 165-174.

Córdoba, M. A.; Bruno, C. I.; Costa, J. L.; Peralta, N. R.; Balzarini, M. G. 2016. Protocol for multivariate homogeneous zone delineation in precision agriculture. **Biosystems Engineering**, 143, pp. 95-107.

Carroll, S. S.; Cressie, N. 1996. A comparison of geostatistical methodologies used to estimate snow water equivalent. **Journal of the American Water Resources Association**, 32 (2), pp. 267-278.

Clark, I. 1979. **Practical geostatistics**. Applied Science Publishers, London.

Coutinho, M. A. N.; Alari, F. O.; Ferreira, M. M. C.; Amaral, L. R. 2019. Influence of soil sample preparation on the quantification of NPK content via spectroscopy. **Geoderma**, 338 (2), pp. 401-409.

Cressie, N. A. C. 1993. **Statistics for Spatial Data**. Wiley-Interscience Publication, New York.

Cressie, N.; Hawkins, D. M. 1980. Robust estimation of the variogram. **Journal of the International Association of Mathematical Geology**, 12, pp. 115-125.

Dall'agnol, R. W.; Michelon, G. K.; Bazzi, C. L.; Magalhães, P. S. G.; Souza, E. G.; Betzek, N. M., Sobjak, R. 2020. Web applications for spatial analyses and thematic map generation. **Computers and Electronics in Agriculture**, 172.

Diggle, P. J., Ribeiro Jr., P. J. (Eds.). 2007. **Model-based geostatistics**. Springer, New York.

Doerge, T. A. 2000. Management Zone Concepts. **Site-Specific Management Guidelines**. Potash and Phosphate Institute. University South Dakota, Brokings.

Dowd, P. A. 1984. The variogram and kriging: Robust and resistant estimators. In: Verly, G.; David, M.; Journel, A. G.; Marechal, A. (Eds.), **Geostatistics for natural resources characterization**. Springer, Dordrecht, pp. 91-106.

Eldeiry, A. A.; Garcia, L. A. 2012. Evaluating the performance of ordinary kriging in mapping soil salinity. **Journal of Irrigation and Drainage Engineering**, 138 (12), pp. 1046-1059.

Faraco, M. A., Uribe-Opazo, M. A., Silva, E. A. A., Johann, J. A., Borssoi, J. 2008. Selection criteria of spatial variability models used in thematical maps of soil physical attributes and soybean yield. **Revista Brasileira de Ciência do Solo**, 32 (2), pp. 463-476.

Ferguson, R. B.; Hergert, G. W. 2009. Soil Sampling for Precision Agriculture. **Precision Agriculture**, pp. 1-4.

Franzen, D. W.; Hopkins, D. H.; Sweeney, M. D.; Ulmer, M. K.; Halvorson, A. D. 2002. Evaluation of Soil Survey Scale for Zone Development of Site-Specific Nitrogen Management. **Agronomy Journal**, 94 (2), pp. 381-389.

Fraser, B. T.; Congalton, R. G. 2019. Evaluating the Effectiveness of Unmanned Aerial Systems (UAS) for Collecting Thematic Map Accuracy Assessment Reference Data in New England Forests. **Forests**, 10 (1), pp. 1-17.

Genton, M. G. 1998. Highly robust variogram estimation. **Mathematical Geology**, 30, pp. 213-221.

Gojiya, K. M.; Gontia, N. K.; Patel, K. C. 2018. Generation of Thematic Maps of a Forest Watershed using Remote Sensing and GIS. **International Journal of Current Microbiology and Applied Sciences**, 7 (12), pp. 2952-2962.

Goovaerts, P. 1997. **Geostatistics for Natural Resources Evaluation**. Applied Geostatistics Series. Oxford University Press, New York & Oxford.

Han, C.; Wang, J.; Zheng, M., Wang, E., Xia, J., Li, G., Choe, S. 2016. New variogram modeling method using MGGP and SVR. **Earth Science Informatics**, 9, pp. 197-213.

Isaaks, E. H.; Srivastava, R. M. 1989. **Applied geostatistics**. Oxford University Press, New York.

Journell, A. G.; Huijbregts, C. J. 1978. **Mining geostatistics**. Academic Press, London.

Konopatzki, M. R.; Souza, E. G.; Nóbrega, L. H.; Uribe-Opazo, M. A.; Suszek, G. 2012. Spatial variability of yield and other parameters associated with pear trees. **Engenharia Agrícola**, 32 (2), pp. 381-392.

Lark, R. M. 2000. Estimating variograms of soil properties by the method-of-moments and maximum likelihood. **European Journal of Soil Science**, 51, pp. 717-728.

Li, Z.; Zhang, X.; Clarke, K. C.; Liu, G.; Zhu, R. 2018. An automatic variogram modeling method with high reliability fitness and estimates. **Computers & Geosciences**, 120, pp. 48-59.

Matheron, G. 1963. Principles of geostatistics. **Economic Geology**, 58 (8), p. 1246-1266.

Mazzini, P. L. F.; Schettini, C. A. F. 2009. Avaliação de metodologias de interpolação espacial aplicadas a dados hidrográficos costeiros quase sinópticos. **Brazilian Journal of Aquatic Science and Technology**, 13 (1), pp. 53-64.

McBratney, A. B.; Pringle, M. J. 1999. Estimating Average and Proportional Variograms of Soil Properties and Their Potential Use in Precision Agriculture. **Precision Agriculture**, 1, pp. 125-152.

Michelon, G. K.; Bazzi, C. L.; Upadhyaya, S.; Souza, E. G.; Magalhães, P. S. G.; Borges, L. F.; Schenatto, K.; Sobjak, R.; Gavioli, A.; Betzek, N. M. 2019. Software AgDataBox-Map to precision agriculture management. **SoftwareX**, 10.

Mikula, K.; Izydorczyk, G.; Skrzypczak, D.; Mironiuk, M.; Moustakas, K.; Witek-Krowiak, A.; Chojnacka, K. 2020. Controlled release micronutrient fertilizers for precision agriculture - A review. **Science of The Total Environment**, 712, pp. 1-9.

Mueller, T. G.; Pusuluri, N. B.; Mathias, K. K.; Cornelius, P. L.; Barnhisel, R. I.; Shearer, S. A. 2004. Map quality for ordinary Kriging and inverse distance weighted interpolation. **Soil Science Society of America Journal**, 68 (6), pp. 2042-2047.

Oliver, M. A.; Webster, R. 2015. **Basic steps in geostatistics**: the variogram and Kriging. Springer-Verlag, London.

Pimentel-Gomes, F. 2009. **Curso de estatística experimental**, 15. FEALQ, Piracicaba.

Reza, S. K.; Sarkar, D.; Daruah, U.; Das, T. H. 2010. Evaluation and comparison of ordinary Kriging and inverse distance weighting methods for prediction of spatial variability of some chemical parameters of Dhalai district, Tripura. **Agropedology**, 20 (1), pp. 38-48.

Rodrigues, M. S.; Alves, D. C.; Souza, V. C. de; Melo, A. C. DE; Lima, A. M. do N. 2018. Spatial interpolation techniques for site-specific irrigation management in a mango orchard. **Comunicata Scientiae**, 9 (1), pp. 93-101.

Souza, E. G. et al. Comparison of yield maps of three fields. **Computers and Electronics in Agriculture,** 2021. (in analysis).

Souza, E. G.; Bazzi, C. L.; Khosla, R.; Uribe-Opazo, M. A.; Reich, R. M. 2016. Interpolation type and data computation of crop yield maps is important for precision crop production. **Journal of Plant Nutrition**, 39 (4), pp. 531-538.

Souza, E. G.; Schenatto, K.; Bazzi, C. L. 2018. Creating thematic maps and management zones for agriculture fields. In: **Proceedings of the 14th International Conference On Precision Agriculture** (IPCA).

Uribe-Opazo, M. A., Borssoi, J. A., Galea, M. 2012. Influence diagnostics in Gaussian spatial linear models. **Journal of Applied Statistics** 39 (3), pp. 615–630.

Vieira, S. R. 2000. Geoestatística em estudos de variabilidade espacial do solo. In: Novais, R. F. de; Alvarez V., V. H.; Schaefer, C. E. G. R. (Eds.), **Tópicos em ciência do solo**. Sociedade Brasileira de Ciência do Solo, Viçosa.

Wackernargel, H. 2003. **Multivariate geostatistic**: an introduction with applications, 3. Springer-Verlag Berlin Heidelberg.

Webster, R. 1985. **Quantitative spatial analysis of soil in the field**. In: Stewart, B. A. (Ed.), Advance in soil science, 3. Spriger-Verlag, New York.

Webster, R.; Oliver, M. A. 1990. **Statistical methods in soil and land resource survey**. Oxford University Press, Oxford.

Wikle, C. K.; Zammit-Mangion, A.; Cressie, N. 2019. **Spatio-Temporal Statistics with R**. Chapman & Hall/CRC, Boca Raton, FL.

Wollenhaupt, N. C.; Wolkowski, R. P.; Clayton, M. K. 1994. Mapping soil test phosphorus and potassium for variable-rate fertilizer application. **Journal of Production Agriculture**, 7 (4), pp. 441-448.

Zhang, N.; Wang, M.; Wang, N. 2002. Precision agriculture - a worldwide overview. **Computers and Electronics in Agriculture**, 36 (2-3), pp. 113-132.

8 FINAL CONSIDERATIONS

The microservices architecture (MSA) of the AgDataBox (ADB) digital platform made available the functionalities for the thematic maps (TMs) creation and management zones (MZs) delineation in a satisfactory way. Around this, different applications can be developed.

ADB-Map application consumes ADB-MSA resources, it is free for access (http://adb.md.utfpr.edu.br/map), and can be used by technicians and researchers for commercial, educational, or research activities. In addition, this application allows performing all steps to easily, and friendly create TMs and delineate MZs.

ADB-MSA and ADB-Map application contribute to the new agriculture's phase, called digital agriculture, favoring the precision agriculture techniques adoption, assisting in data analysis and processing from farms, and allowing agricultural management decision-making. The case studies have indicated that it is possible to estimate the site-specific fertilizer and lime requirements based on soil attributes availability.

The interpolator selection process and its parameters have been improved in this new ADB version based on the inclusion of new selection criteria constituted of (i) effective spatial dependence index (%ESDI) > 25%, (ii) the first semivariance significance index ($\%\gamma(1)$) < 50%, and (iii) slope of the model ends index (%SMEI) > 20%. These criteria should be applied during the interpolator selection analysis and followed the application of interpolator selection index (ISI).

9 FUTURE WORKS

- Add functionality of definition of automated sampling grids from aerial images established by using level curves.
- Development of computational module for economic analysis and cost of agricultural production (ADB-Economic).
- Improve MZs rectification module to specify a minimum size of each zone and allow MZs realignment demarcations according to the field topography and level curves.
- Delimitation of rectangular and homogeneous management zones with multiple widths of the operational range.