WESTERN PARANÁ STATE UNIVERSITY – UNIOESTE

CENTER FOR EXACT AND TECHNOLOGICAL SCIENCES – CAMPUS CASCAVEL

GRADUATE PROGRAM IN AGRICULTURAL ENGINEERING

JORGE AIKES JUNIOR

AGDATABOX-MAP-FAST TRACK: WEB APPLICATION MODULE FOR AUTOMATIC CREATION OF THEMATIC MAPS AND MANAGEMENT ZONES

CASCAVEL

2022

JORGE AIKES JUNIOR

AGDATABOX-MAP-FAST TRACK: WEB APPLICATION MODULE FOR AUTOMATIC CREATION OF THEMATIC MAPS AND MANAGEMENT ZONES

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

Advisor: PhD. Eduardo Godoy de Souza
CoAdvisor: PhD. Claudio Leones Bazzi

CASCAVEL – PARANÁ – BRAZIL

2022

Ficha de identificação da obra elaborada através do Formulário de Geração Automática do Sistema de Bibliotecas da Unioeste[1].

Aikes Junior, Jorge
AgDatabox-Map-Fast Track: Web Application Module for Automatic Creation of Thematic Maps and Management Zones / Jorge Aikes Junior; orientador Eduardo Godoy de Souza; coorientador Claudio Leones Bazzi. -- Cascavel, 2022.
183 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, 2022.

1. Agricultura de precisão. 2. Zonas de manejo. 3. Softwares para agricultura. I. Souza, Eduardo Godoy de, orient. II. Bazzi, Claudio Leones, coorient. III. Título.

[1] Revisão de português, inglês e normas realizada por Ana Maria Vasconcelos, em 16 de maio de 2022.

JORGE AIKES JUNIOR

AgDataBox-Map-FastTrack: Módulo computacional para delineamento automático de zonas de manejo.

Tese apresentada ao Programa de Pós-Graduação em Engenharia Agrícola em cumprimento parcial aos requisitos para obtenção do título de Doutor em Engenharia Agrícola, área de concentração Sistemas Biológicos e Agroindustriais, linha de pesquisa Geoprocessamento, Estatística Espacial e Agricultura de Precisão, APROVADO(A) pela seguinte banca examinadora:

Orientador(a) – Eduardo Godoy de Souza

Universidade Estadual do Oeste do Paraná – Campus de Cascavel (UNIOESTE)

Claudio Leones Bazzi

Universidade Tecnológica Federal do Paraná (UTFPR)

Ricardo Sobjak

Universidade Tecnológica Federal do Paraná (UTFPR)

Marcio Antonio Vilas Boas

Universidade Estadual do Oeste do Paraná – Campus de Cascavel (UNIOESTE)

Alan Gavioli

Universidade Tecnológica Federal do Paraná (UTFPR)

Cascavel, 06 de abril de 2022

BIOGRAPHY

Jorge Aikes Junior, born on November 14th, 1986 in Florianópolis/SC. He holds a degree in Computer Science from the *Centro de Ensino Superior de Foz do Iguaçu* (Cesufoz), under the counselling of Professor Igor Mussoi de Lima, completed in 2008, and an academic master's degree in Engineering of Dynamic and Energetic Systems, from the Western Paraná State University (UNIOESTE) *Campus* of Foz do Iguaçu, under the counselling of Professor Huei Diana Lee, whose conclusion was in April 2012.

He has been a professor of Higher Education at the Academic Department of Computing at the Federal University of Technology in Paraná (UTFPR) *campus* of Medianeira, since 2015, working in teaching, research, and extension activities. He joined the *Stricto Sensu* Graduate course in Agricultural Engineering, Doctorate level, in March 2018, at UNIOESTE, in the area of concentration on Biological and Agroindustrial Systems, whose theme regards precision agriculture, under the counselling of Professor Eduardo Godoy de Sousa.

DEDICATION

To my wife Livia and my daughter Ana Clara.

Thank you very much; I Love you ineffably!

ACKNOWLEDGEMENTS

I thank my wife, Livia Willemann Peres, for all love, understanding, patience, and support on this journey. Everything, in every aspect of my life, would be so much harder without you, and for that, I will be forever grateful.

To my daughter Ana Clara Willemann Aikes for being the light in the midst of the storm, for making me a better person and immeasurably happier.

To my advisor Eduardo Godoy de Souza, and my co-advisor, Claudio Leones Bazzi, for the rich counselling throughout the research development, your enormous availability and dedication, and the possibility of learning from your experience and professionalism.

To UNIOESTE, in particular the Graduate Program in Agricultural Engineering, for the opportunity for professional growth.

To all professors and colleagues from PGEAGRI for the countless learning opportunities.

To all my friends and colleagues who somehow contributed to the development of this research, THANK YOU SO MUCH for the support and collaboration.

To the Federal University of Technology Paraná - UTFPR, for the opportunity to take half of my absence from professional activities and give the opportunity to dedicate myself to this doctoral program.

RESUMO

Aikes Junior, Jorge. **AgDatabox-Map-Fast Track: Módulo Computacional para Geração de Mapas Temáticos e Delineamento de Zonas de Manejo Automático**. Orientador: Eduardo Godoy de Souza; Coorientador: Claudio Leones Bazzi. 2022. 183 f. Tese (Doutorado em Engenharia Agrícola) – Universidade Estadual do Oeste do Paraná, Cascavel - Paraná, 2022.

A agricultura de precisão consiste na aplicação de insumos em quantidades necessárias, no local e no momento adequados, de maneira a maximizar a produtividade e reduzir o impacto ambiental. Duas tarefas importantes desse processo consistem na criação de mapas temáticos (TM) e zonas de manejo (MZ). Apesar de sumamente importantes, os TMs e o delineamento de MZs dependem de grande conhecimento técnico para sua construção, assim dificulta sua utilização em especial por produtores sem acesso a equipes especializadas, devido à necessidade de uma equipe multidisciplinar especializada. Neste sentido, este trabalho apresenta um módulo (ADB-MAP-FT) que permite delinear MZs e gerar TMs de maneira automática, integrado à aplicação *AgDataBox-Map* (ADB-Map) da plataforma *web AgDataBox* (ADB; https://adb.md.utfpr.edu.br/). ADB é uma plataforma *web* para integração de dados, *software*, procedimentos e metodologias para agricultura digital. Esta tese foi formatada em três documentos: Livro 1: apresenta conceitos, protocolos, softwares e diversos exemplos de uso tanto de TMs quanto de MZs de maneira a se ter um entendimento de ambos. Este livro está dividido em duas partes principais: a primeira apresenta as TMs, com suas características, importância, uso, definições para a melhor escolha do esquema de cores e vários exemplos e a segunda apresenta as MZs e diversos exemplos. Como o delineamento de ZMs apresenta várias possibilidades, as definições, protocolos, retorno econômico, opções e softwares mais comuns utilizados baseiam-se em um estudo sistemático da literatura, constituído a partir da união de técnicas sistemáticas de mapeamento de literatura e snowball. Artigo 1: foi desenvolvido o módulo computacional *web* (ADB-MAP-FT) para a geração de TMs, delineamento de MZs automáticos e implementação de protocolos definidos com base em extensa pesquisa bibliográfica. O módulo e suas características são apresentados, utilizando-se dados reais em um estudo de caso. Artigo 2: Foi conduzido um comparativo entre o módulo desenvolvido e os softwares mais utilizados na literatura bem como os softwares a partir do estado de arte existente. As características e capacidades técnicas dos softwares são comparadas e validadas em um estudo de caso utilizando duas áreas comerciais distintas, comparando também os resultados do processo.

PALAVRAS-CHAVE: AgDataBox-Map, agricultura de precisão, software.

ABSTRACT

Aikes Junior, Jorge. **AgDataBox-Map-Fast Track: Web Application Module for Automatic Creation of Thematic Maps and Management Zones**. Advisor: Eduardo Godoy de Souza; CoAdvisor: Claudio Leones Bazzi. 2022. 183 f. Dissertation (PhD in Agricultural Engineering) – Universidade Estadual do Oeste do Paraná, Cascavel - Paraná, 2022.

Precision agriculture consists of applying inputs in the right quantities at the right time and place to maximize yield. There are two important tasks in this process: the generation of thematic maps (TM) and management zones (MZ). Although the TMs and the MZs delineation are extremely important, they depend on great technical knowledge for their construction, so, their use is difficult, especially by producers who do not have access to specialized teams, due to the need for a specialized multidisciplinary team. Thus, this work presents a module (ADB-MAP-FT) able of delineating MZs and generating TMs automatically, integrated with AgDataBox-Map (ADB-Map) application of AgDataBox web platform (ADB; https://adb. md.utfpr.edu.br/). ADB is a web platform to integrate data, software, procedures, and methodologies for digital agriculture. This dissertation was organized in three documents: Book 1: presents concepts, protocols, software, and several examples of use and understand both TMs and MZs. This book is divided into two main parts: the first one presents TMs, their characteristics, importance, use, definitions for the best choice of color scheme, and several examples. The second one presents MZs and several examples. As ZMs design presents several possibilities, definitions, protocols, economic return, options, and most commonly used software are based on a systematic study of literature, constituted from the association of systematic techniques of literature mapping and snowball. Article 1: the web computational module (ADB-MAP-FT) was developed to automatically create TMs and delineate MZs, implementing protocols defined based on extensive bibliographic research. The module and its features are presented using real data in a case study. Paper 2: It was carried out a comparison between the developed module and the most used software in the literature based on the existing state-of-the-art software. The characteristics and technical abilities of software are compared and validated in a case study using two different commercial areas, also comparing the results of this process.

KEYWORDS: AgDataBox-Map, precision agriculture, software.

SUMMARY

LIST OF FIGURES

BOOK 1

PAPER 1

PAPER 2

LIST OF TABLES

BOOK 1

PAPER 1

PAPER 2

LIST OF ABBREVIATIONS AND ACRONYMS

ACPA	Australian Centre for Precision Agriculture
ADB	AgDataBox
ADB-MAP	AgDataBox-Map
ADB-MAP-FT	ADB-Map Fast-Track
AE	Average Error
ANOVA	Analysis of Variance
API	Application Programming Interface
ARS	Agricultural Research Service
ASC	Average Silhouette Coefficient
BILP	Binary Integer Linear Programming
BPMN	Business Process Model Notation
BSB	Backward snowballing
CAFe	Federated Academic Community
CAPES	Coordination for the Upgrading of Higher Education Personnel
CEC	Cation Exchange Capacity
CR	Constant Rate
CRD	Coefficient of Relative Deviation
CSS	Cascade Style Sheets
CSU	Colorado State University
CSV	Comma-Separated Values
CV	Coefficient of Variation
CVI	Coefficient of Variation at the ith Sampling point
DBMS	Database Management System
DD	Discarding duplicates
ECa	Apparent Electrical Conductivity of the Soil
ECI	Error comparison index

EDA	Exploratory data analysis
EMBRAPA	Brazilian Agricultural Research Corporation
FCM	Fuzzy C-means
FI	Fragmentation Index
FPI	Fuzziness Performance Index
FSB	Forward Snowballing
GA	Global Accuracy
Gc	Gaybille-Caruso
GCS	Geographic coordinate system
Gg	Gibbon loam
GIS	Geographic Information System
GNSS	Global Navigation Satellite Systems
GPS	Global Positioning System
GQI	Global Quality Index
GSC	Group Silhouette Coefficient
ha	Hectare
HACC	Hierarchical Agglomerative Constrained Clustering
HACC-SPATIAL	Hierarchical Agglomerative Constrained Clustering Spatial
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
HRFA	Homogeneous Rate Fungicide Application
HZ	Homogeneous Zones
ICVI	Improved Cluster Validation Index
ID	Inverse of the Distance
IDS	Inverse of the Square Distance
IDW	Inverse Distance Weighting
IQR	Interquartile Range
ISI	Interpolator Selection Index
JSON	JavaScript Object Notation

K	Kappa
kPa	Kilopascal
Kg	Kilogram
KRG	Kriging
LAI	Leaf Area Index
LM	Local Moran Index
MA	Moving Average
MAD	Mean Absolute Difference
MC	Management Class
ME	Mean Error
MPE	Modified Partition Entropy
MULTISPATI-PCA	Multivariate spatial Analysis Based on Moran's Index PCA
MZ	Management Zone
MZA	Management Zone Analyst
NC	Number of Classes
NCE	Normalized Classification Entropy
NDVI	Normalized Difference Vegetation Index
NMZ	Number of Zones
OM	Organic Matter
OK	Ordinary Kriging
OS	Operating system
PA	Precision Agriculture
PC	Principal Components
PCA	Principal Component Analysis
Q1	1º Quarter
Q3	3º Quarter
RE	Reduced average error
REST	Representational State Transfer
RS	Remote Sensing

SAE	Standard deviation of the average error
SAR	Selection by abstract reading
SB	Snowballing
SD	Standard Deviations
SDME	Standard Deviation of Mean Error
SDUM	Software for Defining Management Zones
SH	Selective Harvest
SI	Smoothness Index
SLM	Systematic Literature Mapping
SLR	Systematic Literature Review
SLS	Systematic Literature Study
SPR	Soil Penetration Resistance
STR	Selection by title reading
SRE	Standard deviation of the reduced error
SWC	Soil Water Content
Tk	Tukey's test
TM	Thematic Map
TXT	Text File Format
UAV	Unmanned Aerial Vehicles
UC Davis	University of California Davis
URA	Uniform Nitrogen Rate Application Based on an Agronomic Optimum
URE	Uniform Nitrogen Rate Application Based on an Economic Optimum
USDA	United States Department of Agriculture
USP	University of São Paulo
UTFPR	Federal University of Technology - Paraná
UTM	Universal Transverse Mercator
UNIOESTE	Western Paraná State University
VAR	Variance
VI	Vegetation Index

VR	Variance Reduction
VRSA	Variable-rate Fungicide Application
WGS84	World Geodetic System 84
XB	Xie Beni Index
YM	Yield Maps

1 INTRODUCTION

The growing population (ONU, 2020) has demanded large amounts of food, and environmental issues, for both environment preservation and better rational use of all elements on food production chain. So, there is a huge interest by producers to make the most optimized use of their lands and inputs (Baudron and Giller, 2014). Precision agriculture (PA) is a management system that aims at optimizing the use of agricultural inputs, and meeting this need for more profitability with less environmental damage. Several tools can be used to support this, among them, thematic maps and management zones stand out.

Generally, thematic maps (TMs) are used to identify different cartographic representations, and they represent not only the land but also some associated characteristics. TMs development is related to data collection, analysis, interpretation, and representation of information on a map. They facilitate identification of similarities and enable visualization of spatial correlations.

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 is choropleth maps that use color to show ranges of a specific variable 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 indexes). Based on samples collected before, during, and after a crop life span , TMs are created to identify the variability of topography, soil, and plant properties and to be compared with yield (Souza; Schenatto; Bazzi, 2018). Several tasks need to be performed in order to create TMs, such as (i) selection of a coordinate system, to store, retrieve, manage, display, and analyze all types of geographic and spatial data; (ii) data normalization, where the most common methods are the standard score, range, and mean (Schenatto et al., 2017b); (iii) exploratory data analysis (EDA), where it is employed a variety of techniques (mostly graphical) to maximize insight into a data set; uncover underlying structure; extract important variables; detect inliers and outliers (atypical values) and anomalies; underlying test assumptions; develop parsimonious models; and to determine optimal factor settings (NIST/SEMATECH, 2013); (iv) data interpolationis usually applied to result in a dense and regular grid to create TMs and management zones (MZ) that are continuous and smooth; (v) creation of TMs, where someone must decide the number of classes, the method for breaking data into ranges and color scheme.

One practical way to apply PA in a field is to divide it into homogeneous areas, called management zones (MZs). Each zone is a subregion of a field that expresses a functional 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 variable-rate application machines could be used, MZs usually involve conventional machinery. After delineation, MZs can be used in smart sampling, where one-smart composite sampling is obtained per zone to delineate the field soil variability. This approach is likely to reduce laboratory costs while maintaining reliability (Ferguson and Hergert, 2009; Mallarino and Wittry, 2004).

MZs application is economically and productively viable in several situations, showing results of cost reduction, increase in yield, and improvement of product quality parameters (Bernardi et al., 2018; Cid-Garcia; Ibarra-Rojas, 2019; Kyaw et al., 2008; Li et al. 2013; Roberts et al., 2012; Robertson et al., 2008; Schwalbert et al., 2018; Velandia et al., 2008; Vitharana et al., 2008; Whetton et al., 2018). Thus, its application often leads to an increased profitability and reduces costs with inputs, consequently results on fewer environmental impacts.

However, there are still several outstanding issues, such as: (i) what is the ideal protocol for MZs delineation , (ii) what is the best delineation algorithm?; (iii) which software allows you to handle all the stages during the process?; (iv) is there a way to simplify and/or automatize the process facilitating MZs adoption? So, the task of defining ideal MZs is still a challenge.

The main content of this work is structured in papers (chapters 3 to 5):

- Book 1 (chapter 3): This book is intended to assist in understanding both tools, TMs and MZs. The objective is to define them and present an ideal protocol for their development, with examples in both cases. This book is divided into two main parts: one that presents TMs, their characteristics, importance, usage, definitions for the best choice of color scheme, and several examples. The second one presents the MZs. As MZs delineation presents several possibilities, definitions, protocols, economic return, the most common options and the applied software are based on a systematic literature study, constituted from the union of systematic literature mapping, snowball, and systematic literature review techniques. This ensures that the main procedures and trends are achieved, gathering an extensive summary of classic works and the most recent ones;
- Paper 1 (chapter 4): this work presents a new computational module, called ADB-Map Fast-Track (ADB-MAP-FT), which allows TMs

creation and automatic MZs delineation (following the best protocols - a result from the research performed in book 1 by a web-friendly interface platform, ideal for users who do not have all the technical knowledge necessary for MZs delineation);

- Paper 2 (chapter 5): this work compares ADB-MAP-FT with the main and the most modern software for MZs delineation, both in technical aspects and in a case study using two actual areas.

2 OBJECTIVES

2.1 General objectives

Develop a computational module to create TMs and delineate MZs automatically.

2.2 Specific objectives

- Define a protocol for the automatic design of TMs and MZs;
- Implement a computational module capable of automatically TMs creation and MZs delineation;
- Evaluate the computational module for MZs delineation developed by carrying out a case study;
- Compare the MZs delineation module to another traditional and state-of-the-art software.

3 BOOK 1 – THEMATIC MAPS AND MANAGEMENT ZONES FOR PRECISION AGRICULTURE: SYSTEMATIC LITERATURE STUDY, PROTOCOLS, AND PRACTICAL CASES

3.1 INTRODUCTION

The growing population has demanded large amounts of food, and environmental issues, for both environment conservation and a better rational use of all elements of the food production chain. Thus, producers are motivated to optimize the use of land and its inputs as best as they can (Baudron and Giller, 2014). Precision agriculture (PA) is a management system that aims at optimizing the use of agricultural inputs, and meeting this need for better profitability with less environmental damage.

Climatic, topographic, and biological variations, in both spatial and temporal domains, are factors that induce yield variations in the field. The premise of PA is to know these variations and provide support for punctual and localized crop management. Several tools can be used to support this, and among them, thematic maps and management zones stand out.

Thematic maps (TMs) are used to illustrate themes as well as represent the terrain. Generally, TMs are used to identify different cartographic representations, and they represent not only the land but also associated characteristics. TMs development is associated to data collection, analysis, interpretation, and representation of information on a map. They make the identification of similarities easier and enable the visualization of spatial correlations. Based on samples collected before, during and after the life period of a culture, TMs are usually generated to identify properties variability of of topography, soil, and plants and compare with yield. However, firstly, it is necessary to interpolate data into a dense and regular grid to provide values for locations that were not sampled. This task is performed with interpolation methods support, and kriging was the most used interpolation method.

Timlin et al. (1998) showed that yield and other field attributes presenting spatial variability could be effectively used in site-specific management (precision agriculture, PA) to increase fertilizer efficiency and environmental sustainability, although it is often costly (Khosla et al., 2008). Typically, soil samples are analyzed to determine soil nutrient levels. Sampling, therefore, should be dense enough to allow nutrient variability determination on soil so that fertilizers can be used profitably and in an environmentally sustainable way (Ferguson and Hergert, 2009; Franzen et al., 2002). Time and available budget for sampling should be considered to determine the right soil sampling density in an area.

Traditional farm management uses a whole-field approach, in which each field is treated as a homogeneous area (Srinivasan, 2006), and variability in soil, topography, local weather conditions, and land use are not considered (Nawar et al., 2017). In this management, inputs are applied uniformly across the field, and it is attractive to farmers because it is easy and fast. However, it is possible to achieve more economical and environmentally-friendly management with a site-specific input application. PA uses this kind of application, and it is defined 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 estimated variability for improved resource use efficiency, productivity, quality, profitability and sustainability of agricultural production (ISPA, 2019).

One practical way to apply PA in a field is to divide it into homogeneous areas, called management zones (MZs). Each zone 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 variable-rate application machines could be used, MZs usually involve conventional machinery. After delineation, MZs can be used in smart sampling, where one-smart composite sampling is obtained per zone to delineate the field soil variability. This approach is likely to reduce laboratory costs while maintaining reliability level (Ferguson and Hergert, 2009; Mallarino and Wittry, 2004). Smart sampling has been shown to improve nutrient efficiency use while keeping or increasing yield and potentially reducing nutrient overloading into the environment (Moshia et al., 2014; Khosla et al., 2002). Many studies related to sampling density have been performed (Journel and Huijbregts, 1978; Demattê et al., 2014; Wollenhaupt, Wolkowski, and Clayton, 1994; Franzen et al., 2002; Ferguson and Hergert, 2009; Doerge, 2000), which resulted in a suggested minimum density of one sample per hectare (Ferguson and Hergert, 2009) to 2.5 samples per hectare (Journel and Huijbregts, 1978; Doerge, 2000), which should be composed of at least eight individual samples (Wollenhaupt, Wolkowski, and Clayton, 1994).

Several kinds of sample data can be used to delineate MZs; however, to produce more stable MZs, it is advantageous to use a set of multivariate attributes data that do not vary significantly over time (topography, electrical conductivity, soil physical properties) and that are correlated with the target variable (usually yield) (Buttafuoco et al., 2010; Doerge, 2000). That is important because MZs are usually applied for many years. Nevertheless, there are other situations in which the purpose is to use MZs immediately but just at once. It is the case of MZs for agrochemical applications.

MZs use is economically and productively viable in several situations, showing results of cost reduction, increase in yield, and improvement of product quality parameters (Kyaw et al., 2008; Robertson et al., 2008; Velandia et al., 2008; Vitharana et al., 2008b; Roberts et al., 2012; Li et al., 2013; Bernardi et al., 2018; Schwalbert et al., 2018; Whetton et al., 2018). Thus, their application often takes to an increase in profitability and reduction of costs with inputs, consequently leading to fewer environmental impacts.

However, there are still several outstanding issues, such as: (i) what is the ideal protocol for MZs delineation?; (ii) what is the best delineation algorithm?; (iii) which software allows you to handle all the stages in the process? Therefore, the task of defining the ideal MZs is still a challenge.

This book aims at assisting in understanding both tools, TMs and MZs. The objective is to define them and present an ideal protocol for their development, with examples in both cases. This book is divided into two main parts: Chapter 2 presents TMs, with their characteristics, importance, usage, definitions for the best choice of color scheme, and several examples. Chapter 3 presents MZs. As MZs delineation presents several possibilities, definitions, protocols, economic return, and the most common options and software used are based on a systematic study of literature, constituted from the union of systematic literature mapping and snowball techniques. This ensures that the main procedures and trends are achieved, gathering an extensive summary of classic works and the most recent ones. At the end of this chapter, there are also several examples of MZs to offer the reader several possibilities.

3.2 THEMATIC MAPS

Maps that represent the land and a topic associated with it are called thematic maps (TMs), and they aim at informing by graphic symbols where a specific geographical phenomenon occurs. TM development is linked to data collection, analysis, interpretation, and representation of information on a map, facilitating the identification of similarities and enabling spatial correlations visualization. The information presented in TMs may include, for example, maximum temperature or maximum precipitation at a given date, amount of calcium and potassium in soil, and soybean yield at a given agricultural area. Figure 1 shows a TM of world apple production in 2009.

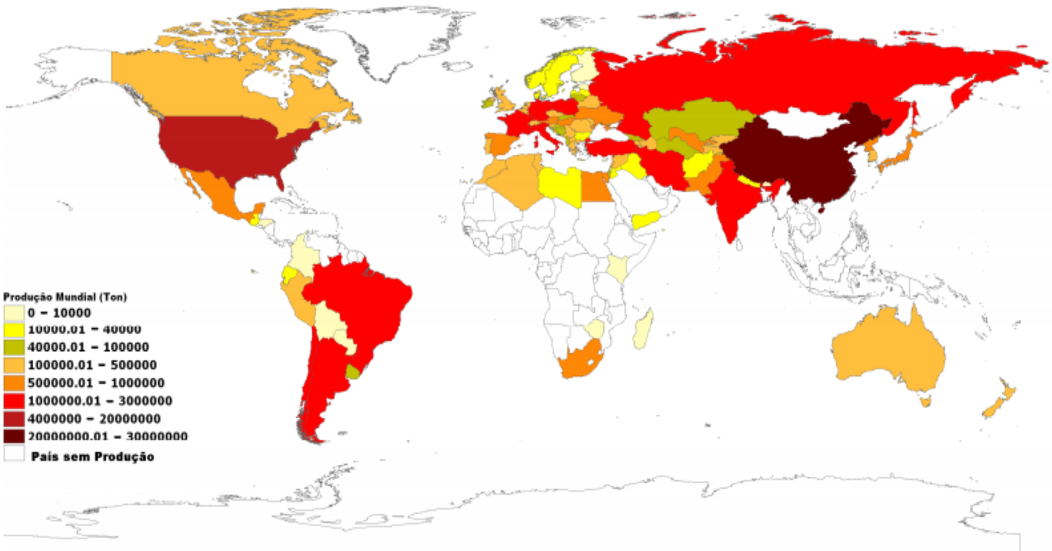
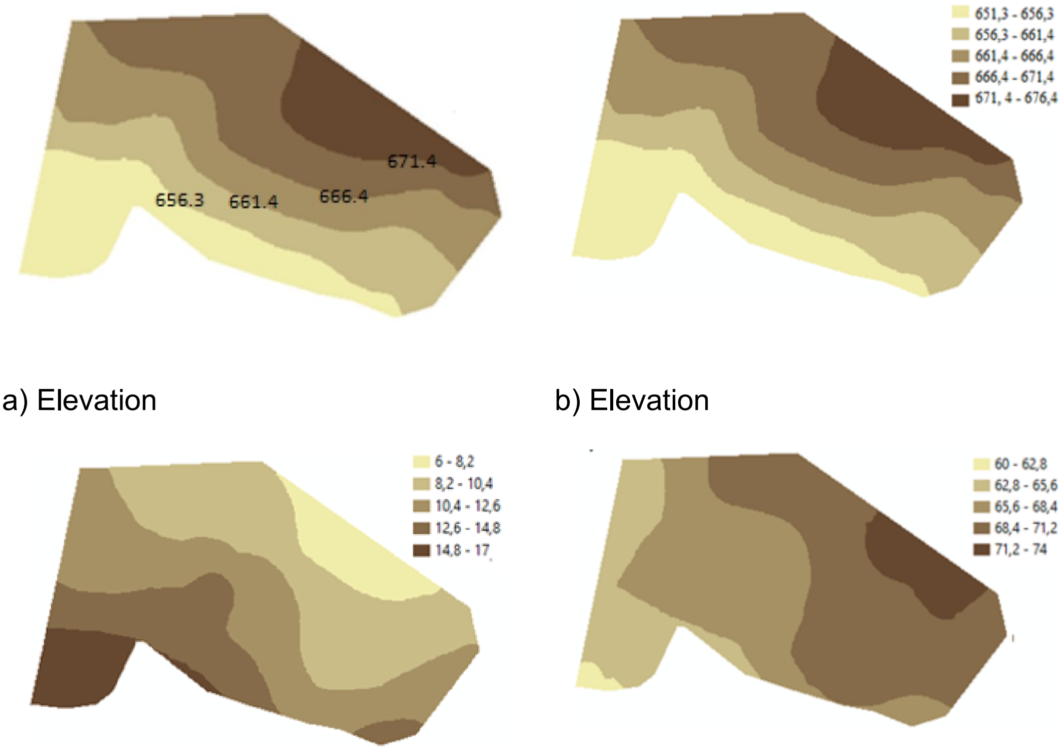


Figure 1. Thematic map of world apple production in 2009
Source: Carvalho (2011).

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. While choropleth maps use color to show ranges of a specific variable 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 indexes). Figure 2 shows examples of contour and choropleth maps. Thus, it is necessary to follow a protocol to construct TMs about attributes collected in agriculture fields,like the one presented in Figure 3.



a) Elevation b) Elevation

c) sand d) Clay

Figure 2. Examples of contour map: a) elevation (m), and choropleth maps: b) elevation (m), c) sand (%), d) Clay (%)

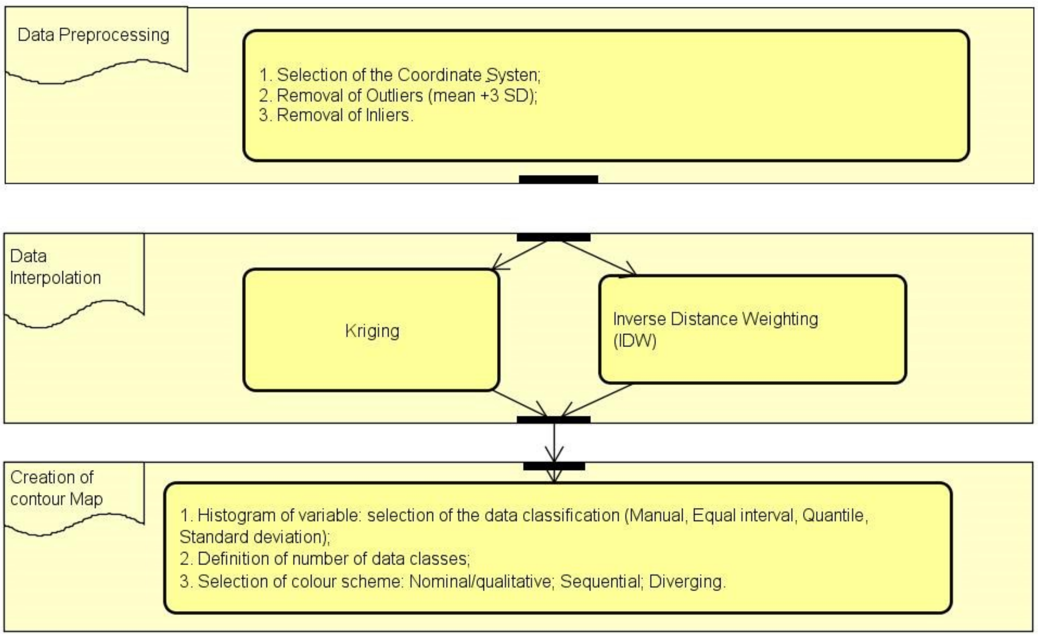


Figure 3. Flowchart of the typical protocol to create a thematic map

I. Selection of the coordinate system - A geographic information system (GIS Software) is designed to store, retrieve, manage, display, and analyze all kinds of geographic and spatial data. Thus, it is necessary a GIS software, and a file with at least three columns representing X (longitude) and Y (latitude) coordinates to construct 2-D TMs as well as the value of the measured attribute (for 3-D, we need one more coordinate, Z (altitude)). The most typical coordinate systems are the geographic coordinate system (GCS) and universal transverse Mercator (UTM). GCS is associated with a model of the Earth shape (reference ellipsoid) called a datum. The datum WGS84 (World Geodetic System 84) is the most commonly used. Units are in degrees, minutes, and seconds with GCS and meters for UTM.

II. Data normalization - Variables normalization is interesting when someone wants to construct and compare TMs of a variable that has been measured several times. This is the case of yield from an area measured for several years and/or with several crops. The most common methods are the standard score, range, and mean (Schenatto et al., 2017b).

III. Exploratory data analysis (EDA) - is the summarization of data set by their main characteristics. EDA employs a variety of techniques (mostly graphical) to maximize insight into a data set; uncover underlying structure; extract important variables; detect inliers, outliers (atypical values) and anomalies; test underlying assumptions; develop parsimonious models; and determine optimal factor settings (NIST/SEMATECH, 2013). When constructing TMs, the essential use of EDA is to detect and remove outliers. According to Amidan et al. (2005), data outliers can have a significant impact upon data-driven decisions, and in many cases, they do not reflect the true nature of data and, hence, should not be included in the analyses. They proposed an outlier detection method using Chebyshev's inequality to form a data-driven outlier detection method that is not dependent upon knowing data distribution. According to Córdoba et al. (2016), values outside the mean ± 3 SD (standard deviation) are identified as outliers and should be removed (also Haghverdi et al., 2015). They remarked that even though real data could belong to this interval, upper and lower limits should be modified to obtain robust variance estimators. It is also necessary the removal of inliers, data that differ significantly from their neighborhood but lie within the variation range of data set (Córdoba et al., 2016). Moreover, additional care should be taken for yield data obtained with yield monitor. Many approaches for yield data cleaning have already been proposed just like by Blackmore and Moore (1999) to eliminate errors associated with unknown header width, combine filling/emptying times, the time lag of grain with the combine, positional errors, fast changes, and others (Sudduth and Drummond, 2007). Vega et al. (2019) proposed a protocol to automate error removal from yield maps divided into two steps: (1) removal of yield data with values equal to zero, removal edge values and potential end-of-field yield monitor errors, and removal of yield data that are outside the mean ± 3 SD; and (2) use of the local Moran's spatial autocorrelation index and Moran's plot to identify and remove data that are inconsistent with their neighbor points. The protocol was evaluated on 595 real yield datasets with good results and can be used with other geo-referenced variables in precision agriculture.

VI. Data interpolation - The sample data are usually interpolated in a dense and regular grid to generate TMs and MZs that are continuous and smooth. This task is performed by interpolating methods. The inverse distance weighting (IDW) and kriging are the interpolation methods commonly used in PA. They are differentiated by how weights are assigned to different samples, influencing estimated values (Reza et al., 2010). Several software packages are available to perform data interpolation, such as Surfer (Golden Software, LLC) and ArcGIS (ESRI, Environmental Systems Research Institute).

Kriging is considered the best method of data interpolation when data present spatial dependence. Nevertheless, first, the appropriated geostatistical model for data needs to be found out by cross-validation. This technique compares theoretical values with those obtained from sampling, then analyzes the estimation errors, and chooses the best model (Arlot and Celisse, 2010; Kohavi, 1995). Faraco et al. (2008) considered cross-validation better to evaluate the adjustment of theoretical spatial models than Akaike's and Filiben's information criteria and the maximum logarithm value of likelihood function. The following measures are calculated using cross-validation: average error (AE), reduced average error ($\overline{RE}$), standard deviation of average error (SAE), and standard deviation of reduced error (SRE) (Cressie, 1993; McBratney and Webster, 1986). According to non-tendentiousness criteria, values for AE and $\overline{RE}$ should be as close to zero as possible to choose the best-adjusted model, SAE value should be as small as possible, and SRE value should be close to 1 (Cressie, 1993; McBratney and Webster, 1986). Souza et al. (2016) proposed the error comparison index (ECI, Equation 1) since cross-validation makes it possible the occurrence of ambiguous situations. As lower ECI is, the better semivariogram is.

$$ECI_i = \frac{ABS(\overline{RE})_i}{\max\ |_{i=1}^{j}[ABS(\overline{RE})]} + \frac{ABS(SRE-1)_i}{\max\ |_{i=1}^{j}[ABS(SRE-1)]} \qquad \text{Eq. 1}$$

where ECI_i is the error comparison index for model *i*, $ABS(\overline{RE})$ is the module value of the reduced average error, and $\max\ |_{i=1}^{j}$ is the highest value among the compared *j* semivariograms.

One recurrent issue when interpolating agricultural data is choosing between deterministic and stochastic methods of interpolation. Bier and Souza (2017) proposed the interpolation selection index (ISI, Equation 2), which assumes a lower value as better the interpolator is.

$$ISI = \left\{ \frac{ABS(AE)}{max\Big|_{i\,=\,1}^{\;j}[ABS(AE)]} + \frac{\Big[SAE - min\Big|_{i\,=\,1}^{\;j}SAE\Big]}{max\Big|_{i\,=\,1}^{\;j}[ABS(SAE)]} \right\} \qquad \text{Eq. 2}$$

where n is the number of data; $ABS(AE)$ is the module value of the average error of crossed validation; $min|_{i=1}^{j}$ is the lowest value obtained among the compared j models; $max|_{i=1}^{j}$ is the highest value obtained among the compared j models.

V. TMs Creation – after data interpolation, we must decide both the number of classes and the method to splitdata into ranges and draw TMs with our data. The goal is to group similar observations and split up substantially different observations (Indiemapper, 2016). The first decision-making is to look at the histogram (or scatterplot) to determine the 'form' of the researcher’s observations. This critical step to build a map, how we can dramatically change a map perspective, and thus, its message, it is one of the easiest ways to "lie with maps". There is no escape from the cartographic paradox: to present a useful and truthful picture, since an accurate map must tell white lies (Monmonier, 1996). There are many ways to classify data systematically and each GIS software will offer some of them. The most popular are described below (Indiemapper, 2016; ESRI ArcGIS 9, Help Menu, Standard Classification Schemes):

Manual interval: we set one or all of the class breaks manually. We use this method when the others do not give a good solution. So, a good way is to start with one of the standard classifications and make adjustments as they are needed;

Equal interval: we divide data into equal size classes, and it works well on data that is generally spread across the entire range. This classification should be avoided if data are skewed to one end or there are one or two large outlier values;

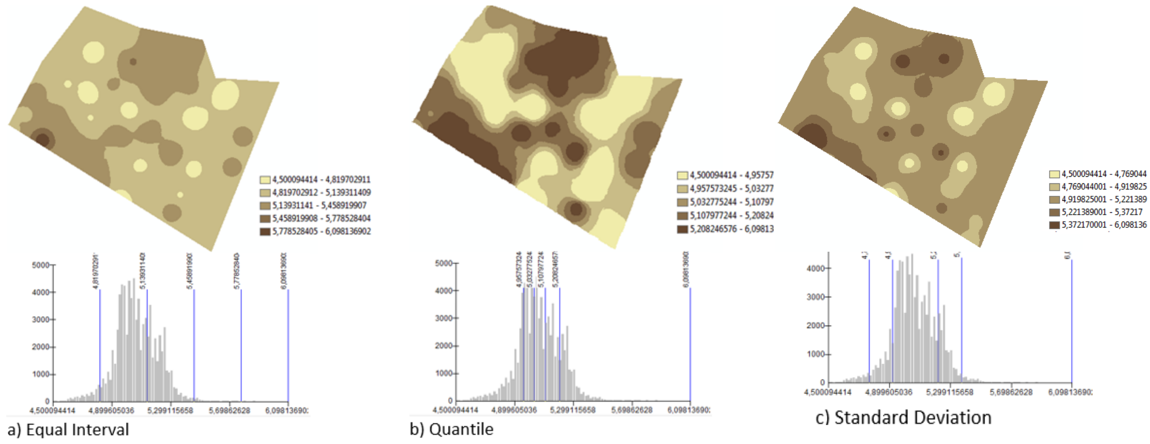
Quantile: we divide it into classes with an equal number of features, and it works well on data that is linearly distributed across the entire range. Nevertheless, the resulting map can be misleading, with similar features placed in adjacent classes, or widely different values put in the same class;

Standard deviation: it is a particular case of equal interval where the class size is a multiple of standard deviation. It works well with data that has normal distribution. It is good to see which features are above or below an average value.

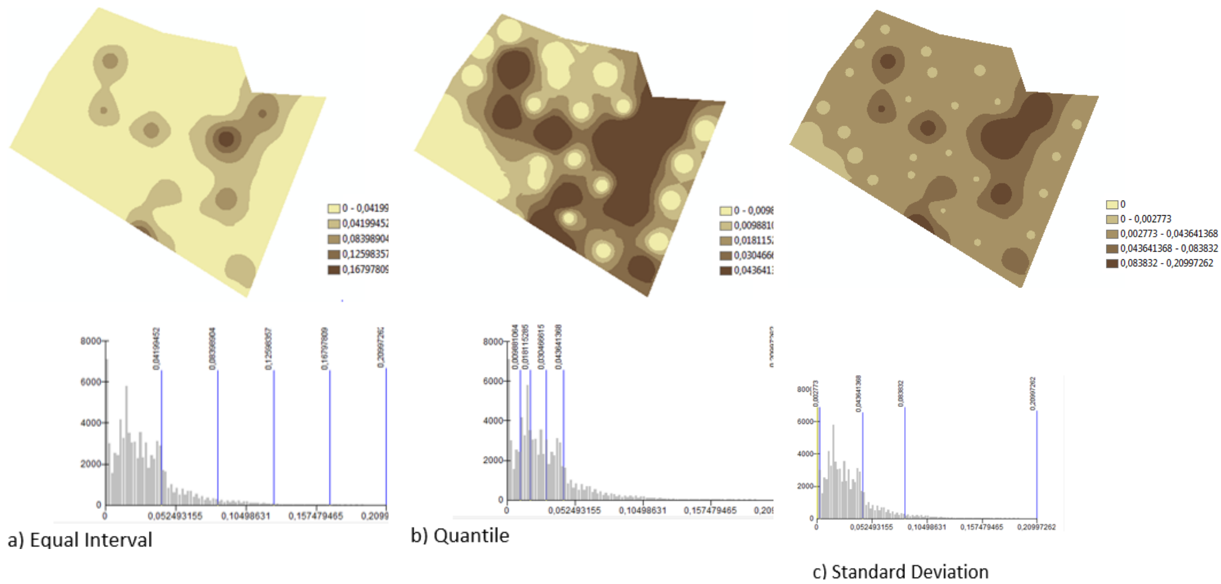
The number of data classes is also an essential part of map design. When there is an increase on the number of data classes, it will result on a more revealing map, but this requires more colors. Generally, it is advised not to exceed seven classes.

Examples of choropleth maps are presented in Figure 4. Each case presents the map using five classes, classified by equal interval, quantile, and standard deviation, and its corresponding histogram. For example, in pH (Figure 4a), there is an attribute with a distribution close to normal, and the equal interval classification looks like the best choice, but the standard deviation classification is also good. However, with the map of aluminum (Figure 4b), the distribution is moderately skewed right, and then quantile is visually the best option.

After we selected how data should be classified, it is crucial choosing an effective color scheme for TM. A good color scheme needs to be attractive but also support the map's message and be appropriately matched to the nature of data (Harrower and Brewer, 2003), therefore, it is relevant to choose three dimensions of color: hue, lightness, and saturation. There are three kinds of color scheme: **nominal/qualitative** (unorderable data, like land use, Figure 5a): different hues that keep lightness and saturation constant should be used; **sequential** (orderable, like numerical data (or low/med/high), like yieldFigure 5b): single or multi-hue with different lightness/saturation should be used; **diverging** (when there is a mid-point, like zero, or if we want to compare with an average, like profitFigure 5c). Harrower and Brewer (2003) designed an online tool “ColorBrewer.org” to help users on selecting the appropriate color schemes for their specific mapping needs.Figure 6 presents some practical examples of color schemes application.

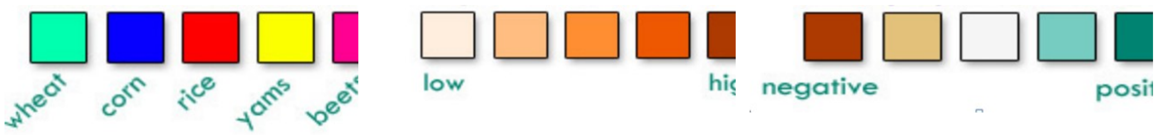


a) pH



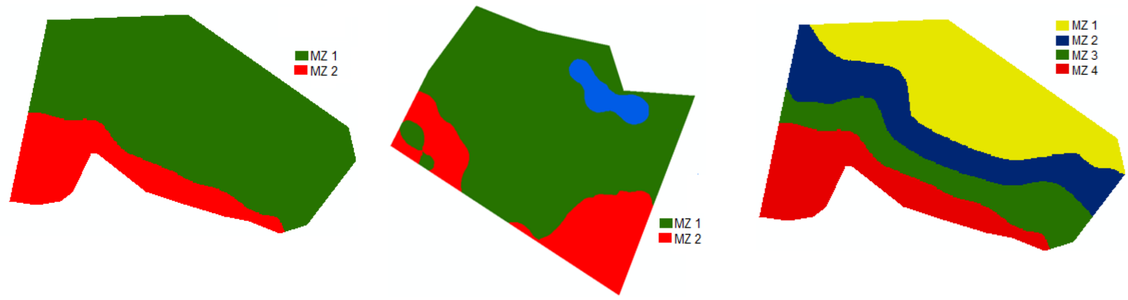
b) Aluminum

Figure 4. Thematic Maps for pH (a), and aluminium (b) using three forms of classification (equal interval, quantile, and standard deviation)



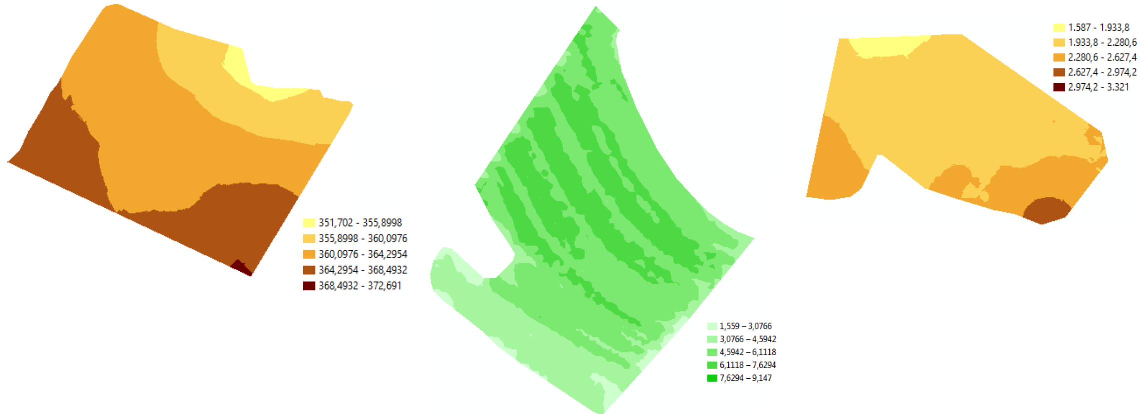
a) Nominal Color Scheme b) Sequential Color Scheme c) Diverging Color Scheme

Figure 5. Three kinds of color scheme: nominal/qualitative (a), sequential (b), and diverging (c)



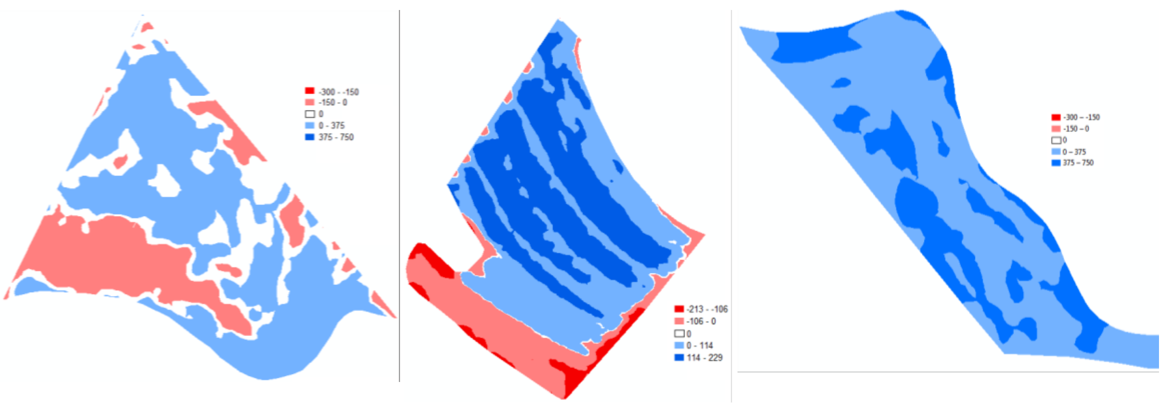
a.1) a.2) a.3)

a) Nominal Color Scheme: maps with two (a.1), three (a.2), and four (a.3) management zones (MZs)



b.1) Altitude b.2) Yield b.3) SPR 0.0 -0.1 m

b) - Sequential Color Scheme: maps of altitude (b.1), yield (b.2), and Soil Penetration Resistance – SPR (b.3)



c.1) c.2) c.3)

c) - Diverging Color Scheme – Profit Maps (US$ ha^{-1})

Figure 6. Examples of color scheme: nominal/qualitative (a), sequential (b), and diverging (c)

Contour maps use a continuous scale - despite being the most common map using a discrete scale, some people prefer continuous scale. The problem with a color ramp is that the perception of color intensity is not linear, consequently, the user could make a false assumption about what data value it represented. Basso et al. (2009) studied the effects of landscape position and rainfall on spatial variability of wheat yield and protein on a 10-ha field with the rolling landscape of Southern Italy, and presented an interpolated map of wheat yield (Figure 7) using a continuous scale.

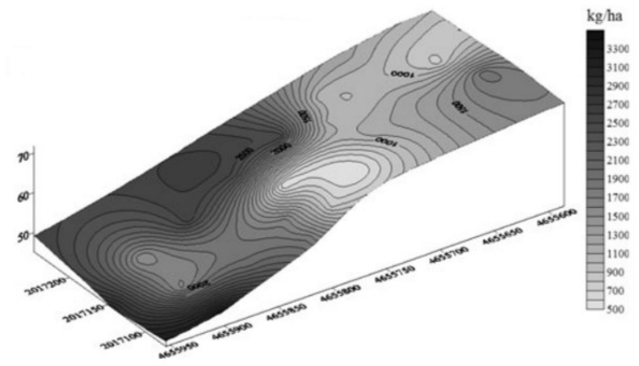


Figure 7. 3D interpolated map of wheat yield (kg ha^{-1}) for 2003
Source: Basso et al. (2009).

3.2.1 Examples of thematic maps

In order to demonstrate several situations in which TMs can be used, in sequence, several examples of TMs will be presented, associated with a brief discussion of data that have originated them.

3.2.1.1 Yield, protein, and oil content maps

Silva (2016) carried out a spatial analysis of quality parameters (protein and oil content) for soybean and corn crops in two experimental areas (field A - 10.0 ha, and field B - 23.8 ha) and two agricultural years (2012/2013 and 2013/2014.). Figure 8 shows the thematic maps of soybean yield and the corresponding protein and oil content. Statistical analysis using Moran's bivariate spatial autocorrelation statistic showed that soybean protein and oil content were inversely correlated for both experimental areas and agronomic years (2012/13 and 2013/14). It can be highlighted how important it is to choose the right color scheme. In this case, variables are quantitative, and therefore, the scheme should be sequential (single-hue with different lightness/saturation). Only to compare, the same variable is presented using a nominal color scheme, and map readability is reduced.

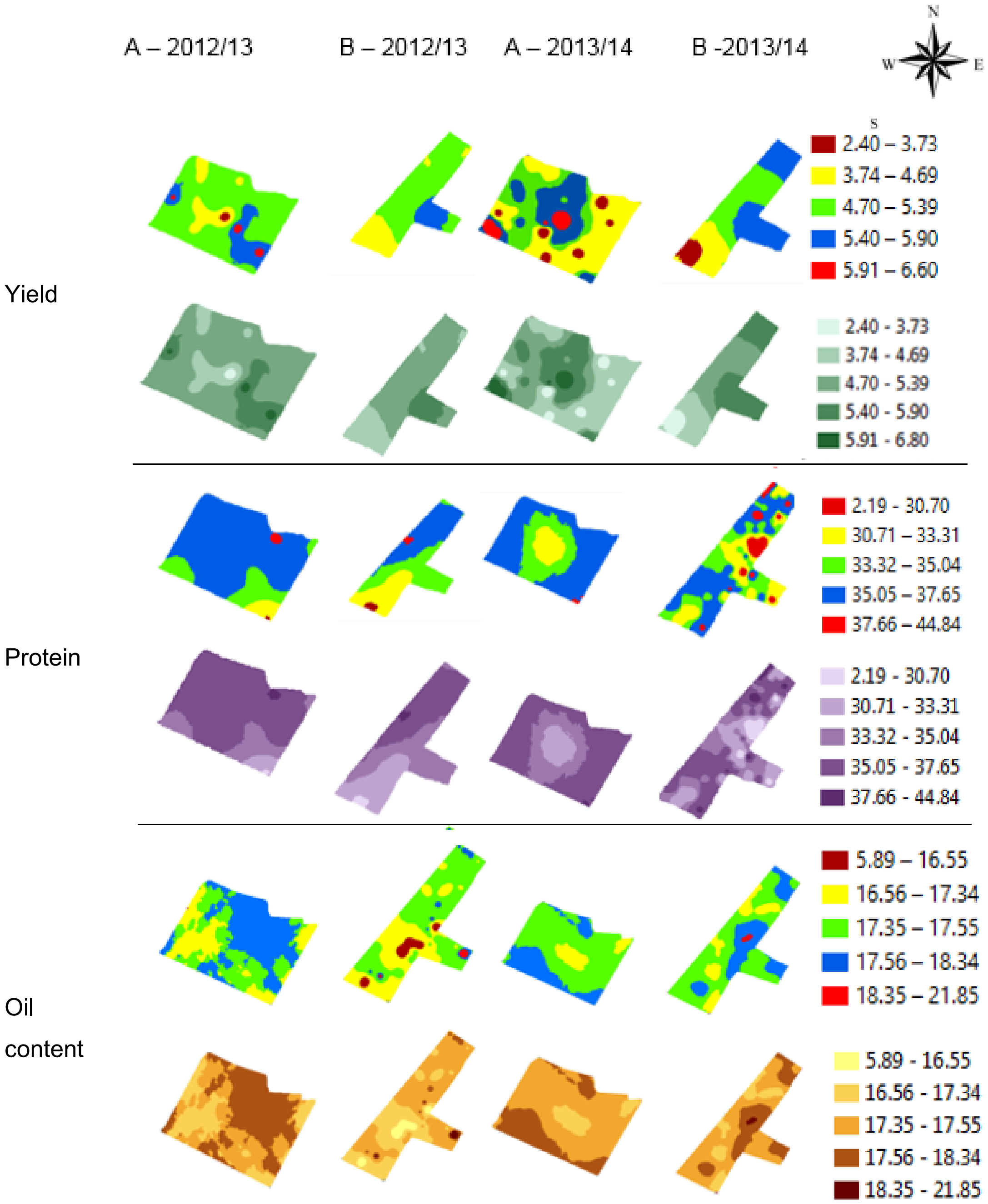


Figure 8. Thematic maps of soybean yield and the corresponding protein and oil content for fields A and B in 2012 and 2013, using a sequential (different colors) and nominal (single-hue with different lightness/saturation) color scheme
Source: adapted from Silva (2016).

3.2.1.2. Yield, profit, and profitability maps

Bazzi et al. (2015) studied the economic viability of agricultural products using profit and profitability maps. For each data set, yield, profit, and profitability maps (Figure 9) were generated using the following interpolation methods: inverse of distance (ID), inverse of square distance (IDS), and kriging (KRG). The authors concluded that profit and profitability maps are important tools to diagnose the spatial variability of economic return because they have assisted farmers in management decision-making. The impact of the interpolator type was less than 200 kg ha^{-1} for yield, US$ 30 ha^{-1} for profit, and 7% for profitability. Figure 9 shows that, in this 45-ha area, there are variations from 2.5 to 5.5 t ha^{-1} for yield, from -300 to 450 $ ha^{-1} for profit, and from -45 to 45% for profitability.

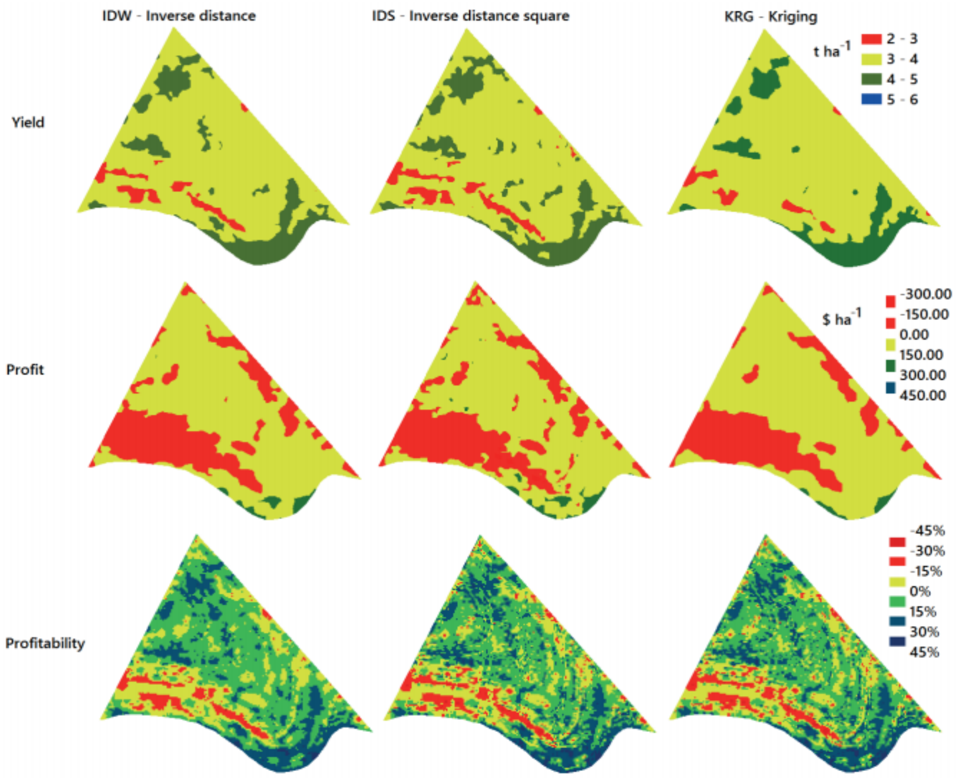


Figure 9. Yield, profit, and profitability maps for the 2006 Soybean harvest using the interpolation methods (i) inverse distance weighted (IDW), (ii) inverse distance weighted squared (IDS) and (iii) kriging (KRG). The production cost and sale prices of the product were obtained in the harvest month in a 45-ha field.
Source: Bazzi et al. (2015).

3.2.1.3. Grape yield maps

Martínez-Casasnovas and Bordes (2005) used information obtained from multispectral images to estimate crop vigor and to forecast yield (Figure 10) in Spain, at the wine farm of Raimat (Lleida).

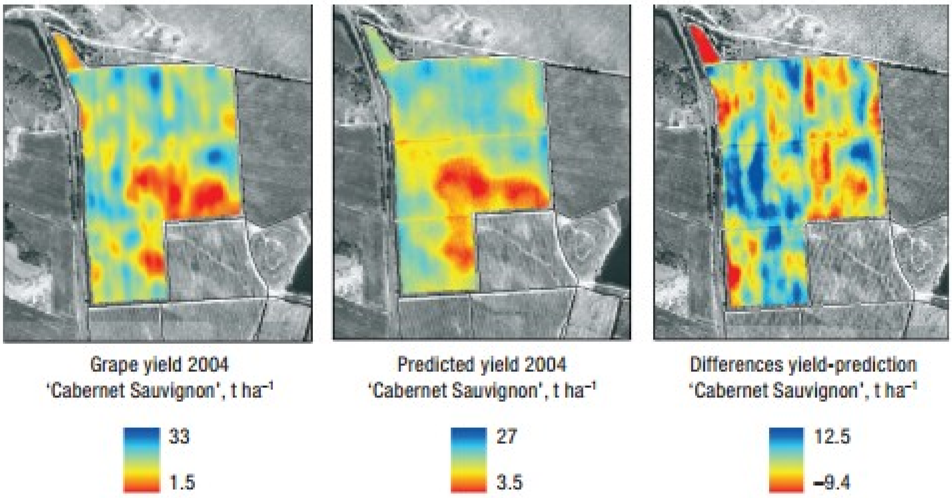
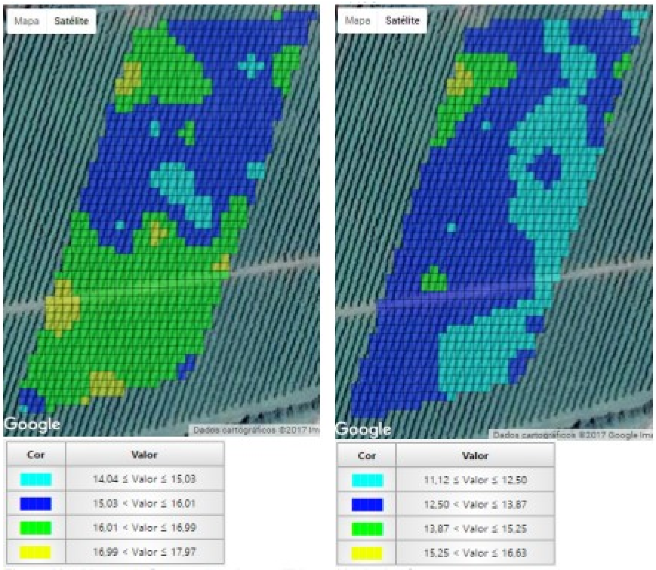


Figure 10. Comparison of the 2004 yield map of a 'Cabernet Sauvignon' plot (left) with the map obtained from a prediction model using NDVI (normalized difference vegetation index) from a QuickBird-2 multispectral image acquired one month before harvesting (center) (R^2 = 0.72). The map on the right shows the differences of both maps
Source: Martínez-Casasnovas and Bordes (2005).

3.2.1.4. Apple attributes maps

Longo (2017) developed a tool (apple show) to map the apple quality indices georeferenced and turn them into a graphics variable to provide support in the orchard management. Figure 11 presents the firmness of fruit pulp and total soluble solids of fruits in an area of 3.13 hectares.



a) b)

Figure 11. The firmness of fruit pulp (a) and total soluble solids of apple fruits (b) in a 3.13-ha area
Source: adapted from Longo (2017).

3.2.1.5. Weed infestation maps

Balastreire and Baio (2001) evaluated a practical method for weed mapping by driving over the patch contour with an all-terrain vehicle. Figure 12 presented a weed map showing three infestations levels. An important conclusion obtained was that timing to perform the weed mapping is a crucial factor to be considered for site-specific chemical applications.

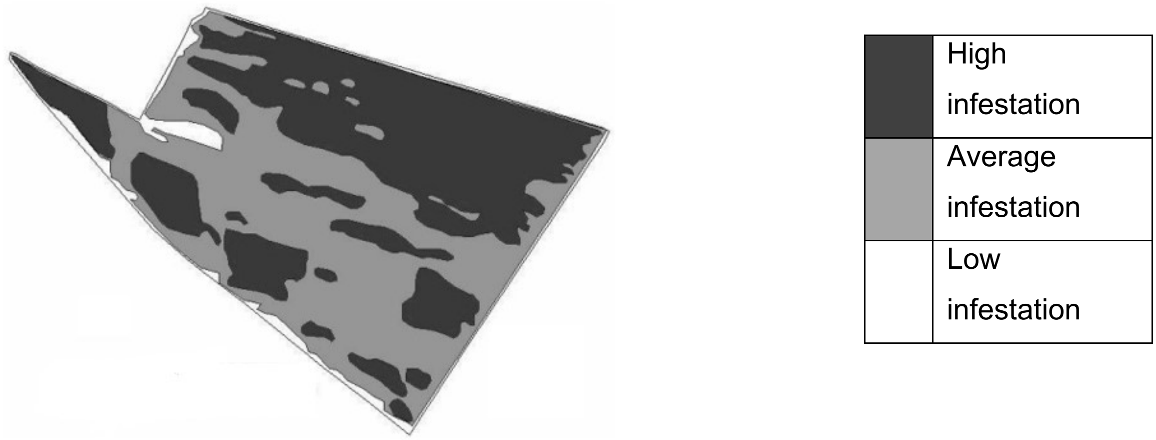


Figure 12. Weed maps showing three infestations levels from a 72-ha flat terrain, planted in no-tillage system and with soil covered by soybean stubble
Source: adapted from Balastreire and Baio (2001).

3.2.1.6. Dry matter yield, stocking rate, and milk yield maps

Bernardi et al. (2016) evaluated the spatial variability of soil properties, yield, lime and fertilizer needs, and economic return of an alfalfa grassland. The study was conducted in a 5.3-ha irrigated alfalfa grassland in São Carlos, SP, Brazil, directly grazed and intensively managed in a 270-paddock rotational system. According to them, the stocking rate is a key-management variable to determine productivity and profitability of grazing systems, and Figure 13 illustrates that the simulation based on dry matter yield allowed estimation of stocking rates and milk yield within the area. Therefore, these kinds of maps may be used to avoid over-or under-grazing. In addition, this study showed the methodology's advantages that allow identifying areas for differentiated paddocks management instead of homogeneous fertilizer application.

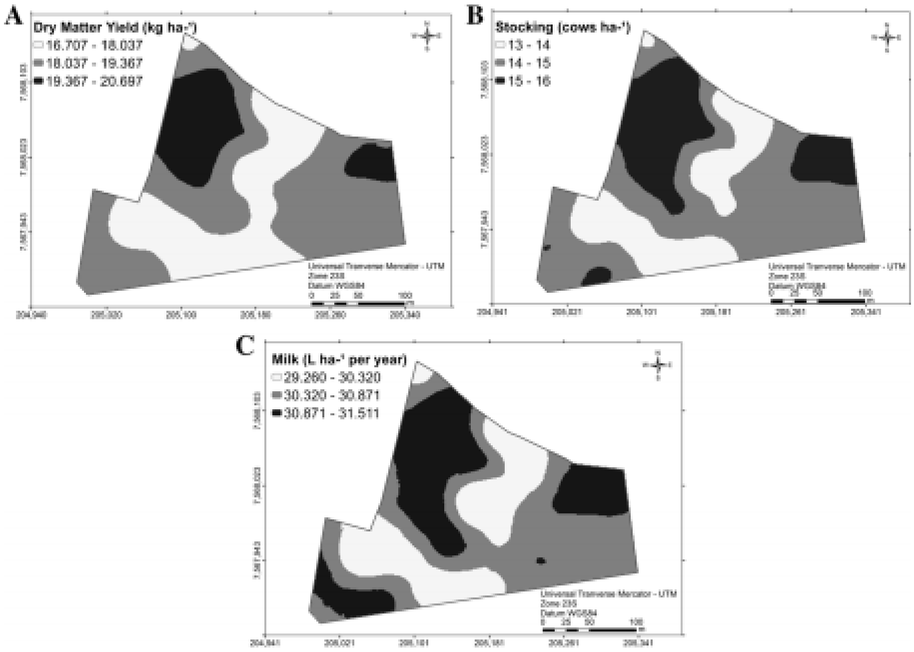


Figure 13. Kriged maps for dry matter yield (a), stocking rate (b), and milk yield (c) of a grazed alfalfa pasture in Brazil
Source: Bernardi et al. (2016).

3.3 MANAGEMENT ZONES (MZs)

MZ is a kind of choropleth map that is a sub-region of a field that expresses a functionally homogeneous combination of yield-limiting factors. However, despite this original concept of an MZ, the target agricultural variables can be other than yield, like pest and disease infestation, water content, Brix, soil resistance to penetration, and crop quality. An MZ can be used for one year, but also for several ones (usually three to five). This fact is essential when we are choosing variables. If we are planning to use only once, as in weed infestation, we can use variables that are not temporally stable to delineate the MZs. However, in most cases, we want to use MZs during several years, and we should use relatively temporally stable variables like topography data (elevation and slope) and physical data (Bulk density, soil texture, soil penetration resistance – SPR).

Considering the importance of MZs delineation in the current PA context, we made a systematic literature study (SLS) that had as the primary focus to identify researches about MZs delineation , as well as reporting the results of their use and synthesizing evidence that allows a common understanding of this research area. In this SLS, we used three techniques: (i) systematic literature mapping (SLM), which identifies searches in a given topic by choosing keywords and conducting database searches, (ii) snowballing (SB), which expands the initial selection by adding new studies to the classification process, consulting the references of the selected studies, (iii) systematic literature review (SLR), which summarizes the studies identified with SLM and SB.

3.3.1 Systematic Literature Study (SLS)

As mentioned, three steps were followed for the study:

Step 1 – Systematic Literature Mapping (SLM): The SLM was developed according to the following sequence of steps: keywords definition , scientific databases selection, determination of study selection criteria, study analysis, and synthesis methodology (Kitchenham and Charters, 2007; Talavera et al., 2017).

The following questions were asked to define the keywords: (1) what are the procedures and protocols to delineate MZs? (2) what are the most common algorithms to delineate MZs? (3) how the ideal number of MZs classes was found out? (4) what are the economic or environmental advantages to adopt MZs? (5) which software was used to delineate MZs? The site of Coordination for the Upgrading of Higher Education Personnel (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES) was accessed to obtain some scientific databases by remote access platform of the Federated Academic Community, CAFe (*Comunidade Acadêmica Federada*) (Fig. **14**). We searched for four considered relevant databases on this researched area: Scopus, Science Direct, Web of Science, and Wiley.

The period covered by SLM searches was limited from 2008 to 2019 to present the most recent papers. However, SB had rescued relevant researches that were left behind. The standardized information extracted from all papers were: title, authors, journal, publication volume, the country in which the research was developed, year of publication, abstract, DOI, software used, and results.

Step 2 - Snowballing (SB): The SB is characterized by the addition of new references to classify the process by consulting the references of selected studies and share references from people with knowledge in the area, thus, characterizing a sample of chain references (Biernacki and Waldorf, 1981; Cohen and Arieli, 2011). In this book, snowballing was used as a complementary strategy to increase efficiency and quality of the search, reducing the chances to obtain a search bias (Cohen and Arieli, 2011) and rescue important classic texts from the period before 2008. It is important to observe that no survey method is 100% effective, but the combination of both techniques is expected to reduce omission problems.

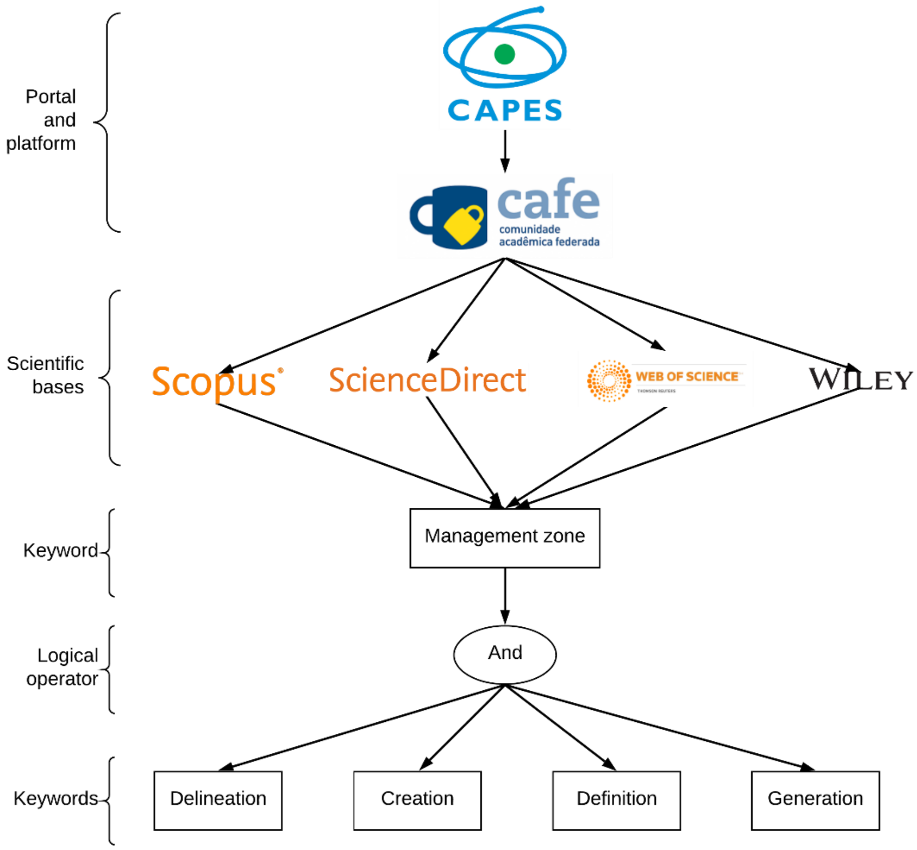


Fig. 14. Workflow used with the Systematic Literature Mapping (SLM)

Snowballing can be categorized as backward snowballing (BSB) or forward snowballing (FSB). New papers are included with BSB, based on the list of references raised with SLM. Nevertheless, with FSB, new studies are included from the list of references of ones selected by BSB (Wohlin, 2014).

The flow to select works used in SLS is shown in Figure 15 and Figure 16. Table 1 presents the 165 studies selected based on the research technique (SLM, BSB, and FSB). Only researchers directly related to agriculture (excluding, for example, those related to forestry or geological management) and that have explained the MZs delineation process were kept. Studies that delineate MZs using algorithms based only on images (without other layers of information or indexes) were also excluded.

Table 1 Clustering of studies selected by research technique

Techniques	studies
Systematic literature mapping (SLM) 96 references	(Ikenaga and Inamura, 2008; Kyaw et al., 2008; Mishra et al., 2008; Molin and de Castro, 2008; Robertson et al., 2008; Velandia et al., 2008; Vitharana et al., 2008b; Vitharana et al., 2008a; Li et al., 2008; Morari et al., 2009; Song et al., 2009; Xin-Zhong et al., 2009; Buttafuoco et al., 2010; Castrignanò et al., 2010; Fu et al., 2010; Guastaferro et al., 2010; Moral et al., 2010; Aimrun et al., 2011; Arno and Martinez-Casasnovas, 2011; Moral et al., 2011; Salami et al., 2011; Suszek et al., 2011; Jiang et al., 2011; Bansod et al., 2012; Davatgar et al., 2012; Jiang et al., 2012; McClymont et al., 2012; Roberts et al., 2012; Valente et al., 2012; Aggelopooulou et al., 2013; Alves et al., 2013; Bazzi et al., 2013; Benedetto et al., 2013b; Benedetto et al., 2013a; Cid-Garcia et al., 2013; Córdoba et al., 2013; Diacono et al., 2013; Li et al., 2013; Lin et al., 2013; Meirvenne et al., 2013; Peralta et al., 2013; Peralta and Costa, 2013; Ruß, 2013; Santesteban et al., 2013; Santi et al., 2013; Scudiero et al., 2013; Tagarakis et al., 2013; Chang et al., 2014; Galambošová et al., 2014; Gozdowski et al., 2014; Patil et al., 2014; Urretavizcaya et al., 2014; Yao et al., 2014; Bazzi et al., 2015; Caires et al., 2015; Landrum et al., 2015; Rodrigues and Corá, 2015; Santos and Saraiva, 2015; Tripathi et al., 2015; Peralta et al., 2015; Boluwade et al., 2016; Cavallo et al., 2016; Córdoba et al., 2016; Damian et al., 2016; Gavioli et al., 2016; Oldoni and Bassoi, 2016; Ortuani et al., 2016; Schenatto et al., 2016b; Shaddad et al., 2016; Shamal et al., 2016; Sobjak et al., 2016; Xiaohu et al., 2016; Bottega et al., 2017; Buttafuoco et al., 2017; Gili et al., 2017; González-Fernández et al., 2017; Jacintho et al., 2017; Schenatto et al., 2017b; Servadio et al., 2017; Yari et al., 2017; Shukla et al., 2017; Agati et al., 2018; Albornoz et al., 2018; Behera et al., 2018; Bernardi et al., 2018; Betzek et al., 2018; Karlik et al., 2018; Miao et al., 2018; Schwalbert et al., 2018; Scudiero et al., 2018; Whetton et al., 2018; Martínez-Casasnovas et al., 2018; Khan et al., 2018; Verma et al., 2018; González-Fernández et al., 2019; Moral et al., 2019)
Backward Snowballing (BSB) 18 *references*	(MacQueen, 1967; Bezdek, 1981; McBratney and Moore, 1985; Hotelling, 1933; Odeh et al., 1992 Gnanadesikan et al., 1995; Dobermann et al., 2003; Hornung et al., 2006; Dray et al., 2008; Schenatto et al., 2016a; Nawar et al., 2017; Souza et al., 2018; Albornoz et al., 2019; Betzek et al., 2019; Gavioli et al., 2019; Loisel et al., 2019; Bazzi et al., 2019; Nascimento et al., 2019)
Forward Snowballing	(Biernacki and Waldorf, 1981; Rousseeuw, 1987; Webster, 1990; Blackmore and Moore, 1999; Khosla and Alley, 1999; Blackmore,

(FSB) 51 *references*	2000; Doerge, 2000; Fleming et al., 2000; Fridgen et al., 2000; Fraisse et al.,, 2001; Boydell and McBratney 2002; Franzen et al., 2002; Khosla et al., 2002; Kitchen et al., 2002; Minasny and McBratney, 2002; Molin, 2002; Ping and Dobermann, 2003; Taylor et al., 2003; Adamchuk et al., 2004; Fridgen et al., 2004; Amidan et al., 2005; Brock et al., 2005; Jaynes et al., 2005; Kitchen et al., 2005; Frogbrook and Oliver, 2007; Kitchenham and Chartes, 2007; Li et al., 2007; Sudduth and Drummond, 2007; Taylor et al., 2007; Xiang et al., 2007; Gonzales and Woods, 2008; Inman et al., 2008; Coelho et al., 2009; Kitchenham et al., 2009; Zhang et al., 2010; Cohen and Arieli, 2011; Kuang et al., 2012; NIST/SEMATECH, 2012; Hörbe et al., 2013; Baudron and Giler, 2014; Wohlin, 2014; Mieza et al., 2016; Mulla and Khosla, 2016; Arango et al., 2017; Talavera et al., 2017; Schemberger et al., 2017; Schenatto et al., 2017a; Yang et al., 2017; Ortuani et al,. 2019; Reyes et al., 2019; Vega et al., 2019)

After the selection, the studies (165) were clustered in chronological order of publication (Figure 17). A smooth growth tendency was observed until 2013, which presented the largest number of studies (19). In 2014, there was a decrease on the number of these studies (8 studies), but with a tendency to increase in later years.

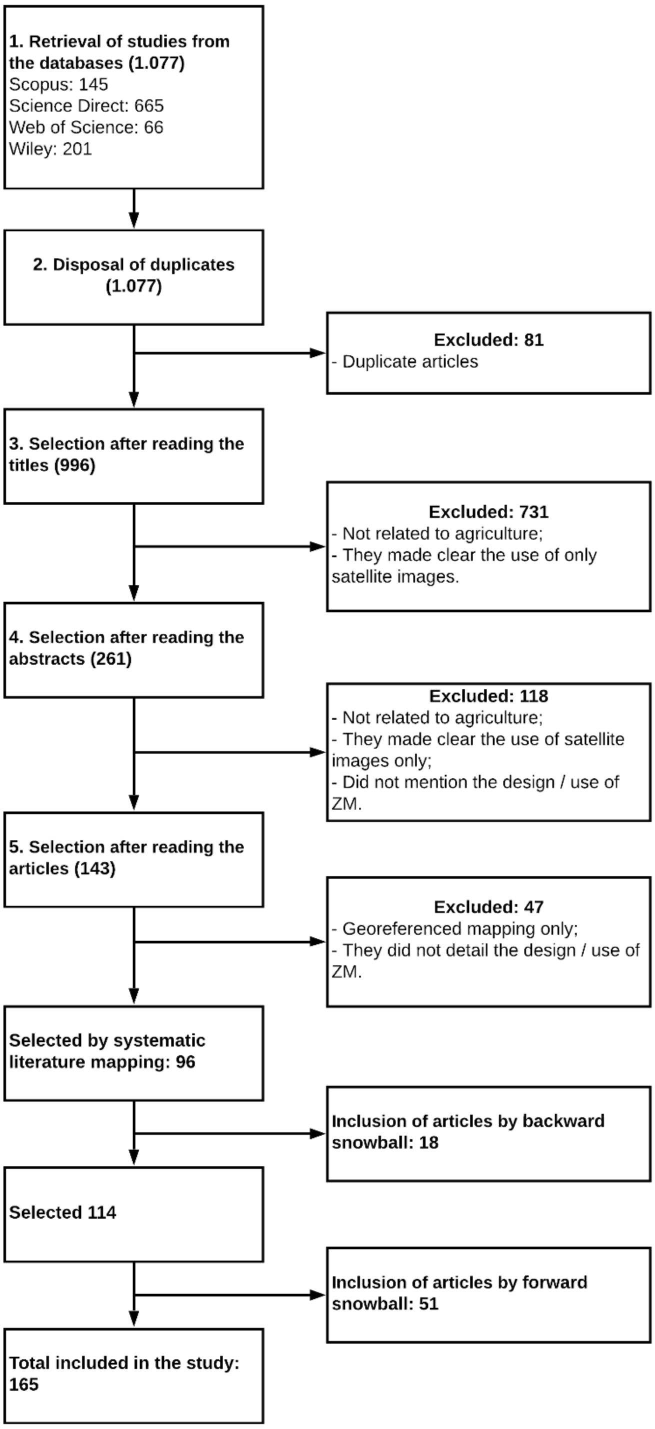


Figure 15. Workflow used for the Systematic Literature Study (SLS)

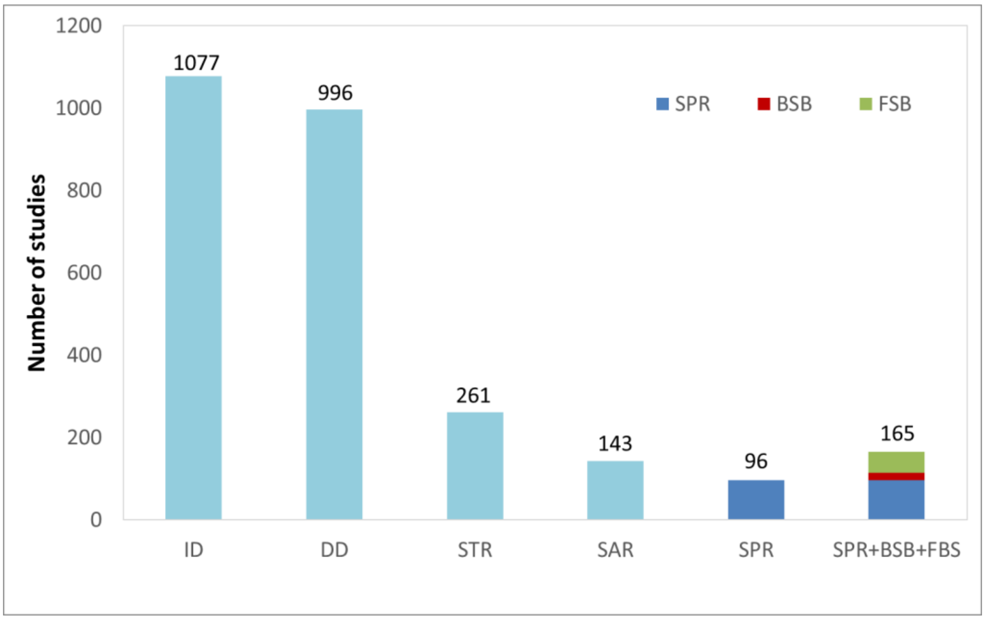


Figure 16. Stages followed in the systematic literature study (SLS) to select the primary papers: Identification (ID); discarding duplicates (DD), selection by title reading (STR); selection by abstract reading (SAR), selection by paper reading (SPR), adding by Backward Snowballing (BSB), and adding by Forward Snowballing (FSB)

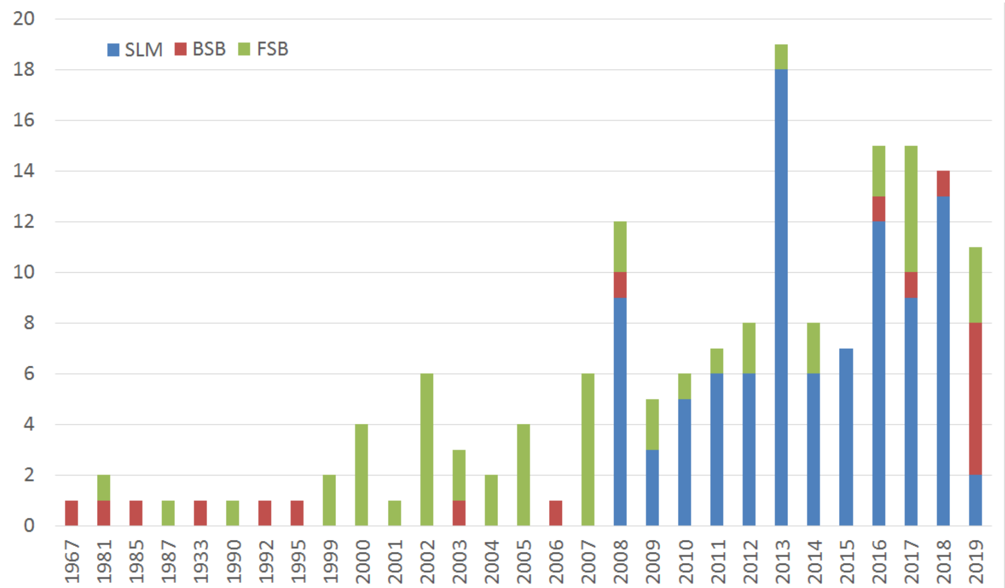


Figure 17. Quantity of selected studies classified in chronological order and type of research technique, Systematic Literature Mapping (SLM), Backward Snowballing (BSB), and Forward Snowballing (FSB)

Except for Antarctica, all continents were represented by at least one of the selected studies (Fig. 18). They were classified by the country where the authors conducted the research and when is only a theoretical manuscript, where it was published. Regarding the distribution vehicle (Figure 19), the journals *Computers and Electronics in Agriculture* and *Precision Agriculture* presented the best selected studies, with 17 and 14%, respectively. Journals and publishers that showed less than three studies in the review were clustered into a single item named as “others”.

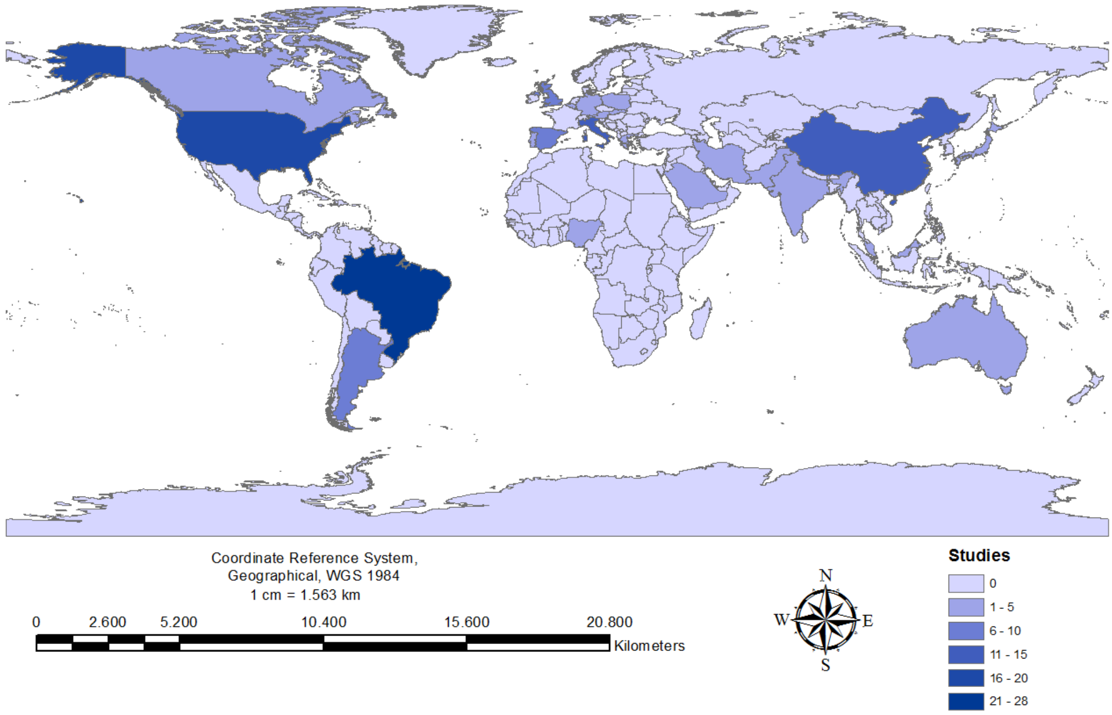


Figure 18. Distribution of selected studies by country of study (classified by the country where the authors conducted the research, and when it is only a theoretical manuscript, where it was published)

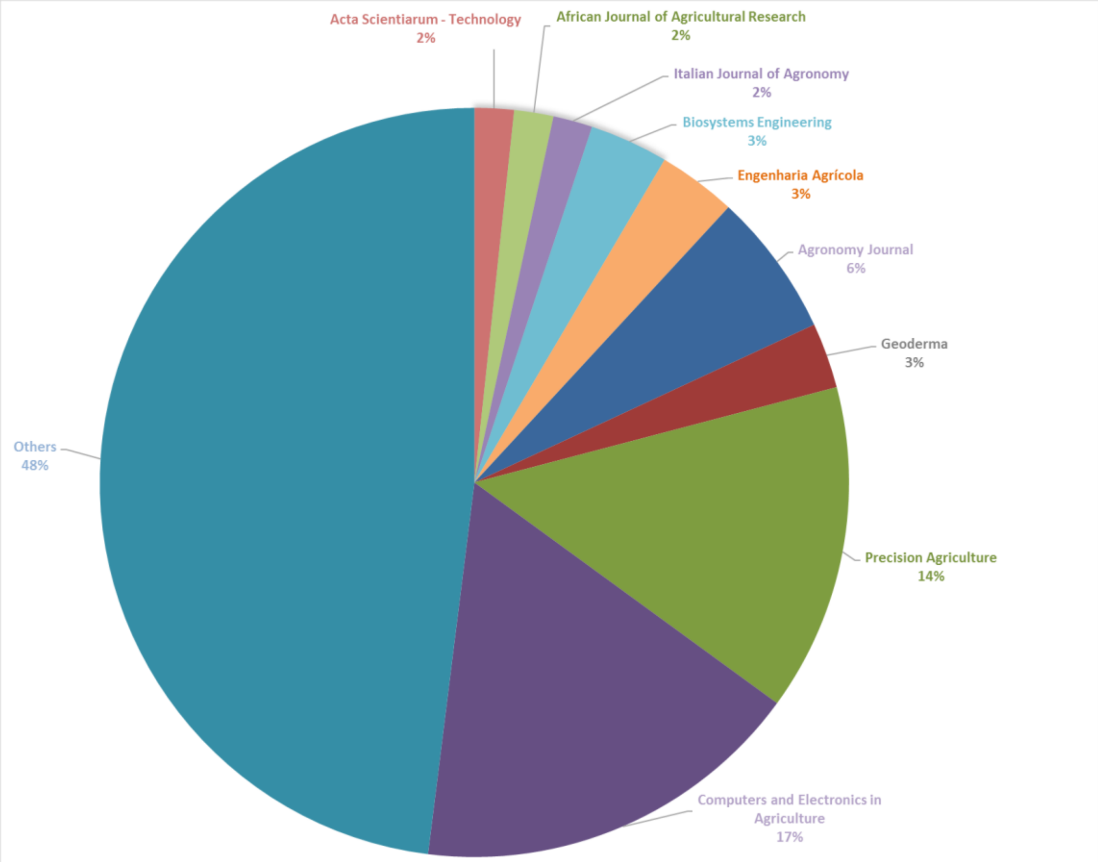


Figure 19 Distribution of selected papers by journals in which the selected studies were published

Step 3 - Systematic literature review (SLR): After identifying the relevant scientific papers using SLM and SB, SLR (Kitchenham et al., 2009) was conducted to aggregate the existing information on each researched question.

3.3.1.2 Results and discussion of SLS

The terms MZ and management class (MC) are frequently used in PA literature and often interchangeable terms. However, these terms are not identical. An MC is an area in which a particular treatment may be applied. A management zone is a spatially contiguous area to which a specific treatment may be used. Thus, an MC may consist of numerous zones, whereas an MZ can contain only one MC (Taylor et al., 2007).

Procedures and protocols for management zones delineation (Question 1)

Much effort has been made and it is being made to define the best delineation process for MZs. While some studies focus on creating a protocol that includes the entire process, from the initial treatment of the variables to the evaluation of the result, others work on specific parts of the process.

In this survey, we found out four studies that define a complete protocol to delineate MZs, but only one considers temporal issues. The first one, developed by Santos e Saraiva (2015), uses the Business Process Model and Notation (BPMN) to facilitate the interpretation. The authors proposed five macro steps: (1) data collection, (2) data filtering, (3) data selection, (4) data clustering, and (5) map evaluation. Each macro step is subdivided into several steps, some with sequential and others with iterative flow. Córdoba et al. (2016) proposed a seven-step protocol: (1) conversion of spatial coordinates, (2) removal of outliers, (3) removal of inliers, (4) spatial interpolation, (5) multivariate site classification, (6) smoothing of classification results, and (7) smoothing of classification results. A script in the R language is also available since it contains codes ready to execute each of the steps.

Souza et al. (2018) presented a more specific protocol, divided into nine main stages: (1) selection of the coordinate system, (2) removal of the outliers and inliers, (3) data normalization, (4) variable selection which will be used to delineate MZs, (5) data interpolation, (6) MZs delineation, (7) MZs rectification , and (8) selection of optimal number of MZs, and (9) MZs evaluation. Although there are subtle differences among the cited protocols, all primarily perform the same tasks and are very similar. The protocol proposed by Souza et al. (2018), considered more completed, is presented in Figure 20.

Differently of the three other protocols, the one outlined by Scudiero et al. (2018) takes into account variations between soil-plant, and consists of four main steps: (1) soil and time-specific plant spatial information acquisition, pre-processing interpretation, and interpolation, 2. time-specific sub-field soil-plant modeling, (3) time-specific MZ delineation with cluster analysis, and (4) evaluation and interpretation of MZs. The authors point out that traditional MZ delineation methods create static zones that are not ideal since the spatial patterns of soil-plant relationship change over time due to weather changes and/or other transient factors.

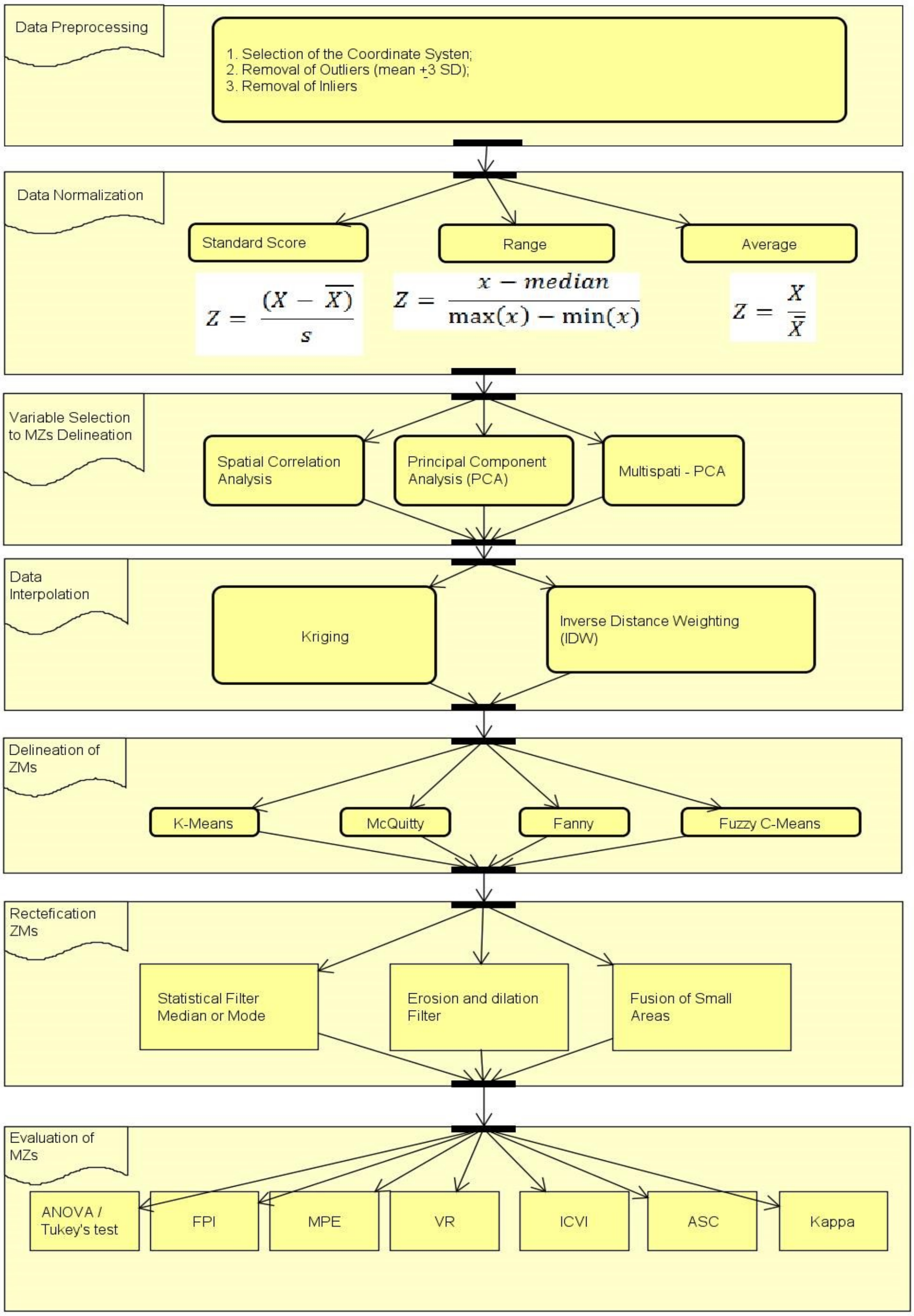


Figure 20. The protocol of management zones delineation , according to Souza et al. (2018). ANOVA: analysis of variance, SD: standard deviation, MZ: Management Zone, SD: standard deviation, ANOVA: analysis of variance, FPI: Fuzziness Performance Index), MPE: Modified Partition Entropy, VR: variance reduction, ICVI: improved cluster validation index, ASC: average silhouette coefficient
Source: Souza et al. (2018).

In addition to previous efforts to define a complete protocol to delineate MZs, some authors addressed specific issues at each stage, that is, they carried out studies aimed at improving part of the process. Thus, the studies selected by the research are organized below according to the sequence in the process:

1. Acquisition of variables: According to Nawar et al. (2017), the seven most common properties that can be used as an input variable to delineate MZs are related to:

Farmer knowledge – this knowledge may allow identifying different MZs in a field, based on the production history (Fleming et al., 2000, Khosla et al., 2002, Hörbe et al., 2013, Schenatto et al., 2017a).

Geomorphology – elevation is the most used topographic variable to delineate MZs. However, other variables such as elevation, slope, plan curvature, aspect, and depression depth have been successfully used (Jaynes et al., 2005). Another possibility is the topographic position index (Mieza et al., 2016).

Soil chemical and physical analyses – the soil chemical variables are often discarded to delineate MZs to be used for several years (Doerge, 2000) because of their temporal variability. However, they can be very interesting to delineated MZs to be used only once, as in the variable-rate fertilizer application. Nevertheless, soil physical variables, such as sand, silt and clay contents, organic matter, and soil water content are often used to delineate MZs (Doerge, 2000; Buttafuoco, 2010).

Soil class – the general sense is that soil maps, even with high resolution, are alone not enough to reliably identify crop productivity MZs since in a zone with the same soil series, many other variables can influence yields (as topography and chemical attributes) (Khosla and Alley, 1999; Franzen et al., 2002; Brock et al., 2005). In addition, Franzen et al. (2002) further reported that Order 1 soil survey maps (i.e., map scales of 1:5000 to 1:10 000) were helpful to develop Nitrogen-MZs.

Yield maps – they are the complete information to visualize the spatial variability of crops (Molin, 2002). However, their temporal variation hinder when a single-year yield map is used to delineate MZs reliably. Blackmore (2000) and Molin (2002) used normalized data from multiple years to make up for this problem. Although one-year yield data alone are not directly suitable for MZs determination, because their availability and low cost make them a valuable possibility to improve effectiveness of MZs delineation based on other information (Nawar, 2017). Two approaches are commonly used to delineate MZs using yield maps (Xiang et al., 2007): (1) the empirical method, which uses frequency distribution of yield and expertise knowledge to divide the field usually in three or four MZs (Blackmore, 2000), and (2) cluster analysis such as K-means and fuzzy c-means (FCM) (Taylor et al., 2003; Taylor, Mcbratney, and Whelan, 2007; Li et al., 2007) (USAR VÍRGULAS - APAGUE). and/or iterative self-organizing of data analysis technique (Fridgen et al., 2000; Kitchen et al., 2002).

Crop coverage – the most used information about crop coverage are vegetation indices (VI) and leaf area index (LAI). Both of them can be manually measured and with remote sensing (RS) methods. Traditional methods to acquire crop traits (plant height, leaf color, LAI, chlorophyll content, biomass, yield) rely on manual sampling, which is time-consuming and laborious (Yang et al., 2017). However, RS platforms, like unmanned aerial vehicles (UAV), equipped with different sensors, are currently an important approach. The most common RS application in PA is detecting spatial and temporal patterns in crop nutrient deficiencies (Mulla and Khosla, 2016), and it can provide information about photosynthetically active biomass – ie canopy health and vigor. Several authors use RS data to delineate MZs based on RS data alone (Inman et al., 2008; Song et al., 2009; Chang et al., 2014) or to improve effectiveness of MZs delineation based on other information (Li et al., 2007; Inman et al. 2008; Song et al. 2009; Ortuani et al., 2019, Tagarakis et al,. 2013).

Proximal soil sensors – conventional soil sampling and analyses have shown mixed economic returns due to the high costs associated with labor-intensive sampling and analysis procedures, which map uncertainties might accompany. Therefore, conventional laboratory methods are being replaced or complemented with analytical soil sensing techniques (Kuang et al., 2012). Typically, sensor sampling is taken at fixed intervals using a vehicle while driving along straight parallel lines, thus, there is a result in a regular grid of sample points, which produces a fine-resolution spatial data that can reveal detailed spatial patterns of measured parameters (e.g., electrical, optical, mechanical, electrochemical, acoustic, and pneumatic) (Nawar, 2017; Adamchuk et al., 2004; Kuang et al., 2012).

Scudiero et al. (2013) emphasized the potential of using multiple-sensor platforms to delineate MZs. For example, they combined two proximal-sensing (the apparent electrical conductivity of soil (ECa) and bare-soil NDVI) data and FCM algorithm to divide a 21-ha cornfield into five zones. The authors highlighted that even when measurements as ECa and bare-soil NDVI are not directly correlated to a corn yield, their combined use could help on classifying the soil according to its fertility.

2. *outliers and inliers Removal*: exploratory data analysis (EDA) is a summary of data set according to their main characteristics and employs a variety of techniques (mostly graphical) to maximize insight into a data set (NIST/SEMATECH, 2013). When constructing MZs, the essential use of EDA is to detect and remove outliers. According to Amidan et al. (2005), data outliers can significantly impact data-driven decisions, and

in many cases, they do not reflect the true nature of data and, hence, they should not be included in the analyses. According to Córdoba et al. (2016), the values outside the mean ± 3 SD are identified as outliers and should be removed. They remarked that even though real data could belong to this interval, the upper and lower limits should be modified to obtain robust variance estimators. It is also necessary to remove inliers, data that differ significantly from their neighborhood but lie within the variation range of data set (Córdoba et al., 2016). An additional care should be taken for yield data obtained with yield monitor. Many approaches for yield data removal were already being proposed (like by Blackmore and Moore, 1999) to eliminate errors associated with unknown header width, combine filling/emptying times, time lag of grain with the combine, positional errors, rapid velocity changes, and others (Sudduth and Drummond, 2007). Vega et al. (2019) proposed a protocol to automate error removal from yield maps divided in two steps: (1) removal of yield data with values equal to zero, removal edge values and potential end-of-field yield monitor errors, and removal of yield data that are outside the mean ± 3 SD; and (2) use of the local Moran's spatial autocorrelation index and Moran's plot to identify and remove data that are inconsistent with their neighbor points. The protocol was evaluated on 595 real yield datasets with good results and can be used with other geo-referenced variables in precision agriculture.

3. Data normalization: some clustering techniques such as FCM algorithm with Euclidean are sensitive to characteristics of the input variables. Fridgen et al. (2004) reported that Euclidean distance should be used only for statistically independent variables to show equal variances. In this sense, when the Euclidean distance is used to clustering, normalization data can be a crucial step before creating MZs (Schenatto et al. 2017b). Schenatto et al. (2017b) evaluated the influence of using three data normalization methods (amplitude, mean, and standard score) to delineate MZs with FCM algorithm using Euclidean distance, with corn and soybean data. The authors concluded that the amplitude normalization method was the most appropriate.

4. Selection of input variables: The selection of variables that are most related to the target variable, usually crop yield, can be done before or after delineating MZs. According to Gnanadesikan et al. (1995), the weighting and selection of variables are the most challenging issues in cluster analysis. However, the capacity of clustering software to process a large number of variables encourages users to be generous in the number of variables used in the process. Furthermore, the variable choice (as well as their weights) can and often will influence clustering (MZs delineation) (Gozdowski, 2014; Sobjak, 2016). Sobjak et al. (2016) showed that with FCM algorithm, no combination of variables produced statistically better performance than the MZ delineated only with non-redundant variables. Therefore, the selection of variables before the delineation process is recommended.

4.1 Selection of variables before the delineation process: in this case, techniques are applied to reduce the variables' number and/or dimensionality. The use of redundant variables decreases clustering performance and increases computational time (Bazzi et al., 2013; Schenatto et al., 2016a; Sobjak et al., 2016). Good results were obtained with multivariate techniques to reduce the dimensionality of variables and promote orthogonality among them (Hotelling, 1933; Dray et al., 2008; Gavioli et al., 2016). Three-variable selection techniques (Table 2) that are most used in combination with FCM algorithm are:

Spatial correlation analysis (Bazzi et al., 2013): is a method using Moran's bivariate spatial autocorrelation statistic to build a spatial correlation matrix. The procedure was: (1) elimination of variables with no significant spatial autocorrelation at 5% significance; (2) removal of variables that were not correlated with yield (or other target variables); (3) decreasing ordination of the remaining variables, considering the degree of correlation with yield; and (4) elimination of variables which are correlated with each other, with preference to remove those variables with lower correlation with yield.

Principal component analysis (PCA) (Hotelling, 1933): is a multivariate technique that consists of building a new set of orthogonal synthetic variables denominated principal components (PC) and is the most frequently reported technique (Table 2) in the process of selection/reduction of variables to delineate MZs. These PCs are linear combinations of the original variables whose results come from transformations that prioritize the representation of data variability in the first components. Thus, if the original variables have a high degree of dependence among them, it is possible to reduce data dimensionality using the first PCs. Another possibility is to select only the variables that had the most significant influence on PCs delineation.

Multivariate spatial analysis based on Moran's index PCA (MULTISPATI-PCA) (Dray et al., 2008; Córdoba et al., 2013; Gavioli et al., 2016): is an extension of PCA that adds spatial restriction considering data georeferencing (and, therefore, spatial dependence) by adding a spatial weighting matrix created with Moran's bivariate spatial autocorrelation statistic. The MULTISPATI-PCA aims to maximize the spatial autocorrelation between the points, while the traditional PCA, aims to maximize the total variance.

Table 2 Main techniques for variables' selection and management zones (MZs) delineation

Selection of techniques	N° of papers	Papers
Spatial correlation analysis	10	(Bazzi et al., 2013; Gavioli et al., 2016; Schenatto et al., 2016a; Schenatto et al., 2016b; Sobjak et al., 2016; Jacintho et al., 2017; Schenatto et al., 2017b; Betzek et al., 2018; Bazzi et al., 2019; Betzek et al., 2019)
Principal Component Analysis (PCA)	38	(Fraisse et al., 2001 ; Li et al., 2007; Molin and de Castro, 2008; Vitharana et al., 2008a; Morari et al., 2009; Xin-Zhong et al., 2009; Buttafuoco et al., 2010; Castrignanò et al., 2010; Guastaferro et al., 2010; Moral et al., 2010 ; Salami et al., 2011; Jiang et al., 2011; Bansod et al., 2012; Davatgar et al., 2012; Jiang et al., 2012; Benedetto et al., 2013b; Li et al., 2013; Lin et al., 2013; Meirvenne et al., 2013; Peralta et al., 2013; Peralta and Costa, 2013; Urretavizcaya et al., 2014; Yao et al., 2014; Caires et al., 2015; Landrum et al., 2015; Tripathi et al., 2015; Córdoba et al., 2016 ; Gavioli et al., 2016; Ortuani et al., 2016; Buttafuoco et al., 2017; Gili et al., 2017; Shukla et al., 2017; Behera et al., 2018; Schwalbert et al., 2018; Scudiero et al., 2018; Verma et al., 2018; González-Fernández et al., 2019; Reyes et al., 2019)
MULTISPATI-PCA	4	(Córdoba et al., 2013; Peralta et al., 2015; Gavioli et al., 2016; Gili et al., 2017)

Gavioli et al. (2016) evaluated the efficiency of each of these three techniques (spatial correlation analysis, PCA, MULTISPATI-PCA) and two new methods proposed by them. One, the MPCA-SC, is based on the combined use of spatial correlation analysis and MULTISPATI-PCA, and the other, PCA-SC, which applies PCA only to the stable variables that showed significant spatial correlation with the target variable (selected by the spatial correlation matrix). They found out that MPCA-SC provided the best performance for FCM algorithm, reduced data dimensionality without losing essential information in most cases.

4.2 Selection of variables after the delineation process: Although selecting variables before the delineation process is the most common, some authors decided to proceed it after. These are the cases:

Kitchen et al. (2005) compared the productivity zones (SPZ) delineated using ECa and elevation with the ones from yield map data (YPZ). Using overall accuracy and Kappa coefficient K, they found out the best combination of ECa and/or elevation data combinations. They considered it a promising level of agreement (until 60–70%), especially due to many other yield-limiting factors unrelated to ECa and elevation.

Gozdowski et al. (2014) used logistic regression to find out which soil attributes were more correlated with the MZs delineated according to the multiyear mean standardized yield divided each field into two zones, one above and the other below the mean yield. They analyzed several variables, including soil chemical and physical properties and topographic attributes, and concluded that soil sand and organic carbon content produced the most proper delineation of MZs.

Bottega et al. (2017) delineated MZs based on the one-year yield data and MZs based on ECa, coarse sand, fine sand, silt, clay, and combinations among them. They concluded that ECa provided the best agreement by using Kappa coefficient.

Miao et al. (2018) evaluated three approaches to delineate MZs on two no-till corn-soybean rotation fields: (1) ROSE-YSTTS, using relative elevation, organic matter (OM), slope, ECa), yield spatial trend map, and yield temporal stability map, (2) ROSE, using soil and landscape information (relative elevation, OM, slope, and ECa), and (3) CMYYM, using clustering multiple-year yield maps corn-soybean turnover. They evaluated the accuracy of different approaches using relative variance (Dobermann et al., 2003) and concluded that the ROSE-YSTTS approach could overcome the weaknesses of approaches using only soil, landscape, or yield information and is more robust for MZ delineation.

5*. Data interpolation:* data to delineate MZs are usually interpolated to delineate MZs that are continuous and smooth. Typically, this task is performed with inverse distance weighting (IDW) or kriging interpolation methods. Kriging is the best interpolator when a minimum spatial dependence is confirmed; otherwise, IDW presents an advantage (Betzek et al., 2019).

6. *MZs Delineation* - Two approaches are commonly used to delineate MZs: (1) the empirical method, which uses frequency distribution of target variable (usually yield) to divide the MZs field (Blackmore, 2000), and (2) cluster analysis such as K-means and FCM (Taylor et al., 2003; Taylor et al., 2007; Li et al., 2007). The cluster analysis methods divide data points of an agricultural area into classes, which are also termed groups, by employing a similarity evaluation function for this division. In practice, these classes are applied to delineate MZs, which can be subsequently delimited in field (Boydell and McBratney, 2002; Córdoba et al., 2016).

7. MZs Rectification : After their delineation, MZs often present isolated pixels, small regions, or even a transition border among very irregular zones, making it difficult or even impossible to operate in a field. Thus, a smoothing process called rectification can be applied to optimize the zones. Betzek et al. (2018) presented a solution based on the filters mode and median application with 3 × 3 and 5 × 5 pixel masks. The best results were obtained with masks of 5x5 pixels, regardless if it is used mode or median. Gonzalez and Woods (2008), Córdoba et al. (2016), and Albornoz et al. (2018) used median and dilatation filters and erosion to reduce MZs fragmentation .

8. Evaluation of delineated MZs: the performance of a delineation process can be assessed using indices and analysis of variance (ANOVA). The most used statistics are: (1) variance reduction (VR) (Ping and Dobermann, 2003), (2) the fuzziness performance index (FPI) (Fridgen et al., 2004), (3) modified partition entropy (MPE) (McBratney and Moore, 1985), (4) normalized classification entropy (NCE) (Bezdek, 1981), (5) improved cluster validation index (ICVI) (Gavioli et al., 2016), (6) smoothness index (SI) (Gavioli et al., 2016), (7) average silhouette coefficient (ASC) (Rousseeuw, 1987), (8) Kappa coefficient (K) (Cohen, 1960), and (9) coefficient of relative deviation (CRD) (Coelho et al., 2009). Depending on the MZ delineation approach, only some of these indices can be used: FPI, MPE, NCE, and ICVI can only be used with fuzzy logic. These measures aim to quantify how heterogeneous the zones are across the studied field (important for MZ delineation or similarity among the zones (important for most segmentation algorithms, in their zone fusion stage), but not simultaneously.

In this sense, Loisel et al. (2019) presented a criterion that considers both questions, conducting tests on 50 hypothetical and one real database. Their results showed relevancy of the methodology to compare maps with different zones and to sort them and provided a ranked set of possible maps with different within-field zones.

Algorithms for management zones delineation (Question 2)

Many techniques and algorithms are available for each stage of the MZs delineation. Choosing the best algorithm is not a trivial task, and it should be conducted based on empirical analysis. However, although several statistical or even empirical approaches exist, the cluster methods, more specifically FCM, and k-means, are the most applied ones (Table 3).

FCM unsupervised classification algorithm (Bezdek, 1981), sometimes also named as Fuzzy K-means, produces a continuous cluster of objects considering the principles of fuzzy logic. It minimizes the variability within the cluster while it maximizes variability among them in order to create homogeneous clusters. In addition, the fuzzy logic principle allows a specific element to be contained in more than one cluster simultaneously by assigning degrees of permanence in each one, which reduces some possible distortion caused by outliers.

Table 3 Algorithms used for management zones (MZs) delineation

Algorithm	N° of papers	Papers
Fuzzy c-means (FCM)	74	(Bezdek, 1981; Boydell and McBratney, 2002; Kitchen et al.,, 2002; Fridgen et al.,, 2004; Kitchen et al.,, 2005; Li et al., 2007; Kyaw et al., 2008; Mishra et al., 2008; Molin and de Castro, 2008; Vitharana et al., 2008b; Vitharana et al., 2008a; Li et al., 2008; Morari et al., 2009; Song et al., 2009; Xin-Zhong et al., 2009; Fu et al., 2010; Guastaferro et al., 2010; Moral et al., 2010; Zhang et al., 2010; Arno and Martinez-Casasnovas, 2011; Jiang et al., 2011; Bansod et al., 2012; Davatgar et al., 2012; Jiang et al., 2012; McClymont et al., 2012; Roberts et al., 2012; Valente et al., 2012; Bazzi et al., 2013; Córdoba et al., 2013; Li et al., 2013; Lin et al., 2013; Meirvenne et al., 2013; Scudiero et al., 2013; Tagarakis et al., 2013; Chang et al., 2014; Patil et al., 2014; Urretavizcaya et al., 2014; Yao et al., 2014; Bazzi et al., 2015; Caires et al., 2015; Rodrigues and Corá, 2015; Santos and Saraiva, 2015; Tripathi et al., 2015; Peralta et al., 2015; Boluwade et al., 2016; Gavioli et al., 2016; Oldoni and Bassoi, 2016; Ortuani et al., 2016; Schenatto et al., 2016b; Schenatto et al., 2016a; Sobjak et al., 2016; Bottega et al., 2017; Gili et al., 2017; Schemberger et al., 2017; Schenatto et al., 2017b; Servadio et al., 2017; Yari et al., 2017; Shukla et al., 2017; Albornoz et al., 2018; Behera et al., 2018; Betzek et al., 2018; Miao et al., 2018; Schwalbert et al., 2018; Scudiero et al., 2018; Martínez-Casasnovas et al., 2018; Khan et al., 2018; Verma et al., 2018; Bazzi et al., 2019; Betzek et al., 2019; Gavioli et al., 2019; González-Fernández et al., 2019; Nascimento et al., 2019; Ortuani et al., 2019; Reyes et al., 2019)
K-means	18	(Taylor et al., 2003; Jaynes et al., 2005; Hornung et al., 2006; Xiang et al., 2007; Ikenaga and Inamura, 2008; Inman et al., 2008; Robertson et al., 2008; Arno and Martinez-Casasnovas, 2011; Meirvenne et al., 2013; Galambošová et al., 2014; Santos and Saraiva, 2015; Damian et al., 2016; Schemberger et al., 2017; Agati et al., 2018; Karlik et al., 2018; Whetton et al., 2018; Gavioli et al., 2019; Loisel et al., 2019)
non-parametric estimate of probability density function	5	(Castrignanò et al., 2010; Guastaferro et al., 2010; Aggelopooulou et al., 2013; Benedetto et al., 2013b; Benedetto et al., 2013a; Diacono et al., 2013)

Ordinary kriging / Factorial kriging / Factorial cokriging / Multicollocated cokriging / Multicollocated factor cokriging	5	(Buttafuoco et al., 2010; Landrum et al., 2015; Cavallo et al., 2016; Shaddad et al., 2016; Buttafuoco et al., 2017)
Ward	5	(Fleming et al., 2000; Salami et al., 2011; Santesteban et al., 2013; Galambošová et al., 2014; Gavioli et al., 2019)
ISODATA	3	(Fraisse et al., 2001; Guastaferro et al., 2010; González-Fernández et al., 2017)
RASCH	2	(Moral et al., 2011; Moral et al., 2019)
others	21	(Blackmore, 2000; Franzen et al., 2002; Molin, 2002; Frogbrook and Oliver, 2007; Velandia et al., 2008; Fu et al., 2010; Suszek et al., 2011; Bansod et al., 2012; Cid-Garcia et al., 2013; Hörbe et al., 2013; Peralta and Costa, 2013; Ruß, 2013; Gozdowski et al., 2014; Shamal et al., 2016; Xiaohu et al., 2016; Jacintho et al., 2017; Nawar et al., 2017; Schemberger et al., 2017; Bernardi et al., 2018; Gavioli et al., 2019; Reyes et al., 2019)

Three matrices are needed to develop FCM (McBratney and Moore, 1985). The first one, matrix **X**, consists of the data to be classified. The second one, Matrix **V**, is the matrix with centroids of clusters, and consists of k centroids of clusters contained in the searched space defined by matrix **X**. The third one, matrix **U**, consists of assigning the pertinence value of each cluster in **V** for each point in **X**, considering that the sum of pertinence values of each observation must be equal to 1. An ideal fuzzy k partitioning is defined as a weighted minimization of the square distance between the observation points and the centroid of the classes, according to Equation 3:

$$j_m(U,v) = \sum_{j=0}^{n}\sum_{i=1}^{k}\left(u_{ij}\right)^m\left(d_{ij}\right)^2 \qquad \text{Eq. 3}$$

where *m* is the fuzzy weighting coefficient (1≤*m*<∞) that controls the pertinence value shared among classes. The closer to 1, the smaller the sharing among classes; the closer to infinity, the greater the value of sharing pertinence resulting in less distinct classes; *u* represents the pertinence of an element in a class; and $(d_{ij})^2$ is the square of distance (usually Euclidean distance) in the space between point *j* and the centroid of class *i*.

Despite the lack of an explicit criterion to choose the parameter *m*, when related to agriculture, values between 1 and 2 are generally used, so that 1.3 and 1.5 are the most recurrent ones (Table 4).

Table 4 Values adopted for fuzzy weighting coefficient in fuzzy c-means algorithm

Exponent	N° of papers	Papers
1	1	(Servadio et al., 2017)
1,3	24	(Kitchen et al., 2002; Kitchen et al., 2005; Kyaw et al., 2008; Vitharana et al., 2008b; Morari et al., 2009; Moral et al., 2010; Arno and Martinez-Casasnovas, 2011; Roberts et al., 2012; Bazzi et al., 2013; Córdoba et al., 2013; Meirvenne et al., 2013; Tagarakis et al., 2013; Patil et al., 2014; Rodrigues and Corá, 2015; Peralta et al., 2015; Gili et al., 2017; Schenatto et al., 2017b; Yari et al., 2017; Betzek et al., 2018; Schwalbert et al., 2018; Martínez-Casasnovas et al., 2018; Khan et al., 2018; Betzek et al., 2019; Reyes et al., 2019)
1,5	13	(Fridgen et al., 2004; Vitharana et al., 2008a; Xin-Zhong et al., 2009; Davatgar et al., 2012; Jiang et al., 2012; Lin et al., 2013; Scudiero et al., 2013; Chang et al., 2014; Tripathi et al., 2015; Cavallo et al., 2016; Shukla et al., 2017; Albornoz et al., 2018; Behera et al., 2018)
2	2	(Valente et al., 2012; Alves et al., 2013)

Although Euclidean distance (Table 5) is usually used as a parameter for both FCM and k-means, it generates spherical clusters, hardly present in soil data, and it is sensitive to the amplitude (thus requiring data normalization) of variables when two or more input variables are used (Bezdek, 1981; Fridgen et al., 2004, Schenatto et al., 2017b). Mahalanobis distance is often used as an alternative, especially when clustering multivariate data since it adds intra-class variance restrictions to the calculation (Bezdek, 1981; McBratney and Moore, 1985).

Table 5 Types of distances used in Fuzzy c-means algorithm

Distance	N° of papers	Papers
Euclidian Distance	28	(Fraisse et al., 2001; Kyaw et al., 2008; Molin and de Castro, 2008; Robertson et al., 2008; Morari et al., 2009; Xin-Zhong et al., 2009; Guastaferro et al., 2010; Davatgar et al., 2012; Jiang et al., 2012; Roberts et al., 2012; Aggelopooulou et al., 2013; Alves et al., 2013; Benedetto et al., 2013b; Benedetto et al., 2013a; Lin et al., 2013; Scudiero et al., 2013; Tagarakis et al., 2013; Chang et al., 2014; Galambošová et al., 2014; Rodrigues and Corá, 2015; Tripathi et al., 2015; Damian et al., 2016; Oldoni and Bassoi, 2016; Ortuani et al., 2016; Gili et al., 2017; Whetton et al., 2018; González-Fernández et al., 2019; Reyes et al., 2019)
Mahalanobis	17	(Kitchen et al., 2002; Fridgen et al., 2004; Kitchen et al., 2005; Mishra et al., 2008; Vitharana et al., 2008a; Arno and Martinez-Casasnovas, 2011; Jiang et al., 2012; McClymont et al., 2012; Roberts et al., 2012; Córdoba et al., 2013; Tagarakis et al., 2013; Ortuani et al., 2016; Servadio et al., 2017; Yari et al., 2017; Martínez-Casasnovas et al., 2018; Khan et al., 2018; González-Fernández et al., 2019)

The k-means unsupervised clustering algorithm (MacQueen, 1967) aims to separate data set elements by clustering them into k sets. Initially, the algorithm chooses the position of k initial centroid points, usually randomly, within the set of points in matrix **X** and calculates the distance of all points (typically using Euclidean distance) to the centroids and assigns the location to the nearest centroid. That is, considering $x_j \in X$, it is associated with cluster C_{j^*} that has the closest z_i centroid (Equation 4). Once this assignment is made, the average distance from all points connected to a centroid is calculated. Subsequently, the centroid is repositioned at the average distance from all points connected to that centroid. This change can cause some points to be attributed to another centroid since it is always the nearest centroid. This procedure is repeated until no centroid has its position changed. This will occur when all the centroids are in the central position of the distance among the points take part of that centroid.

$$j^* = \underset{i=1,\dots,k}{\operatorname{argmin}}\{|x_j - z_i|\} \qquad \text{Eq. 4}$$

We must consider that fuzzy algorithms c-means and k-means are available in most software, which contributes to the preference of their use over the others. Despite this, validations are still necessary to determine the clustering algorithm considering agricultural data. Gavioli et al. (2019) evaluated twenty different clustering algorithms, including FCM and k-means, to delineate MZs with data of three commercial agricultural fields cultivated with soybean and corn. They used elevation, clay, sand, silt, soil penetration resistance, slope, and bulk density. McQuitty's Method and Fanny obtained the best results in the three areas, but the results were equivalent to FCM and k-means in two.

Other algorithms, such as RASCH, kriging and derivatives, and linear programming, are also being researched. Still, present works are low enough to allow a direct and more in-depth comparison in multiple situations.

Guastaferro et al. (2010) evaluated ISODATA, FCM algorithm, and a density-based non-parametric clustering method to delineate MZs in wheat. They considered that, although ISODATA presents a lower computational cost and a better visual distinction of MZs, the lack of information on the transition areas was a problem.

Gili et al. (2017) stated that choosing the ideal algorithm of MZ delineation depends on the objectives to use MZs. In their research with corn, they used MULTISPATI-PCA to produce synthetic variables (PCs) and three clustering strategies: (1) S1- the first PC and the Jenk's natural rupture method, (2) S2- FCM using directly on soil variables (Clay + silt, OM, pH, ECa, and organic matter index) data, and (3) S3- FCM using the first three PCs. The different strategies resulted in a different number of zones with different characteristics: for fertilization management zonification, it might be prioritized OM differentiation, available P contents and the S3 use; if water were the main limiting factor, there should be two management zones according to S1 or S2, responding to textural and altimetry differences.

Boluwade et al. (2016) analyzed irrigation on MZs delineation, employing ECa and elevation, with FCM, regionalization and partitioning clustering algorithms. Their results indicate that the use of both algorithms presents very similar results.

It is also possible to combine algorithms in sequence. Galambošová et al. (2014), clustering yield and electromagnetic data of a 17 hectares, used Ward's method (to obtain information on clusters as the ideal number of clusters) followed by k-means clustering method. The delineated MZs showed more quality and information on clusters than if both algorithms had been applied separately.

The adaptation of traditional algorithms to consider new spatial constraints can also be performed. An example is the adaptation of the Hierarchical Agglomerative

Constrained Clustering algorithm (HACC) to analyze spatial data (HACC-SPATIAL), which has been tested on wheat data, and demonstrated its viability (Ruß, 2013).

Another possibility is the modification of algorithms that were not initially developed for MZs delineation. Zhang et al. (2016) introduced improvements to a method that uses Binary Integer Linear Programming (BILP) and semivariograms, aiming to delineate rectangular MZs, due to it is easy its handling in a field. Their results, based on rice data, demonstrate the effectiveness of using this method. Cid-Garcia et al. (2013) used the computational technique of the Integer Linear Programming Management Zone to delineate MZs in a rectangular format. Albornoz et al. (2019) extended the approach Cid-Garcia et al. (2013), adding temporal variability restrictions, which improved the process.

Definition of the ideal number of classes of management zones (Question 3)

Most clustering techniques allow the user to choose the number of classes (MCs), making it possible to test several subdivisions in the area. Thus, it must be defined a way to select the most appropriate MCs, usually the one that presents the most significant reduction in the overall variance of the target variable (typically yield) (Frogbrook and Oliver, 2007; Nawar et al., 2017). Zhang et al. (2010) proposed a two criteria method to decide the optimal number of zones: (1) overall reduction of variance is >50%, and (2) progressive reduction of variance is <20%. More advanced analysis regarding clustering process performance can be assessed using indices and analysis of variance (ANOVA).

According to Souza et al. (2018), it is logical to divide the entire field into MZs with a statistically distinct target variable. They proposed that after confirming that there is no spatial dependence within each class, an ANOVA is conducted in the average values of the target variable (usually, the yield), using Tukey test. Secondly, it is calculated indices of performance.

Table 6 presents several measures (showing only the ones used in three or more studies) used in this task, but, in most cases, they are related and restricted in conjunction with the algorithm used in the delineation. Considering that this is the most frequently used MZ delineation algorithm in clustering applications, it would be expected that they are among the most used measures.

Fuzziness Performance Index (FPI) measures the degree of separation among the fuzzy partitions of **X**. Their values range from 0 to 1. Values close to zero indicate distinct classes with only a small value of the shared pertinence function, and values close to 1 indicate that there is no distinction among the classes, presenting a high value of shared pertinence function (Equation 5):

$$FPI = 1 - \frac{c}{c-1}\left[1 - \frac{\sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik})^2}{n}\right] \quad \text{Eq. 5}$$

where c is the number of clusters; n is the number of elements in the data set; and u_{ik} is the element of fuzzy pertinence matrix.

FPI, NCE, and MPE measures (Bezdek, 1981; McBratney and Moore, 1985; Odeh et al., 1992) are strongly connected to FCM algorithm. The Normalized Classification Entropy (NCE) aims to model the amount of disorganization of a fuzzy partition c, and can be defined by (Equation 6):

$$NCE = \frac{n}{n-c}\left[-\frac{\sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}\log_a(u_{ik})}{n}\right] \quad \text{Eq. 6}$$

where $\log_a$ is the logarithmic base, a is any positive integer.

Table 6 The most used measures to choose the number of management zones (MZs)

Measure	N° of papers	Papers
Fuzziness Performance Index (FPI)	62	(McBratney and Moore, 1985; Boydell and McBratney, 2002; Kitchen et al., 2002; Fridgen et al., 2004; Li et al., 2007; Kyaw et al., 2008; Mishra et al., 2008; Molin and de Castro, 2008; Vitharana et al., 2008a; Li et al., 2008; Morari et al., 2009; Song et al., 2009; Xin-Zhong et al., 2009; Guastaferro et al., 2010; Moral et al., 2010; Arno and Martinez-Casasnovas, 2011; Jiang et al., 2011; Bansod et al., 2012; Davatgar et al., 2012; Jiang et al., 2012; Roberts et al., 2012; Valente et al., 2012; Alves et al., 2013; Bazzi et al., 2013; Córdoba et al., 2013; Li et al., 2013; Lin et al., 2013; Meirvenne et al., 2013; Scudiero et al., 2013; Tagarakis et al., 2013; Chang et al., 2014; Patil et al., 2014; Urretavizcaya et al., 2014; Yao et al., 2014; Caires et al., 2015; Rodrigues and Corá, 2015; Tripathi et al., 2015; Peralta et al., 2015; Boluwade et al., 2016; Gavioli et al., 2016; Ortuani et al., 2016; Schenatto et al., 2016a; Schenatto et al., 2016b; Sobjak et al., 2016; Bottega et al., 2017; Gili et al., 2017; Schenatto et al., 2017b; Servadio et al., 2017; Yari et al., 2017; Shukla et al., 2017; Albornoz et al., 2018; Behera et al., 2018; Betzek et al., 2018; Miao et al., 2018; Schwalbert et al., 2018; Martínez-Casasnovas et al., 2018; Khan et al., 2018; Verma et al., 2018; Bazzi et al., 2019; Betzek et al., 2019; González-Fernández et al., 2019; Ortuani et al., 2019)
Analysis of Variance (ANOVA)	46	(Fleming et al., 2000; Jaynes et al., 2005; Ikenaga and Inamura, 2008; Inman et al., 2008; Molin and de Castro, 2008; Vitharana et al., 2008b; Xin-Zhong et al., 2009; Zhang et al., 2010; Aimrun et al., 2011; Arno and Martinez-Casasnovas, 2011; Jiang et al., 2011; Davatgar et al., 2012; Jiang et al., 2012; McClymont et al., 2012; Bazzi et al., 2013; Córdoba et al., 2013; Li et al., 2013; Lin et al., 2013; Peralta et al., 2013; Santesteban et al., 2013; Scudiero et al., 2013; Chang et al., 2014; Urretavizcaya et al., 2014; Yao et al., 2014; Bazzi et al., 2015; Santos and Saraiva, 2015; Tripathi et al., 2015; Peralta et al., 2015; Damian et al., 2016; Gavioli et al., 2016; Oldoni and Bassoi, 2016; Ortuani et al., 2016; Schenatto et al., 2016b; Schenatto et al., 2016a; Sobjak et al., 2016; González-Fernández et al., 2017; Schenatto et al., 2017b; Shukla et al., 2017; Betzek et al., 2018; Martínez-Casasnovas et al., 2018; Khan et al., 2018; Verma et al., 2018; Betzek et al., 2019; Gavioli et al., 2019; Bazzi et al., 2019; Reyes et al., 2019)
Normalized Classification Entropy (NCE)	45	(Kitchen et al., 2002; Fridgen et al., 2004; Li et al., 2007; Kyaw et al., 2008; Mishra et al., 2008; Vitharana et al., 2008a; Li et al., 2008; Morari et al., 2009; Xin-Zhong et al., 2009; Guastaferro et al., 2010; Moral et al., 2010; Arno and Martinez-Casasnovas, 2011; Jiang et al., 2011;

		Bansod et al., 2012; Davatgar et al., 2012; Jiang et al., 2012; Roberts et al., 2012; Alves et al., 2013; Córdoba et al., 2013; Li et al., 2013; Lin et al., 2013; Scudiero et al., 2013; Tagarakis et al., 2013; Chang et al., 2014; Patil et al., 2014; Caires et al., 2015; Rodrigues and Corá, 2015; Santos and Saraiva, 2015; Tripathi et al., 2015; Peralta et al., 2015; Boluwade et al., 2016; Ortuani et al., 2016; Gili et al., 2017; Servadio et al., 2017; Yari et al., 2017; Shukla et al., 2017; Albornoz et al., 2018; Behera et al., 2018; Miao et al., 2018; Schwalbert et al., 2018; Martínez-Casasnovas et al., 2018; Khan et al., 2018; Verma et al., 2018; González-Fernández et al., 2019; Ortuani et al., 2019)
Modified Partition Entropy (MPE)	17	(Boydell and McBratney, 2002; Molin and de Castro ,2008; Song et al., 2009; Valente et al., 2012; Meirvenne et al., 2013; Urretavizcaya et al., 2014; Yao et al., 2014; Gavioli et al., 2016; Oldoni and Bassoi, 2016; Schenatto et al., 2016a; Schenatto et al., 2016b; Sobjak et al., 2016; Bottega et al., 2017; Schenatto et al., 2017b; Betzek et al., 2018; Betzek et al., 2019; Bazzi et al., 2019)
Variance Reduction (VR)	9	(Gavioli et al., 2016; Schenatto et al., 2016b; Schenatto et al., 2016a; Sobjak et al., 2016; Schenatto et al., 2017b; Betzek et al., 2018; Betzek et al., 2019; Gavioli et al., 2019; Bazzi et al., 2019)
Smoothness Index (SI)	6	(Gavioli et al., 2016; Schenatto et al., 2016a; Schenatto et al., 2017b; Betzek et al., 2018; Betzek et al., 2019; Bazzi et al., 2019)
Relative Variance (RV)	4	(Xiang et al., 2007; Dobermann et al., 2003, Ping and Dobermann, 2003; Miao, Mulla, and Robert, 2018)
Improved Cluster Validation Index (ICVI)	4	(Arango et al., 2017; Gavioli et al., 2019; Schenatto et al., 2016a; Betzek et al., 2019)
Average silhouette coefficient (ASC)	3	(Rousseeuw, 1987; Gavioli et al. 2019; Reyes et al., 2019)

The Modified Partition Entropy (MPE) estimates the difficulty level in organizing *c* clusters, with values close to 0. This indicates some low difficulty in organizing the clusters. It can be defined by (Equation 7):

$$MPE = \frac{-\sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}\log(u_{ik})/n}{\log c} \qquad \text{Eq. 7}$$

Combinations of more than one measure of FPI with NCE and FPI with MPE are common (Table 7). In both cases, it must be sought the number of clusters in which values of both measures are the lowest ones. Unfortunately, there may be times when these measures do not agree. In these situations, the choice of value may be subjective or require help of other measures, such as ICVI.

Table 7 Combination of the most used measures to choose the number of management zones (MZs)

Measures	N° of papers	Papers
Fuzziness Performance Index (FPI) e *Modified Partition Entropy* (MPE)	39	(Fridgen et al., 2004; Li et al., 2007; Kyaw et al., 2008; Vitharana et al., 2008a; Li et al., 2008; Morari et al., 2009; Xin-Zhong et al., 2009; Guastaferro et al., 2010; Moral et al., 2010; Arno and Martinez-Casasnovas, 2011; Jiang et al., 2011; Bansod et al., 2012; Davatgar et al., 2012; Jiang et al., 2012; Roberts et al., 2012; Alves et al., 2013; Córdoba et al., 2013; Li et al., 2013; Lin et al., 2013; Scudiero et al., 2013; Chang et al., 2014; Patil et al., 2014; Caires et al., 2015; Rodrigues and Corá, 2015; Santos and Saraiva, 2015; Tripathi et al., 2015; Peralta et al., 2015; Boluwade et al., 2016; Ortuani et al., 2016; Gili et al., 2017; Servadio et al., 2017; Yari et al., 2017; Shukla et al., 2017; Behera et al., 2018; Schwalbert et al., 2018; Martínez-Casasnovas et al., 2018; Khan et al., 2018; Verma et al., 2018; González-Fernández et al., 2019)
Fuzziness Performance Index (FPI) e *Normalized Classification Entropy* (NCE)	17	(Kitchen et al., 2002; Molin and de Castro, 2008; Song et al., 2009; Valente et al., 2012; Meirvenne et al., 2013; Urretavizcaya et al., 2014; Yao et al., 2014; Gavioli et al., 2016; Oldoni and Bassoi, 2016; Schenatto et al., 2016a; Schenatto et al., 2016b; Sobjak et al., 2016; Bottega et al., 2017; Schenatto et al., 2017b; Betzek et al., 2018; Betzek et al., 2019; Bazzi et al., 2019).

The Variance Reduction (VR) (Gavioli et al., 2016) is a relative variance (RV) change proposed by Webster and Oliver (1990) (Xiang et al., 2007, Dobermann et al., 2003, Ping and Dobermann, 2003). It is calculated for the target variable, with the expectation that the sum of data variances from delineated MZs is smaller than the total variance (Equation 8):

$$VR = \left(1 - \frac{\sum_{i=1}^{c} W_i * V_{mzi}}{V_{field}}\right) * 100 \qquad \text{Eq. 8}$$

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

The Improved Cluster Validation Index (ICVI) (Gavioli et al., 2016) was proposed to solve the possible problem of non-agreement of FPI, MPE, and VR measures in MZs delineation. The higher VR value and the lower FPI and MPE values are, the closer ICVI will be to 0; the lower ICVI, the better the result of the clustering method. It can be determined as follows (Equation 9):

$$ICVI_i = \frac{1}{3} * \left(\frac{FPI_i}{Max\{FPI\}} + \frac{MPE_i}{Max\{MPE\}} + \left(1 - \frac{VR_i}{Max\{VR\}}\right)\right) \quad \text{Eq. 9}$$

where FPI_i is FPI value of i-th (variable selection) method; MPE_i is the MPE value of i-th method; VR_i is the VR value of i-th method; and $Max\{index\_X\}$ represents the maximum value of *index_X* among *n* variable selection methods.

Smoothness Index (SI) (Gavioli et al., 2016) gives pixel-by-pixel frequency of classes change in a thematic map in horizontal and vertical directions and along the diagonal. For maps with more uniform classes, SI tends to 100, while maps with many exchanges among classes tend to lower values. It can be calculated by (Equation 10):

$$SI = 100 - \left(\frac{\sum_{i=1}^{k} C_{H_i}}{4P_H} + \frac{\sum_{j=1}^{k} C_{V_j}}{4P_V} + \frac{\sum_{l=1}^{k} C_{DR_l}}{4P_{DR}} + \frac{\sum_{m=1}^{k} C_{DL_m}}{4P_{DL}}\right) * \ 100 \quad \text{Eq. 10}$$

where C_{H_i} is the number of changes on row i (horizontal); C_{V_j} is the number of changes in column j (vertical); C_{RD_l} is the number of changes on diagonal l (diagonal right - DR); C_{LD_m} is the number of lines on diagonal m (diagonal left - DL); 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_{DR} is the possibility of changing in the right diagonal; and P_{DL} is the possibility of changes in the left diagonal.

Average silhouette coefficient (ASC) is obtained from the silhouette coefficient (SC) (Rousseeuw, 1987), which is an evaluation index that measures both levels of satisfactory internal formation and external separation of groups. The SC value for point p, denoted by scp, is calculated using the average of intra-group distances ap and the average of inter-group distances bp (Equation 11).

$$sc_p = \frac{b_p - a_p}{Max(a_p, b_p)} \quad \text{Eq. 11}$$

where a_p is the average of distances among point p and all other points in the same group, and b_p is the average of distances among point p and all points in the closest group containing p.

The group silhouette coefficient (GSC) is obtained by calculating the average of silhouette coefficients for the points of this group, and the value that corresponds to ASC coefficient of k groups is obtained by calculating the average of GSC values of k groups. ASC values vary from -1 to 1; -1 indicates an incorrect grouping, and 1 indicates groups with the best intra-group formation and the best possible inter-group separation.

Based on the Analysis of Variance (ANOVA), the target variable (usually yield) is compared among classes by using the average target variable and performing the Tukey range test to identify whether the generated classes showed significant differences (first, we confirmed that there was no spatial dependence within each class).

Despite plurality of measures, the most usual measures, used together or not, are mainly related to the clustering methods based on FCM. A few, such as VR, ASC, SI, and Tukey test (ANOVA) can be used regardless of the algorithm used for delineation. It is expected that, with the increase in research related to other algorithms for MZs delineation , there will also be an increase in the number of measures used to define the ideal number of MZs.

Although they have not appeared consistently in SLM, some of the measures often used that deserve to be highlighted are:

Fragmentation index (FI%): it takes into account how higher is the number of zones (NMZ) in comparison with the number of classes (NC). If each class corresponds to a single zone, then the estimated fragmentation by FI% will be zero. If, for example, for a four-class design, five zones are created, then FI% will be 25%. The higher the fragmentation of delineation is, the higher FI% is (Equation 12):

$$FI\% = 100\frac{NMZ - NC}{NC} \quad \text{Eq. 12}$$

Global Quality Index (GQI): it looks for finding the best number of classes during ZMs' delineation, taking into account the values of ICVI, SIr and FIr (Equation 13):

$$GQI_i = \frac{ICVI_i * (100 + FIr_i)}{SIr_i} \qquad \text{Eq. 13}$$

Kappa coefficient (K) (Cohen, 1960): this index is not used to validate the clustering process but to compare the agreement of two MZ delineation approach. Landis and Koch (1977) proposed the following classification: 0 < K ≤ 0.2 indicates no agreement, 0.2 < K ≤ 0.4 weak agreement, 0.4 < K ≤ 0.6 moderate agreement, 0.6 < K ≤ 0.8 strong agreement, and 0.8 < K ≤ 1 very strong agreement.

Coefficient of relative deviation (CRD) (Coelho et al., 2009): it calculates the mean difference in modulus of the interpolated values on a thematic map when compared to a map taken as a reference (Equation 14):

$$CRD = \sum_{i=1}^{n} ABS(\frac{Zi_B - Zi_A}{Zi_A}) \qquad \text{Eq. 14}$$

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

Mean absolute difference (MAD, Equation 15): it computes the mean absolute difference among values on the two maps.

$$MAD = \frac{\sum_{i=1}^{n} ABS(Zi_B - Zi_A)}{n} \qquad \text{Eq. 15}$$

Possible economic or environmental advantages of management zones adoption (Question 4)

Despite the complexity involved in the procedure, delineating MZs in itself is not an end goal. Instead, its premise is to serve as a subsidy for decision-making on how to allocate better resources in the field, aiming at a more rational use with less environmental impact and higher profitability. Despite this, most studies only present the ideal number of MZs and the MZs map as the final product, often omitting if the zones are significant and the possible economic or environmental advantages of their adoption. This remark was also made by Nawar et al. (2017).

It is important to perform a statistical analysis of MZs delineated to validate the zones division. One way to do this is with ANOVA where the target variable (usually yield) is compared among classes by using the average target variable and performing the Tukey range test to identify whether the generated classes showed significant differences (first, we confirmed that there was no spatial dependence within each class). Observing Table 6, it can be observed that only 46 (approximately 28%) of the selected papers did a consistent statistical analysis (ANOVA) to validate the existence of considerable differences among the resulting zones to justify this division. An even smaller number (8) of studies analyzes the economic impact of adopting the use of MZs.

Kyaw et al. (2008) worked with five areas with chlorosis-prone soybeans and corn to delineate MZs for its control, concluding that the control of chlorosis using MZs did not increase yield but reduced Fe application considerably. In one case, the application was reduced to just 43% of the total area, in another, to 41%, lowering the average cost per hectare. Robertson et al. (2008) conducted a study on wheat with 199 properties, ranging from 10 to 172 ha, and found out an economic benefit between US$ 5.00/ha and US$ 40.00/ha when they adopted MZs. This benefit represents a significant differential for producers in Western Australia since the region has a margin of around US$ 100.00 ha^{-1}.

Velandia et al. (2008) analyzed the economic impact of four approaches of N application in cotton: (1) uniform N rate application based on an agronomic optimum (URA), (2) uniform N rate application based on an economic optimum (URE), (3) variable-rate N application based on the economic optimum for each of the management zones established by our spatial procedure above (VRN, developed at this work), and (4) variable-rate N application based on landscape position (VRL). Their results demonstrated that VRN application could result in net returns over US5.28ha^{-1}$, US$ 6.17 ha^{-1}, and US$ 7.28 ha^{-1}, when compared to VRL, URE, and URA, respectively.

In a study involving six producers, whose fields were cultivated with corn, Roberts et al. (2012) developed MZs to control Nitrogen. Two areas showed no correlation between yield and N, while in the other four, they found out that the variable-rate N application according to soil-based MZ showed a gain of –US$ 33 ha^{-1}, US$ 145 ha^{-1}, US$ 0, and US$ 32 ha^{-1}. Hörbe et al. (2013) assessed the efficiency of variable-rate seeding of corn with delineated MZs, split into low, medium, and high crop performance zones. They reduced the recommended plant population by 31 % in the low management zone resulted in a yield increase of 1.5 Mg ha^{-1} and induced an increase of US$ 342 ha^{-1} in partial net economic return. Increasing the recommended plant population by 13% in the high management zone resulted in an increase of 0.91 Mg ha^{-1} in grain yield and induced an increase of US$ 113 ha^{-1} in partial net economic return. Also, working with corn, Bernardi et al. (2018) registered three as the ideal number of classes. The class with the highest profitability recorded 12% profit higher than the class with the lowest profitability.

Whetton et al. (2018) evaluated the economic viability and environmental benefit of adopting a variable-rate fungicide application (VRSA) and selective harvest (SH) in winter wheat. Results showed that in this study, VRFA allowed for fungicide reductions from 22 to 26% when compared to homogeneous rate fungicide application (HRFA). The net saving after considering sensing costs was £67 ha^{-1}, which is roughly equivalent to €80 or \$90 ha^{-1} per year. This study was restricted to a single field but demonstrates the potential of fungicide reductions and economic viability of MZ concept.

Schwalbert et al. (2018) compared four different wheat fertilization strategies from two producers' fields: (1) traditional N fertilization management (constant rate, CR), (2) variable-rate N application based on crop remote sensing (CS), (3) VNR based on MZs (MZs), (4) integrated approach combining MZs and crop sensing (MZ+CS). They concluded that the integrated version (MZ+CS) presented an average economic return of US$ 42 ha-1 (field 1) and US$ 32 ha-1 (field 2) higher than the CR. However, when considering only the highest yield MZ, the values change to US$ 80 ha-1 and US$ 40.00 ha-1 for fields 1 and 2, respectively. Despite the small number of studies validating the economic return of using MZs, the advantages of their adoption in all cases were verified.

Software used to delineate management zones (Question 5)

Three main questions must be addressed for an efficient delineation of MZs: (1) what data set should be used?; (2) what algorithm should be used to delineate the MZs?; and (3) what is the optimal number of MZ classes? (Fridgen et al., 2004). Although they seem to be simple questions, each unfolds in virtually dozens of options, with specific advantages and disadvantages. For a correct understanding and analysis, they often require the knowledge of several areas, and create great difficulty to adopt MZs in agriculture.

Some of these difficulties can be reduced by using specialized software. Despite many software for PA, few are directed to delineate MZs (Table 8). Golden Software Surfer, ESRI ArcGIS, and R software package are commonly used. Despite allowing MZs delineation, they do not have all the desired functionality since this is not the focus of these products. So, it is required to go to other computer programs to perform the entire process. Furthermore, when they have all the necessary functionalities, they are not user-friendly. Another determining factor to hinder access to software is because most of them present only paid commercial licenses, and this discourage their adoption by non-specialized people since they may not realize the advantages of their use at first.

Among the specific software to delineate MZs, the following are well-known (organized by release date): (1) FuzME (Minasny and McBratney, 2002), Management Zone Analyst (MZA) (Fridgen et al., 2004), (2) Software for Management Zones Definition (SDUM) (Bazzi et al., 2013; Bazzi et al. 2019), (3) ZoneMAP (Zhang et al., 2010), and (4) automatic software to delineate MZs proposed by Albornoz et al. (Albornoz et al, 2018).

Table 8 Software used for management zones delineation by the papers included in this review

Names	N*	main functions used	License	OS*	Developer	Site
ArcGis/ ArcMap	41	Maps, classification	Commercial (paid)	Windows Web	ESRI	https://www.arcgis.com/
MZA	31	Delineation of MZ	free	Windows	Cropping Systems and Water Quality Research	https://www.ars.usda.gov/research/software/download/?softwareid=24&modecode=50-70-10-00
SAS	20	Statistical analysis, classification	Commercial (paid)	Windows Linux z/OS	SAS	https://www.sas.com
SPSS	18	Statistical analysis	Commercial (paid)	Windows Linux Mac	IBM	https://www.ibm.com/spss
R	14	Statistical analysis, classification Selection of variables	free (Open Source)	Windows Linux Mac	r-Project (Open Source)	https://www.r-project. org/
FuzMe	13	Delineation of MZ	free	Windows	Precision Agriculture Laboratory, University	https://sydney.edu.au/ agriculture/pal/

					of Sydney	
GS+	9	Geostatistical analysis, interpolation	Commercial (paid)	Windows	Gamma-design	https://geostatistics.com
ISATIS	8	Geostatistical analysis	Commercial (paid)	Windows Linux	Geovarian-ces	https://www.geovariances.com
Surfer	6	Maps	Commercial (paid)	Windows	Golden Software	https://www.goldensoftware.com
MatLab	5	Mathematical analysis, modeling	Commercial (paid)	Windows Linux Mac	MathWorks	https://www.mathworks.com/
Statistica	5	Statistical analysis	Commercial (paid)	Windows	StatSoft	http://www.statsoft.com
Vesper	5	Interpolation	Share-ware	Windows	Precision Agriculture Laboratory, University of Sydney	https://sydney.edu.au/ agriculture/pal/
SDUM	5	Statistical analysis, Statistical and geostatistical analysis, maps	free	Windows	Grupo Agricultura de Precisão da Região Oeste do Paraná	http://ppat.md.utfpr.edu.br/
ERDAS Imagine	4	Maps, image Analysis	Commercial (paid)	Windows	Hexagon Geospatial	https://www.hexagongeospatial.com
Unscrambler	3	Statistical analysis, modeling	Commercial (paid)	Windows	Camo Analytics	https://www.camo.com
Krig-ME	3	Delineation of MZ	***	3		
Not specified	7					
Others	38					

* N: number of papers using software. ** OS: operating system. *** Download not available to collect information. Only software that has been used in at least 3 papers is mentioned. A paper can use more than one software.

FuzME is a software provided by the Precision Agriculture Laboratory (PA Lab) of the Australian Centre for Precision Agriculture (ACPA), from University of Sydney, in Australia. It is available for Microsoft Windows 95/NT or superior, and its most current version is 3.5c. The used algorithm is FCM (with a few variants), and outputs are all in text files. The software features are a simplified graphical interface, and consist essentially of three toolbars. The first toolbar presents the options to select the input files with the respective variables, output files, internal control files, and analysis title. The second one presents the options to create clusters, such as distance metrics and fuzzy exponents, among others. The third one presents the options to allow resampling by using the bootstrap and Jackknife methods. Among the possible options to adjust the clustering algorithm, there are: (1) choice of distance metric (Euclidean, diagonal, and Mahalanobis); (2) choice of fuzzy exponent; (3) definition of minimum and maximum number of classes (from 1 to 100); (4) fuzzy discriminants analysis; (5) configuration of the initial random values of the members' definition and the number of attempts, stopping criterion, maximum number of iterations, and (6) choice of algorithm (classic FCM, extra-grade FCM, equal-area FCM, and FCM with covariance matrix).

Although the simplified interface is a positive point for its use, as well as the definition of some standardized parameters, it is impossible to: (1) visualize delineated MZs, (2) perform interpolations, (3) adjust sample size, (4) visualize the behavior of input variables, (5) calculate statistics of MZ quality; and (6) export the results. Another limiting factor is the need to run on computers using a specific operating system (PC Windows environment), considering the dissemination of ubiquitous computing nowadays.

MZA is the most used one among the specific software for MZs delineation (Table 8). It is available by the Agricultural Research Service (ARS) of the United States Department of Agriculture (USDA), USA. It is available for Microsoft Windows 95/NT or superior, and its most current version is 1.0.1. MZA and FuzME also used FCM algorithm.

It also presents a simplified graphical interface and, to perform classification, the instructions must be followed in a sequence of four menus that present the definition of parameters step by step. Initially (start window), you must provide the input file in CSV text format (comma-separated values), containing variables and their values. In the same window, one or more variables must be chosen to be used. The following window, Explore Data, allows descriptive data statistics to be computed and saved in a text file: the mean, standard deviation, coefficient of variation, minimum and maximum, sums of squares, and variance and covariance matrices. The third window, Delineate Zones, presents the options to perform the classification with FCM: fuzzy exponent; measure of similarity (Euclidean, diagonal, or Mahalanobis); maximum number of iterations; convergence criterion; minimum and maximum number of MZs; and location and name of output data file.

The last window, Post Classification Analysis, presents two graphs of performance indices (NCE and FPI) as a function of the number of zones. The authors consider this last window to be one of the most critical differentials of MZA because it helps to choose the ideal number of zones, and avoids subjectivity. It is worth remembering that the ideal number of these measures may still not be following the restrictions of field mechanization, considered purely mathematical analyses of the generated clusters. As in FuzME, the user-friendly interface and the definition of a precise sequence of steps to delineate MZs are positive points. Another coincident factor of both software is the lack of data processing tools, such as interpolation and data size adjustment for a common grid. It is also important that, depending on the characteristics of the input data, the resulting MZs can contain much-fragmented information, which will require the use of external software to smooth and visualize MZs. A third problematic element concerns one of its main advantages: choosing the ideal number of clusters. NCE and FPI measures cannot necessarily agree on the ideal number of clusters, thus, the analyst returns to subjectivity since the software does not indicate which is preferred over the other. A final limitation is the need to run on computers using a specific operating system (PC Windows environment).

SDUM (Bazzi et al., 2019) is software available by the Paraná Precision Agriculture Team, from the Western region of Paraná, Brazil. It is available for Microsoft Windows XP platform or superior, and the current version is 1.0. The execution outputs can be given in text, images, PDF, and KLM (Google Earth) formats. The software allows the insertion of one or more layers of georeferenced sample data. Entries are in the text file and have a user-friendly data importer. It also allows data interpolation by inverse distance weighting, moving average, and nearest neighbor.

Thematic maps can be generated with the interpolated data. To do this, we define the type of geometry that can be continuous surfaces or points; interpolator parameter; and radius parameter, consisting of the distance that samples will be selected for interpolation. There are also tools for descriptive statistical analysis and statistical analysis of spatial correlation.

MZs can be delineated by empirical methods (data normalization by means and standard deviation) and clustering (k-means and FCM). The number of classes and the number of iterations must be defined when using the k-means method. When using FCM, the number of classes, fuzzy exponent, and maximum error are defined.

SDUM also calculated performance indices (FPI, VR, and MPE), ANOVA, and Tukey test.

The user-friendly interface and the definition of a clear sequence of steps to delineate MZs, just like the previous software, are important positive points in SDUM. In addition, important features when compared to previous programs: data interpolation tools, spatial correlation analysis, more evaluation measures of MZs quality (FPI, VR, MPE, ANOVA, and Tukey test), generation of thematic maps, and MZs maps, organization as projects, and data storage in a database. Finally and most importantly, SDUM can present the delineated MZs, while FuzME and MZA must use a GIS or desktop mapping software. The presence of these elements in a simple interface increases considerably the user's independence regarding the use of complementary software and in knowledge domain from other areas.

As for disadvantages, we can highlight that, as in the previous software, depending on the characteristics of the input data, the resulting MZs can contain much-fragmented information, requiring the use of external software to smooth MZs. A final limitation is the need to run on computers using a specific operating system. ZoneMap was unavailable when this paper was under development, due to financial reasons, according to its developers. Thence, it could not be evaluated.

The automatic software to delineate MZs proposed by Albornoz et al. (2018) is a software available (in test version) by the Faculty of Engineering and Water Sciences (Facultad de Ingeniería y Ciencias Hídricas) of the National University of the Coast (Universidad Nacional del Litoral), Argentina. According to the authors, there are desktop and web versions of such software. The delineation algorithm is FCM, and outputs are in ESRI shapefile. The web version has a straightforward interface, following only a sequence of steps. The first step is to upload the file containing the variables for analysis. Vector data (such as yield or apparent soil conductivity) must be in text files (CSV, dat, or txt), and raster data in GeoTiff format.

All input variables are interpolated in the next step (screen) to the same user-defined grid by the Sibson (without using squares), Sibson (with squares), Farin, or Quadratic methods. This defined the map boundaries by the largest coincident area for all variables. The third screen defines the parameters of FCM algorithm: minimum and maximum number of zones, fuzzy exponent, and convergence value.

On the next screen, the MZs maps are presented for each number of zones. Also, on this screen, there is a table with the three evaluation measures of MZs quality (NCE, FPI, and Xie and Beni (XB)), as well as a graph of the Euclidean distance of these measures (EcD=$\sqrt{FPI^2 + NCE^2 + XB^2}$) as a function of the number of MZs. This distance was implemented to avoid the subjectivity of individual measures if they disagreed on a minimum value of MZs. This screen also presents the option for how many classes someone wishes to continue the process. On the next screen, there are the options to filter (rectification) the map: mask size (3x3, 5x5, and 7x7 pixels), the type of filter (medium or mode), minimum size of the area in m^2, and the number of running times of erosion and dilation.

The final screen presents the results of the original map image and the filtered map image and the option to download the resulting ESRI shapefile. The graphical interface of this software is highly minimalist and user-friendly, showed a precise sequence of steps to delineate MZs and, therefore, considered decisive positive points. There are two notable highlights of this software to FuzME and MZA: (1) the availability of data interpolation and conversion tools for a common size sample grid, in a fully automated form, (2) more comprehensive evaluation measures of MZs quality (NCE, FPI, XB, EcD). Another substantial differential is the possibility to smooth the maps generated using algorithms from digital image processing, which aimed at creating smoother transition edges and eliminating small MZs that, in practice, cannot be worked on in the field.

Another positive point to be highlighted is the existence of the desktop version and a web platform version. This gives independence to the user platform, working on virtually any operating system. Another element is the transference of the processing load, usually high in this kind of procedure, from the user's machine to a web server. The counterpart is the requirement of a stable connection to Internet, server's availability , and additional issues of data security/confidentiality .

As for disadvantages, we can highlight the lack of tools to conduct statistical and geostatistical analyses. However, interpolation tools, for example, are already advantageous when compared to FuzME and MZA. Table 9 compares the main features of the specific software for MZs delineation.

Table 8 Features of specific software for management zones (MZs) delineation

Software / Feature	FuzME	MZA	SDUM	Albornoz et al. (2018)
Multiplataform				x (Web)
Input data visualization / data description tools		x	x	
Pre-processing tools			x	x
Results export type	Text	Text	Text, image, PDF,	Shape file

			KLM	
MZ evaluation		x	x	x
Map Generation			x	x
Intuitive interface	x	x	x	x

Including new papers

After selecting papers according to the systematic literature and snowballing, four papers about a new web platform AgDataBox (ADB) were obtained (Bazzi et al., 2019; Michelon et al., 2019; Dall'Agnol et al., 2020; Borges et al., 2020). This platform aims at integrating data, software, procedures, and methodologies for Digital Agriculture. It is a joint project coordinated by the Western Paraná State University (Unioeste) and the Federal University of Technology - Paraná (UTFPR) with the cooperation of the Colorado State University (CSU), the United States Agricultural Research Service (USDA) in Columbia, the University of California Davis (UC Davis), the University of São Paulo (ESALQ/USP), and the Brazilian Agricultural Research Corporation (Embrapa). This platform is a continuation of the project for SDUM software (Bazzi et al., 2013). This web Platform has an Application Programming Interface (API), which consists of a set of resources accessible by the Hypertext Transfer Protocol (HTTP) to transfer both request and response messages expressed in JavaScript Object Notation (JSON) format. ADB-API, where data and processing routines are centered, enables interoperability of several applications. Four applications are under development: (1) ADB-Mobile, (2) ADB-Map, (3) ADB-Admin, and (4) ADB-IoT. The application ADB-Map works with spatial data aiming at creating thematic maps and management zones. Among ADB-Map functionalities, there are: (1) data importing/exporting, (2) data analysis and filtering, (3) data normalization, (4) data interpolation and generation of thematic maps, (5) delineation and evaluation of management zones, encompassing variable selection methods, empirical and data clustering methods, and evaluation statistics, (6) management zone rectification methods, (7) application map generation and exporting, and (8) optimal placement of proximal sensors for PA. Since this platform is not already available on web, it was impossible a detailed discussion of its performance.

3.3.2. Examples of management zones

Several examples of ZMs will be presented, associated to a brief discussion of data that originated them, to demonstrate several situations in which ZMs can be used.

3.3.2.1. Target values of the management zones

Yield (productivity) management zones

Usually, on MZs delineation, yield is used as target values. Kitchen et al. (2005) researched two Missouri claypan soil fields to answer the question of whether ECa and elevation could be used to delineate productivity zones (SPZ) that would agree with productivity zones delineated from yield map data (YPZ). Figure 21 presents the results for Field 2 that showed the best performing combinations of ECa and elevation variables, which gave a 60-70% agreement (overall accuracy) between YPZ and SPZ.

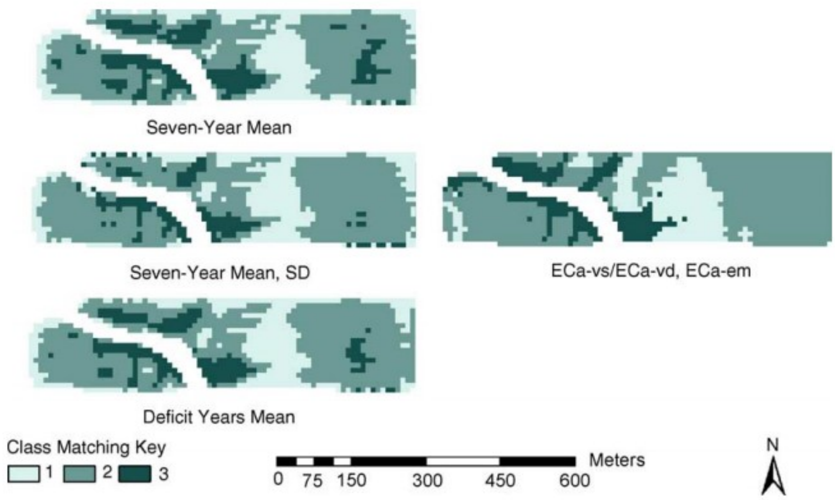
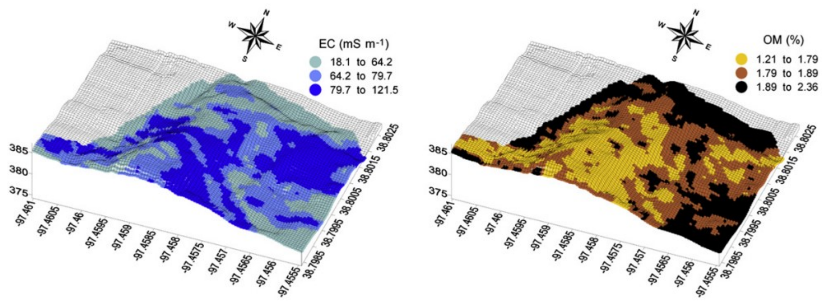


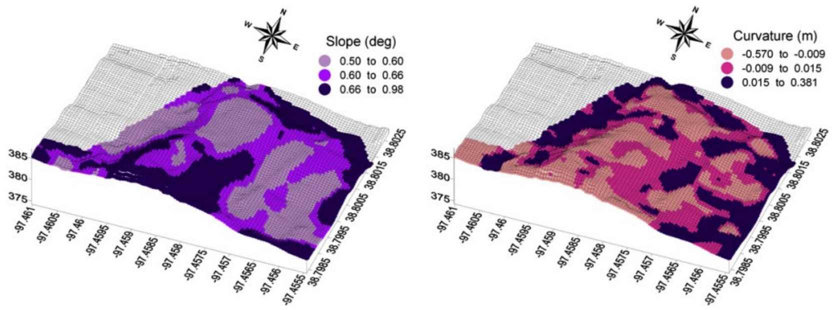
Figure 21. Reference yield zone maps (left) compared to the best performing productivity zone map derived from unsupervised clustering of ECa and elevation (right)
Source: Kitchen et al. (2005).

Kweon (2012) developed a delineation procedure for site-specific productivity zones with a fuzzy logic system using soil properties obtained from on-the-go electrical conductivity (ECa) and organic matter (OM) sensors and topographic attributes in two typical central Kansas upland fields (Field 1, 57 ha, and Field 2, 18 ha). EC, OM, slope, and curvature were used as input variables, and yield was set as an output variable. Using the quantile classification, the authors divided all thematic maps into three classes (low, medium, and high) (each class has the same number of data points). Figure 22 shows continuous EC and OM maps, and Figure 23, the maps of terrain slope and curvature. They constructed three types of MZs: 1) 5-year mean normalized yield map (Figure 24a); 2) Productivity map, generated by a producer's decision-making knowledge and the fuzzy logic system (Figure 24b); and 3) FCM map using EC, OM, slope, and curvature (Figure 24c). The spatial agreement among productivity and the 5-year-mean yield maps showed an overall accuracy and kappa coefficients of 0.57 and 0.35. The productivity map presented a better agreement with the normalized yield map than FCM map. All the presented figures are for Field 2.



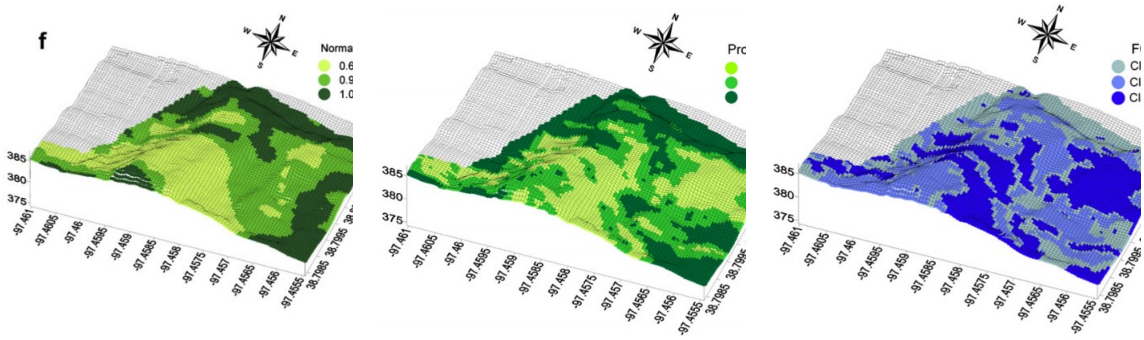
a) b)

Figure 22. EC and OM maps generated by the on-the-go sensor for Field 2
Source: Kweon (2012).



a) b)

Figure 23. Terrain slope and curvature maps for Field 2
Source: Kweon (2012).



a) normalized yield map

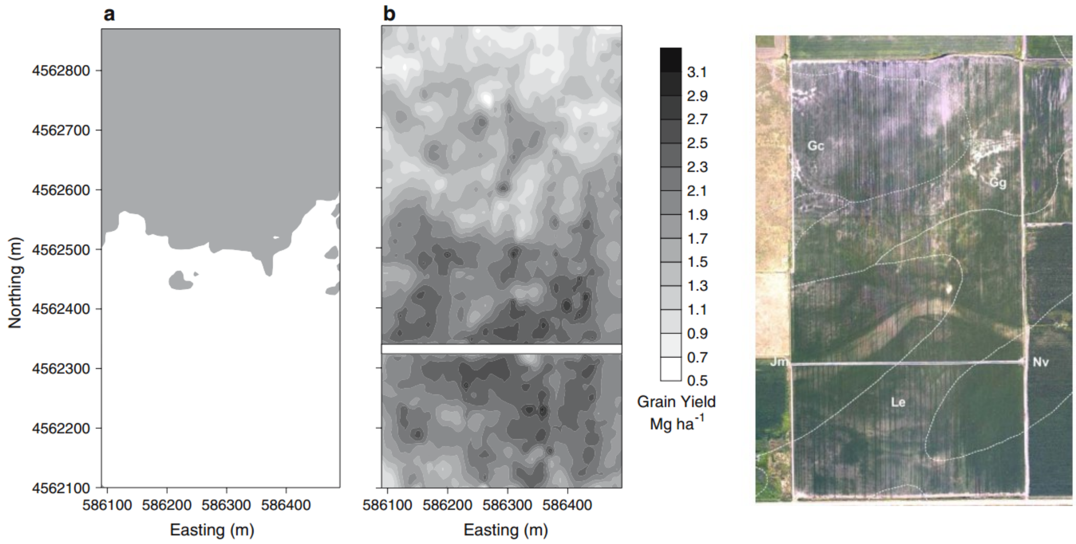
b) productivity MZs

c) FCM map.

Figure 24. Five-year mean normalized yield map (a). Productivity map generated by the developed fuzzy logic system (b). FCM map (c) (all figures are for Field 2)
Source: Kweon (2012).

3.3.2.2. Chlorosis management zones

Kyaw et al. (2008) evaluated delineating chlorosis MZs using VI derived from aerial imagery, on-the-go measurement of soil pH, and ECa. The study was conducted at six sites in 2004 and 2005, and generally, the yield was best predicted with the combination of NDVI and deep ECa. The delineation of chlorosis MZs from aerial imagery combined with soil ECa seems to be a useful tool for the site-specific management of iron chlorosis. Figure 25 illustrates the relationship of chlorosis zones to grain yield, and, in general, the northern part of this field can be considered chlorosis-prone. This area generally coincides with Gibbon loam (Gg) and Gayville-Caruso (Gc) soil series, fairly poorly drained, with salt accumulation in Gayville series occasionally causing dispersion of soil colloids (classified as Leptic Natrustolls).



(a) (b)

Figure 25. Chlorosis-prone area (a) (zone 1, gray shading) delineated from the combination of ECa and NDVI; soybean yield (b) (2005); and aerial photograph (2005), with soil series boundaries superimposed. All figures from site 1
Source: adapted from Kyaw et al. (2008).

3.3.2.3. Apparent electrical conductivity management zones

Yan et al. (2007a) studied a 10.5-ha site and measured ECa. Measurements were performed thrice *in situ* on topsoil (0-20 cm) across the field to identify MZs. The results indicated high coefficients of variation for topsoil salinity over all three samplings. However, the spatial structure of salinity variability remained relatively stable with time. Kriged choropleth maps, drawn based on spatial variance structure of data, showed the spatial trend of salinity distribution and revealed areas of consistently high or consistently low salinity (Figure 26); a temporal stability map indicated some stable and unstable regions (Figure 27). Cluster analysis divided the site into three potential MZs (Fig. **28**a) based on the spatiotemporal characteristics, each one with different characteristics that could impact the way the field was managed. Visually, the pattern of cotton yield appeared to correspond quite well with the trend of management classes (Fig. **28**b). Generally, the highest yields occurred in class 1, and the lowest ones in class 3.

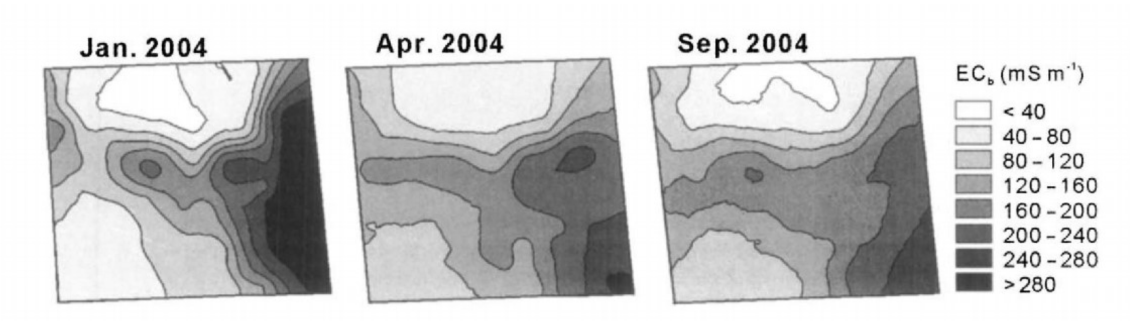
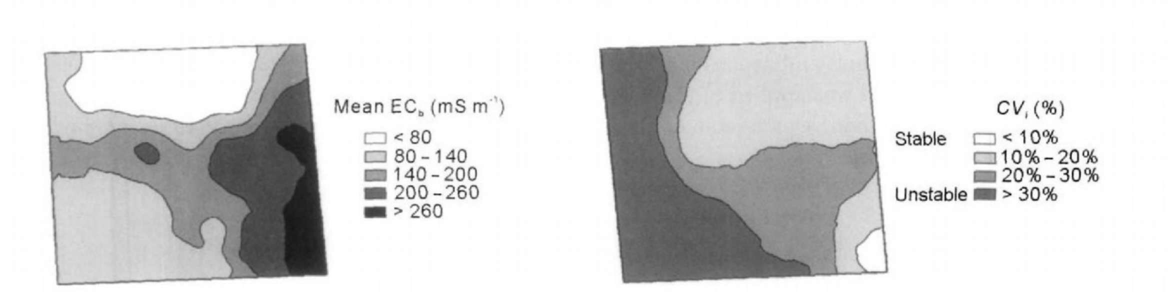
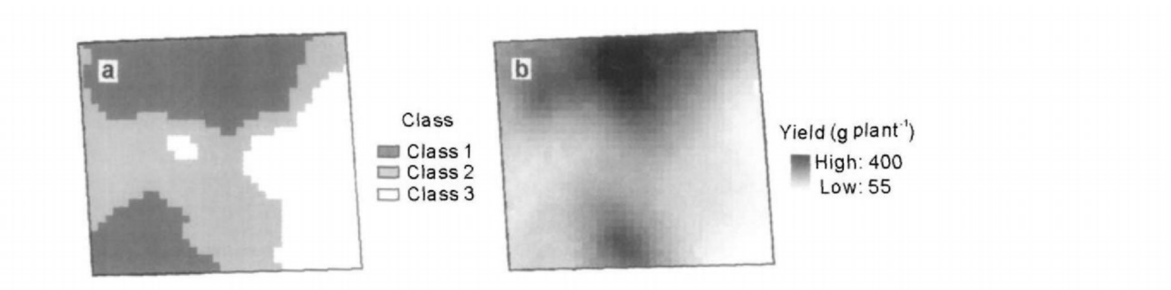


Figure 26. Smoothed choropleth maps produced by ordinary kriging for apparent electrical conductivity (ECa) at three different sampling dates
Source: Yan et al. (2007a).



(a) (b)

Figure 27. Spatial trend map composed of the mean apparent electrical conductivity (ECa) (a) and temporal stability map produced for ECa based on the CVi (coefficient of variation at the ith sampling point) (b)
Source: Yan et al. (2007a).



(a) (b)

Fig. 28. Spatial distribution of the three classes of practical management zones across the field using cluster analysis (a) and the spatial distribution of cotton yield interpolated by kriging (b).
Source: Yan et al. (2007a).

3.3.2.4. Soil available water content management zones

De Lara et al. (2018) studied the characterization of spatial distribution of soil water content (SWC) at a field scale by ECa. They found out that the delineated soil ECa MZs (Figure 29) can effectively characterize macro-scale in-field SWC variability among zones throughout the crop season. Furthermore, the inclusion of OM and salt content data significantly improved SWC assessment according to the ANOVA test.

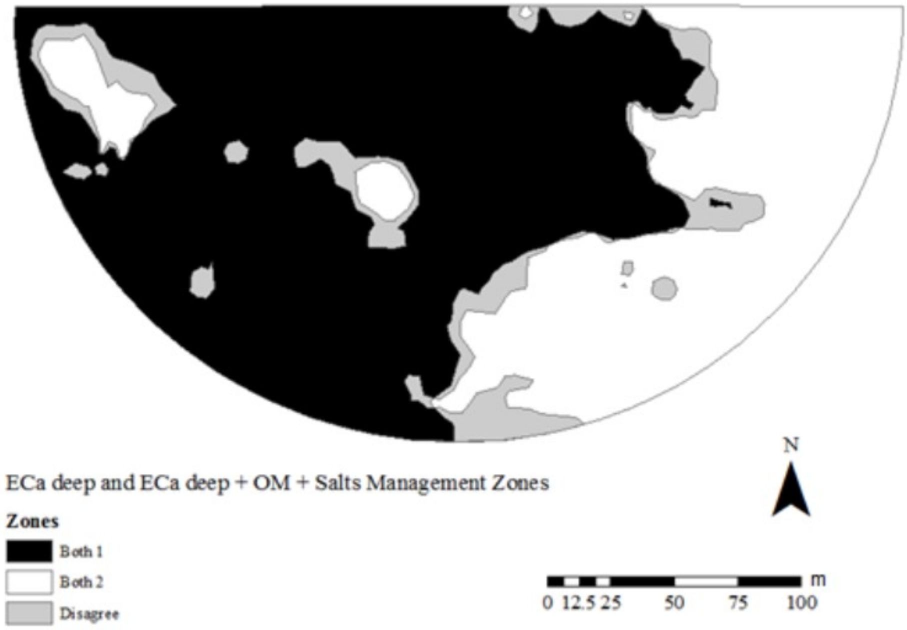


Figure 29. Comparison of management zones delineated with soil ECa measured up to 1.5-m depth and management zones delineated using soil ECa measured up to 1.5-m depth in addition to organic matter and soil salinity for ARDEC. Differences in both techniques are presented as gray, referred to as "disagree" in the legend.
Source: De Lara et al. (2018).

3.3.2.5. Quality-based management zones

Tagarakis et al. (2013) delineated MZs using fuzzy clustering techniques in a 1.0-ha commercial vineyard in Central Greece during 2009 and 2010. They used ECa, NDVI at different stages (NDVI 1, NDVI 2, NDVI 3, NDVI 4, and NDVI 5) during the vine growth cycle, yield, and grape quality index (sugar/acidity ratio of grape). Soil properties, yield, and grape composition parameters showed high spatial variability. Maps of two MZs were produced using MZA software.Figure 30 shows the yield-based MZs using soil depth, NDVI 1, NDVI 2, NDVI 3, and NDVI 4 (Figure 30a), and quality-based MZs using ECa, NDVI1, NDVI 2, NDVI3, and NDVI 4 (b). They concluded that these maps presented a high degree of agreement, from 79.2 to 89.6%.

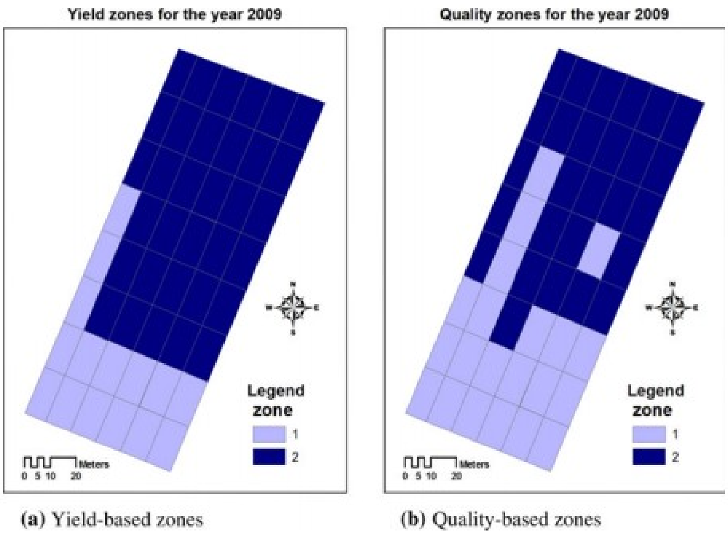


Figure 30. Yield-based management zones (soil depth, NDVI 1, NDVI 2, NDVI 3, NDVI 4) and quality-based management zones (ECa, NDVI 1, NDVI 2, NDVI 3, NDVI 4) using fuzzy clustering from a commercial vineyard in Central Greece. Data from the 2009 agricultural year
Source: Tagarakis et al. (2013).

3.3.2.6. Weed Management Zones

Regarding MZs for agrochemicals applications, the purpose is to use them immediately and just once.Figure 31 andFig. 32 present the delineated MZs using small, and large leaves weed plants, respectively (Rodrigues, 2009).

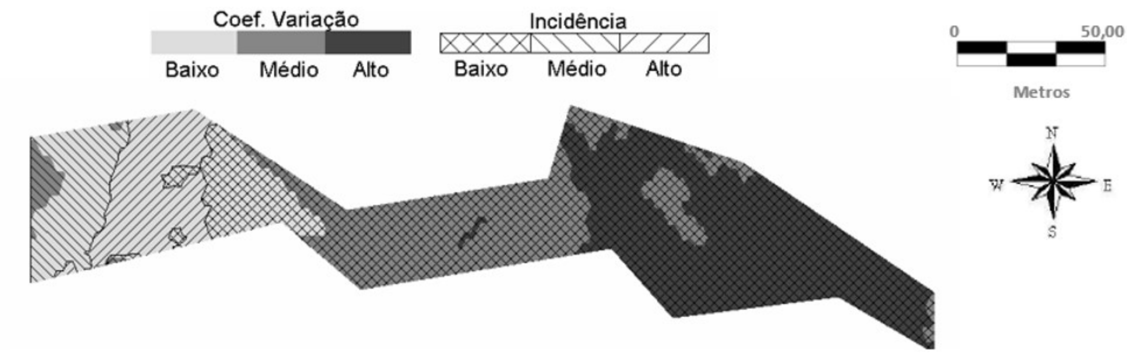


Figure 31. Management zones of small leaves weed plants in a 1.24-ha pear orchard
Source: Rodrigues, 2009.

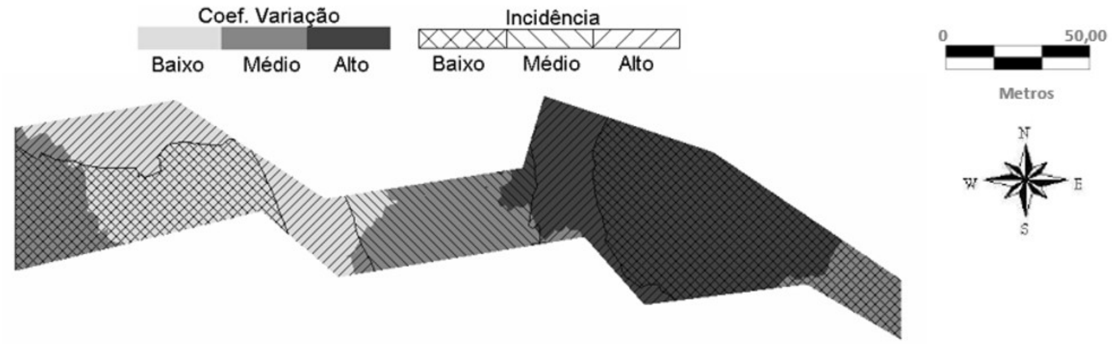


Fig. 32 Management zones of large leaves weed plants in a 1.24-ha pear orchard
Source: Rodrigues, 2009.

3.3.2.7. Vegetation Indices Management Zones

Regarding MZs for VI classification, like MZs for agrochemicals applications, the purpose of the research conducted by Costa et al. (2019) was to use them immediately and just once. Using geostatistics and multivariate analysis, they delimited homogeneous zones (HZs) of different VIs to identify vegetation patterns in Cabernet Franc and Cabernet Sauvignon vineyards. Using Crop Circle ACS-430 active sensor and simultaneously measuring crop spectral reflectance at 670 nm (ρR, red), 30 nm (ρRE, red edge), and 780 nm (ρNIR, near-infrared). Despite the variations of VIs spatial distribution patterns, multivariate analysis resulted in a representative categorization of grapevine vegetative vigor and HZs delimitation for this characteristic.

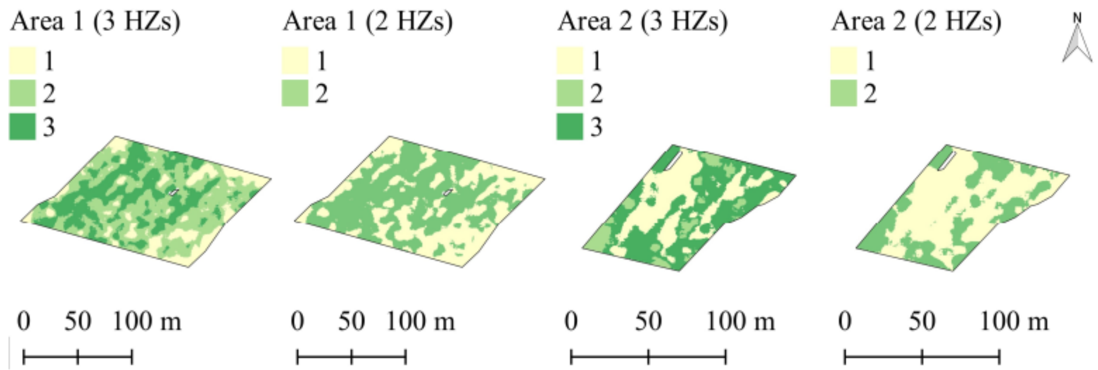


Figure 33. Homogenous zones resulting from clustering analysis of VIs, calculated based on ρ, for two studied areas. Reflectance was measured at canopy height using an ACS-430 active sensor. The studied areas were cropped with Cabernet Franc and Cabernet Sauvignon vines
Source: Costa et al. (2019).

3.3.2.8. Used variables to delineate Management Zones

Satellite imagery data

Zhang et al. (2010) developed a web-based decision support tool to automatically determine the optimal number of MZs and delineate them using satellite imagery and field data. In this tool, currently discontinued, application rates, such as fertilizer, could be prescribed for each zone and downloaded in several formats to ensure compatibility with GNSS-enabled farming equipment. Figure shows results from a 45.3-ha field in Potter County, South Dakota, where the rotation of crops from 2003 to 2005 was with corn, sunflowers, and spring wheat. The farmer delineated four subfield zones (Figurec) using a 2003-yield map (Figurea) and a 25-August-2004-Landsat NDVI map (Figureb) to determine urea application rates for the next year. As a result of this variable-rate application, the spring wheat planted in 2005 delivered a much more uniform yield (Figured). While the mean yields of each crop were about the same, 7.33 t ha^{-1} for corn and 7.17 t ha^{-1} for spring wheat, standard deviation was reduced from 1.93 t ha^{-1} for corn in 2003 to 1.23 t ha^{-1} for spring wheat in 2005.

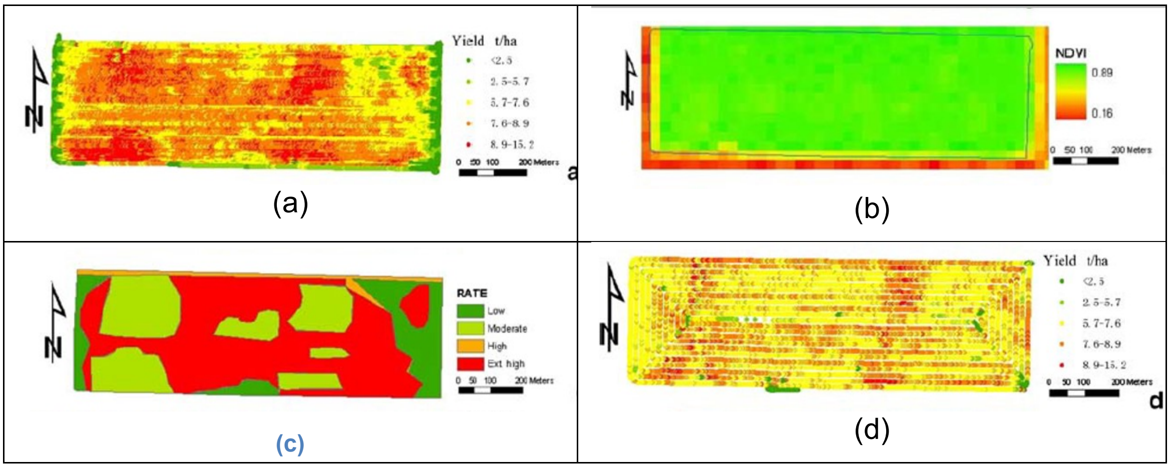


Figure 34. Using the 2003 yield map of corn (a) and 2004 NDVI map by Landsat of August 25, 2004 (b), the farmer delineated the management zones (c) as a basis to determine variable rate fertilizer application resulting in a more uniform yield for 2005 spring wheat (d)
Source: adapted from Zhang et al. (2010).

Active canopy sensor data

Chang et al. (2014) analyzed NDVI data at five growth stages of tobacco growth cycle measured by using a GreenSeeker handheld crop sensor at the location of each sample point. Three soil properties (OM, AP, and Fe) and two stages of NDVI measured (NDVI_40 and NDVI_60) were the critical factors for the tobacco yield. They compared delineation of two MZ methods of : (1) using soil properties (Figure 34a); and (2) using tobacco RS data (Figure 34b). They concluded that it is feasible to use an active canopy sensor to delineate MZs for tobacco-planting fields.

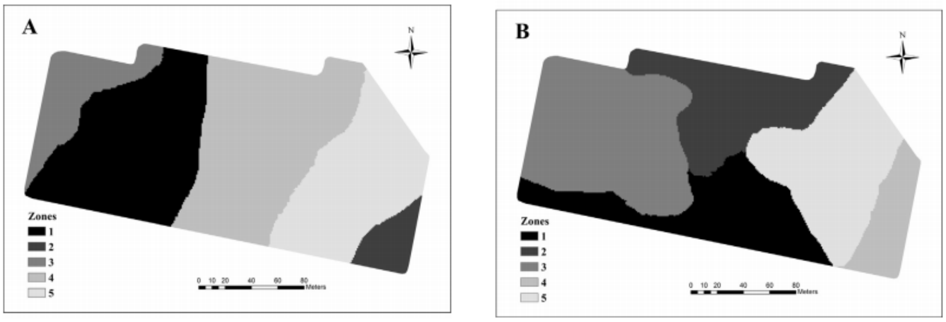


Figure 34. Map of management zones based on soil properties (A) and NDVI measurements (B)
Source: adapted from Chang et al. (2014).

Yield data

Arnó et al. (2005) used normalized yield maps from three years (2002, 2003, and 2004) to delineate a reclassified yield map (zones, Figure 35) in a parcel at Raimat (Lleida, Spain).

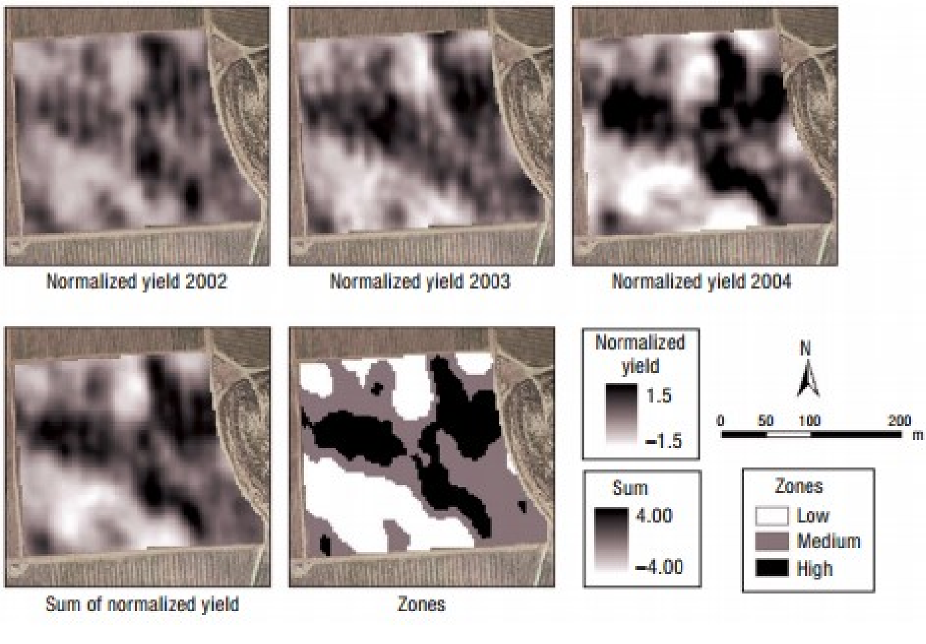


Figure 35. Yield management zones delineated using grape normalized yields
Source: adapted from Arnó et al. (2005).

Topography, electrical conductivity, and soil properties

Molin and Castro (2008) delineated MZs using ECa and eleven other soil properties (P, OM, pH, K, Ca, Mg, SB (sum of bases), CEC (cation exchange capacity), V% (base saturation), Clay, and Sand) in a 35.8-ha area, in Southen Brazil. PCA was used to group variables, and FCM was used to delineate MZs (Figure 36). Results had confirmed ECa utility of to define MZs and feasibility of the proposed method.

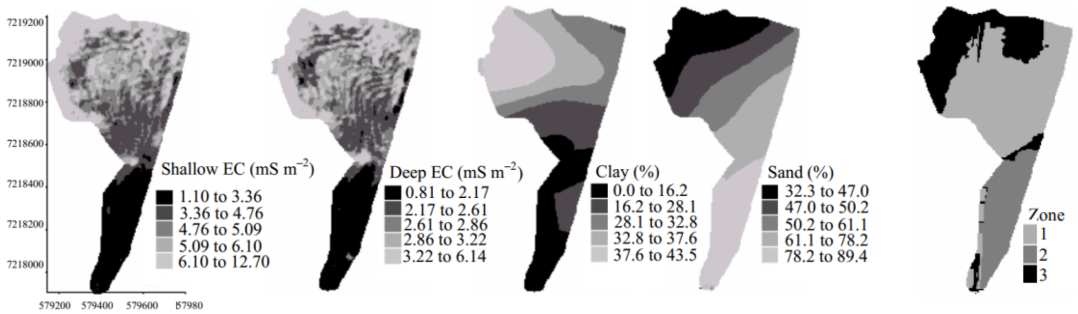


Figure 36. Shallow (0 – 0.3 m) and deep-reading (0 – 0.9 m) soil EC maps, soil clay and sand content maps, and Management zones
Source: adapted from Molin and Castro (2008).

Jaynes et al. (2005) applied cluster analysis of five-year soybean (Glycine max [L.] Merr.) yield to partition a field into a few groups or clusters with similar temporal yield patterns and investigated the relationships among these yield clusters and the easily measured and derived properties (elevation, E; slope, SL; plan curvature, PL; aspect, AS; and depression depth, DD) and ECa (Figure 37). The terrain attributes SL, PL, AS, DD, and ECa effectively identified yield cluster membership for 80% of the 224 transect plots.

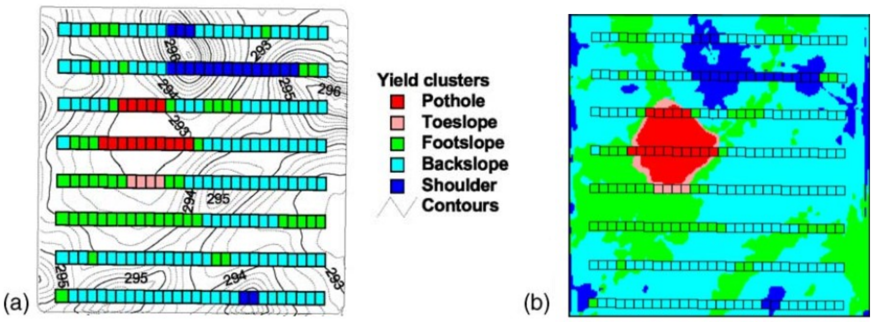


Figure 37. Soybean-yield cluster classification for the 224 transect plots overlaid on elevation contours (a) and the predicted yield zones (b). Transect plots are shown 3× actual width for better visibility.
Source: adapted from Jaynes et al. (2005).

3.3.2.9. Methods to select the variables used in the clustering process

Spatial correlation analysis

Bazzi et al. (2013) used physical and chemical properties of soil and yield from a 19.8-ha commercial farming area in Brazil to delineate MZs by FCM algorithm (Figure 38). The division of the area into two MZs was considered appropriate since it provided distinct averages of most soil properties and yields.

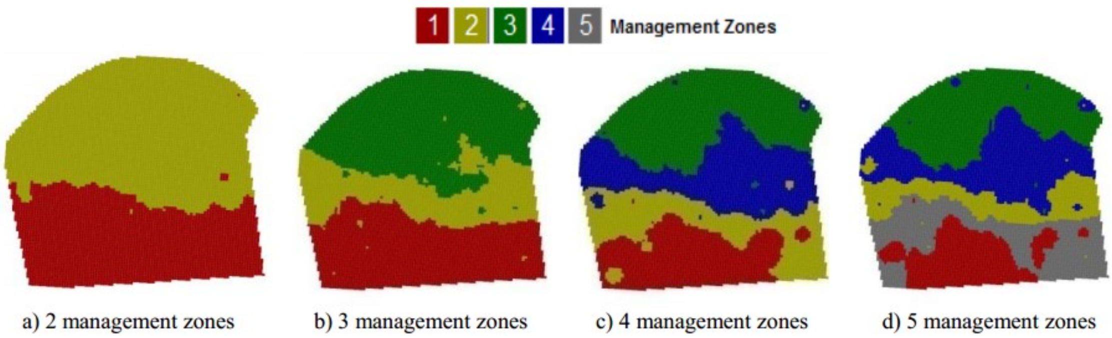


Figure 38. Division of the area into management zones using the Fuzzy C-means algorithm with variables selected with the spatial correlation matrix approach
Source: Bazzi et al. (2013).

Principal component analysis

Molin and Castro (2008) sampled ECa and eleven other soil properties in a 35.8-ha area in Southen Brazil, aiming to delineate MZs with these variables. PCA was used to group variables, and FCM classification was used to cluster the transformed variables (Figure 39). The results confirmed ECa utility to define MZs and feasibility of the proposed method.

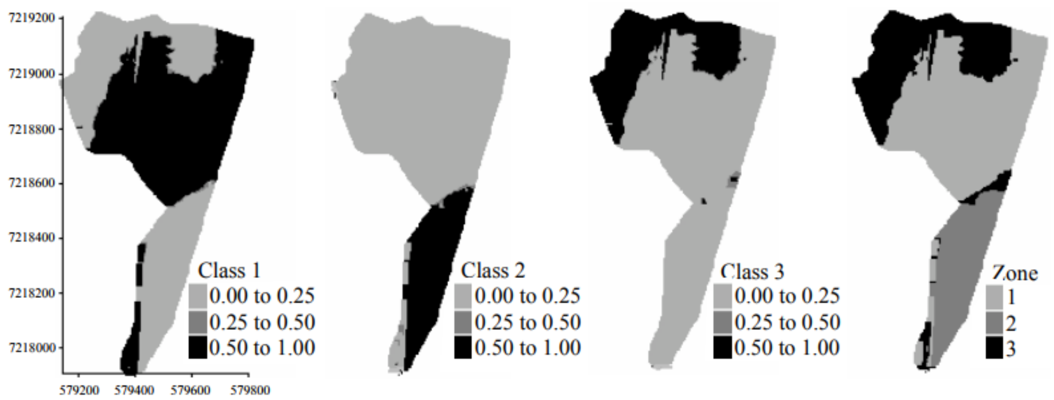


Figure 39. Spatial distribution of participation function values for each individual in the three classes generated after classification by the fuzzy-k-means algorithm of two principal components selected and a corresponding map showing the resulting management zones
Source: Molin and Castro (2008).

Multivariate spatial analysis based on Moran's index PCA (MULTISPATI-PCA)

Córdoba et al. (2016) delineated MZs with ECa, elevation, and soil depth as input variables. MZs were validated using yield, OM, and clay. The field was a rain-fed wheat crop (60 ha) from the Argentine Pampas. They used MULTISPATI-PCA to group variables and FCM clusterization technique and concluded that the best classification was with two zones (Figure 40).

Figure 40. Map with two (left), three (center), and four (right) within-field management classes
Source: Córdoba et al. (2016).

Comparing methods to select the variables

Gavioli et al. (2016) compared the efficiency of six techniques variable selection techniques: (1) All-Attributes: no disposal of stable variables; (2) Spatial-Matrix (Spatial correlation analysis); (3) PCA-All (traditional PCA); (4) MPCA-All (traditional MULTISPATI-PCA); (5) PCA-SC (PCA applied only on the stable variables that showed significant spatial correlation with the yield); and (6) MPCA-SC (MPCA applied only on the stable variables that showed significant spatial correlation with the yield). The methods were used in conjunction with FCM clustering method using data collected from 2010 to 2014 from three agricultural areas in Southern Brazil. The delineated MZs are presented in Figure 41. They founded out that MPCA-SC provided the best performance to define MZs, with greater internal homogeneity, making them more viable for field management.

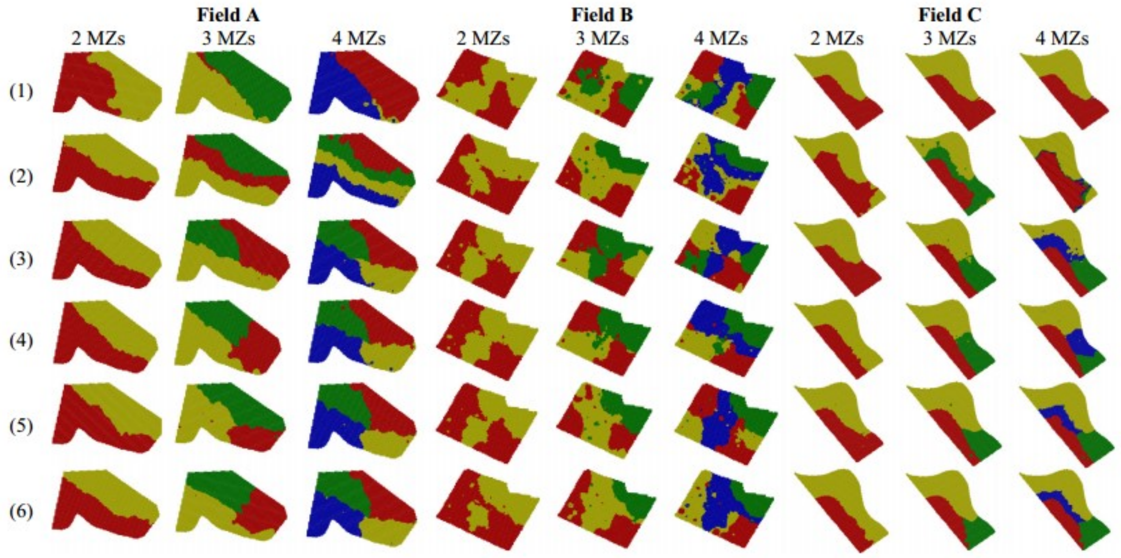


Figure 41. Managements zones generated by the six approaches: (1) All-Attrib; (2) Spatial-Matrix; (3) PCA-All; (4) MPCA-All; (5) PCA-SC; (6) MPCA-SC
Source: Gavioli et al. (2016).

3.3.2.10. Methods to delineate Management Zones

Gavioli et al. (2018), with data, obtained from 2010 to 2015 in three commercial agricultural fields cultivated with soybean and corn in Brazil, evaluated the use of 20 clustering algorithms presented to delineate these subareas: Average Linkage, Bagged Clustering, Centroid Linkage, Clara, Complete Linkage, Diana, Fanny, FCM, Fuzzy C-shells, Hard Competitive Learning, Hybrid Hierarchical Clustering, K-means, McQuitty’s Method, Median Linkage, Neural Gas, Partitioning Around Medoids, Single Linkage, Spherical K-means, Unsupervised Fuzzy Competitive Learning and Ward’s Method. Figure 36 presents the MZs Maps of MZs delineated with the application of 17 (three were discarded, Table 10) clustering algorithms for the three fields. McQuitty’s Method and Fanny were considered the best algorithm because they produced the most significant reductions in the variance of yield in the three fields. In addition, these methods generated classes with high internal homogeneity and delimited MZs without spatial fragmentation (suitable for field operations). The classic FCM and K-means developed significantly different subareas in only two fields, in which the obtained results were similar to the results of McQuitty's Method and Fanny (Figure 42).

Table 9 Clustering methods implemented and compared to define MZs

Method	Acronym	References
Average Linkage[a]	AVG	Jain and Dubes (1988)
Centroid Linkage[a]	CEN	Jain and Dubes (1988)
Complete Linkage[a]	COM	Jain and Dubes (1988)
Divisive Analysis (Diana)[a]	DIA	Kaufman and Rousseeuw (1990)
Hybrid Hierarchical Clustering[a]	HHC	Chipman and Tibshirani (2006)
Median Linkage[a]	MED	Jain and Dubes (1988)
McQuitty's Method (McQuitty)[a]	MCQ	McQuitty (1966)
Ward's Method (Ward)[a]	WAR	Ward (1963)
Single Linkage[a]	SIN	Jain and Dubes (1988)
Bagged Clustering[b]	BCL	Leisch (1999)
Clustering Large Applications (Clara)[b]	CLA	Kaufman and Rousseeuw (1990)
Fuzzy Analysis Clustering (Fanny)[b]	FNY	Kaufman and Rousseeuw (1990)
Fuzzy C-means[b]	FCM	Bezdek (1981)
Fuzzy C-shells[b]	FCS	Dave (1992)
Hard Competitive Learning[b]	HCL	Xu and Wunsch (2009)
K-means[b]	KME	MacQueen (1967)
Neural Gas[b]	NGA	Martinetz et al. (1993)
Partitioning Around Medoids[b]	PAM	Kaufman and Rousseeuw (1990)
Spherical K-means[b]	SKM	Dhillon and Modha (2001)
Unsupervised Fuzzy Competitive Learning[b]	UFCL	Pal et al. (1996)

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

Source: Gavioli et al. (2018).

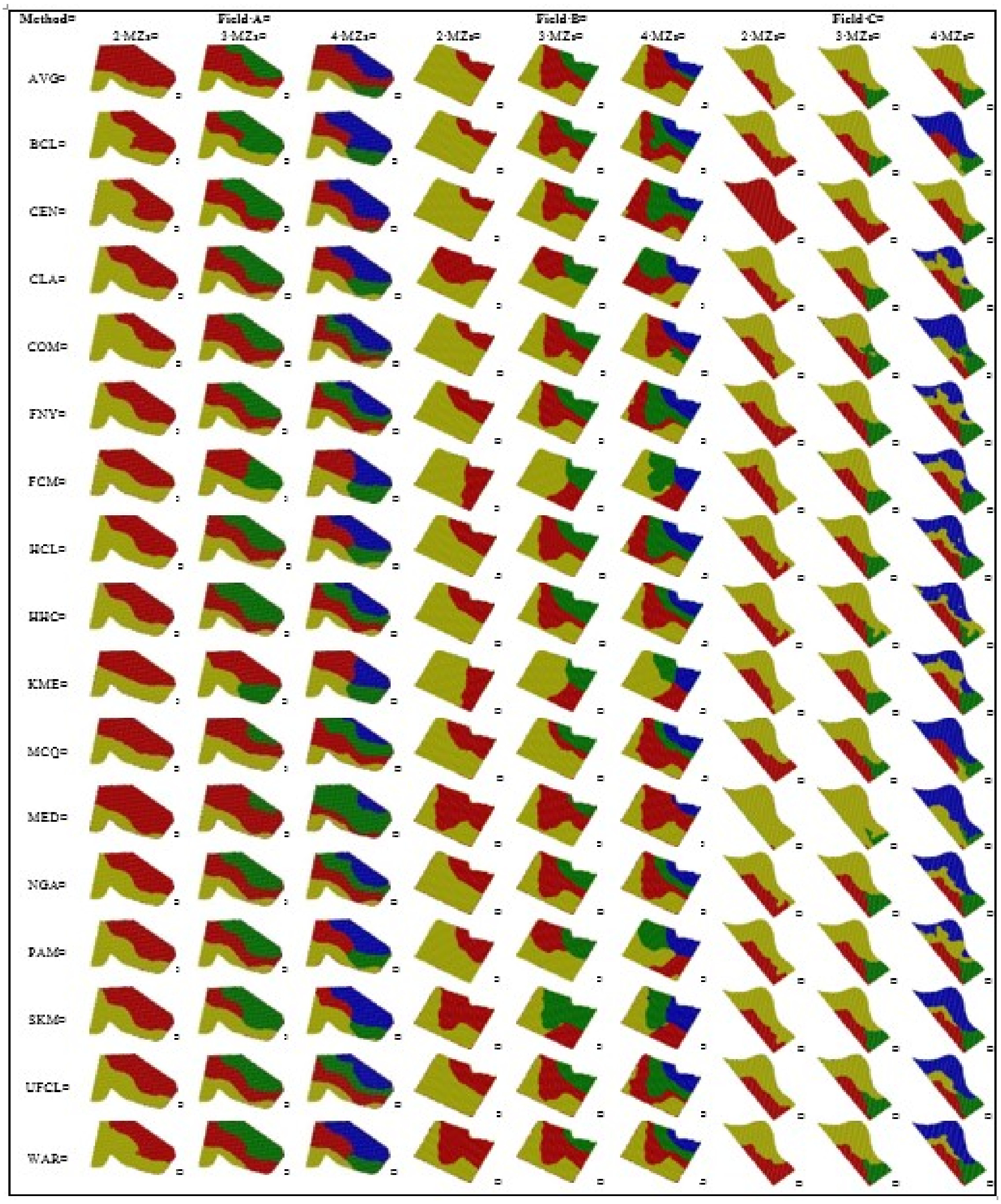
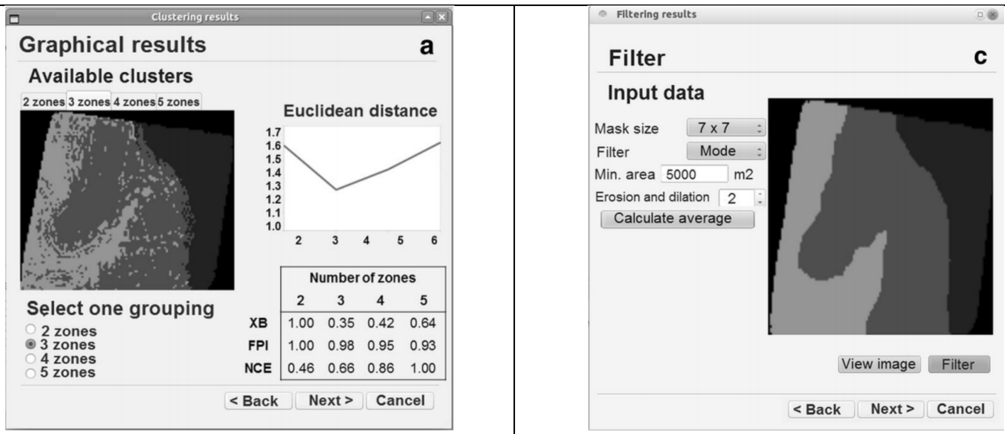


Figure 42. Maps of management zones delineated with the application of 17 clustering algorithms for the three fields
Source: Gavioli et al. (2018).

3.3.2.11. Management Zones Rectification

Albornoz et al. (2018) developed a user-friendly automatic software that integrated all steps to delineate MZs and make prescription files. A careful combination of options in the automatic post-processing methods was selected to reduce fragmentation, including a mode filter with a 7 x 7 mask, erosion and dilation filter, and the fusion of areas smaller than a minimum size of 0.5 hectare. These procedures allow removing all the isolated small areas and improving the border definition and compactness of the delineated zones (Figure 43).



(a) (b)

Figure 43. Zones fragmentation for the delineated management zones (Site 1) before (a) and after (b) the automatic filtering post-processing techniques
Source: adapted from Albornoz et al. (2018).

3.3.2.12. Evaluation of Management Zones Quality

Analysis of Variance, Variance Reduction, Fuzziness Performance Index, Modified Partition Entropy, Smoothness Index, and Improved Cluster Validation Index

As reported before, Gavioli et al. (2016) compared the efficiency of six techniques variable selection techniques (All-Attrib, Spatial-Matrix, PCA-All, MPCA-All, PCA-SC, and MPCA-SC) using these indices: VR, FPI, MPE, SI, ICVI and ANOVA (Table 10). The first analysis to be made is Tukey range test to discard ZMs whose target variable means (in this case yield) are not all statistically distinct. As a result, for field A, it must be considered that it is only advisable to divide it into two ZMs, and the approach all-Attributes is not advised. Regarding the indices, the higher RV and SI, and the lower FPI, MPE, and ICVI, the better MZs; this implies that for area A, the best approach was the Spatial-Matrix.

Table 10 Results for ANOVA (Tukey range test), VR, FPI, MPE, SI, and ICVI, for field A significant at 0.05 confident level

Classes	Approach	ANOVA (Tukey's test)				VR (%)	FPI	MPE	SI (%)	ICVI
		C1	C2	C3	C4					
	All-Attrib	a	a			0.0	0.500	0.079	98.4	1
	Spatial-Matrix	a	b			42.7	0.091	0.018	98.3	0.137
2	PCA-All	a	b			42.5	0.185	0.035	98.5	0.273
	MPCA-All	a	b			25.5	0.161	0.030	98.6	0.368
	PCA-SC	a	b			24.4	0.177	0.032	98.4	0.396
	MPCA-SC	a	b			28.8	0.153	0.029	98.6	0.333
	All-Attrib	a	a	a		0.0	0.667	0.125	97.7	1
	Spatial-Matrix	a	b	b		22.6	0.156	0.032	96.8	0.307
3	PCA-All	a	a	b		39.8	0.287	0.058	97.6	0.298
	MPCA-All	a	a	b		16.7	0.212	0.043	97.5	0.414
	PCA-SC	a	b	a		28.4	0.200	0.042	97.7	0.307
	MPCA-SC	a	b	b		33.6	0.210	0.043	97.7	0.272
	All-Attrib	a	a	a	a	0.0	0.750	0.158	97.1	1
	Spatial-Matrix	a	b	b	a	39.1	0.213	0.044	95.0	0.254
4	PCA-All	a	b	b	a	28.1	0.314	0.069	96.9	0.427
	MPCA-All	a	ab	b	a	20.8	0.215	0.048	96.5	0.388
	PCA-SC	a	b	a	b	48.9	0.178	0.038	97.0	0.159
	MPCA-SC	a	a	b	b	33.7	0.182	0.041	97.2	0.271

Source: Gavioli et al. (2016).

Average silhouette coefficient (ASC)

The indices FPI, MPE, SI, and ICVI cannot be used to evaluate MZs that were not delineated by the clustering process. In this case, a good choice is the coefficient ASC. As reported before, Gavioli et al. (2018) evaluated 20 clustering algorithms, and Table 12 presents the results of 17 methods (three were discarded) in the generation of two, three, and four classes for field A. The clustering process quality was performed by ANOVA (Tukey range test), VR index, and ASC coefficient. The Tukey range test (0.05 level) showed that it was possible to divide the field only with two classes. McQuitty yielded both the highest values for ASC and VR but FCM and K-means also had similar performance.

Table 11 Results of the clustering methods evaluation to generate two, three and four classes by ANOVA (Tukey range test), VR index and ASC coefficient, for field A

Method	2 classes				3 classes					4 classes					
	C_1	C_2	VR%	ASC	C_1	C_2	C_3	VR%	ASC	C_1	C_2	C_3	C_4	VR%	ASC
Average Linkage	a	b	15.9	0.55	a	b	b	18.4	0.45	a	ab	bc	c	20.6	0.46
Bagged Clustering	a	b	16.7	0.58	a	b	b	36.3	0.45	a	b	ab	b	21.3	0.55
Centroid Linkage	a	b	18.2	0.57	a	ab	b	20.4	0.45	a	a	a	a	0	0.41
Clustering Large Applications	a	b	21	0.59	a	b	b	25.3	0.47	a	ab	b	b	19.5	0.55
Complete Linkage	a	a	9.5	0.55	a	ab	b	15	0.46	a	ab	b	b	22.2	0.38
Fuzzy Analysis Clustering (Fanny)	a	b	21.2	0.59	a	b	b	30.2	0.46	a	ab	c	bc	29.6	0.39
Fuzzy C-means (FCM)	a	b	34.1	0.59	a	b	b	35.5	0.46	a	a	b	b	35.6	0.54
Hard Competitive Learning	a	b	21.6	0.59	a	a	b	26.2	0.46	a	b	ab	b	19.9	0.54
Hybrid Hierarchical Clustering	a	b	21.6	0.59	a	a	b	21.4	0.48	a	ab	b	b	21.5	0.38
K-means	a	b	33.8	0.59	a	b	a	23.8	0.46	a	a	b	b	35.8	0.39
McQuitty's Method (McQuitty)	a	b	39.2	0.59	a	b	b	38.3	0.43	a	ab	c	bc	37.4	0.35
Median Linkage	a	b	16.2	0.56	a	b	b	14.4	0.42	a	ab	bc	c	13.2	0.33
Neural Gas	a	b	21.4	0.59	a	b	a	25.8	0.46	ac	b	c	ab	29.7	0.38
Partitioning Around Medoids	a	b	20.9	0.59	a	b	b	29.3	0.46	a	ab	b	b	23.5	0.54
Spherical K-means	a	b	22.4	0.59	a	b	a	41.6	0.47	a	b	b	a	46.9	0.49
Unsupervised Fuzzy Competitive Learning	a	b	21.7	0.59	a	b	b	25.8	0.46	a	ab	bc	c	30.7	0.39
Ward's Method	a	b	19.8	0.58	a	a	b	21.3	0.47	a	ab	c	bc	29.3	0.54

C_i: class *i*; VR: variance reduction index; ASC: average silhouette coefficient.

Source: Gavioli et al. (2018)

Kappa coefficient

Kappa coefficient (K) is applied to measure the degree of agreement among MZ maps generated by the clustering algorithms. As reported before, Kitchen et al. (2005) compared the productivity zones (SPZ) delineated using ECa and elevation with the ones delineated from yield map data (YPZ, Figure 21). Using K, they found out a 60–70% agreement between YPZ and SPZ. They considered this level of agreement promising, especially considering many other yield-limiting factors unrelated to ECa and elevation.

Coefficient of relative deviation and mean absolute difference

Souza et al. (2016) studied the influence of three interpolation methods (i.e., the inverse of distance, inverse of square distance, and the ordinary kriging) commonly used in developing yield maps. They found out that mean absolute difference (MAD) varied from 0.04 to 0.32 t ha^{-1} and corresponded to a relative deviation (CRD) from 1.20 to 7.53%, meaning that the management decisions can differ in some cases on each kind of interpolation implemented.

3.4 REFERENCES

ADAMCHUK, V.; HUMMEL, J.; MORGAN, M.; UPADHYAYA, S. On-the-go soil sensors for precision agriculture. **Comput Electron Agric,** v. 44, p.71–91, 2004.

AGATI, G.; SOUDANI, K.; TUCCIO, L.; FIERINI, E.; BEN GHOZLEN, N.; FADAILI, E.M.; ROMANI, A.; CEROVIC, Z. G. Management Zone Delineation for Winegrape Selective Harvesting Based on Fluorescence-Sensor Mapping of Grape Skin Anthocyanins. **J Agric Food Chem,** v. 66, p. 5778–5789, 2018

AGGELOPOOULOU, K.; CASTRIGNANÒ, A.; GEMTOS, T.; BENEDETTO, D. Delineation of management zones in an apple orchard in Greece using a multivariate approach. **Comput Electron Agric,** v. 90, p. 119–130, 2013.

AIMRUN, W.; AMIN, M. S. M.; EZRIN, M. H.; MASTURA, M. Paddy soil properties and yield characteristics based on apparent electrical conductivity zone delineation for a humid tropical rice farm. **African J Agric Res**, v. 6, p. 5339–5350, 2011.

ALBORNOZ, E. M.; KEMERER, A. C.; GALARZA, R.; MASTAGLIA, N.; MELCHIORI, R.; MARTÍNEZ, C. E. Development and evaluation of an automatic software for management zone delineation. **Precis Agric** v. 19, p. 463–476, 2018.

ALBORNOZ, V. M.; ÑANCO, L. J.; SÁEZ, J. L. Delineating robust rectangular management zones based on column generation algorithm. **Comput Electron Agric**, p. 194–201, 2019

ALVES, S. M. DE F.; ALCÂNTARA, G. R.; QUEIROZ, D. M.; VALENTE, D. S. M. S. M. Definition of Management Zones from Maps of Electrical Conductivity and Organic Matter. **Biosci J** v. 29, p. 104–114, 2013.

AMIDAN, B.; FERRYMAN, T.; COOLEY, S. Data outlier detection using the Chebyshev theorem. **In IEEE Aerosp. Conf**, p. 3814-3819, 2005.

ANDERSON-COOK, C. M.; ALLEY, M. M.; ROYGARD, J. K. F.; KHOSLA, R.; NOBLE, R. B.; DOOLITTLE, J. A. Differentiating soil types using electromagnetic conductivity and crop yield maps. **Soil Science Society of America Journal**, v. 66, p. 1562-1570, 2002.

ARANGO, R. B.; CAMPOS, A. M.; COMBARRO, E. F.; CANAS, E. R.; DÍAZ, I. Identification of agricultural management zones through clustering algorithms with thermal and multispectral satellite imagery. **International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems**, v. 25, p. 121-140, 2017.

ARLOT, S.; CELISSE, A. A survey of cross-validation procedures for model selection. **Statist. Surv**. v. 4, p. 40–79, 2010.

ARNÓ J.; BORDES X.; RIBES-DASI M.; BLANCO R.; ROSELL J.R.; ESTEVE J. Obtaining grape yield maps and analysis of within-field variability in Raimat (Spain). **Proc V ECPA-Eur Conf on Precision Agriculture**. Uppsala, Sweden, p. 899-906, 2005.

ARNO, J.; MARTINEZ-CASASNOVAS, J. Clustering of grape yield maps to delineate site-specific management zones. **Spanish J Agric Res** v. 9, p. 721–729, 2011.

BALASTREIRE, L. A.; BAIO, F. H. R. Avaliação de uma metodologia prática para o mapeamento de plantas daninhas. **Revista Brasileira Engenharia Agrícola e Ambiental**, v. 5, n. 2, p. 349-352, 2001.

BANSOD, B. S.; PANDEY, O. P.; RAJESH, N. L. Analysis and delineation of spatial variability using geo-sensed apparent electrical conductivity and clustering techniques. **Int J Agric Biol** v. 14, p. 481–491, 2012.

BASSO, B.; CAMMARANO, D.; CHEN, D.; CAFIERO, G.; AMATO, M.; BITELLA, G.; ROSSI, R., BASSO, F. Landscape position and precipitation effects on spatial variability of wheat yield and grain protein in Southern Italy. **J Agron Crop Sci**, v. 195, p. 301–312, 2009.

BAUDRON, F.; GILLER, K. E. Agriculture and nature: Trouble and strife?. **Biol Conserv** v. 170, p. 232–245, 2014.

BAZZI, C. L.; JASSE, E. P.; MAGALHÃES, P. S. G.; MICHELON, G. K.; DE SOUZA, E. G.; SCHENATTO, K.; SOBJAK, R. AgDataBox API–Integration of data and software in precision agriculture. **SoftwareX**, v. 10, 8 pp, 2019.

BAZZI, C. L.; SOUZA, E. G.; KONOPATZKI, M. R.; PEREIRA NÓBREGA, L. H.; URIBE-OPAZO, M. A. Management zones applied to pear orchard. **J Food, Agric Environ** v. 13, p. 98–104, 2019.

BAZZI, C. L.; SOUZA, E. G.; SCHENATTO, K.; BETZEK, N. M.; GAVIOLI, A. A software for the delineation of crop management zones (SDUM). **Aust J Crop Sci,** v. 13, p. 26–34, 2019.

BAZZI, C.L.; SOUZA, E.G.; KHOSLA, R.; URIBE-OPAZO, M.A.; SCHENATTO, K. Profit maps for precision agriculture. **Ciencia e Investigación Agraria** (en línea. Edición en Español), v. 42, p. 385–396, 2015.

BAZZI, C. L.; SOUZA, E. G.; URIBE-OPAZO, M. A.; NÓBREGA, L. H. P.; ROCHA, D. M. Management zones definition using soil chemical and physical attributes in a soybean area. **Eng Agrícola**, v. 33, p. 952–964, 2013.

BEHERA, S. K.; MATHUR, R. K.; SHUKLA, A. K.; SURESH, K.; PRAKASH, C. 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**, v. 165, p. 251–259.

BENEDETTO, D. DE, CASTRIGNANO, A.; DIACONO, M.; RINALDI, M.; RUGGIERI, S.; TAMBORRINO, R. Field partition by proximal and remote sensing data fusion. **Biosyst Eng**, v. 114, p. 372–383, 2013a.

BENEDETTO, D. DE, CASTRIGNANÒ, A.; RINALDI, M.; RUGGIERI, S.; SANTORO, F.; FIGORITO, B.; GUALANO, S.; DIACONO, M.; TAMBORRINO, R. An approach for delineating homogeneous zones by using multi-sensor data. **Geoderma**, v. 199, p. 117–127, 2013b.

BERNARDI, A. C. DE C.; TUPY, O.; EDUARDA LIMA SANTOS, K.; MAZZUCO, G. G.; BETTIOL, G. M.; RABELLO, L. M.; INAMASU, R. Y. Mapping of yield, economic return, soil electrical conductivity, and management zones of irrigated corn for silage. **Pesq agropec bras,** v. 53, p. 1289–1298, 2018.

BERNARDI, A. C. C.; BETTIOL G. M.; FERREIRA R. P.; SANTOS K. E. L, RABELLO L. M. AND INAMASU R. Y. Spatial variability of soil properties and yield of a grazed alfalfa pasture in Brazil. **Precision Agriculture**, p. 1–16, 2016.

BETZEK, N. M.; SOUZA, E. G.; BAZZI, C. L.; SCHENATTO, K.; GAVIOLI, A.Rectification methods for optimization of management zones. **Comput Electron Agric**, v. 146, p. 1–11, 2018.

BETZEK, N. M.; SOUZA E. G.; BAZZI, C. L.; SCHENATTO, K.; GAVIOLI, A.; MAGALHÃES, P. S. G. G. Computational routines for the automatic selection of the best parameters used by interpolation methods to create thematic maps. **Comput Electron Agric**, v. 157, p. 49–62, 2019.

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

BIER, V. A.; SOUZA, E. G. Interpolation selection index for delineation of thematic maps. **Comput Electron Agric**, v. 136, p. 202-209, 2017.

BIERNACKI, P.; WALDORF, D. A. N. Snowball Sampling: Problems and Techniques of Chain Referral Sampling. **Sociol Methods Res**, v.10, p. 141–163, 1981.

BLACKMORE, S. The interpretation of trends from multiple yield maps. **Comput Electron Agric**, v. 26, p. 37–51, 2000.

BLACKMORE, S.; MOORE, M. Remedial correction of yield map data. **Precision agriculture,** v. 1, p. 53-66, 1999.

BOBRYK, C. W.; MYERS, D. B.; KITCHEN, N. R.; SHANAHAN, J. F.; SUDDUTH, K. A.; DRUMMOND, S. T.; GUNZENHAUSER, B.; GOMEZ RABOTEAUX, N. N. Validating a digital soil map with corn yield data for precision agriculture decision support. **Agron J**, v. 108, p. 957–965, 2016.

BOLUWADE, A.; MADRAMOOTOO, C.; YARI, A. Application of unsupervised clustering techniques for management zone delineation: case study of variable rate irrigation in southern Alberta, Canada. **J Irrig Drain Eng,** v. 142, p. 05015007, 2016.

BORGES, L. F.; BAZZI, C. L.; SOUZA, E. G. D.; MAGALHÃES, P. S. G.; MICHELON, G. K. Web software to create thematic maps for precision agriculture. **Pesq agropec bras**, v. 55, p. 11 pp, 2020.

BOTTEGA, E. L.; QUEIROZ, D. M. DE, PINTO, F. DE A. DE C.; SOUZA, C. M. A. DE, VALENTE, D. S. M. Precision agriculture applied to soybean: Part I - Delineation of management zones. **Aust J Crop Sci**, v. 11, p. 573–579, 2017.

BOYDELL, B.; MCBRATNEY, A.B. Identifying potential within-field management zones from cotton-yield estimates. **Precis. Agric**, v. 3, p. 9–23, 2002.

BROCK, A.; BROUDER, S.M.; BLUMHOFF, G.; HOFMANN, B.S. Defining yield-based management zones for corn–soybean rotations. **Agron J**, v. 97, p. 1115-1128, 2005.

BUTTAFUOCO, G.; CASTRIGNANÒ, A.; COLECCHIA, A. S.; RICCA, N. Delineation of management zones using soil properties and a multivariate geostatistical approach. **Ital J Agron**, v. 5, p. 323–332, 2010.

BUTTAFUOCO, G.; CASTRIGNANÒ, A.; CUCCI, G.; LACOLLA, G.; LUCÀ, F. Geostatistical modelling of within-field soil and yield variability for management zones delineation: a case study in a durum wheat field. **Precis Agric**, v. 18, p. 37–58, 2017

CAIRES, S. A. DE, WUDDIVIRA, M. N.; BEKELE, I. Spatial analysis for management zone delineation in a humid tropic cocoa plantation. **Precis Agric**, v.16, p. 129–147, 2015.

CARVALHO, V. R. F. **Cadeia produtiva da maçã no Brasil: limitações e potencialidades**. Porto Alegre, BRDE, 2011.

CASTRIGNANÒ, A.; BENEDETTO, D.; GIRONE, G.; GUASTAFERRO, F.; SOLLITTO, D. Characterization, delineation and visualization of agro-ecozones using multivariate geographical clustering. **Ital J Agron**, v. 5, p. 121–132, 2010.

CAVALLO, G.; BENEDETTO, D. DE, CASTRIGNANÒ, A.; QUARTO, R.; VONELLA, A. V.; BUTTAFUOCO, G. Use of geophysical data for assessing 3D soil variation in a durum wheat field and their association with crop yield. **Biosyst Eng**, v. 152, p. 28–40, 2016.

CHANG, D.; ZHANG, J.; ZHU, L.; GE, S. H.; LI, P. Y.; LIU, G. S. Delineation of management zones using an active canopy sensor for a tobacco field. **Comput Electron Agric**, v. 109, p. 172–178, 2014.

CID-GARCIA, N. M.; ALBORNOZ, V.; RIOS-SOLIS, Y. A.; ORTEGA, R. Rectangular shape management zone delineation using integer linear programming. **Comput Electron Agric**, v. 93, p. 1–9, 2013.

COELHO, E. C.; SOUZA, E. G. D.; URIBE-OPAZO, M. A.; PINHEIRO NETO, R. Influência da densidade amostral e do tipo de interpolador na elaboração de mapas temáticos. **Acta Scientiarum. Agronomy**, v. 31, p. 165-174, 2009.

COHEN, J. A. Coefficient of agreement for nominal scales. **Educ Psychol Measur**, v. 20, p. 37–46, 1960.

COHEN, N.; ARIELI, T. Field research in conflict environments: Methodological challenges and snowball sampling. **J Peace Res**, v. 48, p. 423–435, 2011.

COHEN, S.; COHEN, Y.; ALCHANATIS, V.; LEVI, O. Combining spectral and spatial information from aerial hyperspectral images for delineating homogenous management zones. **Biosyst Eng,** v. 114, p. 435–443, 2013

CÓRDOBA, M.; BRUNO, C.; COSTA, J.; BALZARINI, M. Subfield management class delineation using cluster analysis from spatial principal components of soil variables. **Comput Electron Agric,** v. 97, p. 6–14, 2013.

CÓRDOBA, M. A.; BRUNO, C. I.; COSTA, J. L.; PERALTA, N. R.; BALZARINI, M. G. Protocol for Multivariate Homogeneous Zone Delineation in Precision Agriculture. **Biosyst Eng**, v. 143, p. 95–107, 2016.

COSTA, B. R.; OLDONI, H.; JUNIOR, R.; ROMERO, C. AND BASSOI, L. H. Delimitation of homogeneous zones in vineyards using geostatistics and multivariate analysis of different vegetation indices. **Engenharia Agrícola**, v. 39, p.13-22, 2019.

CRESSIE, N. **Statistics for spatial data**. Revised ed. John Wiley & Sons, New York, 1993.

CZAPLEWSKI, R. L.; REICH, R. M. **Expected value and variance of Moran's bivariate spatial autocorrelation statistic for a permutation test**. Fort Collins CO, 1993.

DALL'AGNOL, R. W.; MICHELON, G. K.; BAZZI, C. L.; MAGALHÃES, P. S. G.; DE SOUZA, E. G.; BETZEK, N. M.; SOBJAK, R. Web applications for spatial analyses and thematic map generation. **Comput Electron Agric**, v. 172, 13 p.; 2020.

DAMIAN, M.; PIAS, O. H. D. C.; SANTI, A. L.; DI VIRGILIO, N.; BERGHETTI, J.; BARBANTI, L.; MARTELLI, R. Delineating management zones for precision agriculture applications: A case study on wheat in sub-tropical Brazil. **Ital J Agron**, v. 11, p. 171–179, 2016

DAVATGAR, N.; NEISHABOURI, M. R.; SEPASKHAH, A. R. Delineation of site specific nutrient management zones for a paddy cultivated area based on soil fertility using fuzzy clustering. **Geoderma**, p. 111–118, 2012.

DAVIDSON, D. (2014). Evaluating the quality of your soil. **Crop Soils**, v. 2, p. 4–13, 2014.

DE LARA, A.; KHOSLA, R. AND LONGCHAMPS, L. Characterizing spatial variability in soil water content for precision irrigation management. **Agronomy**, v. 5, 59 p.; 2018.

DEMATTÊ, J. A. M.; DEMATTÊ, J. L. I.; ALVES, E. R.; BARBOSA, R. N.; MORELLI, J. L. Precision agriculture for sugarcane management: a strategy applied for Brazilian conditions. **Acta Sci Agron**, v. 36, p. 111, 2014.

DIACONO, M.; BENEDETTO, D.; CASTRIGNANÒ, A.; RUBINO, P.; VITTI, C.; ABDELRAHMAN, H. M.; SOLLITTO, D.; COCOZZA, C.; VENTRELLA, D. A combined approach of geostatistics and geographical clustering for delineating homogeneous zones in a durum wheat field in organic farming. **NJAS - Wageningen J Life Sci**, v. 65, p. 47–57, 2013

DOBERMANN, A.; PING, J. L.; ADAMCHUK, V. I.; SIMBAHAN, G. C.; FERGUSON, R. B. Classification of Crop Yield Variability in Irrigated Production Fields. **Agron J**, v. 95, p. 1105, 2003.

DOERGE, T.A. **Site-Specific Management Guidelines**. Potash & Phosphate Institute, Norcross, 2000.

DRAY, S.; SAÏD, S.; DÉBIAS, F. Spatial ordination of vegetation data using a generalization of Wartenberg's multivariate spatial correlation. **J Veg Sci**, v. 19, p. 45–56, 2008.

ESRI. **ArcGIS Desktop: Release 9**. Redlands, CA: Environmental Systems Research Institute, 2009.

FARACO, M. A.; URIBE-OPAZO, M. A.; DA SILVA, E. A. A.; JOHANN, J. A. AND BORSSOI, J. A. Seleção de modelos de variabilidade espacial para elaboração de mapas temáticos de atributos físicos do solo e produtividade da soja. **Revista Brasileira de Ciência do Solo**, v. 32, p. 463-476, 2008.

FERGUSON, R. B.; HERGERT, G. W. Soil sampling for precision agriculture. **Ext. Precis. Agric**, p. 1–4, 2009.

FLEMING, K. L.; WESTFALL, D. G.; WIENS, D. W.; BRODAHL, M. C. Evaluating farmer defined management zone maps for variable rate fertilizer application. **Precis Agric**, v. 2, n. 2, p. 201-215, 2000.

FLEMING, K. L.; HEERMANN, D.F.; WESTFALL, D.G. Evaluating soil color with farmer input and apparent soil electrical conductivity for management zone delineation. **Agron J**, v. 96, p. 1581–1587, 2004.

FLEMING, K. L.; WESTFALL, D. G.; WIENS, D. W.; BRODAHL, M. C. Evaluating farmer defined management zone maps for variable rate fertilizer application. **Precis Agric,** v. 2, p. 201-215, 2000.

FRAISSE, C. W.; SUDDUTH, K. A.; KITCHEN, N. R. Delineation of Site-Specific Management Zones by Unsupervised Classification of Topographic Attributes and Soil Electrical Conductivity. **Trans ASAE**, v. 44, p. 155–166, 2001.

FRANZEN, D. W.; HOPKINS, D. H.; SWEENEY, M. D.; ULMER, M. K.; HALVORSON, A. D. Evaluation of Soil Survey Scale for Zone Development of Site-Specific Nitrogen Management. **Agron J**, v. 94, p. 381, 2002.

FRIDGEN, J. J.; KITCHEN, N. R.; SUDDUTH, K. A.; DRUMMOND, S. T.; WIEBOLD, W. J.; FRAISSE, C. W. Management Zone Analyst (MZA): Software for Subfield Management Zone Delineation. **Agron J**, v. 96, p. 100–108, 2004.

FRIDGEN, J.J.; KITCHEN, N.R.; SUDDUTH, K. A. Variability of soil and landscape attributes within sub-field management zones. In P. C. Roberts, et al. (Eds.), Precision agriculture. **Proceedings of the 5th International Conference of the ASA, CSSA, and SSSA**, Madison, WI, USA, 2000.

FROGBROOK, Z. L.; OLIVER, M. A. Identifying management zones in agricultural fields using spatially constrained classification of soil and ancillary data. **Soil Use Manag**, v. 23, p. 40–51, 2007.

FU, Q.; WANG, Z.; JIANG, Q. Delineating soil nutrient management zones based on fuzzy clustering optimized by PSO. **Math Comput Model**, v. 51, p. 1299–1305, 2010.

GALAMBOŠOVÁ, J.; RATAJ, V.; PROKEINOVÁ, R.; PREŠINSKÁ, J. Determining the management zones with hierarchic and non-hierarchic clustering methods. **Res Agric Eng**, v. 60, p. S44–S51, 2014.

GAVIOLI, A.; SOUZA, E. G.; BAZZI, C. L.; GUEDES, L. P. C.; SCHENATTO, K. Optimization of management zone delineation by using spatial principal components. **Comput Electron Agric**, v. 127, p. 302–310, 2016

GAVIOLI, A.; SOUZA E. G.; BAZZI, C. L.; SCHENATTO, K.; BETZEK, N. M. Identification of management zones in precision agriculture: An evaluation of alternative cluster analysis methods. **Biosyst Eng**, v. 181, p. 86–102, 2019.

GILI, A.; ÁLVAREZ, C.; BAGNATO, R.; NOELLEMEYER, E. Comparison of three methods for delineating management zones for site-specific crop management. **Comput Electron Agric**, v. 139, p. 213–223, 2017.

GNANADESIKAN, R.; KETTENRING, J. R.; TSAO, S. L. Weighting and selection of variables for cluster analysis. **J Classif**, v. 12, p. 113–136, 1995.

GONZALEZ, R.C.; WOODS, R. **Digital image processing**, third ed. Pearson Prentice Hall. New Jersey, 2008.

GONZÁLEZ-FERNÁNDEZ, A. B.; RODRÍGUEZ-PÉREZ, J. R.; ABLANEDO, E. S.; ORDOÑEZ, C. Vineyard zone delineation by cluster classification based on annual grape and vine characteristics. **Precis Agric**, v. 18, p. 525–573.

GONZÁLEZ-FERNÁNDEZ, A. B.; RODRÍGUEZ-PÉREZ, J. R.; SANZ-ABLANEDO, E.; VALENCIANO, J. B.; MARCELO, V. Delineating vineyard zones by fuzzy K-means algorithm based on grape sampling variables. **Sci Hortic (Amsterdam)**, v. 243, p. 559–566, 2019

GOZDOWSKI, D.; STEPIEŃ, M.; SAMBORSKI, S.; DOBERS, E. S.; SZATYŁOWICZ, J.; CHORMAŃSKI, J. Determination of the Most Relevant Soil Properties for the Delineation of Management Zones in Production Fields. **Commun Soil Sci Plant Anal**, v. 45, p. 2289–2304, 2014.

GUASTAFERRO, F.; CASTRIGNANÒ, A.; BENEDETTO, D.; SOLLITTO, D.; TROCCOLI, A.; CAFARELLI, B. A comparison of different algorithms for the delineation of management zones. **Precis Agric**, v. 11, p. 600–620, 2010

HAGHVERDI, A.; LEIB, B. G.; WASHINGTON-ALLEN, R. A.; AYERS, P. D.; BUSCHERMOHLE, M. J. Perspectives on delineating management zones for variable rate irrigation. **Comput Electron Agric**, v. 117, p. 154–167, 2015.

HARROWER M.; BREWER C. A. ColorBrewer.org: an online tool for selecting colour schemes for maps. **Cartogr J**, v. 40, p. 27-37, 2003.

HORNUNG, A.; KHOSLA, R.; REICH, R.; INMAN, D.; WESTFALL, D. G. Comparison of Site-Specific Management Zones. **Agron J**, v. 98, p. 407, 2006

HÖRBE, T.D.A.; AMADO, T.J.C.; FERREIRA, A.D.O.; ALBA, P.J. Optimization of corn plant population according to management zones in Southern Brazil. **Precis Agric**, v. 14, p. 450-465, 2013.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. **J Educ Psychol**, v. 6, p. 417–441, 1933

IKENAGA, S.; INAMURA, T. Evaluation of site-specific management zones on a farm with 124 contiguous small paddy fields in a multiple-cropping system. **Precis Agric**, v. 9, p. 147–159, 2008.

INDIEMAPPER. **The basics of data classification**. Available in http://indiemapper.com/app/learnmore.php?l=classification. Accessed at 16 may 2019, 2016.

INMAN, D.; KHOSLA, R.; REICH, R.; WESTFALL, D.G. Normalized difference vegetation index and soil color-based management zones in irrigated maize. **Agron J**, v. 100, p. 60-66, 2008.

ISAAKS, E. H; SRIVASTAVA, R. M. **Applied geostatistics**. New York: Oxford University Press, 1989.

ISPA, INTERNATIONAL SOCIETY FOR PRECISION AGRICULTURE. **ISPA Forms Official Definition of 'Precision Agriculture'**. Available in https://www.precisionag.com/market-watch/ispa-forms-official-definition-of-precision-agriculture/. Accessed at 16 set 2019, 2019.

JACINTHO, J. L.; FERRAZ, G. A. E S.; SILVA, F. M. DA, SANTOS, S. A. Management zones in coffee cultivation. **Rev Bras Eng Agrícola e Ambient**, v. 21, p. 94–99, 2017.

JAYNES, D. B.; COLVIN, T.S. AND KASPAR, T. C. Identifying potential soybean management zones from multi-year yield data. **Comput Electron Agric**, v. 46, p. 309-327, 2005.

JIANG, H. L.; LIU, G. S.; LIU, S. D.; LI, E. H.; WANG, R.; YANG, Y. F.; HU, H. C. Delineation of site-specific management zones based on soil properties for a hillside field in central China. **Arch Agron Soil Sci**, v. 58, p. 1075–1090, 2012.

JIANG, Q.; FU, Q.; WANG, Z. Delineating site-specific irrigation management zones. **Irrig Drain**, v. 60, p. 464–472, 2011.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**, sixth ed. Pearson, New Jersey, 2007.

JOURNEL, A. G.; HUIJBREGTS, C. J. **Mining Geostatistics**. London, New York, San Francisco: Academic Press, 1978.

KARLIK, L.; MARIÁN, G.; FALŤAN, V.; HAVLÍČEK, M. Vineyard zonation based on natural terroir factors using multivariate statistics – Case study Burgenland (Austria). **OENO One**, v. 52, p. 105–117, 2018.

KHAN, M. S.; MEHDI, S. M.; SHEIKH, A. A.; KHAN, A. G.; SHAKAR, M.; ALVI, A.; AHMAD, H. R.; DAVID, M. Delineation of nutrient management zones in Chunian using soil attributes database developed by soil fertility research institute, Punjab, Pakistan. **Soil Enviromnment**, v. 37, p. 116–122, 2018.

KHOSLA, R.; ALLEY, M. M. Soil-specific nitrogen management on mid-Atlantic coastal plain soils. **Better Crops**, v. 83, p. 6-7, 1999.

KHOSLA, R.; FLEMING, K.; DELGADO, J. A.; SHAVER, T. M.; WESTFALL, D. G. Use of site-specific management zones to improve nitrogen management for precision agriculture. **Soil Water**, v. 57, p. 513–518, 2002.

KHOSLA, R.; INMAN, D.; WESTFALL, D. G.; REICH, R. M.; FRASIER, M.; MZUKU, M.; ... & HORNUNG, A. A synthesis of multi-disciplinary research in precision agriculture: site-specific management zones in the semi-arid western Great Plains of the USA. **Precis Agric**, v. 9, p. 85-100, 2008.

KHOSLA, R.; WESTFALL, D. G.; REICH, R. M.; MAHAL, J. S. AND GANGLOFF, W. J. Spatial variation and site-specific management zones. **In Geostatistical applications for precision agriculture**, p. 195-219. Springer, Dordrecht, 2010.

KITCHEN, N.R.; SUDDUTH, K.A.; DRUMMOND, S.T.; FRIDGEN, J.J.; FRAISSE, C.W. Procedures for evaluating unsupervised classification to derive management zones. p. 330-345. **In P.C. Robert et al. (ed.) Proc. 6th Int. Conf. on Precision Agriculture**, Minneapolis, MN. 14-17 July 2002. ASA, CSSA, and SSSA, Madison, WI, 2002.

KITCHEN, N. R.; SUDDUTH, K. A.; MYERS, D. B.; DRUMMOND, S. T.; HONG, S. Y. Delineating productivity zones on claypan soil fields using apparent soil electrical conductivity. **Comput Electron Agric**, v. 46, p. 285–308, 2005.

KITCHENHAM, B.; CHARTERS, S. Guidelines for performing Systematic Literature reviews in Software Engineering. **Engineering**, v. 45, p. 1051, 2007.

KITCHENHAM, B.; PEARL BRERETON, O.; BUDGEN, D.; TURNER, M.; BAILEY, J.; LINKMAN, S. Systematic literature reviews in software engineering - A systematic literature review. **Inf. Softw. Technol**, v. 51, p. 7–15, 2009.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Mellish, C.S. (Ed.), **Proceedings of the 14th International Joint Conference on Artificial Intelligence**, p. 1137–1143, 1995.

KUANG, B.; MAHMOOD, H.S.; QURAISHI, M.Z.; HOOGMOED, W.B.; MOUAZEN, A.M.; VAN HENTEN, E.J. Sensing soil properties in the laboratory, in situ, and on-line: a review. **In Advances in agronomy**, v. 114, p. 155-223. Academic Press, 2012.

KWEON, G. Delineation of site-specific productivity zones using soil properties and topographic attributes with a fuzzy logic system. **Biosystems Engineering**, v. 112, n. 4, p. 261-277, 2012

KYAW, T.; FERGUSON, R. B.; ADAMCHUK, V. I.; MARX, D. B.; TARKALSON, D. D.; MCCALLISTER, D. L. Delineating site-specific management zones for pH-induced iron chlorosis. **Precis Agric**, v. 9, p. 71–84, 2008.

LANDRUM, C.; CASTRIGNANÒ, A.; MUELLER, T.; ZOURARAKIS, D.; ZHU, J.; BENEDETTO, D. DE. An approach for delineating homogeneous within-field zones using proximal sensing and multivariate geostatistics. **Agric Water Manag**, v. 147, p. 144–153, 2015.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, p. 159–174, 1977.

LI, Y.; SHI, Z.; LI, F.; LI, H. Y. Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. **Comput Electron Agric**, v. 56, p. 174–186, 2007.

LI, Y.; SHI, Z.; WU, C.; LI, H.; LI, F. Determination of potential management zones from soil electrical conductivity, yield and crop data. **J Zhejiang Univ B**, v. 9, p. 68–76, 2008.

LI, Y.; SHI, Z.; WU, H. X.; LI, F.; LI, H. Y. Definition of management zones for enhancing cultivated land conservation using combined spatial data. **Environ Manage**, v. 52, p. 792–806, 2013.

LIN, Q.; LI, H.; LUO, W.; LIN, Z.; LI, B. Optimal soil-sampling design for rubber tree management based on fuzzy clustering. **For Ecol Manage**, v. 308, p. 214–222, 2013.

LOISEL, P.; CHARNOMORDIC, B.; JONES, H.; TISSEYRE, B. An optimisation-based approach to generate interpretable within-field zones. **Precis Agric**, v. 20, p. 101–117, 2019.

LONGO, R. D. Desenvolvimento de sistema de informação geográfica para mapeamento da qualidade de maçã. 91 f. Dissertação (Mestrado em Tecnologias Computacionais para o Agronegócio). **Universidade Tecnológica Federal do Paraná**, 2017.

MACQUEEN, J.Some methods for classification and analysis of multivariate observations. **In: Proceedings of the Fifht Berkeley Symposium on Mathematics**, p. 281–296, 1967

MALLARINO, A. P.; WITTRY, D. J. Efficacy of grid and zone soil sampling approaches for site-specific assessment of phosphorus, potassium, pH, and organic matter. **Precis Agric**, v. 5, p. 131–144, 2004.

MARTÍNEZ CASASNOVAS, J. A.; BORDES AYMERICH, X. Viticultura de precisión: Predicción de cosecha a partir de variables del cultivo e índices de vegetación. **Revista de teledetección**, v. 24, p. 67-71, 2005.

MARTÍNEZ-CASASNOVAS, J.; ESCOLÀ, A.; ARNÓ, J. Use of Farmer Knowledge in the Delineation of Potential Management Zones in Precision Agriculture: A Case Study in Maize (Zea mays L.). **Agriculture**, v. 8, p. 84, 2018

MCBRATNEY, A. B.; MOORE, A. W. Application of fuzzy sets to climatic classification. **Agric For Meteorol**, v. 35, p. 165–185, 1985

MCBRATNEY, A.; WEBSTER, R.; Choosing functions for semi-variograms of soil properties and fitting them to sample estimates. **J Soil Sci**, v. 37, p. 617–639, 1986

MCCLYMONT, L.; GOODWIN, I.; MAZZA, M.; BAKER, N.; LANYON, D. M.; ZERIHUN, A.; CHANDRA, S.; DOWNEY, M. O. Effect of site-specific irrigation management on grapevine yield and fruit quality attributes. **Irrig Sci**, v. 30, p. 461–470, 2012

MEIRVENNE, M. VAN, ISLAM, M. M.; SMEDT, P. DE, MEERSCHMAN, E.; VIJVER, E. VAN DE, SAEY, T. Key variables for the identification of soil management classes in the aeolian landscapes of north–west Europe. **Geoderma**, v. 199, p. 99–105, 2013.

MIAO, Y.; MULLA, D. J.; ROBERT, P. C. An integrated approach to site-specific management zone delineation. **Front Agric Sci Eng**, v. 5, p. 432–441, 2008.

MICHELON, G. K.; BAZZI, C. L.; UPADHYAYA, S.; DE SOUZA, E. G.; MAGALHÃES, P. S. G.; BORGES, L. F.; BETZEK, N. M. Software AgDataBox-Map to precision agriculture management. **SoftwareX**, 10, 100320, 7 pp, 2019.

MIEZA, M. S.; CRAVERO, W. R.; KOVAC, F. D.; BARGIANO, P. G. Delineation of site-specific management units for operational applications using the topographic position index in La Pampa, Argentina. **Comput Electron Agric**, v. 127, p. 158-167, 2016.

MINASNY, B.; MCBRATNEY, A. B. **FuzME version 3.0**. Australian Centre for Precision Agriculture, The University of Sydney, Australia, 2002..

MISHRA, U.; CLAY, D.; TROOIEN, T.; DALSTED, K.; MALO, D.; CARLSON, C. G. Assessing the value of using a remote sensing-based evapotranspiration map in site-specific management. **J Plant Nutr**, v. 31, p. 1188–1202, 2008.

MOLIN, J. P. Definição de Unidade de Manejo a Partir de Mapas de Produtividade. **Eng Agrícola**, v. 22, p. 83–92, 2002.

MOLIN, J. P.; CASTRO, C. N. DE. Establishing management zones using soil electrical conductivity and other soil properties by the fuzzy clustering technique. **Sci Agric**, v. 65, p. 567–573, 2008.

MONMONIER M. **How to Lie with Maps**. University of Chicago Press, Chicago, IL, 1996.

MORAL, F. J.; REBOLLO, F. J.; SERRANO, J. M. Estimating and mapping pasture soil fertility in a portuguese montado based on a objective model and geostatistical techniques. **Comput Electron Agric**, v. 157, p. 500–508, 2019

MORAL, F. J.; TERRÓN, J. M.; REBOLLO, F. J. Site-specific management zones based on the Rasch model and geostatistical techniques. **Comput Electron Agric**, v. 75, p. 223–230, 2011.

MORAL, F. J.; TERRÓN, J. M.; SILVA, J. R. M. DA. Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. **Soil Tillage Res**, v. 106, p. 335–343, 2010.

MORARI, F.; CASTRIGNANÒ, A.; PAGLIARIN, C. Application of multivariate geostatistics in delineating management zones within a gravelly vineyard using geo-electrical sensors. **Comput Electron Agric**, v. 68, p. 97–107, 2009.

MOSHIA, M. E.; KHOSLA, R.; LONGCHAMPS, L.; REICH, R.; DAVIS, J. G.; WESTFALL, D. G. Precision manure management across site-specific management zones: grain yield and economic analysis. **Agron J**, v. 106, p. 2146–2156, 2014.

MULLA, D.; KHOSLA, R. **Historical evolution and recent advances in precision farming**. Soil specific farming. CRS Press, Boca Raton, FL, 2016.

MZUKU, M.; KHOSLA, R.; REICH, R.; INMAN, D.; SMITH, F.; MACDONALD, L. Spatial variability of measured soil properties across site-specific management zones. **Soil Sci Soc Am J**, v. 69, p. 1572–1579, 2005.

NASCIMENTO, A. L. DO, SOUSA, E. D. T. DOS S.; SANTOS, F. F. L. DOS, ALMEIDA, S. L. H. DE, VALENTE, D. S. M. Confused Pixels Interference in Maps of Agricultural Management Zones. **J Exp Agric Int**, v. 30, p. 1–10, 2019.

NAWAR, S.; CORSTANJE, R.; HALCRO, G.; MULLA, D.; MOUAZEN, A. M. Delineation of Soil Management Zones for Variable-Rate Fertilization: A Review. **In: Sparks, DL (ed) Advances in Agronomy**, v. 143, p.175–245, 2017.

NIST/SEMATECH, 2012. **E-Handbook of Statistical Methods**, http://www.itl.nist.gov/div898/handbook/. Accessed 23 April 2020, 2012.

ODEH, I. O. A.; CHITTLEBOROUGH, D. J.; MCBRATNEY, A. B. Soil Pattern Recognition with Fuzzy-c-means: Application to Classification and Soil-Landform Interrelationships. **Soil Sci Soc Am J**, v. 56, p. 505, 1992.

OLDONI, H.; BASSOI, L. H. Delineation of irrigation management zones in a Quartzipsamment of the Brazilian semiarid region. **Pesqui Agropecu Bras**, v. 51, p. 1283–1294, 2016.

ORTUANI, B.; CHIARADIA, E. A.; PRIORI, S.; L'ABATE, G.; CANONE, D.; COMUNIAN, A.; GIUDICI, M.; MELE, M.; FACCHI, A. Mapping Soil Water Capacity Through EMI Survey to Delineate Site-Specific Management Units Within an Irrigated Field. **Soil Sci**, v. 181, p. 252–263, 2016.

ORTUANI, B.; SONA, G.; RONCHETTI, G.; MAYER, A.; FACCHI, A. Integrating Geophysical and Multispectral Data to Delineate Homogeneous Management Zones within a Vineyard in Northern Italy. **Sensors**, v. 19, p. 3974, 2019.

PATIL, V. C.; AL-GAADI, K. A.; MADUGUNDU, R.; TOLA, E. K. H. M.; MAREY, S. A.; AL-OMRAN, A. M.; KHOSLA, R.; UPADHYAYA, S. K.; MULLA, D. J.; AL-DOSARI, A. Delineation of management zones and response of spring wheat (Triticum aestivum) to irrigation and nutrient levels in Saudi Arabia. **Int J Agric Biol**, v.16, p. 104–110, 2014.

PERALTA, N. R.; COSTA, J. L. Delineation of management zones with soil apparent electrical conductivity to improve nutrient management. **Comput Electron Agric**, v. 99, p. 218–226, 2013.

PERALTA, N. R.; COSTA, J. L.; BALZARINI, M.; ANGELINI, H. Delineation of management zones with measurements of soil apparent electrical conductivity in the southeastern pampas. **Can J Soil Sci**, v. 93, p. 205–218, 2013.

PERALTA, N. R.; COSTA, J. L.; BALZARINI, M.; CASTRO FRANCO, M.; CÓRDOBA, M.; BULLOCK, D.; Delineation of management zones to improve nitrogen management of wheat. **Comput Electron Agric**, v. 110, p. 103–113, 2015.

PING, J.L.; DOBERMANN, A. Creating spatially contiguous yield classes for site-specific management. **Agron J**, v. 95, p. 1121–1131, 2003.

REYES, J.; WENDROTH, O.; MATOCHA, C.; ZHU, J. Delineating site-specific management zones and evaluating soil water temporal dynamics in a farmer's field in Kentucky. **Vadose Zone J**, v. 1, p. 18, 2019.

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

REZA, S. K.; SARKAR, D.; DARUAH, U.; DAS, T. H. 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**, v. 20, p. 38–48, 2010.

ROBERTS, D. F.; FERGUSON, R. B.; KITCHEN, N. R.; ADAMCHUK, V. I.; SHANAHAN, J. F. Relationships between soil-based management zones and canopy sensing for corn Nitrogen management. **Agron J** ,v. 104, p. 119–129, 2012.

ROBERTSON, M. J.; LYLE, G.; BOWDEN, J. W. Within-field variability of wheat yield and economic implications for spatially variable nutrient management. **F Crop Res**, v. 105, p. 211–220, 2008.

RODRIGUES, S. Variabilidade espacial e manejo localizado de plantas daninhas em Pomar de Pereiras. Dissertação (Mestrado em Mestrado Em Engenharia Agrícola) - **Universidade Estadual do Oeste do Paraná**, 2009.

RODRIGUES, M. S.; CORÁ, J. E. Management zones using fuzzy clustering based on spatial-temporal variability of soil and corn yield. **Eng Agrícola**, v. 35, p. 470–483, 2015.

ROUSSEEUW, P. (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. **J Comput Appl Math**, v. 20, p. 53-65, 1987.

RUß, G. **From spatial data mining in precision agriculture to environmental data mining**. Springer-Verlag, Berlin, 2013.

SALAMI, B. T.; OKOGUN, J. A.; SANGINGA, N. Delineation of management zones by classification of soil physico-chemical properties in the Northern Savanna of Nigeria. **African J Agric Res**, v. 6, p. 1572–1579, 2013.

SANTESTEBAN, L. G.; GUILLAUME, S.; ROYO, J. B.; TISSEYRE, B. Are precision agriculture tools and methods relevant at the whole-vineyard scale? **Precis Agric**, v. 14, p. 2–17, 2013.

SANTI, A. L.; AMADO, T. J. C.; EITELWEIN, M. T.; CHERUBIN, M. R.; SILVA, R. F. DA, ROS, C. O. DA. Definição de zonas de produtividade em áreas manejadas com agricultura de precisão. **Rev Bras Ciencias Agrar**, v. 8, p. 510–515, 2013.

SANTOS, R. T.; SARAIVA, A. M. A Reference Process for Management Zones Delineation in Precision Agriculture. **IEEE Lat Am Trans**, v. 13, p. 727–738, 2015.

SCHEMBERGER, E. E.; FONTANA, F. S.; JOHANN, J. A.; SOUZA, E. G. Data Mining for the Assessment of Management Areas in Precision Agriculture. **Eng Agrícola**, v. 37, p. 185–193, 2017.

SCHENATTO, K.; SOUZA, E. G.; BAZZI, C. L.; BETZEK, N. M.; GAVIOLI, A.; BENEDUZZI, H. M. Use of the farmer’s experience variable in the generation of management zones. **Semina: Ciências Agrárias**, v. 38, p. 2305-2321, 2017a.

SCHENATTO, K.; SOUZA, E. G.; BAZZI, C. L.; BIER, V. A.; BETZEK, N. M.; GAVIOLI, A. Data interpolation in the definition of management zones. **Acta Sci – Technol**, v. 38, p. 31–34, 2016b.

SCHENATTO, K.; SOUZA, E. G.; BAZZI, C. L.; BETZEK, N. M.; GAVIOLI, A. Two approach comparison to define crop management zones (MZs). **African J Agric Res**, v. 11, p. 3680–3689, 2016a.

SCHENATTO, K.; SOUZA, E. G.; BAZZI, C. L.; GAVIOLI, A.; BETZEK, N. M.; BENEDUZZI, H. M. Normalization of data for delineating management zones. **Comput Electron Agric**, v. 143, p. 238–248, 2017b.

SCHEPERS, A. R.; SHANAHAN, F. J.; LIEBIG, M. A.; SCHEPERS, J. S.; JOHNSON, S. H.; LUCHIARI, J. A. Appropriateness of management zones for characterizing spatial variability of soil properties and irrigated corn yields across years. **Agronomy J**, v. 96, p. 195–203, 2004.

SCHWALBERT, R. A.; AMADO, T. J. C.; REIMCHE, G. B.; GEBERT, F. Fine-tuning of wheat (Triticum aestivum, L.) variable nitrogen rate by combining crop sensing and management zones approaches in southern Brazil. **Precis Agric**, v. 20, p. 1–22, 2018.

SCUDIERO, E.; TEATINI, P.; CORWIN, D. L.; DEIANA, R.; BERTI, A.; MORARI, F. Delineation of site-specific management units in a saline region at the Venice Lagoon margin, Italy, using soil reflectance and apparent electrical conductivity. **Comput Electron Agric**, v. 99, p. 54–64, 2013.

SCUDIERO, E.; TEATINI, P.; MANOLI, G.; BRAGA, F.; SKAGGS, T.; MORARI, F. Workflow to Establish Time-Specific Zones in Precision Agriculture by Spatiotemporal Integration of Plant and Soil Sensing Data. **Agronomy**, v. 8, p. 253, 2018.

SERVADIO, P.; BERGONZOLI, S.; VEROTTI, M. Delineation of management zones based on soil mechanical-chemical properties to apply variable rates of inputs throughout a field (VRA). **Eng Agric Environ Food**, v. 10, p. 20–30, 2017

SHADDAD, S. M.; MADRAU, S.; CASTRIGNANÒ, A.; MOUAZEN, A. M. Data fusion techniques for delineation of site-specific management zones in a field in UK. **Precis Agric**, v. 17, p. 200–217, 2016.

SHAMAL, S. A. M.; ALHWAIMEL, S. A.; MOUAZEN, A. M. Application of an on-line sensor to map soil packing density for site specific cultivation. **Soil Tillage Res**, v. 162, p. 78–86, 2016.

SHUKLA, A. K.; SINHA, N. K.; TIWARI, P. K.; PRAKASH, C.; BEHERA, S. K.; LENKA, N. K.; SINGH, V. K.; DWIVEDI, B. S.; MAJUMDAR, K.; KUMAR, A.; SRIVASTAVA, P. C.; PACHAURI, S. P.; MEENA, M. C.; LAKARIA, B. L.; SIDDIQUI, S. Spatial Distribution and Management Zones for Sulphur and Micronutrients in Shiwalik Himalayan Region of India. **L Degrad Dev**, v. 28, p. 959–969, 2017.

SILVA, F. O. Espacialização dos teores de proteína e óleo de soja e milho e sua correlação com os atributos do solo e produtividade. Dissertação (Mestrado em Engenharia Agrícola) – **Universidade Estadual do Oeste do Paraná**, 2016.

SINFIELD, J. V.; FAGERMAN, D. AND COLIC, O.; Evaluation of sensing technologies for on-the-go detection of macro-nutrients in cultivated soils. **Comput Electron Agric**, v. 70, p.1-18, 2010.

SOBJAK, R.; SOUZA, E. G. DE, BAZZI, C. L.; URIBE-OPAZO, M. A.; BETZEK, N. M. Redundant variables and the quality of management zones. **Eng Agrícola**, v. 36, p. 78–93, 2016.

SONG, X.; WANG, J.; HUANG, W.; LIU, L.; YAN, G.; PU, R. The delineation of agricultural management zones with high resolution remotely sensed data. **Precis Agric**, v. 10, p. 471–487, 2009.

SOUZA, E. G. DE, SCHENATTO, K.; BAZZI, C. L. Creating thematic maps and management zones for agriculture fields. **In: 14th International Conference on Precision Agriculture**. Montreal, p. 1–17, 2018.

SOUZA, E. G.; C. L. BAZZI, R. KHOSLA, M. A. URIBE-OPAZO, AND R. M. REICH. Interpolation type and data computation of crop yield maps is important for precision crop production. **Journal of Plant Nutrition**, v. 39, p. 531–538, 2016.

SRINIVASAN. **Handbook of Precision Agriculture: Principles and Applications**. Food Products Press, New York, NY. 708 p, 2006.

SUDDUTH, K.A.; DRUMMOND, S.T. Yield editor. **Agron J**, v. 99, p. 1471-1482, 2007.

SUSZEK, G.; SOUZA, E. G. DE, URIBE-OPAZO, M. A.; NOBREGA, L. H. P. Determination of management zones from normalized and standardized equivalent produtivity maps in the soybean culture. **Eng Agrícola**, v. 31, p. 895–905, 2011.

TAGARAKIS, A.; LIAKOS, V.; FOUNTAS, S.; KOUNDOURAS, S.; GEMTOS, T. A. Management zones delineation using fuzzy clustering techniques in grapevines. **Precis Agric**, v. 14, p. 18–39, 2013.

TALAVERA, J. M.; TOBÓN, L. E.; GÓMEZ, J. A.; CULMAN, M. A.; ARANDA, J. M.; PARRA, D. T.; QUIROZ, L. A.; HOYOS, A.; GARRETA, L. E. Review of IoT applications in agro-industrial and environmental fields. **Comput Electron Agric**, v. 142, p. 283–297, 2017.

TAYLOR, J. A.; MCBRATNEY, A. B.; WHELAN, B. M. Establishing Management Classes for Broadacre Agricultural Production. **Agron J**, v. 99, p. 1366, 2007.

TAYLOR, J. C.; WOOD, G. A.; EARL, R.; GODWIN, R. J. Soil Factors and their Influence on Within-field Crop Variability, Part II: Spatial Analysis and Determination of Management Zones. **Biosyst Eng**, v. 4, p. 441-453, 2003.

TIMLIN, D. J.; PACHEPSKY, Y.; SNYDER, V. A.; BRYANT, R. B. Spatial and temporal variability of corn grain yield on a hillslope. **Soil Sci Soc Am J**, v. 62, p. 764, 1998.

TRIPATHI, R.; NAYAK, A. K.; SHAHID, M.; LAL, B.; GAUTAM, P.; RAJA, R.; MOHANTY, S.; KUMAR, A.; PANDA, B. B.; SAHOO, R. N. Delineation of soil management zones for a rice cultivated area in eastern India using fuzzy clustering. **Catena**, v. 133, p. 128–136, 2015.

URRETAVIZCAYA, I.; SANTESTEBAN, L. G.; TISSEYRE, B.; GUILLAUME, S.; MIRANDA, C.; ROYO, J. B. Oenological significance of vineyard management zones delineated using early grape sampling. **Precis Agric**, v. 15, p. 111–129, 2014.

VALENTE, D. S. M.; QUEIROZ, D. M. DE, PINTO, F. DE A. DE C.; SANTOS, N. T.; SANTOS, F. L. Definition of management zones in coffee production fields based on apparent soil electrical conductivity. **Sci Agric**, v. 69, p. 173–179, 2012.

VEGA, A.; CÓRDOBA, M.; CASTRO-FRANCO, M.; BALZARINI, M. Protocol for automating error removal from yield maps. **Precis Agric**, v. 20, p. 1030-1044, 2019.

VELANDIA, M.; REJESUS, R. M.; BRONSON, K.; SEGARRA, E. Economics of management zone delineation in cotton precision agriculture. **J Cotton Sci**, v. 12, p. 210–227, 2008.

VERMA, R. R.; MANJUNATH, B. L.; SINGH, N. P.; KUMAR, A.; ASOLKAR, T.; CHAVAN, V.; SRIVASTAVA, T. K.; SINGH, P. Soil mapping and delineation of management zones in the Western Ghats of coastal India. **L Degrad Dev**, v. 29, p. 4313–4322, 2018.

VISCARRA ROSSEL R. A.; MCBRATNEY A. B. Laboratory evaluation of a proximal sensing technique for simultaneous measurement of clay and water content. **Geoderma**, v. 85, p. 19-39, 1998.

VITHARANA, U. W. A.; MEIRVENNE, M. VAN, SIMPSON, D.; COCKX, L.; BAERDEMAEKER, J. DE. Key soil and topographic properties to delineate potential management classes for precision agriculture in the European loess area. **Geoderma**, v. 143, p. 206–215, 2008a.

VITHARANA, U. W. A.; MEIRVENNE, M. VAN, SIMPSON, D.; COCKX, L.; HOFMAN, G. Agronomic consequences of potential management zones delineated on the basis of EM38DD measurements. **Near Surf Geophys**, v.6, p. 289–296, 2008b.

WEBSTER, R.; OLIVER, M. A. **Statistical methods in soil and and land resource survey**. Oxford Univ. Press, New York, 1990.

WHETTON, R. L.; WAINE, T. W.; MOUAZEN, A. M. Evaluating management zone maps for variable rate fungicide application and selective harvest. **Comput Electron Agric**, v. 153, p. 202–212, 2018.

WOHLIN, C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. **In: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering**, p. 1–10, 2014.

WOLLENHAUPT, N. C.; WOLKOWSKI, R. P.; CLAYTON, M. K. Mapping soil test phosphorus and potassium for variable-rate fertilizer application. **J Prod Agric** v. 7, p. 441–448, 1994.

XIANG, L.; YU-CHUN, P.; ZHONG-QIANG, G.; CHUN-JIANG, Z. Delineation and Scale Effect of Precision Agriculture Management Zones Using Yield Monitor Data Over Four Years. **Agric Sci China**, v. 6, p. 180–188, 2007.

XIAOHU, Z.; LI, J.; XIAOLEI, Q.; JIANXIU, Q.; JUAN, W.; YAN, Z.; ZHANG, X.; JIANG, L.; QIU, X.; QIU, J.; WANG, J.; ZHU, Y. An improved method of delineating rectangular management zones using a semivariogram-based technique. **Comput Electron Agric**, v. 121, p. 74–83, 2016.

XIN-ZHONG, W.; GUO-SHUN, L.; HONG-CHAO, H.; ZHEN-HAI, W.; QING-HUA, L.; XU-FENG, L.; WEI-HONG, H.; YAN-TAO, L. Determination of management zones for a tobacco field based on soil fertility. **Comput Electron Agric**, v. 65, p. 168–175, 2009.

YAN, L.; ZHOU, S.; AND FENG, L. Delineation of site-specific management zones based on temporal and spatial variability of soil electrical conductivity. **Pedosphere**, v. 17, p. 156–164, 2007.

YAN, L.; ZHOU, S.; FENG L.; AND L. HONG-YI. Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. **Comput Electron Agric**, v. 56, p. 174-186, 2017.

YANG, G.; LIU, J.; ZHAO, C.; LI, Z.; HUANG, Y.; YU, H.; XU, B.; YANG, X.; ZHU, D.; ZHANG, X.; ZHANG, R. Unmanned aerial vehicle remote sensing for field-based crop phenotyping: current status and perspectives. **Front Plant Sci**, v. 8, p. 1111, 2017.

YAO, R.; YANG, J. S.; ZHANG, T.; GAO, P.; WANG, X.; HONG, L.; WANG, M. Determination of site-specific management zones using soil physico-chemical properties and crop yields in coastal reclaimed farmland. **Geoderma**, v. 232, p. 381–393, 2014.

YARI, A.; MADRAMOOTOO, C. A.; WOODS, S. A.; ADAMCHUK, V. I.; HUANG, H. Assessment of Field Spatial and Temporal Variabilities to Delineate Site-Specific Management Zones for Variable-Rate Irrigation. **J Irrig Drain Eng**, v. 143, p. 04017037, 2017.

ZHANG, X.; SHI, L.; JIA, X.; SEIELSTAD, G.; HELGASON, C. Zone mapping application for precision-farming: A decision support tool for variable rate application. **Precis Agric**, v. 11, p. 103–114, 2010.

4 PAPER 1 - AGDATABOX-MAP-FAST TRACK: WEB APPLICATION FOR AUTOMATIC CREATION OF THEMATIC MAPS AND MANAGEMENT ZONES

ABSTRACT: Precision agriculture consists of inputs application in the right amounts at the right time in order to maximize productivity. Two important tasks of this process are the generation of thematic maps (TM) and management zones (MZs). Although extremely important, TMs and MZs delineation depend on great technical knowledge for their construction, which ends up making their use difficult, due to the need for a specialized multidisciplinary team. Thus, this work presents a computational module with a modern interface able to delineate MZs and TMs automatically. The computational module uses state-of-the-art protocols, algorithms and parameters to perform both tasks and the results demonstrate its feasibility and ease of use even by inexperienced users

KEYWORDS: Software, AgDataBox-Map, Automatic protocol

4.1 INTRODUCTION

It is a challenge of nowadays agriculture to make the most optimized use of the land and its inputs as well as to face the pressure to supply the world with much food, fiber, and fuel, as the global population is projected to increase by 3 billion people from now to 2050 (ONU, 2020). According to these data, producers are motivated for both environment preservation and better rational use of all elements from the food production chain (Baudron and Giller, 2014). So, one possibility of a management system that aims at optimizing the use of agricultural inputs, meeting this need for greater profitability with less environmental damage, is precision agriculture (PA). Climatic, topographic, and biological variations, in both spatial and temporal domains, are factors that have induced yield variations in the field. Thus, the premise of PA is to know these variations and provide support for punctual and localized crop management. To achieve that, PA makes use of several tools and techniques. These include thematic maps (TMs) and management zones (MZs).

Thematic maps represent the land and a topic associated with them, and they aim to inform based on graphic symbols where a specific geographical phenomenon occurs. TMs development is linked to data collection, analysis, interpretation, and representation of the information on a map, consequently, similarities are easily identified and spatial correlations visualization can be enabled (Souza et al., 2018). Contour maps are among the most used TMs related to agriculture. 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 the geographic space. While choropleth maps use color to show ranges of a specific variable 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 indexes).

Several questions must be considered when building contour/choropleth maps:

1. Data pre-processing:
    a. Selection of the correct coordinate system,
    b. Conduction of an exploratory data analysis (EDA) to (i) remove outliers, which, in many cases, do not reflect the true nature of data and can have a significant impact up to data-driven decisions (Amidan et al., 2005) and (ii) remove inliers, data that differ significantly from their neighborhood, but lie within the variation range of data set (Córdoba et al., 2016);
2. Data normalization to scale all variables to have the same range;
3. Data interpolation, to provide a dense and regular grid and provide values in places where data were not originally sampled. Choosing the best interpolator for each situation is usually among ordinary kriging, cokriging, and inverse distance weighting (IDW), the most common interpolation methods for agricultural data (Betzek et al., 2019). The decision about the best approach and its parameters depend on the characteristics of data; and
4. Group similar observations and split apart considerably different observations
    a. Choose the best method to break data into ranges, usually manual interval, equal interval, quantile or standard deviation,
    b. Define the number of data classes, usually from five to seven,
    c. Choose the color scheme to show, which depends on the information to be explained. Common options are nominal/qualitative (unorderable data), sequential (orderable data), and diverging (Souza et al., 2018).

Management zones, which can graphicaly be seen as a particular kind of TM, is a subregion of a field that have similar soil and topographic characteristics, consequently, they require equal amounts of inputs, and allow for an optimized uniform management in this sub-area (Doerge, 2000; Moral et al., 2010; Bobryk et al., 2016). As a result, MZs is economically and productively viable in several situations, showing results of cost reduction, increase in yield, and improvement of product quality parameters (Kyaw et al., 2008; Robertson et al., 2008; Velandia et al., 2008; Vitharana et al., 2008; Roberts et al., 2012; Li et al., 2013; Bernardi et al., 2018; Schwalbert et al., 2018; Whetton et al., 2018). Even though several outstanding issues remain, such as: (i) what is the ideal protocol for MZs delineation?, (ii) what is the best algorithm delineation?, (iii) which software allows you to handle all the stages in the process? Each one of these issues unfolds in several others. So, the task of defining ideal MZs is still a challenge (Aikes Junior et al., 2021).

In a previous work (Aikes Junior et al., 2021), we present an extensive study of protocols, both dealing with the complete process of MZs delineation and protocols that deal with only part of the process, main techniques and algorithms used as well as several software used in delineation for a vast amount of papers. Among all, the MZ protocol, proposed by Souza et al. (2018), was considered the most complete, and it divided the process into eight main stages:

1. Acquisition of variables: According to (Nawar et al., 2017), there are seven common properties that are used as an input variable for delineation: (i) farmer knowledge, based on production history (Fleming et al., 2000; Khosla et al., 2002; Hörbe et al., 2013; Schenatto et al., 2017b), (ii) geomorphology, like elevation, slope, plan curvature, aspect, and depression (Jaynes et al., 2005) or the topographic position index (Mieza et al., 2016), (iii) soil chemical and physical analyses (Doerge, 2000; Buttafuoco et al., 2010), (iv) soil class (Khosla and Alley, 1999; Franzen et al., 2002; Brock et al., 2005), (v) yield maps (Blackmore, 2000; Molin, 2002), and (vi) crop coverage, such vegetation indices and leaf area (Chang et al., 2014; Yang et al., 2017), (vii) proximal soil sensors (Adamchuk et al., 2004; Kuang et al,. 2012; Nawar et al., 2017).
2. Remotion of outliers and inliers: Similar to construct TMs, the presence of this kind of data may not reflect the true nature of data and, hence, should not be included in the analyses (Amidan et al., 2005). According to (Córdoba et al., 2016; Vega et al., 2019), the values outside the mean ± three standard deviations (SD) are identified as outliers and should be removed. It is also necessary to remove inliers, data that differ significantly from their neighborhood but lie within the data set variation range (Córdoba et al., 2016). An additional care should be taken for yield data obtained with yield monitor, as well as to eliminate errors associated with unknown header width, to combine filling/emptying times, the time lag of grain by the combine, positional errors, rapid velocity changes, and others (Blackmore and Moore, 1999; Sudduth and Drummond, 2007; Vega et al., 2019).

3. Data normalization: some clustering techniques such as FCM algorithm with Euclidean are sensitive to the input variables' characteristics. Fridgen et al. (2004) reported that Euclidean distance should be used only for statistically independent variables demonstrating equal variances. In this sense, when the Euclidean distance is used to clusteri, data normalization can be crucial before creating MZs (Schenatto et al., 2017a).
4. Selection of input variables: The selection of variables, most related to the target variable, usually crop yield, can be done before (Bazzi et al., 2013; Schenatto et al., 2016; Sobjak et al., 2016) or after (Kitchen et al., 2005; Gozdowski et al., 2014a; Bottega et al., 2017; Miao et al., 2018) MZs delineation. The first approach is the most common, and it seems to reduce the variables' number and/or dimensionality. Furthermore, redundant variables decrease the clustering's performance and increase computational time (Bazzi et al., 2013; Schenatto et al., 2016; Sobjak et al., 2016). According to Gnanadesikan et al. (1995), weighting and selection of variables are the most challenging cluster analyses issue since the variable choice (as well as their weights) can and often will influence the clustering (MZs delineation) (Gozdowski et al., 2014b; Sobjak et al., 2016). On the other hand, good results were obtained with multivariate techniques to reduce variables' dimensionality and promote orthogonality among them (Hotelling, 1933; Dray et al., 2008; Gavioli et al., 2016). The three most used variable selection techniques (Aikes Junior et al., 2021) are spatial correlation analysis (Bazzi et al., 2013), principal component analysis (PCA) (Hotelling, 1933), and multivariate spatial analysis based on Moran's index PCA (MULTISPATI-PCA) (Dray et al., 2008; Córdoba et al., 2013; Gavioli et al., 2016).
5. Data interpolation: usually, data used to delineate MZs are interpolated to do the same on continuous and smooth MZs. Although there are several interpolation methods, in most cases, this task is performed with IDW or kriging interpolation methods. According to Betzek et al. (2019), kriging is the best interpolator when a minimum spatial dependence is confirmed; otherwise, IDW should be used.
6. MZs Delineation - Two approaches are commonly used to delineate MZs (Aikes Junior et al., 2021): (i) empirical method, which uses frequency distribution of target variable (usually yield) to divide the field (Blackmore, 2000), and (ii) cluster analysis such as K-means and FCM (Taylor et al., 2003; Li et al., 2007; Taylor et al., 2007). The second one intended to

divide data points of an agricultural area into classes by employing a similarity evaluation function for this division. In practice, these classes are applied to delineate MZs, subsequently delimited in the field (Boydell and McBratney, 2002; Córdoba et al., 2016).

7. MZs Rectification: After their delineation, MZs often present isolated pixels, small regions, or even a transition border among irregular zones, making them difficult or even impossible to operate in the field. Thus, a smoothing process called rectification can be applied to optimize the zones, usually based on the application of filters mode and median with 3 × 3 and 5 × 5 pixel mask (Betzek et al., 2018) or dilatation filters (Gonzales and Woods, 2008; Córdoba et al., 2016; Albornoz et al., 2018).
8. Evaluation of delineated MZs: the performance of delineation process can be assessed using indices and analysis of variance (ANOVA). These measures aim to quantify how heterogeneous the zones are across the studied field. The most used statistics are (Aikes Junior et al., 2021): (i) variance reduction (VR) (Dobermann et al., 2003), (ii) the fuzziness performance index (FPI) (Fridgen et al., 2004), (iii) modified partition entropy (MPE) (McBratney and Moore, 1985), (iv) normalized classification entropy (NCE) (Bezdek, 1981), (v) improved cluster validation index (ICVI) (Gavioli et al., 2016), (vi) smoothness index (SI) (Gavioli et al., 2016), (vii) average silhouette coefficient (ASC) (Rousseeuw, 1987), (viii) Kappa coefficient (K) (Cohen, 1960), and (ix) coefficient of relative deviation (CRD) (Coelho et al., 2009). In Sobjak (2021) are still listed the Global Quality Index (GQI) and Modified Global Quality Index (MGQI). Some of these indices (FPI, MPE, NCE, and ICVI) can only be used with clustering algorithms that employ fuzzy logic.

Although the adoption of a protocol helps MZs delineation, each of the stages of the protocol unfolds in several options, which have specific advantages and disadvantages, and which, for a correct understanding and analysis, often require the knowledge of several areas, creating difficulty to adopt MZs in agriculture. Some of these difficulties can be reduced by using specialized software. Despite the existence of many software for PA, few are directed to delineate MZs. Golden Software Surfer, ESRI ArcGIS, and R software package are commonly used to delineate MZs, but they do not have all the desired functionality (Aikes Junior et al., 2021). When they have all the necessary functionalities, they are not user-friendly (Albornoz et al., 2018). Another determining factor to hinder access to software is because most present only paid commercial licenses, discouraging its adoption by non-specialized people since they may not realize the advantages of its use at first.

Among the specific software to delineate MZs, the following were well-known (organized by release date): (i) FuzME (Minasny and McBratney, 2002), (ii) Management Zone Analyst (MZA) (Fridgen et al., 2004), (iii) Software for the Definition of Management Zones (SDUM) (Bazzi et al., 2013; Bazzi et al., 2019), (iv) ZoneMAP (Zhang et al., 2010), (v) automatic software to delineate MZs proposed by Albornoz et al. (Albornoz et al., 2018), and (vi) FastMapping (Paccioretti et al., 2020). There are features in the last two ones that allow MZs delineation in a semi-automatic way, to make easy MZs adoption in situations where the operator does not have complete technical knowledge of all the processes and optimal parameters for each step of the process.

The cited computer programs are limited without offering correct coordinate systems, so, they removed outliers (except FastMapping), and rectified the delineated MZs. To address this, AgDataBox (ADB) (Michelon et al., 2019; Borges et al., 2020; Dall'Agnol et al., 2020) aims at integrating data, software, procedures, and methodologies for Digital Agriculture. It is a joint project coordinated by the Western Paraná State University (Unioeste) and the Federal University of Technology - Paraná (UTFPR) with the cooperation of the Colorado State University (CSU), the United States Agricultural Research Service (USDA) in Columbia, the University of California Davis (UC Davis), the University of São Paulo (ESALQ/USP), and the Brazilian Agricultural Research Corporation (Embrapa). This platform is a continuation of software SDUM (Bazzi et al., 2013; Bazzi et al., 2019) and offers a plataform of microservices, accessible by its Application Programming Interface (API), which consists of a set of resources accessible by the Hypertext Transfer Protocol (HTTP) to transfer request and answer messages expressed in JavaScript Object Notation (JSON) format. The ADB-API, in which the data and processing routines are centered, enables several applications' interoperability. Five applications are under development: (i) ADB-Mobile, (ii) ADB-Map, (iii) ADB-Admin, (iv) ADB-IoT, and (v) ADB-RS (Remote Sensing). The ADB-Map application works with spatial data aiming at creating TMs and MZs and has the following functionalities: (i) data importing/exporting, (ii) data analysis and filtering, (iii) data normalization, (iv) data interpolation and creation of TMs, (v) delineation and evaluation of MZs, encompassing variable selection methods, empirical and data clustering methods, and evaluation statistics, (vi) management zone rectification methods, (vii) application map creation and exporting, and (viii) optimal placement of proximal sensors for PA.

Thus, this work presents a new module for ADB, called ADB-Map-Fast Track (ADB-MAP-FT), which allows creating TMs and delineating MZs automatically using a web-friendly interface platform, ideal for users who do not have all the technical knowledge necessary to delineate MZs.

4.2 MATERIAL AND METHODS

ADB-MAP-FT functionalities are divided into different layers, composed of a back-end, which contains business's algorithms and rules of operation, and the front-end, interface of interaction with the user. The back-end is composed of R language and R packages (https://www.r-project.org), specialized in statistics and the analyses of spatial data, Node.JS (https://nodejs.org) to build Representational State Transfer (REST) resources and organize the flow of requests and responses in ADB-API and Apache Tomcat (https://tomcat.apache.org) as an application server. The Angular framework version 9 (https://angular.io), TypeScript programming language (https://www.typescriptlang.org), and NPM package manager (https://www.npmjs.com) were used to write the front-end. In addition, there is a central system using PostgreSQL (https://www.postgresql.org) Database Manager System (DBMS), with PostGIS extension (https://postgis.net), to store data managed by ADB-API. Moreover, each microservice can still use NoSQL database MongoDB (https://www.mongodb.com/), which acts out on storing documents as JavaScript objects. An Internet connection and a web browser are enough to access ADB-MAP-FT by the link https://adb.md.utfpr.edu.br/map, where it can be created a completely free account; there is no need to download or install any program. The application is compatible with different platforms (e.g., Windows, Linux, macOS, Android, and iOS).

4.2.1 Data

The study was conducted within a 15.5-ha area in Céu Azul, Paraná, Southern Brazil (Fig. 1), with the geometric center at coordinates (WGS84) 25º06'32'S and 53º49'55''W. The field has an average altitude of 662 m, and a 1.21º-average slope, using a no-tillage system. At this particular site, a succession of different crops was cultivated for more than ten years. Soybeans and corn were grown during summer harvest or off-season period, and wheat or oats were used as cover crops during the winter. The soil of the area was classified as Rhodic Ferralsol. Forty points (2.67 points ha^{-1}) in an irregular sampling grid were located on the imaginary centreline among the contour lines of each field. As the area has a certain degree of declivity and has contour lines, it was decided to use an irregular sampling grid, defining the points in places that did not coincide with the curves. This is due to the possibility of their influence on productivity. Sampling density was greater than 2.5 points ha^{-1}, following the recommendation of Doerge (2000) and Nanni et al. (2011) to enable the detection of spatial dependence among samples.

Only stable attributes, that is, those recommended for studying MZs delineation (Doerge, 2000), were collected and analyzed (Table 1). Locations of the sampling points were obtained by Global Navigation Satellite System (GNSS) receiver (Juno SB, Trimble Navigation Limited, Sunnyvale, CA, USA), and the elevations were obtained using a total station (GPT-7505, Topcon Corporation, Tokyo, Japan). Soil penetration resistance (SPR) measurements were taken around each point delineated on the sampling grid, using an electronic penetrometer (PenetroLOG, Falker, Porto Alegre, Brazil). The means of measurements were subsequently calculated to represent the sampling value average at depths of 0–0.1, 0.1–0.2, and 0.2–0.3 m. At the same locations, eight subsamples of soil were collected at a 0–0.2m depth within a 3-m radius from the point determined on the grid (adapted from Wollenhaupt et al., 1994). Subsequently, the samples were taken to be analyzed at the laboratory and to obtain data on soil texture (clay, silt, and sand). Soybean and corn yields were determined at the same points in which the soil samples were taken, the harvest and threshing of which occurred manually in a 0.9-m^2 area. Subsequently, yield values were calculated and converted to a 13%-moisture content.

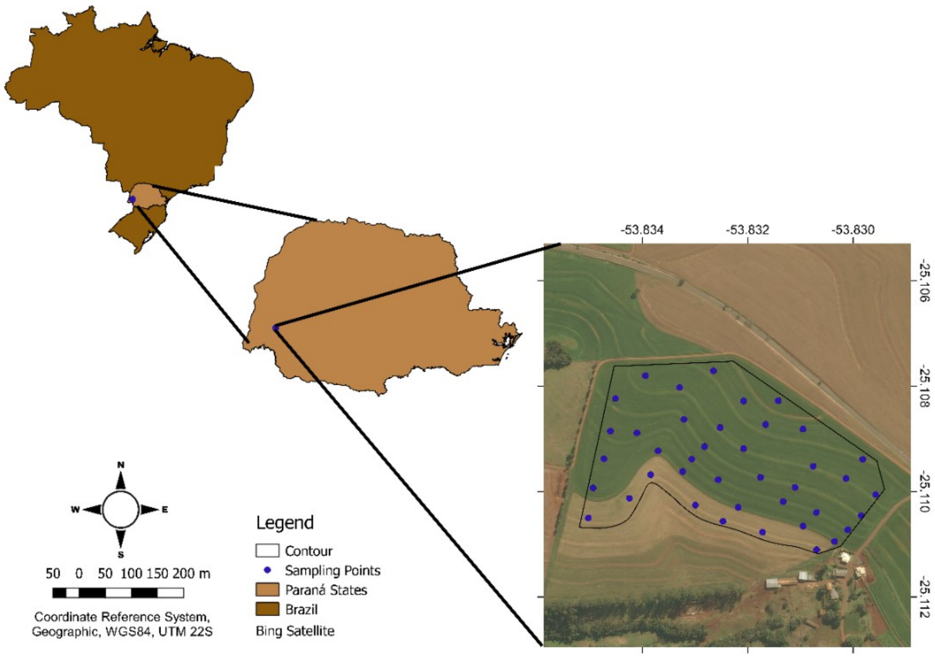


Figure 1. Location of the experimental area in Paraná state, Southern Brazil

Table 1 Attributes collected at each sample point

Attribute / year	2012	2013	2015	2016	2018
Soybean yield (t ha^{-1})	X	X	X	X	X
SPR 0.0-0.1 m (kPa)		X	X		
SPR 0.1-0.2 m (kPa)		X	X		
SPR 0.2-0.3 m (kPa)		X	X		
Altitude (m)		X			

SPR: soil penetration resistance

4.3. SOFTWARE DESCRIPTION

Once authenticated in the system, the user must select which operation he wants to perform: a TM or an MZ (Figure 2). Different procedures and purposes of use, the information requested, and the automatic operations to be carried out will vary according to the applicable option.

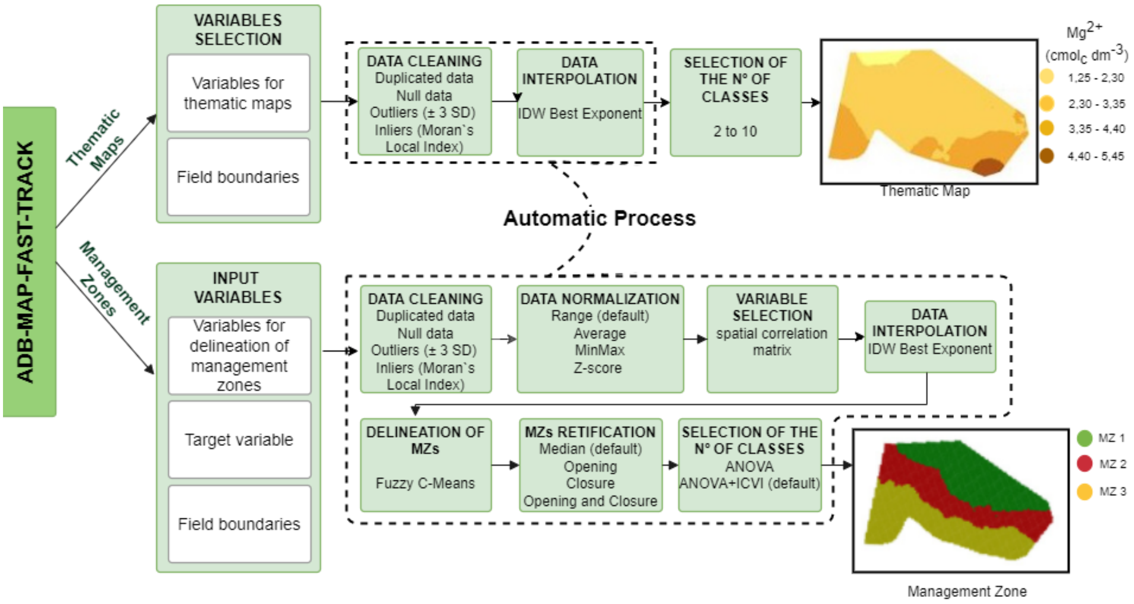


Figure 2. Overview of ADB-Map-Fast Track (ADB-MAP-FT) operation

4.3.1 Thematic map

The user must inform the input variables and boundary file (or create a boundary inside of ADB-Map) to develop a TM, and this information must have been previously imported. ADB-Map allows importing data in text file formats (txt), Comma-Separated Values (CSV), and GeoJSON (JSON), enabling broad compatibility with the file formats most used by geographic information software (GIS). If the user chooses more than one input variable, a TM is generated for each one.

Once the input variables and boundary are informed, the module software will proceed in the following sequence:

1. *Data cleaning:* All layers' coordinate systems (input variables and field boundaries) are validated for a start since ADB allows importing data in different coordinate systems. If the layers are in other coordinate systems, ADB-MAP-FT performs automatic conversion of the coordinate system of all layers to the coordinate system of the open project. Then, duplicate data, zeros, outliers (values outside the mean ± 3 SD), and inliers (using Moran's local index) (Córdoba et al., 2016; Vega et al., 2019) are removed;
2. *Data interpolation*: after data normalization, data are interpolated. For this, IDW interpolator is used (default grid is 1/100 of the largest dimension in North-South and East-West directions), which data are

weighted such that the influence of each sampled point is inversely proportional to the distance raised to the power of the point to be estimated (Isaaks and Srivastava, 1989; Kerry et al., 2010; Betzek et al., 2019). Some parameters must be defined for its use, such as the exponent and the number of neighbors. The greater the exponent is, the lower the influences of points with greater distances. The main difficulty in IDW is choosing the correct exponent to be used in the interpolation (Betzek et al., 2019). Thus, ADB-MAP-FT employs the method proposed by Betzek et al. (2019) to choose the best exponent and number of neighbors. It consists of performing interpolation using a combinator of the exponent from 1 to 6, with jumps of interval 0.5 and with a minimum number of 6 and a maximum of 12 neighbors. The result of these interpolations is measured by the interpolation selection index (ISI) (Bier and de Souza, 2017), which is used to choose the result with better parameters;

3. *The number of classes*: finally, the user selects the number of classes he wants to divide TM, and the map is then presented to the user, and saved in the database, to be available to the user on any device he uses to connect to ADB in a persistent manner, until the moment he decides to delete it. With the map displayed, the user can choose how to classify the observations using manual interval, equal interval, quantile or standard deviation, legend captions, and the color scheme (Souza et al., 2018).

4.3.2 Management zones

The user must inform the input variables to develop MZs, the target variable (to be used in Tukey test to evaluate whether MZS are statistically different), and the field boundaries. This information must have been previously imported with the same file formats supported, similar to TM procedure. Next, the module software will proceed in the following sequence:

1. *Data cleaning*: likewise the first phase of TM creation, we need to select the coordinate system and remove outliers and inliers (all is marked as default);

2. *Data normalization*: the resulting data after cleaning is then passed on to the normalization stage. The default is a ranged method (as indicated by Schenatto et al., 2017b), but the user can change to the other three methods: average, Z-score, and MinMax;
3. *Variable selection:* as mentioned before, the use of redundant variables decreases clustering performance and increases computational time (Bazzi et al., 2013; Schenatto et al., 2016; Sobjak et al., 2016). ADB-MAP-FT adopts the spatial correlation analysis (Bazzi et al., 2013), using Moran's bivariate spatial autocorrelation statistic to build a spatial correlation matrix. Although PCA appears in Aikes Junior et al. (2021) as the most used attribute selection technique, this is essentially a dimensionality reduction technique. Thus, the second most popular technique was chosen (spatial correlation analysis). The procedure consists of (1.) eliminating variables with no significant spatial autocorrelation at 5% significance; (2.) removing variables that were not correlated with the target variable; (3.) decreasing ordination of the remaining variables, considering the correlation degree with the target variable; and (4.) eliminating variables which are correlated with each other, with preference to remove those variables with lower correlation with the target variable;
4. *Interpolation*: the remaining variables are then interpolated by IDW (the default grid is 1/100 of the largest dimension in the North-South and East-West directions). The best exponent and number of neighbors are selected by employing the method proposed by Betzek et al. (2019), similar to TMs module;
5. *MZs Delineation* MZs delineation uses FCM algorithm (Bezdek, 1981). FCM was selected because it is the most popular algorithm for MZs delineation (Aikes Junior et al., 2021). 2 to 5 MZs (most of the works recorded in Aikes Junior et al. (2021) present these limits for practical reasons in the field and software configurations) are created in the standard configuration, with 500 interactions (Minasny and McBratney, 2002; Fridgen et al., 2004) and 1.3-fuzzy exponents (Aikes Junior et al., 2021). The resulting MZs are then sent to the rectification step;
6. *MZs rectification:* here, rectification is performed, by default, employing the best method indicated by Betzek et al. (2018), which consists of

the median filter, with a square mask format with a kernel size of 5 x 5 pixels and five interactions;

7. *Automatic selection of number of MZs:* it is logical to divide the entire field into MZs with a statistically distinct target variable (Souza et al., 2018). Then, the following procedure was taken: (i) Tukey test was applied to identify whether the generated classes showed significant differences in terms of normalized average target variable; (ii) for the selected classes, it was chosen as the best combination of the one with the lowest ICVI (Gavioli et al., 2016).

Finally, MZs are then presented to the user, who can handle the data, change color schemes, caption titles, and so on forth. Thus, as in TMs, the resulting MZs are automatically saved in ADB database, allowing users to consult and manipulate them on their various devices.

It is important to notice that the standard behavior of each stage, that is, algorithms and their parameters, both to create TMs and MZs delineation, was defined based on research results that point them as more appropriate for most cases or, in their absence, the most popular were chosen according to previous research (Aikes Junior et al., 2021). This was done to facilitate users with little technical knowledge, to delineate MZs with real parameters without extensive research. However, let's suppose the user has the necessary technical knowledge (or wants to test different parameters) and intends to use the facilities of ADB-MAP-Fast Track with its integrated steps. In that case, this is still possible since most of the parameters of each step can be customized for the user. For example, if the user wants to change the size of interpolation grid, or fuzzy exponent for FCM algorithm, this is possible by adjusting ADB-MAP-FT settings (more details in the results section). ADB-MAP-FT, due to technical limitations of not implementing all possible algorithms/parameters, on each step, cannot guarantee that the best possible MZ will be created. However, for advanced users, ADB-MAP allows manual execution of each step with significantly more algorithms/parameters.

4.4 RESULTS AND DISCUSSIONS

4.4.1 Descriptive Statistics

The descriptive statistics of values of variables (Table 2) indicate that all variables, except Soybean Yield of 2013 and Slope (2013), were normally distributed. Since the field was almost flat, altitude was the most homogeneous variable, with a coefficient of variation (CV) that is considered low (CV < 10%), followed by Soybeand Yield of 2015 and SPR 0.1-0.2 m of 2013. The SPR 0.0–0.1 m (2013), SPR 0.2–0.3 m (2013), SPR 0.1–0.2 m (2015), SPR 0.2–0.3 m (2015) and Soybean Yield (2016) were considered as having medium CV (10% ≤ CV ≤ 20%). The remaining variables were considered as having a high CV (CV > 20%) (Pimentel-Gomes, 2009).

Table 2 Descriptive statistics of values of variables

Variables	Years	Minimum	Mean	Median	Maximum	CV (%)
Soybean Yield (t ha^{-1})	2012	0.69	2.31	2.00	3.82	40 (h)
Soybean Yield* (t ha^{-1})	2013	2.30	4.05	3.95	8.33	23 (h)
Altitude (m)	2013	651	663	662	676	1 (l)
SPR 0.0–0.1 m (kPa)	2013	1587	2238	2212	3321	17 (m)
SPR 0.1–0.2 m (kPa)	2013	1689	2262	2246	2724	10 (l)
SPR 0.2–0.3 m (kPa)	2013	1337	1919	1900	2294	12 (m)
Soybean Yield (t ha^{-1})	2015	4.14	4.61	4.62	5.14	5 (l)
SPR 0.0–0.1 m (kPa)	2015	657	1480	1458	2387	27 (h)
SPR 0.1–0.2 m (kPa)	2015	1963	2868	2874	3561	11 (m)
SPR 0.2–0.3 m (kPa)	2015	2197	2981	2940	4478	12 (m)
Soybean Yield (t ha^{-1})	2016	2.20	3.67	3.71	4.69	16 (m)
Soybean Yield (t ha^{-1})	2018	1.21	3.10	2.92	8.04	34 (h)

CV, coefficient of variation; h, high; l, low; m, medium; SPR, soil penetration resistance.
* No normality at 5% significance level (Pimentel-Gomes, 2009).

4.4.2 Thematic Maps

After the user has logged in with his credentials, ADB initial screen will be displayed (Figure 3), where the user's projects are listed on the left side. If it is decided, the user can create a new project using the “+ New project” button and enter the registration information such as a name, description, area, and Datum (Figure 4). Once the project is created and/or if you want to open a project, just click on its name in the user's list of projects. After opening the project, as mentioned before, the first step to generate a TM is to import data into ADB platform. Each dataset is created as a layer in the platform. You can click on the “+ New Layer” button to import a layer (Fig. 5), and among the options presented, click on the “Sampling grid” option.

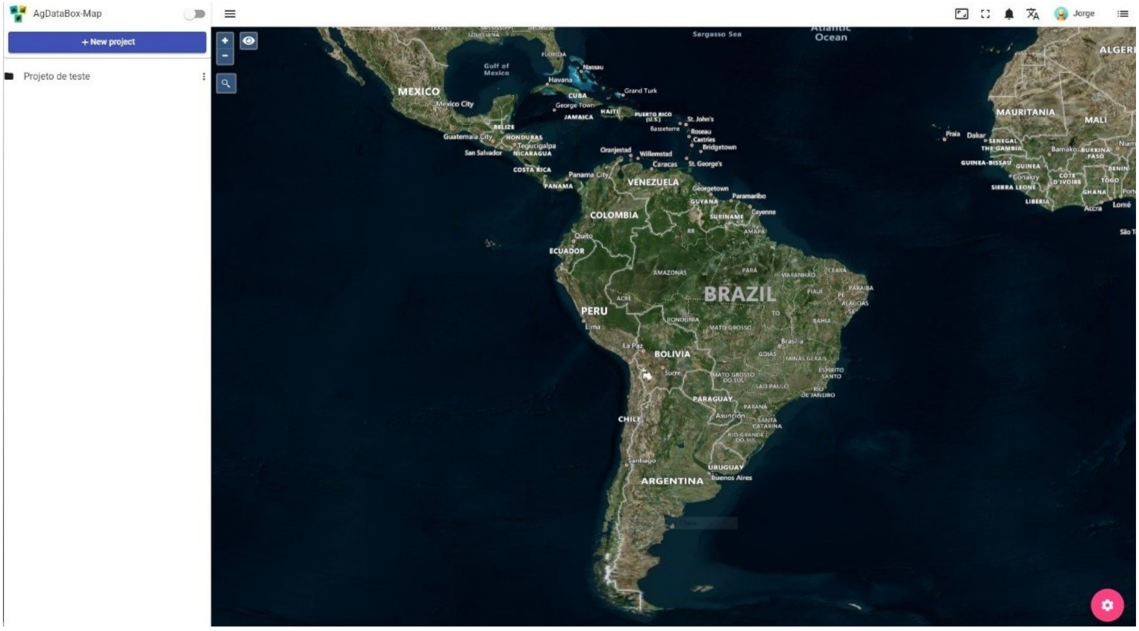


Figure 3 AgDataBox projects screen

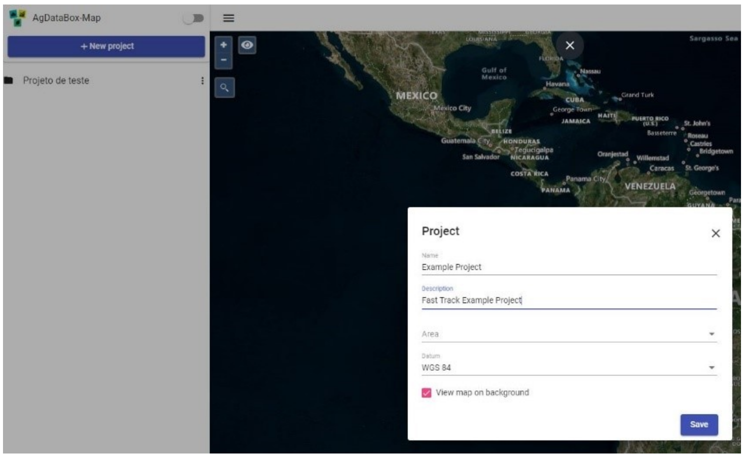


Figure 4 Creation of a new project

Then, it must be chosen whether to import a disk file on the user's device or import data previously saved in the internal ADB API. For this example, the file option will be used, and then the screen to open files of the device operating system is displayed, which can be used to find the desired file. Once the file is chosen, ADB will upload it and present the data import screen for the new layer (Fig. 6). At the top of this screen, the file's name is displayed for confirmation and then the option to inform whether the imported file (in this case, a CSV) has a header that informs the column's name in the file. This option will consider the first line as a data line instead of the column name if unchecked. Below, you can choose the character used as a separator (tab, semicolon, comma, dot, or space), Datum (several options), column containing X coordinate, column containing Y coordinate, and lastly, a list of all other columns in the file. In this list, you can enable the import of a column or do not use the checkbox immediately to the left of its name and provide a name and description for each column. By default, the column name of the file is shown, but here you can change the name to any most convenient string. Once the desired options are configured, you can click on the "Create the layers (import the selected variables)" button, and the new layer must be created and displayed in the sampling list (Fig. 7). Then, if necessary, the same procedure is done to import all the desired data.

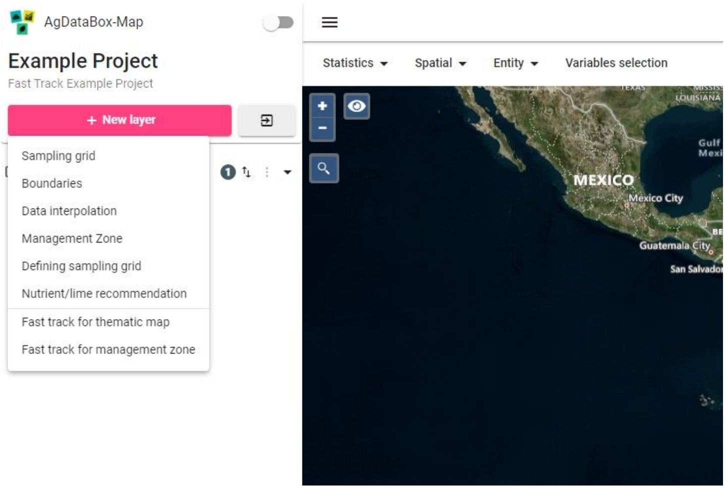


Figure 5 New Layer creation menu

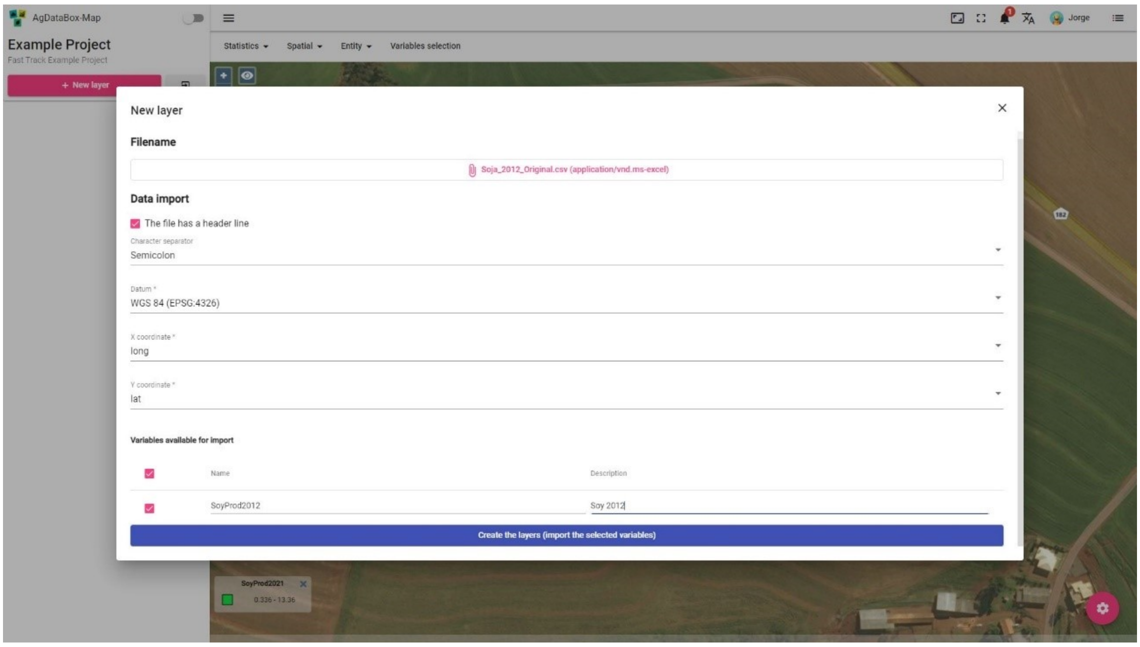


Figure 6 Import data (create layer) options

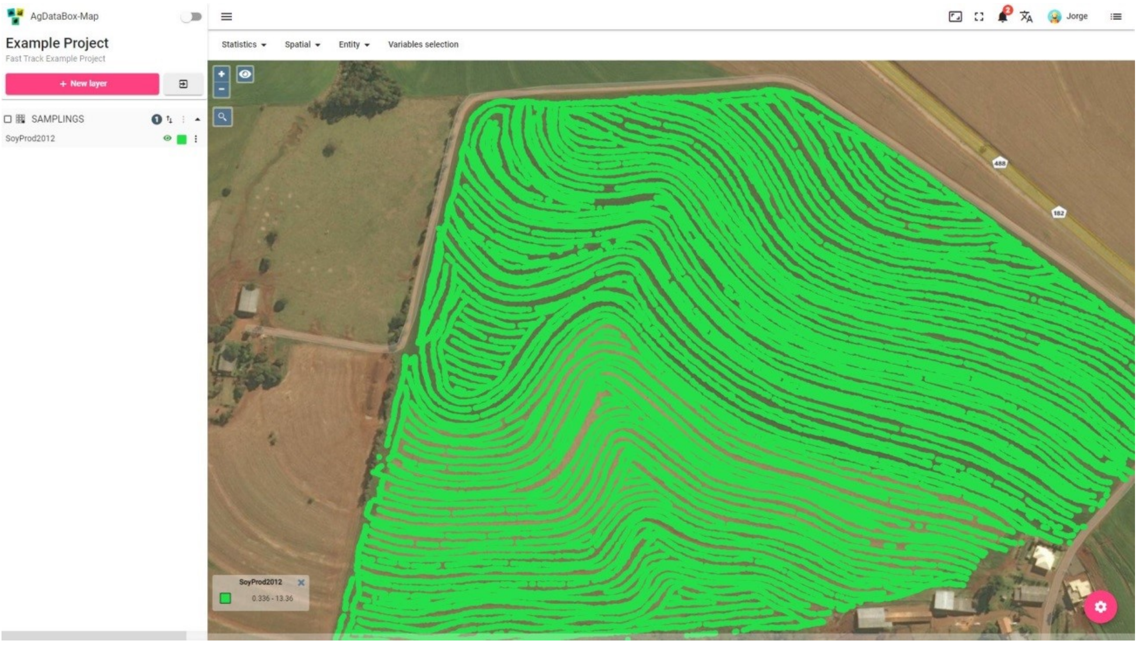


Figure 7 Imported data visualization

There are several controls besides the layer name, such as the possibility to display or not the layer overlaying the map (Fig. 7 - panel on the right), changing the color of the overlaying data, and the ellipsis button, with several options such as normalization, discretization, interpolation, calculation of statistics, edits of data set.

As previously described, for the automatic TMs creation, ADB-MAP-FT can perform data interpolation. For that, field boundaries are necessary. You can similarly import the field boundaries data as importing data layers by clicking on the option "+ New Layer - Boundaries" (Fig. 5), choose the option to import by files or by ADB API, or still perform the delimitation manually using the help map and the option "Delemit new contour". Once the contour has been imported or delimited, it appears in the "Boundaries" section on the left side of the screen and can be viewed similarly to any layer (Fig. 8).

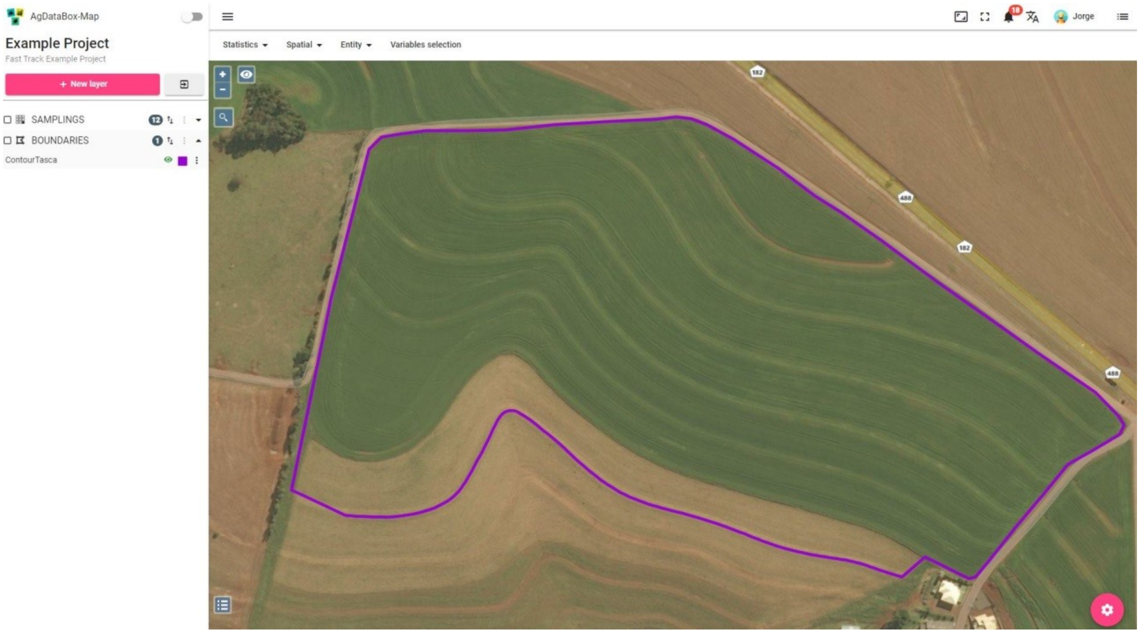


Figure 8 Boundaries (contours) visualization

The option "+ New layer - Fast track for thematic map" can be used to create TM (Fig. 5). The screen of Figure 9 will then be displayed, in which the user will choose one or more layers of samples to create TM(s), the contour to be used as a limiting factor, and the number of classes. In this way, ADB-MAP-FT will generate a TM with the mentioned default settings.

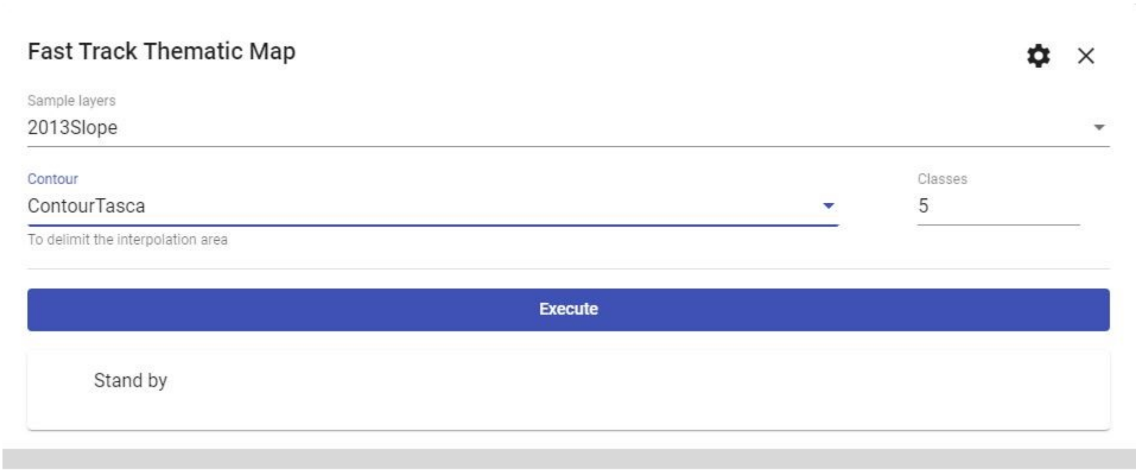


Figure 9 AgDataBox-Map-Fast Track for thematic maps main screen

It is also possible to define some additional settings, if they are intended. So, to do that, the gear button on the upper right side of the window can be used, and the optional settings will be displayed (Fig. 10).

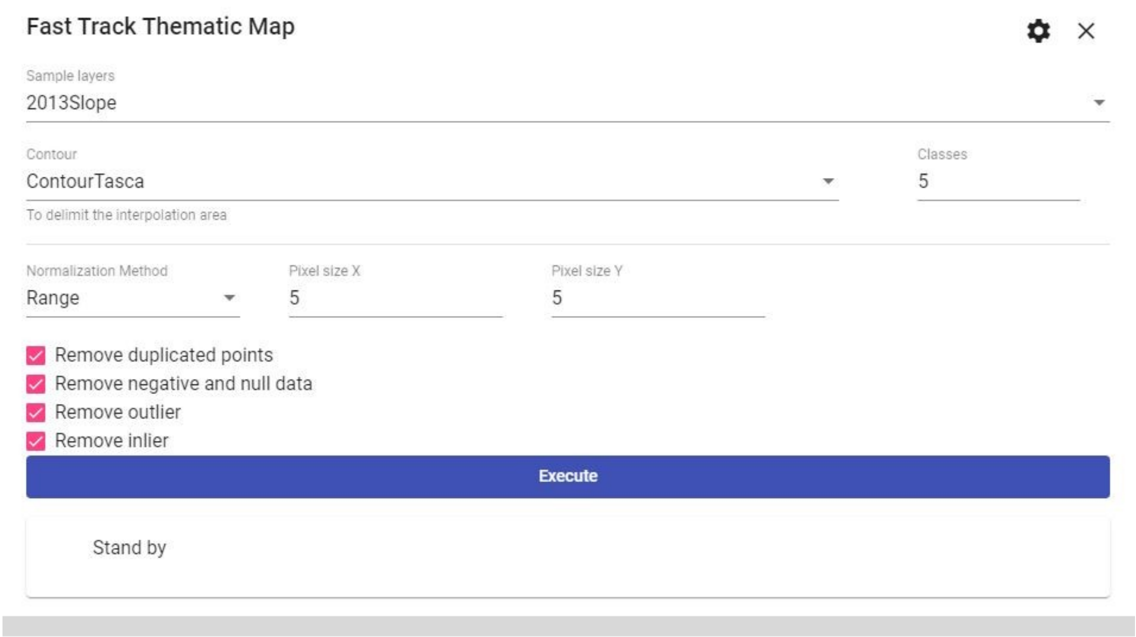


Figure 10 AgDataBox-Map-Fast Track for thematic maps optional configurations

TMs are created by default without data normalization. However, the main screen of ADB-MAP-FT allows choosing data normalization method (Range, Average, MinMax, Z-score), as well as (i) editing the interpolation pixel sizes (X and Y) and (ii) data cleaning options. After clicking the “Execute” button, TMs are created without the user’s intervention. Depending on data amount and settings chosen, the process may take a few minutes. The process evolution can be followed by messages about the steps progress at the bottom of the main window (Fig. 11).

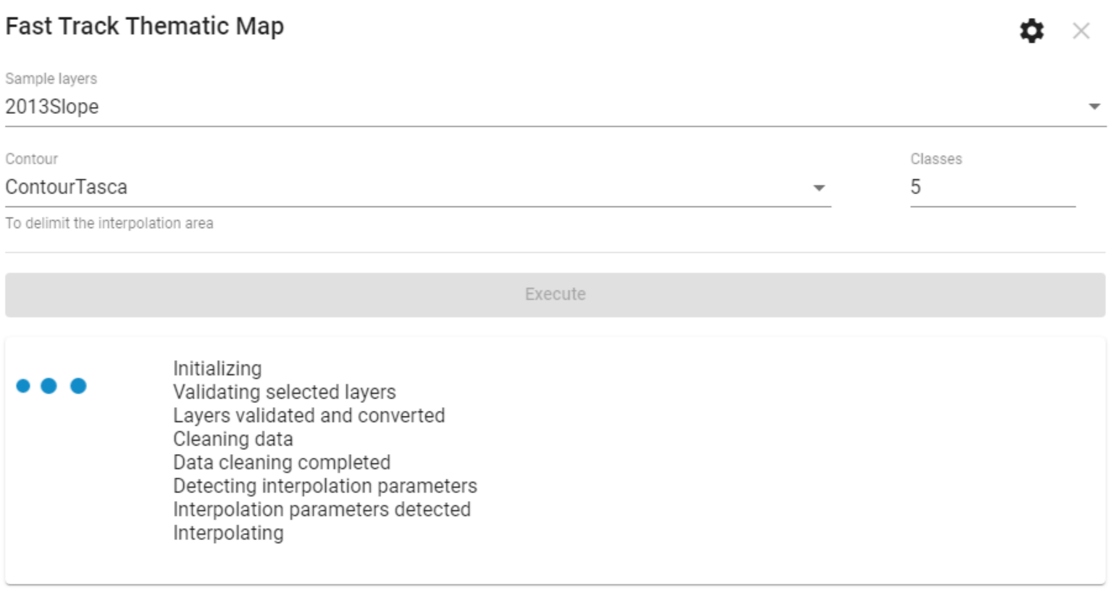


Figure 11 AgDataBox-Map-Fast Track for thematic maps executing message status

When the process is complete, the main window of ADB-MAP-FT for TMs will be automatically closed, and the user will be notified of the new layer created for each TM by the notifications icon located in the upper right corner of the main ADB screen (Fig. 12).

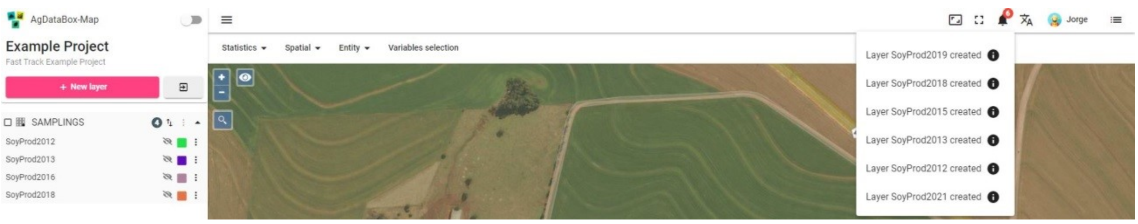


Figure 12 Notifications of concluded operations

The created TMs are stored in the interpolations section on the left side menu of the main screen. Thus, the user should proceed similarly to visualize any layer as well as the generated map (Fig. 13). Then, a new ADB layer is created, and with it, any natively possible layer operation can be performed. For example, the edit layer style allows changing the type of interval classification, the number of classes, color palette, labels to the legend, icon shape, and size (Fig. 14). It can also be displayed the descriptive statistics, data normality tests, discretization and many others, as well as the properties of the resulting dataset, such as its name, the description that informs all the procedures used to generate TM and, in dataset tab, the possibility to visualize and edit each value in each coordinate (Fig. 15). The generated TMs can also be exported in CSV, TXT, and image formats.

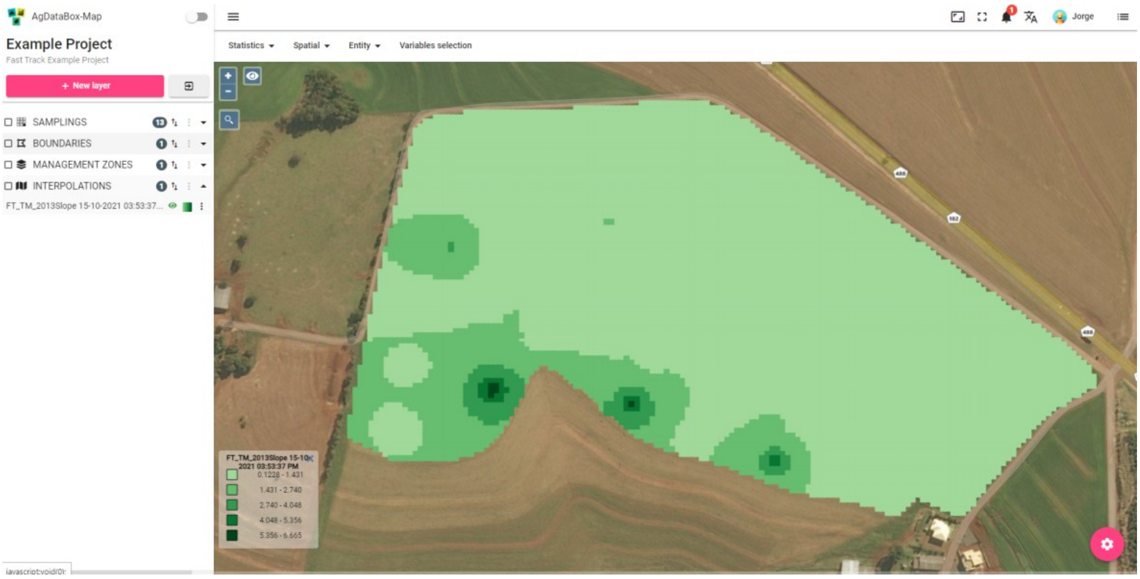


Figure 13 Example of thematic map generated

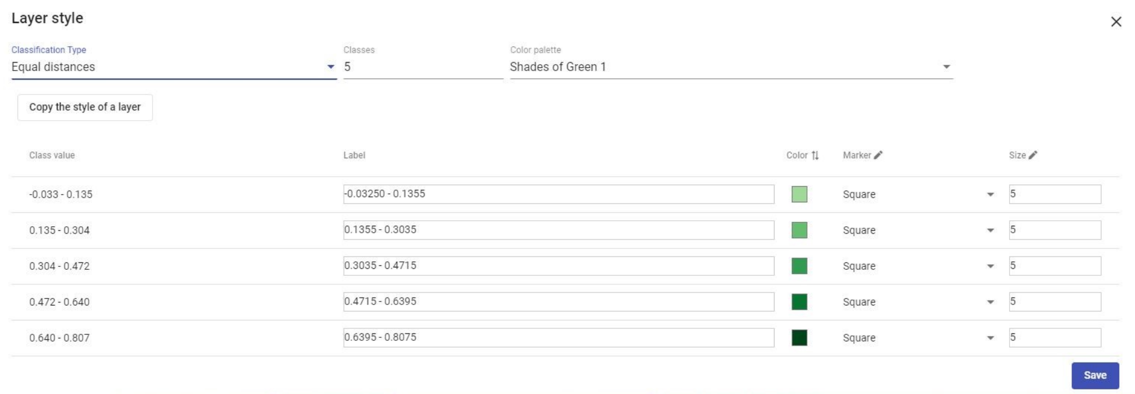


Figure 14 Layer style options

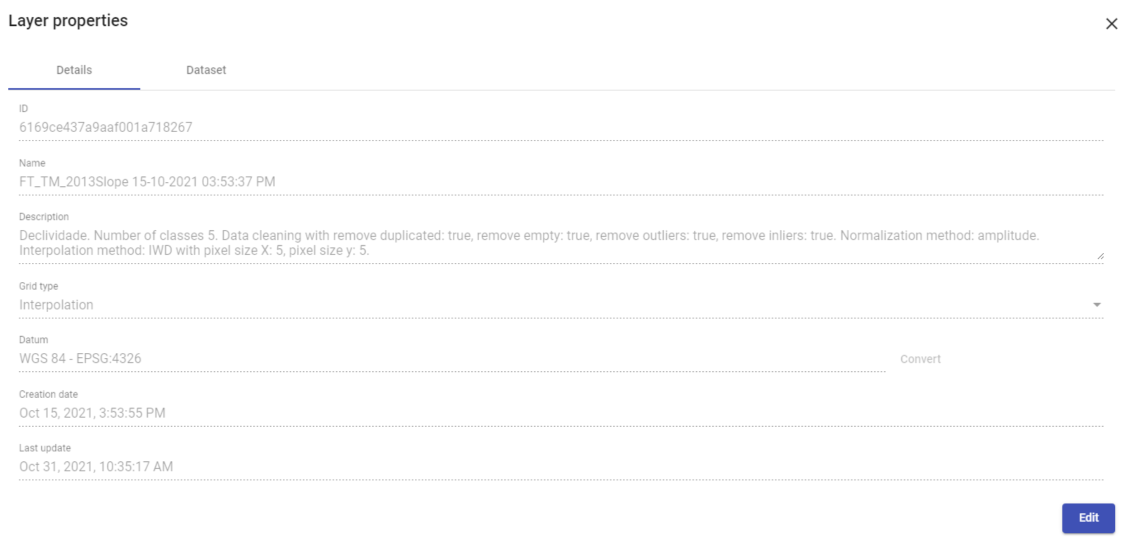


Figure 15 Layer properties of the resulting thematic map

4.4.3 Management Zones

Importing data takes place in a similar way to that performed to create TMs in order to delineate MZs. After data have been imported, the option "+ New layer - Fast track for management zone" (Fig. 5) can be used to access the main screen of the ADB-MAP-FT for MZs (Fig. 16).

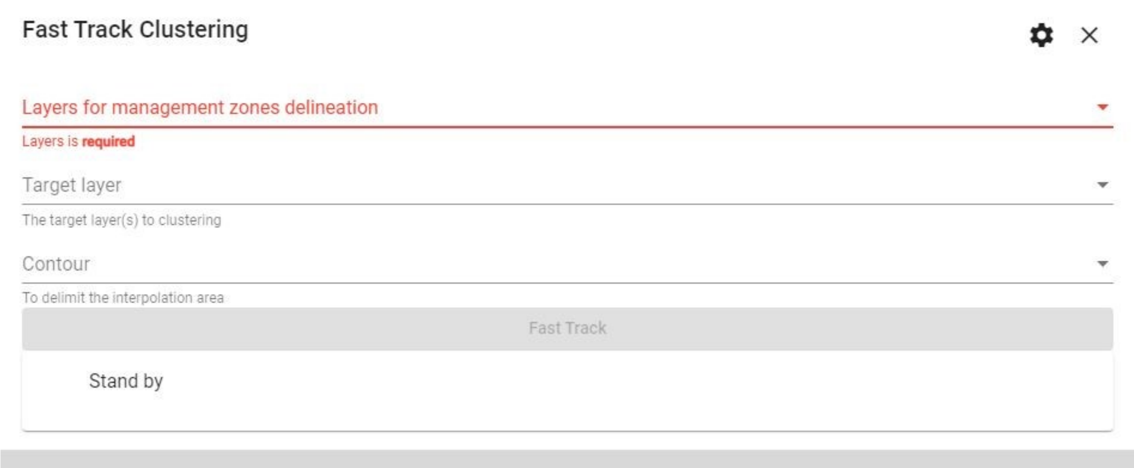


Figure 16 AgDataBox-Map-Fast Track for management zones main screen

In this screen, the user will choose one or more candidate layers to delineate MZs, one layer to be used as target variable, and the boundaries to be used as a limiting factor. In this way, MZs will be generated with the mentioned default settings. However, it is also possible to define several additional settings. To do that, the user can click on the gear button on the upper right side of the window, and the optional settings will be displayed (Fig. 17).

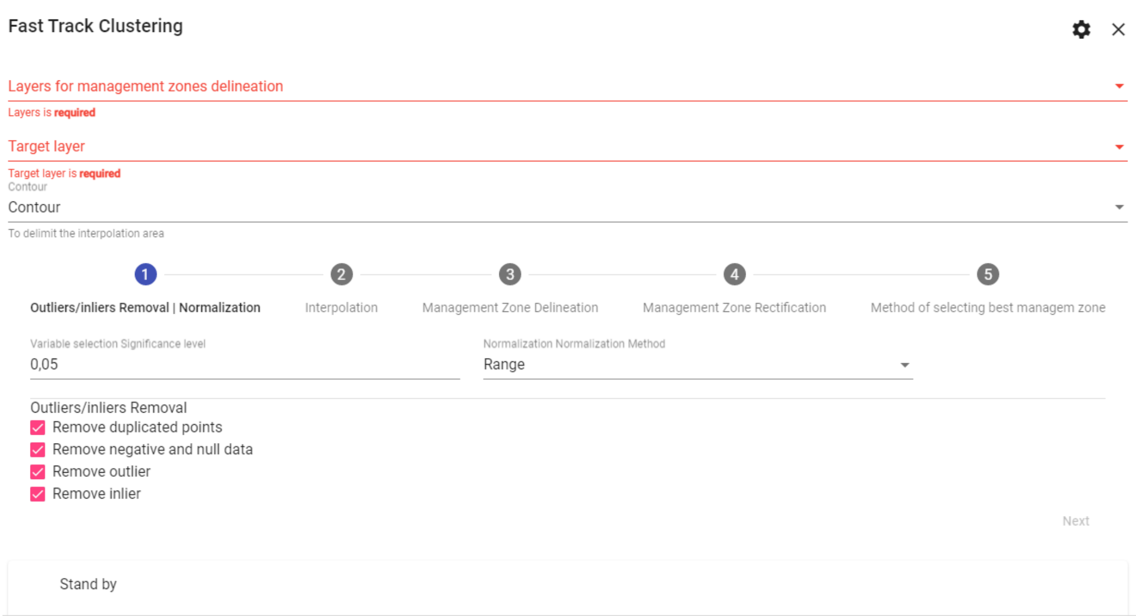


Figure 17 AgDataBox-Map-Fast Track for management zones optional configurations

The possible configurations are presented step-by-step, similar to a tutorial to avoid overloading the user with many options displayed at once,. The settings are divided into:

1. Outliers/inliers | Normalization: in this step, the user can define the level of significance to select variables, the normalization method (Range, Average, MinMax, Z-score), and for data cleaning, the possibility to remove or not duplicate points, negative values and nulls, outliers and inliers;
2. Interpolation: here, the user can define the pixel sizes (X and Y) to interpolate algorithms;
3. Management zone delineation: for now, only the Fuzzy C-means method is available to delineate MZ. It is available to define the number of interactions, the degree fuzzy and the number of management classes (2 to 10);
4. Management zone rectification: here, the user can choose the rectification method (Median, Opening, Closure, Opening, and

Closure), mask format (Square, Circle, Cross), kernel size (3, 5, 7, 9, 11) and number of interactions (1 to 10). The default is median (rectification method), square (mask format), kernel size of 5, and five interactions;

5. Method to select the number of classes: ANOVA, using Tukey test, or ANOVA + ICVI.

Once all the desired settings have been defined, the “Execute Fast Track” button can be clicked, and ADB-MAP-FT for MZs will perform all operations without user’s intervention. Depending on the amount of data and chosen settings, the process may take several minutes. Then, so that the user can follow the evolution of the process, at the bottom of the main window, messages about the steps' progress appear (Fig. 18).

Thus, in order to exemplify ADB-MAP-FT functionalities for MZs, Altitude (2013), Slope (2013), SPR 0.0-0.1 m (2013), SPR 0.1-0.2 m (2013), SPR 0.2-0.3 m (2013), SPR 0.0-0.1 m (2015), SPR 0.1-0.2 m (2015) and SPR 0.2-0.3 m (2015) variables were selected as possible input layers. The target layer was the normalized average soybean yield of 2012, 2013, 2015, 2016, and 2018. All configurations were kept as default (Section 4.3.2).

Once the process is concluded, the resulting MZs are stored in the Management Zones section on the left side menu of the main screen. The user can proceed similarly to visualize any layer as well as the delineated MZs, (Fig. 19). A new ADB layer is created for each validated MZ, and with it, any natively possible layer operation can be performed. For example, the edit layer style allows (i) changing the type of interval classification, (ii) the number of classes, (iii) color palette, (iv) icon shape, (v) label to legend, and (vi) size, as well as the properties of the resulting dataset (its name, the description that informs all the procedures used to generate MZs and, in the dataset tab, the possibility to visualize and edit each value in each coordinate (Fig. 20). Some quality measures on delineated MZs can be observed in the Delineated MZs Quality table (Fig. 21).

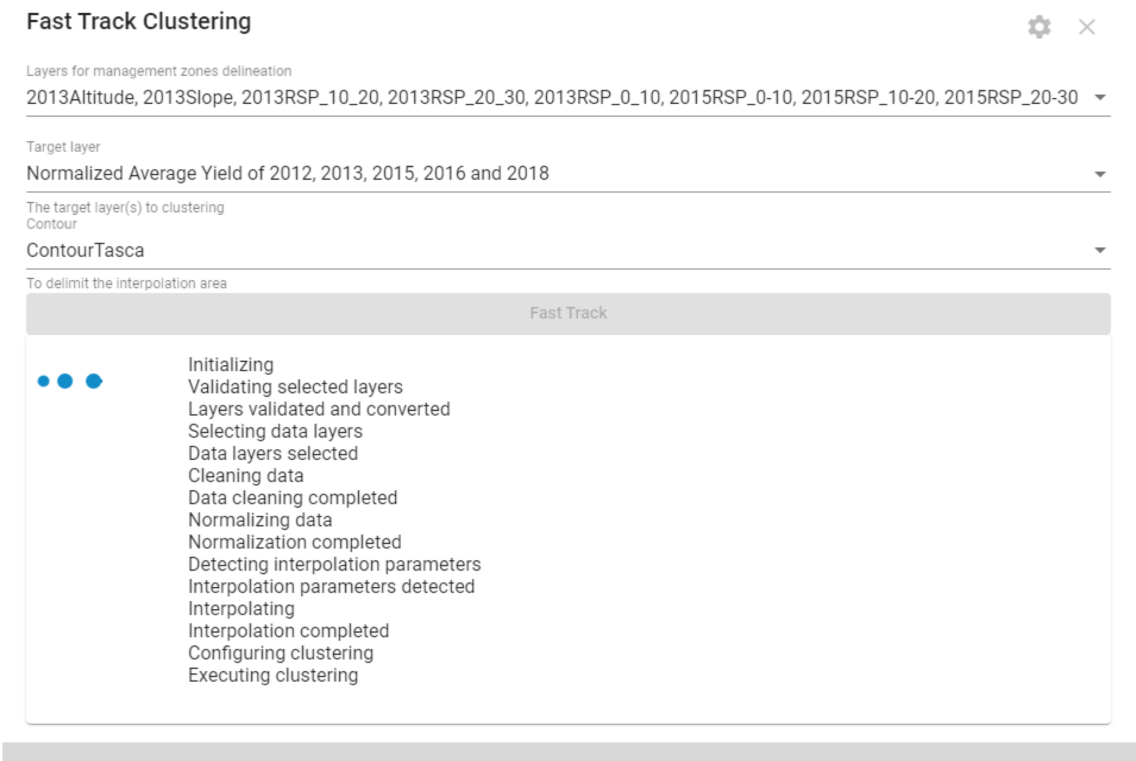


Figure 18 AgDataBox-Map-Fast Track for management zones executing message status

Figure 19 Example of delineated management zone

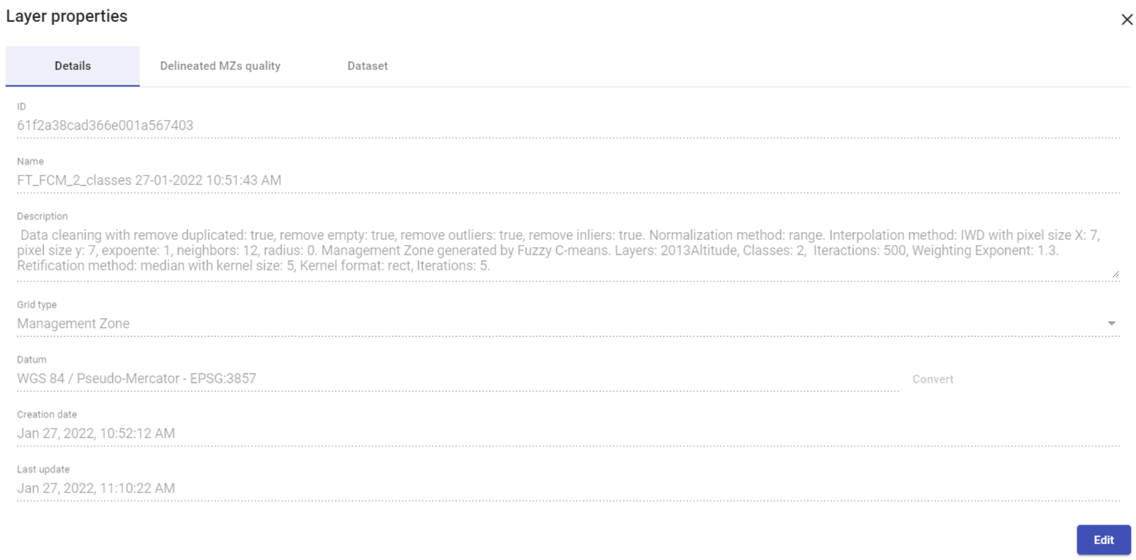


Figure 20 Layer properties of the resulting management zone

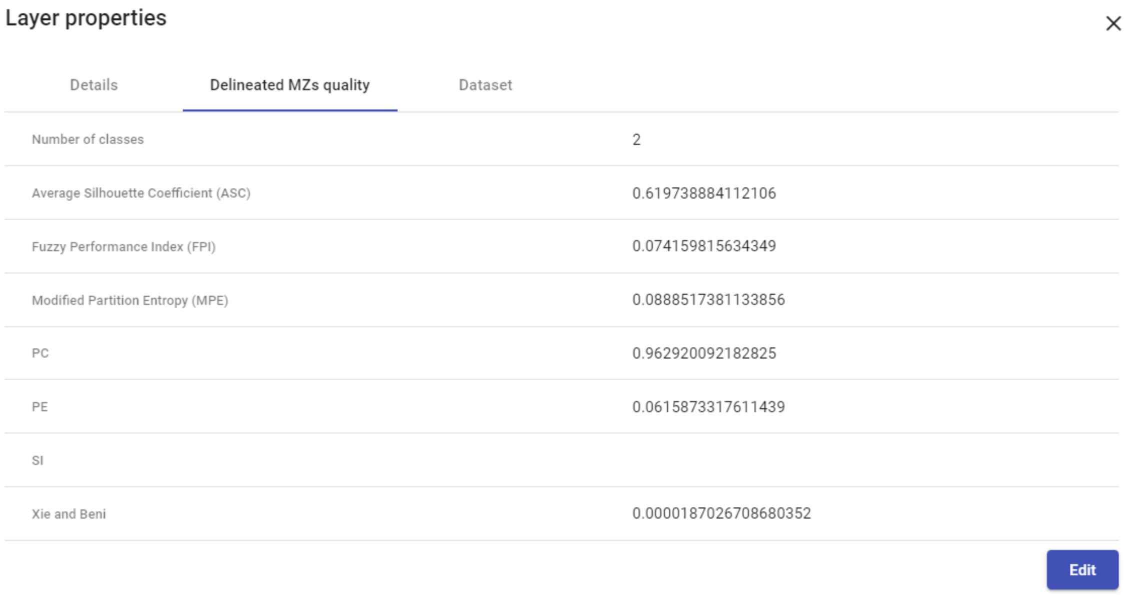
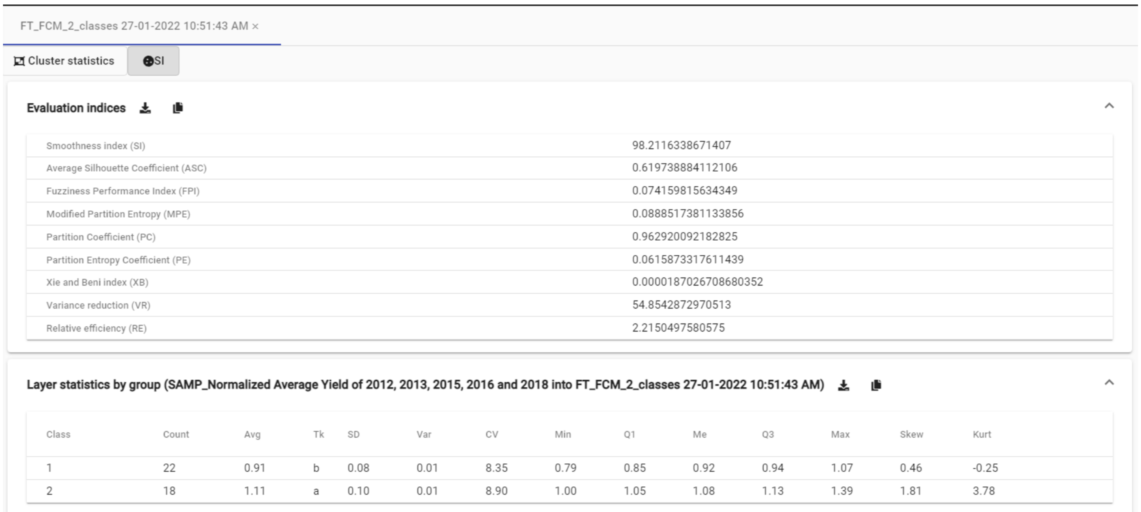


Figure 21 Management zone layer properties: delineated management zones quality

As a result of the provided data, ADB-MAP-FT for MZs defined the subdivision into two management classes, with one MZ in each class. Figure 19 presents the result using a blue color palette (default), where the darker is the class, the higher is the mean target variable, as seen in the legend. In Figure 20, it is possible to observe in the field description some details of the procedures chosen to delineate MZs. Data cleaning was applied to remove duplicates, empty data, outliers, and inliers for the resulting layer. The normalization method applied was range, with interpolation by IDW, with pixel size X and Y of 7, exponent 1 and 12 neighbors. The MZ was delineated using the Fuzzy C-Means algorithm, with 500 interactions and 1.3-Weighting Exponent, using only the Altitude (2013) layer. The ANOVA + ICVI procedure to define the quality of MZs determined that the ideal number of management classes is two, so only this division was sent to the rectification procedure. Rectification was applied using a median with a five-sized kernel, five interactions and rectangular shape.

The user can access the menu beside the layer name and choose the option Statistics for a complete view of the statistics. A screen (Fig. 22) will then be displayed that allows calculating several measures on the delineated MZs, being Smoothness Index (SI), Average Silhouette Coefficient (ASC), Fuzziness Performance Index (FPI), Modified Partition Entropy (MPE), Partition Coefficient (PC), Partition Entropy Coefficient (PE), Xie and Beni Index (XB), Variance Reduction (VR) and Relative Efficiency (RE). It is also possible to visualize the results and statistics for each MZ to the target variable, displaying the Class, with its sample point count (Count), average (Avg), Tukey test result (Tk), Standard Deviation (SD), Variance (VAR), Coefficient of Variation (CV), minimum value (Min), 1st quarter (Q1), Median (Me), 3rd quarter (Q3), Maximum value (Max), Skew and Kurtosis. The generated MZs can also be exported in CSV, TXT, and image formats.


FT_FCM_2_classes 27-01-2022 10:51:43 AM ×

Cluster statistics | SI

Evaluation indices

Smoothness index (SI)	98.2116338671407
Average Silhouette Coefficient (ASC)	0.619738884112106
Fuzziness Performance Index (FPI)	0.074159815634349
Modified Partition Entropy (MPE)	0.0888517381133856
Partition Coefficient (PC)	0.962920092182825
Partition Entropy Coefficient (PE)	0.0615873317611439
Xie and Beni index (XB)	0.0000187026708680352
Variance reduction (VR)	54.8542872970513
Relative efficiency (RE)	2.2150497580575

Layer statistics by group (SAMP_Normalized Average Yield of 2012, 2013, 2015, 2016 and 2018 into FT_FCM_2_classes 27-01-2022 10:51:43 AM)

Class	Count	Avg	Tk	SD	Var	CV	Min	Q1	Me	Q3	Max	Skew	Kurt
1	22	0.91	b	0.08	0.01	8.35	0.79	0.85	0.92	0.94	1.07	0.46	-0.25
2	18	1.11	a	0.10	0.01	8.90	1.00	1.05	1.08	1.13	1.39	1.81	3.78

Figure 22 General and per management zone statistics

The metrics presented in Figure 22 validate the choices made by ADB-MAP-FT for MZs. For example, the division into two management classes supported by Tukey test, and FPI and MPE (measures commonly adopted in conjunction with Fuzzy

C-means) indicate good quality in delineated MZs and their feasibility in-field application.

We highlight that the entire process was performed without any user's intervention, except by selecting data layers and contour. This proves that it is an easy process to be used with very little technical knowledge about the MZ delineation procedure. Furthermore, the possibility of configuring parameters and algorithms in each of the steps also allows more advanced users to perform fine tunings, taking advantage of the integrated procedure although it is not an obligation to perform each step individually, also streamlining the workflow for this audience.

4.5 CONCLUSIONS

This work presented the AgDataBox-Map-Fast Track (ADB-MAP-FT), a web application to automatically create thematic maps and management zones. This is an integral module of the AgDataBox platform that aims to facilitate the required process to carry out each task.

The choice of the algorithms sequence and their parameters in each process were based on a literature research, where the best algorithms/parameters were identified for most cases, or, in their absence, the most popular algorithms/parameters. Both generation of TMs and MZs can be performed by people with little knowledge of the area since, after properly selecting data, ADB-MAP-FT is in charge of carrying out all the procedures without user's intervention.

For people with advanced knowledge, or for those who simply want to test other configurations or algorithms besides those ones defined in the standard protocol, this is still possible since ADB-MAP-FT, both allows a series of configuration definitions, presented sequentially following the process steps in order to facilitate the user's understanding to generate TMs and delineate MZs. Once the new settings are selected, ADB-MAP-FT can carry out the entire procedure sequentially without the user's intervention.

As the resulting TM or MZ is an AgDataBox-Map layer, they can be readjusted entirely according to the user's interest, such as changing color schemes, subtitle texts, subtitle ranges, and can also be exported in CSV, TXT and image formats. The platform is free to use, it stores all project information in the cloud, allowing for later demands and, as it is a web platform, it removes the processing cost on the client-side, since the procedure is performed on the server-side, having as access restriction only a device with a web browser and an internet connection.

4.6 ACKNOWLEDGMENTS

The authors would like to thank the Western Paraná State University (UNIOESTE) and the Federal University of Technology of Paraná (UTFPR) for funding this project.

4.7 REFERENCES

ADAMCHUK, V.; HUMMEL, J.; MORGAN, M.; UPADHYAYA, S. On-the-go soil sensors for precision agriculture. **Comput Electron Agric,** v. 44, p.71–91, 2004.

AIKES JUNIOR, J.; SOUZA, E. G. DE; SOBJAK, R.; BAZZI, C. L. **Thematic Maps and Management Zones for Precision Agriculture: Systematic Literature Study, Protocols, and Practical Cases**. Poncã, Curitiba, 2021.

ALBORNOZ, E. M.; KEMERER, A. C.; GALARZA, R.; MASTAGLIA, N.; MELCHIORI, R.; MARTÍNEZ, C. E. Development and evaluation of an automatic software for management zone delineation. **Precis Agric** v. 19, p. 463–476, 2018.

AMIDAN, B.; FERRYMAN, T.; COOLEY, S. Data outlier detection using the Chebyshev theorem. **In IEEE Aerosp. Conf**, p. 3814-3819, 2005.

BAUDRON, F.; GILLER, K. E. Agriculture and nature: Trouble and strife?. **Biol Conserv** v. 170, p. 232–245, 2014.

BAZZI, C. L.; SOUZA, E. G.; SCHENATTO, K.; BETZEK, N. M.; GAVIOLI, A. A software for the delineation of crop management zones (SDUM). **Aust J Crop Sci,** v. 13, p. 26–34, 2019.

BAZZI, C. L.; SOUZA, E. G.; URIBE-OPAZO, M. A.; NÓBREGA, L. H. P.; ROCHA, D. M. Management zones definition using soil chemical and physical attributes in a soybean area. **Eng Agrícola**, v. 33, p. 952–964, 2013.

BERNARDI, A. C. DE C.; TUPY, O.; EDUARDA LIMA SANTOS, K.; MAZZUCO, G. G.; BETTIOL, G. M.; RABELLO, L. M.; INAMASU, R. Y. Mapping of yield, economic return, soil electrical conductivity, and management zones of irrigated corn for silage. **Pesq agropec bras,** v. 53, p. 1289–1298, 2018.

BETZEK, N. M.; SOUZA, E. G.; BAZZI, C. L.; SCHENATTO, K.; GAVIOLI, A.Rectification methods for optimization of management zones. **Comput Electron Agric**, v. 146, p. 1–11, 2018.

BETZEK, N. M.; SOUZA E. G.; BAZZI, C. L.; SCHENATTO, K.; GAVIOLI, A.; MAGALHÃES, P. S. G. G. Computational routines for the automatic selection of the best parameters used by interpolation methods to create thematic maps. **Comput Electron Agric**, v. 157, p. 49–62, 2019.

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

BIER, V. A.; SOUZA, E. G. Interpolation selection index for delineation of thematic maps. **Comput Electron Agric**, v. 136, p. 202-209, 2017.

BLACKMORE, S. The interpretation of trends from multiple yield maps. **Comput Electron Agric**, v. 26, p. 37–51, 2000.

BLACKMORE, S.; MOORE, M. Remedial correction of yield map data. **Precision agriculture,** v. 1, p. 53-66, 1999.

BOBRYK, C. W.; MYERS, D. B.; KITCHEN, N. R.; SHANAHAN, J. F.; SUDDUTH, K. A.; DRUMMOND, S. T.; GUNZENHAUSER, B.; GOMEZ RABOTEAUX, N. N. Validating a digital soil map with corn yield data for precision agriculture decision support. **Agron J**, v. 108, p. 957–965, 2016.

BORGES, L. F.; BAZZI, C. L.; SOUZA, E. G. D.; MAGALHÃES, P. S. G.; MICHELON, G. K. Web software to create thematic maps for precision agriculture. **Pesq agropec bras**, v. 55, p. 11 pp, 2020.

BOTTEGA, E. L.; QUEIROZ, D. M. DE, PINTO, F. DE A. DE C.; SOUZA, C. M. A. DE, VALENTE, D. S. M. Precision agriculture applied to soybean: Part I - Delineation of management zones. **Aust J Crop Sci**, v. 11, p. 573–579, 2017.

BOYDELL, B.; MCBRATNEY, A.B. Identifying potential within-field management zones from cotton-yield estimates. **Precis. Agric**, v. 3, p. 9–23, 2002.

BROCK, A.; BROUDER, S.M.; BLUMHOFF, G.; HOFMANN, B.S. Defining yield-based management zones for corn–soybean rotations. **Agron J**, v. 97, p. 1115-1128, 2005.

BUTTAFUOCO, G.; CASTRIGNANÒ, A.; COLECCHIA, A. S.; RICCA, N. Delineation of management zones using soil properties and a multivariate geostatistical approach. **Ital J Agron**, v. 5, p. 323–332, 2010.

CHANG, D.; ZHANG, J.; ZHU, L.; GE, S. H.; LI, P. Y.; LIU, G. S. Delineation of management zones using an active canopy sensor for a tobacco field. **Comput Electron Agric**, v. 109, p. 172–178, 2014.

COELHO, E. C.; SOUZA, E. G. D.; URIBE-OPAZO, M. A.; PINHEIRO NETO, R. Influência da densidade amostral e do tipo de interpolador na elaboração de mapas temáticos. **Acta Scientiarum. Agronomy**, v. 31, p. 165-174, 2009.

COHEN, J. A. Coefficient of agreement for nominal scales. **Educ Psychol Measur**, v. 20, p. 37–46, 1960.

CÓRDOBA, M.; BRUNO, C.; COSTA, J.; BALZARINI, M. Subfield management class delineation using cluster analysis from spatial principal components of soil variables. **Comput Electron Agric,** v. 97, p. 6–14, 2013.

CÓRDOBA, M. A.; BRUNO, C. I.; COSTA, J. L.; PERALTA, N. R.; BALZARINI, M. G. Protocol for Multivariate Homogeneous Zone Delineation in Precision Agriculture. **Biosyst Eng**, v. 143, p. 95–107, 2016.

DALL'AGNOL, R. W.; MICHELON, G. K.; BAZZI, C. L.; MAGALHÃES, P. S. G.; DE SOUZA, E. G.; BETZEK, N. M.; SOBJAK, R. Web applications for spatial analyses and thematic map generation. **Comput Electron Agric**, v. 172, 13 p.; 2020.

DOBERMANN, A.; PING, J. L.; ADAMCHUK, V. I.; SIMBAHAN, G. C.; FERGUSON, R. B. Classification of Crop Yield Variability in Irrigated Production Fields. **Agron J**, v. 95, p. 1105, 2003.

DOERGE, T.A. **Site-Specific Management Guidelines**. Potash & Phosphate Institute, Norcross, 2000.

DRAY, S.; SAÏD, S.; DÉBIAS, F. Spatial ordination of vegetation data using a generalization of Wartenberg's multivariate spatial correlation. **J Veg Sci**, v. 19, p. 45–56, 2008.

FLEMING, K. L.; WESTFALL, D. G.; WIENS, D. W.; BRODAHL, M. C. Evaluating farmer defined management zone maps for variable rate fertilizer application. **Precise Agric**, v. 2, n. 2, p. 201-215, 2000.

FRANZEN, D. W.; HOPKINS, D. H.; SWEENEY, M. D.; ULMER, M. K.; HALVORSON, A. D. Evaluation of Soil Survey Scale for Zone Development of Site-Specific Nitrogen Management. **Agron J**, v. 94, p. 381, 2002.

FRIDGEN, J. J.; KITCHEN, N. R.; SUDDUTH, K. A.; DRUMMOND, S. T.; WIEBOLD, W. J.; FRAISSE, C. W. Management Zone Analyst (MZA): Software for Subfield Management Zone Delineation. **Agron J**, v. 96, p. 100–108, 2004.

FRIDGEN, J.J.; KITCHEN, N.R.; SUDDUTH, K. A. Variability of soil and landscape attributes within sub-field management zones. In P. C. Roberts, et al. (Eds.), Precision agriculture. **Proceedings of the 5th International Conference of the ASA, CSSA, and SSSA**, Madison, WI, USA, 2000.

GAVIOLI, A.; SOUZA, E. G.; BAZZI, C. L.; GUEDES, L. P. C.; SCHENATTO, K. Optimization of management zone delineation by using spatial principal components. **Comput Electron Agric**, v. 127, p. 302–310, 2016

GNANADESIKAN, R.; KETTENRING, J. R.; TSAO, S. L. Weighting and selection of variables for cluster analysis. **J Classif**, v. 12, p. 113–136, 1995.

GONZALEZ, R.C.; WOODS, R. **Digital image processing**, third ed. Pearson Prentice Hall. New Jersey, 2008.

GOZDOWSKI, D.; STEPIEŃ, M.; SAMBORSKI, S.; DOBERS, E. S.; SZATYŁOWICZ, J.; CHORMAŃSKI, J. Determination of the Most Relevant Soil Properties for the Delineation of Management Zones in Production Fields. **Commun Soil Sci Plant Anal**, v. 45, p. 2289–2304, 2014.

HÖRBE, T.D.A.; AMADO, T.J.C.; FERREIRA, A.D.O.; ALBA, P.J. Optimization of corn plant population according to management zones in Southern Brazil. **Precis Agric**, v. 14, p. 450-465, 2013.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. **J Educ Psychol**, v. 6, p. 417–441, 1933

ISAAKS, E. H; SRIVASTAVA, R. M. **Applied geostatistics**. New York: Oxford University Press, 1989.

JAYNES, D. B.; COLVIN, T.S. AND KASPAR, T. C. Identifying potential soybean management zones from multi-year yield data. **Comput Electron Agric**, v. 46, p. 309-327, 2005.

KERRY, R.; OLIVER, M. A.; FROGBROOK, Z. L. Geostatistical Applications for Precision Agriculture. **In: Oliver MA (ed) Geostatistical Applications for Precision Agriculture**. Springer Netherlands, Dordrecht, p. 35–63, 2010.

KHOSLA, R.; ALLEY, M. M. Soil-specific nitrogen management on mid-Atlantic coastal plain soils. **Better Crops**, v. 83, p. 6-7, 1999.

KHOSLA, R.; FLEMING, K.; DELGADO, J. A.; SHAVER, T. M.; WESTFALL, D. G. Use of site-specific management zones to improve nitrogen management for precision agriculture. **Soil Water**, v. 57, p. 513–518, 2002.

KITCHEN, N.R.; SUDDUTH, K.A.; DRUMMOND, S.T.; FRIDGEN, J.J.; FRAISSE, C.W. Procedures for evaluating unsupervised classification to derive management zones. p. 330-345. **In P.C. Robert et al. (ed.) Proc. 6th Int. Conf. on Precision Agriculture**, Minneapolis, MN. 14-17 July 2002. ASA, CSSA, and SSSA, Madison, WI, 2002.

KITCHEN, N. R.; SUDDUTH, K. A.; MYERS, D. B.; DRUMMOND, S. T.; HONG, S. Y. Delineating productivity zones on claypan soil fields using apparent soil electrical conductivity. **Comput Electron Agric**, v. 46, p. 285–308, 2005.

KUANG, B.; MAHMOOD, H.S.; QURAISHI, M.Z.; HOOGMOED, W.B.; MOUAZEN, A.M.; VAN HENTEN, E.J. Sensing soil properties in the laboratory, in situ, and on-line: a review. **In Advances in agronomy**, v. 114, p. 155-223. Academic Press, 2012.

KYAW, T.; FERGUSON, R. B.; ADAMCHUK, V. I.; MARX, D. B.; TARKALSON, D. D.; MCCALLISTER, D. L. Delineating site-specific management zones for pH-induced iron chlorosis. **Precis Agric**, v. 9, p. 71–84, 2008.

LI, Y.; SHI, Z.; LI, F.; LI, H. Y. Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. **Comput Electron Agric**, v. 56, p. 174–186, 2007.

LI, Y.; SHI, Z.; WU, H. X.; LI, F.; LI, H. Y. Definition of management zones for enhancing cultivated land conservation using combined spatial data. **Environ Manage**, v. 52, p. 792–806, 2013.

MCBRATNEY, A. B.; MOORE, A. W. Application of fuzzy sets to climatic classification. **Agric For Meteorol**, v. 35, p. 165–185, 1985

MIAO, Y.; MULLA, D. J.; ROBERT, P. C. An integrated approach to site-specific management zone delineation. **Front Agric Sci Eng**, v. 5, p. 432–441, 2008.

MICHELON, G. K.; BAZZI, C. L.; UPADHYAYA, S.; DE SOUZA, E. G.; MAGALHÃES, P. S. G.; BORGES, L. F.; BETZEK, N. M. Software AgDataBox-Map to precision agriculture management. **SoftwareX**, 10, 100320, 7 pp, 2019.

MIEZA, M. S.; CRAVERO, W. R.; KOVAC, F. D.; BARGIANO, P. G. Delineation of site-specific management units for operational applications using the topographic position index in La Pampa, Argentina. **Comput Electron Agric**, v. 127, p. 158-167, 2016.

MINASNY, B.; MCBRATNEY, A. B. **FuzME version 3.0**. Australian Centre for Precision Agriculture, The University of Sydney, Australia, 2002..

MOLIN, J. P. Definição de Unidade de Manejo a Partir de Mapas de Produtividade. **Eng Agrícola**, v. 22, p. 83–92, 2002.

MORAL, F. J.; TERRÓN, J. M.; SILVA, J. R. M. DA. Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. **Soil Tillage Res**, v. 106, p. 335–343, 2010.

NAWAR, S.; CORSTANJE, R.; HALCRO, G.; MULLA, D.; MOUAZEN, A. M. Delineation of Soil Management Zones for Variable-Rate Fertilization: A Review. **In: Sparks, DL (ed) Advances in Agronomy**, v. 143, p.175–245, 2017.

NANNI, M.R.; POVH, F.P.; DAMATTÊ, J.A.M.; OLIVEIRA, R.B. de; CHICATI, M.L.; CEZAR, E. Optimum size in grid soil sampling for variable rate application in site-specific management. **Scientia Agricola**, v.68, p.386-392, 2011.

ONU, 2020. World population prospects 2020, **Department of Economic and Social Affairs.** World Population Prospects 2020.

PACCIORETTI, P.; CÓRDOBA, M.; BALZARINI, M. FastMapping: Software to create field maps and identify management zones in precision agriculture. **Comput Electron Agric**, v. 175, p.105556, 2020.

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

ROBERTS, D. F.; FERGUSON, R. B.; KITCHEN, N. R.; ADAMCHUK, V. I.; SHANAHAN, J. F. Relationships between soil-based management zones and canopy sensing for corn Nitrogen management. **Agron J** ,v. 104, p. 119–129, 2012.

ROBERTSON, M. J.; LYLE, G.; BOWDEN, J. W. Within-field variability of wheat yield and economic implications for spatially variable nutrient management. **F Crop Res**, v. 105, p. 211–220, 2008.

ROUSSEEUW, P. (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. **J Comput Appl Math**, v. 20, p. 53-65, 1987.

SCHENATTO, K.; SOUZA, E. G.; BAZZI, C. L.; BETZEK, N. M.; GAVIOLI, A.; BENEDUZZI, H. M. Use of the farmer's experience variable in the generation of management zones. **Semina: Ciências Agrárias**, v. 38, p. 2305-2321, 2017a.

SCHENATTO, K.; SOUZA, E. G.; BAZZI, C. L.; BETZEK, N. M.; GAVIOLI, A. Two approach comparison to define crop management zones (MZs). **African J Agric Res**, v. 11, p. 3680–3689, 2016.

SCHENATTO, K.; SOUZA, E. G.; BAZZI, C. L.; GAVIOLI, A.; BETZEK, N. M.; BENEDUZZI, H. M. Normalization of data for delineating management zones. **Comput Electron Agric**, v. 143, p. 238–248, 2017b.

SOBJAK, R.; SOUZA, E. G. DE, BAZZI, C. L.; URIBE-OPAZO, M. A.; BETZEK, N. M. Redundant variables and the quality of management zones. **Eng Agrícola**, v. 36, p. 78–93, 2016.

SOBJAK, R. Incorporation of computational modules as microservices in the AgDataBox platform and development of agdatabox-map application. Tese (Doutorado em Doutorado em Engenharia Agrícola) - Universidade Estadual do Oeste do Paraná. 256 p., 2021.

SOUZA, E. G. DE, SCHENATTO, K.; BAZZI, C. L. Creating thematic maps and management zones for agriculture fields. **In: 14th International Conference on Precision Agriculture**. Montreal, p. 1–17, 2018.

SOUZA, E. G.; C. L. BAZZI, R. KHOSLA, M. A. URIBE-OPAZO, AND R. M. REICH. Interpolation type and data computation of crop yield maps is important for precision crop production. **Journal of Plant Nutrition**, v. 39, p. 531–538, 2016.

SUDDUTH, K.A.; DRUMMOND, S.T. Yield editor. **Agron J**, v. 99, p. 1471-1482, 2007.

SUSZEK, G.; SOUZA, E. G. DE, URIBE-OPAZO, M. A.; NOBREGA, L. H. P. Determination of management zones from normalized and standardized equivalent produtivity maps in the soybean culture. **Eng Agrícola**, v. 31, p. 895–905, 2011.

TAYLOR, J. A.; MCBRATNEY, A. B.; WHELAN, B. M. Establishing Management Classes for Broadacre Agricultural Production. **Agron J**, v. 99, p. 1366, 2007.

TAYLOR, J. C.; WOOD, G. A.; EARL, R.; GODWIN, R. J. Soil Factors and their Influence on Within-field Crop Variability, Part II: Spatial Analysis and Determination of Management Zones. **Biosyst Eng**, v. 4, p. 441-453, 2003.

VEGA, A.; CÓRDOBA, M.; CASTRO-FRANCO, M.; BALZARINI, M. Protocol for automating error removal from yield maps. **Precis Agric**, v. 20, p. 1030-1044, 2019.

VELANDIA, M.; REJESUS, R. M.; BRONSON, K.; SEGARRA, E. Economics of management zone delineation in cotton precision agriculture. **J Cotton Sci**, v. 12, p. 210–227, 2008.

VITHARANA, U. W. A.; MEIRVENNE, M. VAN, SIMPSON, D.; COCKX, L.; HOFMAN, G. Agronomic consequences of potential management zones delineated on the basis of EM38DD measurements. **Near Surf Geophys**, v.6, p. 289–296, 2008.

WHETTON, R. L.; WAINE, T. W.; MOUAZEN, A. M. Evaluating management zone maps for variable rate fungicide application and selective harvest. **Comput Electron Agric**, v. 153, p. 202–212, 2018.

WOLLENHAUPT, N. C.; WOLKOWSKI, R. P.; CLAYTON, M. K. Mapping soil test phosphorus and potassium for variable-rate fertilizer application. **J Prod Agric** v. 7, p. 441–448, 1994.

XIANG, L.; YU-CHUN, P.; ZHONG-QIANG, G.; CHUN-JIANG, Z. Delineation and Scale Effect of Precision Agriculture Management Zones Using Yield Monitor Data Over Four Years. **Agric Sci China**, v. 6, p. 180–188, 2007.

YANG, G.; LIU, J.; ZHAO, C.; LI, Z.; HUANG, Y.; YU, H.; XU, B.; YANG, X.; ZHU, D.; ZHANG, X.; ZHANG, R. Unmanned aerial vehicle remote sensing for field-based crop phenotyping: current status and perspectives. **Front Plant Sci**, v. 8, p. 1111, 2017.

ZHANG, X.; SHI, L.; JIA, X.; SEIELSTAD, G.; HELGASON, C. Zone mapping application for precision-farming: A decision support tool for variable rate application. **Precis Agric**, v. 11, p. 103–114, 2010.

5 PAPER 2 – SOFTWARES TO DELINEATE MANAGEMENT ZONES: A COMPARATIVE CASE STUDY

ABSTRACT: Management zones (MZ) delineation is not a trivial task although it has been very important for precision agriculture. It involves several steps and knowledge of several areas. One way to mitigate the difficulty in its construction is specialized software. However, several software are listed in literature, each one presents different characteristics and resources, but they often perform only part of the process, so the user has to choose multiple software and the best workflow for its use. This work compares the most used software as well as the state-of-the-art software for MZs delineation. In addition, the softwares are compared for their technical characteristics and MZs delineation results based on a case study using two commercial areas.

KEYWORDS: AgDataBox-Map-FT, MZA, FuzME, FastMapping, Comparison.

5.1 INTRODUCTION

A management zone (MZ) is a subregion of a field with similar soil and topographic characteristics, consequently, it requires the same amounts of inputs, thus, it allows an optimized uniform management in this sub-area (Bobryk et al., 2016; Moral et al., 2010; Moshia et al., 2014; Zeraatpisheh et al., 2022). MZs can be used in several situations such as smart sampling, where one composite sampling is obtained per zone to evaluate the field variability. This approach will likely reduce laboratory costs while maintains reliability level (Ferguson and Hergert, 2009; Mallarino and Wittry, 2004). In addition, smart sampling has improved nutrient efficiency use while has kept or increased yield and potentially reduced the nutrient overloading into the environment (Moshia et al., 2014; Khosla et al., 2002). This approach can enable precision agriculture (PA) for more producers because the homogeneous rate in each sub-area allows some conventional agricultural machines. It has already presented cost decrease, yield increase, and improvement of product quality parameters (Bernardi et al., 2018; Cid-Garcia; Ibarra-Rojas, 2019; Kyaw et al., 2008; Li et al. 2013; Roberts et al., 2012; Robertson et al., 2008; Schwalbert et al., 2018; Velandia et al., 2008; Vitharana et al., 2008; Whetton et al., 2018).

However, it is not an usual task to develop MZs quality since it is a process that involves several tasks from different knowledge areas. Also, a given task can present different results since several algorithms and different parameters combinations can be used.

Some of the presented difficulties can be reduced by using specialized software. There are many software for PA, but few are focused on MZs delineation. Golden Software Surfer, ESRI ArcGIS, and R software package are commonly used for

MZs delineation, but they do not have all desired functionalities (Aikes Junior et al., 2021). When they have all the required functionalities, they are not user-friendly (Albornoz et al., 2018). Sometimes, it delineates zones that cannot be validated since it does not explicitly indicate the data analytics behind the resulting variability maps (Paccioretti et al., 2020). Another determining factor to hinder access to software is because most of them have only paid commercial licenses, which discourages its adoption by non-specialized people since they may not realize the advantages of its use at first.

Among the specific software for MZs delineation, the following were well-known (organized by release date): (i) FuzME (Minasny; McBratney, 2002), (ii) Management Zone Analyst (MZA) (Fridgen et al., 2004), (iii) ZoneMAP (Zhang et al., 2010), (iv) Software for Management Zones Definition (SDUM) (Bazzi et al., 2013; Bazzi et al., 2019), (v) automatic software to delineate MZs proposed by Albornoz et al. (Albornoz et al., 2018), (vi) FastMapping (Paccioretti et al., 2020) and (vii) AgDataBox (ADB) (Bazzi et al., 2019; Borges et al., 2020; Dall'Agnol et al., 2020; Michelon et al., 2019), which have a specific module for automatic delineation of management zones AgDataBox-Map-Fast-Track (ADB-FT) (Paper 2 - is not published yet).

Each mentioned software has different functionalities, and brings specific advantages/disadvantages. Thus, it is essential to choose the appropriate MZs delineation protocol and suitable software. So, this trial aimed at carrying out, based on a case study, a comparative analysis of some software used to delineate MZs.

5.2 MATERIAL AND METHODS

5.2.1 Data

The study was carried out from data collected in two agricultural fields (Fig. 1). Field A is a 15.5-ha area in Céu Azul, Paraná, Southern Brazil, in latitude -25.1092, longitude -53.8319 (WGS84). The soil of the area was classified as Rhodic Ferralsol. The field has a 662-m average altitude and the slope varies from 0.0 to 10.1%, with a mean of 1.2%, using no-till farming, and a succession of different cultures have been cropped for more than ten years. Soybeans and corn were grown during the summer harvest or off-season period, and wheat or oats were used as a cover crop during the winter. Forty points (2.67 points ha^{-1}) in an irregular sampling grid were located on an imaginary centerline among the contour lines of each field. As the field A has a certain degree of declivity and has contour lines, it was decided to use an irregular sampling grid to define the points in places that did not coincide with the curves. This is due to the possibility of their influence on productivity. Field B is 13.0-ha area in Centralia, Missouri, The United States of America, in latitude 39.2346, longitude -92.1469 (WGS84). The soil of the area was classified as a claypan. The field has a 262-m average altitude, a field slope of 1.8% and it was managed with no-tillage system, mostly cropped in a corn-soybean rotation. The exception was two years of continuous soybean (1998 and 1999). The soil showed a clay content higher than 500 g kg^{-1} in the argillic horizon that comprises smectitic (high shrink-swell) clay minerals. Hydrolocally, water flow is but preferential through cracks after profile drying (i.e., late summer and early fall) (Jamison et al., 1968, Kitchen et al. 2005).

The sampling density was greater than 2.5 points ha^{-1}, following the recommendation of Nanni et al. (2011) to enable the detection of spatial dependence among samples. Only stable attributes (Doerge, 2000) were used on MZs delineation (Table 1).

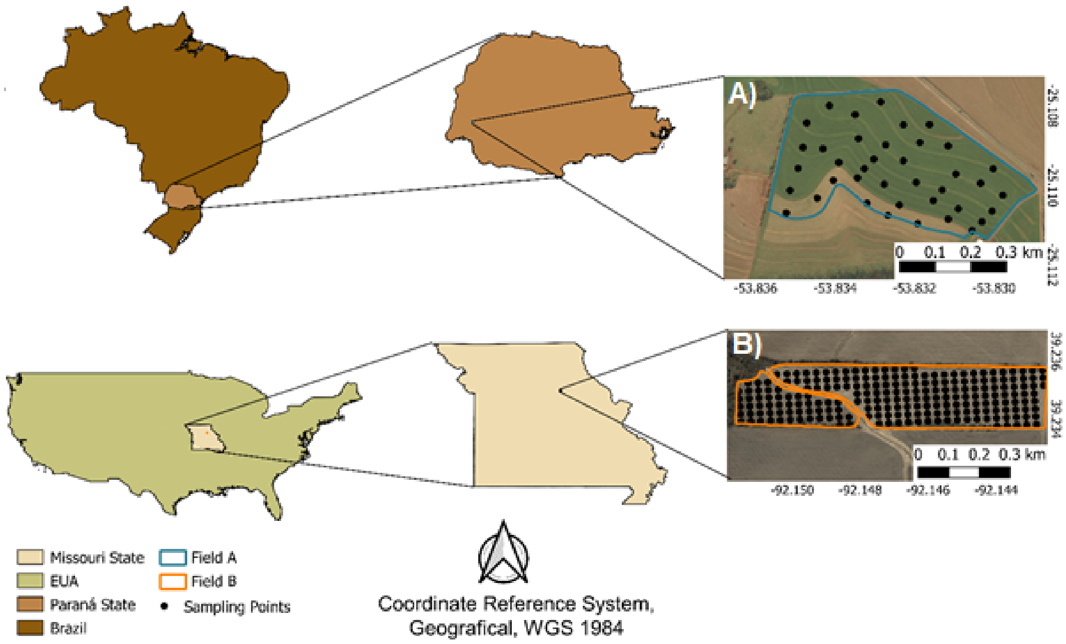


Figure 1 Location of the experimental fields

5.2.2 Softwares

- FuzME: software provided by the Precision Agriculture Laboratory (PA Lab) of the Australian Centre for Precision Agriculture (ACPA), University of Sydney, Australia. It is available for Microsoft Windows 95/NT or superior, and its most current version is 3.5c. The used algorithms are the fuzzy c-means (FCM) (with a few variants), and the outputs are all in text files. The software presents a simple tab-organized user interface.

Table 1 Variables collected in each experimental field to delineate management zones.

Variables (Attribute)	Field A		Field B	
	Year	Methodology/Instrument	Year	Methodology/Instrument
Elevation	2013	Topcon GPT-7505	Not available	Ashtech Z-12 RTK
EM38			Not available	Geonics EM38
VSH				Veris 3100
VDP				
DualA				Dualem 2S
DualC				
SPR 0-10	2013, 2015	Electronic penetrometer PenetroLOG		
SPR 10-20				
SPR 20-30				
Corn Yield			1997, 2000, 2002	Gleaner R42 combine harvester with AgLeader yield monitor
Soybean Yield	2012, 2013, 2015, 2016, 2018	Locations obtained with GNSS receiver Juno SB	1996, 1998 1999, 2001, 2003, 2004	

EM38: vertical apparent profile soil electrical conductivity (ECa) using EM38 or EM38-MK2 (Geonics Limited, Mississauga, ON, Canada); VSH: "shallow" ECa using Veris 3100 (Veris Technologies, Salina, KS, USA); VDP: "deep" ECa using Veris 3100; DualA: "deep" ECa using Dualem 2S (Dualem, Milton, ON, Canada); DualC: "shallow ECa using Dualem 2S; SPR 0-10: Soil penetration resistance from 0 to 10 cm using Electronic penetrometer PenetroLOG (Falker, Porto Alegre, Brazil); SPR 10-20: SPR from 10 to 20 cm; SPR 20-30: SPR from 20 to 30 cm; GNSS receiver Juno SB (Trimble Navigation Limited, Westminster, CO, USA), total station Topcon GPT-7505 (Topcon Corporation, Tokyo, Japan), 6-row Case combine harvester model Case IH 1660 (Case Corporation, Racine, WI, USA) equipped with an AgLeader (AgLeader Technology, Ames, IA, USA) yield-monitoring system.

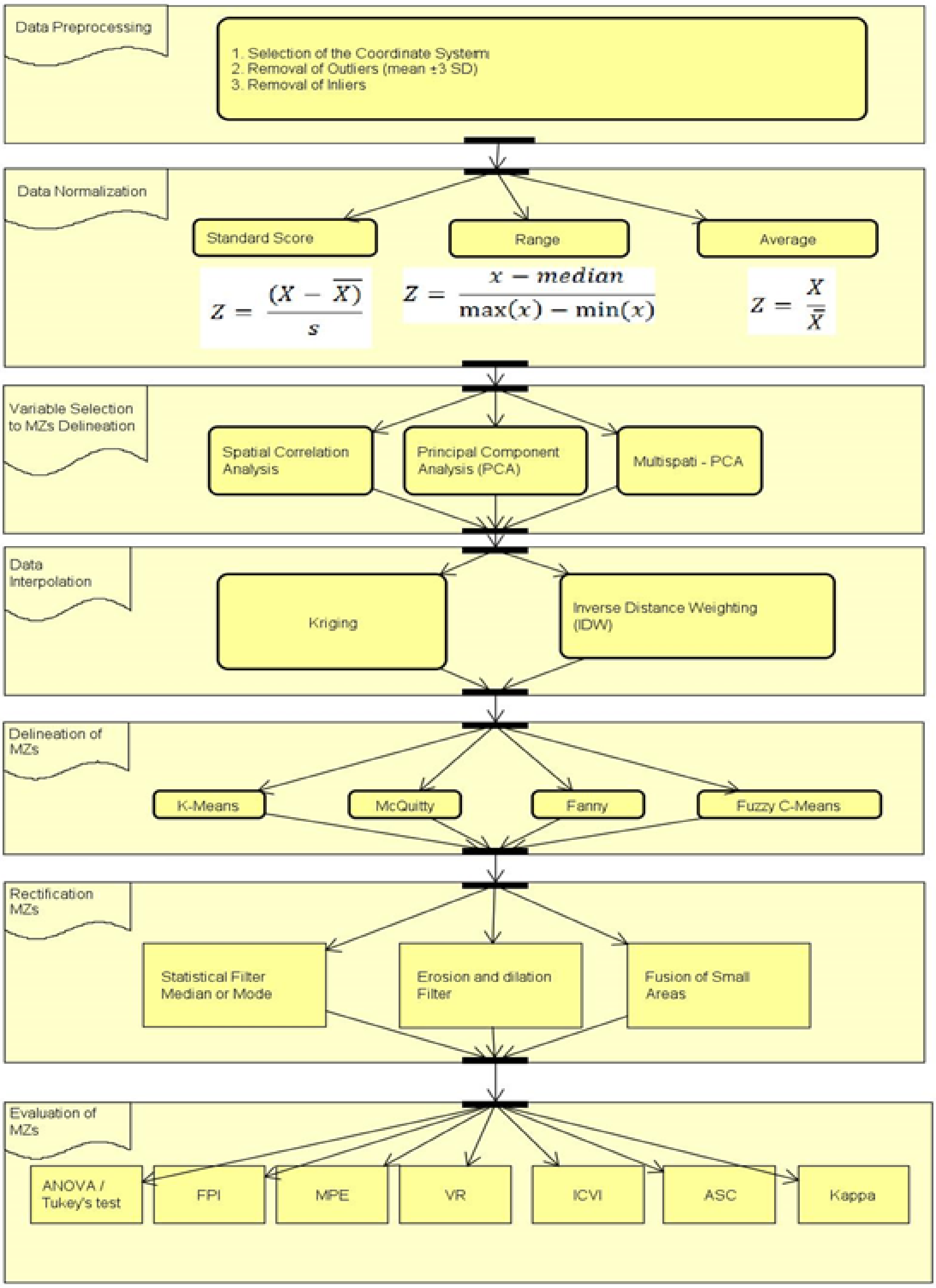
- MZA: it is the most used among the specific software to delineate MZs (Aikes Junior et al., 2021). It is made available by the Agricultural Research Service (ARS) of the United States Department of Agriculture (USDA), USA. It is available for Microsoft Windows 95/NT or superior, and its most current version is 1.0.1. It implements FCM algorithm. It also presents a simplified graphical interface and, to perform data clustering, you must follow the instructions in a sequence of four menus that present the definition of the parameters step by step.

- ZoneMap: it was unavailable when developing this paper, due to financial reasons, according to its developers. Consequently, it could not be evaluated.
- SDUM: software provided by the Paraná Precision Agriculture Team, from the western region of Paraná, Brazil. It is available for Microsoft Windows XP platform or superior, and the current version is 1.0. Entries are in the text file and have a user-friendly data importer. The outputs can be given in the text, images, PDF, and KLM (Google Earth) formats. It offers descriptive statistical analysis, spatial correlation analysis, and interpolation tools. MZs can be delineated by empirical methods (data normalization by means and standard deviation) and clustering (k-means and FCM). SDUM will not be used in the comparison as AgDataBox has replaced it.
- The Albornoz's automatic software to delineate MZs (Albornoz et al., 2018): software provided (in test version) by the Faculty of Engineering and Water Sciences (Facultad de Ingeniería y Ciencias Hídricas) of the National University of the Coast (Universidad Nacional del Litoral), Argentina. According to the authors, there are desktop and web versions of the software, but only web version is available for public tests. The delineation algorithm is FCM, and outputs are in the ESRI shapefile with a user friendly interface.
- FastMapping: it is a web application, provided by a partnership between the Faculty of Agricultural Sciences (Facultad de Ciencias Agropecuarias) of the National University of Córdoba (Universidad Nacional de Córdoba) and the Unit of Phytopathology and Agricultural Modeling (Unidad de Fitopatología y Modelización Agrícola), Argentina. It offers tools for cleaning, interpolation, statistical analysis of spatial data, MZs delineation, MZs evaluation and data exportation with a user-friendly graphical interface. Entries are in the text file and have a user-friendly data importer. The delineation algorithm is FCM and outputs are in text and HTML formats.
- AgDataBox (ADB): is a joint project coordinated by the Western Paraná State University (Unioeste) and the Federal University of Technology - Paraná (UTFPR) with the cooperation of the Colorado State University (CSU), The United States Agricultural Research

Service (USDA) in Columbia, the University of California Davis (UC Davis), the University of São Paulo (ESALQ/USP), and the Brazilian Agricultural Research Corporation (Embrapa). It is a web Platform with an open Application Programming Interface (API) that enables the interoperability of several applications. Five applications are under development: (1) ADB-Mobile, (2) ADB-Map, (3) ADB-Admin, and (4) ADB-IoT. ADB-Map application works with spatial data aiming to create thematic maps and management zones, offering data importing/exporting, data analysis, filtering data normalization, data interpolation and creation of thematic maps, delineation and evaluation of MZs, encompassing variable selection methods, empirical and data clustering methods, and evaluation statistics, management zone rectification methods and application map creation and exporting. ADB-MAP application offers a module called ADB-Map Fast-Track (ADB-MAP-FT), which allows TMs creation and MZs delineation automatically by a web-friendly interface platform. Entries are in the text file and have a user-friendly data importer and outputs are in image text formats. The default settings of all evaluated software were used.

5.2.3 Protocol to Delineate MZ

The search for an ideal protocol for MZs delineation has been still ongoing. So, there is no well-defined delineation protocol. In Aikes Junior et al. (2021), several protocols were presented, which despite having similar tasks in general, there are divergences or gaps among them. Among them all, the MZ protocol proposed by Souza et al. (2018) (Fig. 2) was considered more complete, and it divided the process into (1) variables acquisition of, (2) outliers and inliers remotion, (3) data normalization, (4) input variables selection, (5) data interpolation, (6) MZs delineation (7) MZs rectification, and (8) evaluation of delineated MZs.



ANOVA: analysis of variance, SD: standard deviation, MZ: Management Zone, SD: standard deviation, FPI: Fuzziness Performance Index), MPE: Modified Partition Entropy, VR: variance reduction, ICVI: improved cluster validation index, ASC: average silhouette coefficient.

Figure 2 The protocol of management zones delineation, according to Souza et al. (2018).

The data entry in all software has been performed by importing files in text format, using a graphical interface. ADB-MAP-FT, MZA, and FastMapping have assistants selecting which attributes will be imported, while for FuzME, the file must be previously treated only with the data to be used.

Only ADB-MAP-FT performs variable selection. SCM was used from the three available methods (Spatial correlation matrix – SCM, Principal component analysis – PCA, and MULTISPATI-PCA) (Gavioli et al., 2016). It uses Moran's bivariate spatial autocorrelation statistic to build a spatial correlation matrix. The procedure consists of (1) eliminating variables with no significant spatial autocorrelation at 5% significance, (2) removing variables that were not correlated with the target variable, (3) decreasing ordination of the remaining variables, considering the degree of correlation with the target variable, and (4) eliminating variables that are correlated with each other, with preference to remove those variables with lower correlation with the target variable. FuzME and FastMapping present the possibility of applying PCA to reduce dimensionality of the input variables, but not their use to select variables. Thus, the variables selected by ADB-MAP-FT will be used for all software as an input variable, for field A the Elevation variable and for field B the VSH variable.

MZA and FuzME also do not perform outliers and inliers removal or data normalization and interpolation. Thus, a file with these tasks performed in ADB-MAP-FT was offered as input to these software. The default values of ADB-MAP-FT parameters were used as following: (1) the values outside the mean ± three standard deviations (SD) were considered outliers (Córdoba et al., 2016; Vega et al., 2019) and removed, (2) to remove inliers, the Local Moran’s index of spatial autocorrelation (LI) (Anselin, 1995) was used (Córdoba et al., 2016; Vega et al., 2019), (3) normalization was performed by the range method (Schenatto et al., 2017), (4) interpolation was performed with the inverse distance weighted (IDW) interpolator (Betzek et al., 2019; Isaaks; Srivastava, 1989; Kerry et al., 2010), and (5) the default grid is 1/100 of the largest dimension in the North-South and East-West directions. IDW parameters were defined using the method proposed by Betzek et al. (2019). ADB-MAP-FT performs all these procedures automatically, without user intervention.

FastMapping adopts, when selected univariate data, the possibility of the user to define upper and lower threshold limits to eliminate observations that fall outside the general distribution of data set and automatically removes values below or equal to zero by default. Data points are also removed for a distance of 20 meters from field edges to eliminate edge effects. Outliers are removed when outside the mean ± 3 SD. Inliers are identified using LI and Moran scatterplot. The interpolation is performed by Kriging (Webster; Oliver, 2007), and the root mean square error (RMSE, obtained by a k-fold cross-validation process with k = 10) is used to select the best data variogram automatically. The neighborhood number is defined as a minimum of 7 and a maximum of 25 of the nearest data to the target.

All software employs as default FCM algorithm, fuzzy exponent to 1.3, and the Euclidean distance. MZA and FuzME determine the interaction parameter at 300, while ADB-MAP-FT at 500 and FastMapping at 1000. ADB-MAP-FT and FuzME delineate considering from 2 to 5 clusters and MZA and FastMapping from 2 to 6 clusters.

The best option was considered the management zones that: 1) significant difference in productivity, among management zones, verified by means of ANOVA; 2) showed greater reduction in the coefficient of variation (VR%); 3) lower FPI and MPE.

5.2.4 Evaluation of the management zones quality

The delineated MZs were evaluated using quality indices described in the Appendix A. MZA, FuzME, and FastMapping have always presented all the MZs grouping, allowing the user to choose the most appropriate subdivision. So, in order to help the user, MZA offers Normalized Classification Entropy (NCE) and Fuzziness Performance Index (FPI) as indices, FuzME presents FPI and Modified Partition Entropy (MPE), FastMapping presents Partition Coefficient (PC), FPI, NCE, Xie Beni index (XB) and ANOVA, in addition to the graphical display. ADB-MAP-FT, unlike the others, does not display all the subdivisions delineated, but only those ones that showed significant statistical differences among the management classes, that is, it uses a combination of ANOVA and Improved Cluster Validation Index (ICVI) to present to the user only the subdivisions that are statistically viable. Results can be evaluated graphically or by FPI, MPE, XB, ICVI, Variance Reduction (VR), Smoothness Index (SI), Average Silhouette Coefficient (ASC), Partition Entropy (PC), Partition Entropy Coefficient (PE), Relative efficiency (RE) and ANOVA (Tukey range test).

Another characteristic restricted to ADB-MAP-FT, considering the tested software, is the fact that it automatically applies MZs rectification to remove isolated pixels, small regions, or even a transition border among very irregular zones, that make difficult or even impossible to operate in the field. The rectification methods available are median, opening, closure, opening/closure. In addition, it is possible to choose the kernel format (square, circle, or cross) and the kernel size (from 3x3 to 11x11) (Sobjak, 2021). Rectification is performed, by default, employing the best method indicated by Betzek et al. (2018), which consists of the median filter, with a square mask format with a kernel size of 5 x 5 pixels and five interactions.

5.3 RESULTS AND DISCUSSIONS

Although all software programs have the same objective, there are significant similarities and differences (Table 2). MZA tries to make easier MZ delineation by presenting a graphical interface in steps. Initially (start window), the user must provide the input file in text format. In the same window, one or more variables must be chosen to be used. The following window, Explore Data, allows descriptive data statistics to be computed and saved in a text file. The third window, Delineate Zones, presents the options to performe the classification with FCM, location and name of the output data file. The last window, Post Classification Analysis, presents two graphs of the performance indices (NCE and FPI) as a function of the number of zones. The authors consider this last window to be one of the most critical differentials of MZA because it helps to choose the ideal number of zones, avoiding subjectivity.

Table 2 Features of specific software for management zones delineation (MZs) with default parameters.

Software / Feature	FuzME	MZA	FastMapping	ADB-MAP-FT
Input data graphical visualization				x
Input data description tools		x	x	x
Data selection for MZ				automatic
Cleaning data (outlier, inliers, null)			automatic	automatic
Data normalization			automatic	automatic
Data interpolation			automatic	automatic
MZ delineation	x	x	x	x
MZ evaluation	x	x	x	automatic
MZ rectification				automatic
Map creation			automatic	automatic
Save project				x
Results export type	Text	Text	Text/HTML with images	Text/Image
License	Free	Free	Free	Free
Operational system	Windows	Windows	Web	Web
Intuitive interface	Simple	Simple	Modern	Modern

HyperText Markup Language (HTML); Management Zone (MZ).

FuzME features a simplified graphical interface consisting of three toolbars in a precise sequence of steps to delineate MZs. The first one presents the options to select the input file with the respective variables, output files, internal control files, and analysis title. The second presents the options to create clusters, such as distance metrics and fuzzy exponents. Finally, the third one presents the options to allow resampling using the bootstrap and Jackknife methods. It is interesting to observe that FuzME requires that the input file contains a sample identification column and that the file contains only variables that will be used in the delineation process. All other software allows the choice of which variables from the input file will be used to delineate.

Despite the attractive graphical interface of FuzME and MZA, as well as the definition of some standardized parameters, it is impossible to: (1) visualize the delineated MZs, (2) perform interpolations, (3) adjust the sample size, (4) visualize the behavior of the input variables, (5) calculate statistics of MZ quality, (6) export the results graphically, (7) Perform data pre-processing such as cleaning (removal of outliers, inliers, null values) and normalization, (8) choice of the best (or set of best) input variables for MZs delineation concerning the target variable, and (9) the smoothing of fragmented information. All these tasks must be performed in external software. Another limiting factor is the requirement to run on computers using a specific operating system (Microsoft Windows environment), considering the dissemination of ubiquitous computing nowadays.

Another problematic element concerns choosing the ideal number of clusters. Measures such as NCE and FPI, on MZA, and FPI e MPE, on FuzME cannot necessarily agree on the ideal number of clusters, returning subjectivity to the analyst since the software does not indicate which is preferred over the other. It is worth remembering that the ideal number of these measures may still not be following the restrictions of field mechanization, considered purely mathematical analyses of the generated clusters.

FastMapping presents a modern and intuitive graphical interface, and organizes the MZ delineation process in stages. The first one (after Home), Dataset, allows uploading the file containing the input data and optionally the file with edges of the field. This tab allows choosing variables to be used and a tabular view of data. The remaining steps vary with the chosen variables. If the input data is multivariate, the multivariate analysis tab appears, which is subdivided into (1) Parameters for KM-sPC classification, to define the parameters of PCA and FCM, (2) Classification results, to present the results, (3) Cluster Plot to visualize the maps and (4) Validation to verify the statistical tests. When univariate (like it was used in this work) it subdivided into (1)

Depuration for data cleaning, (2) Prediction to select the parameters to choose the best interpolator by kriging, (3) Results with several subdivisions to visualize the kriging models and the result of data cleaning, (4) Cluster with several subdivisions to visualize and validate the clusters data, with their indexes and download the results, as well as statistical validation. Finally, regardless of whether it is univariate or multivariate, the Report tab allows downloading the results of all stages in HTML (HyperText Markup Language) format.

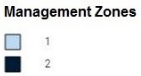
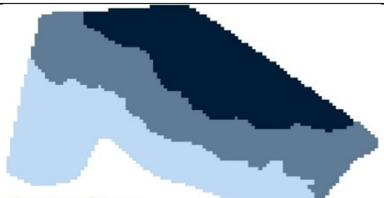
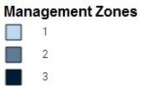
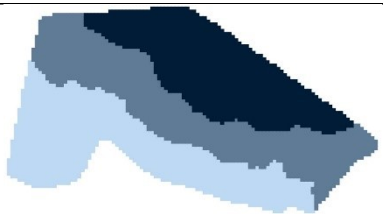
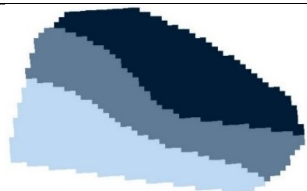
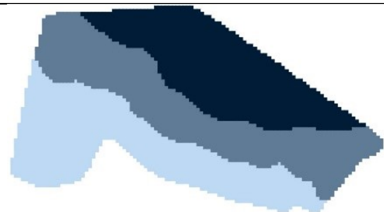
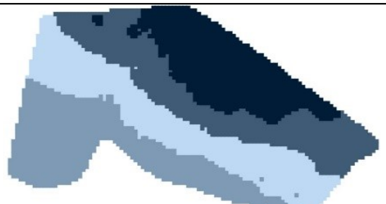
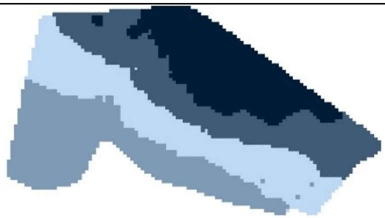
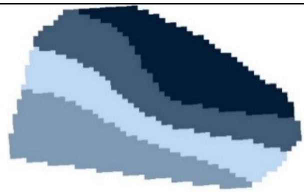
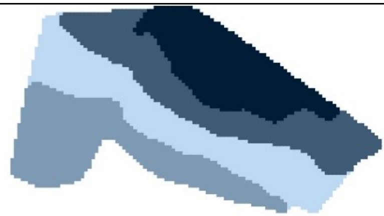
Despite the good graphical interface, FastMapping performs the processing (or at least the data request) on the server one step at a time, that is, each time a step is selected (viewing the map with two clusters, for example), the processing of the respective step takes place. This leads to a constant delay each time you navigate among tabs. This is not a major issue, but when combined with the fact that (during testing on several different days for this work) the user is constantly disconnected from the server and has to start the procedure all over again, as there are no options to save the project to continue later, ends up negatively impacting the user's experience. Furthermore, the lack of integrated tools to rectify MZs and select variables (according to the website still under development) requires external software for a complete process. Since it is a web platform, any device can access it with a web browser and internet. In addition, the processing is performed on the server, taking the load off the user's device. This can be considered positive because MZ delineation tasks usually depend on high processing loads. Unfortunately, despite more indices than MZA and FuzME, it is still up to the user to interpret the results and choose the best subdivision of the field.

ADB-MAP-FT is an ADB-MAP module for the automatic MZ delineation. Just like FastMapping is a web application with a modern and intuitive graphical interface. The user must create a project for its use and then use the data import wizard. He can perform visualizations and pre-processing with the imported data if desired. For MZ delineation, the user chooses the option of a new management zone, and among the options, he will choose Fast Track. A single window is presented, in which the user will select all the input variables, the area boundaries, and the target variable (usually yield). All the process of all the steps of the protocol chosen in this work is then carried out automatically, without user's intervention. In the end, new layers are added with MZs that meet the statistical criteria.

A point to highlight is that the parameters default values, used in ADB-MAP-FT, are a result from previous research to choose the best value or, in the absence of this, the most used value in the literature (Paper 2 - not published yet). This favors the use by inexperienced users or those who do not know all the areas necessary for a good MZ delineation, offering, at least, a good starting point for fine tunings. Another characteristic that makes it different from the others is that the user can delineate MZs within the context of projects, and the user can have several projects if desired. Projects and all their data and performed processing are saved in the cloud. This allows the user to revisit all their data, start the procedure and continue where they left off, or even be able to compare directly in the software, both graphically and using the indices offered, MZs delineated with different combinations of parameters/algorithms. As it is a web application, like FastMapping, it can be accessed from any device with a browser and internet, and data processing is performed on the server, to remove the processing load from the user's machine. ADB-MAP-FT was the only one that presented all the necessary tools for the complete execution of the protocol without any requirement for an external software.

Table 3 presents the delineated MZs using software MZA, FuzME, FastMapping, and ADB-MAP-FT for 2 to 4 classes for each field. Since ADB-Map-FT selects automatically the best number of classes, this functionality was disabled to permit it to show all MZs. A remarkable similarity can be visually perceived in MZs delineated by MZA, FuzME, and ADB-MAP-FT for both fields, confirmed by Kappa and Global Accuracy (GA) index (MZA is always used as a reference since it is the most used software). Despite being very similar, in MZA and FuzME, we can see isolated pixels in MZ 2 and MZ 4 and irregular transitions among zones in all cases of field A, making it difficult for the operational implementation. Such problems were solved by the rectification step in the ADB-MAP-FT, making its use more feasible, especially by machinery without an automated variable rate. In the case of FastMapping, its functionality to remove pixels 20 meters from the edges and, in the default configuration, making an interpolation with a lower resolution than ADB-MAP-FT made its result visually different (there is a option to upload a contour file to delimit the area limits, however, despite the successful upload and visualization of the area limits data, whenever this option was used, FastMapping presented errors when performing the clustering task). Still, the division of zones in all cases is smooth. The results are also very similar for field B, with a slight decrease in K and GA indices of ADB-MAP-FT to MZA. This small decrease in agreement is due to MZs rectification routines of ADB-MAP-FT, which smoothed the transitions of zones.

1 **Table 3** Management zones delineated using software MZA, FuzMe, FastMapping, and ADB-Map.*

Area	Nº of classes	Software			
		MZA	FuzME	FastMapping	ADB-MAP-FT
A	2	 			
	K/GA	-/-	1.00 / 1.00	-/-	0.99 / 1.00
	3	 			
	K/GA	-/-	1.00 / 1.00	-/-	0.98 / 0.99
	4	 			
	K/GA	-/-	1.00 / 1.00	-/-	0.97 / 0.98

		MZA	FuzME	FastMapping	ADB-MAP-FT
B	2				
	K/GA	-/-	1.00 / 1.00	-/-	0.99 / 0.99
	3				
	K/GA	-/-	1.00 / 1.00	-/-	0.97 / 0.98
	4				
	K/GA	-/-	1.00 / 1.00	-/-	0.98 / 0.98

* the ADB-Map was used instead of ADB-FT because the last one only presents the best delineated management zones. Kappa index (K), Global accuracy (GA).

Table 4 presents MZ statistics by class and software in each field. Despite this table presents MZs with two, three, and four classes for all software, ADB-MAP-FT selected two classes for both fields as the best, then presenting to the user only this option. In this work, zones with three and four classes were also generated in ADB-MAP just to be compared with the other softwares.

Table 4 Management zone statistics per software

Field	Software	Tukey's test (ha^{-1})				VR	FPI	MPE	NCE	XB	ICVI*	SI
		C1	C2	C3	C4	%						%
A	MZA	0.92a	1.11b			48.59	0.035		0.012			98.12
		0.92a	1.01ab	1.11b		30.53	0.039		0.018			96.38
		0.91a	0.93a	1.10b	1.11b	45.93	0.034		0.018			94.43
	FuzME	0.92a	1.11b			48.59	0.298	0.362			0.67	98.12
		0.92a	1.01ab	1.11b		30.53	0.304	0.337			0.77	96.38
		0.92a	0.92a	1.10b	1.11b	45.66	0.301	0.313			0.64	94.43
	FastMapping	0.91a	1.10b			51.15				5.2 E-5		
		0.92a	0.99a	1.10b		26.50				5.7 E-5		
		0.91a	0.94a	1.07b	1.12b	43.66				6.3 E-5		
	ADB-MAP-FT	0.91a	1.11b			54.85	0.074	0.089		1.9 E-5	0.64	98.21
		0.93a	1.00a	1.11b		27.46	0.077	0.076		2.0 E-5	0.80	96.60
		0.90a	0.93a	1.09b	1.12b	53.64	0.069	0.063		1.9 E-5	0.52	95.24
B	MZA	0.94a	1.01b			4.42	0.011		0.004			98.60
		0.91a	1.00b	1.04b		10.03	0.015		0.008			96.85
		0.91a	0.99b	1.00b	1.06b	9.16	0.017		0.009			95.81
	FuzME	0.94a	1.01b			4.42	0.188	0.235			0.78	98.60
		0.91a	1.00b	1.04b		10.03	0.197	0.228			0.60	96.85
		0.91a	0.99b	1.00b	1.06b	9.16	0.232	0.247			0.70	95.82
	FastMapping	0.92a	1.01b			8.15				1.5 E-5		
		0.91a	1.00b	1.01b		7.05				1.9 E-5		
		0.85a	1.00b	1.00b	1.06b	17.03				3.0 E-5		
	ADB-MAP-FT	0.94a	1.01b			4.42	0.038	0.049		1.3 E-5	0.84	98.63
		0.91a	1.00b	1.04b		10.03	0.038	0.039		1.4 E-5	0.58	97.14
		0.91a	0.99b	1.00b	1.06b	9.16	0.040	0.037		2.3 E-5	0.61	95.93

variance reduction (VR); fuzziness performance index (FPI); modified partition entropy index (MPE); normalized classification entropy (NCE); Xie e Beni (XB); Improved Cluster Validation Index (ICVI); smoothness index (SI). * The ICVI calculation was performed for each area, and for each software. Ex. ICVI for FuzME of area A is calculated from the VR, FPI and MPE of classes 2, 3 and 4 (using the 3 data sets) of FuzME.

All software showed a statistically significant difference in Tukey test for the two classes in both fields, indicating that the ideal division is statistically into two classes in both cases. Unfortunately, only FastMapping and ADB-MAP-FT present the statistical test (among other metrics) for this finding, so the user will be faced with the need to analyze one or several metrics to decide which division is ideal. Only ADB-MAP-FT presents VR, SI, and ICVI indices. For the other software, these indices, when present, were calculated using ADB-MAP.

For field A, using MZA, FPI indices indicate that the best division consists of four classes but by only one-cent the division into two classes. NCE indicates two classes. As MZA presents only these two indices, the user would be in doubt which division is the ideal one. SI and VR indices also agree with the division into two classes. FuzME presents FPI that indicates two classes and MPE indices, which indicates the division into four classes, and again the user would be in doubt which division is the ideal one. ICVI, heavily influenced by the MPE and FPI, also agrees with the division into four classes, but with a very close value to the division with two classes. VR and SI indices indicate the division into two classes. For FastMapping, considering XB and VR, the ideal division is into two classes. While for ADB-MAP-FT, FPI, MPE and ICVI indices indicate the division into four classes. XB index presents the same values for two and four classes, while SI and VR indices indicate two classes. As ICVI depends on MPE and FPI results, not presented by MZA and FastMapping, it was impossible to calculate this index for the respective software. If Tukey test shows no statistical difference among the classes, there is no meaning in dividing the field into more classes; so, the ideal division is in two classes.

For field B, considering MZA, FPI and NCE indices offered to agree on the division into two classes. The same occurs for SI index, while for VR, the ideal division would be in three classes. For FuzME, FPI and MPE indices differ in the number of ideal classes, two for the first one and three for the second, this leaves to the user the decision of the ideal division. SI index indicates the division into two classes, while ICVI and VR indices indicate the division into three classes. For FastMapping, XB indicates the division into two classes, while VR indicates four classes. For ADB-MAP-FT, the indices, FPI, XB, and SI, indicate two classes as the ideal division (with a draw of FPI to three classes), ICVI and VR three classes, and MPE four classes. It can be seen that the lack of MZA and FuzME on displaying a statistical validation hinders a final decision by the user and the need to offer several indices to help the user guide his choice.

Very close values can be seen on average per class for all software. This demonstrates that all software presents similar results in direct comparison when considering two, three, and four classes.

When considering VR, for field A, for two classes, ADB-MAP-FT has greater variability reduction, followed by FastMapping, while MZA and FuzME have the same value. When divided into three classes, MZA and FuzME show the greatest reductions, followed by ADB-MAP-FT and FastMapping. ADB-MAP-FT shows the greatest reduction for four classes, followed by MZA, FuzME, and FastMapping. For field B, FastMapping shows the greatest reduction indicated by VR for two and four classes, with a tie for the other software. For three classes, the result is exactly the opposite.

Considering SI index, for field A, two, three, and four classes, ADB-MAP-FT presents the highest values, followed by MZA and FuzME draws, indicating ADB-MAP-FT feasibility. The result is repeated for field B, except for the divergence of one-tenth between FuzME and MZA for four classes. Unfortunately, it was impossible to calculate SI of FastMapping, since this index depends on pixels being aligned in their vertical/horizontal coordinates, and FastMapping allows the interpolation result to insert pixels in diagonal coordinates making the calculation unfeasible.

Comparing ICVI of ADB-MAP-FT and FuzME, in field A, considering two and four classes, the former presents the lowest values. As for the division into three classes, FuzME presents the lowest values. In Field B, for the division into two classes, FuzME presents the lowest values, while for three and four classes, ADB-MAP-FT presents the lowest values. Considering ADB-MAP-FT and FastMapping, the advantages of both platforms' web architecture, the greater integration of pre-and post-data processing tools, the viability of both platforms to MZA and FuzME, more traditional in literature, is demonstrated. ADB-MAP-FT still has the advantage of being only necessary to select the input, boundaries, and target variables for ZM delineation. The process, including statistical validation, is done without human intervention in the automatic protocol, without external tools; it also features the most export tools and comparative indexes for both maps and MZs, presenting itself as the most complete tool.

5.4 CONCLUSIONS

The process of MZs delineation is complex and involves many areas of knowledge. Thus, specific software is crucial to make the process viable. Although MZA and FuzME present themselves as the most used software in literature, considering the progress of research since their respective releases, they end up not presenting all the necessary tools to delineate MZs. Much modern software that provides more pre- and post-data processing tools and takes advantage of state-of-art architectures, such as the web, has advantages. Among them, FastMapping and ADB-MAP-FT have several interesting features, especially their easy way to be used with modern, intuitive interfaces and automated processes. ADB-MAP-FT presented itself as the most complete solution, as it was the only one that did not require any external software for the delineation process, following the chosen protocol, in addition to the fact that it is also the easiest program to be used since all the steps are performed automatically with parameters selected from research in the specific literature.

5.5 ACKNOWLEDGMENTS

The authors would like to thank the Western Paraná State University (UNIOESTE) and the Federal University of Technology of Paraná (UTFPR).

5.6 REFERENCES

ADAMCHUK, V. .; HUMMEL, J. .; MORGAN, M. .; UPADHYAYA, S. On-the-go soil sensors for precision agriculture. **Computers and Electronics in Agriculture**, v. 44, n. 1, p. 71–91, 2004.

ANDERBERG, M. R. **Cluster Analysis for Applications**. New York: Academic Press, 1973.

AIKES JUNIOR, J.; SOUZA, E. G. DE; SOBJAK, R.; BAZZI, C. L. **Thematic Maps and Management Zones for Precision Agriculture: Systematic Literature Study, Protocols, and Practical Cases**. Curitiba: Poncã, 2021.

ALBORNOZ, E. M.; KEMERER, A. C.; GALARZA, R.; et al. Development and evaluation of an automatic software for management zone delineation. **Precision Agriculture**, v. 19, n. 3, p. 463–476, 2018.

AMIDAN, B. G.; FERRYMAN, T. A.; COOLEY, S. K. Data outlier detection using the Chebyshev theorem. 2005 IEEE Aerospace Conference. **Anais...** . p.3814–3819, 2005. Big Sky, MT, USA: IEEE.

ANSELIN, L. Local Indicators of Spatial Association-LISA. **Geographical Analysis**, v. 27, n. 2, p. 93–115, 1995.

BAZZI, C. L.; SOUZA, E. G. DE; SCHENATTO, K.; BETZEK, N. M.; GAVIOLI, A. A software for the delineation of crop management zones (SDUM). **Australian Journal of Crop Science**, v. 13, n. 01, p. 26–34, 2019.

BAZZI, C. L.; SOUZA, E. G.; URIBE-OPAZO, M. A.; NÓBREGA, L. H. P.; ROCHA, D. M. Management zones definition using soil chemical and physical attributes in a soybean area. **Engenharia Agrícola**, v. 33, n. 5, p. 952–964, 2013.

BERNARDI, A. C. DE C.; TUPY, O.; EDUARDA LIMA SANTOS, K.; et al. Mapping of yield, economic return, soil electrical conductivity, and management zones of irrigated corn for silage. **Pesq. agropec. bras**, v. 53, n. 12, p. 1289–1298, 2018.

BETZEK, N. M.; SOUZA, E. G. DE; BAZZI, C. L.; et al. Computational routines for the automatic selection of the best parameters used by interpolation methods to create thematic maps. **Computers and Electronics in Agriculture**, v. 157, n. January 2018, p. 49–62, 2019. Elsevier.

BETZEK, N. M.; SOUZA, E. G. DE; BAZZI, C. L.; SCHENATTO, K.; GAVIOLI, A. Rectification methods for optimization of management zones. **Computers and Electronics in Agriculture**, v. 146, n. February, p. 1–11, 2018.

BEZDEK, J. C. **Pattern Recognition with Fuzzy Objective Function Algorithms**. New

York: Plenum Press, 1981.

BIER, V. A.; DE SOUZA, E. G. Interpolation selection index for delineation of thematic maps. **Computers and Electronics in Agriculture**, v. 136, p. 202–209, 2017.

BLACKMORE, S. The interpretation of trends from multiple yield maps. **Computers and Electronics in Agriculture**, v. 26, n. 1, p. 37–51, 2000.

BOBRYK, C. W.; MYERS, D. B.; KITCHEN, N. R.; et al. Validating a digital soil map with corn yield data for precision agriculture decision support. **Agronomy Journal**, v. 108, n. 3, p. 957–965, 2016.

BORGES, L. F.; BAZZI, C. L.; SOUZA, E. G. DE; MAGALHÃES, P. S. G.; MICHELON, G. K. Web software to create thematic maps for precision agriculture. **Pesquisa Agropecuária Brasileira**, v. 55, 2020.

BOTTEGA, E. L.; DE QUEIROZ, D. M.; PINTO, F. DE A. DE C.; DE SOUZA, C. M. A.; VALENTE, D. S. M. Precision agriculture applied to soybean: Part I - Delineation of management zones. **Australian Journal of Crop Science**, v. 11, n. 5, p. 573–579, 2017.

BROCK, A.; BROUDER, S. M.; BLUMHOFF, G.; HOFMANN, B. S. Defining Yield-Based Management Zones for Corn-Soybean Rotations. **Agronomy Journal**, v. 97, n. 4, p. 1115–1128, 2005.

BUTTAFUOCO, G.; CASTRIGNANÒ, A.; COLECCHIA, A. S.; RICCA, N. Delineation of management zones using soil properties and a multivariate geostatistical approach. **Italian Journal of Agronomy**, v. 5, n. 4, p. 323–332, 2010.

CHANG, D.; ZHANG, J.; ZHU, L.; et al. Delineation of management zones using an active canopy sensor for a tobacco field. **Computers and Electronics in Agriculture**, v. 109, p. 172–178, 2014.

CID-GARCIA, N. M.; IBARRA-ROJAS, O. J. An integrated approach for the rectangular delineation of management zones and the crop planning problems. **Computers and Electronics in Agriculture**, v. 164, n. July, p. 104925, 2019. Elsevier.

COELHO, E. C.; DE SOUZA, E. G.; URIBE-OPAZO, M. A.; NETO, R. P. Influência da densidade amostral e do tipo de interpolador na elaboração de mapas temáticos. **Acta Scientiarum - Agronomy**, v. 31, n. 1, p. 165–174, 2009.

COHEN, J. A Coefficient of Agreement for Nominal Scales. **Educational and Psychological Measurement**, v. 20, n. 1, p. 37–46, 1960.

CÓRDOBA, M. A.; BRUNO, C. I.; COSTA, J. L.; PERALTA, N. R.; BALZARINI, M. G. Protocol for Multivariate Homogeneous Zone Delineation in Precision Agriculture. **Biosystems Engineering**, v. 143, p. 95–107, 2016.

CÓRDOBA, M.; BRUNO, C.; COSTA, J.; BALZARINI, M. Subfield management class delineation using cluster analysis from spatial principal components of soil variables. **Computers and Electronics in Agriculture**, v. 97, p. 6–14, 2013.

DALL'AGNOL, R. W.; MICHELON, G. K.; BAZZI, C. L.; et al. Web applications for spatial analyses and thematic map generation. **Computers and Electronics in Agriculture**, v. 172, n. March, p. 105374, 2020. Elsevier.

DOBERMANN, A.; PING, J. L.; ADAMCHUK, V. I.; SIMBAHAN, G. C.; FERGUSON, R. B. Classification of Crop Yield Variability in Irrigated Production Fields. **Agronomy Journal**, v.

95, n. 5, p. 1105, 2003.

DOERGE, T. A. Site-Specific Management Guidelines. In: C. et Al (Org.); , 2000. Norcros: Potash e Phosphate Institute.

DRAY, S.; SAÏD, S.; DÉBIAS, F. Spatial ordination of vegetation data using a generalization of Wartenberg's multivariate spatial correlation. **Journal of Vegetation Science**, v. 19, n. 1, p. 45–56, 2008.

FERGUSON, R. B.; HERGERT, G. W. Soil Sampling for Precision Agriculture. **Extension: Precision Agriculture**, p. 1–4, 2009.

FLEMING, K. L.; WESTFALL, D. G.; WIENS, D. W.; BRODAHL, M. C. Evaluating farmer defined management zone maps for variable rate fertilizer application. **Precision Agriculture**, v. 2, n. 2, p. 201–215, 2000.

FOODY, G. M. Status of land cover classification accuracy assessment. Remote sensing of environment, v. 80, n. 1, p. 185-201, 2002.

FRANZEN, D. W.; HOPKINS, D. H.; SWEENEY, M. D.; ULMER, M. K.; HALVORSON, A. D. Evaluation of Soil Survey Scale for Zone Development of Site-Specific Nitrogen Management. **Agronomy Journal**, v. 94, n. 2, p. 381, 2002.

FRIDGEN, J. J.; KITCHEN, N. R.; SUDDUTH, K. A. Variability of soil and landscape attributes within sub-field management zones. Proceedings of the 5th International Conference on Precision Agriculture. **Anais...** . p.1–16, 2000. Bloomington, Minnesota, USA.

FRIDGEN, J. J.; KITCHEN, N. R.; SUDDUTH, K. A.; et al. Management Zone Analyst (MZA): Software for Subfield Management Zone Delineation. **Agronomy Journal**, v. 96, n. 1, p. 100–108, 2004.

GAVIOLI, A.; DE SOUZA, E. G.; BAZZI, C. L.; GUEDES, L. P. C.; SCHENATTO, K. Optimization of management zone delineation by using spatial principal components. **Computers and Electronics in Agriculture**, v. 127, p. 302–310, 2016.

GONZALES, R. C.; WOODS, R. E. **Digital Image Processing**. 3º ed. Upper Saddle River: Pearson Prentice Hall, 2008.

GOZDOWSKI, D.; STEPIEŃ, M.; SAMBORSKI, S.; et al. Determination of the Most Relevant Soil Properties for the Delineation of Management Zones in Production Fields. **Communications in Soil Science and Plant Analysis**, v. 45, n. 17, p. 2289–2304, 2014.

HÖRBE, T. A. N.; AMADO, T. J. C.; FERREIRA, A. O.; ALBA, P. J. Optimization of corn plant population according to management zones in Southern Brazil. **Precision Agriculture**, v. 14, n. 4, p. 450–465, 2013.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. **Journal of Educational Psychology**, v. 6, n. 24, p. 417–441, 1933.

ISAAKS, E. H.; SRIVASTAVA, R. M. **Applied Geostatistics**. 1º ed. New York: Oxford, 1989.

JAMISON, V.C., SMITH, D.D., THORNTON, J.F. Soil and water research on a claypan soil. USDA-ARS Technical Bulletin 1379, U.S. Government Printing Office, Washington, DC., 1968.

JAYNES, D. B.; COLVIN, T. S.; KASPAR, T. C. Identifying potential soybean management zones from multi-year yield data. **Computers and Electronics in Agriculture**, v. 46, n. 1-3 SPEC. ISS., p. 309–327, 2005.

KERRY, R.; OLIVER, M. A.; FROGBROOK, Z. L. Geostatistical Applications for Precision Agriculture. In: M. A. Oliver (Org.); **Geostatistical Applications for Precision Agriculture**. p.35–63, 2010. Dordrecht: Springer Netherlands.

KHOSLA, R.; ALLEY, M. M. Soil-Specific Nitrogen Management On Mid-Atlantic Coastal Plain Soils. **Better Crops**, v. 83, n. 3, p. 6–7, 1999.

KHOSLA, R.; FLEMING, K.; DELGADO, J. A.; SHAVER, T. M.; WESTFALL, D. G. Use of site-specific management zones to improve nitrogen management for precision agriculture. **Soil and Water**, v. 57, n. 6, p. 513–518, 2002.

KITCHEN, N. R.; SUDDUTH, K. A.; DRUMMOND, S. T.; et al. Procedures of Evaluating Unsupervised Classification to Derive Management Zones. Proceedings of the Six International Conference on Precision Agriculture. **Anais...** , 2002.

KITCHEN, N. R.; SUDDUTH, K. A.; MYERS, D. B.; DRUMMOND, S. T.; HONG, S. Y. Delineating productivity zones on claypan soil fields using apparent soil electrical conductivity. **Computers and Electronics in Agriculture**, v. 46, p. 285–308, 2005.

KUANG, B.; MAHMOOD, H. S.; QURAISHI, M. Z.; et al. Sensing Soil Properties in the Laboratory, In Situ, and On-Line. **Advances in Agronomy**. p.155–223, 2012. Academic Press.

KYAW, T.; FERGUSON, R. B.; ADAMCHUK, V. I.; et al. Delineating site-specific management zones for pH-induced iron chlorosis. **Precision Agriculture**, v. 9, n. 1–2, p. 71–84, 2008. VAN GODEWIJCKSTRAAT 30, 3311 GZ DORDRECHT, NETHERLANDS: SPRINGER.

LANDIS, J.R., KOCH, G.G. The measurement of observer agreement for categorical data. Biometrics 33, 159–174, 1977

LARSCHEID, G.; BLACKMORE, B. S. Interactions Between Farm Managers and Information Systems with Respect to Yield Mapping. **Proceedings of the Third International Conference on Precision Agriculture**. p.1153–1163, 1996. Minneapolis: Springer.

LI, Y.; SHI, Z.; LI, F.; LI, H. Y. Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. **Computers and Electronics in Agriculture**, v. 56, n. 2, p. 174–186, 2007.

LI, Y.; SHI, Z.; WU, H. X.; LI, F.; LI, H. Y. Definition of management zones for enhancing cultivated land conservation using combined spatial data. **Environmental Management**, v. 52, n. 4, p. 792–806, 2013.

MALLARINO, A. P.; WITTRY, D. J. Efficacy of grid and zone soil sampling approaches for site-specific assessment of phosphorus, potassium, pH, and organic matter. **Precision Agriculture**, v. 5, n. 2, p. 131–144, 2004.

MCBRATNEY, A. B.; MOORE, A. W. Application of fuzzy sets to climatic classification. **Agricultural and Forest Meteorology**, v. 35, n. 1–4, p. 165–185, 1985.

MIAO, Y.; MULLA, D. J.; ROBERT, P. C. An integrated approach to site-specific management zone delineation. **Frontiers of Agricultural Science and Engineering**, v. 5, n. 4, p. 432–441, 2018.

MICHELON, G. K.; BAZZI, C. L.; UPADHYAYA, S.; et al. Software AgDataBox-Map to precision agriculture management. **SoftwareX**, v. 10, p. 100320, 2019. Elsevier B.V.

MIEZA, M. S.; CRAVERO, W. R.; KOVAC, F. D.; BARGIANO, P. G. Delineation of site-specific management units for operational applications using the topographic position index in La Pampa, Argentina. **Computers and Electronics in Agriculture**, v. 127, p. 158–167, 2016.

MILLIGAN, G. W.; COOPER, M. C. A study of standardization of variables in cluster analysis. **Journal of Classification**, v. 5, n. 2, p. 181–204, 1988.

MINASNY, B.; MCBRATNEY, A. B. FuzME version 3.0. , 2002. Australian Centre for Precision Agriculture, The University of Sydney.

MOLIN, J. P. Definição de Unidade de Manejo a Partir de Mapas de Produtividade. **Engenharia Agrícola**, v. 22, n. 1, p. 83–92, 2002.

MORAL, F. J.; TERRÓN, J. M.; SILVA, J. R. M. DA. Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. **Soil and Tillage Research**, v. 106, n. 2, p. 335–343, 2010.

MOSHIA, M. E.; KHOSLA, R.; LONGCHAMPS, L.; et al. Precision manure management across site-specific management zones: Grain yield and economic analysis. **Agronomy Journal**, v. 106, n. 6, p. 2146–2156, 2014.

NAWAR, S.; CORSTANJE, R.; HALCRO, G.; MULLA, D.; MOUAZEN, A. M. Delineation of Soil Management Zones for Variable-Rate Fertilization: A Review. In: Sparks, DL (Org.); **Advances in Agronomy**. v. 143, p.175–245, 2017.

NANNI, M.R.; POVH, F.P.; DAMATTÊ, J.A.M.; OLIVEIRA, R.B. de; CHICATI, M.L.; CEZAR, E. Optimum size in grid soil sampling for variable rate application in site-specific management. **Scientia Agricola**, v.68, p.386-392, 2011.

ODEH, I. O. A., Chittleborough, D. J., McBratney, A. B. Soil Pattern Recognition with Fuzzy-c-means: Application to Classification and Soil-Landform Interrelationships. **Soil Sci Soc Am J**. v. 56, p. 505, 1992.

PACCIORETTI, P.; CÓRDOBA, M.; BALZARINI, M. FastMapping: Software to create field maps and identify management zones in precision agriculture. **Computers and Electronics in Agriculture**, v. 175, n. May, p. 105556, 2020. Elsevier.

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

ROBERTS, D. F.; FERGUSON, R. B.; KITCHEN, N. R.; ADAMCHUK, V. I.; SHANAHAN, J. F. Relationships between soil-based management zones and canopy sensing for corn Nitrogen management. **Agronomy Journal**, v. 104, n. 1, p. 119–129, 2012.

ROBERTSON, M. J.; LYLE, G.; BOWDEN, J. W. Within-field variability of wheat yield and economic implications for spatially variable nutrient management. **Field Crops Research**, v. 105, n. 3, p. 211–220, 2008.

ROUSSEEUW, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. **Journal of Computational and Applied Mathematics**, v. 20, p. 53–65, 1987.

SCHENATTO, K.; DE SOUZA, E. G.; BAZZI, C. L.; et al. Normalization of data for delineating management zones. **Computers and Electronics in Agriculture**, v. 143, p.

238–248, 2017.

SCHENATTO, K.; SOUZA, E. G. DE; BAZZI, C. L.; BETZEK, N. M.; GAVIOLI, A. Two approach comparison to define crop management zones (MZs). **African Journal of Agricultural Research**, v. 11, n. 38, p. 3680–3689, 2016.

SCHEPERS, A. R.; SHANAHAN, J. F.; LIEBIG, M. A.; SCHEPERS, J. S.; JOHNSON, S. H.; LUCHIARI, J. A. Appropriateness of management zones for characterizing spatial variability of soil properties and irrigated corn yields across years. **Agronomy Journal,** v. 96, p. 195-203, 2004.

SCHWALBERT, R. A.; AMADO, T. J. C.; REIMCHE, G. B.; GEBERT, F. Fine-tuning of wheat (Triticum aestivum, L.) variable nitrogen rate by combining crop sensing and management zones approaches in southern Brazil. **Precision Agriculture**, v. 20, n. 1, p. 1–22, 2018.

SOBJAK, R. Incorporation of computational modules as microservices in the AgDataBox platform and development of agdatabox-map application. Tese (Doutorado em Doutorado em Engenharia Agrícola) - Universidade Estadual do Oeste do Paraná. 256 p., 2021.

SOBJAK, R.; SOUZA, E. G. DE; BAZZI, C. L.; URIBE-OPAZO, M. A.; BETZEK, N. M. Redundant variables and the quality of management zones. **Engenharia Agrícola**, v. 36, n. 1, p. 78–93, 2016.

SOUZA, E. G.; SCHENATTO, K.; BAZZI, C. L. Creating thematic maps and management zones for agriculture fields. 14th International Conference on Precision Agriculture. **Anais...** . p.1–17, 2018. Montreal.

SWINDELL, J. 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. **BIOSScientific**. p. 827-834,1997.

TAYLOR, J. A.; MCBRATNEY, A. B.; WHELAN, B. M. Establishing Management Classes for Broadacre Agricultural Production. **Agronomy Journal**, v. 99, n. 5, p. 1366, 2007.

TAYLOR, J. C.; WOOD, G. A.; EARL, R.; GODWIN, R. J. Soil factors and their influence on within-field crop variability, Part II: Spatial analysis and determination of management zones. **Biosystems Engineering**, v. 84, n. 4, p. 441–453, 2003.

VEGA, A.; CÓRDOBA, M.; CASTRO-FRANCO, M.; BALZARINI, M. Protocol for automating error removal from yield maps. **Precision Agriculture**, v. 20, n. 5, p. 1030–1044, 2019.

VELANDIA, M.; REJESUS, R. M.; BRONSON, K.; SEGARRA, E. Economics of management zone delineation in cotton precision agriculture. **Journal of Cotton Science**, v. 12, n. 3, p. 210–227, 2008.

VITHARANA, U. W. A.; VAN MEIRVENNE, M.; SIMPSON, D.; COCKX, L.; HOFMAN, G. Agronomic consequences of potential management zones delineated on the basis of EM38DD measurements. **Near Surface Geophysics**, v. 6, n. 5, p. 289–296, 2008.

WEBSTER, R.; OLIVER, M. A. **Geostatistics for Environmental Scientists**. Chichester, UK: John Wiley & Sons, Ltd, 2007.

WHETTON, R. L.; WAINE, T. W.; MOUAZEN, A. M. Evaluating management zone maps for variable rate fungicide application and selective harvest. **Computers and Electronics in Agriculture**, v. 153, p. 202–212, 2018.

WOLLENHAUPT, N. C.; WOLKOWSKI, R. P.; CLAYTON, M. K. Mapping Soil Test Phosphorus and Potassium for Variable-Rate Fertilizer Application. **Journal of Production**

Agriculture, v. 7, n. 4, p. 441–448, 1994.

XIE, X. L.; BENI, G. A validity measure for fuzzy clustering. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 13, n. 8, p. 841–847, 1991.

XIANG, L.; YU-CHUN, P.; ZHONG-QIANG, G.; CHUN-JIANG, Z. Delineation and Scale Effect of Precision Agriculture Management Zones Using Yield Monitor Data Over Four Years. **Agricultural Sciences in China**, v. 6, n. 2, p. 180–188, 2007.

YANG, G.; LIU, J.; ZHAO, C.; et al. Unmanned Aerial Vehicle Remote Sensing for Field-Based Crop Phenotyping: Current Status and Perspectives. **Frontiers in Plant Science**, v. 8, 2017.

ZERAATPISHEH, M.; BOTTEGA, E. L.; BAKHSHANDEH, E.; et al. Spatial variability of soil quality within management zones: Homogeneity and purity of delineated zones. **Catena**, v. 209, n. P1, p. 105835, 2022. Elsevier B.V.

ZHANG, X.; SHI, L.; JIA, X.; SEIELSTAD, G.; HELGASON, C. Zone mapping application for precision-farming: A decision support t tool for variable rate application. **Precision Agriculture**, v. 11, n. 2, p. 103–114, 2010.

Appendix A

Data normalization methods

- **Range** (Anderberg, 1973; Milligan and Cooper, 1988 – Equation A1):

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

where, Z_{iN} – normalized observation i; X_i – original data value i; $\mathrm{Min}(X)$ – minimum value of data set; $\mathrm{Max}(X)$ – maximum value of 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; $\overline{X}$ – sample mean of data set.

- **Standard Score** or **Z-Score** (Larscheid and Blackmore, 1996 – Equation A3):

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

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

- **Min-Max** method (Milligan and Cooper, 1988 – Equation A4):

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

where, Z_{iN} – normalized observation i; X_i – original data value i; $\mathrm{Min}(X)$ – minimum value of data set; $\mathrm{Max}(X)$ – 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 ($i = 1, 2, \ldots, n$), to get a zero average by the formulas $y_i = (y_i - \overline{Y})$ and $z_j = (z_j - \overline{Z})$, where $\overline{Y}$ and $\overline{Z}$ 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):

$$\mathrm{ISI} = \left\{\frac{\mathrm{abs(ME)}}{\max\Big|_{i=1}^{j}[\mathrm{abs(ME)}]} + \frac{\left[\mathrm{SDME} - \min\Big|_{i=1}^{j}\mathrm{SDME}\right]}{\max\Big|_{i=1}^{j}[\mathrm{abs(SDME)}]}\right\}, \tag{A6}$$

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.

$$\mathrm{ME} = \frac{1}{n}\sum_{i=1}^{n} Z(s_i) - \hat{Z}(s_i), \tag{A7}$$

$$\mathrm{SDME} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Z(s_i) - \hat{Z}(s_i)\right)^2}, \tag{A8}$$

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.

Data interpolation - Inverse Distance Weighting (IDW – Equation A9): is calculated by:

$$\hat{Z}_i = \frac{\sum_{i=1}^{n}\left(\frac{1}{d_i^p} * Z_i\right)}{\sum_{i=1}^{n}\left(\frac{1}{d_i^p}\right)}, \tag{A9}$$

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

Indices for evaluation of the management zones quality

a) **Variance reduction** (VR% – Xiang et al., 2007; Schenatto et al., 2017 – Equation A10): 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.

$$\mathrm{VR\%} = \left(1 - \frac{\sum_{i=1}^{c} W_i * V_{MZ_i}}{V_{field}}\right) * 100, \tag{A10}$$

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 A11): measures the degree of separation between the fuzzy c groups generated from a data set. FPI varies between 0 and 1.

$$\mathrm{FPI} = 1 - \frac{\mathrm{c}}{(\mathrm{c}-1)}\left[1 - \sum_{\mathrm{j}=1}^{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{c}}(\frac{(\mathrm{m_{ij}})^2}{\mathrm{n}}\right], \quad \text{(A11)}$$

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 A12): estimates the difficulty level to organize *c* groups.

$$\mathrm{MPE} = \frac{-\sum_{\mathrm{j}=1}^{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{c}} \mathrm{m_{ij}}\log(\mathrm{m_{ij}})/\mathrm{n}}{\log \mathrm{c}}, \quad \text{(A12)}$$

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 A13): is a composition of FPI, MPE, and VR% indices.

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

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 A14): 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.

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

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

g) **Average Silhouette Coefficient** (ASC – Rousseeuw, 1987 – Equation A15): 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:

$$\mathrm{sc_p} = \frac{\mathrm{b_p} - \mathrm{a_p}}{\mathrm{Max}(\mathrm{a_p}, \mathrm{b_p})}, \tag{A15}$$

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) **Normalized Classification Entropy** (NCE): this coeficiente models the amount of disorganization of a fuzzy c-partition of *Y* (Bezdek (1981); Odeh et al., 1992). The classification entropy (*H*) is defined by Equation A16:

$$H(\boldsymbol{U}; c) = -\sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik} log_a(u_{ik})/n \tag{A16}$$

where ***U*** is the fuzzy membership matrix, *c are the data entry partitions* and the logarithmic base a is any positive integer. Values of *H* will range from 0 to $log_a(c)$. Bezdek (1981) reported that the endpoints of *H* range do not accurately represent the amount of disorganization present (i.e., at *c* = 1, *H* = 0; at *c* = *n*, *H* = 0). To remedy this issue, Fridgen et al. (2004) suggested NCE variation in Equation A17:

$$NCE = \frac{H(\boldsymbol{U}; \boldsymbol{c})}{1 - (c/n)} \tag{A17}$$

NCE values will be similar to those of *H* when *c* is relatively small compared with *n* [i.e., (*c/n*) ap proaching 0]. However, in situations where (*c/n*) is large (i.e., pproaching 1), NCE will produce substantially different results.

i) **Xie and Beni index** (XB – Equation A18): this index is focus on separation and compactness. Separation is a measure of the distance between one cluster and another cluster and compactness is a measure of proximity between data points in a cluster. According to this method, the optimal *c* is the one with the smallest XB value (Xie e Beni, 1991). The function of this method is :

$$V_{XB} = \frac{\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^{2}||V_i - X||^2}{nmin_{i,j}||V_i - X_j||^2} \tag{A18}$$

where μ is the degree of membership, *V* is the centroid and *X* is the data entry.

Indices for comparison between thematic maps and between management zones

a) **Kappa coefficient** (K) (Cohen, 1960): this index is not used to validate the clustering process but to compare the agreement of two MZ delineation approach. Landis and Koch (1977) proposed the following **classification**: 0 < K ≤ 0.2 indicates no agreement, 0.2 < K ≤ 0.4 weak agreement, 0.4 < K ≤ 0.6 moderate agreement, 0.6 < K ≤ 0.8 strong agreement, and 0.8 < K ≤ 1 very strong agreement.

b) **Global accuracy** (GA – Foody, 2002 – Equation A19): like K, GA measures the degree of agreement among maps (MZs) and corresponds to the simple percent agreement.

$$\mathrm{GA} = \frac{\sum_{i=1}^{c} x_{ii}}{n}, \tag{A19}$$

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).

6 FINAL CONSIDERATIONS

The developed computational module correctly applies the protocols for automatically constructing thematic maps and the management zones delineation. The case study demonstrated its ease of use from a modern and user-friendly graphical interface and the advantages of data persistence and cloud processing. The comparison with other traditional and state-of-the-art software showed greater ease of use, greater completeness of integrated tasks, better time efficiency, better data persistence, and a greater amount of analysis tools in the module developed in this work.

The developed computational module is already integrated with the free web platform AgDataBox (https://adb.md.utfpr.edu.br/map/auth/login) and can be used by technicians and researchers for commercial and educational activities or research.

7 FUTURE WORKS

As suggestions for future works that can be incorporated to the module, a researcher can add the option of choosing the best interpolator by the kriging algorithm in addition to IDW. It is also possible to add other algorithms besides FCM to MZs delineation in the configurations, since ADB-MAP has thirteen more delineation algorithms.