JORGE AIKES JUNIOR . EDUARDO GODOY DE SOUZA . CLAUDIO LEONES BAZZI . RICARDO SOBJAK

THEMATIC MAPS AND MANAGEMENT ZONES FOR PRECISION AGRICULTURE

SYSTEMATIC LITERATURE STUDY, PROTOCOLS, AND PRACTICAL CASES

Coordenação Editorial e Capa: Thais Serpa
Fotografia da capa: Ivan Bandura
sob Unsplash License

Catalogação na publicação (CIP)
Ficha catalográfica elaborada pelo editor

A291t Aikes Junior, Jorge [1986 -]
Thematic Maps and Management Zones for Precision Agriculture: Systematic Literature Study, Protocols, and Practical Cases / Jorge Aikes Junior; Eduardo Godoy de Souza; Claudio Leones Bazzi; Ricardo Sobjak -- Curitiba: Poncã, 2021.

ISBN: 978-65-80756-02-5

1. Zona de manejo. 2. Mapas temáticos. 3. Agricultura de precisão.
I. Título. II. Autor.
•

CDD: 630

Índice para catálogo sistemático:
1. Agricultura e tecnologias relacionadas 630

THEMATIC MAPS AND MANAGEMENT ZONES FOR PRECISION AGRICULTURE

SYSTEMATIC LITERATURE STUDY, PROTOCOLS, AND PRACTICAL CASES

List of Figures

Fig. 1. Thematic map of world apple production in 2009 13
Fig. 2. Examples of contour map: a) elevation (m), and choropleth maps: b) elevation (m), c) sand (%), d) Clay (%) 14
Fig. 3. Flowchart of the typical protocol to create a thematic map 15
Fig. 4. Thematic Maps for pH (a), and aluminium (b) using three forms of classification (equal interval, quantile, and standard deviation) 20
Fig. 5. Three kinds of color scheme: nominal/qualitative (a), sequential (b), and diverging (c) 20
Fig. 6. Examples of color scheme: nominal/qualitative (a), sequential (b), and diverging (c) 21
Fig. 7. 3D interpolated map of wheat yield (kg ha-1) for 2003 22
Fig. 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 24
Fig. 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 25
Fig. 10. Comparison of the 2004 yield map of a ‘Cabernet Sauvignon’ plot (left) with the map obtained from a prediction model using the 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 26
Fig. 11. The firmness of fruit pulp (a) and total soluble solids of apple fruits (b) in a 3.13-ha area 27
Fig. 12. Weed maps showing three infestations levels from a 72-ha flat terrain, planted in the no-tillage system and with soil covered by soybean stubble 28
Fig. 13. Kriged maps for dry matter yield (a), stocking rate (b), and milk yield (c) of a grazed alfalfa pasture in Brazil 29
Fig. 14. Workflow used with the Systematic Literature Mapping (SLM) 32
Fig. 15. Workflow used for the Systematic Literature Study (SLS) 34

Fig. 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).. 35
Fig. 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)....................................... 35
Fig. 18. Distribution of selected studies by country of study (classified by the country where the authors conducted the research, and when is only a theoretical manuscript, where it was published)... 36
Fig. 19. Distribution of selected articles by journals in which the selected studies were published 37
Fig. 20. The protocol of the delineation of management zones, 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 40
Fig. 21. Reference yield productivity zone maps (left) compared to the best performing productivity zone map derived from unsupervised clustering of ECa and elevation (right)... 72
Fig. 22. EC and OM maps generated by the on-the-go sensor for Field 2 73
Fig. 23. Terrain slope and curvature maps for Field 2 74
Fig. 24. 5-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)74
Fig. 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 75
Fig. 26. Smoothed choropleth maps produced by ordinary kriging for apparent electrical conductivity (ECa) at three different sampling dates 76
Fig. 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)........................... 77

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).. 77
Fig. 29. Comparison of management zones delineated with using 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 the two techniques are presented as gray color referred to as “disagree” in the legend. 78
Fig. 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 79
Fig. 31. Management zones of small leaves weed plants in a 1.24 ha pear orchard 79
Fig. 32. Management zones of large leaves weed plants in a 1.24 ha pear orchar 80
Fig. 33. Homogenous zones resulting from the clustering analysis of VIs, calculated based on ρ, for two study 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 81
Fig. 34. Using 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 for determination of variable rate fertilizer application resulting in a more uniform yield for 2005 spring wheat (d) 82
Fig. 35. Map of management zones based on soil properties (A) and NDVI measurements (B)... 83
Fig. 36. Yield management zones delineated using grape normalized yields .. 83
Fig. 37. Shallow (0 – 0.3 m) and deep-reading (0 – 0.9 m) soil EC maps, soil clay and sand content maps, and Management zones 84
Fig. 38. Soybean-yield cluster classification for the 224 transect plots overlaid on the elevation contours (a) and the predicted yield zones (b). Transect plots are shown 3× actual width for better visibility 85
Fig. 39. Division of the area into management zones using the Fuzzy C-means algorithm with variables selected with spatial correlation matrix approach....... 85

Fig. 40. 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. 86
Fig. 41. Map with two (left), three (center), and four (right) within-field management classes 87
Fig. 42. 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 88
Fig. 43. Maps of the management zones delineated with the application of 17 clustering algorithms for the three fields 89
Fig. 44. Zones fragmentation for the delineated management zones (Site 1) before (a) and after (b) the automatic filtering post-processing techniques 91

List of Tables

Table 1. Clustering of studies selected by research technique 33
Table 2. Main techniques for the selection of variables for delineation of management zones (MZs) 45
Table 3. Algorithms used for the delineation of management zones (MZs) 49
Table 4. Values adopted for the fuzzy weighting coefficient in the fuzzy c-means algorithm 50
Table 5. Types of distances used in the Fuzzy c-means algorithm 51
Table 6. Most used measures for choosing the number of management zones (MZs) 56
Table 7. Combination of most used measures for choosing the number of management zones (MZs) 57
Table 8. Software used for the delineation of management zones by the papers included in this review 65
Table 9. Features of specific software for delineation of management zones (MZs) 70
Table 10. Clustering methods implemented and compared for the definition of MZs 90
Table 11. Results for ANOVA (Tukey's range test), VR, FPI, MPE, SI, and ICVI, for field A 92
Table 12. Results of the evaluation of the clustering methods in the generation of two, three and four classes by the ANOVA (Tukey's range test), VR index and ASC coefficient, for field A 93

Summary

1. Introduction ... 10
2. Thematic Maps ... 13
2.1. Examples of thematic maps ... 22
2.1.1 Yield, protein and oil content maps ... 22
2.1.2. Yield, profit, and profitability maps ... 25
2.1.3. Grape yield maps ... 26
2.1.4. Apple attributes maps ... 26
2.1.5. Weed infestation maps ... 27
2.1.6. Dry matter yield, stocking rate, and milk yield maps ... 28
3. Management Zones (MZs) ... 30
3.1 Systematic Literature Study (SLS) ... 30
3.1.2 Results and discussion of the SLS ... 37
3.2. Examples of management zones ... 72
3.2.1. Target values of the management zones ... 72
3.2.2. Chlorosis management zones ... 74
3.2.3. Apparent electrical conductivity management zones ... 76
3.2.4. Soil available water content management zones ... 77
3.2.5. Quality-based management zones ... 78
3.2.6. Weed Management Zones ... 79
3.2.7. Vegetation Indices Management Zones ... 80
3.2.8. Used variables for delineating Management Zones ... 81
3.2.9. Methods for selecting the variables used in the clustering process ... 85
3.2.10. Methods for delineating Management Zones ... 88
3.2.11. Rectification of Management Zones ... 90
3.2.12. Evaluation of the quality of Management Zones ... 91

1. Introduction

The growing population, demanding large amounts of food, and environmental issues, for both the conservation of the environment itself and for the more rational use of all elements of the food production chain, are motivating producers to make the most optimized use of the land and its inputs (Baudron and Giller 2014). Thus, producers need to make the most optimized use of the land and its inputs. Precision agriculture (PA) is a management system that aims to optimize the use of agricultural inputs, meeting this need for more profitability with less environmental damage.

Climatic, topographic, and biological variations, both in 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. Among them, thematic maps and management zones stand out.

In addition to representing the terrain, thematic maps (TMs) are used to illustrate themes. Generally, TMs are used to identify different cartographic representations, and they represent not only the land but also associated characteristics. The development of TMs is linked to data collection, analysis, interpretation, and representation of the information on a map. They facilitate the identification of similarities and enable the visualization of spatial correlations. Based on samples collected before, during e after the life period of the culture, TMs are usually generated to identify the variability of properties of the topography, soil, and plants and compare with the yield. However, first, it is necessary to interpolate the data into a dense and regular grid to provide values for locations that were not sampled. This task is performed with the aid of interpolation methods, being kriging 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 in the 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 the variability in soil, topography, local weather conditions, and land use is not considered (Nawar et al., 2017). In this management, inputs are applied uniformly across the field, and it is attractive to growers because it is easy and speedy. However, with a site-specific input application is possible to achieve more economical and environmentally-friendly management. 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 the level of reliability (Ferguson and Hergert, 2009; Mallarino and Wittry, 2004). Smart sampling has been shown to improve nutrient efficiency use while keeping or increasing the yield and potentially reducing the nutrient overloading into the environment (Moshia et al., 2014; Khosla et al., 2002). Many studies related to the 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), resulting in a suggested minimum density of one sample per ha (Ferguson and Hergert, 2009) to 2.5 samples per ha (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 target variable (usually yield) (Buttafuoco et al., 2010; Doerge, 2000). That is important because, usually, we want to use the MZs for many years. Nevertheless, there are other situations in which the purpose is to use immediately and just once MZs. It is the case of MZs for agrochemical applications.

The use of 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. 2008b; Roberts et al. 2012; Li et al. 2013; Bernardi et al. 2018; Schwalbert et al. 2018; Whetton et al. 2018). Thus, its application often leads 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 the delineation of MZs, (ii) what is the best delineation algorithm, (iii) which software allows you to handle all the stages in the process. Because of this, the task of defining ideal MZs is still challenging.

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: Chapter 2 presents the TMs, with their characteristics, importance, usage, definitions for the best choice of color scheme, and several examples. Chapter 3 presents the MZs. As the delineation of MZs presents several possibilities, the definitions, protocols, economic return, and most common options and software used are based on a systematic study of the 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 various possibilities.

2. Thematic Maps

Maps that represent the land and a topic associated with it are called thematic maps (TMs), and they aim to inform through graphic symbols where a specific geographical phenomenon occurs. TM development is linked to data collection, analysis, interpretation, and representation of the 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 the soil, and soybean yield at a given agricultural area. Fig. 1 shows a TM of world apple production in 2009.

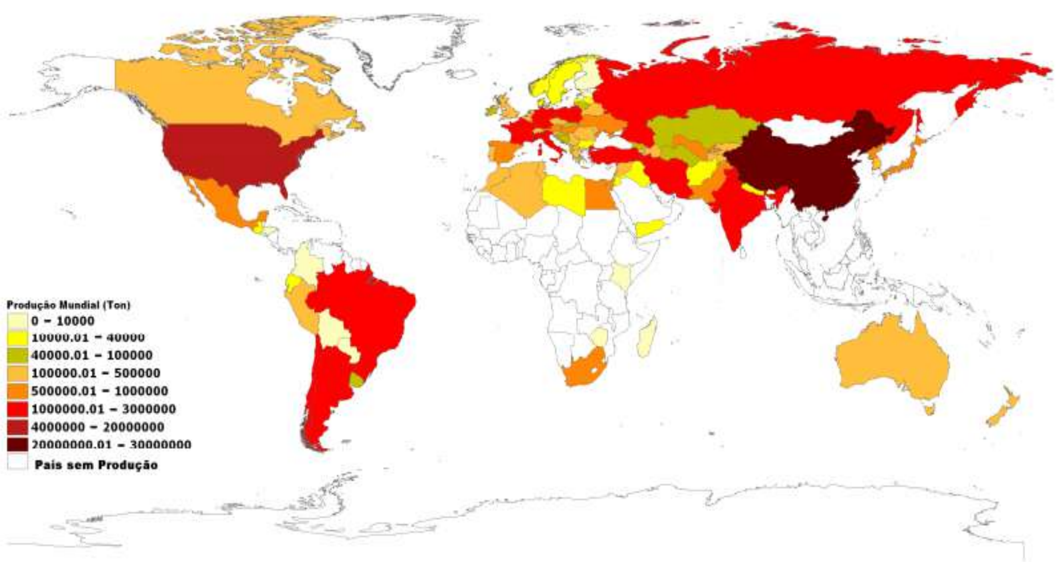
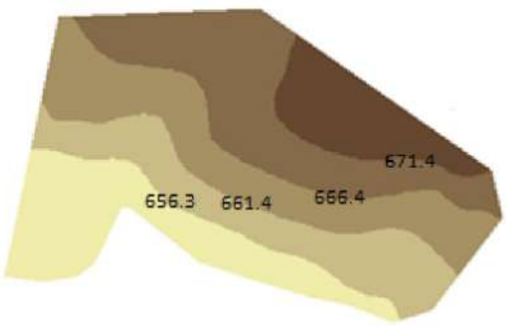


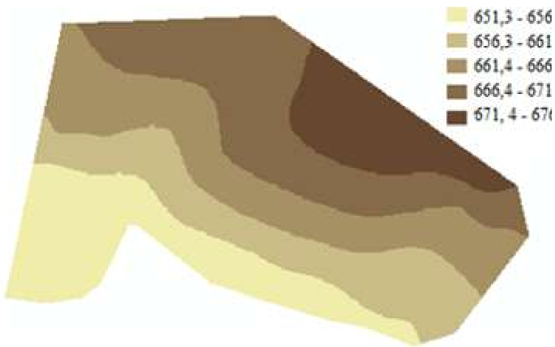
Fig. 1. Thematic map of world apple production in 2009

Source: Carvalho (2011).

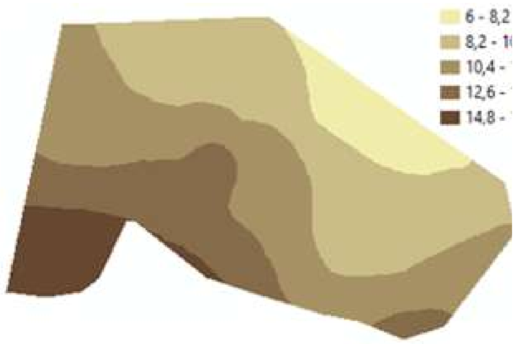
One specific case of TMs is contour maps built by connecting points of the same value and applying them to geographical phenomena that show continuity in the geographic space. Another is choropleth maps that use color to show ranges of a specific variable within a defined geographic area. Contour and choropleth maps can be built from categorical data (elevation, temperature, precipitation, humidity, and atmospheric pressure) or relative data (density, percentages, and indexes). Fig. 2 shows examples of contour and choropleth maps.



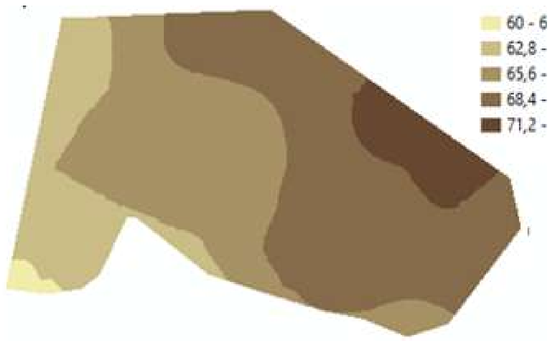
a) Elevation



b) Elevation



c) sand



d) Clay

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

To construct TMs about attributes collected in agriculture fields, it is necessary to follow a protocol like the one presented in Fig. 3.

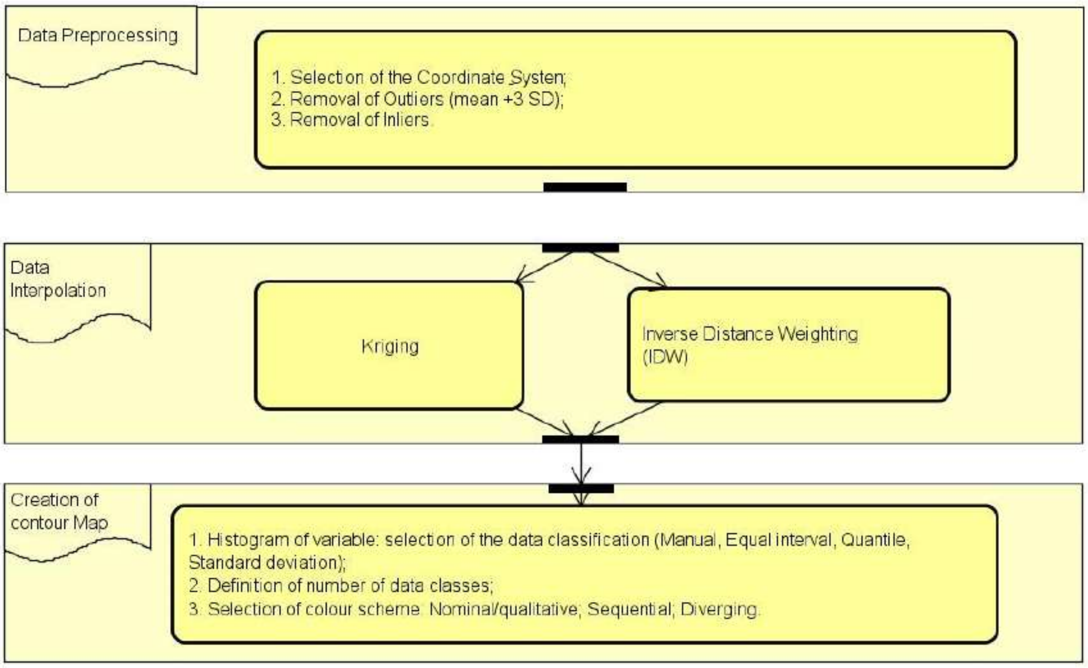


Fig. 3. Flowchart of the typical protocol to create a thematic map

I. Selection of the coordinate system - A geographic information system (GIS Software) is designed to store, retrieve, manage, display, and analyze all types of geographic and spatial data. To construct 2-D TMs is necessary a GIS software, and a file with at least three columns representing the X (longitude) and Y (latitude) coordinates and 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). The GCS is associated with a model of the Earth shape (reference ellipsoid) called a datum. The datum WGS84 (World Geodetic System 84) is most commonly used. The units are in degrees, minutes, and seconds with GCS and meters for UTM.

II. Data normalization - The normalization of variables is interesting when one wants to construct and compare TMs of a variable that has been measured several times. This is the case of the yield of 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 the data set through 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 and 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 the 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 the data distribution. According to Córdoba et al. (2016), the 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, the upper and lower limits should be modified to obtain robust variance estimators. Also necessary is the removal of inliers, data that differ significantly from their neighborhood but lie within the variation range of the data set (Córdoba et al., 2016). For yield data obtained with yield monitor, additional care should be taken. Many approaches for yield data cleaning were already being proposed (like by Blackmore and Moore, 1999) to eliminate errors associated with unknown header width, combine filling/emptying times, the time lag of grain through the combine, positional errors, rapid velocity changes, and others (Sudduth and Drummond, 2007). Vega et al. (2019) proposed a protocol for automating 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 the 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 with the aid of interpolation methods. The inverse distance weighting (IDW) and kriging are the interpolation methods commonly used in PA. They are differentiated by how the weights are assigned to the different samples, influencing the estimated values (Reza et al., 2010). Various software packages are available for performing 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 the data needs to be found through cross-validation. This technique compares theoretical values with those obtained from sampling and then analyzing the estimation errors and choosing the best model (Arlot and Celisse, 2010; Kohavi, 1995). Faraco et al. (2008) considered cross-validation a better way to evaluate the adjustment of theoretical spatial models than Akaike's and Filiben's information criteria and the maximum logarithm value of the likelihood function. With cross-validation are calculated the following measures: the average error (AE), the reduced average error ($\overline{RE}$), the standard deviation of the average error (SAE), and the standard deviation of the reduced error (SRE) (Cressie, 1993; McBratney and Webster, 1986). According to non-tendentiousness criteria, to choose the best-adjusted model, values for AE and $\overline{RE}$ should be as close to zero as possible, the value of SAE should be as small as possible, and the value of SRE should be close to 1 (Cressie, 1993; McBratney and Webster, 1986). Because cross-validation makes it possible for ambiguous situations to occur, Souza et al. (2016) proposed the error comparison index (ECI, Equation 1). As lower ECI is, the better the semivariogram is.

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

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 question 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.

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

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

V. Creation of TMs – after the data interpolation, to draw TMs with our data, we must decide both the number of classes and the method for breaking the data into ranges. The goal is to group similar observations and split apart substantially different observations (Indiemapper, 2016). The first thing to do is looking at the histogram (or scatterplot) to determine the 'form' of your observations. This critical step of map creation and how we do that can dramatically change the look of the map, and thus, its message, and 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, an accurate map must tell white lies (Monmonier, 1996). They are many ways to classify data systematically and each GIS software will offer some of them. The most popular are (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 others do not give a good solution. A good way is to start with one of the standard classifications and make adjustments as needed;
- Equal interval: we divide the 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 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

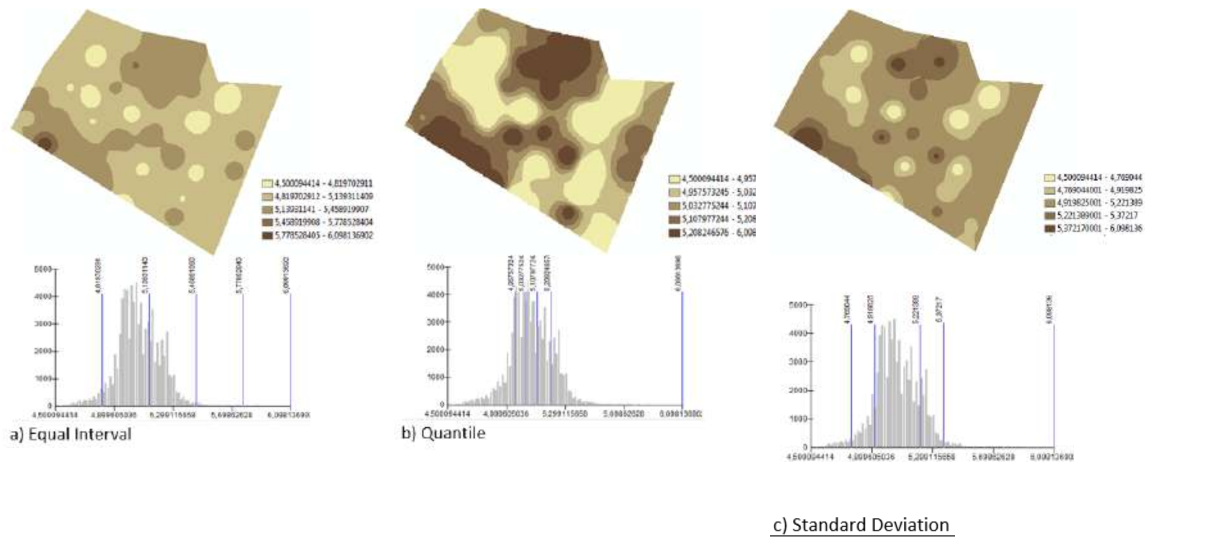
features placed in adjacent classes, or widely different values put in the same class;

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

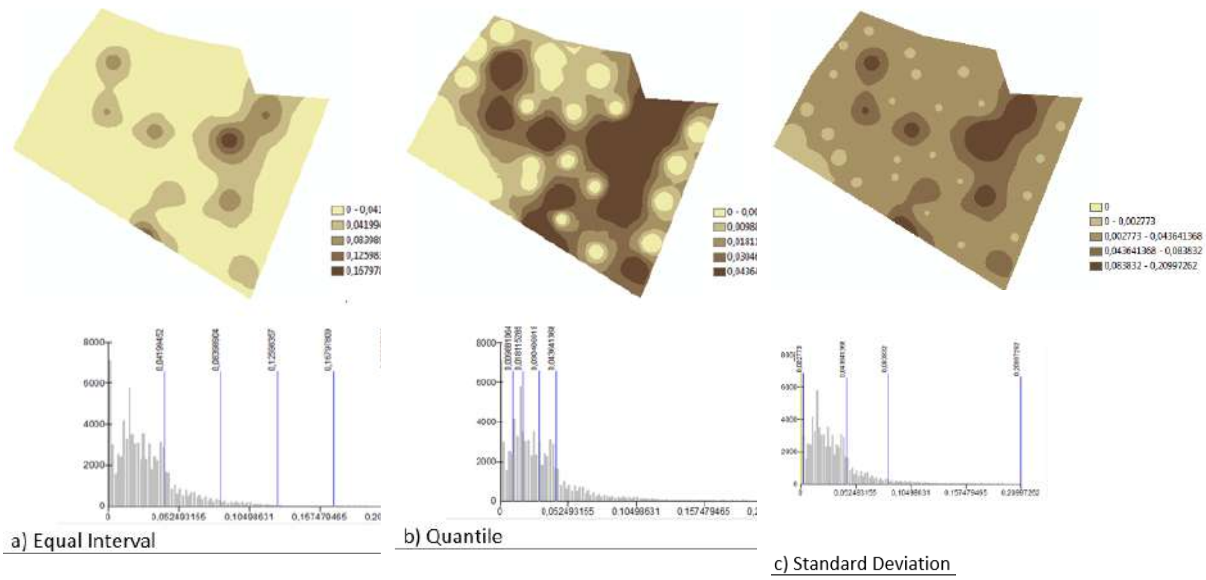
The number of data classes is also an essential part of map design. Increasing the number of data classes will result in a more revealing map but require more colors. Generally, it is advised not to exceed seven classes.

Examples of choropleth maps are presented in Fig. 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 (Fig. 4a), we have 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 (Fig. 4b), the distribution is moderately skewed right, and then the quantile is visually the best option.

After we selected how to classify the data, choosing an effective color scheme for the TM is crucial. A good color scheme needs to be attractive but also support the map's message and be appropriately matched to the nature of the data (Harrower and Brewer, 2003), being necessary to choose three dimensions of color: hue, lightness, and saturation. There are three kinds of color scheme: **nominal/qualitative** (unorderable data, like land use, Fig. 5a): different hues that keep lightness and saturation constant should be used; **sequential** (orderable, like numerical data (or low/med/high), like yield, Fig. 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 profit, Fig. 5c). Harrower and Brewer (2003) designed an online tool “ColorBrewer.org” to help users select appropriate color schemes for their specific mapping needs. Fig. 6 presents some practical examples of the application of the color schemes.

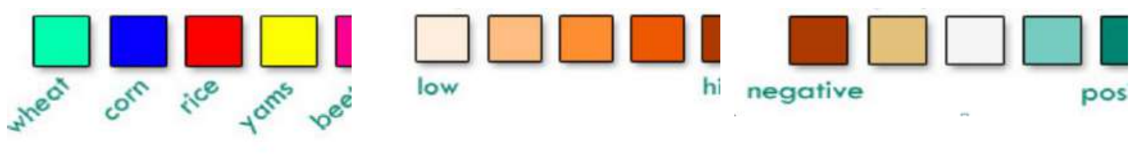


a) Ph



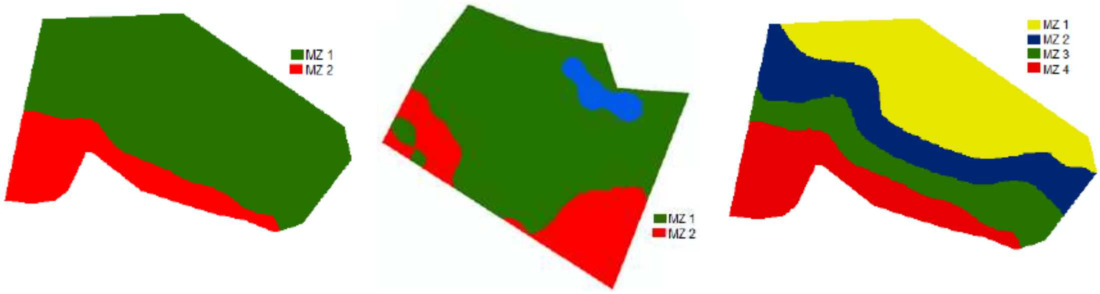
b) Aluminum

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



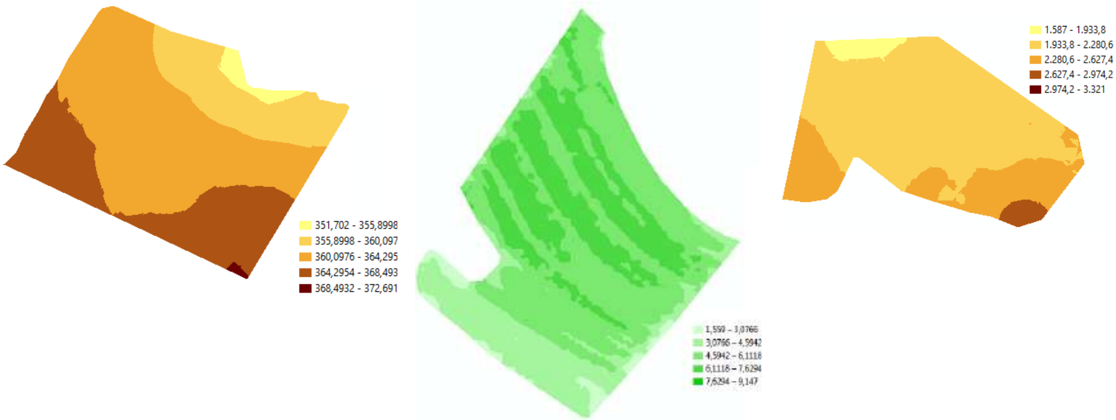
a) Nominal Color Scheme b) Sequential Color Scheme c) Diverging Color Scheme

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



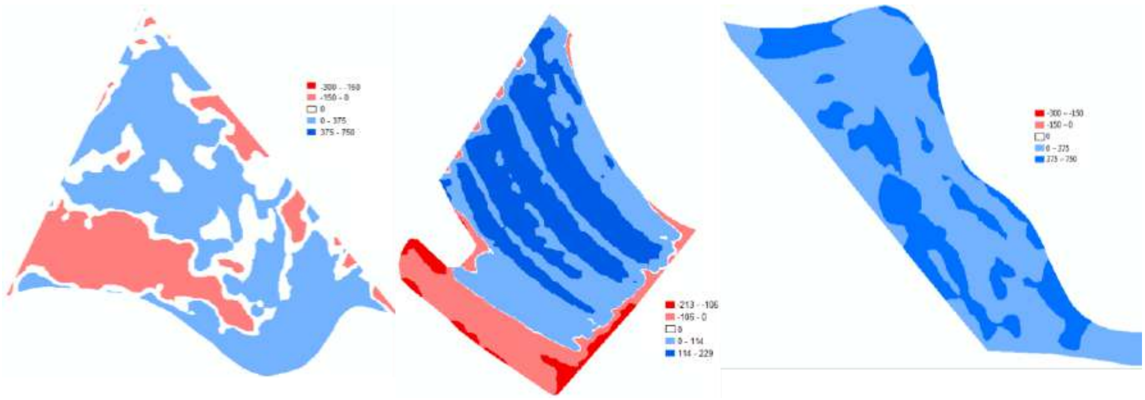
a.1) a.2) a.3)

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})

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

Contour maps using a continuous scale - despite being most common using a discrete scale, some people prefer continuous scale. The problem with a color ramp is that perception of color intensity is not linear, and consequently, the user could make a false assumption about what data value it represented. Basso et al. (2009), studying 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, presented an interpolated map of wheat yield (Fig. 7) using a continuous scale.

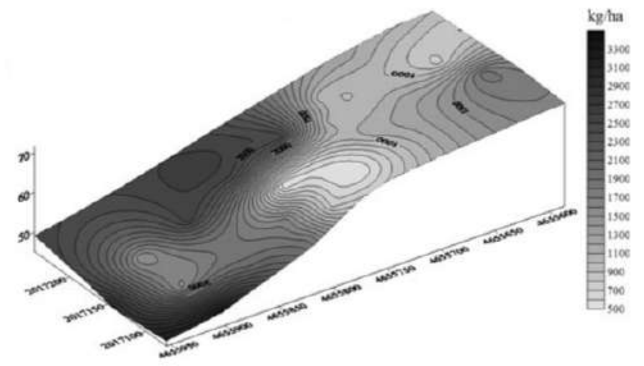


Fig. 7. 3D interpolated map of wheat yield (kg ha^{-1}) for 2003

Source: Basso et al. (2009).

2.1. Examples of thematic maps

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

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 in two experimental areas (field A - 10.0 ha, and field B - 23.8 ha) and two agricultural years (2012/2013 and 2013/2014.). Fig. 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.

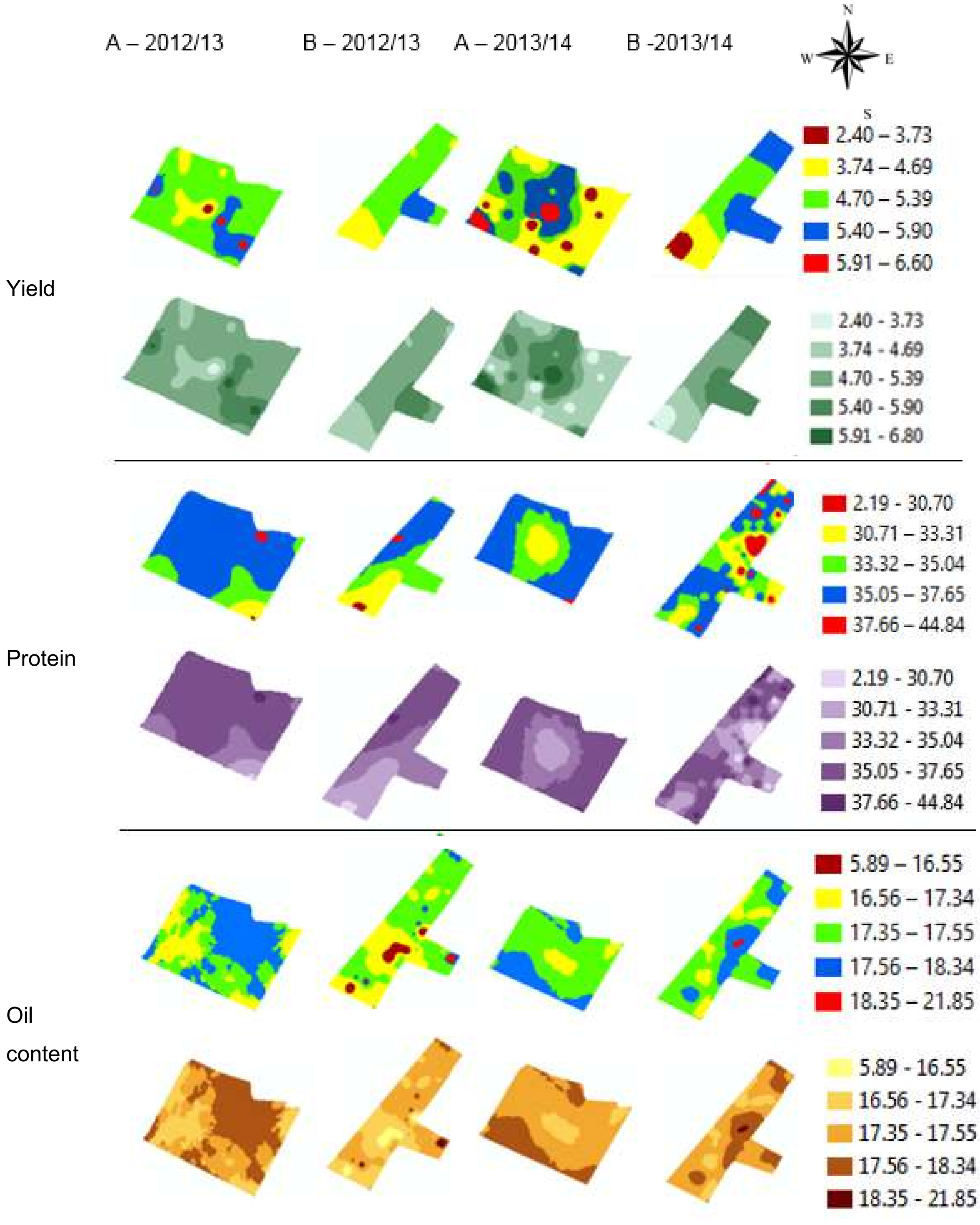


Fig. 8. Thematic maps of soybean yield and the corresponding protein and oil content for fields A and B in 2012 and 2013, using a sequential (different colors) and nominal (single-hue with different lightness/saturation) color scheme

Source: adapted from Silva (2016).

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 (Fig. 9) were generated using the following interpolation methods: inverse of the distance (ID), inverse of the square distance (IDS), and kriging (KRG). They concluded that profit and profitability maps are important tools for the diagnosis of spatial variability of economic return because they assist 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. Fig. 9 shows that there are in this 45-ha area 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.

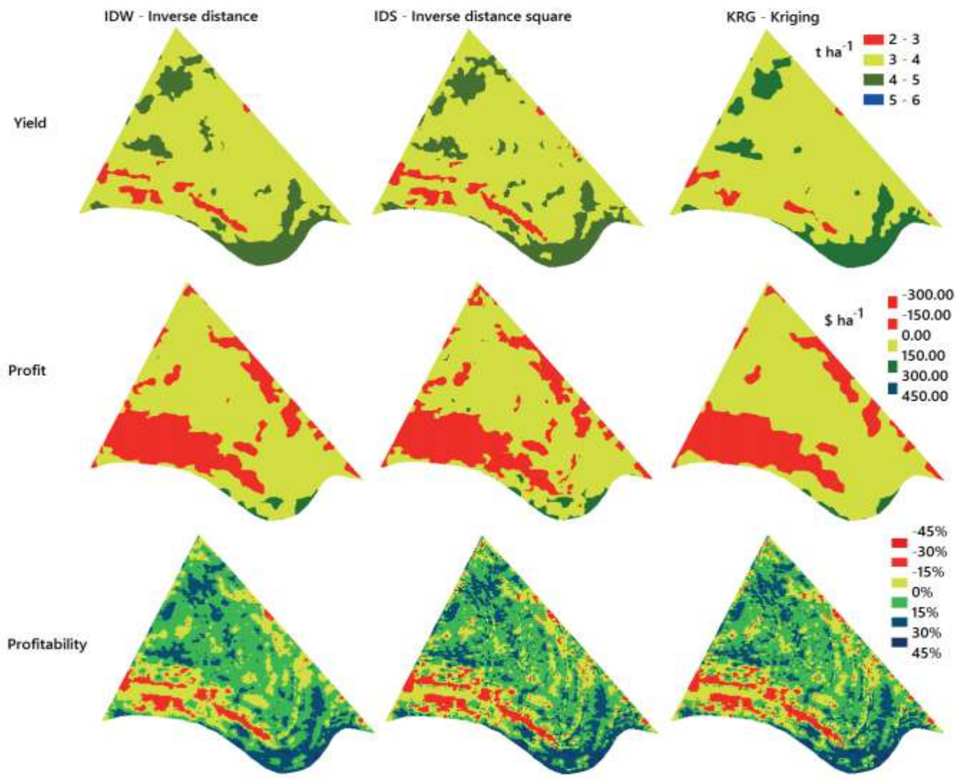


Fig. 9. Yield, profit, and profitability maps for the 2006 Soybean harvest using the interpolation methods (i) inverse distance weighted (IDW), (ii) inverse distance weighted squared (IDS) and (iii) kriging (KRG). The production cost and sale prices of the product were obtained in the harvest month in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

Source: Bazzi et al. (2015).

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 (Fig. 10) in Spain, at the wine farm of Raimat (Lleida).

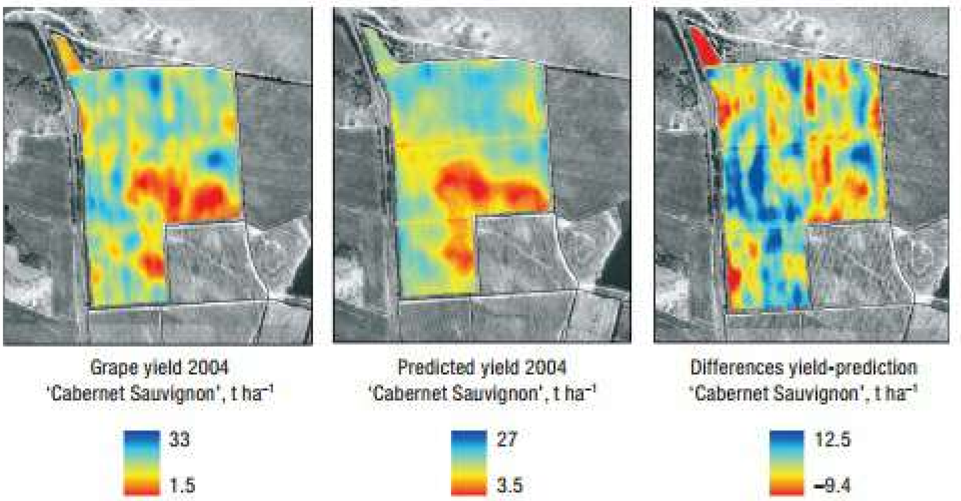
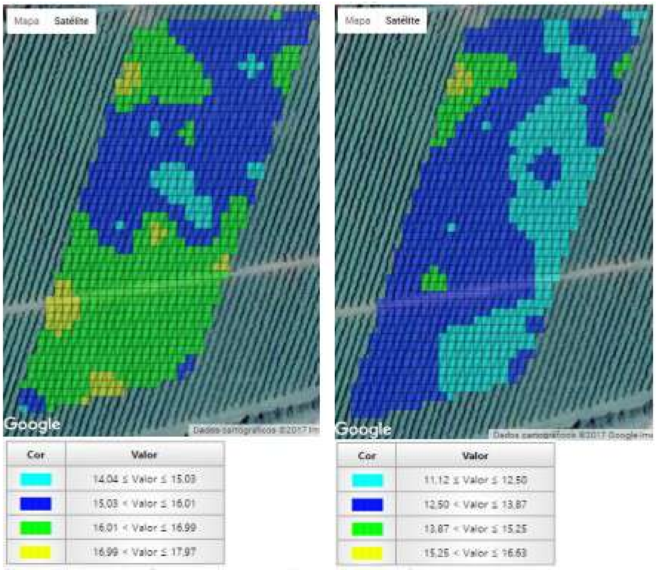


Fig. 10. Comparison of the 2004 yield map of a 'Cabernet Sauvignon' plot (left) with the map obtained from a prediction model using the NDVI (normalized difference vegetation index) from a QuickBird-2 multispectral image acquired one month before harvesting (center) (R^2 = 0.72). The map on the right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

Source: Martínez-Casasnovas and Bordes (2005).

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. Fig. 11 presents the firmness of fruit pulp and total soluble solids of the fruits in an area of 3.13 ha.



a) b)

Fig. 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).

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. Fig. 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

	High infestation
	Average infestation
	Low infestation

Fig. 12. Weed maps showing three infestations levels from a 72-ha flat terrain, planted in the no-tillage system and with soil covered by soybean stubble

Source: adapted from Balastreire and Baio (2001).

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 pasture. The study was conducted in a 5.3-ha irrigated alfalfa pasture 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 for determining productivity and profitability of grazing systems, and Fig. 13 illustrates that the simulation based on dry matter yield allowed estimation of stocking rates and milk yield within the area. Therefore, maps of this type may be used to avoid over-or under-grazing. In addition, this study showed the methodology's advantages that allow the identification of areas for differentiated paddocks management instead of homogeneous fertilizer application.

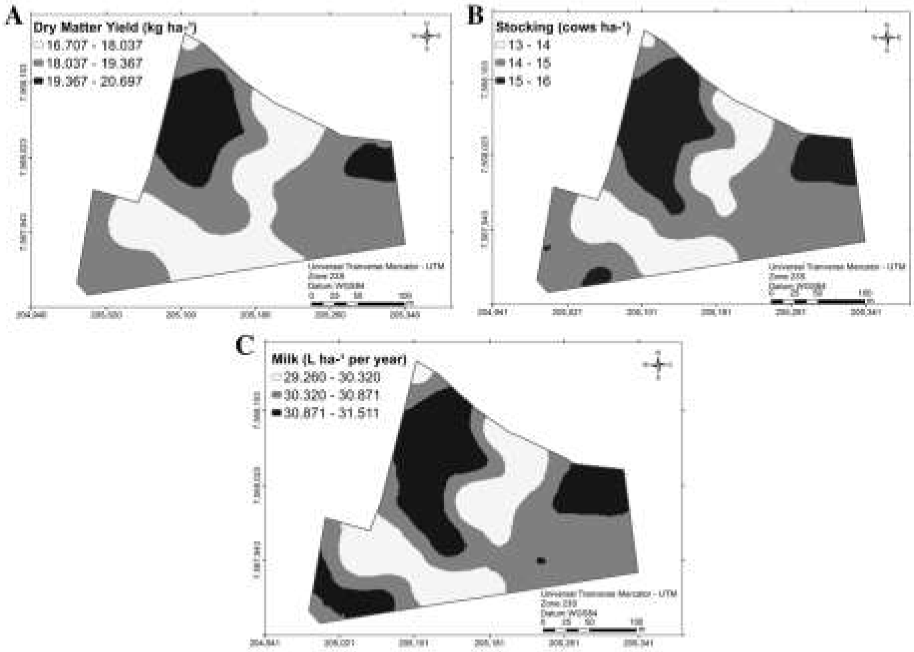


Fig. 13. Kriged maps for dry matter yield (a), stocking rate (b), and milk yield (c) of a grazed alfalfa pasture in Brazil

Source: Bernardi et al. (2016).

3. 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 or several years (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 the MZs for multiple 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 the delineation of MZs in the current context of PA, we made a systematic literature study (SLS) that had as the primary focus to identify researches about the delineation of MZs, 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.

3.1 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: definition of keywords, choice of scientific databases, determination of study selection criteria, study analysis, and synthesis methodology (Kitchenham and Charters, 2007; Talavera et al., 2017).

To define the keywords, we raised the following questions: (1) what are the procedures and protocols for the delineation of MZs? (2) what are the most common algorithms for the delineation of MZs? (3) how to find the ideal number of MZs classes? (4) what are the economic or environmental advantages of the adoption of MZs? (5) what software is used for delineating MZs? To obtain access to scientific databases, we used the portal of the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Coordination for the Upgrading of Higher Education Personnel, CAPES) through the remote access platform of the Comunidade Acadêmica Federada (Federated Academic Community, CAFe) (Fig. 14). We searched four databases considered relevant to this research area: Scopus, Science Direct, Web of Science, and Wiley.

The period covered by the SLM searches was limited from 2008 to 2019 to present the most recent articles. However, the SB had rescued relevant researches left behind. The standardized information extracted from all articles 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 the classification process by consulting the references of the selected studies and sharing 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 the efficiency and quality of the search, reducing the chances of obtaining a search bias (Cohen and Arieli 2011) and rescuing important classic texts referring to the period before 2008. It is important to note that no survey method is 100% effective, but the combination of both techniques is expected to reduce omission problems.

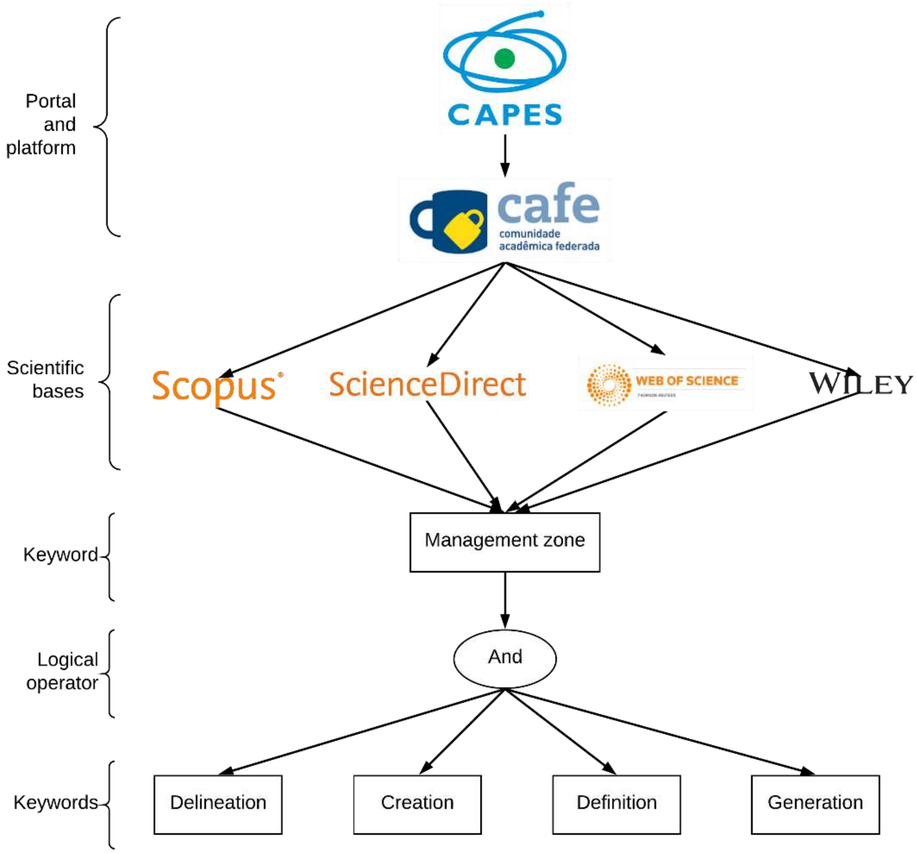


Fig. 14. Workflow used with the Systematic Literature Mapping (SLM)

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

The flow of selection of works used in the SLS is shown in Fig. 15 and Fig. 16. Table 1 presents the 165 studies selected by type of research technique (SLM, BSB, and FSB). Only ones directly related to agriculture (excluding, for example, those related to forestry or geological management) and which have explained the process of delineation of MZs were kept. Studies that delineate management zones 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

Technique	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 (FSB) 51 *references*	(Biernacki and Waldorf 1981; Rousseeuw 1987; Webster 1990; Blackmore and Moore 1999; Khosla and Alley 1999; Blackmore 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 (Fig. 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)Fig. 17). We observed a smooth growth tendency until the year 2013, which presented the largest number of studies (19). In 2014, there was a drop (8 studies), but with a tendency for growth in later years.

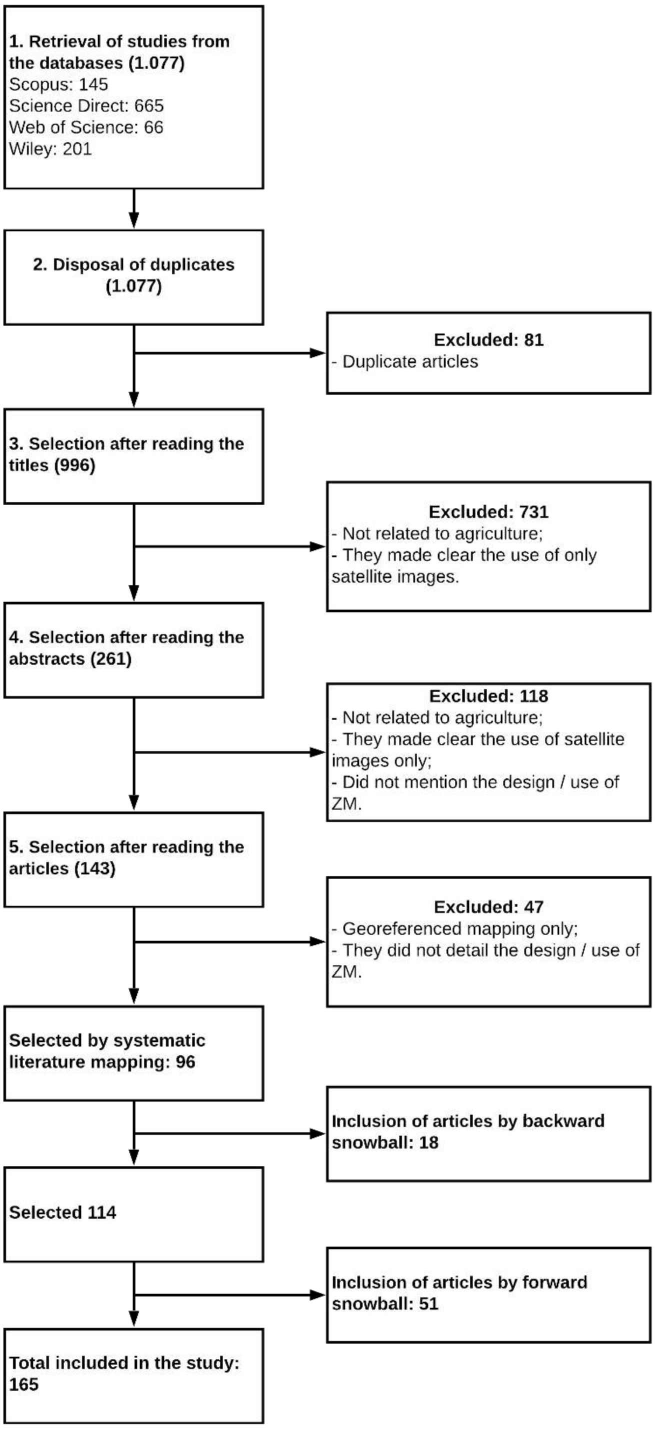


Fig. 15. Workflow used for the Systematic Literature Study (SLS)

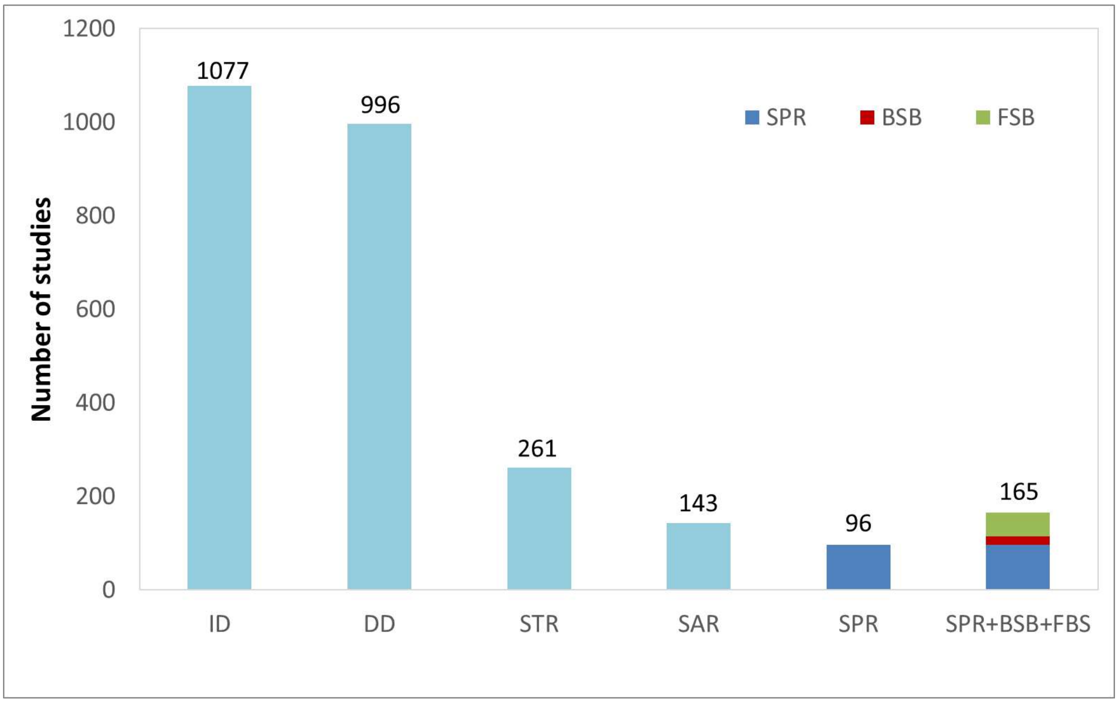


Fig. 16. Stages followed in the systematic literature study (SLS) to select the primary papers: Identification (ID); discarding duplicates (DD), selection by title reading (STR); selection by abstract reading (SAR), selection by paper reading (SPR), adding by Backward Snowballing (BSB), and adding by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

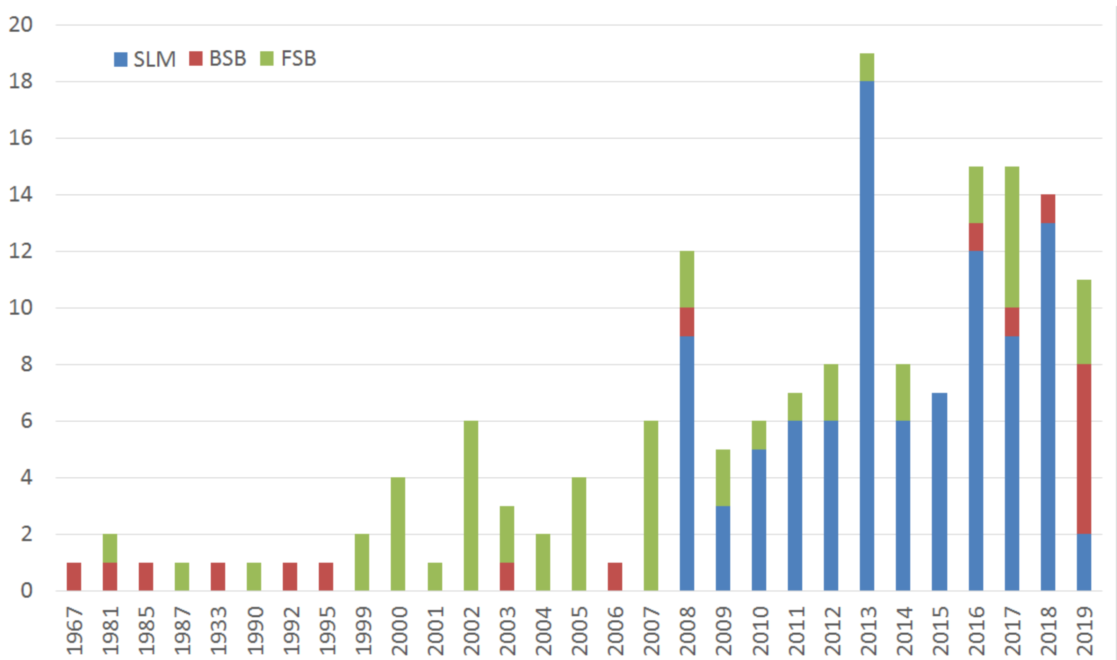


Fig. 17. Quantity of selected studies classified in chronological order and type of research technique, Systematic Literature Mapping (SLM), Backward Snowballing (BSB), and Forward Snowballing (FSB)

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 (Fig. 19), the journals *Computers and Electronics in Agriculture* and *Precision Agriculture* presented the most studies selected, with 17 and 14%, respectively. The journals and publishers that showed less than three studies in the review were clustered into a single item “others”.

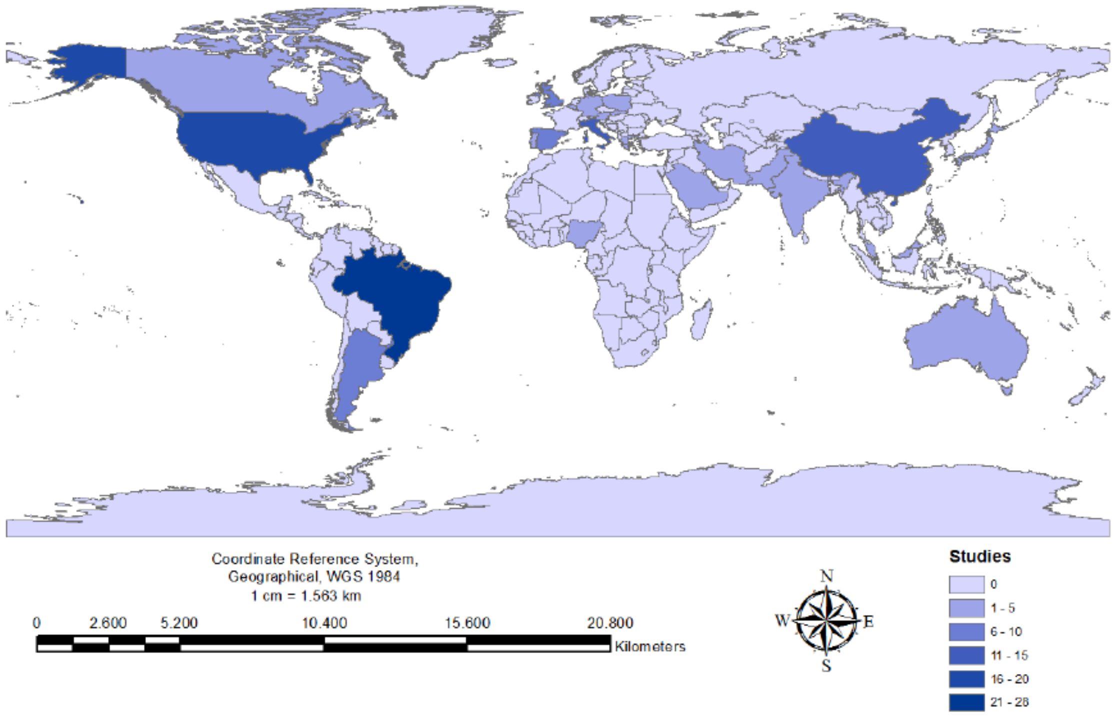


Fig. 18. Distribution of selected studies by country of study (classified by the country where the authors conducted the research, and when is only a theoretical manuscript, where it was published)

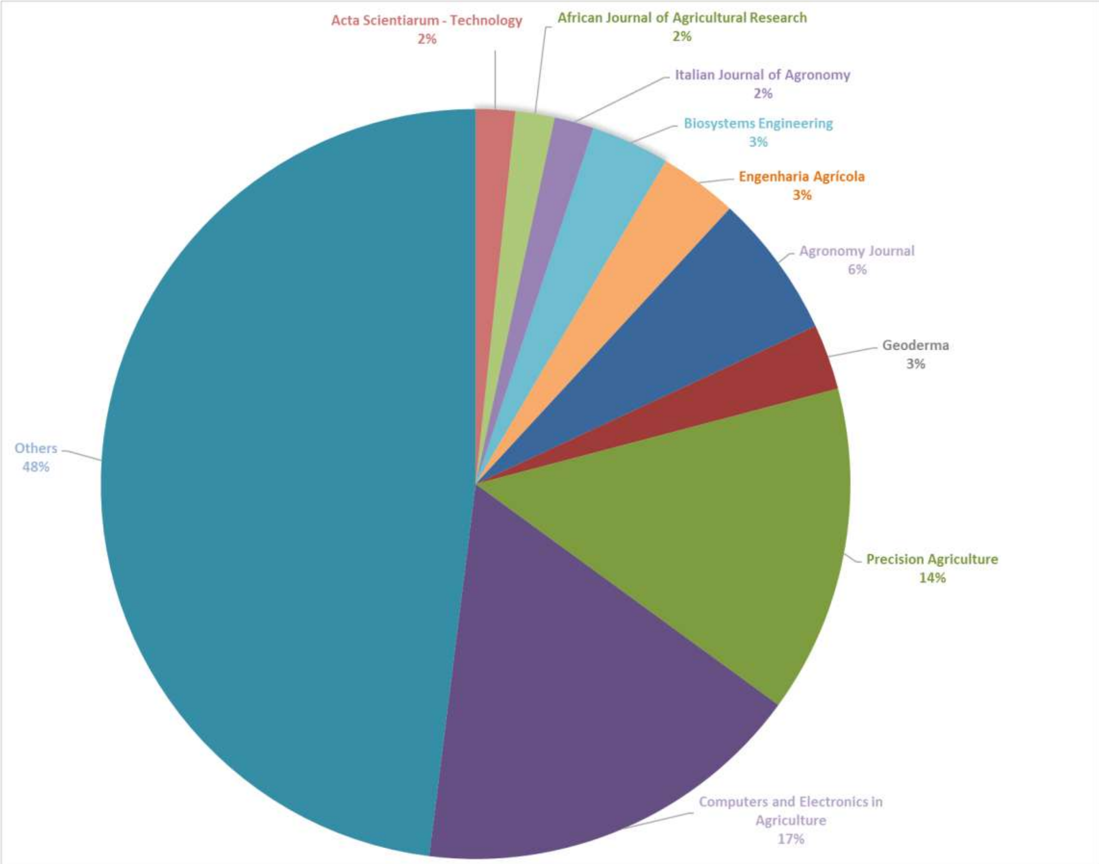


Fig. 19. Distribution of selected articles by journals in which the selected studies were published

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

3.1.2 Results and discussion of the 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 the delineation of management zones (Question 1)

Much effort has been made and is being made in defining the best delineation process for MZs. While some studies focus on creating a protocol that encompasses 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 four studies that define a complete protocol for delineating MZs, but only one considers temporal issues. The first, 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 containing codes ready for executing each of the steps is also available.

Souza et al. (2018) presented a more specific protocol, divided into nine main stages: (1) selection of the coordinate system, (2) remotion of the outliers and inliers, (3) data normalization, (4) variable selection which will be used for delineating MZs, (5) data interpolation, (6) delineation of MZs, (7) rectification of the MZs, and (8) selection of the optimal number of MZs, and (9) evaluation of the MZs. Although there are subtle differences between 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 Fig. 20.

Differently of three other protocols, the one outlined by Scudiero et al. (2018) takes into account variations between the soil-plant, consisting 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 the MZs. The authors comment that traditional

MZ delineation methods create static zones that are not ideal since the spatial patterns of the soil-plant relationship change over time due to weather changes and/or other transient factors.

In addition to previous efforts to define a complete protocol for delineating MZs, some authors addressed specific issues at each stage, that is, they perform 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 for delineating MZs are related to:

- Farmer knowledge – this knowledge may allow the identification of 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).

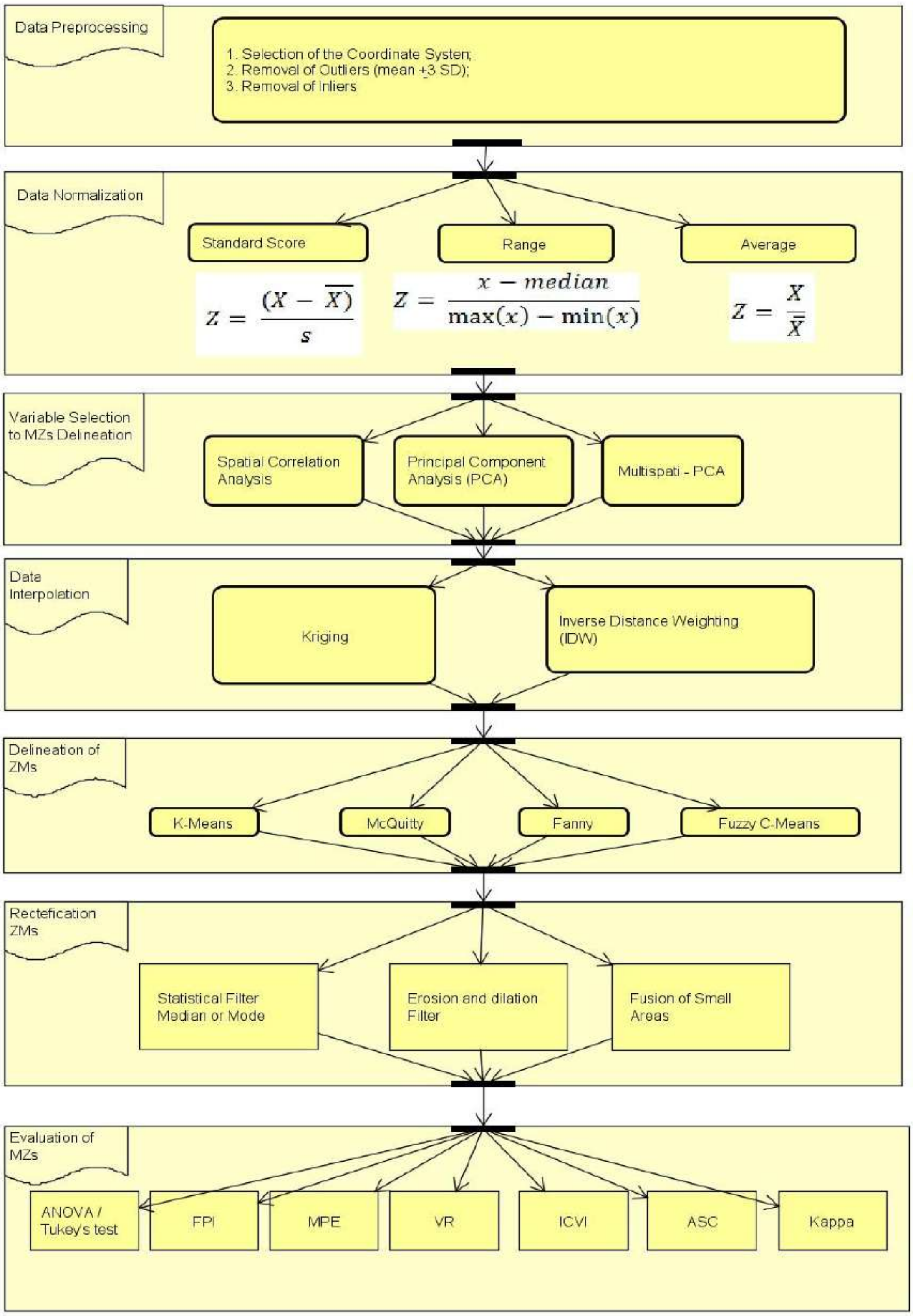


Fig. 20. The protocol of the delineation of management zones, according to Souza et al. (2018). ANOVA: analysis of variance, SD: standard deviation, MZ: Management Zone, SD: standard deviation, ANOVA: analysis of variance, FPI: Fuzziness Performance Index), MPE: Modified Partition Entropy, VR: variance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

- Geomorphology – elevation is the most used topographic variable to delineate MZs. However, other variables like 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, like 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 insufficient to reliably identify crop productivity MZs since in a zone with the same soil series, many other variables can influence yields (like 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 for developing Nitrogen-MZs.
- Yield maps – they are the complete information to visualize the spatial variability of crops (Molin 2002). However, its temporal variation complicates using a single-year yield map to delineate MZs reliably. Blackmore (2000) and Molin (2002) used normalized data from multiple years to compensate for this problem. Although one-year yield data alone are not directly suitable for MZs determination, their availability and low cost make them a valuable possibility for improving the effectiveness of MZs delineation based on other information (Nawar 2017). Two approaches are commonly used for delineating 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) 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 is vegetation indices (VI) and leaf area index (LAI). Both can be measured manually and using remote sensing (RS) methods. Traditional methods for acquiring 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 delineated MZs based on RS data alone (Inman et al. 2008; Song et al. 2009; Chang et al. 2014) or for improving the 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 resulting in a regular grid of sample points, producing 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 the soil (ECa) and bare-soil NDVI) data and the FCM algorithm to divide a 21-ha cornfield into five zones. The authors commented that even when measurements like ECa and bare-soil NDVI are not directly correlated to the corn yield, their combined use could help classify the soil according to its fertility.

2. *Remotion of the outliers and inliers*: exploratory data analysis (EDA) is the summarization of the data set through 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 the data and, hence, 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. Also necessary is the removal of inliers, data that differ significantly from their neighborhood but lie within the variation range of the data set (Córdoba et al., 2016). For yield data obtained with yield monitor, additional care should be taken. Many approaches for yield data cleaning were already being proposed (like by Blackmore and Moore, 1999) to eliminate errors associated with unknown header width, combine filling/emptying times, the time lag of grain through the combine, positional errors, rapid velocity changes, and others (Sudduth and Drummond, 2007). Vega et al. (2019) proposed a protocol for automating 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 the 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 like the 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 demonstrating equal variances. In this sense, when the Euclidean distance is used to clustering, the 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) for the delineation of MZs with the 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 the MZs. According to Gnanadesikan et al. (1995), the weighting and selection of variables is the most challenging issue 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 the clustering (delineation of MZs) (Gozdowski 2014, Sobjak 2016). Sobjak et al. (2016) showed that with the 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 the performance of the clustering and increases the computational time (Bazzi et al. 2013; Schenatto et al. 2016a; Sobjak et al. 2016). Good results were obtained with multivariate techniques for reducing the dimensionality of variables and promoting orthogonality between them (Hotelling 1933; Dray et al. 2008; Gavioli et al. 2016). Three variable selection techniques (Table 2) that are most used in combination with the 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 95% significance; (2) removal of the 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 for delineating MZs. These PCs are linear combinations of the original variables resulting from transformations that prioritize the representation of the data variability in the first components. Thus, if the original variables have a high degree of dependence between them, one can reduce the dimensionality of the data using the first PCs. Another possibility is to select only the variables that had the most significant influence on the PCs for the 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 the PCA which adds spatial restriction considering data georeferencing (and, therefore, the 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, the total variance.

Table 2. Main techniques for the selection of variables for delineation of management zones (MZs)

Selection technique	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, based on the combined use of spatial correlation analysis and MULTISPATI-PCA, and 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 the MPCA-SC provided the best performance for the FCM algorithm, reducing the dimensionality of the data 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 after. These are the cases of:

- 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 the best combination of ECa and/or elevation data combinations. They considered this level of agreement (until 60–70%) promising, especially considering many other yield-limiting factors unrelated to ECa and elevation.
- Gozdowski et al. (2014) used logistic regression to find witch soil attributes were more correlated with the MZs delineated thought the multiyear mean standardized yield divided each field into two zones, one above and one 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. Using the Kappa coefficient, they concluded that ECa provided the best agreement.
- Miao et al. (2018) evaluated three approaches to delineating 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:* usually, the data to delineate MZs is 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, the IDW presents an advantage (Betzek et al. 2019).

6. *Delineation of the MZs* - Two approaches are commonly used for delineating MZs: (1) the empirical method, which uses frequency distribution of target variable (usually yield) to divide the field MZs (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 are intended to divide the 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 the field (Boydell and McBratney, 2002; Córdoba et al., 2016).

7. Rectification of the MZs: After their delineation, the MZs often present isolated pixels, small regions, or even a transition border between very irregular zones, making it difficult or even impossible to operate in the field. In this sense, 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 the fragmentation of MZs.

8. Evaluation of the delineated MZs: the performance of the 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 study field (important for MZ delineation or the similarity between 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 the 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 delineation of management zones (Question 2)

Many techniques and algorithms are available for each stage of the delineation of MZs. 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 used (Table 3).

The 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 maximizing variability between them, seeking 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, reducing the possible distortion caused by outliers.

Table 3. Algorithms used for the delineation of management zones (MZs)

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 the FCM (McBratney and Moore 1985). The first, matrix **X**, consists of the data to be classified. The second, Matrix **V**, is the matrix with the centroids of the clusters, consisting of k centroids of clusters contained in the search space defined by matrix **X**. The third, matrix **U**, consists of assigning the pertinence value of each cluster in **V** for each point in **X**, considering that the sum of the 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:

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

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

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

Table 4. Values adopted for the fuzzy weighting coefficient in the 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 normalization of the data) of the variables when two or more input variables are used (Bezdek 1981; Fridgen et al. 2004, Schenatto et al. 2017b). The 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 the 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 the 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 the matrix **X** and calculates the distance of all points (typically using the Euclidean distance) to the centroids and assigns the location to the nearest centroid. That is, considering $x_j \in X$, it is associated with the 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 between the points part of that centroid.

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

We must consider that the 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 20 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 the ISODATA, FCM algorithm, and a density-based non-parametric clustering method for delineating MZs in wheat. They considered that, although ISODATA presents a lower computational cost and a better visual distinction of the MZs, the lack of information on the transition areas was a problem.

Gili et al. (2017) stated that the choice of the ideal algorithm of MZ delineation depends on the objectives of the use of the MZs. In their research on corn, they used the 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- the FCM using directly on the soil variables (Clay + silt, OM, pH, ECa, and organic matter index) data, and (3) S3- the FCM using the first three PCs. The different strategies resulted in a different number of zones with different characteristics: for fertilization management zonification might prioritize the differentiation of OM and available P contents and use S3; if water were the main limiting factor, the management zones would be two according to S1 or S2, responding to textural and altimetry differences.

Boluwade et al. (2016) analyzed de delineation of irrigation MZs, employing ECa and elevation, with the FCM and the 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 ha, used Ward's method (to obtain information on the clusters like the ideal number of clusters) followed by the k-means clustering method. The delineated MZs had more quality and information on the 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), having been tested on wheat data, demonstrating its viability (Ruß 2013).

Another possibility is the modification of algorithms that were not initially developed for the delineation of MZs. 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 its ease of operation in the 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 the 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, one must define 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 of the performance of the clustering process 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's test. Secondly, it is calculated performance indices.

Table 6 presents several measures (showing only 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.

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

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

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

The FPI, NCE, and MPE measures (Bezdek 1981; McBratney and Moore 1985; Odeh et al. 1992) are strongly connected to the 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):

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

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

Table 6. Most used measures for choosing 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 level of difficulty in organizing the c clusters, with values close to 0 indicating low difficulty in organizing the clusters. It can be defined by (Equation 7):

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

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

Table 7. Combination of most used measures for choosing the number of management zones (MZs)

Measure	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 modification of the relative variance (RV) 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):

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

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

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

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

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

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

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

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 the 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 diagonal right ; and P_{DL} is the possibility of changes in the diagonal left .

The 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 the intra-group distances ap and the average of the inter-group distances bp (Equation 11).

$$sc_p = \frac{b_p - a_p}{Max(a_p, b_p)} \tag{11}$$

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

The group silhouette coefficient (GSC) is obtained by calculating the average of the silhouette coefficients for the points of this group, and the value that corresponds to the ASC coefficient of k groups is obtained by calculating the average of the GSC values of the k groups. The 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.

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

Despite the 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's test (ANOVA), can be used regardless of the algorithm used for the delineation. It is expected that, with the increase in research related to other algorithms for the delineation of MZs, 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 the FI% will be 25%. The higher the fragmentation of delineation, the higher the FI% (Equation 12):

$$FI\% = 100\frac{NMZ - NC}{NC} \tag{12}$$

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):

$$GQI_i = \frac{ICVI_i * (100 + FIr_i)}{SIr_i} \tag{13}$$

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: $0 < K \leq 0.2$ indicates no agreement, $0.2 < K \leq 0.4$ weak agreement, $0.4 < K \leq 0.6$ moderate agreement, $0.6 < K \leq 0.8$ strong agreement, and $0.8 < K \leq 1$ 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):

$$CRD = \sum_{i=1}^{n} ABS\left(\frac{Zi_B - Zi_A}{Zi_A}\right) \tag{14}$$

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

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

$$MAD = \frac{\sum_{i=1}^{n} ABS(Zi_B - Zi_A)}{n} \tag{15}$$

Possible economic or environmental advantages of the adoption of management zones (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 resources in the field better, 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 the MZs delineated to validate the zones division. One way to do this is with ANOVA where the target variable (usually yield) is compared between classes by using the average target variable and performing the Tukey's 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, one sees that only 46 (approximately 28%) of the selected papers did a consistent statistical analysis (ANOVA) to validate the existence of considerable differences between 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 with the adoption of 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 the 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 through our spatial procedure above (VRN, developed at this work), and (4) variable-rate N application based on landscape position (VRL). Their results demonstrated that the VRN application could result in net returns over US$ 5.28 ha^{-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 producer fields cultivated with corn, Roberts et al. (2012) developed MZs for the N control. Two areas showed no correlation between yield and N, while in the other four, they found 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) found three as the ideal number of classes. The class with the highest profitability had a profit of 12% 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. The SH of the high and low-quality grain categories would result in a reduced risk of mycotoxin contaminated grain, reaching a human consumer. This study was restricted to a single field but demonstrates the potential of fungicide reductions and the economic viability of the MZ concept.

Schwalbert et al. (2018) compare four different wheat fertilization strategies two producer 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 for delineating 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 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, which have specific advantages and disadvantages. For a correct understanding and analysis, they often require the knowledge of several areas, creating great difficulty for adopting MZs in agriculture.

Some of these difficulties can be reduced by using specialized software. Despite many software for the PA, few are directed to delineating MZs (Table 8). Golden Software Surfer, ESRI ArcGIS, and the R software package are commonly used. Despite allowing the delineation of MZs, they do not have all the desired functionality since this is not the focus of these products, requiring 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 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 for delineation of MZs, the following were well-known (organized by release date): (1) FuzME (Minasny and McBratney 2002), Management Zone Analyst (MZA) (Fridgen et al. 2004), (2) Software for the Definition of Management Zones (SDUM) (Bazzi et al. 2013; Bazzi et al. 2019), (3) ZoneMAP (Zhang et al. 2010), and (4) automatic software for delineating MZs proposed by Albornoz et al. (Albornoz et al. 2018).

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

Name	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 of Sydney	https://sydney.edu.au/agriculture/pal/
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 the 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.

The FuzME is a 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 algorithm is the FCM (with a few variants), and the outputs are all in text files. The software features a simplified graphical interface, consisting essentially of three toolbars. The first presents the options for selecting the input files with the respective variables, output files, internal control files, and analysis title. The second presents the options for creating clusters, such as distance metrics and fuzzy exponents, among others. The third presents the options to allow resampling using the bootstrap and Jackknife methods. Among the possible options for adjusting the clustering algorithm are: (1) choice of the distance metric (Euclidean, diagonal, and Mahalanobis); (2) choice of the fuzzy exponent; (3) definition of the minimum and the maximum number of classes (between 1 and 100); (4) analyzing fuzzy discriminants; (5) configuration of the initial random values of the definition of the members and the number of attempts, the stopping criterion, the maximum number of iterations, and (6) choice of the 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: (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 the 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 among the specific software for the delineation of MZs (*Table 8*). 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. MZA, like FuzME, also used the FCM algorithm.

It also presents a simplified graphical interface and, to perform classification, you must follow the instructions in a sequence of four menus that present the definition of the parameters step by step. Initially (start window), you must provide the input file in CSV text format (comma-separated values), containing the variables and their values. In this same window, one or more variables to be used must be chosen. 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 for performing the classification with FCM: the fuzzy exponent; the measure of similarity (Euclidean, diagonal, or Mahalanobis); the maximum number of iterations; the convergence criterion; the minimum and maximum number of MZs; and the 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. 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 for delineating 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. Also important is that, depending on the characteristics of the input data, the resulting MZs can contain much-fragmented information, requiring the use of external software for smoothing and visualizing the MZs. A third problematic element concerns one of its main advantages: choosing the ideal number of clusters. The NCE and FPI measures 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. 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 made 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 the 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 through 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, which can be continuous surfaces or points; the interpolator parameter; and the radius parameter, consisting of the distance the 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, the fuzzy exponent, and the maximum error are defined. The SDUM also calculated performance indices (FPI, VR, and MPE), ANOVA, and Tukey's test.

Like previous software, the user-friendly interface and the definition of a clear sequence of steps for delineating MZs 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 the MZs quality (FPI, VR, MPE, ANOVA, and Tukey's test), generation of thematic maps, and the maps of the MZs, organization in the form of projects, and data storage in a database. Finally and most importantly, the 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 considerably increases the user's independence regarding the use of complementary software and in the domain of knowledge 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 the MZs. A final limitation is the need to run on computers using a specific operating system.

ZoneMap was unavailable when developing this paper, due to financial reasons, according to its developers. Therefore, it could not be evaluated.

The automatic software for delineation of MZs proposed by Albornoz et al (2018) is a software made 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 the software. The delineation algorithm is the FCM, and the outputs are in the 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.

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

On the next screen, the MZs maps are presented for each number of zones. Also, on this screen, there is a table given the three evaluation measures of the 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 one wishes to continue the process. On the next screen are the options for filtering (rectification) the map: the mask size (3x3, 5x5, and 7x7 pixels), the type of filter (medium or mode), the 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, presenting a precise sequence of steps for the delineation of 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 the MZs quality (NCE, FPI, XB, EcD**)**. Another substantial differential is the possibility of smoothing the maps generated using algorithms from digital image processing, which aim to create smoother transition edges and eliminate 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 transfer of the processing load, usually high in this type of procedure, from the user's machine to a web server. The counterpart is the necessity of a stable connection to the Internet, the availability of the server, and additional issues of security/confidentiality of the data.

As for the disadvantages, we can highlight the lack of tools to conduct statistical and geostatistical analysis. 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 the delineation of MZs.

Table 9. Features of specific software for delineation of management zones (MZs)

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, KLM	Shape file
MZ evaluation		x	x	x
Map Generation			x	x
Intuitive interface	x	x	x	x

Including new articles

After selecting articles through systematic literature and snowballing, four papers about a new web platform AgDataBox (ADB) were found (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 software SDUM (Bazzi et al. 2013). This web Platform has an Application Programming Interface (API), which consists of a set of resources accessible through the Hypertext Transfer Protocol (HTTP) for transferring request and response messages expressed in JavaScript Object Notation (JSON) format. The ADB-API, where the data and processing routines are centered, enables the interoperability of several applications. Five applications are under development: (1) ADB-Mobile, (2) ADB-Map, (3) ADB-Admin, and (4) ADB-IoT. The application ADB-Map is the application that works with spatial data aiming to create thematic maps and management zones. Among the functionalities of ADB-Map 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. Because this platform is not already available on the web, a detailed discussion of its performance was impossible.

3.2. Examples of management zones

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

3.2.1. Target values of the management zones

Yield (productivity) management zones

Usually, on MZs delineation, the 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). Fig. 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.

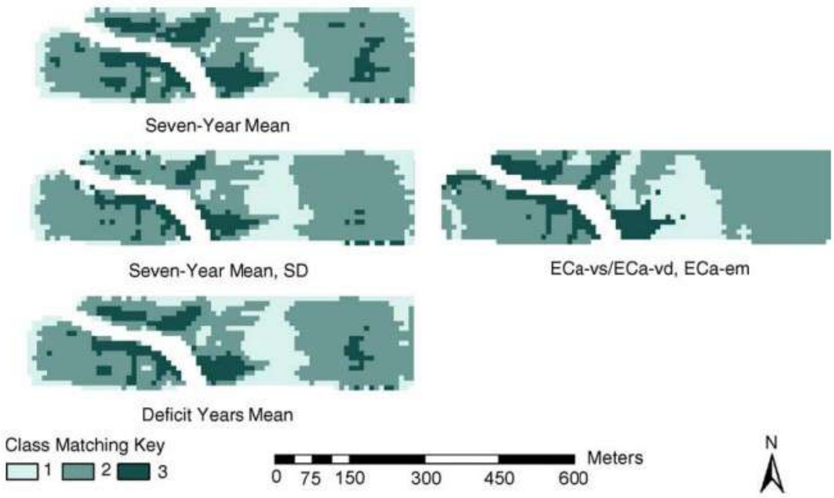
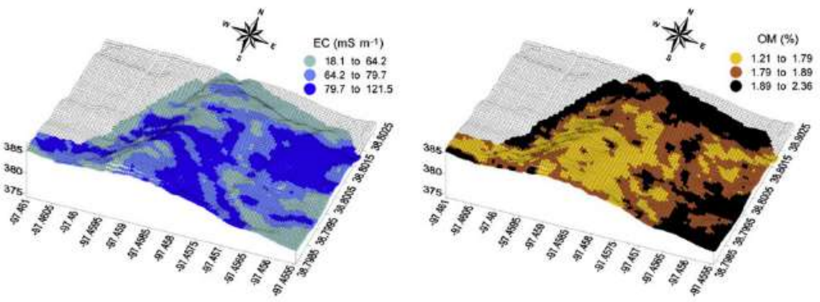


Fig. 21. Reference yield zone maps (left) compared to the best performing productivity zone map derived from unsupervised clustering of ECa and elevation (right)

Source: Kitchen et al. (2005).

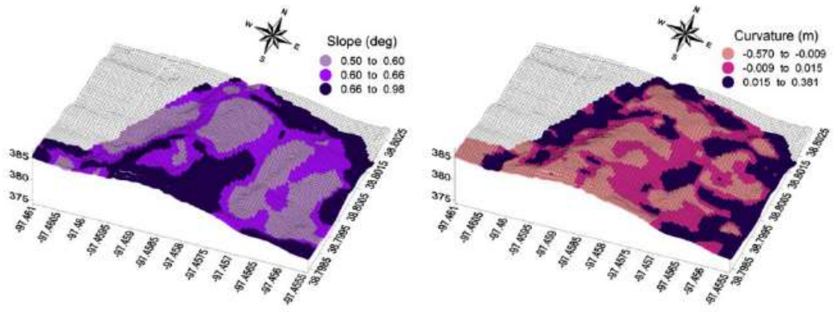
Kweon (2012) developed a delineation procedure for site-specific productivity zones with a fuzzy logic system using soil properties obtained from on-the-go electrical conductivity (ECa) and organic matter (OM) sensors and topographic attributes in two typical central Kansas upland fields (Field 1, 57 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). Fig. 22 shows continuous EC and OM maps, and Fig. 23, the maps of terrain slope and curvature. They constructed three types of MZs: 1) 5-year mean normalized yield map (Fig. 24a); 2) Productivity map, generated by a producer's decision-making knowledge and the fuzzy logic system (Fig. 24b); and 3) FCM map using EC, OM, slope, and curvature (Fig. 24c). The spatial agreement between the productivity and the 5-year-mean yield maps showed an overall accuracy and kappa coefficient of 0.57 and 0.35. The productivity map presented a better agreement with the normalized yield map than the FCM map. All the presented figures are for Field 2.



a) b)

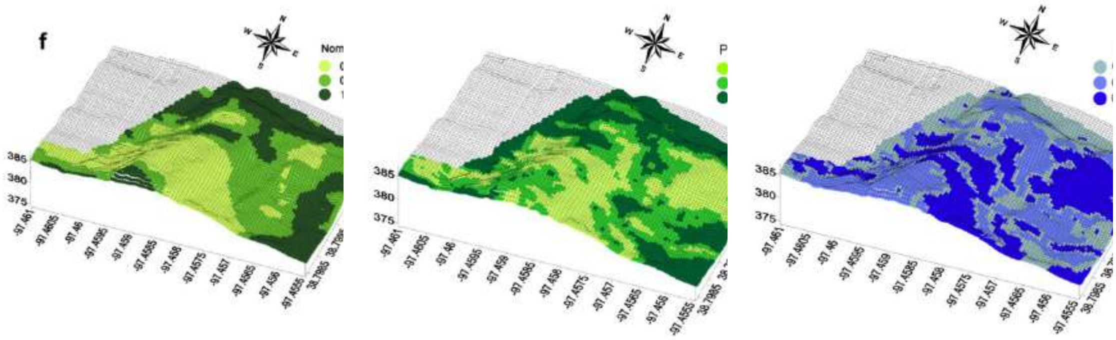
Fig. 22. EC and OM maps generated by the on-the-go sensor for Field 2

Source: Kweon (2012).



a) b)

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

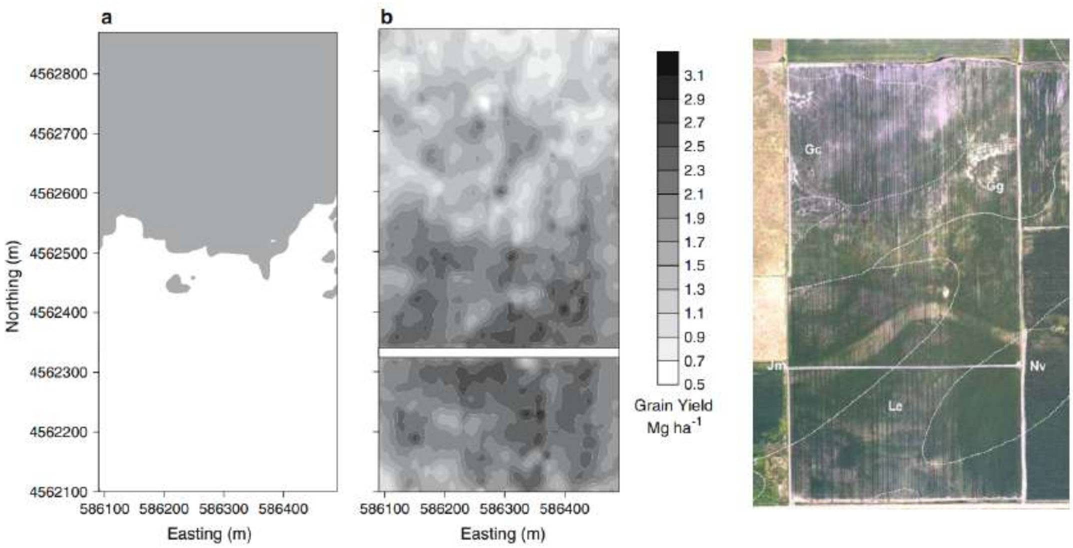


a) normalized yield map b) productivity MZs c) FCM map.

Fig. 24. 5-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).

3.2.2. 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 predicted best with the combination of NDVI and deep ECa. The delineation of chlorosis MZs from aerial imagery combined with soil ECa appears to be a useful tool for the site-specific management of iron chlorosis. Fig. 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 the Gibbon loam (Gg) and Gayville-Caruso (Gc) soil series, fairly poorly drained, with salt accumulation in the Gayville series occasionally causing dispersion of the soil colloids (classified as Leptic Natrustolls).



(a) (b)

Fig. 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).

3.2.3. Apparent electrical conductivity management zones

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

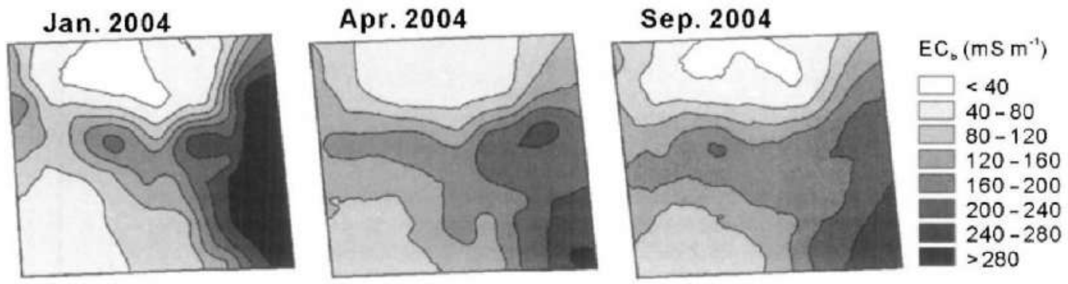
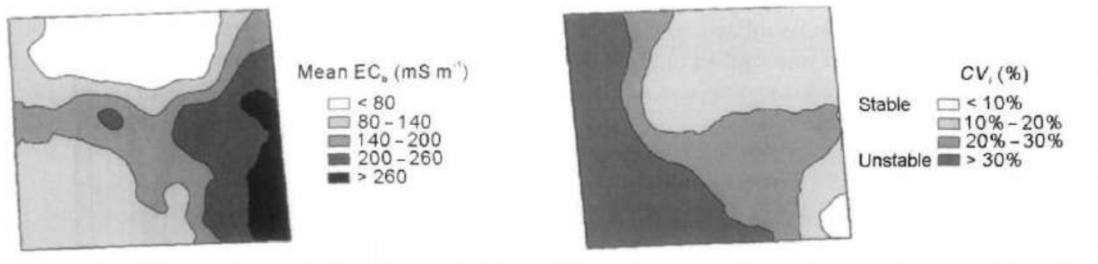


Fig. 26. Smoothed choropleth maps produced by ordinary kriging for apparent electrical conductivity (ECa) at three different sampling dates

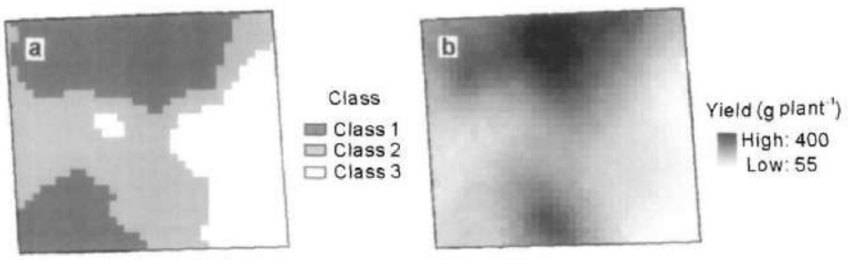
Source: Yan et al. (2007a).



(a) (b)

Fig. 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).

3.2.4. Soil available water content management zones

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

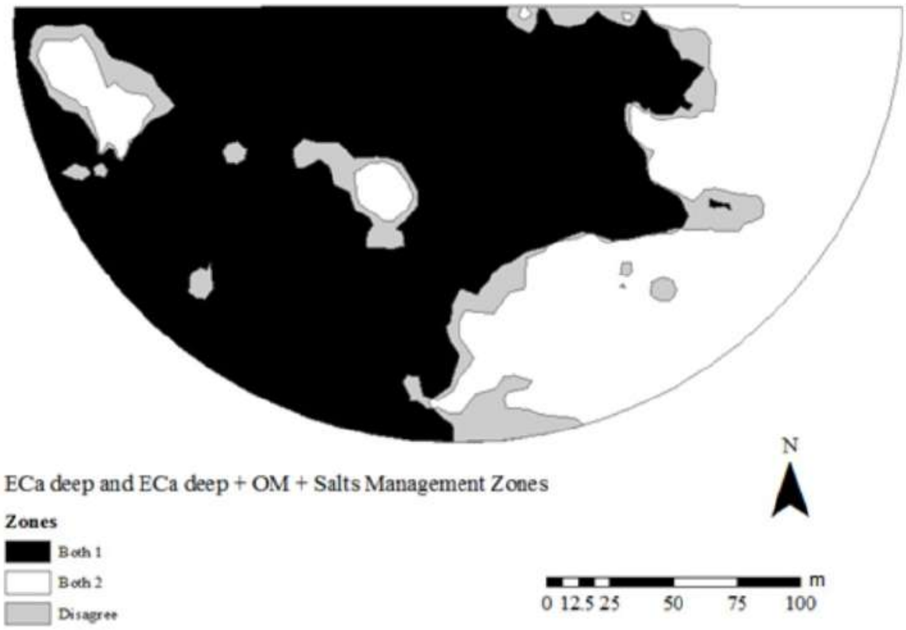


Fig. 29. Comparison of management zones delineated with using soil ECa measured up to 1.5 m depth and management zones delineated using soil ECa measured up to 1.5 m depth in addition to organic matter and soil salinity for ARDEC. Differences in the two techniques are presented as gray, referred to as “disagree” in the legend.

Source: De Lara et al. (2018).

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 the grape must). Soil properties, yield, and grape composition parameters showed high spatial variability. Maps of two MZs were produced using the MZA software. Fig. 30 shows the yield-based MZs using soil depth, NDVI 1, NDVI 2, NDVI 3, and NDVI 4 (Fig. 30a), and quality-based MZs using ECa, NDVI1, NDVI 2, NDVI3, and NDVI 4 (Fig. 30b). They concluded that these maps presented a high degree of agreement, from 79.2 to 89.6%.

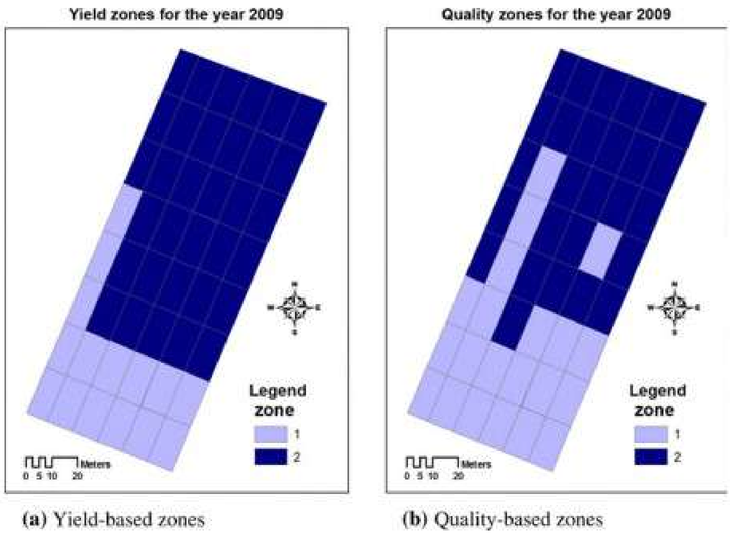


Fig. 30. Yield-based management zones (soil depth, NDVI 1, NDVI 2, NDVI 3, NDVI 4) and quality-based management zones (ECa, NDVI 1, NDVI 2, NDVI 3, NDVI 4) using fuzzy clustering from a commercial vineyard in Central Greece. Data from the 2009 agricultural year

Source: Tagarakis et al. (2013).

3.2.6. Weed Management Zones

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

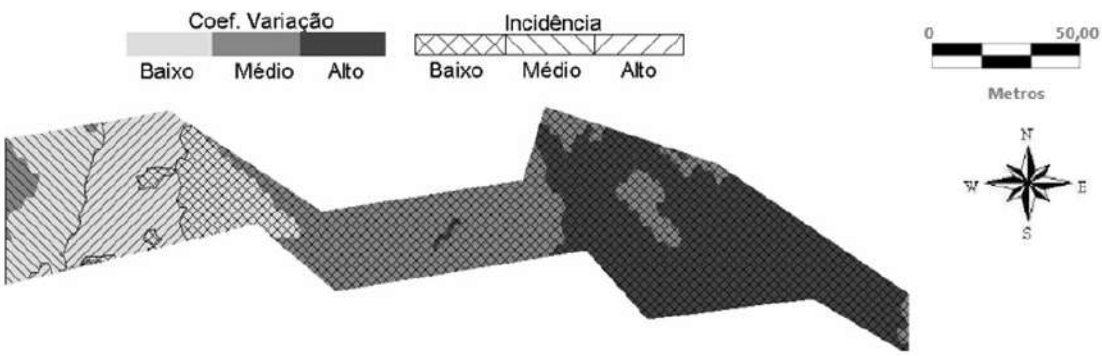


Fig. 31. Management zones of small leaves weed plants in a 1.24 ha pear orchard

Source: Rodrigues, 2009.

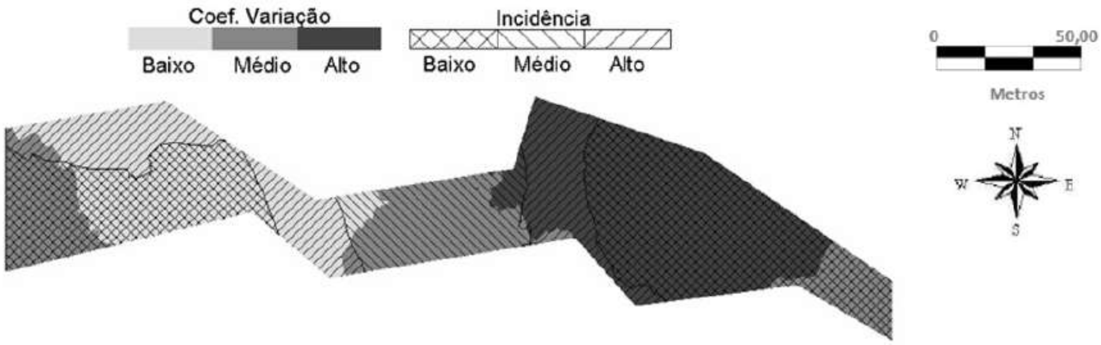


Fig. 32. Management zones of large leaves weed plants in a 1.24 ha pear orchar
Source: Rodrigues, 2009.

3.2.7. Vegetation Indices Management Zones

In the case of 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 the VIs spatial distribution patterns, the multivariate analysis resulted in a representative categorization of the grapevine vegetative vigor and delimitation of HZs for this characteristic.

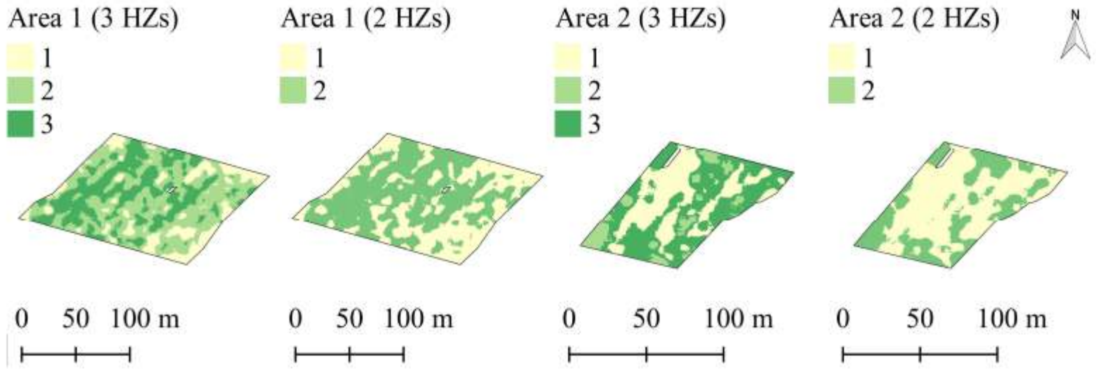


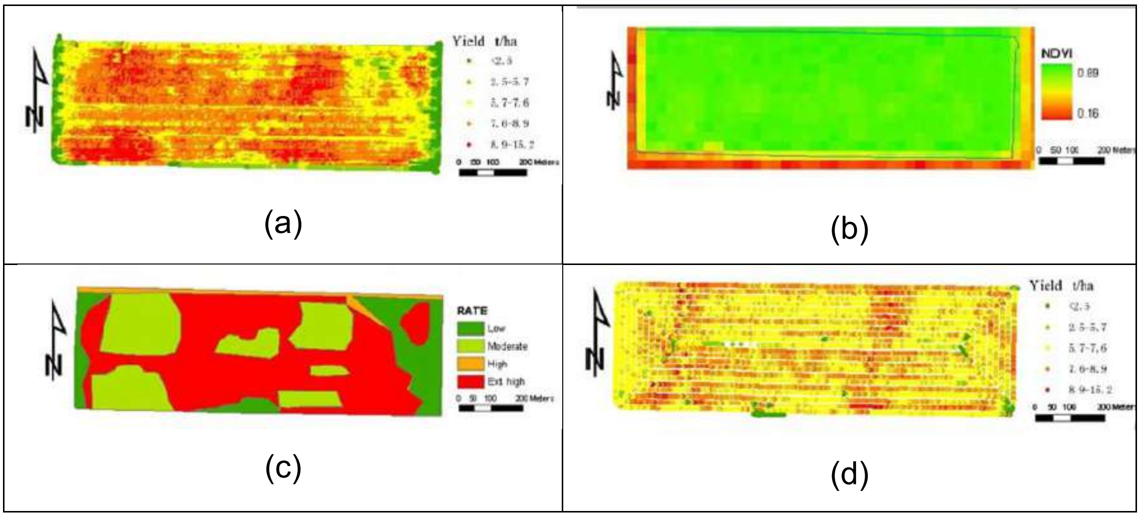
Fig. 33. Homogenous zones resulting from the clustering analysis of VIs, calculated based on ρ, for two study areas. Reflectance was measured at canopy height using an ACS-430 active sensor. The studied areas were cropped with Cabernet Franc and Cabernet Sauvignon vines

Source: Costa et al. (2019).

3.2.8. Used variables for delineating 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 various formats to ensure compatibility with GNSS-enabled farming equipment. Fig. 34 shows results from a 45.3-ha field in Potter County, South Dakota, where the rotation of crops from 2003 to 2005 was corn, sunflowers, and spring wheat. The farmer delineated four subfield zones (Fig. 34c) using a 2003-yield map (Fig. 34a) and a 25-August-2004-Landsat NDVI map (Fig. 34b) 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 (Fig. 34d). 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, the 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.



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

Fig. 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 for the determination of 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 the 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 two methods of MZs: (1) using soil properties (Fig. 35a); and (2) using tobacco RS data (Fig. 35b). They concluded that it is feasible to use an active canopy sensor to delineate MZs for tobacco-planting fields.

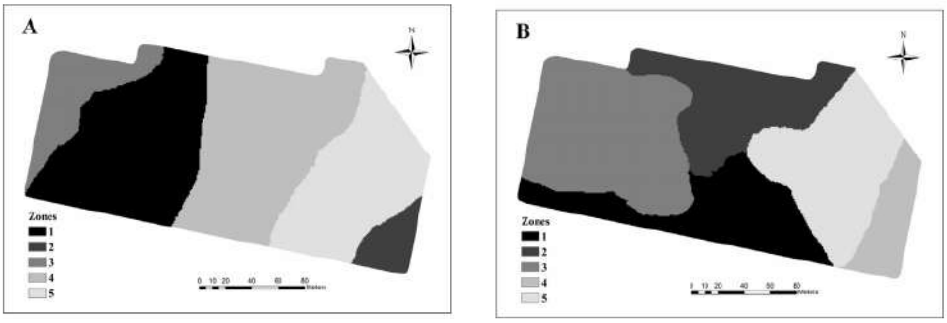


Fig. 35. Map of management zones based on soil properties (A) and NDVI measurements (B)

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, Fig. 36) in a parcel at Raimat (Lleida, Spain).

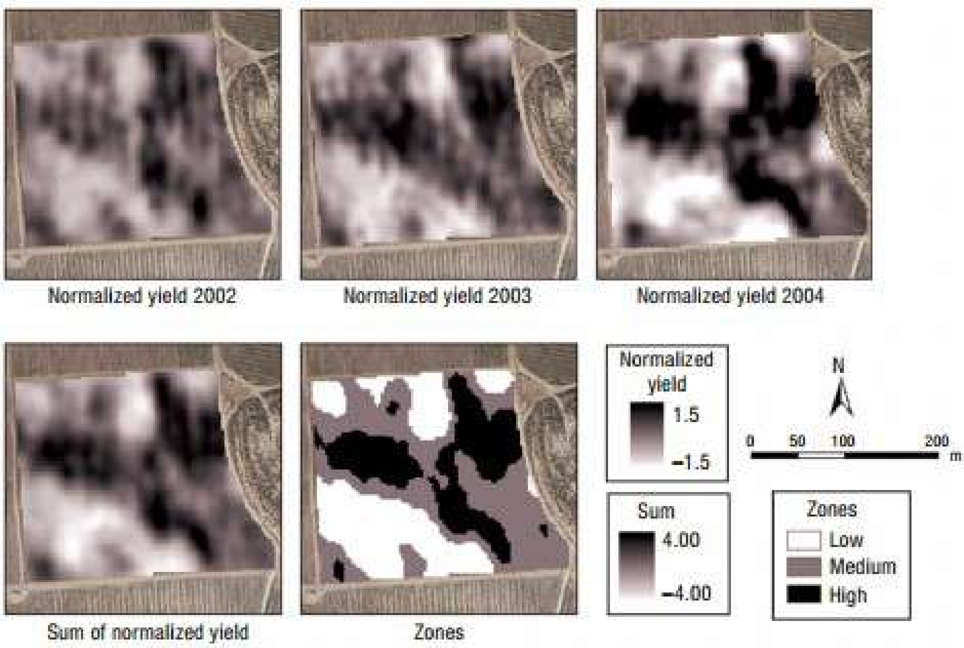


Fig. 36. Yield management zones delineated using grape normalized yields

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, located in Southen Brazil. PCA was used to group variables, and FCM was used to delineate MZs (Fig. 37). Results had confirmed the utility of ECa in the definition of MZs and the feasibility of the proposed method.

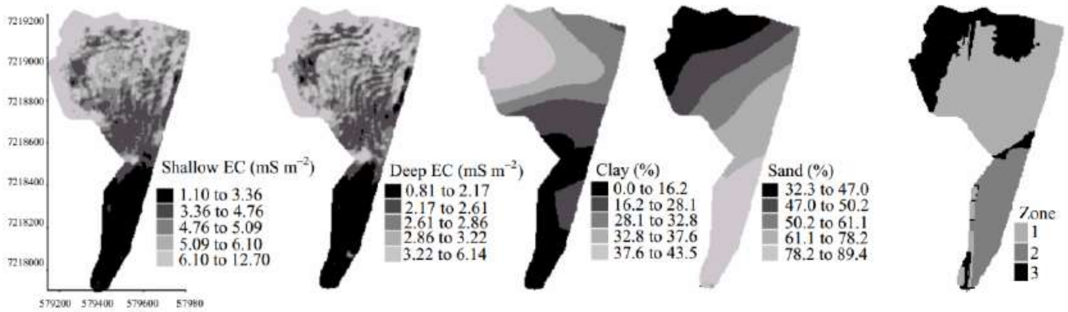


Fig. 37. Shallow (0 – 0.3 m) and deep-reading (0 – 0.9 m) soil EC maps, soil clay and sand content maps, and Management zones
Source: adapted from Molin and Castro (2008).

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 between 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 (Fig. 38). The terrain attributes SL, PL, AS, DD, and ECa effectively identified yield cluster membership for 80% of the 224 transect plots.

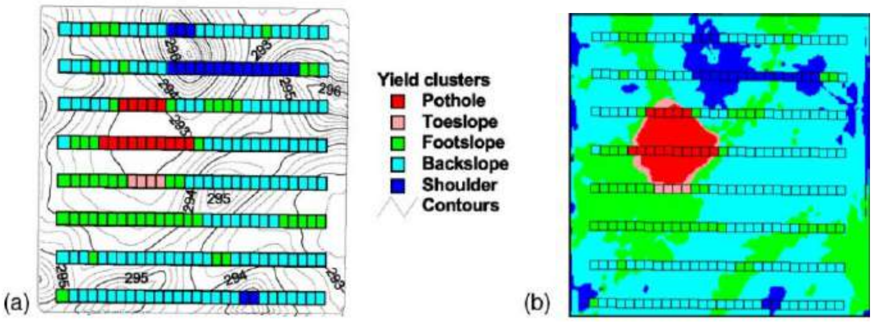


Fig. 38. Soybean-yield cluster classification for the 224 transect plots overlaid on the elevation contours (a) and the predicted yield zones (b). Transect plots are shown 3× actual width for better visibility.

Source: adapted from Jaynes et al. (2005).

3.2.9. Methods for selecting the variables used in the clustering process

Spatial correlation analysis

Bazzi et al. (2013) used the physical and chemical properties of soil and yield from a 19.8-ha commercial farming area in Brazil to delineate MZs by the FCM algorithm (Fig. 39). The division of the area into two MZs was considered appropriate since it provided distinct averages of most soil properties and yields.

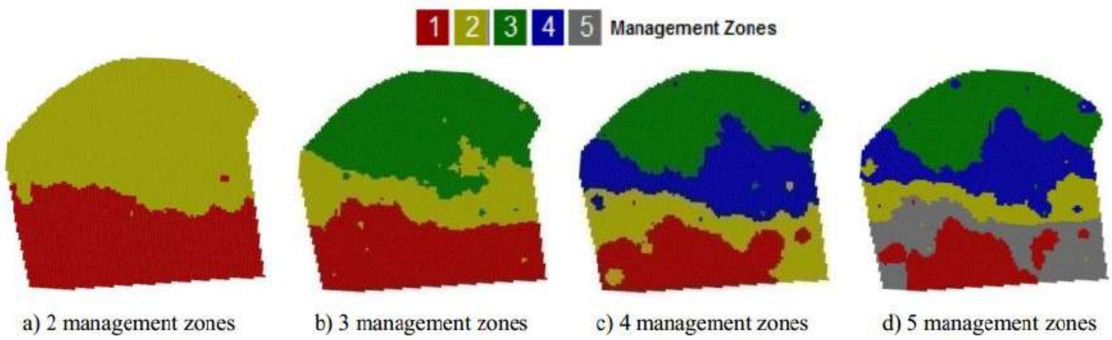


Fig. 39. Division of the area into management zones using the Fuzzy C-means algorithm with variables selected with the spatial correlation matrix approach.

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 located in Southen Brazil, aiming to delineate MZs with these variables. PCA was used to group variables, and FCM classification was used for clustering the transformed variables (Fig. 40). The results confirmed the utility of ECa in the definition of MZs and the feasibility of the proposed method.

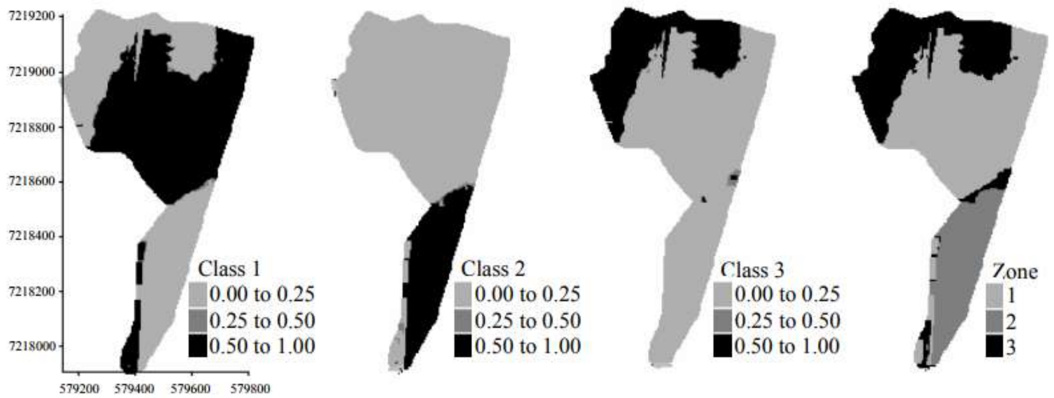


Fig. 40. Spatial distribution of participation function values for each individual in the three classes generated after classification by the fuzzy-k-means algorithm of two principal components selected and a corresponding map showing the resulting management zones.

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. The 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 for grouping variables and the FCM clusterization technique and concluded that the best classification was with two zones (Fig. 41).

Fig. 41. Map with two (left), three (center), and four (right) within-field management classes

Source: Córdoba et al. (2016).

Comparing methods for selecting 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 the FCM clustering method using data collected from 2010 to 2014 from three agricultural areas in Southern Brazil. The delineated MZs are presented in Fig. 42. They founded that MPCA-SC provided the best performance defining MZs, with greater internal homogeneity, making them more viable for field management.

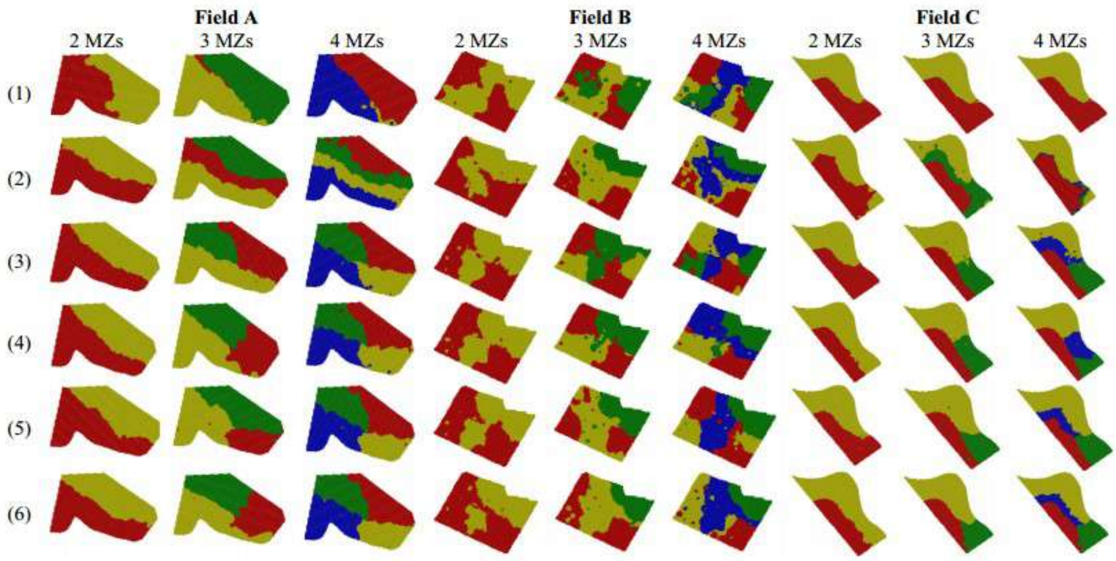


Fig. 42. Managements zones generated by the six approaches: (1) All-Attrib; (2) Spatial-Matrix; (3) PCA-All; (4) MPCA-All; (5) PCA-SC; (6) MPCA-SC

Source: Gavioli et al. (2016).

3.2.10. Methods for delineating Management Zones

Gavioli et al. (2018), with data obtained between 2010 and 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 the 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 (Fig. 43).

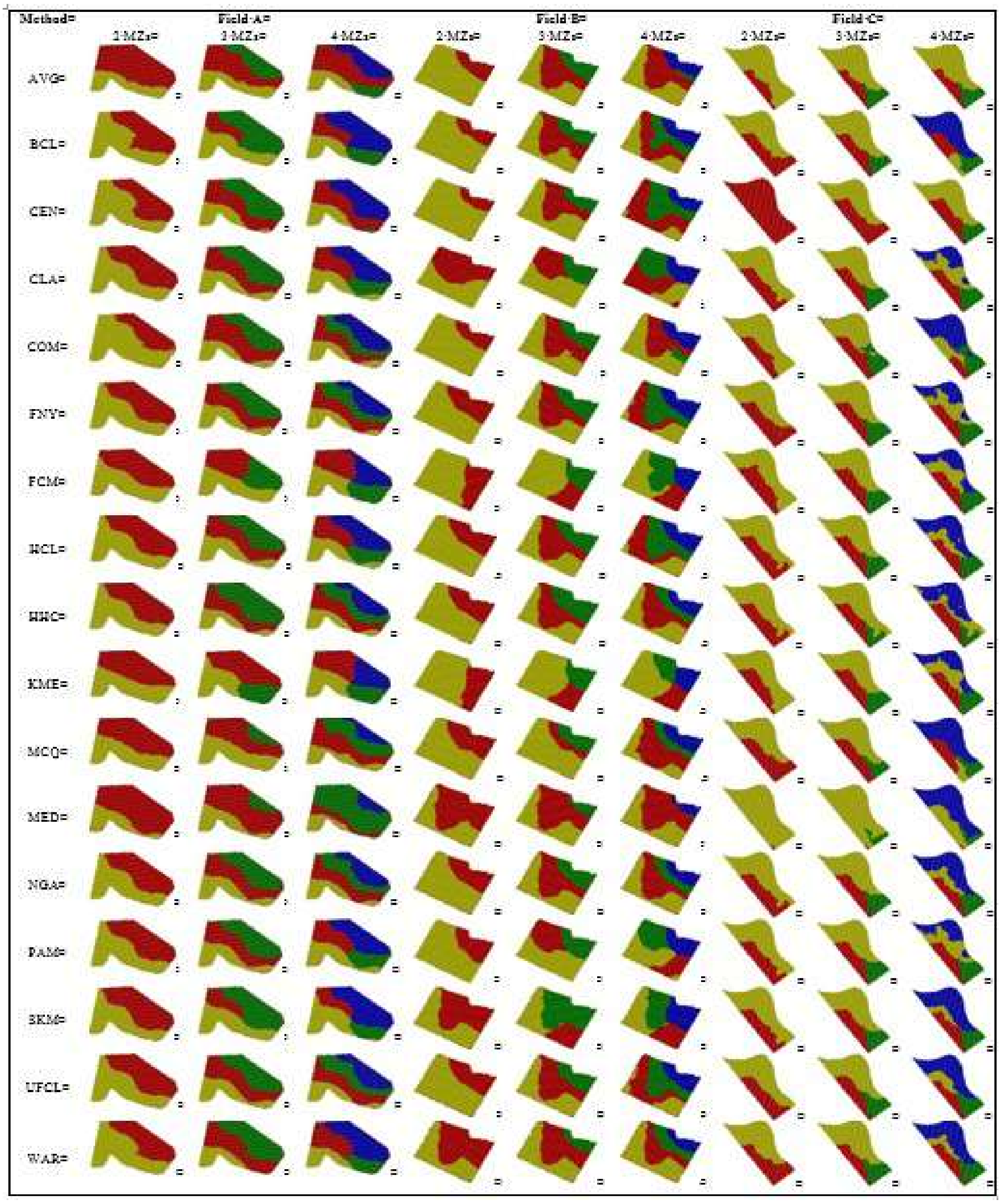


Fig. 43. Maps of the management zones delineated with the application of 17 clustering algorithms for the three fields

Source: Gavioli et al. (2018).

Table 10. Clustering methods implemented and compared for the definition of MZs

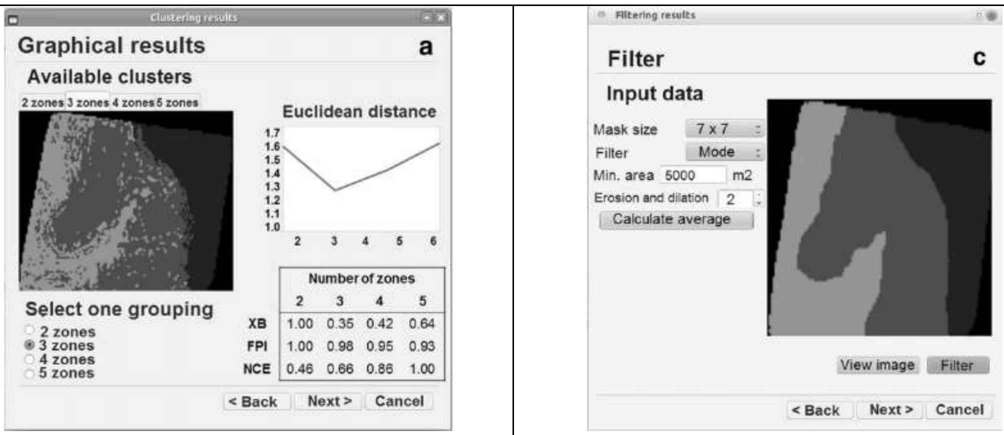
Method	Acronym	References
Average Linkage[a]	AVG	Jain and Dubes (1988)
Centroid Linkage[a]	CEN	Jain and Dubes (1988)
Complete Linkage[a]	COM	Jain and Dubes (1988)
Divisive Analysis (Diana)[a]	DIA	Kaufman and Rousseeuw (1990)
Hybrid Hierarchical Clustering[a]	HHC	Chipman and Tibshirani (2006)
Median Linkage[a]	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).

3.2.11. Rectification of Management Zones

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 ha. These procedures allow removal of all the isolated small areas and improve the border definition and compactness of the zones delineated (Fig. 44).



(a) (b)

Fig. 44. 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).

3.2.12. Evaluation of the quality of Management Zones

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 the indices VR, FPI, MPE, SI, ICVI and ANOVA (Table 11). The first analysis to be made is Tukey's range test to discard ZMs whose target variable means (in this case yield) are not all statistically distinct. As a result, for field A, one has to consider 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 the MZs; this implies that for area A, the best approach was the Spatial-Matrix.

Table 11. Results for ANOVA (Tukey's range test), VR, FPI, MPE, SI, and ICVI, for field A

Classes	Approach	ANOVA (Tukey's test)				VR (%)	FPI	MPE	SI (%)	ICVI
		C1	C2	C3	C4					
2	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
	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
3	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
	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
4	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
	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

significant at 0.05 confident level

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 quality of the clustering process was performed by the ANOVA (Tukey's range test), VR index, and ASC coefficient. The Tukey's 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 12. Results of the evaluation of the clustering methods in the generation of two, three and four classes by the ANOVA (Tukey's 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

The 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, Fig. 21). Using K, they found 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 the type of interpolation implemented.

References

Adamchuk, V., Hummel, J., Morgan, M., Upadhyaya, S. (2004). On-the-go soil sensors for precision agriculture. *Comput. Electron. Agric* 44:71–91.

Agati, G., Soudani, K., Tuccio, L., Fierini, E., Ben Ghozlen, N., Fadaili, E.M., Romani, A., Cerovic, Z. G. (2018) Management Zone Delineation for Winegrape Selective Harvesting Based on Fluorescence-Sensor Mapping of Grape Skin Anthocyanins. *J Agric Food Chem* 66:5778–5789. https://doi.org/10.1021/acs.jafc.8b01326

Aggelopooulou, K., Castrignanò, A., Gemtos, T., Benedetto, D. (2013) Delineation of management zones in an apple orchard in Greece using a multivariate approach. *Comput Electron Agric* 90:119–130. https://doi.org/10.1016/j.compag.2012.09.009

Aimrun, W., Amin, M. S. M., Ezrin, M. H., Mastura, M. (2011) Paddy soil properties and yield characteristics based on apparent electrical conductivity zone delineation for a humid tropical rice farm. *African J Agric Res* 6:5339–5350. https://doi.org/10.5897/AJAR11.370

Albornoz, E. M., Kemerer, A. C., Galarza, R., Mastaglia, N., Melchiori, R., Martínez, C. E. (2018) Development and evaluation of an automatic software for management zone delineation. *Precis Agric* 19:463–476. https://doi.org/10.1007/s11119-017-9530-9

Albornoz, V. M., Ñanco, L. J., Sáez, J. L. (2019) Delineating robust rectangular management zones based on column generation algorithm. *Comput Electron Agric* 194–201. https://doi.org/10.1016/j.compag.2019.01.045

Alves, S. M. De F., Alcântara, G. R., Queiroz, D. M., Valente, D. S. M. S. M., (2013) Definition of Management Zones from Maps of Electrical Conductivity and Organic Matter. *Biosci J* 29:104–114.

Amidan, B., Ferryman, T., Cooley, S (2005). Data outlier detection using the Chebyshev theorem. In IEEE Aerosp. Conf (pp. 3814e3819).

Anderson-Cook, C. M., Alley, M. M., Roygard, J. K. F., Khosla, R., Noble, R. B. and Doolittle, J. A., (2002). Differentiating soil types using electromagnetic conductivity and crop yield maps. *Soil Science Society of America Journal*, 66(5), pp.1562-1570.

Arango, R. B., Campos, A. M., Combarro, E. F., Canas, E. R., Díaz, I. (2017) Identification of agricultural management zones through clustering algorithms with thermal and multispectral satellite imagery. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 25:121-140.

Arlot, S., Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statist. Surv*. 4, 40–79.

Arnó J., Bordes X., Ribes-Dasi M., Blanco R., Rosell J.R., Esteve J. (2005). Obtaining grape yield maps and analysis of within-field variability in Raimat (Spain). *Proc V ECPA-Eur Conf on Precision Agriculture*. Uppsala, Sweden, June 8-11. pp. 899-906.

Arno, J., Martinez-Casasnovas, J. (2011) Clustering of grape yield maps to delineate site-specific management zones. *Spanish J Agric Res* 9:721–729. https://doi.org/10.5424/sjar/20110903-456-10.

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

Bansod, B. S., Pandey, O. P., Rajesh, N. L. (2012) Analysis and delineation of spatial variability using geo-sensed apparent electrical conductivity and clustering techniques. *Int J Agric Biol* 14:481–491.

Basso, B., Cammarano, D., Chen, D., Cafiero, G., Amato, M., Bitella, G., Rossi, R., Basso, F. (2009). Landscape position and precipitation effects on spatial variability of wheat yield and grain protein in Southern Italy. *J. Agron. Crop Sci*. 195, 301–312.

Baudron, F., Giller, K. E. (2014) Agriculture and nature: Trouble and strife? *Biol Conserv* 170:232–245. https://doi.org/10.1016/j.biocon.2013.12.009

Bazzi, C. L., Jasse, E. P., Magalhães, P. S. G., Michelon, G. K., de Souza, E. G., Schenatto, K., Sobjak, R. (2019) AgDataBox API–Integration of data and software in precision agriculture. *SoftwareX*, 10, 100327, 8 pp.

Bazzi, C. L., Souza, E. G., Konopatzki, M. R., Pereira Nóbrega, L. H., Uribe-Opazo, M. A. (2015) Management zones applied to pear orchard. *J Food, Agric Environ* 13:98–104.

Bazzi, C. L., Souza, E. G., Schenatto, K., Betzek, N. M., Gavioli, A. (2019) A software for the delineation of crop management zones (SDUM). *Aust J Crop* Sci 13:26–34. https://doi.org/10.21475/ajcs.19.13.01.p1026

Bazzi, C.L., Souza, E.G., Khosla, R., Uribe-Opazo, M.A., Schenatto, K. (2015). Profit maps for precision agriculture. *Ciencia e Investigación Agraria (en línea*. Edición en Español) 42, 385–396.

Bazzi, C. L., Souza, E. G., Uribe-Opazo, M. A., Nóbrega, L. H. P., Rocha, D. M. (2013) Management zones definition using soil chemical and physical attributes in a soybean area. *Eng Agrícola* 33:952–964. https://doi.org/10.1590/S0100-69162013000500007

Behera, S. K., Mathur, R. K., Shukla, A. K., Suresh, K., Prakash, C. (2018) Spatial variability of soil properties and delineation of soil management zones of oil palm plantations grown in a hot and humid tropical region of southern India. *Catena* 165:251–259. https://doi.org/10.1016/j.catena.2018.02.008

Benedetto, D. De, Castrignano, A., Diacono, M., Rinaldi, M., Ruggieri, S., Tamborrino, R. (2013a) Field partition by proximal and remote sensing data fusion. *Biosyst Eng* 114:372–383. https://doi.org/https://doi.org/10.1016/j.biosystemseng.2012.12.001

Benedetto, D. De, Castrignanò, A., Rinaldi, M., Ruggieri, S., Santoro, F., Figorito, B., Gualano, S., Diacono, M., Tamborrino, R. (2013b) An approach for delineating homogeneous zones by using multi-sensor data. *Geoderma* 199:117–127. https://doi.org/https://doi.org/10.1016/j.geoderma.2012.08.028

Bernardi, A. C. De C., Tupy, O., Eduarda Lima Santos, K., Mazzuco, G. G., Bettiol, G. M., Rabello, L. M., Inamasu, R. Y. (2018) Mapping of yield, economic return, soil electrical conductivity, and management zones of irrigated corn for silage. *Pesq agropec bras* 53:1289–1298. https://doi.org/10.1590/S0100-204X2018001200001

Bernardi, A. C. C., Bettiol G. M., Ferreira R. P., Santos K. E. L, Rabello L. M. and Inamasu R. Y. (2016). Spatial variability of soil properties and yield of a grazed alfalfa pasture in Brazil. Precision Agriculture 1–16.

Betzek, N. M., Souza, E. G., Bazzi, C. L., Schenatto, K., Gavioli, A. (2018) Rectification methods for optimization of management zones. *Comput Electron Agric* 146:1–11. https://doi.org/https://doi.org/10.1016/j.compag.2018.01.014

Betzek, N. M., Souza E. G., Bazzi, C. L., Schenatto, K., Gavioli, A., Magalhães, P. S. G. G. (2019) Computational routines for the automatic selection of the best parameters used by interpolation methods to create thematic maps. *Comput Electron Agric* 157:49–62. https://doi.org/10.1016/j.compag.2018.12.004

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

Bier, V. A.; Souza, E. G. (2017). Interpolation selection index for delineation of thematic maps. *Comput Electron Agric*, 136, 202-209.

Biernacki, P., Waldorf, D. A. N. (1981) Snowball Sampling: Problems and Techniques of Chain Referral Sampling. *Sociol Methods Res* 10:141–163. https://doi.org/https://doi.org/10.1177/004912418101000205

Blackmore, S. (2000). The interpretation of trends from multiple yield maps. *Comput Electron Agric* 26:37–51. https://doi.org/10.1016/S0168-1699(99)00075-7

Blackmore, S.; Moore, M., (1999). Remedial correction of yield map data. *Precision agriculture* 1:53-66.

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

Boluwade, A., Madramootoo, C., Yari, A. (2016) Application of unsupervised clustering techniques for management zone delineation: case study of variable rate irrigation in southern Alberta, Canada. *J Irrig Drain Eng* 142:05015007. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000936

Borges, L. F., Bazzi, C. L., Souza, E. G. D., Magalhães, P. S. G., Michelon, G. K. (2020) Web software to create thematic maps for precision agriculture. *Pesq agropec bras*, 55, 11 pp.

Bottega, E. L., Queiroz, D. M. De, Pinto, F. De A. De C., Souza, C. M. A. De, Valente, D. S. M. (2017) Precision agriculture applied to soybean: Part I - Delineation of management zones. *Aust J Crop Sci* 11:573–579. https://doi.org/10.21475/ajcs.17.11.05.p381

Boydell, B., McBratney, A.B. (2002) Identifying potential within-field management zones from cotton-yield estimates. *Precis. Agric* 3:9–23.

Brock, A., Brouder, S.M., Blumhoff, G., Hofmann, B.S. (2005) Defining yield-based management zones for corn–soybean rotations. *Agron J* 97:1115-1128.

Buttafuoco, G., Castrignanò, A., Colecchia, A. S., Ricca, N. (2010) Delineation of management zones using soil properties and a multivariate geostatistical approach. *Ital J Agron* 5:323–332. https://doi.org/10.4081/ija.2010.323

Buttafuoco, G., Castrignanò, A., Cucci, G., Lacolla, G., Lucà, F. (2017) Geostatistical modelling of within-field soil and yield variability for management zones delineation: a case study in a durum wheat field. *Precis Agric* 18:37–58. https://doi.org/10.1007/s11119-016-9462-9

Caires, S. A. De, Wuddivira, M. N., Bekele, I. (2015) Spatial analysis for management zone delineation in a humid tropic cocoa plantation. *Precis Agric* 16:129–147. https://doi.org/10.1007/s11119-014-9366-5

Carvalho, V. R. F. Cadeia produtiva da maçã no Brasil: limitações e potencialidades. Porto Alegre, BRDE, 2011. 44 p.

Castrignanò, A., Benedetto, D., Girone, G., Guastaferro, F., Sollitto, D. (2010) Characterization, delineation and visualization of agro-ecozones using multivariate geographical clustering. *Ital J Agron* 5:121–132. https://doi.org/10.4081/ija.2010.121

Cavallo, G., Benedetto, D. De, Castrignanò, A., Quarto, R., Vonella, A. V., Buttafuoco, G. (2016) Use of geophysical data for assessing 3D soil variation in a durum wheat field and their association with crop yield. *Biosyst Eng* 152:28–40. https://doi.org/10.1016/j.biosystemseng.2016.07.002

Chang, D., Zhang, J., Zhu, L., Ge, S. H., Li, P. Y., Liu, G. S. (2014) Delineation of management zones using an active canopy sensor for a tobacco field. *Comput Electron Agric* 109:172–178. https://doi.org/10.1016/j.compag.2014.09.019

Cid-Garcia, N. M., Albornoz, V., Rios-Solis, Y. A., Ortega, R. (2013) Rectangular shape management zone delineation using integer linear programming. *Comput Electron Agric* 93:1–9. https://doi.org/10.1016/j.compag.2013.01.009

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

Cohen, J. A. (1960). Coefficient of agreement for nominal scales. *Educ. Psychol*. *Measur*. 20, 37–46.

Cohen, N., Arieli, T. (2011) Field research in conflict environments: Methodological challenges and snowball sampling. *J Peace Res* 48:423–435. https://doi.org/10.1177/0022343311405698

Cohen, S., Cohen, Y., Alchanatis, V., Levi, O. (2013). Combining spectral and spatial information from aerial hyperspectral images for delineating homogenous management zones. *Biosyst. Eng.* 114, 435–443.

Córdoba, M., Bruno, C., Costa, J., Balzarini, M. (2013) Subfield management class delineation using cluster analysis from spatial principal components of soil variables. *Comput Electron Agric* 97:6–14. https://doi.org/https://doi.org/10.1016/j.compag.2013.05.009

Córdoba, M. A., Bruno, C. I., Costa, J. L., Peralta, N. R., Balzarini, M. G. (2016) Protocol for Multivariate Homogeneous Zone Delineation in Precision Agriculture. *Biosyst Eng* 143:95–107. https://doi.org/10.1016/j.biosystemseng.2015.12.008

Costa, B. R., Oldoni, H., Junior, R., Romero, C. and Bassoi, L. H. (2019). Delimitation of homogeneous zones in vineyards using geostatistics and multivariate analysis of different vegetation indices. *Engenharia Agrícola*, 39(SPE), pp.13-22.

Cressie, N. (1993). Statistics for spatial data. Revised ed. John Wiley & Sons, New York.

Czaplewski, R. L., Reich, R. M. (1993). Expected value and variance of Moran's bivariate spatial autocorrelation statistic for a permutation test. Fort Collins CO.

Dall'Agnol, R. W., Michelon, G. K., Bazzi, C. L., Magalhães, P. S. G., de Souza, E. G., Betzek, N. M., Sobjak, R. (2020) Web applications for spatial analyses and thematic map generation. *Comput Electron Agric*, 172, 105374, 13 pp.

Damian, M., Pias, O. H. D. C., Santi, A. L., Di Virgilio, N., Berghetti, J., Barbanti, L., Martelli, R. (2016) Delineating management zones for precision agriculture applications: A case study on wheat in sub-tropical Brazil. *Ital J Agron* 11:171–179. https://doi.org/10.4081/ija.2016.713

Davatgar, N., Neishabouri, M. R., Sepaskhah, A. R. (2012) Delineation of site specific nutrient management zones for a paddy cultivated area based on soil fertility using fuzzy clustering. *Geoderma* 173–174:111–118. https://doi.org/10.1016/j.geoderma.2011.12.005

Davidson, D. (2014). Evaluating the quality of your soil. *Crop Soils*, 47(2), 4–13.

De Lara, A., Khosla, R. and Longchamps, L. (2018). Characterizing spatial variability in soil water content for precision irrigation management. Agronomy, 8(5), p.59.

Demattê, J. A. M., Demattê, J. L. I., Alves, E. R., Barbosa, R. N., Morelli, J. L. (2014). Precision agriculture for sugarcane management: a strategy applied for Brazilian conditions. *Acta Sci*. Agron. 36, 111.

Diacono, M., Benedetto, D., Castrignanò, A., Rubino, P., Vitti, C., Abdelrahman, H. M., Sollitto, D., Cocozza, C., Ventrella, D. (2013) 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* 64–65:47–57. https://doi.org/https://doi.org/10.1016/j.njas.2013.03.001

Dobermann, A., Ping, J. L., Adamchuk, V. I., Simbahan, G. C., Ferguson, R. B. (2003) Classification of Crop Yield Variability in Irrigated Production Fields. *Agron J* 95:1105. https://doi.org/10.2134/agronj2003.1105

Doerge, T.A. (2000). Site-Specific Management Guidelines. Potash & Phosphate Institute, Norcross.

Dray, S., Saïd, S., Débias, F. (2008) Spatial ordination of vegetation data using a generalization of Wartenberg's multivariate spatial correlation. *J Veg Sci* 19:45–56. https://doi.org/10.3170/2007-8-18312

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

Faraco, M. A., Uribe-Opazo, M. A., Da Silva, E. A. A., Johann, J. A. and Borssoi, J. A. (2008). 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*, Viçosa 32:463-476.

Ferguson, R. B., Hergert, G. W. (2009). Soil sampling for precision agriculture. *Ext. Precis. Agric*. 1–4.

Fleming, K. L.; Westfall, D. G.; Wiens, D. W.; Brodahl, M. C. (2000) Evaluating farmer defined management zone maps for variable rate fertilizer application. *Precision Agriculture*, Dordrecht, v. 2, n. 2, p. 201-215.

Fleming, K. L., Heermann, D.F., Westfall, D.G. (2004). Evaluating soil color with farmer input and apparent soil electrical conductivity for management zone delineation. Agron. J. 96, 1581–1587.

Fleming, K. L.; Westfall, D. G.; Wiens, D. W.; Brodahl, M. C. (2000) Evaluating farmer defined management zone maps for variable rate fertilizer application *Precis Agric* 2:201-215.

Fraisse, C. W., Sudduth, K. A., Kitchen, N. R. (2001) Delineation of Site-Specific Management Zones by Unsupervised Classification of Topographic Attributes and Soil Electrical Conductivity. *Trans ASAE* 44:155–166. https://doi.org/10.13031/2013.2296

Franzen, D. W., Hopkins, D. H., Sweeney, M. D., Ulmer, M. K., Halvorson, A. D. (2002) Evaluation of Soil Survey Scale for Zone Development of Site-Specific Nitrogen Management. *Agron J* 94:381. https://doi.org/10.2134/agronj2002.0381

Fridgen, J. J., Kitchen, N. R., Sudduth, K. A., Drummond, S. T., Wiebold, W. J., Fraisse, C. W. (2004) Management Zone Analyst (MZA): Software for Subfield Management Zone Delineation. *Agron J* 96:100–108

Fridgen, J.J., Kitchen, N.R., Sudduth, K. A. (2000). 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.

Frogbrook, Z. L., Oliver, M. A. (2007) Identifying management zones in agricultural fields using spatially constrained classification of soil and ancillary data. *Soil Use Manag* 23:40–51. https://doi.org/10.1111/j.1475-

2743.2006.00065.x

Fu, Q., Wang, Z., Jiang, Q. (2010) Delineating soil nutrient management zones based on fuzzy clustering optimized by PSO. *Math Comput Model* 51:1299–1305. https://doi.org/10.1016/j.mcm.2009.10.034

Galambošová, J., Rataj, V., Prokeinová, R., Prešinská, J. (2014) Determining the management zones with hierarchic and non-hierarchic clustering methods. *Res Agric Eng* 60:S44–S51. https://doi.org/10.1094/CM-2008-0925-01-RS

Gavioli, A., Souza, E. G., Bazzi, C. L., Guedes, L. P. C., Schenatto, K. (2016) Optimization of management zone delineation by using spatial principal components. Comput Electron Agric 127:302–310 . https://doi.org/10.1016/j.compag.2016.06.029

Gavioli, A., Souza E. G., Bazzi, C. L., Schenatto, K., Betzek, N. M. (2019) Identification of management zones in precision agriculture: An evaluation of alternative cluster analysis methods. *Biosyst Eng* 181:86–102. https://doi.org/10.1016/j.biosystemseng.2019.02.019

Gili, A., Álvarez, C., Bagnato, R., Noellemeyer, E. (2017) Comparison of three methods for delineating management zones for site-specific crop management. *Comput Electron Agric* 139:213–223. https://doi.org/https://doi.org/10.1016/j.compag.2017.05.022

Gnanadesikan, R., Kettenring, J. R., Tsao, S. L. (1995) Weighting and selection of variables for cluster analysis. *J Classif* 12:113–136. https://doi.org/10.1007/BF01202271

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

González-Fernández, A. B., Rodríguez-Pérez, J. R., Ablanedo, E. S., Ordoñez, C .(2017) Vineyard zone delineation by cluster classification based on annual grape and vine characteristics. *Precis Agric* 18:525–573. https://doi.org/10.1007/s11119-016-9475-4

González-Fernández, A. B., Rodríguez-Pérez, J. R., Sanz-Ablanedo, E., Valenciano, J. B., Marcelo, V. (2019) Delineating vineyard zones by fuzzy K-means algorithm based on grape sampling variables. *Sci Hortic (Amsterdam)* 243:559–566. https://doi.org/https://doi.org/10.1016/j.scienta.2018.09.012

Gozdowski, D., Stepień, M., Samborski, S., Dobers, E. S., Szatyłowicz, J., Chormański, J. (2014) Determination of the Most Relevant Soil Properties for the Delineation of Management Zones in Production Fields. *Commun Soil Sci Plant Anal* 45:2289–2304.

https://doi.org/10.1080/00103624.2014.912289

Guastaferro, F., Castrignanò, A., Benedetto, D., Sollitto, D., Troccoli, A., Cafarelli, B. (2010) A comparison of different algorithms for the delineation of management zones. *Precis Agric* 11:600–620. https://doi.org/10.1007/s11119-010-9183-4

Haghverdi, A., Leib, B. G., Washington-Allen, R. A., Ayers, P. D., Buschermohle, M. J. (2015). Perspectives on delineating management zones for variable rate irrigation. *Comput. Electron. Agric*. 117, 154–167.

Harrower M., Brewer C. A. (2003) ColorBrewer.org: an online tool for selecting colour schemes for maps. *Cartogr. J*.40:27-37.

Hornung, A., Khosla, R., Reich, R., Inman, D., Westfall, D. G. (2006) Comparison of Site-Specific Management Zones. *Agron J* 98:407. https://doi.org/10.2134/agronj2005.0240

Hörbe, T.D.A., Amado, T.J.C., Ferreira, A.D.O., Alba, P.J. (2013). Optimization of corn plant population according to management zones in Southern Brazil. *Precis Agric* 14:450-465.

Hotelling, H. (1933) Analysis of a complex of statistical variables into principal components. *J Educ Psychol* 6:417–441. https://doi.org/http://dx.doi.org/10.1037/h0071325

Ikenaga, S., Inamura, T. (2008) Evaluation of site-specific management zones on a farm with 124 contiguous small paddy fields in a multiple-cropping system. *Precis Agric* 9:147–159. https://doi.org/10.1007/s11119-008-9062-4

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

Inman, D., Khosla, R., Reich, R., Westfall, D.G. (2008). Normalized difference vegetation index and soil color-based management zones in irrigated maize. *Agron J* 100:60-66.

Isaaks, E. H; Srivastava, R. M. Applied geostatistics. New York: Oxford University Press, 1989. 561p.

ISPA, International Society for Precision Agriculture, 2019. 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.

Jacintho, J. L., Ferraz, G. A. e S., Silva, F. M. Da, Santos, S. A. (2017) Management zones in coffee cultivation. *Rev Bras Eng Agrícola e Ambient*

21:94–99. https://doi.org/10.1590/1807-1929/agriambi.v21n2p94-99

Jaynes, D. B., Colvin, T.S. and Kaspar, T. C. (2005). Identifying potential soybean management zones from multi-year yield data. *Comput Electron Agric* 46:309-327.

Jiang, H. L., Liu, G. S., Liu, S. D., Li, E. H., Wang, R., Yang, Y. F., Hu, H. C. (2012) Delineation of site-specific management zones based on soil properties for a hillside field in central China. *Arch Agron Soil Sci* 58:1075–1090. https://doi.org/10.1080/03650340.2011.570337

Jiang, Q., Fu, Q., Wang, Z. (2011) Delineating site-specific irrigation management zones. *Irrig Drain* 60:464–472. https://doi.org/10.1002/ird.588

Johnson, R. A., Wichern, D. W. (2007). Applied Multivariate Statistical Analysis, sixth ed. Pearson, New Jersey.

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

Karlik, L., Marián, G., Falťan, V., Havlíček, M. (2018) Vineyard zonation based on natural terroir factors using multivariate statistics – Case study Burgenland (Austria). *OENO One* 52:105–117. https://doi.org/10.20870/oeno-one.2018.52.2.1907

Khan, M. S., Mehdi, S. M., Sheikh, A. A., Khan, A. G., Shakar, M., Alvi, A., Ahmad, H. R., David, M. (2018) Delineation of nutrient management zones in Chunian using soil attributes database developed by soil fertility research institute, Punjab, Pakistan. *Soil Enviromnment* 37:116–122. https://doi.org/10.25252/SE/18/61526

Khosla, R., Alley, M. M. (1999). Soil-specific nitrogen management on mid-Atlantic coastal plain soils. *Better Crops* 83:6-7.

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

Khosla, R., Inman, D., Westfall, D. G., Reich, R. M., Frasier, M., Mzuku, M., ... & Hornung, A. (2008). 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*, 9(1-2), 85-100.

Khosla, R., Westfall, D. G., Reich, R. M., Mahal, J. S. and Gangloff, W. J. (2010). Spatial variation and site-specific management zones. In *Geostatistical applications for precision agriculture* (pp. 195-219). Springer, Dordrecht.

Kitchen, N.R., Sudduth, K.A., Drummond, S.T., Fridgen, J.J., Fraisse, C.W.

(2002) 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.

Kitchen, N. R., Sudduth, K. A., Myers, D. B., Drummond, S. T., Hong, S. Y. (2005) Delineating productivity zones on claypan soil fields using apparent soil electrical conductivity. *Comput Electron Agric* 46:285–308. https://doi.org/10.1016/j.compag.2004.11.012

Kitchenham, B., Charters, S. (2007) Guidelines for performing Systematic Literature reviews in Software Engineering. *Engineering* 45:1051. https://doi.org/10.1145/1134285.1134500

Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., Linkman, S. (2009) Systematic literature reviews in software engineering - A systematic literature review. *Inf. Softw. Technol* 51:7–15

Kohavi, R., (1995). 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*, pp. 1137–1143.

Kuang, B., Mahmood, H.S., Quraishi, M.Z., Hoogmoed, W.B., Mouazen, A.M., van Henten, E.J. (2012). Sensing soil properties in the laboratory, in situ, and on-line: a review. In Advances in agronomy 114:155-223. Academic Press.

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

Kyaw, T., Ferguson, R. B., Adamchuk, V. I., Marx, D. B., Tarkalson, D. D., McCallister, D. L. (2008) Delineating site-specific management zones for pH-induced iron chlorosis. *Precis Agric* 9:71–84. https://doi.org/10.1007/s11119-008-9059-z

Landrum, C., Castrignanò, A., Mueller, T., Zourarakis, D., Zhu, J., Benedetto, D. De (2015) An approach for delineating homogeneous within-field zones using proximal sensing and multivariate geostatistics. *Agric Water Manag* 147:144–153. https://doi.org/https://doi.org/10.1016/j.agwat.2014.07.013

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

Li, Y., Shi, Z., Li, F., Li, H. Y. (2007) Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. *Comput Electron Agric* 56:174–186. https://doi.org/10.1016/j.compag.2007.01.013

Li, Y., Shi, Z., Wu, C., Li, H., Li, F. (2008) Determination of potential management zones from soil electrical conductivity, yield and crop data. *J Zhejiang Univ B* 9:68–76. https://doi.org/10.1631/jzus.B071379

Li, Y., Shi, Z., Wu, H. X., Li, F., Li, H. Y. (2013) Definition of management zones for enhancing cultivated land conservation using combined spatial data. *Environ Manage* 52:792–806. https://doi.org/10.1007/s00267-013-0124-7

Lin, Q., Li, H., Luo, W., Lin, Z., Li, B. (2013) Optimal soil-sampling design for rubber tree management based on fuzzy clustering. *For Ecol Manage* 308:214–222. https://doi.org/https://doi.org/10.1016/j.foreco.2013.07.028

Loisel, P., Charnomordic, B., Jones, H., Tisseyre, B. (2019) An optimisation-based approach to generate interpretable within-field zones. *Precis Agric* 20:101–117. https://doi.org/10.1007/s11119-018-9584-3

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

MacQueen, J. (1967) Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifht Berkeley Symposium on Mathematics. pp 281–296

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

Martínez Casasnovas, J. A., & Bordes Aymerich, X. (2005). Viticultura de precisión: Predicción de cosecha a partir de variables del cultivo e índices de vegetación. *Revista de teledetección*, 2005, vol. 24, p. 67-71.

Martínez-Casasnovas, J., Escolà, A., Arnó, J. (2018) Use of Farmer Knowledge in the Delineation of Potential Management Zones in Precision Agriculture: A Case Study in Maize (Zea mays L.). *Agriculture* 8:84. https://doi.org/10.3390/agriculture8060084

McBratney, A. B., Moore, A. W. (1985) Application of fuzzy sets to climatic classification. *Agric For Meteorol* 35:165–185. https://doi.org/10.1016/0168-1923(85)90082-6

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

McClymont, L., Goodwin, I., Mazza, M., Baker, N., Lanyon, D. M., Zerihun, A., Chandra, S., Downey, M. O. (2012) Effect of site-specific irrigation management on grapevine yield and fruit quality attributes. *Irrig Sci* 30:461–470. https://doi.org/10.1007/s00271-012-0376-7

Meirvenne, M. Van, Islam, M. M., Smedt, P. De, Meerschman, E., Vijver, E. Van De, Saey, T. (2013) Key variables for the identification of soil management classes in the aeolian landscapes of north–west Europe. *Geoderma* 199:99–105. https://doi.org/https://doi.org/10.1016/j.geoderma.2012.07.017

Miao, Y., Mulla, D. J., Robert, P. C. (2018) An integrated approach to site-specific management zone delineation. *Front Agric Sci Eng* 5:432–441. https://doi.org/10.15302/J-FASE-2018230

Michelon, G. K., Bazzi, C. L., Upadhyaya, S., de Souza, E. G., Magalhães, P. S. G., Borges, L. F., Betzek, N. M. (2019) Software AgDataBox-Map to precision agriculture management. *SoftwareX*, 10, 100320, 7 pp.

Mieza, M. S., Cravero, W. R., Kovac, F. D., Bargiano, P. G. (2016) Delineation of site-specific management units for operational applications using the topographic position index in La Pampa, Argentina. *Comput. Electron. Agric* 127:158-167.

Minasny, B., McBratney, A. B. (2002) FuzME version 3.0. Australian Centre for Precision Agriculture, The University of Sydney, Australia.

Mishra, U., Clay, D., Trooien, T., Dalsted, K., Malo, D., Carlson, C. G. (2008) Assessing the value of using a remote sensing-based evapotranspiration map in site-specific management. *J Plant Nutr* 31:1188–1202. https://doi.org/10.1080/01904160802134491

Molin, J. P. (2002) Definição de Unidade de Manejo a Partir de Mapas de Produtividade. *Eng Agrícola 22*:83–92.

Molin, J. P., Castro, C. N. De (2008) Establishing management zones using soil electrical conductivity and other soil properties by the fuzzy clustering technique. *Sci Agric* 65:567–573. https://doi.org/10.1590/S0103-90162008000600001

Monmonier M. (1996). How to Lie with Maps. University of Chicago Press, Chicago, IL.

Moral, F. J., Rebollo, F. J., Serrano, J. M. (2019) Estimating and mapping pasture soil fertility in a portuguese montado based on a objective model and geostatistical techniques. *Comput Electron Agric* 157:500–508. https://doi.org/https://doi.org/10.1016/j.compag.2019.01.033

Moral, F. J., Terrón, J. M., Rebollo, F. J. (2011) Site-specific management zones based on the Rasch model and geostatistical techniques. *Comput Electron Agric* 75:223–230. https://doi.org/10.1016/j.compag.2010.10.014

Moral, F. J., Terrón, J. M., Silva, J. R. M. Da (2010) Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. *Soil Tillage Res* 106:335–343. https://doi.org/10.1016/j.still.2009.12.002

Morari, F., Castrignanò, A., Pagliarin, C. (2009) Application of multivariate geostatistics in delineating management zones within a gravelly vineyard using geo-electrical sensors. *Comput Electron Agric* 68:97–107. https://doi.org/10.1016/j.compag.2009.05.003

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

Mulla, D., Khosla, R. (2016) Historical evolution and recent advances in precision farming. Soil specific farming. CRS Press, Boca Raton, FL. https://www. taylorfrancis. com/books/9781482245349.

Mzuku, M., Khosla, R., Reich, R., Inman, D., Smith, F., MacDonald, L. (2005). Spatial variability of measured soil properties across site-specific management zones. *Soil Sci. Soc. Am. J.* 69, 1572–1579.

Nascimento, A. L. Do, Sousa, E. D. T. dos S., Santos, F. F. L. Dos, Almeida, S. L. H. De, Valente, D. S. M. (2019) Confused Pixels Interference in Maps of Agricultural Management Zones. *J Exp Agric Int* 30:1–10. https://doi.org/10.9734/JEAI/2019/46239

Nawar, S., Corstanje, R., Halcro, G., Mulla, D., Mouazen, A. M. (2017) Delineation of Soil Management Zones for Variable-Rate Fertilization: A Review. In: Sparks, DL (ed) Advances in Agronomy, vol 143. Elsevier academic press.175–245.

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

Odeh, I. O. A., Chittleborough, D. J., McBratney, A. B. (1992) Soil Pattern Recognition with Fuzzy-c-means: Application to Classification and Soil-Landform Interrelationships. *Soil Sci Soc Am J* 56:505. https://doi.org/10.2136/sssaj1992.03615995005600020027x

Oldoni, H., Bassoi, L. H. (2016) Delineation of irrigation management zones in a Quartzipsamment of the Brazilian semiarid region. *Pesqui Agropecu Bras* 51:1283–1294. https://doi.org/10.1590/S0100-204X2016000900028

Ortuani, B., Chiaradia, E. A., Priori, S., L'Abate, G., Canone, D., Comunian, A., Giudici, M., Mele, M., Facchi, A. (2016) Mapping Soil Water Capacity Through EMI Survey to Delineate Site-Specific Management Units Within an Irrigated Field. *Soil Sci* 181:252–263. https://doi.org/10.1097/SS.0000000000000159

Ortuani, B., Sona, G., Ronchetti, G., Mayer, A., Facchi, A. (2019). Integrating Geophysical and Multispectral Data to Delineate Homogeneous Management Zones within a Vineyard in Northern Italy. *Sensors*, 19:3974.

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. (2014) Delineation of management zones and response of spring wheat (Triticum aestivum) to irrigation and nutrient levels in Saudi Arabia. *Int J Agric Biol* 16:104–110

Peralta, N. R., Costa, J. L. (2013) Delineation of management zones with soil apparent electrical conductivity to improve nutrient management. *Comput Electron Agric* 99:218–226. https://doi.org/10.1016/j.compag.2013.09.014

Peralta, N. R., Costa, J. L., Balzarini, M., Angelini, H. (2013) Delineation of management zones with measurements of soil apparent electrical conductivity in the southeastern pampas. *Can J Soil Sci* 93:205–218. https://doi.org/10.4141/cjss2012-022

Peralta, N. R., Costa, J. L., Balzarini, M., Castro Franco, M., Córdoba, M., Bullock, D., (2015) Delineation of management zones to improve nitrogen management of wheat. *Comput Electron Agric* 110:103–113. https://doi.org/10.1016/j.compag.2014.10.017

Ping, J.L., Dobermann, A. (2003) Creating spatially contiguous yield classes for site-specific management. *Agron J* 95:1121–1131.

Reyes, J., Wendroth, O., Matocha, C., Zhu, J. (2019) Delineating site-specific management zones and evaluating soil water temporal dynamics in a farmer's field in Kentucky. *Vadose Zone J* 18(1).

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

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

Roberts, D. F., Ferguson, R. B., Kitchen, N. R., Adamchuk, V. I., Shanahan, J. F. (2012) Relationships between soil-based management zones and canopy sensing for corn Nitrogen management. *Agron J* 104:119–129. https://doi.org/10.2134/agronj2011.0044

Robertson, M. J., Lyle, G., Bowden, J. W. (2008) Within-field variability of wheat yield and economic implications for spatially variable nutrient management. *F Crop Res* 105:211–220. https://doi.org/10.1016/j.fcr.2007.10.005

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

Rodrigues, M. S., Corá, J. E. (2015) Management zones using fuzzy clustering based on spatial-temporal variability of soil and corn yield. *Eng Agrícola* 35:470–483. https://doi.org/10.1590/1809-4430-Eng.Agric.

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

Ruß, G. (2013) From spatial data mining in precision agriculture to environmental data mining. Springer-Verlag, Berlin

Salami, B. T., Okogun, J. A., Sanginga, N. (2011) Delineation of management zones by classification of soil physico-chemical properties in the Northern Savanna of Nigeria. *African J Agric Res* 6:1572–1579. https://doi.org/10.5897/AJAR10.573

Santesteban, L. G., Guillaume, S., Royo, J. B., Tisseyre, B. (2013) Are precision agriculture tools and methods relevant at the whole-vineyard scale? *Precis Agric* 14:2–17. https://doi.org/10.1007/s11119-012-9268-3

Santi, A. L., Amado, T. J. C., Eitelwein, M. T., Cherubin, M. R., Silva, R. F. Da, Ros, C. O. Da (2013) Definição de zonas de produtividade em áreas manejadas com agricultura de precisão. *Rev Bras Ciencias Agrar* 8:510–515. https://doi.org/10.5039/aqraria.v8i3a2489

Santos, R. T., Saraiva, A. M. (2015) A Reference Process for Management Zones Delineation in Precision Agriculture. *IEEE Lat Am Trans* 13:727–738. https://doi.org/10.1109/TLA.2015.7069098

Schemberger, E. E., Fontana, F. S., Johann, J. A., Souza, E. G.(2017) Data Mining for the Assessment of Management Areas in Precision Agriculture. *Eng Agrícola* 37:185–193. https://doi.org/10.1590/1809-4430-eng.agric.v37n1p185-193/2017

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

Schenatto, K., Souza, E. G., Bazzi, C. L., Bier, V. A., Betzek, N. M., Gavioli, A. (2016b) Data interpolation in the definition of management zones. *Acta Sci - Technol* 38:31–34. https://doi.org/10.4025/actascitechnol.v38i1.27745

Schenatto, K., Souza, E. G., Bazzi, C. L., Betzek, N. M., Gavioli, A. (2016a) Two approach comparison to define crop management zones (MZs). *African J Agric Res* 11:3680–3689. https://doi.org/10.5897/ajar2016.11453

Schenatto, K., Souza, E. G., Bazzi, C. L., Gavioli, A., Betzek, N. M., Beneduzzi, H. M. (2017b) Normalization of data for delineating management zones. *Comput Electron Agric* 143:238–248. https://doi.org/10.1016/j.compag.2017.10.017

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

Schwalbert, R. A., Amado, T. J. C., Reimche, G. B., Gebert, F. (2018) Fine-tuning of wheat (Triticum aestivum, L.) variable nitrogen rate by combining crop sensing and management zones approaches in southern Brazil. *Precis Agric* 20:1–22. https://doi.org/10.1007/s11119-018-9581-6

Scudiero, E., Teatini, P., Corwin, D. L., Deiana, R., Berti, A., Morari, F. (2013) 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* 99:54–64. https://doi.org/https://doi.org/10.1016/j.compag.2013.08.023

Scudiero, E., Teatini, P., Manoli, G., Braga, F., Skaggs, T., Morari, F. (2018) Workflow to Establish Time-Specific Zones in Precision Agriculture by Spatiotemporal Integration of Plant and Soil Sensing Data. *Agronomy* 8:253. https://doi.org/10.3390/agronomy8110253

Servadio, P., Bergonzoli, S., Verotti, M. (2017) Delineation of management zones based on soil mechanical-chemical properties to apply variable rates of inputs throughout a field (VRA). *Eng Agric Environ Food* 10:20–30. https://doi.org/10.1016/j.eaef.2016.07.001

Shaddad, S. M., Madrau, S., Castrignanò, A., Mouazen, A. M. (2016) Data fusion techniques for delineation of site-specific management zones in a field in UK. *Precis Agric* 17:200–217. https://doi.org/10.1007/s11119-015-9417-6

Shamal, S. A. M., Alhwaimel, S. A., Mouazen, A. M. (2016) Application of an on-line sensor to map soil packing density for site specific cultivation. *Soil Tillage Res* 162:78–86. https://doi.org/10.1016/j.still.2016.04.016

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. (2017) Spatial

Distribution and Management Zones for Sulphur and Micronutrients in Shiwalik Himalayan Region of India. *L Degrad Dev* 28:959–969. https://doi.org/10.1002/ldr.2673

Silva, F. O. Espacialização dos teores de proteína e óleo de soja e milho e sua correlação com os atributos do solo e produtividade. 2016. 1009 f. Dissertação (Mestrado em Engenharia Agrícola) – Universidade Estadual do Oeste do Paraná, 2016.

Sinfield, J. V., Fagerman, D. and Colic, O., 2010. Evaluation of sensing technologies for on-the-go detection of macro-nutrients in cultivated soils. *Comput Electron Agric*, 70(1), pp.1-18.

Sobjak, R., Souza, E. G. De, Bazzi, C. L., Uribe-Opazo, M. A., Betzek, N. M. (2016) Redundant variables and the quality of management zones. *Eng Agrícola* 36:78–93. https://doi.org/10.1590/1809-4430-Eng.Agric.v36n1p78-93/2016

Song, X., Wang, J., Huang, W., Liu, L., Yan, G., Pu, R. (2009) The delineation of agricultural management zones with high resolution remotely sensed data. *Precis Agric* 10:471–487. https://doi.org/10.1007/s11119-009-9108-2

Souza, E. G. De, Schenatto, K., Bazzi, C. L. (2018) Creating thematic maps and management zones for agriculture fields. In: 14th International Conference on Precision Agriculture. Montreal, pp 1–17

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

Srinivasan (2006). Handbook of Precision Agriculture: Principles and Applications. Food Products Press, New York, NY. 708 p.

Sudduth, K.A., Drummond, S.T. (2007). Yield editor. *Agron J* 99:1471-1482.

Suszek, G., Souza, E. G. De, Uribe-Opazo, M. A., Nobrega, L. H. P. (2011) Determination of management zones from normalized and standardized equivalent produtivity maps in the soybean culture. *Eng Agrícola* 31:895–905. https://doi.org/10.1590/S0100-69162011000500007

Tagarakis, A., Liakos, V., Fountas, S., Koundouras, S., Gemtos, T. A. (2013) Management zones delineation using fuzzy clustering techniques in grapevines. *Precis Agric* 14:18–39. https://doi.org/10.1007/s11119-012-9275-4

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. (2017) Review of IoT applications in agro-industrial and environmental fields. *Comput Electron Agric* 142:283–297. https://doi.org/10.1016/j.compag.2017.09.015

Taylor, J. A., McBratney, A. B., Whelan, B. M. (2007) Establishing Management Classes for Broadacre Agricultural Production. *Agron J* 99:1366. https://doi.org/10.2134/agronj2007.0070

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

Timlin, D. J., Pachepsky, Y., Snyder, V. A., Bryant, R. B. (1998). Spatial and temporal variability of corn grain yield on a hillslope. *Soil Sci. Soc. Am. J.* 62, 764.

Tripathi, R., Nayak, A. K., Shahid, M., Lal, B., Gautam, P., Raja, R., Mohanty, S., Kumar, A., Panda, B. B., Sahoo, R. N. (2015) Delineation of soil management zones for a rice cultivated area in eastern India using fuzzy clustering. *Catena* 133:128–136. https://doi.org/10.1016/j.catena.2015.05.009

Urretavizcaya, I., Santesteban, L. G., Tisseyre, B., Guillaume, S., Miranda, C., Royo, J. B. (2014) Oenological significance of vineyard management zones delineated using early grape sampling. *Precis Agric* 15:111–129. https://doi.org/10.1007/s11119-013-9328-3

Valente, D. S. M., Queiroz, D. M. De, Pinto, F. De A. De C., Santos, N. T., Santos, F. L. (2012) Definition of management zones in coffee production fields based on apparent soil electrical conductivity. *Sci Agric* 69:173–179. https://doi.org/10.1590/S0103-90162012000300001

Vega, A., Córdoba, M., Castro-Franco, M., Balzarini, M. (2019). Protocol for automating error removal from yield maps. *Precis Agric* 20:1030-1044.

Velandia, M., Rejesus, R. M., Bronson, K., Segarra, E. (2008) Economics of management zone delineation in cotton precision agriculture. *J Cotton Sci* 12:210–227. https://doi.org/http://www.scopus.com/inward/record.url?eid=2-s2.0-84892084118&partnerID=40&md5=c7c7c2a97aa378a56f66c9c6904ab7b6

Verma, R. R., Manjunath, B. L., Singh, N. P., Kumar, A., Asolkar, T., Chavan, V., Srivastava, T. K., Singh, P. (2018) Soil mapping and delineation of management zones in the Western Ghats of coastal India. *L Degrad Dev* 29:4313–4322. https://doi.org/10.1002/ldr.3183

Viscarra Rossel R. A., McBratney A. B. (1998) Laboratory evaluation of a proximal sensing technique for simultaneous measurement of clay and water content. *Geoderma*, 85:19-39.

Vitharana, U. W. A., Meirvenne, M. Van, Simpson, D., Cockx, L., Baerdemaeker, J. De (2008a) Key soil and topographic properties to delineate potential management classes for precision agriculture in the European loess area. *Geoderma* 143:206–215. https://doi.org/https://doi.org/10.1016/j.geoderma.2007.11.003

Vitharana, U. W. A., Meirvenne, M. Van, Simpson, D., Cockx, L., Hofman, G. (2008b) Agronomic consequences of potential management zones delineated on the basis of EM38DD measurements. *Near Surf Geophys* 6:289–296. https://doi.org/10.3997/1873-0604.2008003

Webster, R., Oliver, M. A. (1990). Statistical methods in soil and and land resource survey. Oxford Univ. Press, New York.

Whetton, R. L., Waine, T. W., Mouazen, A. M. (2018) Evaluating management zone maps for variable rate fungicide application and selective harvest. *Comput Electron Agric* 153:202–212. https://doi.org/https://doi.org/10.1016/j.compag.2018.08.004

Wohlin, C. (2014) 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. pp 1–10

Wollenhaupt, N. C., Wolkowski, R. P., Clayton, M. K. (1994) Mapping soil test phosphorus and potassium for variable-rate fertilizer application. *J. Prod. Agric*. 7, 441–448.

Xiang, L., Yu-chun, P., Zhong-qiang, G., Chun-jiang, Z. (2007) Delineation and Scale Effect of Precision Agriculture Management Zones Using Yield Monitor Data Over Four Years. *Agric Sci China* 6:180–188. https://doi.org/10.1016/S1671-2927(07)60033-9

Xiaohu, Z., Li, J., Xiaolei, Q., Jianxiu, Q., Juan, W., Yan, Z., Zhang, X., Jiang, L., Qiu, X., Qiu, J., Wang, J., Zhu, Y., (2016) An improved method of delineating rectangular management zones using a semivariogram-based technique. *Comput Electron Agric* 121:74–83. https://doi.org/10.1016/j.compag.2015.11.016

Xin-Zhong, W., Guo-Shun, L., Hong-Chao, H., Zhen-Hai, W., Qing-Hua, L., Xu-Feng, L., Wei-Hong, H., Yan-Tao, L. (2009) Determination of management zones for a tobacco field based on soil fertility. *Comput Electron Agric* 65:168–175. https://doi.org/https://doi.org/10.1016/j.compag.2008.08.008

Yan, L., Zhou, S., and Feng, L. (2007a) Delineation of site-specific management zones based on temporal and spatial variability of soil electrical conductivity. *Pedosphere*, 17, 156–164.

Yan, L., Zhou, S., Feng L., and L. Hong-yi (2007b) Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. *Comput Electron Agric*. 56:174-186.

Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H., Xu, B., Yang, X., Zhu, D., Zhang, X., Zhang, R. (2017) Unmanned aerial vehicle remote sensing for field-based crop phenotyping: current status and perspectives. *Front Plant Sci* 8:1111.

Yao, R., Yang, J. S., Zhang, T., Gao, P., Wang, X., Hong, L., Wang, M. (2014) Determination of site-specific management zones using soil physico-chemical properties and crop yields in coastal reclaimed farmland. *Geoderma* 232–234:381–393. https://doi.org/https://doi.org/10.1016/j.geoderma.2014.06.006

Yari, A., Madramootoo, C. A., Woods, S. A., Adamchuk, V. I., Huang, H. (2017) Assessment of Field Spatial and Temporal Variabilities to Delineate Site-Specific Management Zones for Variable-Rate Irrigation. *J Irrig Drain Eng* 143:04017037. https://doi.org/10.1061/(ASCE)IR.1943-4774.0001222

Zhang, X., Shi, L., Jia, X., Seielstad, G., Helgason, C. (2010) Zone mapping application for precision-farming: A decision support tool for variable rate application. *Precis Agric* 11:103–114. https://doi.org/10.1007/s11119-009-9130-4


PONCÃ