ISSN 1807-1929
Revista Brasileira de Engenharia Agrícola e Ambiental
Brazilian Journal of Agricultural and Environmental Engineering
v.28, n.9, e278672, 2024
Campina Grande, PB – http://www.agriambi.com.br – http://www.scielo.br/rbeaa

DOI: http://dx.doi.org/10.1590/1807-1929/agriambi.v28n9e278672

Review

Data processing to remove outliers and inliers: A systematic literature study[1]

Processamento de dados para remoção de pontos outliers e inliers: Estudo sistemático da literatura

Fernando Alves[2], Eduardo G. de Souza[3*], Ricardo Sobjak[4], Claudio L. Bazzi[4], Antonio M. M. Hachisuca[3] & Erivelto Mercante[3]

[1] Research developed at Universidade Estadual do Oeste do Paraná, Cascavel, PR, Brazil
[2] Instituto Federal do Paraná/Departamento de Ciência da Computação, Cascavel, PR, Brazil
[3] Universidade Estadual do Oeste do Paraná/Centro de Ciências Exatas e Tecnológicas/Pós-Graduação em Engenharia Agrícola, Cascavel, PR, Brazil
[4] Universidade Tecnológica Federal do Paraná/Departamento de Ciência da Computação, Medianeira, PR, Brazil

HIGHLIGHTS:

Sample data acquisition often exhibits outliers and inliers.
Outlier and inlier techniques include Chebyshev's inequality, boxplot, principal component analysis, and local Moran's index.
This systematic literature search effectively identified and filtered the research papers.

ABSTRACT: Outliers and inliers often arise during sample data acquisition. While outliers represent anomalous observations, inliers are erroneous data points within the main body of the dataset. It was aimed to conduct a systematic literature study (SLS) to survey methods and software employed for outlier and inlier removal, particularly within exploratory data analysis. The study was conducted in three phases: (i) systematic literature mapping (SLM), (ii) snowballing (SB), and (iii) SLR. Initially, 772 scientific studies were identified, subsequently narrowed down to 86 after applying selection criteria. Backward (BSB) and forward (FSB) snowballing further yielded 16 studies, resulting in a final pool of 102 studies for analysis. It was identified three outlier removal techniques (Chebyshev's inequality, boxplot, and principal component analysis), one inlier removal technique (local Moran's index), and thirteen commonly used software.

Key words: exploratory analysis, precision agriculture, local Moran's index, data cleaning

RESUMO: Outliers e inliers aparecem frequentemente na aquisição de dados amostrais. Outliers são observações anômalas e inliers são dados errôneos no interior do conjunto de dados. Esta pesquisa teve como objetivo realizar um estudo sistemático da literatura (SLS) para levantar os métodos utilizados para remoção de outliers e inliers e os softwares utilizados na análise exploratória dos dados. Este estudo sistemático da literatura foi realizado em três etapas: (i) mapeamento sistemático da literatura (SLM), (ii) bola de neve (SB) e (iii) revisão sistemática da literatura (SLR). Setecentos e setenta e dois estudos científicos foram obtidos e reduzidos para 86 após seleção. Foram acrescentados mais dezessete estudos selecionados por bola de neve (bola de neve para trás (BSB) e bola de neve para frente (FSB)), o que resultou em 102 estudos utilizados nesta pesquisa. Foram observadas três técnicas de remoção de outliers (desigualdade de Chebyshev, boxplot e análise de componentes principais), uma única técnica de remoção de inliers (índice de Moran local) e treze softwares.

Palavras-chave: análise exploratória; agricultura de precisão; índice de Moran local; limpeza de dados

• Ref. 278672 – Received 18 Sept, 2023
* Corresponding author - E-mail: eduardo.souza@unioeste.br
• Accepted 14 Apr, 2024 • Published 10 Jun, 2024
Editors: Toshik Iarley da Silva & Carlos Alberto Vieira de Azevedo

Introduction

Higher crop yields have long been a priority, driven especially by the increasing global population. This has necessitated not only the expansion of productive areas but also the constant evolution of management methods. Extensive research has shown that the spatial variability of yield in agricultural areas has diverse causes (Trevisan et al., 2021).

"Precision Agriculture (PA) is a management strategy that gathers, processes and analyzes temporal, spatial and individual plant and animal 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, 2024). PA is essential for maximizing output while minimizing input use, reducing environmental impacts, and ensuring sustainability (Karunathilake et al., 2023).

Raw and semi-processed agricultural data are typically collected from various sources, including the Internet of Things (IoT), sensors, satellites, weather stations, robots, farm equipment, farmers, and agribusinesses.

Before analyzing collected data, a statistical and exploratory data analysis is essential. This analysis utilizes techniques to gain deeper insights into the dataset, extract key variables, and identify outliers, inliers, and anomalies. Outliers are unexpected observations that deviate significantly from the majority of data points. Outlier detection and prediction are challenging tasks due to the inherent rarity of outliers (Reunanen et al., 2020). Inliers, on the other hand, are data points that differ from their immediate neighbors but still fall within the overall range of variation in the dataset (Córdoba et al., 2016; Vega et al., 2019).

However, some crucial questions remain regarding the descriptive and exploratory statistical information used in PA, particularly the methods and software employed for outlier and inlier removal. Therefore, this systematic literature study (SLS) (Fisch & Block, 2018; Abbasi et al., 2022) focused on identifying research related to data treatments in PA, specifically methods for removing outliers and inliers.

Material and Methods

This study employed three search techniques to identify relevant studies for data cleaning:

(i) Systematic Literature Mapping (SLM), which uses targeted keywords to uncover research on a specific topic across various databases. To do so, it was adhered to well-defined SLM principles outlined by previous studies (Talavera et al., 2017; Aikes Junior et al., 2021; Beneduzzi et al., 2022; Moreira et al., 2022), defining relevant keywords, selecting suitable databases, establishing selection criteria, analyzing the retrieved studies, and synthesizing their findings. To guide this keyword selection, it was formulated three key questions: 1. What are the most frequently used methods for outlier removal in precision agriculture (PA)? 2. What are the most common methods for removing inliers in PA? 3. Which software programs are most widely employed for cleaning harvester data?

The Science Direct database was accessed through the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) portal using the Federated Academic Community (CAFe) remote access platform (Figure 1). To capture the

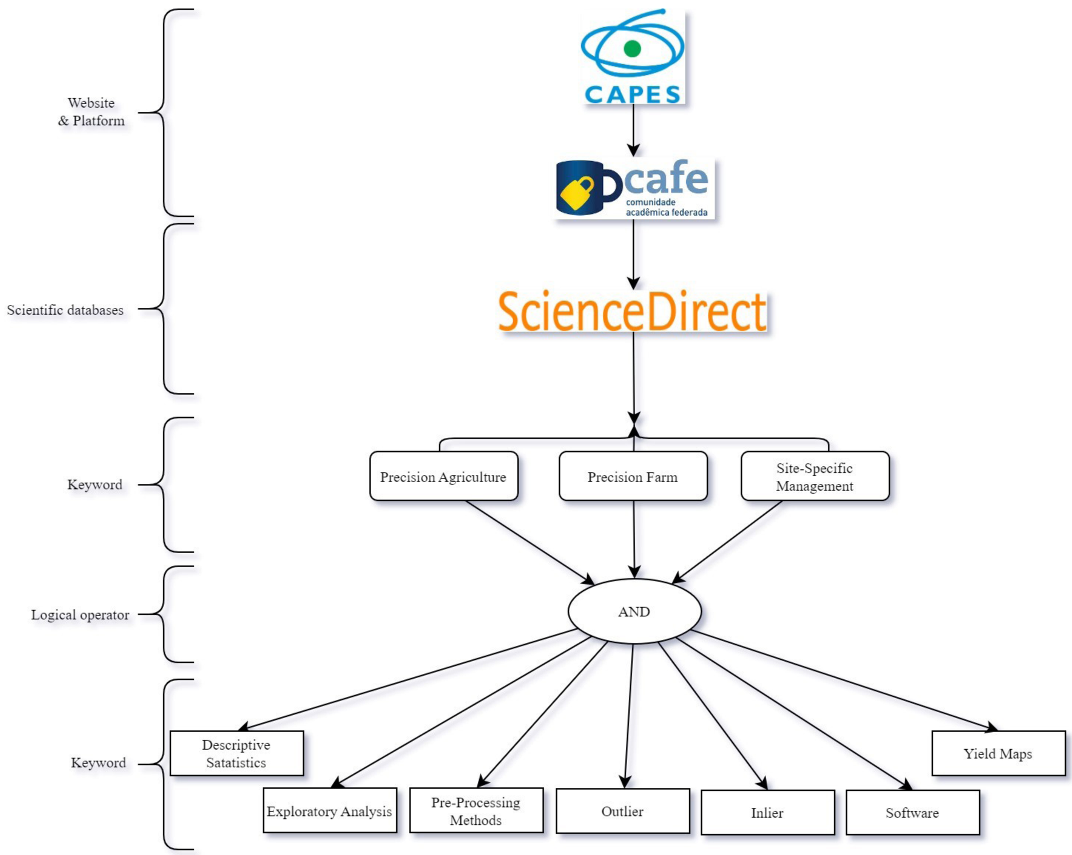

Figure 1. Workflow for searches on scientific bases

most recent advancements, it was searched for publications from 2009 to 2019. From each identified study, it was extracted specific details crucial for this analysis, including title, authors, journal, year of publication, abstract, keywords, any software used, and reported results.

(ii) Snowballing (SB), a complementary technique, leveraged the reference lists of identified studies to uncover additional relevant articles (Li & Zhan, 2022; Moreira et al., 2022). This boosted search efficiency, mitigated potential biases (Aikes Junior et al., 2021; Beneduzzi et al., 2022), and even unearthed classic texts from conventional agriculture (Beneduzzi et al., 2022) that might have been missed in the initial search.

Two types of snowballing were employed: 1. Backward snowballing (BSB), which included new studies related to those found in the initial SLM phase; and 2. Forward snowballing (FSB), which built upon BSB by using the reference lists of those identified studies to find even more relevant articles (Aikes Junior et al., 2021; Beneduzzi et al., 2022; Moreira et al., 2022).

(iii) Systematic literature review (SLR) then integrated the findings from both SLM and SB. To ensure a targeted and efficient review, the study selection followed the following steps (Figure 2): 1. Retrieving and deduplicating articles, where studies were retrieved from the database removing duplicates, resulting in a pool of 772 studies; 2. Title screening, where the titles were reviewed to identify relevant articles, narrowing the selection down to 427; 3. Abstract analysis, which consisted of further refinement through abstract review, leaving 156 studies for in-depth evaluation; 4. Full-text reading and assessment, where the remaining 86 studies were carefully read and assessed; 5. Snowballing integration, where additional studies identified through snowballing were incorporated (10 from BSB and 6 from FSB). This meticulous process yielded a final pool of 102 studies for analysis (Table 1).

In terms of the geographic distribution of the selected studies (Figure 3), every continent except Antarctica was represented by at least one research paper. As for the distribution across different publication outlets (Figure 4), the journals Computers and Electronics in Agriculture and Geoderma exhibited the highest number of selected studies, accounting for 19 and 16%, respectively. Journals featuring only one study in the review were consolidated into a single category labeled "others," comprising 19% of the total.

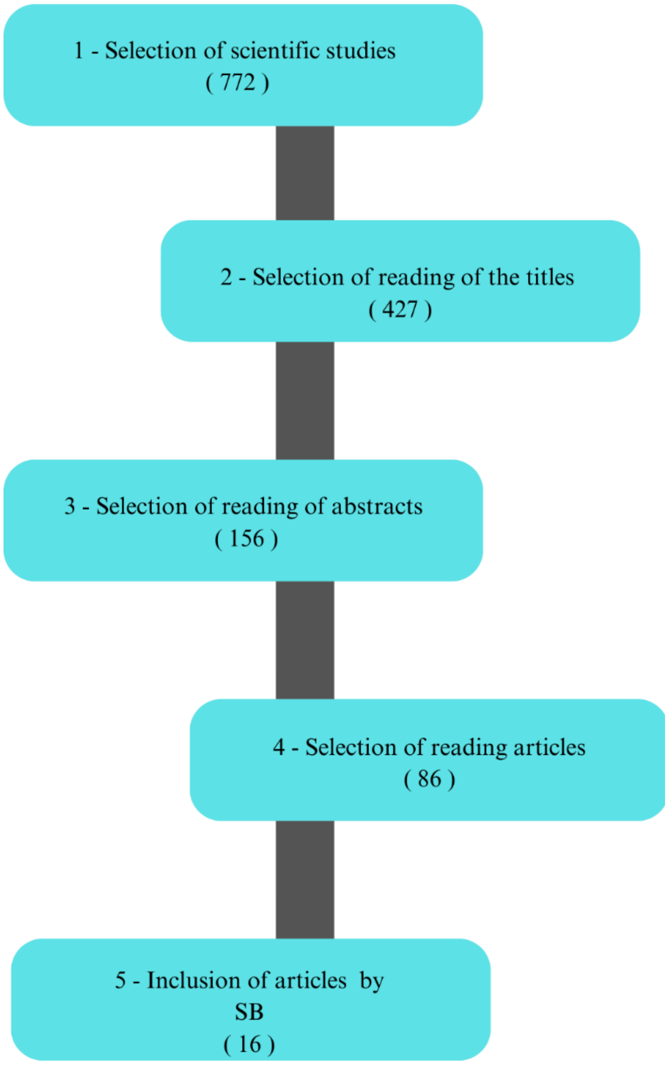

Figure 2. Workflow used for Systematic Literature Study (SLS)

Table 1. Selection after reading the studies (102 selected). Systematic literature mapping (SLM, 86 references); Backward Snowballing (BSB, 10 references); and Forward Snowballing (FSB, 6 references)

Technique	Studies
SLM	(Lamsal et al., 2009); (Liu et al., 2009); (Pei et al., 2010); (Eugster et al., 2010); (Vašát et al., 2010); (Gianquinto et al., 2011); (Amirinejad et al., 2011); (Tesfahunegn et al., 2011) ; (Davatgar et al., 2012) ; (Henriques et al., 2012) ; (Li et al., 2012); (Castrignanò et al., 2012); (Sudduth et al., 2012); (Van Meirvenne et al., 2013); (Chung et al., 2013); (Oliver & Webster, 2014); (Eitel et al., 2014); (Yao et al., 2014); (Mat et al., 2014); (Kanellopoulos et al., 2014); (Waruru et al., 2015); (Landrum et al., 2015); (Kaniu & Angeyo, 2015); (Tripathi et al., 2015); (Wang et al., 2015); (Stockmann et al., 2015); (Calafat et al., 2015); (Haghverdi et al., 2015); (Knadel et al., 2015); (Driemeier et al., 2016); (Turner et al., 2016); (Mieza et al., 2016); (Castaldi et al., 2016); (Páscoa et al., 2016); (Bitencourt et al., 2016); (Morellos et al., 2016); (Gerstmann et al., 2016); (Rodriguez-Moreno et al., 2016) ; (Li et al., 2016) ; (Cavallo et al., 2016); (Córdoba et al., 2016) ; (Jordanova et al., 2016); (Pelosi et al., 2016); (Nouri et al., 2017); (Gili et al., 2017); (Demattê et al., 2017); (Bodner & Robles, 2017); (Rosemary et al., 2017); (Mirzaeitalarposhti et al., 2017); (Adeline et al., 2017); (Medina et al., 2017); (Paraforos et al., 2017); (Castrignanò et al., 2018); (Coelho et al., 2018) ; (Neave et al., 2018) ; (Paraforos et al., 2018) ; (Sirsat et al., 2018) ; (Camino et al., 2018); (Leroux et al., 2018); (Liang et al., 2018); (Fujinuma et al., 2018); (Schönhart et al., 2018); (Squalli & Adamkiewicz, 2018); (Sanches et al., 2018); (Altdorff et al., 2018); (Behera et al., 2018); (Gholizadeh et al., 2018); (Raj et al., 2018); (Uribeetxebarria et al., 2018); (Hong et al., 2019); (Barca et al., 2019); (Krishna et al., 2019); (Zhou et al., 2019); (Betzek et al., 2019); (González-Fernández et al., 2019); (Liu et al., 2019); (Gavioli et al., 2019); (Araujo et al., 2019); (Ali et al., 2019); (Moura-Bueno et al., 2019); (Prasad et al., 2019); (Shaddad et al., 2019); (Sanches et al., 2019); (Uribeetxebarria et al., 2019); (Mura et al., 2019); (Vega et al., 2019)
BSB	(Arslan & Colvin, 2002); (Yield et al., 2004); (Menegatti & Molin, 2004); (Amidan et al., 2005); (Ping & Dobermann, 2005); (Dray et al., 2006); (Taylor et al., 2007); (Sudduth & Drummond, 2007); (Maldaner & Molin, 2020); (Paccioretti et al., 2020)
FSB	(Hotelling, 1933); (Gnanadesikan, 1977); (Barnett & Lewis, 1994); (Anselin, 1995); (Blackmore & Moore, 1999); (Jolliffe, 2005)

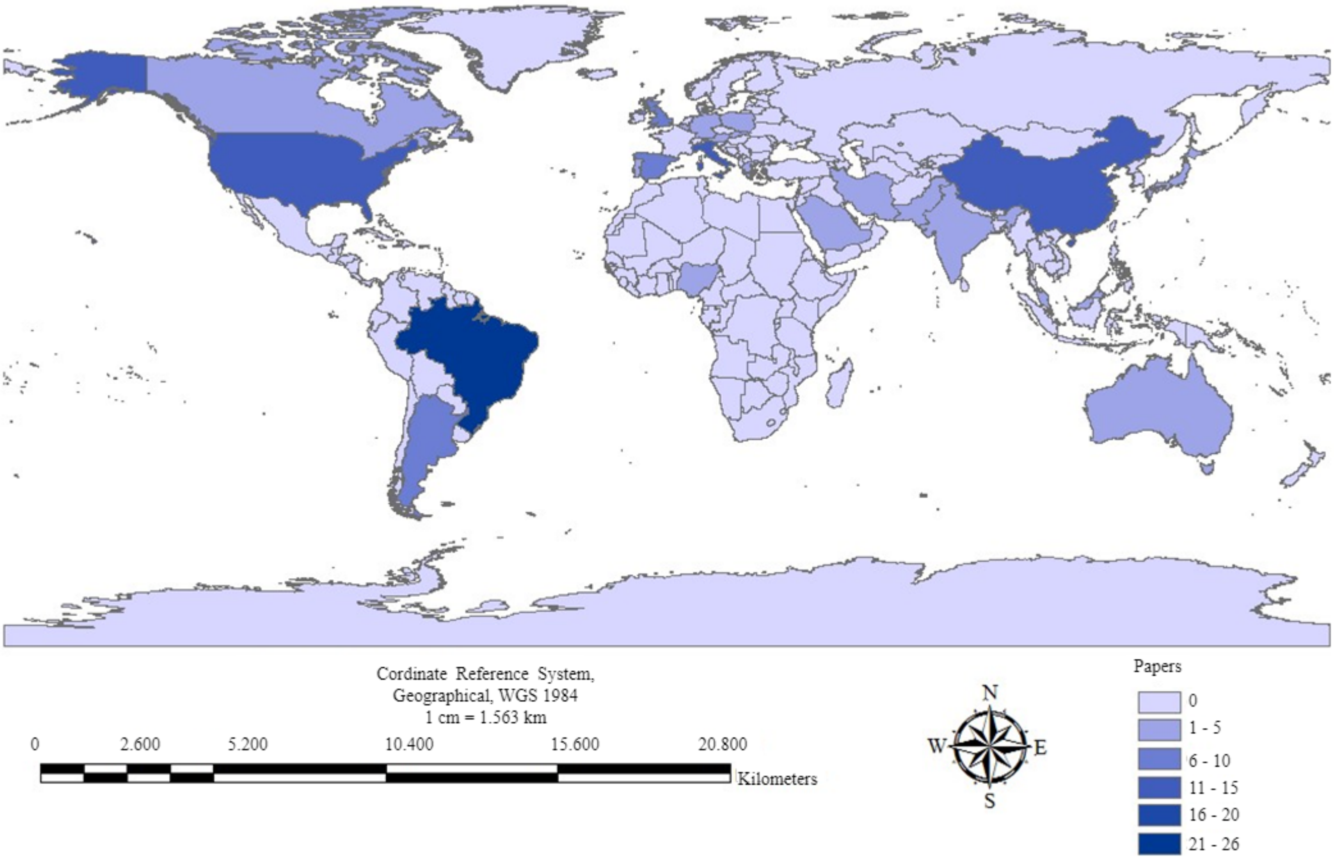

Figure 3. Distribution of selected studies per country (where authors carried out the study and, in cases of theoretical manuscript, where it was published)

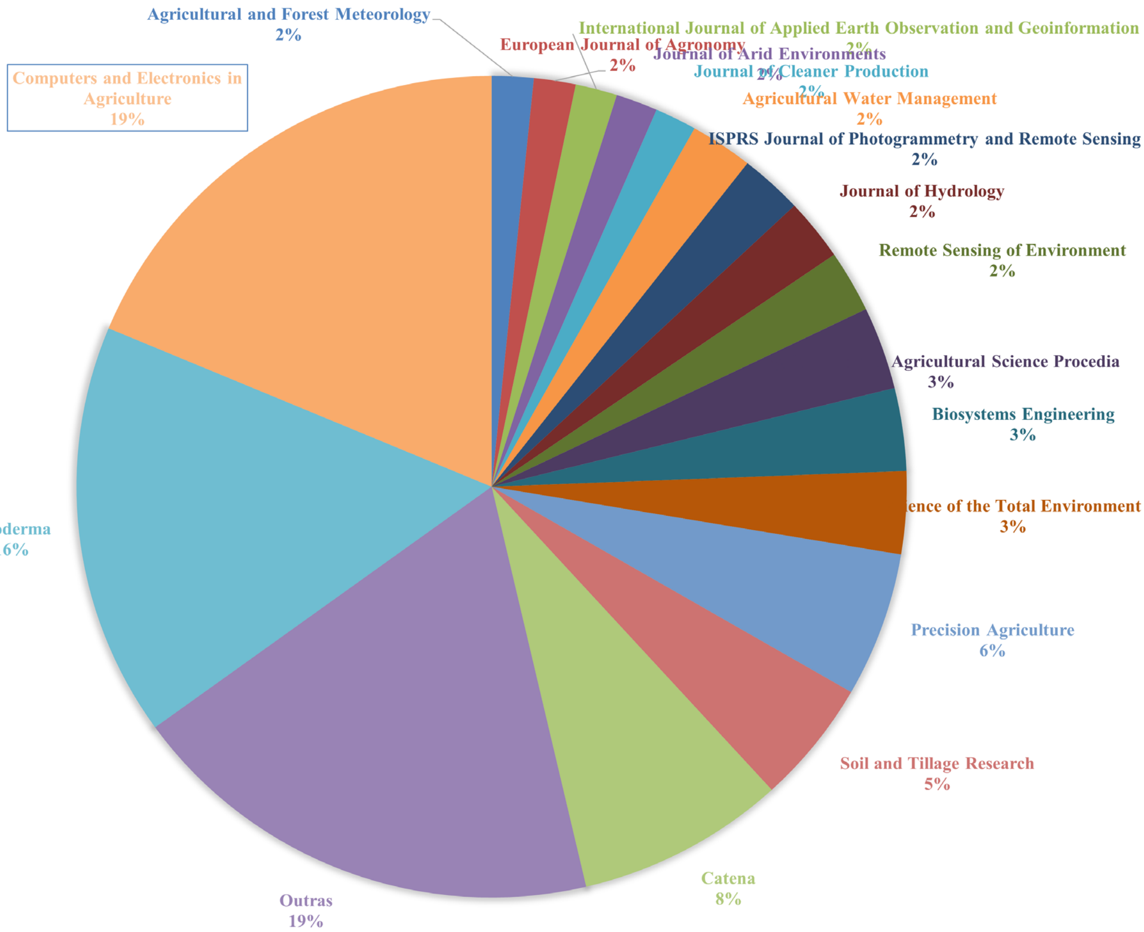

Figure 4. Scientific journals where the selected studies were published

Results and Discussion

Numerous sensors used for retrieving spatial data are integrated into agricultural machinery. However, errors in their data can arise during the construction of digital maps (Spekken et al., 2005). For instance, the movement of the harvester into and out of the crop field, as well as areas with minimal or no crop, can lead to gradual changes in yield readings rather than sudden fluctuations, as expected (Kharel et al., 2019). Another potential source of errors is associated with the GPS (Global Positioning System) signal receiver, now commonly referred to as the GNSS (Global Navigation Satellite System) receiver (Blackmore & Moore, 1999). The yield measurement using a monitor is automated, allowing for the collection of a large volume of data. Hence, errors may be introduced (Menegatti & Molin, 2004), including georeferencing errors, sensor inaccuracies, and operational errors (Sun et al., 2013).

Outlier detection represents a primary task preceding analysis, aiming to identify anomalies within extensive datasets. Various terms, such as noise detection, anomaly detection, deviation detection, or outlier detection, are used to describe this process. According to Barnett & Lewis (1994), outliers are observations that deviate significantly from the general pattern or distribution of the dataset, highlighting the importance of removing outliers before making agronomic decisions (Taylor et al., 2007). The removal of outliers is a crucial step that precedes the construction of thematic maps, ensuring the accuracy of agronomic decisions (Vega et al., 2019).

Identifying and potentially removing outliers during data collection and analysis is often imperative. Thus, it is essential to establish that a given methodology is both objective and efficient in identifying outliers for removal (Amidan et al., 2005). Additionally, several automated methods are available for detecting and eliminating outliers (Table 2).

Chebyshev's inequality was developed to establish a lower bound on the percentage of data within a certain range of standard deviations from the mean. For data following a normal (bell-shaped) distribution, approximately 95% of the data typically lies within two standard deviations from the mean. Consequently, around 5% of the data would be expected to fall outside this range (Amidan et al., 2005).

Chebyshev's inequality (Eq. 1) is applicable when the distribution of the data is unknown (Amidan et al., 2005):

$$P\left(\left|X-\mu\right| \le k\sigma\right) \ge \left(1-\frac{1}{k^2}\right) \tag{1}$$

where: X - data; μ - data mean; σ - standard deviation of the data; and, k - number of standard deviations of the mean.

Chebyshev's theorem asserts that a minimum of 89% of data falls within the range of the mean ± 3 standard deviations, irrespective of the distribution (Córdoba et al., 2016). Table 2 outlines various other proposals for lower and upper limits based on the dataset (Table 2).

Adjusting the lower and upper limits of the standard deviation parameter for a specific field area can be accomplished through the iterative selection of values while observing the impact on the resulting yield data distribution (Sudduth & Drummond, 2007). The interval of ± 3 standard deviations was the most commonly cited in the studies (Table 2).

The boxplot graph (Tukey, 1977) is a useful tool that segments data based on four hidden boundaries: two inner fences and two outer fences (Figure 5). Typically, the interquartile range (IQR) is defined as Q3-Q1. The inner fences are Q1 - 1.5 IQR and Q3 + 1.5 IQR, while the outer fences are Q1 - 3 IQR and Q3 + 3 IQR. The whiskers extend to the most extreme data points within the inner fences. Data points outside the inner fences but within the outer fences are considered mild outliers, each marked with a symbol. Data points outside the outer fences are classified as extreme outliers and are marked with a different symbol. There is no consensus regarding the choice of symbols (Dawson, 2011).

Boxplots are effective tools for visualizing data distribution in terms of quartiles and identifying outliers (Prasad et al., 2019). Consequently, several studies (Table 2) have utilized boxplots for outlier detection and removal.

Principal Components Analysis (PCA) (Hotelling, 1933) is a multivariate technique outlined in Table 2, involving the creation of a new set of orthogonal synthetic variables known as principal components (PC). The primary objective is to reduce the dimensionality of a dataset containing numerous interrelated variables while preserving as much variation as possible (Jolliffe, 2005).

Table 2. Outlier removal techniques

Technique	N	Studies
$\mu \pm 1.5\sigma$	1	(Gerstmann et al., 2016)
$\mu \pm 2\sigma$	4	(Pei et al., 2010); (Gianquinto et al., 2011); (Chung et al., 2013); (Rosemary et al., 2017)
$\mu \pm 2.5\sigma$	6	(Squalli & Adamkiewicz, 2018); (Castrignanò et al., 2012); (Cavallo et al., 2016); (Uribeetxebarria et al., 2018); (Uribeetxebarria et al., 2019); (Coelho et al., 2018)
$\mu \pm 3\sigma$	9	(Simbahan et al., 2004); (Liu et al., 2009); (Haghverdi et al., 2015); (Landrum et al., 2015); (Córdoba et al., 2016); (Driemeier et al., 2016); (Li et al., 2016); (Bodner & Robles, 2017); (Sanches et al., 2018); (Betzek et al., 2019); (Sanches et al., 2019); (Vega et al., 2019)
Boxplot	26	(Eugster et al., 2010); (Tesfahunegn et al., 2011); (Henriques et al., 2012); (Li et al., 2012); (Eitel et al., 2014); (Kanellopoulos et al., 2014); (Oliver & Webster, 2014); (Wang et al., 2015); (Castaldi et al., 2016); (Jordanova et al., 2016); (Pelosi et al., 2016); (Rodriguez-Moreno et al., 2016); (Turner et al., 2016); (Gili et al., 2017); (Medina et al., 2017); (Nouri et al., 2017); (Paraforos et al., 2017); (Fujinuma et al., 2018); (Liang et al., 2018); (Neave et al., 2018); (Paraforos et al., 2018); (Schönhart et al., 2018); (Sirsat et al., 2018); (Barca et al., 2019); (Liu et al., 2019); (Prasad et al., 2019)
PCA	15	(Hotelling, 1933); (Gnanadesikan, 1977); (Jolliffe, 2005); (Mat et al., 2014); (Waruru et al., 2015); (Kaniu & Angeyo, 2015); (Knadel et al., 2015); (Stockmann et al., 2015); (Morellos et al., 2016); (Páscoa et al., 2016); (Adeline et al., 2017); (Demattê et al., 2017); (Mirzaeitalarposhti et al., 2017); (Gholizadeh et al., 2018); (Moura-Bueno et al., 2019)

N - Number of studies

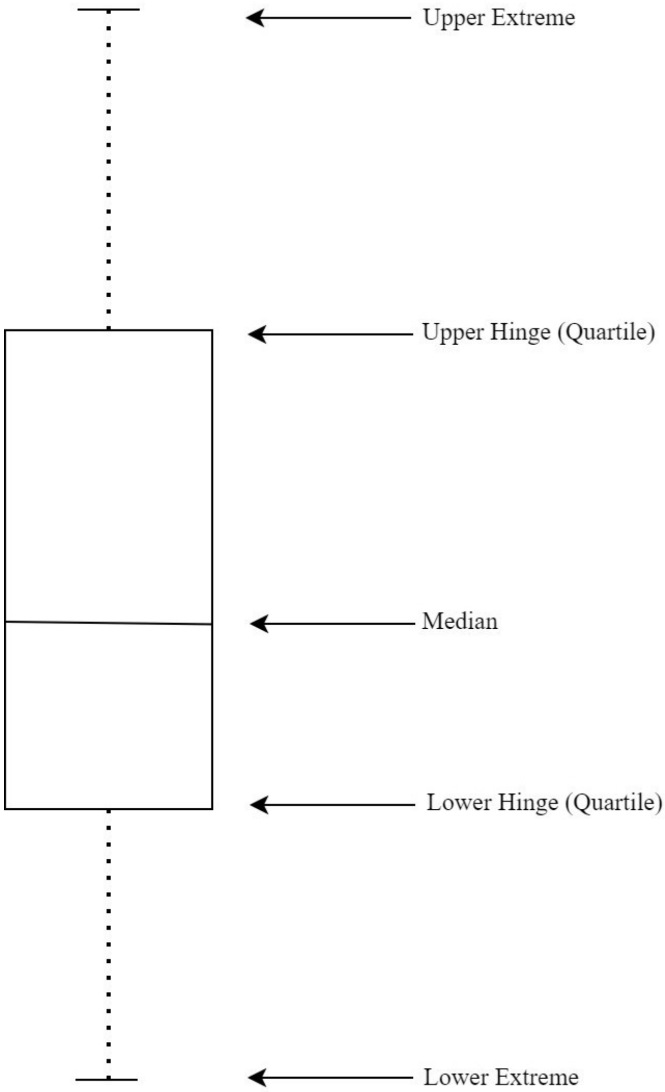

Figure 5. Boxplot structure

As described by Filzmoser et al. (2005), the covariance matrix characterizes the shape and size of multivariate data. The Mahalanobis distance is a commonly used distance measure that takes into account the covariance matrix. The Mahalanobis distance (Eq. 2) for a p-dimensional multivariate sample x_1, ..., x_n is given by:

$$MD_i = \left[\left(x_i - t\right)^T C^{-1}\left(x_i - t\right)\right]^{\frac{1}{2}} \quad \text{for } i = 1,\ldots,n \qquad (2)$$

where: t is the estimated multivariate location, and C is the covariance matrix. Typically, t is the multivariate arithmetic mean, the centroid, while C is the sample covariance matrix.

For multivariate normally distributed data, the squared Mahalanobis distances (MD_i^2) are approximately distributed as chi-square (χ_p^2) with p degrees of freedom. By equating the squared Mahalanobis distance to a specific constant, corresponding to a certain quantile of χ_p^2, it becomes feasible to define ellipsoids that have the same Mahalanobis distance from the centroid (Filzmoser et al., 2005).

Inliers refer to data points that deviate significantly from their local surroundings but still fall within the overall range of variation of the dataset (Córdoba et al., 2016; Vega et al., 2019). The Local Moran's Index of spatial autocorrelation (I_i) (Anselin, 1995) can be employed to identify inliers for each variable within the dataset. As described by Córdoba et al. (2016), I_i represents a Local Moran's Index (LISA) applied individually to each neighborhood, indicating the degree of similarity or dissimilarity between the observation's value and that of its neighbors (Table 3).

According to Vega et al. (2019), the spatial autocorrelation index I_i can be computed for yield data as a tool for identifying inliers, utilizing the following equation (Eq. 3):

$$I_i = \frac{Z_i - \bar{Z}}{\sigma^2} \sum_{i=1,j\neq i}^{n} \left[W_{ij}\left(Z_j - \bar{Z}\right)\right] \text{ for } i = 1,\ldots,n \qquad (3)$$

where: Z is the mean yield value, Z_i is the local yield at location i, Z_j is the yield at other locations ($j \neq i$), σ^2 is the yield variance, n is the sample size, and W_{ij} is the weighted distance between Z_i and Z_j. Neighboring points are defined as continuous to calculate W_{ij} (Anselin, 1995).

A positive Local Moran's (LM) index value indicates spatial clustering of similar values, whereas a negative value suggests clustering of dissimilar values. For instance, a site with a low yield value surrounded by neighbors with high yield values would result in a negative LM index value. The p-value associated with each yield data point evaluates the statistical significance of the LM value. These probability values are adjusted using Bonferroni's criterion to account for the multiplicity of indices evaluated simultaneously within the field. Bonferroni's adjustment is applied to each test performed to maintain the overall study's error rate at 0.05 (Vega et al., 2019).

No studies comparing the efficiency of the reported data-cleaning methods were found. However, it is worth noting that the most recent data-cleansing software predominantly employs statistical methods, likely due to their ease of use and effectiveness.

Software used for detecting and removing outliers and inliers

An extensive array of sensors, such as yield monitors, is commonly integrated with agricultural machinery to collect spatial data. However, these data often contain systematic and random errors, as well as atypical values, necessitating post-processing for their elimination. Various methods and protocols have been developed to apply filter sequences and classify, identify, and remove outliers and inliers from grain yield data (Córdoba et al., 2016; Leroux et al., 2018; Sudduth et al., 2018; Vega et al., 2019). According to Vega et al. (2019), approximately 30% of yield data were removed after data

Table 3. Inlier removal techniques

Technique	N	Studies
Local Moran's Index (LISA)	4	(Lamsal et al., 2009); (Calafat et al., 2015); (Córdoba et al., 2016); (Vega et al., 2019)

N - Number of studies

Table 4. Software programs used in data cleaning

Program	N	Studies	License
FastMapping	1	(Paccioretti et al., 2020)	Free
JMP Pro	1	(Uribeetxebarria et al., 2018)	Proprietary software
MapFilter	-		Free
Matlab	4	(Liang et al., 2018; Paraforos et al., 2018, 2017; Sanches et al., 2019)	Proprietary software
Python	1	(Coelho et al., 2018)	Free
SAS	3	(Chung et al., 2013; Neave et al., 2018; Barca et al., 2019)	Proprietary software
Isatis	2	(Castrignanò et al., 2012; Cavallo et al., 2016)v	Proprietary software
R Core	10	(Sudduth & Drummond, 2007a; Stockmann et al., 2015; Castaldi et al., 2016; Córdoba et al., 2016; Gili et al., 2017; Schönhart et al., 2018; Sirsat et al., 2018; Barca et al., 2019; Moura-Bueno et al., 2019; Vega et al., 2019)	Free
SPSS	5	(Tesfahunegn et al., 2011; Wang et al., 2015; Li et al., 2016; Medina et al., 2017; Rosemary et al., 2017)	Proprietary software
Statgraphics	1	(Gianquinto et al., 2011)	Proprietary software
Statistica	1	(Jordanova & Petrov, 2016)	Proprietary software
YieldEditor	2	(Sudduth & Drummond, 2007b; Sudduth et al., 2012)	Free

N - Number of studies

cleaning. A total of 13 software tools were used in different studies for data cleaning, including both proprietary and free software options (Table 4).

Conclusions

1. The systematic literature study (SLS) proved to be effective in identifying relevant research studies and enhancing understanding of the research subject.

2. The survey conducted through systematic literature mapping (SLM) yielded 772 scientific studies, which were narrowed down to 86 studies following the selection process. Additionally, sixteen more studies identified through snowballing (FSB and BSB) were incorporated, resulting in a total of 102 studies used in this research.

3. Three techniques for outlier removal were identified: (i) Chebyshev's inequality, (ii) boxplot analysis, and (iii) principal component analysis. In terms of removing inliers, all studies employed the local Moran's index. Furthermore, thirteen different software tools were identified across the analyzed studies.

Contribution of authors: Fernando Alves: Systematic literature study, Resources, Formal analysis, Writing – original draft. Eduardo G. de Souza: Conceptualization, Methodology, Writing – review & editing. Ricardo Sobjak: Formal analysis, Writing – review & editing. Claudio L. Bazzi: Methodology, Writing – review & editing. Antonio M. M. Hachisuca: Resources, Investigation. Erivelto Mercante: Resources, Writing – review & editing.

Supplementary documents: There are no supplementary sources.

Conflict of interest: The authors declare no conflict of interest.

Financing statement: There are not financing statements to declare.

Acknowledgements: The authors would like to thank the Universidade Estadual do Oeste do Paraná (UNIOESTE), the Universidade Tecnológica Federal do Paraná (UTFPR), the Instituto Federal do Paraná (IFPR), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES; finance code n° 001), and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Literature Cited

Abbasi, R.; Martinez, P.; Ahmad, R. The digitization of agricultural industry - a systematic literature review on agriculture 4.0. Smart Agricultural Technology, v.2, 100042, 2022. https://doi.org/10.1016/j.atech.2022.100042

Adeline, K. R. M.; Gomez, C.; Gorretta, N.; Roger, J.-M. Predictive ability of soil properties to spectral degradation from laboratory Vis-NIR spectroscopy data. Geoderma, v.288, p.143-153, 2017. https://doi.org/10.1016/j.geoderma.2016.11.010

Aikes Junior, J.; Souza, E.G.; Bazzi, C. L.; Sobjak, R. Thematic maps and management zones for precision agriculture - Systematic literature review, protocols and practical cases. Curitiba: Poncã, v.1, p.1-114, 2021.

Ali, M.; Deo, R. C.; Maraseni, T.; Downs, N. J. Improving SPI-derived drought forecasts incorporating synoptic-scale climate indices in multi-phase multivariate empirical mode decomposition model hybridized with simulated annealing and kernel ridge regression algorithms. Journal of Hydrology, v.576, p.164-184, 2019. https://doi.org/10.1016/j.jhydrol.2019.06.032

Altdorff, D.; Galagedara, L.; Nadeem, M.; Cheema, M.; Unc, A. Effect of agronomic treatments on the accuracy of soil moisture mapping by electromagnetic induction. Catena, v.164, p.96-106, 2018. https://doi.org/10.1016/j.catena.2017.12.036

Amidan, B. G.; Ferryman, T. A.; Cooley, S. K. Data outlier detection using the Chebyshev theorem. IEEE Aerospace Conference Proceedings, v.2005, p.3-8, 2005. https://doi.org/10.1109/AERO.2005.1559688

Amirinejad, A. A.; Kamble, K.; Aggarwal, P.; Chakraborty, D.; Pradhan, S.; Mittal, R. B. Assessment and mapping of spatial variation of soil physical health in a farm. Geoderma, v.160, p.292-303, 2011. https://doi.org/10.1016/j.geoderma.2010.09.021

Anselin, L. Local Indicators of Spatial Association-LISA. Geographical Analysis, v.27, p.93-115, 1995. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Araujo, V.; Mitra, K.; Saguna, S.; Åhlund, C. Performance evaluation of FIWARE: A cloud-based IoT platform for smart cities. Journal of Parallel and Distributed Computing, v.132, p.250-261, 2019. https://doi.org/10.1016/j.jpdc.2018.12.010

Arslan, S.; Colvin, T. S. Grain yield mapping: yield sensing, yield reconstruction, and errors. Precision Agriculture, v.3, p.135-154, 2002. https://doi.org/10.1023/A:1013819502827

Barca, E.; De Benedetto, D.; Stellacci, A. M. Contribution of EMI and GPR proximal sensing data in soil water content assessment by using linear mixed effects models and geostatistical approaches. Geoderma, v.343, p.280-293, 2019. https://doi.org/10.1016/j.geoderma.2019.01.030

Barnett, V.; Lewis, T. Outliers in statistical data. John Wiley & Sons, v.3, p.304, 1994.

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

Beneduzzi, H. M., Souza, E. G. D., Moreira, W. K., Sobjak, R., Bazzi, C. L., Rodrigues, M. Fertilizer recommendation methods for precision agriculture - a systematic literature study. Engenharia Agricola, v.42, p.1-18, 2022. https://doi.org/10.1590/1809-4430-eng.agric.v42n1e20210185/2022

Betzek, N. M.; Souza, E. G. De; Bazzi, C. L.; Schenatto, K.; Gavioli, A.; Magalhães, P. S. G. Computational routines for the automatic selection of the best parameters used by interpolation methods to create thematic maps. Computers and Electronics in Agriculture, v.157, p.49-62, 2019. https://doi.org/10.1016/j.compag.2018.12.004

Bitencourt, D. G. B.; Barros, W. S.; Timm, L. C.; She, D.; Penning, L. H.; Parfitt, J. M. B.; Reichardt, K. Multivariate and geostatistical analyses to evaluate lowland soil levelling effects on physico-chemical properties. Soil and Tillage Research, v.156, p.63-73, 2016. https://doi.org/10.1016/j.still.2015.10.004

Blackmore, S.; Moore, M. Remedial Correction of Yield Map Data. Precision Agriculture, v.1, p.53-66, 1999. https://doi.org/10.1023/A:1009969601387

Bodner, G. S.; Robles, M. D. Enduring a decade of drought: Patterns and drivers of vegetation change in a semi-arid grassland. Journal of Arid Environments, v.136, p.1-14, 2017. https://doi.org/10.1016/j.jaridenv.2016.09.002

Bökle, S.; Karampoiki, M.; Paraforos, D. S.; Griepentrog, H. W. Using an open source and resilient technology framework to generate and execute prescription maps for site-specific manure application. Smart Agricultural Technology, v.5, 100272, 2023. https://doi.org/10.1016/j.atech.2023.100272

Calafat, C.; Gallego, A.; Quintanilla, I. Integrated geo-referenced data and statistical analysis for dividing livestock farms into geographical zones in the Valencian Community (Spain). Computers and Electronics in Agriculture, v.114, p.58-67, 2015. https://doi.org/10.1016/j.compag.2015.03.005

Camino, C.; González-Dugo, V.; Hernández, P.; Sillero, J. C.; Zarco-Tejada, P. J. Improved nitrogen retrievals with airborne-derived fluorescence and plant traits quantified from VNIR-SWIR hyperspectral imagery in the context of precision agriculture. International Journal of Applied Earth Observation and Geoinformation, v.70, p.105-117, 2018. https://doi.org/10.1016/j.jag.2018.04.013

Castaldi, F.; Palombo, A.; Santini, F.; Pascucci, S.; Pignatti, S.; Casa, R. Evaluation of the potential of the current and forthcoming multispectral and hyperspectral imagers to estimate soil texture and organic carbon. Remote Sensing of Environment, v.179, p.54-65, 2016. https://doi.org/10.1016/j.rse.2016.03.025

Castrignanò, A.; Wong, M. T. F. T. F.; Stelluti, M.; De Benedetto, D.; Sollitto, D. Use of EMI, gamma-ray emission and GPS height as multi-sensor data for soil characterisation. Geoderma, v.175, p.78-89, 2012. https://doi.org/10.1016/j.geoderma.2012.01.013

Castrignanò, A.; Buttafuoco, G.; Quarto, R.; Parisi, D.; Viscarra Rossel, R. A.; Terribile, F.; Langella, G.; Venezia, A. A geostatistical sensor data fusion approach for delineating homogeneous management zones in Precision Agriculture. Catena, v.167, p.293-304, 2018. https://doi.org/10.1016/j.catena.2018.05.011

Cavallo, G.; De Benedetto, D.; Castrignanò, A.; Quarto, R.; Vonella, A. V.; Buttafuoco, G. Use of geophysical data for assessing 3D soil variation in a durum wheat field and their association with crop yield. Biosystems Engineering, v.152, p.28-40, 2016. https://doi.org/10.1016/j.biosystemseng.2016.07.002

Chung, S. O.; Sudduth, K. A.; Motavalli, P. P.; Kitchen, N. R. Relating mobile sensor soil strength to penetrometer cone index. Soil and Tillage Research, v.129, p.9-18, 2013. https://doi.org/10.1016/j.still.2012.12.004

Cisternas, I.; Velásquez, I.; Caro, A.; Rodríguez, A. Systematic literature review of implementations of precision agriculture. Computers and Electronics in Agriculture, v.176, 105626, 2020. https://doi.org/10.1016/j.compag.2020.105626

Coelho, A. L. De F.; De Queiroz, D. M.; Valente, D. S. M.; Pinto, F. De A. De C. An open-source spatial analysis system for embedded systems. Computers and Electronics in Agriculture, v.154, p.289-295, 2018. https://doi.org/10.1016/j.compag.2018.09.019

Córdoba, M. A.; Bruno, C. I.; Costa, J. L.; Peralta, N. R.; Balzarini, M. G. Protocol for multivariate homogeneous zone delineation in precision agriculture. Biosystems Engineering, v.143, p.95-107, 2016. https://doi.org/10.1016/j.biosystemseng.2015.12.008

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

Dawson, R. How significant is a boxplot outlier? Journal of Statistics Education, v.19, p.1-13, 2011. https://doi.org/10.1080/10691898.2011.11889610

Demattê, J. A. M.; Ramirez-Lopez, L.; Marques, K. P. P.; Rodella, A. A. Chemometric soil analysis on the determination of specific bands for the detection of magnesium and potassium by spectroscopy. Geoderma, v.288, p.8-22, 2017. https://doi.org/10.1016/j.geoderma.2016.11.013

Dray, S.; Legendre, P.; Peres-Neto, P. R. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecological Modelling, v.196, p.483-493, 2006. https://doi.org/10.1016/j.ecolmodel.2006.02.015

Driemeier, C.; Ling, L. Y.; Sanches, G. M.; Pontes, A. O.; Magalhães, P. S. G. G.; Ferreira, J. E. A computational environment to support research in sugarcane agriculture. Computers and Electronics in Agriculture, v.130, p.13-19, 2016. https://doi.org/10.1016/j.compag.2016.10.002

Eitel, J. U. H.; Magney, T. S.; Vierling, L. A.; Dittmar, G. Assessment of crop foliar nitrogen using a novel dual-wavelength laser system and implications for conducting laser-based plant physiology. ISPRS Journal of Photogrammetry and Remote Sensing, v.97, p.229-240, 2014. https://doi.org/10.1016/j.isprsjprs.2014.09.009

Eugster, W.; Moffat, A. M.; Ceschia, E.; Aubinet, M.; Ammann, C.; Osborne, B., Davis, P. A.; Smith, P.; Jacobs, C.M.J.; Moors, E.J.; Dantec, V.; Béziat, P.; Saunders, M.; Jans, W. W. P.; Grünwald, T.; Rebmann, C.; Kutsch, W. L.; Czerný, R.; Janouš, Dalibor; Moureaux, C.; Buchmann, N. Management effects on European cropland respiration. Agriculture, Ecosystems & Environment v.139, p.346-362, 2010. https://doi.org/10.1016/j.agee.2010.09.001

Filzmoser, P.; Garrett, R. G.; Reimann, C. Multivariate outlier detection in exploration geochemistry. Computers and Geosciences, v.31, p.579-587, 2005. https://doi.org/10.1016/j.cageo.2004.11.013

Fisch, C.; Block, J. Six tips for your (systematic) literature review in business and management research. Management Review Quarterly, v.68, p.103-106, 2018.

Fujinuma, R.; Kirchhof, G.; Ramakrishna, A.; Sirabis, W.; Yapo, J.; Woruba, D.; Gurr, G.; Menzies, N. Intensified sweet potato production in Papua New Guinea drives plant nutrient decline over the last decade. Agriculture, Ecosystems & Environment, v.254, p.10-19, 2018. https://doi.org/10.1016/j.agee.2017.11.012

Gavioli, A.; De Souza, E. G.; Bazzi, C. L.; Schenatto, K.; Betzek, N. M. Identification of management zones in precision agriculture: An evaluation of alternative cluster analysis methods. Biosystems Engineering, v.181, p.86-102, 2019. https://doi.org/10.1016/j.biosystemseng.2019.02.019

Gerstmann, H.; Doktor, D.; Glässer, C.; Möller, M. Phase: A geostatistical model for the kriging-based spatial prediction of crop phenology using public phenological and climatological observations. Computers and Electronics in Agriculture, v.127, p.726-738, 2016. https://doi.org/10.1016/j.compag.2016.07.032

Gholizadeh, A.; Žižala, D.; Saberioon, M.; Borůvka, L. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. Remote Sensing of Environment, v.218, p.89-103, 2018. https://doi.org/10.1016/j.rse.2018.09.015

Gianquinto, G.; Orsini, F.; Fecondini, M.; Mezzetti, M.; Sambo, P.; Bona, S. A methodological approach for defining spectral indices for assessing tomato nitrogen status and yield. European Journal of Agronomy, v.35, p.135-143, 2011. https://doi.org/10.1016/j.eja.2011.05.005

Gili, A.; Álvarez, C.; Bagnato, R.; Noellemeyer, E. Comparison of three methods for delineating management zones for site-specific crop management. Computers and Electronics in Agriculture, v.139, p.213-223, 2017. https://doi.org/10.1016/j.compag.2017.05.022

Gnanadesikan, R. Methods for statistical analysis of multivariate observations. New York: Wiley, 1977.

González-Fernández, A. B.; Rodríguez-Pérez, J. R.; Sanz-Ablanedo, E.; Valenciano, J. B.; Marcelo, V. Delineating vineyard zones by fuzzy k-means algorithm based on grape sampling variables. Scientia Horticulturae, v.243, p.559-566, 2019. https://doi.org/10.1016/j.scienta.2018.09.012

Haghverdi, A.; Leib, B. G.; Washington-Allen, R. A.; Ayers, P. D.; Buschermohle, M. J. Perspectives on delineating management zones for variable rate irrigation. Computers and Electronics in Agriculture, v.117, p.154-167, 2015. https://doi.org/10.1016/j.compag.2015.06.019

Henriques, R.; Bacao, F.; Lobo, V. Exploratory geospatial data analysis using the GeoSOM suite. Computers, Environment and Urban Systems, v.36, p.218-232, 2012. https://doi.org/10.1016/j.compenvurbsys.2011.11.003

Hong, Y.; Liu, Y. Y. Y.; Chen, Y.; Liu, Y. Y. Y.; Yu, L.; Liu, Y. Y. Y.; Cheng, H. Application of fractional-order derivative in the quantitative estimation of soil organic matter content through visible and near-infrared spectroscopy. Geoderma, v.337, p.758-769, 2019. https://doi.org/10.1016/j.geoderma.2018.10.025

Hotelling, H. Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, v.24, p.417-441, 1933. http://dx.doi.org/10.1037/h0071325

ISPA - International Society of Precision Agriculture. Definition of Precision Agriculture, 2024. https://www.ispag.org/about/definition. Accessed on: Jan. 2024.

Johann, J. A.; Opazo, M. A. U.; Souza, E. G. De; Rocha, J. V. Variabilidade espacial dos atributos físicos do solo e da produtividade em um Latossolo Bruno distrófico da região de Cascavel, PR. Revista Brasileira de Engenharia Agrícola e Ambiental, v.8, p.212-219, 2004. https://doi.org/10.1590/S1415-43662004000200008

Jolliffe, I. Principal Component Analysis. Encyclopedia of Statistics in Behavioral Science, v.2, p.15-23, 2005. http://dx.doi.org/10.1037/h0071325

Jordanova, N.; Jordanova, D.; Petrov, P. Soil magnetic properties in Bulgaria at a national scale-Challenges and benefits. Global and Planetary Change, v.137, p.107-122, 2016. https://doi.org/10.1016/j.gloplacha.2015.12.015

Kanellopoulos, A.; Reidsma, P.; Wolf, J.; Van Ittersum, M. K. Assessing climate change and associated socio-economic scenarios for arable farming in the Netherlands: An application of benchmarking and bio-economic farm modelling. European Journal of Agronomy, v.52, p.69-80, 2014. https://doi.org/10.1016/j.eja.2013.10.003

Kaniu, M. I.; Angeyo, K. H. Challenges in rapid soil quality assessment and opportunities presented by multivariate chemometric energy dispersive X-ray fluorescence and scattering spectroscopy. Geoderma, v.241-242, p.32-40, 2015. https://doi.org/10.1016/j.geoderma.2014.10.014

Karunathilake, E. M. B. M.; Le, A. T.; Heo, S.; Chung, Y. S.; Mansoor, S. The path to smart farming: innovations and opportunities in precision agriculture. Agriculture, v.13, p.1-26, 2023. https://doi.org/10.3390/agriculture13081593

Kharel, T. P.; Swink, S. N.; Maresma, A.; Youngerman, C.; Kharel, D.; Czymmek, K. J.; Ketterings, Q. M. Yield monitor data cleaning is essential for accurate corn grain and silage yield determination. Agronomy Journal, v.111, p.509-516, 2019. https://doi.org/10.2134/agronj2018.05.0317

Knadel, M.; Thomsen, A.; Schelde, K.; Greve, M. H. Soil organic carbon and particle sizes mapping using vis-NIR, EC and temperature mobile sensor platform. Computers and Electronics in Agriculture, v.114, p.134-144, 2015. https://doi.org/10.1016/j.compag.2015.03.013

Krishna, G.; Sahoo, R. N.; Singh, P.; Bajpai, V.; Patra, H.; Kumar, S.; Dandapani, R.; Gupta, V. K.; Viswanathan, C.; Ahmad, T.; Sahoo, P. M. Comparison of various modelling approaches for water deficit stress monitoring in rice crop through hyperspectral remote sensing. Agricultural Water Management, v.213, p.231-244, 2019. https://doi.org/10.1016/j.agwat.2018.08.029

Lamsal, S.; Bliss, C. M.; Graetz, D. A. Geospatial mapping of soil nitrate-nitrogen distribution under a mixed-land use system. Pedosphere, v.19, p.434-445, 2009. https://doi.org/10.1016/S1002-0160(09)60136-3

Landrum, C.; Castrignanò, A.; Mueller, T.; Zourarakis, D.; Zhu, J.; De Benedetto, D. An approach for delineating homogeneous within-field zones using proximal sensing and multivariate geostatistics. Agricultural Water Management, v.147, p.144-153, 2015. https://doi.org/10.1016/j.agwat.2014.07.013

Leroux, C.; Jones, H.; Clenet, A.; Dreux, B.; Becu, M.; Tisseyre, B. A general method to filter out defective spatial observations from yield mapping datasets. Precision Agriculture, v.19, p.789-808, 2018. https://doi.org/10.1007/s11119-017-9555-0

Li, Q.; Luo, Y.; Wang, C.; Li, B.; Zhang, X.; Yuan, D.; Gao, X.; Zhang, H. Spatiotemporal variations and factors affecting soil nitrogen in the purple hilly area of Southwest China during the 1980s and the 2010s. Science of The Total Environment, v.547, p.173-181, 2016. https://doi.org/10.1016/j.scitotenv.2015.12.094

Li, T.; Zhan, Z. A systematic review on design thinking Integrated Learning in K-12 education. Applied Sciences, v.12, p.2-34, 2022. https://doi.org/10.3390/app12168077

Li, X.; Lee, W. S.; Li, M.; Ehsani, R.; Mishra, A. R.; Yang, C.; Mangan, R. L. Spectral difference analysis and airborne imaging classification for citrus greening infected trees. Computers and Electronics in Agriculture, v.83, p.32-46, 2012. https://doi.org/10.1016/j.compag.2012.01.010

Liang, W.; Kirk, K. R.; Greene, J. K. Estimation of soybean leaf area, edge, and defoliation using color image analysis. Computers and Electronics in Agriculture, v.150, p.41-51, 2018. https://doi.org/10.1016/j.compag.2018.03.021

Liu, X.; Guanter, L.; Liu, L.; Damm, A.; Malenovský, Z.; Rascher, U.; Peng, D.; Du, S.; Gastellu-Etchegorry, J.-P. Downscaling of solar-induced chlorophyll fluorescence from canopy level to photosystem level using a random forest model. Remote Sensing of Environment, v.231, 110772, 2019. https://doi.org/10.1016/j.rse.2018.05.035

Liu, X.; Zhang, W.; Zhang, M.; Ficklin, D. L.; Wang, F. Spatio-temporal variations of soil nutrients influenced by an altered land tenure system in China. Geoderma, v.152, p.23-34, 2009. https://doi.org/10.1016/j.geoderma.2009.05.022

Maldaner, L. F.; Molin, J. P. Data processing within rows for sugarcane yield mapping. Scientia Agricola, v.77, p.1-8, 2020. https://doi.org/10.1590/1678-992x-2018-0391

Mat, N. N.; Rowshon, K. M.; Guangnan, C.; Troy, J. Prediction of Sugarcane Quality Parameters Using Visible-shortwave Near Infrared Spectroradiometer. Agriculture and Agricultural Science Procedia, v.2, p.136-143, 2014. https://doi.org/10.1016/j.aaspro.2014.11.020

McFadden, J.; Njuki, E.; Griffin, T. Precision Agriculture in the Digital Era: Recent Adoption on U.S. Farms. US Department of Agriculture, Economic Research Service, 2023. 248p. Available at: <https://www.ers.usda.gov>. Accessed on: Jan. 2024.

Medina, H.; De Jong Van Lier, Q.; García, J.; Ruiz, M. E. Regional-scale variability of soil properties in Western Cuba. Soil and Tillage Research, v.166, p.84-99, 2017. https://doi.org/10.1016/j.still.2016.10.009

Menegatti, L. A. A.; Molin, J. P. Remoção de erros em mapas de produtividade via filtragem de dados brutos. Revista Brasileira de Engenharia Agrícola e Ambiental, v.8, p.126-134, 2004. https://doi.org/10.1590/s1415-43662004000100019

Mieza, M. S.; Cravero, W. R.; Kovac, F. D.; Bargiano, P. G. Delineation of site-specific management units for operational applications using the topographic position index in La Pampa, Argentina. Computers and Electronics in Agriculture, v.127, p.158-167, 2016. https://doi.org/10.1016/j.compag.2016.06.005

Milani, L.; Souza, E. G. De; Uribe-Opazo, M. A.; Gabriel Filho, A.; Johann, J. A.; Pereira, J. O. Unidades de manejo a partir de dados de produtividade. Acta Scientiarum. Agronomy, v.28, p.591-598, 2006. https://doi.org/10.4025/actasciagron.v28i4.937

Mirzaeitalarposhti, R.; Demyan, M. S.; Rasche, F.; Cadisch, G.; Müller, T. Mid-infrared spectroscopy to support regional-scale digital soil mapping on selected croplands of South-West Germany. Catena, v.149, p.283-293, 2017. https://doi.org/10.1016/j.catena.2016.10.001

Molin, J.P. Definição de unidades de manejo a partir de mapas de produtividade. Engenharia Agrícola, v.22, p.83-92, 2002.

Monteiro, A., Santos, S.; Gonçalves, P. Precision agriculture for crop and livestock farming-brief review. Animals, v.11, 2345, 2021. https://doi.org/10.3390/ani11082345

Moreira, W K. O.; Godoy, E.; Beneduzzi H. M.; Pereira, H.; Bazzi, C.; Maggi M. F.; Rodrigues, M.; Gavioli, A. Methods to recommend corrective measures for agricultural soils: a systematic literature study. Communications in Soil Science and Plant Analysis, v.54, p.1102-1133, 2022. https://doi.org/10.1080/00103624.2022.2137194

Morellos, A.; Pantazi, X. E.; Moshou, D.; Alexandridis, T.; Whetton, R.; Tziotzios, G.; Wiebensohn, J.; Bill, R.; Mouazen, A. M. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. Biosystems Engineering, v.152, p.104-116, 2016. https://doi.org/10.1016/j.biosystemseng.2016.04.018

Moura-Bueno, J. M.; Dalmolin, R. S. D.; Ten Caten, A.; Dotto, A. C.; Demattê, J. A. M. M. Stratification of a local VIS-NIR-SWIR spectral library by homogeneity criteria yields more accurate soil organic carbon predictions. Geoderma, v.337, p.565-581, 2019. https://doi.org/10.1016/j.geoderma.2018.10.015

Mura, S.; Cappai, C.; Greppi, G. F.; Barzaghi, S.; Stellari, A.; Cattaneo, T. M. P. Vibrational spectroscopy and Aquaphotomics holistic approach to determine chemical compounds related to sustainability in soil profiles. Computers and Electronics in Agriculture, v.159, p.92-96, 2019. https://doi.org/10.1016/j.compag.2019.03.002

Neave, H. W.; Lomb, J.; Weary, D. M.; Leblanc, S. J.; Huzzey, J. M.; Von Keyserlingk, M. A. G. Behavioral changes before metritis diagnosis in dairy cows. Journal of Dairy Science, v.101, p.4388-4399, 2018. https://doi.org/10.3168/jds.2017-13078

Nouri, M.; Gomez, C.; Gorretta, N.; Roger, J. M. Clay content mapping from airborne hyperspectral Vis-NIR data by transferring a laboratory regression model. Geoderma, v.298, p.54-66, 2017. https://doi.org/10.1016/j.geoderma.2017.03.011

Oliver, M. A.; Webster, R. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. Catena, v.113, p.56-69, 2014. https://doi.org/10.1016/j.catena.2013.09.006

Paccioretti, P.; Córdoba, M.; Balzarini, M. FastMapping: Software to create field maps and identify management zones in precision agriculture. Computers and Electronics in Agriculture, v.175, 105556, 2020. https://doi.org/10.1016/j.compag.2020.105556

Paraforos, D. S.; Hübner, R.; Griepentrog, H. W. Automatic determination of headland turning from auto-steering position data for minimising the infield non-working time. Computers and Electronics in Agriculture, v.152, p.393-400, 2018. https://doi.org/10.1016/j.compag.2018.07.035

Paraforos, D. S.; Reutemann, M.; Sharipov, G.; Werner, R.; Griepentrog, H. W. Total station data assessment using an industrial robotic arm for dynamic 3D in-field positioning with sub-centimetre accuracy. Computers and Electronics in Agriculture, v.136, p.166-175, 2017. https://doi.org/10.1016/j.compag.2017.03.009

Páscoa, R. N. M. J.; Lopo, M.; T. Santos, C. A.; Graça, A. R.; Lopes, J. A. Exploratory study on vineyards soil mapping by visible/near-infrared spectroscopy of grapevine leaves. Computers and Electronics in Agriculture, v.127, p.15-25, 2016. https://doi.org/10.1016/j.compag.2016.05.014

Pei, T.; Qin, C.-Z.; Zhu, A.-X.; Yang, L.; Luo, M.; Li, B.; Zhou, C. Mapping soil organic matter using the topographic wetness index: A comparative study based on different flow-direction algorithms and kriging methods. Ecological Indicators, v.10, p.610-619, 2010. https://doi.org/10.1016/j.ecolind.2009.10.005

Pelosi, A.; Medina, H.; Villani, P.; D'urso, G.; Chirico, G. B. Probabilistic forecasting of reference evapotranspiration with a limited area ensemble prediction system. Agricultural Water Management, v.178, p.106-118, 2016. https://doi.org/10.1016/j.agwat.2016.09.015

Ping, J. L.; Dobermann, A. Processing of yield map data. Precision Agriculture, v.6, p.193-212, 2005. https://doi.org/10.1007/s11119-005-1035-2

Pires, J. L. F.; Cunha, G. R. da; Pasinato, A.; França, S.; Rambo, L. Discutindo agricultura de precisão: aspectos gerais. Passo Fundo: EMBRAPA Trigo, 2004. Available at: http://www.agencia.cnptia.embrapa.br/Repositorio/p_do42_000fkl0aoyi02wyiv80sq98yqj0l5xns.pdf. Accessed on: Oct. 2018.

Prasad, R.; Deo, R. C.; Li, Y.; Maraseni, T. Weekly soil moisture forecasting with multivariate sequential, ensemble empirical mode decomposition and Boruta-random forest hybridizer algorithm approach. Catena, v.177, p.149-166, 2019. https://doi.org/10.1016/j.catena.2019.02.012

Raj, A.; Chakraborty, S.; Duda, B. M.; Weindorf, D. C.; Li, B.; Roy, S.; Sarathjith, M. C.; Das, B. S.; Paulette, L. Soil mapping via diffuse reflectance spectroscopy based on variable indicators: An ordered predictor selection approach. Geoderma, v.314, p.146-159, 2018. https://doi.org/10.1016/j.geoderma.2017.10.043

Reunanen, N.; Räty, T.; Jokinen, J. J.; Hoyt, T.; Culler, D. Unsupervised online detection and prediction of outliers in streams of sensor data. International Journal of Data Science and Analytics, v.9, p.285-314, 2020. https://link.springer.com/article/10.1007/s41060-019-00191-3

Robert, P.C. Precision agriculture: a challenge for crop nutrition management. Plant and Soil, v.247, p.143-149, 2002. https://doi.org/10.1023/A:1021171514148

Rodriguez-Moreno, F.; Zemek, F.; Kren, J.; Pikl, M.; Lukas, V.; Novak, J. Spectral monitoring of wheat canopy under uncontrolled conditions for decision-making purposes. Computers and Electronics in Agriculture, v.125, p.81-88, 2016. https://doi.org/10.1016/j.compag.2016.05.002

Rosemary, F.; Vitharana, U. W. A.; Indraratne, S. P.; Weerasooriya, R.; Mishra, U. Exploring the spatial variability of soil properties in an Alfisol soil. Catena, v.150, p.53-61, 2017. https://doi.org/10.1016/j.catena.2016.10.017

Sanches, G. M.; Graziano Magalhães, P. S.; Junqueira Franco, H. C. Site-specific assessment of spatial and temporal variability of sugarcane yield related to soil attributes. Geoderma, v.334, p.90-98, 2019. https://doi.org/10.1016/j.geoderma.2018.07.051

Schönhart, M.; Trautvetter, H.; Parajka, J.; Blaschke, A. P.; Hepp, G.; Kirchner, M.; Mitter, H.; Schmid, E.; Strenn, B.; Zessner, M. Modelled impacts of policies and climate change on land use and water quality in Austria. Land Use Policy, v.76, p.500-514, 2018. https://doi.org/10.1016/j.landusepol.2018.02.031

Shaddad, S. M.; Buttafuoco, G.; Elrys, A.; Castrignanò, A. Site-specific management of salt-affected soils: A case study from Egypt. Science of the Total Environment, v.688, p.153-161, 2019. https://doi.org/10.1016/j.scitotenv.2019.06.214

Simbahan, G. C.; Dobermann, A.; Ping, J. L. Screening yield monitor data improves grain yield maps. Agronomy Journal, v.96, p.1091-1102, 2004. https://doi.org/10.2134/agronj2004.1091

Sirsat, M. S.; Cernadas, E.; Fernández-Delgado, M.; Barro, S. Automatic prediction of village-wise soil fertility for several nutrients in India using a wide range of regression methods. Computers and Electronics in Agriculture, v.154, p.120-133, 2018. https://doi.org/10.1016/j.compag.2018.08.003

Spekken, M. A. R. K.; Anselmi, A. A.; Molin, J. P. A simple method for filtering spatial data. In: Precision agriculture'13. Wageningen Academic Publishers, 2013. p.259-266.

Squalli, J.; Adamkiewicz, G. Organic farming and greenhouse gas emissions: A longitudinal U.S. state-level study. Journal of Cleaner Production, v.192, p.30-42, 2018. https://doi.org/10.1016/j.jclepro.2018.04.160

Stockmann, U.; Malone, B. P.; McBratney, A. B.; Minasny, B. Landscape-scale exploratory radiometric mapping using proximal soil sensing. Geoderma, v.239-240, p.115-129, 2015. https://doi.org/10.1016/j.geoderma.2014.10.005

Sudduth, K. A.; Drummond, S. T. Yield Editor: Software for Removing Errors from Crop Yield Maps. Agronomy Journal, v.99, p.1471-1482, 2007. https://doi.org/10.2134/agronj2006.0326

Sudduth, K. A.; Drummond, S. T.; Myers, D. B. Yield Editor 2.0: Software for Automated Removal of Yield Map Errors. 2012 ASABE Annual International Meeting, v.7004, 14, 2012. https://doi.org/10.13031/2013.41893

Sun, W.; Whelan, B.; Mcbratney, A. B.; Minasny, B. An integrated framework for software to provide yield data cleaning and estimation of an opportunity index for site-specific crop management. Precision Agriculture, v.14, p.376-391, 2013. https://doi.org/10.1007/s11119-012-9300-7

Talavera, J. M.; Tobón, L. E.; Gómez, J. A.; Culman, M. A.; Aranda, J. M.; Parra, D. T.; Quiroz, L. A.; Hoyos, A.; Garreta, L. E. Review of IoT applications in agro-industrial and environmental fields. Computers and Electronics in Agriculture, v.142, p.283-297, 2017. https://doi.org/10.1016/j.compag.2017.09.015

Taylor, J. A. A.; McBratney, A. B. B.; Whelan, B. M. M. Establishing management classes for broadacre agricultural production. Agronomy Journal, v.99, p.1366-1376, 2007. https://doi.org/10.2134/agronj2007.0070

Tesfahunegn, G. B.; Tamene, L.; Vlek, P. L. G. Catchment-scale spatial variability of soil properties and implications on site-specific soil management in northern Ethiopia. Soil and Tillage Research, v.117, p.124-139, 2011. https://doi.org/10.1016/j.still.2011.09.005

Trevisan, R. G.; Bullock, D. S.; Martin, N. F. Spatial variability of crop responses to agronomic inputs in on-farm precision experimentation. Precision Agriculture, v.22, p.342-363, 2021. https://doi.org/10.1007/s11119-020-09720-8

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

Tukey, J. W. Exploratory Data Analysis. Addison-Wesley, 1977.

Turner, P. A.; Griffis, T. J.; Mulla, D. J.; Baker, J. M.; Venterea, R. T. A geostatistical approach to identify and mitigate agricultural nitrous oxide emission hotspots. Science of the Total Environment, v.572, p.442-449, 2016. https://doi.org/10.1016/j.scitotenv.2016.08.094

Uribeetxebarria, A.; Daniele, E.; Escolà, A.; Arnó, J.; Martínez-Casasnovas, J. A. Spatial variability in orchards after land transformation: Consequences for precision agriculture practices. Science of the Total Environment, v.635, p.343-352, 2018. https://doi.org/10.1016/j.scitotenv.2018.04.153

Uribeetxebarria, A.; Martínez-Casasnovas, J. A.; Tisseyre, B.; Guillaume, S.; Escolà, A.; Rosell-Polo, J. R.; Arnó, J. Assessing ranked set sampling and ancillary data to improve fruit load estimates in peach orchards. Computers and Electronics in Agriculture, v.164, 104931, 2019. https://doi.org/10.1016/j.compag.2019.104931

Van Meirvenne, M.; Islam, M. M.; De Smedt, P.; Meerschman, E.; Van De Vijver, E.; Saey, T. Key variables for the identification of soil management classes in the aeolian landscapes of northwest Europe. Geoderma, v.199, p.99-105, 2013. https://doi.org/10.1016/j.geoderma.2012.07.017

Vašát, R.; Heuvelink, G. B. M.; Borůvka, L. Sampling design optimization for multivariate soil mapping. Geoderma, v.155, p.147-153, 2010. https://doi.org/10.1016/j.geoderma.2009.07.005

Vega, A.; Córdoba, M.; Castro-Franco, M.; Balzarini, M. Protocol for automating error removal from yield maps. Precision Agriculture, v.20, p.1030-1044, 2019. https://doi.org/10.1007/s11119-018-09632-8

Wang, N.; Jassogne, L.; Van Asten, P. J. A.; Mukasa, D.; Wanyama, I.; Kagezi, G.; Giller, K. E. Evaluating coffee yield gaps and important biotic, abiotic, and management factors limiting coffee production in Uganda. European Journal of Agronomy, v.63, p.1-11, 2015. https://doi.org/10.1016/j.eja.2014.11.003

Waruru, B. K.; Shepherd, K. D.; Ndegwa, G. M.; Sila, A.; Kamoni, P. T. Application of mid-infrared spectroscopy for rapid characterization of key soil properties for engineering land use. Soils and Foundations, v.55, p.1181-1195, 2015. https://doi.org/10.1016/j.sandf.2015.09.018

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

Zhou, J.; Fu, X.; Zhou, S.; Zhou, J.; Ye, H.; Nguyen, H. T. Automated segmentation of soybean plants from 3D point cloud using machine learning. Computers and Electronics in Agriculture, v.162, p.143-153, 2019. https://doi.org/10.1016/j.compag.2019.04.014