



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

Ricardo Sobjak¹ · Eduardo Godoy de Souza² · Claudio Leones Bazzi¹ · Miguel Angel Uribe Opazo² · Erivelto Mercante² · Jorge Aikes Junior¹

Accepted: 19 February 2023 / Published online: 17 March 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Thematic maps are essential tools in precision agriculture to demonstrate the information of spatially distributed phenomena. A thematic map can be created from sampling data, a standard procedure for soil attributes. Interpolation methods are used to estimate data in unknown locations, such as inverse distance weighting (IDW) and ordinary Kriging (OK). For both interpolators, it is essential to use the appropriate parameters to estimate values in non-sampled locations, either the exponent value and the number of neighbors for IDW, or the theoretical model adjusted to the experimental semivariogram for OK. Thus, this trial aims at adopting additional criteria in selecting interpolators and evaluating their performance. AgDataBox platform's data interpolation module was improved, where the process of selecting the interpolator and determining its parameters considers the criteria (i) effective spatial dependence index, (ii) the first semivariance significance index, and (iii) slope of the model ends index. The experimental data come from an experiment in two agricultural areas in Brazil, using grids with good sampling density (2.7, 2.6, and 3.5 points per ha). It was observed that, usually, the application method of the three new criteria selected different models for a dataset and this must be considered in the interpolator selection process. Thematic maps varied from 0.1 to 64%, according to the coefficient of relative deviation, when comparing the three methods of applying the selection criteria.

Keywords Ordinary Kriging · Inverse distance weighting · Precision agriculture · Spatial dependence · AgDataBox

✉ Ricardo Sobjak
ricardosobjak@utfpr.edu.br

¹ Computer Science Department, Federal University of Technology of Paraná (UTFPR), Medianeira, Paraná, Brazil

² Western Paraná State University (UNIOESTE), PGEAGRI, Technological and Exact Sciences Center, Cascavel, Paraná, Brazil

Introduction

Maps representing a field and a topic associated with it are called thematic maps (TMs) and aim to inform, by graphical symbols, where a specific geographical phenomenon occurs. TMs have become an essential tool in geospatial science to understand spatial information (Fraser & Congalton, 2019), e.g., digital elevation model, slope map, soil map, aspect map, land use/land cover map, and contour map (Gojiya et al., 2018).

In precision agriculture, TM is an essential tool to assist analysts in decision-making, as it allows them to identify spatial variability within the field and manage the area in a localized way. TMs development is associated with data collection, analysis, interpretation, and information representation on a map, facilitating identifying similarities and enabling spatial correlations visualization. One specific case of TMs is contour maps built by connecting points of the same value and applying them to geographical phenomena that show continuity in geographic space. Another is choropleth maps, which use color to show ranges of values of a specific variable within a defined geographic area. Contour and choropleth maps can be built from categorical data (yield, elevation, temperature, precipitation, humidity, and atmospheric pressure) or relative data (density, percentages, and indexes) (Aikes Junior et al., 2021). Usually, both maps are called contour maps.

The advancement of computational technologies allows the creation and analysis of TMs using different techniques, methodologies, and software. For example, geographical information systems (GISs) can store, exhibit, recover, and dissect spatial data in a friendly approach. GIS has been widely used in many studies for spatial and temporal data creation (Gojiya et al., 2018).

Usually, the sampled data are interpolated in a dense and regular grid to generate continuous and smooth TMs. This task is carried out with the aid of interpolation methods. The most used methods in precision agriculture are inverse distance weighted interpolation (IDW—Shepard, 1968) and ordinary Kriging (OK—Cressie, 1993), which are differentiated by how weights are attributed to different samples, and may influence the estimated values (Reza et al., 2010). IDW procedure has been used because it is quick and straightforward; Kriging has been used because it provides the best linear unbiased estimates. However, it is more complex and time-consuming (Mueller et al., 2004). IDW interpolator considers weights at the sample points, which are evaluated during the interpolation process. Each sampled point's influence is inversely proportional to the distance increased to a power from the point to be estimated (Isaaks & Srivastava, 1989). The value of the chosen power predetermines the weight factor; that is, the higher this value, the lower the most distant points' influence.

Kriging has been identified as a Best Linear Unbiased Estimator (BLUE) interpolator (Diggle & Ribeiro, 2007; Isaaks and Srivastava, 1989). However, it must meet the spatial dependence (SD) modeling requests (Oliver & Webster, 2015; Cambardella et al., 1994) to have the correct performance and adequate use in creating a TM. The procedure's performance can be influenced by variability and spatial structure of data, semivariogram model, search radius, and the used number of the closest neighboring points (Reza et al., 2010; Isaaks & Srivastava, 1989). Therefore, the interpolations' quality depends on the variable's spatial structure under study (Amaral & Justina, 2019). The deterministic interpolator IDW does not consider SD and the specific behavior of data, leading to less efficiency in mapping the spatial distribution of a given variable than Kriging (stochastic interpolator) (Betzek et al., 2019). However, when there is no SD (Rodrigues et al., 2018; Cambardella et al., 1994), the use of a deterministic interpolator can be more appropriate.

In geostatistics, semivariograms are not only used as an exploratory tool but allow estimating parameters (Diggle & Ribeiro, 2007). After the experimental semivariogram construction, it is necessary to adjust a theoretical model representing data variability. The curve-fitting can be done “by eye” by trying different values for the model parameters and visually inspecting the fit to the sample variogram (Diggle & Ribeiro, 2007). However, parametric covariance functions can be used to estimate semivariogram parameters. As a result, the variogram parameter estimates minimize the theoretical model’s squared differences and experimental variogram (Li et al., 2018).

Betzek et al. (2019) developed computational routines to determine the best interpolator and its parameters for a data set. The routines determine the best semivariogram model (and its parameters) for OK and the best power and number of neighbors used in the IDW interpolator. The interpolation selection index (Bier & Souza, 2017) enables the selection of the best among several existing mathematical and geostatistical models in a simplified and less subjective manner. It was observed that, in some data sets, the routine implemented to select an interpolator, may mistakenly select a geostatistical model that does not have spatial dependence or consider a model with a lack of adjustment to the experimental semivariogram.

Therefore, this work aims to adopt criteria to guarantee a minimum spatial dependence in the semivariograms applied to the interpolators’ selection process. For that, the indices were proposed (i) the effective spatial dependence index (%ESDI), (ii) the first semivariance significance index (% $\gamma(1)$), and (iii) the slope of the model ends index (%SMEI).

Materials and methods

AgDataBox (ADB, <http://adb.md.utfpr.edu.br>; Michelon et al., 2019, Borges et al., 2020, Dall’agnol et al., 2020) web platform provides tools to create, store, recover, manage, exhibit, and analyze geographic and spatial data of TMs focused on precision agriculture. ADB offers farmers, researchers, and service providers focused on precision agriculture the ability to integrate data, software, procedures, and methodologies to contribute to agriculture development in the country using free technologies. This web platform has a microservices architecture (MSA), called ADB-MSA, which consists of a set of resources accessible remotely, through the hypertext transfer protocol (HTTP), to process and store data from an agricultural environment. ADB-MSA allows interoperability of several applications in which data and processing routines are centralized. The following applications, under development, consume ADB-MSA resources: (1) ADB-Mobile; (2) ADB-Map; (3) ADB-Admin; (4) ADB-IoT; (5) ADB-Remote Sensing.

ADB-Map application is included in ADB web platform and was employed for: (i) descriptive and exploratory analyses, (ii) data interpolation, (iii) selection of the best interpolation method, and (iv) TMs creation. This application aims at mitigating the problem of using different software to create TMs and delineate management zones. In addition, ADB-Map application provides user-friendly interfaces and procedures. This proposal converges to digitize agriculture. The functionalities of ADB-Map application are divided into conceptual modules (Fig. 1).

ADB’s data interpolation module interpolates data by IDW, OK, moving average, and nearest neighbor. Furthermore, it is possible to select the best interpolation method between OK and IDW, in addition to determining its interpolation parameters (Fig. 2). We improved and implemented new features in the module studied and implemented by Betzek

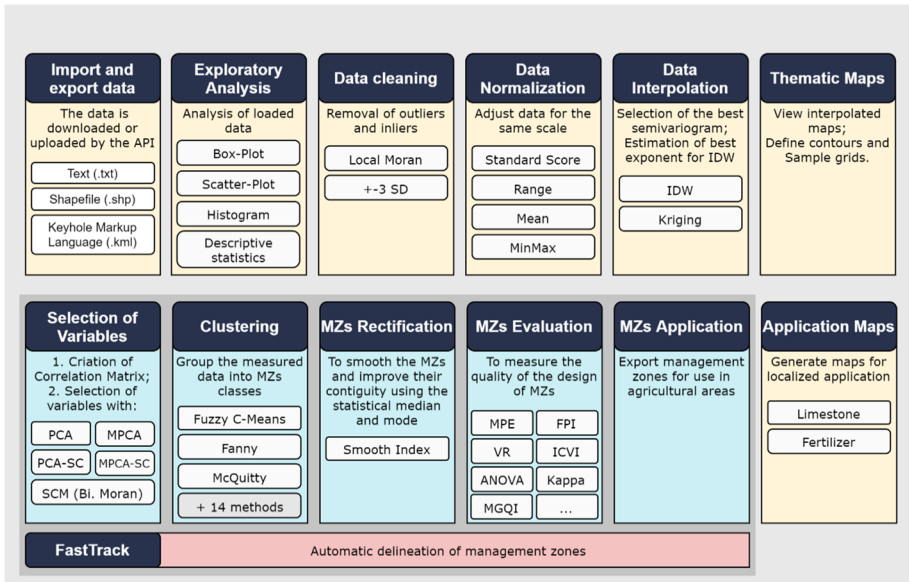


Fig. 1 Overview of modules that make up AgDataBox-Map application

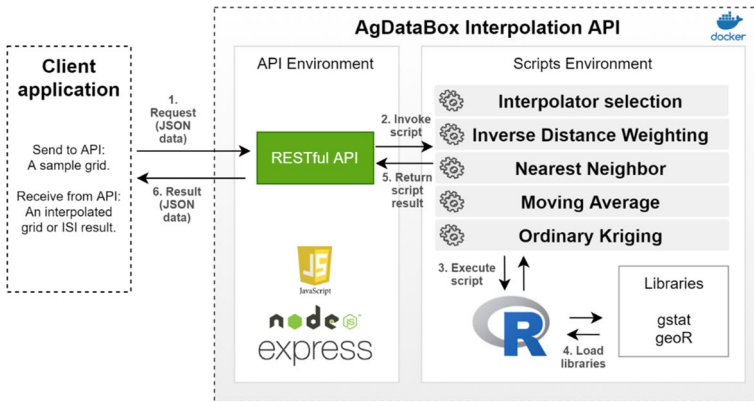


Fig. 2 Architecture of the ADB data interpolation module, representing the components and workflow

et al. (2019). We developed algorithms that make interpolations with R software, using the packages geoR (Ribeiro & Diggle, 2001) and gstat.

Location of the field, data collection, and selection of the coordinate system

Physical and chemical soil attributes were collected based on irregular sampling grids in two agricultural fields located in the municipality of Serranópolis do Iguaçú, western Paraná state, southern Brazil (Fig. 3). The fields have been cultivated under a no-tillage system with a crop succession of soybean and corn. The coordinate systems were the geographic

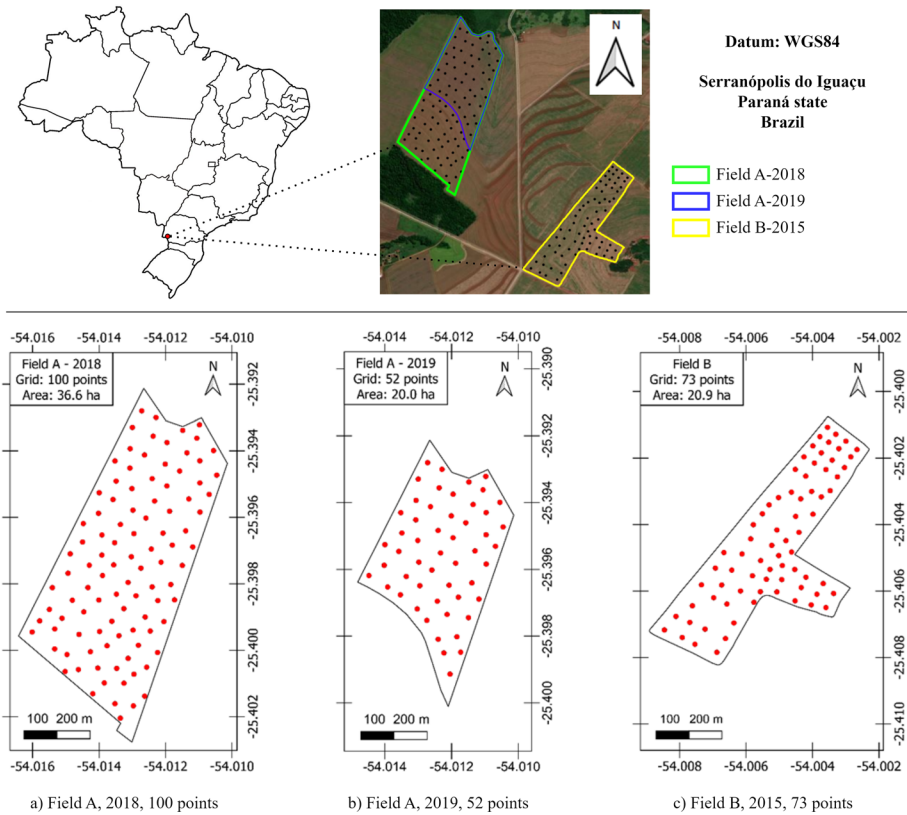


Fig. 3 Location of experimental fields and sampling grids of **a** 100 points in field A-2018; 36.6 ha, **b** 52 points in field A-2019; 20.0 ha, and **c** 73 points in field B-2015; 20.9 ha in the municipality of Serranópolis do Iguaçu, Paraná state, Southern Brazil. Black contour delineates the fields used. Coordinates are in degrees (WGS 1984). The minimum and maximum distances among the sampling points are 41 and 1027 m in field A-2018, 45 and 706 m in field A-2019, and 31 and 838 m in field B-2015

coordinate system (GCS) with WGS 1984 datum. The sampling points' locations were obtained by a GNSS receiver (Juno SB Trimble Navigation Limited, Westminster, CO, USA).

Soil samples were taken from 0 to 0.20 m depth and analyzed in a commercial laboratory. Around each sampling point (using a GNSS Juno SB Trimble Navigation Limited, Westminster, CO, USA) and using a 3-m radius, eight subsamples were randomly collected, two per quadrant, within a symmetrical circle divided into four quadrants. Field A (Fig. 3a, b) was sampled with 100 sampling points in 2018 (36.6 ha) and 52 in 2019 (20.0 ha) and field B (Fig. 3c) was sampled with 73 sampling points (20.9 ha). The minimum and maximum distances among the sampling points are 41 and 1027 m in field A-2018, 45 and 706 m in field A-2019, and 31 and 838 m in field B-2015. Thus, the sampled density corresponds, respectively, to 2.7, 2.6, and 3.5 points ha^{-1} (Table 1), which were considered enough to identify spatial variabilities of the variables of these fields given that they exceed the recommended minimum density of 1 sample ha^{-1} (Ferguson & Hergert, 2009) to 2.5 samples ha^{-1} (Doerge, 2000; Journal & Huijbregts, 1978). However, Oliver and Webster (2015) observed that at least between

Table 1 Details of the study fields

Fields	Areas (ha)	Geographical center coordinates (WGS84)	Elevation (m)	Sample points	Points (ha ⁻¹)
A-2018	36.6	25° 23' 48" S 54° 0' 46" W	345	100	2.7
A-2019	20.0	25° 23' 43" S 54° 0' 44" W	334	52	2.6
B	20.9	25° 24' 28" S 54° 00' 17" W	355	73	3.5

100 and 150 samples are required for a reliable variogram, but Clark (1979) recommended at least 30–50 data points to use Kriging. Nevertheless, the threshold for a sufficient density in one case may not be enough in another. We used different sample densities meeting at least each of the recommendations, 100–150 samples in field A-2018 and 30–50 samples in fields A-2019 and B-2015, to confirm the robustness of ADB's automated procedure and determine whether it can help be employed to determine when to use IDW and when to use OK (i.e., to determine whether the sample density is enough and/or if SD is detected; the pure nugget effect characterizes this case).

Each point sample was composed of eight individual samples (Wollenhaupt et al., 1994). The sampling points were located along an imaginary line among intermediate contour lines with alternated distances and provided a better fit at the smallest lag distances, which is essential in Kriging (Bier & Souza, 2017). The variables obtained from soil analysis were chemical attributes (organic matter (OM; g dm⁻³), zinc (Zn; mg dm⁻³), iron (Fe; mg dm⁻³), manganese (Mn; mg dm⁻³), phosphorus (P; mg dm⁻³), potassium (K; cmol_c dm⁻³), copper (Cu; mg dm⁻³), the potential of hydrogen (pH), calcium (Ca; cmol_c dm⁻³), magnesium (Mg; cmol_c dm⁻³), aluminum (Al; cmol_c dm⁻³), pH of buffer solution Shoemaker–McLean–Pratt (SMP) method, potential acidity (H+Al; cmol_c dm⁻³), the sum of bases (SB; cmol_c dm⁻³), base saturation (V%), aluminum saturation (m%), and physical attributes (clay (%), sand (%), and silt (%)).

Exploratory data analysis

Data were analyzed using descriptive and exploratory statistics and geostatistics. During the descriptive analysis of data, measures of central tendency (mean and median), dispersion [standard deviation (SD) and coefficient of variation (CV)], and normality tests (Kolmogorov–Smirnov and Anderson–Darling tests at 0.05 significance level) were calculated. Data were considered normal when, in at least one of the tests, they presented normality. The coefficient of variation (CV) was classified as low when $CV \leq 10\%$, medium when $10\% < CV \leq 20\%$, high when $20\% < CV \leq 30\%$, and very high when $CV > 30\%$ (Pimentel-Gomes, 2009). The exploratory data analysis (EDA) was used to detect and remove outliers and inliers. Using the module ADB-Map-Clean of platform ADB, duplicate, negative or null points, outliers, and inliers were removed. The outliers were identified as values outside the mean ± 3 SD (Córdoba et al., 2016). The inliers were obtained by Moran's local spatial autocorrelation index (II) (Anselin, 1995).

Analysis of spatial dependence

The semivariogram chart is determined from a set of observed values according to Oliver and Webster (2015) in two stages: (i) the calculation of the empirical semivariogram that summarizes spatial relations in data, and (ii) the adjustment of a mathematical model that best represents semivariances' distribution in each lag distance. Each calculated semivariance for a particular lag (h) is only an estimate of a mean semivariance $\hat{\gamma}(h)$ for that lag. The four main elements are (i) the nugget effect (C_0), (ii) the partial sill (C_1), (iii) the sill ($C_0 + C_1$), and (iv) the range of spatial autocorrelation (R_a).

The Matheron (1963) classic estimator was used to calculate semivariances with at least 30 pairs of points (Journel & Huijbregts, 1978), and the range R_a was limited to half of the maximum distance (MD) among points (cutoff = $0.5 * MD$). The semivariances' calculation should not exceed distances among points greater than half of the maximum distance (Clark, 1979). Points located beyond the cutoff are considered non-influential (Isaaks & Srivastava, 1989). Lag size (h) was defined by calculating the number of lags, the relationship between the cutoff, and the shortest distance among the pairs of points. Therefore, the lag h sizes were 43 m (field A-2018), 44 m (field A-2019), and 30 m (field B-2015), while semivariances 102 and 438 in area A-2018, 53 and 180 in area A-2019, and 55 and 182 in area B-2015. A significant limitation to address in this ADB-Map version is that anisotropy's eventual presence is not considered.

The mathematical model adjustment should describe the spatial variation to estimate or predict values at unsampled places optimally by Kriging (Oliver & Webster, 2015). Only certain mathematical functions are suitable for this purpose, so, choosing and fitting a model must be done with care (Lark, 2000). We selected the most commonly used theoretical models: spherical, exponential, gaussian, and Matérn's family (Uribe-Opazo et al., 2012; Isaaks & Srivastava, 1989).

To evaluate the degree of the SD variable, we used the spatial dependence index (%SDI—Biondi et al., 1994). The %SDI classification (Konopatzki et al., 2012) was adopted: very low for %SDI < 20%; low for $20 \leq \%SDI < 40\%$; medium for $40 \leq \%SDI < 60\%$; high for $60 \leq \%SDI < 80\%$; and very high for %SDI > 80%. This classification has the advantage of having five interpretation levels instead of three as proposed by Cambardela et al. (1994). The classification proposed by Konopatzki et al. (2012) is proportional to the spatial variability (the higher %SDI, the higher SD).

Figure 4 shows hypothetical sample points for which the spherical model was adjusted by routine in R. Considering that C_0 is 1 and C_1 is 9, the associated %SDI is 90%, corresponding to a strong SD. However, all semivariances are in the interval from 7 to 10. In this context, this works presents a new index, the effective spatial dependence index (%ESDI—Eq. 1), a new measure of SD degree. This index considers semivariance ($\gamma(1)$) in the first lag distance ($h(1)$).

$$\%ESDI = \frac{C - \gamma(1)}{C} * 100, \quad (1)$$

where C is the sill (nugget effect + partial sill) and $\gamma(1)$ is the first semivariance of the semivariogram. The %ESDI was classified as %SDI.

The second proposed index was the first semivariance significance index ($\% \gamma(1)$ —Eq. 2), SD fraction due only to ($\% \gamma(1)$).

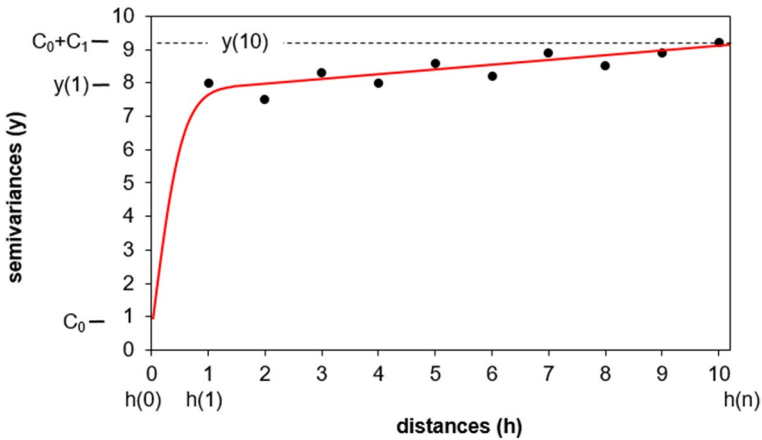


Fig. 4 Example of semivariogram chart adjusted with spherical semivariogram model, where $\gamma(1)$ is the first semivariance, γ_Z is the adjusted theoretic semivariance, $\gamma_Z(0) = C_0$ is the nugget effect, and $\gamma_Z(n)$ is the last adjusted theoretic semivariance

$$\% \gamma(1) = \frac{\gamma(1) - C_0}{C_1} * 100, \tag{2}$$

where C_0 is the nugget effect, C_1 is the partial sill, and $\gamma(1)$ is the first semivariance of the semivariogram.

Furthermore, we also propose a slope of the model ends index (%SMEI—Eq. 3), which aims to assess the inclination degree between the nugget effect and the last adjusted semivariance. When %SMEI is null, it is a pure nugget effect, characterizing a lack of SD.

$$\%SMEI = \left(1 - \frac{\gamma_Z(0)}{10^{-10} + \gamma_Z(n)} \right) * 100 = \left(1 - \frac{C_0}{10^{-10} + \gamma_Z(n)} \right) * 100, \tag{3}$$

where γ_Z is the adjusted theoretic semivariance, $\gamma_Z(0) = C_0$ is the nugget effect, and $\gamma_Z(n)$ is the last adjusted theoretic semivariance, correspondent to the cutoff. The arbitrary constant 10^{-10} was included to avoid division by zero.

Data interpolation

The variables used to generate TM were interpolated using OK and IDW in a 9×9 m grid with pixels. ADB-Map application automatically sets the pixel size based on the area’s size, with the value of 1 hundredth of the longest distance (horizontal or vertical). Computational routines were implemented in R language in ADB-Map application (Betzek et al., 2019).

Inverse distance weighting

IDW (Shepard, 1968) deterministic estimator considers the closest points to the location to be estimated more representative than the most distant one according to the samples’ linear

distances. Twelve different values were used as IDW exponents (p) (0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, and 6.0).

Ordinary Kriging

Variables' semivariograms were adjusted using theoretical models (spherical, gaussian, exponential, Matérn 0.5, Matérn 1.0, Matérn 1.5, and Matérn 2.0) by OLS and WLS methods. WLS weights were considered using the same number of pairs in each bin. Twenty-five different parameter sets (five initial values for the partial sill parameter and five for range) were used for each model, totalizing 350 adjustments.

Determination of the best semivariogram model and its parameters

Bier and Souza (2017) proposed the interpolation selection index (ISI) to automatize the selection of the best interpolation method, which assumes a lower value as better the interpolator is. By cross-validation (Faraco et al., 2008; Isaaks & Srivastava, 1989), mean error (ME) and standard deviation of mean error (SDME) are calculated. ME and SDME values calculated for each parameter set are stored and used to determine ISI that compares the deterministic and stochastic interpolation methods, thus, identifying the best adjustment for each model analyzed.

Statistic called error comparison index (ECI—Souza et al., 2016) was used to determine the best semivariogram fit in each j model analyzed, which assumes a lower value for the model is better stochastic methods of interpolation. The best semivariogram of each j model was used in ISI analysis. The reduced mean error (RME) and the standard deviation of the reduced mean error (SDRME) was determined by ordinary kriging cross-validation.

Computational routines by Betzek et al. (2019) were developed in statistical software R, using the geoR library and functions implemented directly in the PostgreSQL database, to determine the best interpolator (and its parameters) based on ECI and ISI. These computational routines were reimplemented, optimized, and made available on the ADB platform. In the geostatistics module, seven semivariogram models are tested (spherical, gaussian, exponential, Matérn 0.5, Matérn 1.0, Matérn 1.5, and Matérn 2.0), as well as two statistical methods to optimize the semivariogram adjustment, ordinary least squares (OLS) and weighted least squares (WLS—Cressie, 1985), thus totalizing 14 different models. For each model, 25 different parameter sets (five initial values for the partial sill parameter and five for range) are used, totalizing 350 different adjustments being analyzed to find the best one. In the IDW module, is analyzed a range of values for the exponent (0.5, 1.0, ..., n) and a range of values for the number of neighbors (4, 5, ..., n). For selecting the best semivariogram model, ISI is used to identify the best value for the exponent and number of neighbors.

Improving models' selection using effective spatial dependence (%ESD)

Three problems should be addressed when selecting the best semivariogram:

1. A minimum of %ESD should be observed. We proposed that %ESDI must be greater than 25%.
2. The selected semivariogram model should contemplate a fraction of SD due only to ($\% \gamma(1)$) lower than 50%.

Table 2 Criteria to select the best interpolation method

Criterion 1 Minimum of effective spatial dependence	Criterion 2 Spatial dependence due only to the first semivariance	Criterion 3 The model needs to express spatial dependence	The best interpolation method
If % <i>ESDI</i> > 25%	and If % $\gamma(1)$ < 50%	and If %SMEI > 20%	IDW or OK with the lowest ISI
If % <i>ESDI</i> ≤ 25%	or If % $\gamma(1)$ ≥ 50%	or If %SMEI ≤ 20%	IDW with the lowest ISI

%*ESDI* effective spatial dependence index, % $\gamma(1)$ first semivariance significance index, *IDW* inverse distance weighting, *OK* ordinary Kriging, *ISI* interpolator selection index

Table 3 Methods used to select the best interpolation model

Methods	Selection of the best interpolation model
Method 1	Best ISI
Method 2	The three criteria are applied after geostatistics analysis + the best ISI
Method 3	The three criteria are applied during geostatistics analysis + the best ISI

- The inclination degree of between the nugget effect and the last adjusted semivariance, estimated by %SMEI, should be greater than 20%. Otherwise, there is an indication of a pure nugget effect.

We proposed that the selection of the best interpolator model should not depend only on ISI but on the criteria presented on Table 2.

The variable selection process was tested using three methods (Table 3): (i) method 1: best ISI, (ii) method 2 (Fig. 5): the three criteria (Table 2) are applied after geostatistics analysis, (iii) method 3 (Fig. 6): The three criteria are applied during geostatistics analysis.

The main difference between methods 2 and 3 is observed when the three criteria are applied. In method 2, the three criteria are applied to analyze geostatistical models after the ISI determination step and the best interpolator’s indication (Fig. 5). For each semivariogram model and estimation method (Spherical OLS, Spherical WLS, Exponential OLS, Exponential WLS, etc.), all analyses to estimate semivariogram parameters are considered (5 partial sill intervals * 5 range intervals = 25 analysis). In method 3 (Fig. 6), a modification was proposed to filter out unsatisfactory geostatistical models before ECI has determined a semivariogram model’s best fit. Therefore, when selecting the analyses by ECI, only the cleaned models (not discarded) by the new selection criteria are considered.

Selection Methods 2 and 3 can lead to different results. The central aspect of method 3 is to allow another ‘fitted model’ to be selected in an interpolator selection analysis. In geostatistical analysis, for each combination of ‘geostatistical model’ (Spherical, Exponential, etc.) vs. ‘estimation method’ (OLS and WLS), 25 ‘fitted models’ (5 partial sill interval * 5 range intervals) are generated. When applying the selection criteria by Method 2, and eliminating the ‘fitted model’ that was considered the best, it is impossible to use another ‘fitted model’ from the same combination of ‘geostatistical model’ vs. ‘estimation method.’ In this case, the twenty five analyses were eliminated. On the other hand, selection by Method 3 makes it possible to

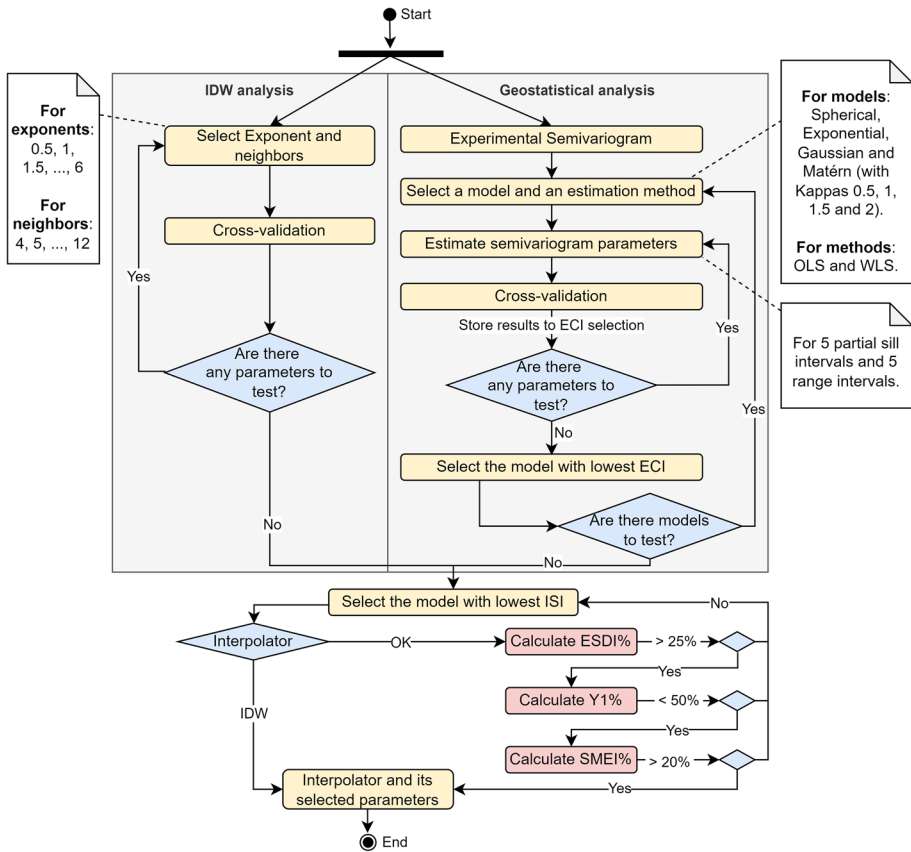


Fig. 5 Selection process of the best interpolator between inverse distance weighting and ordinary Kriging by method 2: the filters using %ESDI, $\gamma(1)$, and %SMEI were applied after geostatistics analysis

use other ‘adjusted models’ within the combined analysis of ‘geostatistical model’ vs. ‘estimation method’.

Map’s evaluation

The interpolated maps were compared using the coefficient of relative deviation (CRD) proposed by Coelho et al. (2009). The coefficient expresses the average absolute percent difference between both maps. The choice of a reference map used for comparison is arbitrary. For this study, the map generated by the best interpolator selected by Method 3 was considered the reference for each variable.

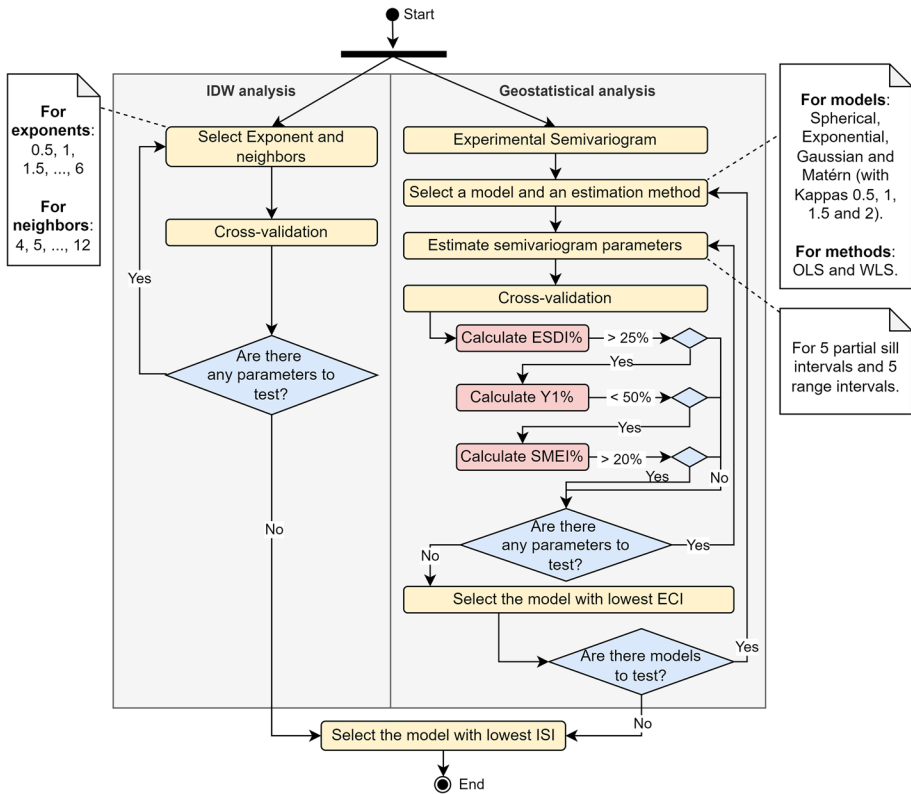


Fig. 6 Selection process of the best interpolator between inverse distance weighting and ordinary Kriging by method 3: the filters using %ESDI, % $\gamma(1)$, and %SMEI were applied during geostatistics analysis

Results and discussion

Descriptive statistics

The descriptive analysis of variables (Tables 4, 5, 6) showed that CV varied from 5% (low, pH SMP) to 118% (very high, Al in field A-2018), 5% (low, pH SMP, and clay) to 123% (very high, aluminum saturation-m% in field A-2019), and from 4% (low, pH SMP, field B-2015) to 146% (very high, Al in field B-2015). Variables Al, C, Ca, Cu, Fe, K, Mg, OM, P, pH (CaCl₂), pH SMP, V, m%, clay, sand, and silt had points that were eliminated after eliminating outliers during EDA. Few outliers were found and eliminated in ten, nine, and twelve variables in fields A-2018, A-2019, and B. In several cases, variables did not present normality at 5% significance level: (i) field A-2018: Al, Cu, H + Al, K, and P; (ii) field A-2019: Al, m%, P, pH (CaCl₂), Zn, and sand; and (iii) field B-2015: Al, C, H + Al, and P.

Table 4 Descriptive statistics of soil attributes in field A-2018 (100 samples)

Variables	Samples remained	Minimum	Means	Medians	Maximum	Standard deviations	CV%
Al* (cmol _c /dm ⁻³)	99	0.00	0.538	0.280	2.450	0.635	118 (VH)
C (g/kg)	98	15.5	21.30	21.43	26.45	2.15	10 (M)
Ca (cmol _c /dm ⁻³)	99	2.07	3.52	3.56	5.47	0.722	21 (H)
Cu* (mg/dm ⁻³)	98	1.86	4.02	3.65	8.66	1.37	34 (VH)
Fe (mg/dm ⁻³)	99	4.88	15.60	15.26	29.04	4.38	28 (H)
H + Al* (cmol _c /dm ⁻³)	100	3.97	8.09	7.76	13.06	1.74	21 (H)
K* (cmol _c /dm ⁻³)	100	0.160	0.439	0.405	0.700	0.130	30 (H)
Mg (cmol _c /dm ⁻³)	100	0.690	1.245	1.270	1.840	0.247	20 (M)
Mn (mg/dm ⁻³)	100	42.67	75.12	76.16	110.41	13.79	18 (M)
OM (g/dm ⁻³)	98	26.72	36.73	36.95	45.60	3.70	10 (M)
P* (mg/dm ⁻³)	98	2.60	9.65	8.40	23.50	4.48	46 (VH)
pH (CaCl ₂)	100	3.58	4.42	4.42	5.21	0.371	8 (L)
pH SMP	99	4.70	5.37	5.40	5.90	0.275	5 (L)
SB (cmol _c /dm ⁻³)	100	3.29	5.23	5.35	7.98	0.990	19 (M)
V%	99	20.85	39.47	39.73	57.23	8.62	22 (H)
Zn (mg/dm ⁻³)	100	4.93	8.12	7.88	12.06	1.85	23 (H)

CV coefficient of variation: low (L) when $CV \leq 10\%$, medium (M) when $10\% < CV \leq 20\%$, high (H) when $20\% < CV \leq 30\%$, and very high (VH) when $CV > 30\%$

Al aluminum, C carbon, Ca calcium, Cu copper, Fe iron, H + Al potential acidity, K potassium, Mg magnesium, Mn manganese, OM organic matter, P phosphorus, pH the potential of hydrogen, pH SMP pH of buffer solution Shoemaker–McLean–Pratt, SB the sum of basis, V% base saturation, Zn zinc

*No normality at 5% significance level

Selection of the best interpolator model

Method 1

The results of selecting the best interpolator model for IDW and OK using ISI for variables of fields A-2018 (Table 9—Appendix), A-2019 (Table 10—Appendix), and B (Table 11—Appendix) showed that the OK one is the best interpolator for 35 variables (9 in field A-2018, 16 in field A-2019, and 10 in field B-2015) and IDW for 15 variables (7 in field A-2018, 3 in field A-2019, and 5 in field B-2015).

During SD analysis, the 50%-cutoff limited range to 513 m (field A-2018), 353 m (field A-2019), and 419 m (field B-2015). Therefore, the correspondent number of lags was twelve (field A-2018), eight (field A-2019), and fourteen (field B-2015), always with a minimum of 30 pairs of points. The first semivariance corresponded to 41 m (field A-2018), 45 m (field A-2019), and 31 m (field B-2015). ISI selected IDW as the best interpolator for (i) field A-2018: H + Al, K, Mn, pH CaCl₂, pH SMP, V%, and Zn, (ii) field A-2019: Ca, Cu, K, m%, and SB, and (iii) field B-2015: Ca, Fe, Mg, Mn, and Zn. For the remained variables, OK was indicated as the best interpolator.

Some variables had their semivariogram models considered unsatisfactory, highlighted in Light Salmon (Tables 9, 10, 11). They did not agree with the criteria defined in Table 2 ($\%ESDI > 25\%$, $\%\gamma(1) < 50\%$, and $\%SMEI > 20$).

Table 5 Descriptive statistics of soil attributes in field A-2019 (52 samples)

Variables	Samples remained	Minimum	Means	Medians	Maximum	Standard deviations	CV%
Al* (cmol _c /dm ⁻³)	51	0.00	0.27	0.15	1.15	0.30	112 (VH)
Ca (cmol _c /dm ⁻³)	52	1.50	4.08	4.20	6.90	1.21	30 (H)
Cu (mg/dm ⁻³)	52	4.3	9.2	8.5	14.1	2.2	24 (H)
Fe (mg/dm ⁻³)	52	36	77	75	121	21	28 (H)
H + Al (cmol _c /dm ⁻³)	52	2.74	5.24	4.96	8.36	1.08	21 (H)
K (cmol _c /dm ⁻³)	51	0.090	0.356	0.330	0.800	0.171	48 (VH)
m%*	51	0.00	5.1	2.3	24.0	6.3	123 (VH)
Mg (cmol _c /dm ⁻³)	52	0.40	1.71	1.70	3.00	0.54	32 (VH)
Mn (mg/dm ⁻³)	52	88	162	159	220	31	19 (M)
OM (g/dm ⁻³)	52	14.7	25.8	26.8	41.6	5.3	21 (H)
P* (mg/dm ⁻³)	51	4.4	18.1	15.8	53.0	11.0	59 (VH)
pH* (CaCl ₂)	51	3.80	4.50	4.50	5.30	0.35	8 (L)
pH SMP	52	5.30	5.96	6.00	6.80	0.29	5 (L)
SB (cmol _c /dm ⁻³)	52	2.2	6.2	6.3	10.6	1.6	27 (H)
V%	52	20.7	53.4	57.1	79.4	11.4	21 (H)
Zn* (mg/dm ⁻³)	50	1.44	3.97	3.77	9.41	1.44	36 (VH)
Clay (%)	51	68.0	74.0	74.0	84.0	3.50	5 (L)
Sand* (%)	50	0.70	2.51	2.60	5.10	0.84	33 (VH)
Silt (%)	51	14.3	23.3	23.2	30.8	3.4	15 (M)

CV coefficient of variation: low (L) when $CV \leq 10\%$, medium (M) when $10\% < CV \leq 20\%$, high (H) when $20\% < CV \leq 30\%$, and very high (VH) when $CV > 30\%$

Al aluminum, Ca calcium, Cu copper, Fe iron, H + Al potential acidity, K potassium, m% aluminum saturation, Mg magnesium, Mn manganese, OM organic matter, P phosphorus, pH the potential of hydrogen, SB the sum of basis, SMP pH of buffer solution Shoemaker–McLean–Pratt, V% base saturation, Zn zinc

*No normality at 5% significance level

The variables' spatial dependences (SD, Fig. 7), measured by the traditional %SDI, were classified, on average, as medium (24%), as high (20%), and very high (30%). However, using %ESDI (Eq. 1), SD was classified, on average, as medium (22%), as high (16%), and very high (12%). That means that the high and very high sum lowered from 50 to 28% and that %SDI masks the actual SD.

According to the visual inspection of each variable semivariogram (Tables 9, 10, 11), there seems to be a lack of adjustment of the model pointed out as the best for some variables in the fields A-2018 (K), A-2019 (Al, H + Al, K, m%, pH SMP, and V%) and B (V%). In other cases, there is an indication of pure nugget effect in field A-2019 (OM and pH CaCl₂) and field B-2015 (Al, Ca, H + Al, P, pH CaCl₂, pH SMP, and SB). Clay and silt can also be included in this list (field A-2019). Among the variables with “doubtful” or “pure nugget effect” adjustment, IDW interpolator was considered the best only for K, fields A-2018, and A-2019, and Ca in field B-2015.

Another aspect observed was the fact that %SDI (Fig. 7) indicated wrongly the presence of strong spatial dependence (high or very high) in some variables in the following areas: (i) field A-2018: K; (ii) field A-2019: Al, H + Al, K, m%, pH SMP, and V%; and (iii) field B-2015: V%. The first semivariance plotted in the semivariograms of these variables shows

Table 6 Descriptive statistics of soil attributes in field B-2015 (73 samples)

Variables	Samples remained	Minimum	Means	Medians	Maximum	Standard deviations	CV%
Al* (cmol _c /dm ⁻³)	72	0.000	0.065	0.020	0.390	0.095	146 (VH)
C* (g/kg)	72	16.9	21.7	21.4	27.7	2.4	11 (M)
Ca (cmol _c /dm ⁻³)	73	3.11	5.35	5.38	8.36	1.03	19 (M)
Cu (mg/dm ⁻³)	72	11.6	14.8	14.8	20.3	1.6	11 (M)
Fe (mg/dm ⁻³)	73	32.3	55.6	53.6	85.0	11.4	20 (H)
H + Al* (cmol _c /dm ⁻³)	73	3.18	5.87	5.76	9.00	1.05	18 (M)
K (cmol _c /dm ⁻³)	72	0.19	0.446	0.430	0.960	0.156	35 (VH)
Mg (cmol _c /dm ⁻³)	72	1.17	2.08	2.06	3.15	0.41	20 (M)
Mn (mg/dm ⁻³)	73	224	316	313	400	49	15 (M)
P* (mg/dm ⁻³)	72	4.8	12.4	11.1	29.9	5.4	43 (VH)
pH (CaCl ₂)	72	4.40	5.04	5.05	5.70	0.29	6 (L)
pH SMP	72	5.20	5.78	5.80	6.20	0.22	4 (L)
SB (cmol _c /dm ⁻³)	73	4.7	7.9	8.0	12.0	1.4	18 (M)
V%	73	34.3	57.2	58.2	79.1	8.4	15 (M)
Zn (mg/dm ⁻³)	73	2.25	4.76	4.59	8.41	1.42	30 (H)

CV coefficient of variation: low (L) when $CV \leq 10\%$, medium (M) when $10\% < CV \leq 20\%$, high (H) when $20\% < CV \leq 30\%$, and very high (VH) when $CV > 30\%$

Al aluminum, C carbon, Ca calcium, Cu copper, Fe iron, H + Al potential acidity, K potassium, Mg magnesium, Mn manganese, P phosphorus, pH the potential of hydrogen, pH SMP pH of buffer solution Shoemaker–McLean–Pratt, SB the sum of basis, V% base saturation, Zn zinc

*No normality at 5% significance level

a high variance of data at the closest distances and that the model was adjusted incorrectly. In these cases, %SDI gives some false feeling of having an adequate model, which presents a strong spatial dependence.

This kind of problem with semivariogram adjustments is due to the model's automatic adjustment to the semivariogram made by geoR package's routines. The automatic adjustment of models to semivariograms is pointed out in literature as a notoriously tricky task (Webster & Oliver, 1990; Goovaerts, 1997). As with any method for adjusting the variogram model, they all assume the model's basic structure in advance and then obtained the predefined model structure's optimal coefficients. Selecting the variogram model and its parameters is the most controversial aspect of geostatistics; shapes of valid variogram models are finite; sometimes, the model's optimal shape cannot be fitted, leading to reduced estimation accuracy (Han et al., 2016). In this sense, it is proposed in this work criteria (using %ESDI, % $\gamma(1)$, and %SMEI) to improve the semivariogram adjustment process, which is presented by Methods 2 and 3.

Method 2

This method was applied to variables with unsatisfactory semivariogram models (Tables 9, 10, 11). As a result, other semivariogram models were selected for variables in field A-2019 (Al, H + Al, m%, pH CaCl₂, m%, and clay). In another case, the IDW

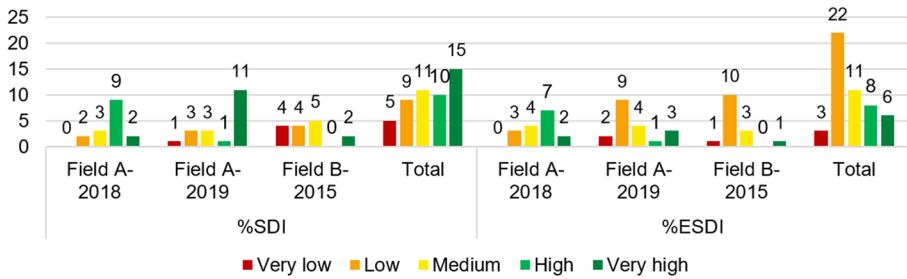


Fig. 7 Number of variables of each class for %SDI and %ESDI (very low, low, medium, high, and very high) for each field (A-2018, A-2019, and B-2015)

interpolator was considered the best for variable SB (field B-2015) (Table 10). It is noteworthy that variables OM and silt, from field A-2019, and C, H + Al, P, pH SMP, and V%, from field B-2015, had all semivariogram models eliminated. In these cases, the IDW interpolator was considered the best one.

IDW interpolator was considered using Method 1 as the best interpolator for variable K, in fields A-2018 (Table 9) and A-2019 (Table 10), and for variable Ca, in field B-2015 (Table 11). However, other semivariogram models’ selection behavior was evaluated regardless of whether IDW was identified as the best. As a result, this allowed us to verify that the variable K, from fields A-2018 and A-2019, and the variable Ca, from field B-2015, could choose another semivariogram model (Table 12—Appendix).

It is essential to highlight that the three criteria must be considered together in the semivariogram models’ selection process. According to the semivariogram structure, a wrong model can be selected when it is not applied in association (see results in Table 7). This issue was the most important in field A-2019 and the least important in field A-2018.

Method 3

This method, like Method 2, was applied to the variables with unsatisfactory semivariogram models (Tables 9, 10, 11). As a result, some models were eliminated in favor of others. In OM and silt variables, from field A-2019, and in C, H + Al, P, and V% variables, from field B-2015, all geostatistical models were eliminated during the geostatistical analysis (Table 13—Appendix). All other variables had changes in semivariogram parameters in comparison to Method 1.

Other semivariogram models were selected for variables in field A-2019 (Al, m%, pH CaCl₂, and clay) and field B-2015 (pH CaCl₂, pH SMP, and SB) (Table 12). In other cases, the IDW interpolator was considered the best one: field A-2019 (OM and silt) and field B-2015 (C, H + Al, P, pH SMP, SB, and V%).

Variable V% (field A-2019) kept the model selected by Method 1 (Spherical – OLS or WLS) but with other semivariogram adjusting parameters. In variables H + Al, K, and pH SMP (field A-2019) and Ca, the model selected by Method 1 (Spherical) remained; however, the method of adjusting the semivariogram changed between OLS and WLS. Variables Al, m%, and clay (field A-2019) and Ca and SB (field B-2015) kept the model selected in Method 2. Despite maintaining the models, variables K (field A-2018) and pH SMP (field A-2019) changed the semivariogram adjustment parameters.

Table 7 Result of selecting the best interpolator model for ordinary Kriging (OK) with Method 2 using each criterion separately and all together for variables of fields A-2018, A-2019, and B-2015

Variables/ Fields	Criterion 1 only %ESDI > 25%	Criterion 2 only % $\gamma(1) < 50\%$	Criterion 3 only %SMEI > 20%	All criteria
K Field A-2018	Spherical – OLS*	Exponential – OLS*	Spherical – OLS*	Exponential – OLS*
Al Field A-2019	Spherical – WLS	Matérn 2 – WLS	Spherical – WLS	Matérn 2 – WLS
H+Al Field A-2019	Spherical – OLS	Exponential – WLS	Spherical – OLS	Exponential – WLS
K Field A-2019	Spherical – OLS*	Gaussian – WLS*	Spherical – OLS*	Gaussian – WLS*
m% Field A-2019	Spherical - WLS	Gaussian - OLS	Spherical - WLS	Gaussian - OLS
OM Field A-2019	All geostatistical models were eliminated	All geostatistical models were eliminated	All geostatistical models were eliminated	All geostatistical models were eliminated
pH CaCl2 Field A-2019	Matérn 1 – OLS	Gaussian – OLS	Gaussian – OLS	Matérn 1 – OLS
pH SMP Field A-2019	Spherical - WLS	Spherical - OLS	Spherical - WLS	Spherical - OLS
V% Field A-2019	Spherical - WLS	Spherical - WLS	Spherical - WLS	Spherical - WLS
Clay Field A-2019	Matérn 2 – OLS	Gaussian – OLS	Gaussian – OLS	Matérn 2 – OLS

As it was expected, Methods 2 and 3 conducted different results. Method 3 allows another ‘fitted model’ to be selected in the geostatistical analysis, and as it was explained in section M&M, it is expected to lead to the best interpolator model (IDW or OK).

Table 7 (continued)

Variables/ Fields	Criterion 1 only %ESDI > 25%	Criterion 2 only % $\gamma(1) < 50\%$	Criterion 3 only %SMEI > 20%	All criteria
Silt Field A-2019	All geostatistical models were eliminated	Gaussian – OLS 	Matérn 1 – OLS 	All geostatistical models were eliminated
C Field B-2015	All geostatistical models were eliminated	Gaussian – WLS 	Gaussian – WLS 	All geostatistical models were eliminated
Ca Field B-2015		Exponential – OLS 	Spherical - OLS 	Spherical - OLS
H+Al Field B-2015		Matérn 1.5 – WLS 	Exponential – WLS 	All geostatistical models were eliminated
P Field B-2015	All geostatistical models were eliminated	All geostatistical models were eliminated	Exponential – WLS 	All geostatistical models were eliminated
pH CaCl2 Field B-2015	Gaussian – WLS 	Gaussian – WLS 	Gaussian – OLS 	Gaussian – OLS
pH SMP Field B-2015	Matérn 2 – WLS 	Matérn 2 – WLS 	All geostatistical models were eliminated	All geostatistical models were eliminated
SB Field B-2015	Gaussian - OLS 	Gaussian - OLS 	Spherical – WLS* 	Spherical – WLS*
V% Field B-2015	Exponential – WLS 	All geostatistical models were eliminated	Exponential – WLS 	All geostatistical models were eliminated

OLS ordinary least squares, WLS weighted least squares, %ESDI effective spatial dependence index, % $\gamma(1)$ first semivariance significance index, %SMEI slope of the model ends index, Al aluminum, C carbon, Ca calcium, H + Al potential acidity, K potassium, m% aluminum saturation, OM organic matter, P phosphorus, pH the potential of hydrogen, pH SMP pH of buffer solution Shoemaker–McLean–Pratt, SB sum of basis, V% base saturation

*The IDW interpolator was considered better than the model adjusted to the semivariogram

Comparison of the three methods

When comparing the interpolator selection result for the variables considered with inadequate geostatistical models, it can be noticed that the selected interpolator might change according to the selection method (Table 8).

The variables OM and silt, from field A-2019, and C, H+Al, P, pH SMP, SB, and V%, from field B-2015 registered that Method 1 had considered OK as the best interpolator, and, after applying the selection criteria by Methods 2 and 3, it started to consider IDW as the best interpolator. Most of these variables had all geostatistical models eliminated after applying the selection criteria, except for variables SB and pH SMP (by Method 3) from field B-2015.

Even with eliminating inappropriate geostatistical models, K, from fields A-2018 and A-2019, and Ca, from field B-2015, kept IDW as the best interpolator. The other variables, Al, H+Al, m%, pH CaCl₂, pH SMP, V%, and clay, from field A-2019, and pH CaCl₂, from field B-2015, kept OK as the best interpolator, as selected by method 1. However, there was the selection of other geostatistical models after selection by Methods 2 and 3.

Thematic maps

Thematic maps (TMs, Table 14—Appendix) were generated by OK using the semivariogram selected by each of three methods and IDW with its best interpolator. The variables are the same as in Table 8. The best interpolator was considered the one selected with Method 3.

Using CRD to compare the maps generated by the interpolator selected by Method 3 (IDW or OK) versus the best semivariogram model indicated by Method 1 (Fig. 8), it can be seen that:

- the selection of other interpolator parameters can result in large differences among the maps. In variable Al, from area A-2019, the best interpolator model, selected by Method 3 (Matérn 2—OLS), deviated by 64% from the map selected by Method 1 (Spherical—WLS);
- the difference was below 5% in eight variables;
- the difference was from 5 to 10% in seven variables;
- over 10% in four variables.

When comparing the maps generated by the interpolator selected by method 3 (IDW or OK) versus the best semivariogram model indicated by method 2 (Fig. 9), it can be seen that:

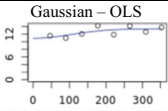
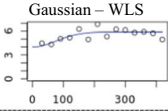
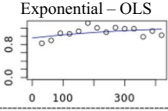
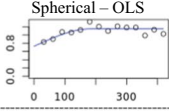
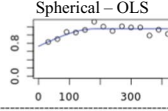
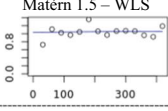
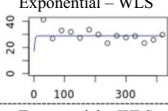
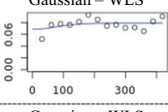
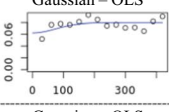
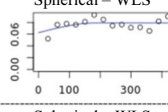
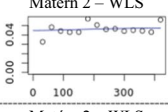
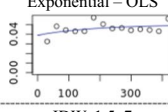
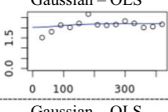
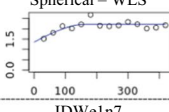
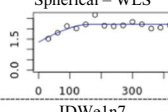
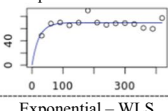
- The most significant difference was observed in variable K (field A-2018; 18%);
- The difference was below 5% in ten variables;
- The difference between 5% and 10% in one variable.

Our study analyzed 50 cases, and in 23 of them, IDW outperformed OK. Consequently, in 27 cases, OK was better than IDW. These results confirm the ones presented by Mueller et al. (2004), i.e., for sample datasets with semivariograms, which did not indicate spatial structure, IDW was a better choice than OK with a nugget model.

Table 8 The best interpolation models selected with each of the three methods

Variables/ Fields		Method 1	Method 2	Method 3
K Field A-2018	Best Semivariogram	Spherical – OLS 	Gaussian – OLS 	Exponential – OLS
	Best Interpolator	IDWe3.5n7	IDWe3.5n7	IDWe3.5n7
Al Field A-2019	Best Semivariogram	Spherical – WLS 	Matérn 2 – WLS 	Matérn 2 – WLS
	Best Interpolator	Spherical – WLS	Matérn 2 – WLS	Matérn 2 – WLS
H+Al Field A-2019	Best Semivariogram	Spherical – OLS 	Exponential – WLS 	Spherical – WLS
	Best Interpolator	Spherical – OLS	Exponential – WLS	Spherical – OLS
K Field A-2019	Best Semivariogram	Spherical – OLS 	Gaussian – WLS 	Spherical – WLS
	Best Interpolator	IDWe1n10	IDWe1n10	IDWe1n10
m% Field A-2019	Best Semivariogram	Spherical – WLS 	Gaussian – OLS 	Gaussian – OLS
	Best Interpolator	Spherical – WLS	Gaussian – OLS	Gaussian – OLS
OM Field A-2019	Best Semivariogram	Matérn 2 – WLS 	All geostatistical models were eliminated	All geostatistical models were eliminated
	Best Interpolator	Matérn 2 – WLS	IDWe1n4	IDWe1n4
pH CaCl2 Field A-2019	Best Semivariogram	Gaussian – OLS 	Matérn 1 – OLS 	Spherical – OLS
	Best Interpolator	Gaussian – OLS	Matérn 1 – OLS	Spherical – OLS
pH SMP Field A-2019	Best Semivariogram	Spherical – WLS 	Spherical – OLS 	Spherical – OLS
	Best Interpolator	Spherical – WLS	Spherical – OLS	Spherical – OLS
V% Field A-2019	Best Semivariogram	Spherical – WLS 	Spherical – OLS 	Spherical – WLS
	Best Interpolator	Spherical – WLS	Spherical – OLS	Spherical – WLS
Clay Field A-2019	Best Semivariogram	Gaussian – OLS 	Matérn 2 – OLS 	Matérn 2 – OLS
	Best Interpolator	Gaussian – OLS	Matérn 2 – OLS	Matérn 2 – OLS

Table 8 (continued)

Variables/ Fields	Method 1	Method 2	Method 3
Silt Field A-2019	Best Semivariogram 	All geostatistical models were eliminated	All geostatistical models were eliminated
	Best Interpolator		
C Field B-2015	Best Semivariogram 	All geostatistical models were eliminated	All geostatistical models were eliminated
	Best Interpolator		
Ca Field B-2015	Best Semivariogram 	Spherical - OLS 	Spherical - OLS 
	Best Interpolator	IDWe1.5n7	IDWe1.5n7
H+Al Field B-2015	Best Semivariogram 	All geostatistical models were eliminated	All geostatistical models were eliminated
	Best Interpolator		
P Field B-2015	Best Semivariogram 	All geostatistical models were eliminated	All geostatistical models were eliminated
	Best Interpolator		
pH CaCl2 Field B-2015	Best Semivariogram 	Gaussian - OLS 	Spherical - WLS 
	Best Interpolator	Gaussian - WLS	Spherical - WLS
pH SMP Field B-2015	Best Semivariogram 	All geostatistical models were eliminated	Exponential - OLS 
	Best Interpolator		
SB Field B-2015	Best Semivariogram 	Spherical - WLS 	Spherical - WLS 
	Best Interpolator	Gaussian - OLS	IDWe1n7
V% Field B-2015	Best Semivariogram 	All geostatistical models were eliminated	All geostatistical models were eliminated
	Best Interpolator		

Method 1 Only the best ISI, *Method 2* the three criteria are applied after geostatistics analysis, *Method 3* the three criteria are applied during geostatistics analysis, *IDWe3.5n7* means inverse distance weighting with exponent 3.5 and 7 neighbors, *OLS* ordinary least squares, *WLS* weighted least squares, *Al* aluminum, *C* carbon, *Ca* calcium, *H + Al* potential acidity, *K* potassium, *m%* aluminum saturation, *OM* organic matter, *P* phosphorus, *pH* the potential of hydrogen, *pH SMP* pH of buffer solution Shoemaker–McLean–Pratt, *SB* sum of basis, *V%* base saturation

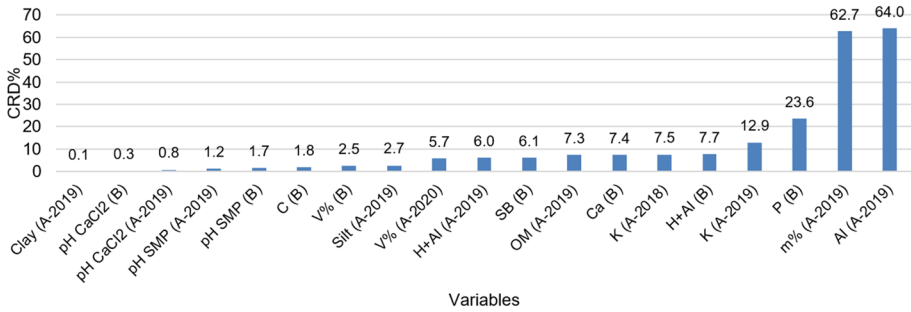


Fig. 8 The coefficient of relative deviation (CRD) between the interpolator selected by method 3 (IDW or OK) versus the best semivariogram model indicated by method 1

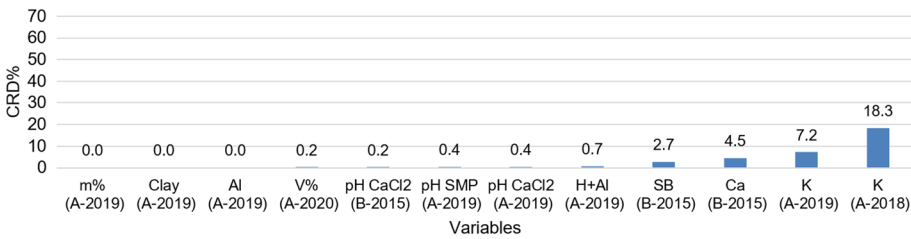


Fig. 9 The coefficient of relative deviation (CRD) between the interpolator selected by method 3 (IDW or OK) versus the best semivariogram model indicated by method 2

Work contribution

Thematic maps in precision agriculture allow identifying the spatial distribution of geographical attributes, soil, and plant productivity (Bazzi et al., 2015). Estimating values for unsampled regions is important to reduce costs with laboratory analysis. More accurate estimates of the interpolated positions contribute to the correct interpretation of the analyzed phenomena, helping the producer in decision-making.

Several precision agriculture applications are available for farmers. However, existing software for creating TMs are not developed specifically for precision agriculture, but for generic data handling (Whelan & Taylor, 2013). Choosing a tool not dedicated to precision agriculture can be challenging (Borges et al., 2020).

As the ADB platform is biased towards precision agriculture, it provides the necessary tools to create TMs without dependence on various software. It allows users with no specific skills to obtain the analysis result, without getting involved in too many process details. On the other hand, it also allows experienced users to choose the analysis settings. The automated routine for interpolator selection calculates, in its default configuration, 398 deterministic and stochastic models, and by ISI selects the best among them. Therefore, this work contributed to the improvement of data interpolation, eliminating the possibility of selecting the wrong model by the automatic selection process, and resulting in more accurate estimates of the data set.

Spatial variability characterization of soil's chemical and physical attributes with greater precision allows, for example, prescription maps creation of fertilizer in variable rates and correctives for the soil and plant. Hence, this may optimize the use of fertilizers and other inputs.

Conclusion

The inclusion of the three criteria (i) effective spatial dependence index (%ESDI) > 25%, (ii) the first semivariance significance index ($\% \gamma(1) < 50\%$ and (iii) the slope of the model ends index (%SMEI) > 20% improved the selection of the best interpolator using only the interpolator selection index (ISI—Bier and Souza, 2017).

The comparison carried out the methodology influence on selecting the best interpolator among the studied thematic maps using three Methods: (i) Method 1—best ISI; (ii) Method 2—the three criteria were applied after geostatistics analysis; Method 3—the three criteria are applied during geostatistics analysis. Method 3 showed as the best approach. The coefficient of relative deviation (CRD) varied from 0.1 to 64% when comparing the maps generated by the three methods.

The newly proposed measurement of the effective spatial dependence index (ESDI) of a semivariogram showed better performance than the usual spatial dependence index (%SDI) widely adopted in the literature.

With the implementation of the methods shown in the ADB platform, it appears that farmers and researchers who work with precision agriculture will have a free tool to carry out analyses in situations where it is difficult to create adequate geostatistical models for the thematic map's creation.

Appendix

See Tables 9, 10, 11, 12, 13 and 14.

Table 9 Result of selecting the best interpolator model for inverse distance weighted interpolation (IDW) and ordinary Kriging (OK) using the interpolator selection index (ISI) for variables of field A-2018, using Method 1: Selection using only ISI

Variables	Models	Geostatistics											IDW					Best Interpolator
		C_0	C_1	Ra	%SDI	%ESDI	$\% \gamma(1)$	%SMEI	ISI	$ME \cdot 10^2$	SDME	Semivariogram	Exp	Neig	ISI	$ME \cdot 10^2$	SDME	
Al	Gaussian WLS	0.158	0.382	176	71 (H)	65 (H)	8	71	0.00223	-0.00650	0.428		5	5	0.0679	0.013	0.458	OK
C	Gaussian OLS	3.44	1.75	234	34 (L)	32 (L)	5	33	0.0935	-0.817	2.031		1	7	0.419	-3.522	2.067	OK
Ca	Matern 1 OLS	0.256	0.495	225	66 (H)	68 (H)	-4	60	0.0238	-0.037	0.529		4.5	5	0.0251	0.00876	0.536	OK
Cu	Gaussian OLS	0.605	2.03	336	77 (H)	83 (VH)	-8	75	0.123	0.591	0.736		6	7	0.540	4.505	0.702	OK
Fe	Exponential WLS	11.3	10.2	244	47 (M)	46 (M)	3	40	0.0243	0.129	4.024		2	11	0.0308	-0.292	3.989	OK
H+Al	Spherical OLS	1.01	2.70	313	73 (H)	54 (M)	26	73	0.131	0.672	1.281		5.5	10	0.0032	-0.00144	1.276	IDW
K	Spherical WLS	0.00361	0.0123	128	77 (H)	27 (L)	65	77	0.1156	-0.0292	0.107		3.5	7	0.00901	-0.000650	0.106	IDW
Mg	Matern 2 OLS	0.0348	0.0565	128	62 (H)	68 (H)	-10	58	0.0575	-0.0268	0.184		5.5	5	0.108	-0.000691	0.203	OK
Mn	Matern 2 OLS	42.06	284	138	87 (VH)	82 (VH)	6	85	0.0695	-1.401	7.579		1.5	10	0.0313	-0.161	7.665	IDW
OM	Gaussian OLS	10.22	5.19	234	34 (L)	32 (L)	5	33	0.0934	-1.408	3.501		1	7	0.419	-6.070	3.564	OK
p	Exponential WLS	7.62	19.4	200	72 (H)	55 (M)	23	70	0.0255	-0.175	3.735		3	9	0.0609	0.154	3.895	OK
pH CaCl2	Matern 2 OLS	0.106	0.0744	128	41 (M)	51 (M)	-23	38	0.0598	-0.0177	0.308		1	8	0.0234	0.000577	0.303	IDW
pH SMP	Matern 2 WLS	0.052	0.0618	128	54 (M)	67 (H)	-24	51	0.0745	-0.0323	0.205		4.5	9	0.00942	-0.00171	0.199	IDW
SB	Exponential OLS	0.524	1.125	513	68 (H)	66 (H)	3	58	0.0159	-0.039	0.754		3.5	5	0.0172	-0.0226	0.759	OK
V%	Exponential OLS	1.94	92.9	120	98 (VH)	64 (H)	35	98	0.168	-4.941	6.104		2.5	8	0.0104	-0.0771	6.141	IDW
Zn	Gaussian OLS	1.29	3.12	250	71 (H)	67 (H)	5	71	0.140	-0.514	1.209		2.5	11	0.0159	0.0107	1.226	IDW

C_0 nugget effect, C_1 partial sill, Ra range, %SDI spatial dependence index, %ESDI effective spatial dependence index, $\% \gamma(1)$ first semivariance significance index, ISI interpolator selection index, ME mean error, SDME standard deviation of mean error, IDW inverse distance weighting, OK ordinary Kriging, Exp exponent, Neig neighbors, OLS ordinary least squares, WLS weighted least squares

Classification of %SDI and ESDI: very low for %SDI < 20%, low for 20 ≤ %SDI < 40%, medium for 40 ≤ %SDI < 60%, high for 60 ≤ %SDI < 80%, and very high for %SDI > 80%. Values highlighted in Light Salmon do not agree with the criteria defined in Table 2 (%ESDI > 25%, $\% \gamma(1)$ < 50%, and %SMEI > 20)

Table 10 Result of selecting the best interpolator model for inverse distance weighted interpolation (IDW) and ordinary Kriging (OK) using the interpolator selection index (ISI) for variables of field A-2019, using Method 1: Selection using only ISI

Variables	Models	C ₀	C ₁	Ra	Geostatistics						Semivariogram	IDW					Best Interpolator	
					%SDI	%ESDI	% $\gamma(1)$	%SMEI	ISI	ME*10 ²		SDME	Exp	Neig	ISI	ME*10 ²		SDME
Al	Spherical WLS	0.00646	0.0885	88	93 (VH)	28 (L)	70	93	0.161	0.159	0.282		1	9	0.435	0.736	0.270	OK
Ca	Matérn 2 WLS	0.732	7.97	353	92 (VH)	91 (VH)	1	67	0.289	-1.295	0.925		1	9	0.223	-1.121	0.892	OK
Cu	Matérn 0.5 WLS	0.64	7.63	353	92 (VH)	82 (VH)	11	88	0.1170	1.898	1.416		4.5	5	0.0594	0.153	1.478	IDW
Fe	Gaussian OLS	253	353	286	58 (M)	56 (M)	5	52	0.0082	0.571	16.366		1.5	6	0.0837	-0.958	17.38	OK
H+Al	Spherical OLS	0.00	1.18	89.1	100 (VH)	30 (L)	70	100	0.111	0.161	16.471		6	4	0.551	2.328	1.105	OK
K	Spherical OLS	0.00	0.0243	88.2	100 (VH)	35 (L)	65	100	0.138	0.0392	1.012		1	10	0.0954	0.0173	0.145	IDW
Mg	Gaussian WLS	0.229	0.183	335	44 (M)	39 (L)	13	35	0.00139	0.000468	0.147		1	6	0.0310	0.0250	0.518	OK
Mn	Exponential WLS	542	506335	260037	100 (VH)	50 (M)	10	56	0.138	-15.778	26.064		6	9	0.221	-24.05	27.08	OK
OM	Matérn 2 WLS	24.0	6.97	170	23 (L)	-3 (VL)	126	13	0.024	-0.434	31.68		1	4	0.481	-7.212	5.814	OK
P	Matérn 2 WLS	79.0	340	337	81 (VH)	79 (H)	3	47	0.00303	-0.217	5.079		1.5	9	0.0645	-0.911	9.950	OK
pH CaCl2	Gaussian OLS	0.0768	0.0511	88.2	40 (L)	22 (L)	46	40	0.0392	-0.0642	9.342		6	8	0.321	-0.272	0.370	OK
m%	Spherical WLS	0.0	40.9	89	100 (VH)	30 (L)	70	100	0.175	3.372	5.772		1	9	0.388	13.025	5.473	OK
SB	Matérn 2 WLS	1.48	12.8	353	90 (VH)	88 (VH)	2	62	0.277	-1.430	1.330		1	10	0.0876	-0.416	1.316	IDW
pH SMP	Spherical WLS	0.00	0.0746	86	100 (VH)	28 (L)	72	100	0.137	-0.0296	0.266		6	4	0.580	-0.550	0.281	OK
V%	Spherical WLS	5.9	118	87	95 (VH)	28 (L)	71	95	0.148	-1.66	10.662		1	9	0.443	-21.31	9.472	OK
Zn	Gaussian OLS	1.16	3.08	446	73 (H)	43 (M)	17	52	0.0229	0.0375	3.286		2.5	5	0.0476	0.0636	1.147	OK
Clay	Gaussian OLS	10.9	4.88	205	31 (L)	25 (L)	18	30	0.0396	-0.678	0.263		1	5	0.142	-0.996	3.639	OK
Sand	Gaussian WLS	0.332	0.435	142.9	57 (M)	45 (M)	21	57	0.0825	0.653	0.659		1	4	0.293	1.853	0.706	OK
Silt	Gaussian OLS	10.9	2.50	146	19 (VL)	14 (VL)	27	19	0.0243	0.277	1.126		1	10	0.0511	-0.407	3.351	OK

C₀ nugget effect, C₁ partial sill, Ra range, %SDI spatial dependence index, %ESDI effective spatial dependence index, % $\gamma(1)$ first semivariance significance index, ISI interpolator selection index, ME mean error, SDME standard deviation of mean error, IDW inverse distance weighting, OK ordinary Kriging, Exp exponent, Neig neighbors, OLS ordinary least squares, WLS weighted least squares

Classification of %SDI and ESDI: very low for %SDI < 20%, low for 20 ≤ %SDI < 40%, medium for 40 ≤ %SDI < 60%, high for 60 ≤ %SDI < 80%, and very high for %SDI > 80%. Values highlighted in Light Salmon do not agree with the criteria defined in Table 2 (%ESDI > 25%, % $\gamma(1)$ < 50%, and %SMEI > 20)

Table 11 Result of selecting the best interpolator model for inverse distance weighted interpolation (IDW) and ordinary Kriging (OK) using the interpolator selection index (ISI) for variables of field B-2015, using Method 1: Selection using only ISI

Variable	Model/method	C ₀	C ₁	Ra	Geostatistics							IDW					Best Interpolator		
					%SDI	%ESDI	%γ(1)	%SMEI	ISI	ME*10 ³	SDME	Semivariogram	Exp	Neig	ISI	ME*10 ³		SDME	
Al	Gaussian WLS	0.00658	0.00209	105	24 (L)	34 (L)	-42	24	0.0456	0.138	0.0926		2.5	4	0.0783	-0.0378	0.0996	OK	
C	Gaussian WLS	3.99	1.89	106	32 (L)	24 (L)	25	32	0.0681	-9.685	2.123		1	7	0.114	-12.84	2.176	OK	
Ca	Exponential OLS	0.957	0.308	341	24 (L)	34 (L)	-41	19	0.0437	-1.157	1.015		1.5	7	0.0347	-0.800	1.010	IDW	
Cu	Spherical OLS	1.836	1.19	419	39 (L)	30 (L)	25	39	0.116	9.89	1.536		3	10	0.646	57.66	1.531	OK	
Fe	Matérn 1.5 WLS	61.4	81.0	94.5	57 (M)	59 (M)	-4	55	0.0689	17.95	9.579		1	5	0.0336	-1.994	9.632	IDW	
H+Al	Matérn 1.5 WLS	1.04	0.02521	105	2 (VL)	32 (L)	-1232	2	0.0272	0.2684	1.051		1.5	7	0.0867	1.122	1.109	OK	
K	Spherical WLS	0.0141	0.0180	350	56 (M)	46 (M)	19	56	0.0026	0.0314	0.130		2	6	0.0344	-0.0992	0.134	OK	
Mg	Gaussian OLS	0.0987	0.0682	105	41 (M)	49 (M)	-20	41	0.0972	-1.444	0.371		1.5	6	0.0533	0.677	0.370	IDW	
Mn	Matérn 2 WLS	625	2706	103	81 (VH)	82 (VH)	-1	79	0.127	-159.4	30.04		1	4	0.00380	-11.04	27.75	IDW	
P	Exponential WLS	17.15	11.86	2.4	41 (M)	-43 (VL)	204	41	0.00000003	-	0.000001	5.404		2	7	0.0691	-2.49	5.810	OK
pH CaCl2	Gaussian WLS	0.0687	0.0103	105	13 (VL)	34 (L)	-162	13	0.0801	-0.645	0.274		1	7	0.230	-2.695	0.273	OK	
pH SMP	Matérn 2 WLS	0.0448	0.00297	105	6 (VL)	32 (L)	-415	5	0.0461	-0.145	0.219		1.5	7	0.0625	0.0570	0.226	OK	
SB	Gaussian OLS	2.019	0.178	118	8 (VL)	31 (L)	-281	8	0.0311	-1.457	1.421		1	7	0.0377	-1.620	1.429	OK	
V%	Exponential WLS	0.00	69.7	22.2	100 (VH)	31 (L)	69	100	0.0759	-48.06	8.254		1	7	0.0768	20.69	8.681	OK	
Zn	Exponential OLS	1.132	1.22	145	52 (M)	32 (L)	39	50	0.0811	5.169	1.260		2.5	12	0.0642	1.654	1.311	IDW	

C₀ nugget effect, C₁ partial sill, Ra range, %SDI spatial dependence index, %ESDI effective spatial dependence index, %γ(1) first semivariance significance index, ISI interpolator selection index, ME mean error, SDME standard deviation of mean error, IDW inverse distance weighting, OK ordinary Kriging, Exp exponent, Neig neighbors, OLS ordinary least squares, WLS weighted least squares

Classification of %SDI and ESDI: very low for %SDI < 20%, low for 20 ≤ %SDI < 40%, medium for 40 ≤ %SDI < 60%, high for 60 ≤ %SDI < 80%, and very high for %SDI > 80%. Values highlighted in Light Salmon do not agree with the criteria defined in Table 2 (%ESDI > 25%, %γ(1) < 50%, and %SMEI > 20)

Table 12 Result of selecting the best interpolator model for ordinary Kriging (OK) using the interpolator selection index (ISI) for variables of fields A-2018, A-2019, and B-2015 using Method 2: The three criteria (%ESDI > 25%, $\% \gamma(1) < 50\%$, and %SMEI > 20) are applied after geostatistics analysis

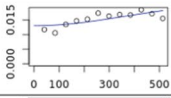
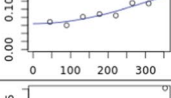
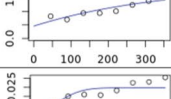
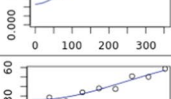
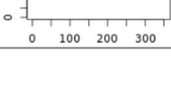
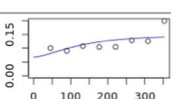
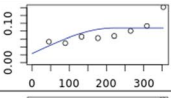
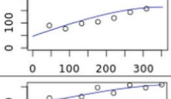
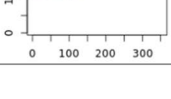

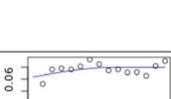
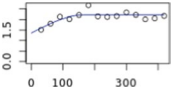
Variables/ Fields	Models	C ₀	C ₁	Ra	%SDI	%ESDI	$\% \gamma(1)$	%SMEI	ISI	ME*10 ²	SDME	Semivariogram	Best Interpolator
K Field A- 2018	Gaussian OLS	0.0130	0.00808	513	38 (L)	45 (M)	-17	28	0.122	-0.015	0.113		IDW
A1 Field A- 2019	Matérn 2 WLS	0.0640	0.372	353	85 (VH)	84 (VH)	1	52	0.1919	0.279	0.272		OK
H+A1 Field A- 2019	Exponential WLS	0.456	1.54	353	77 (H)	59 (M)	24	68	0.2251	1.357	0.916		OK
K Field A- 2019	Gaussian WLS	0.0116	0.0131	88	53 (M)	36 (L)	33	53	0.1474	-0.126	0.136		IDW
m% Field A- 2019	Gaussian OLS	25.73	49.89	353	66 (H)	62 (H)	6	55	0.2179	6.602	5.508		OK
OM Field A- 2019	All geostatistical models were eliminated											IDW	
pH CaCl2 Field A- 2019	Matérn 1 OLS	0.0672	0.0773	88	53 (M)	31 (L)	43	52	0.1255	-0.205	0.313		OK
pH SMP Field A- 2019	Spherical OLS	0.0230	0.0650	221	74 (H)	39 (L)	47	74	0.1827	-0.249	0.236		OK
V% Field A- 2019	Spherical OLS	47.4	116.4	353	71 (H)	45 (M)	36	71	0.2247	-10.551	9.474		OK
Clay Field A- 2019	Matérn 2 OLS	10.8	5.84	96	35 (L)	29 (L)	17	31	0.0452	-0.719	3.290		OK
Silt Field A- 2019	All geostatistical models were removed											IDW	
C Field B- 2015	All geostatistical models were eliminated											IDW	
Ca Field B- 2015	Spherical OLS	0.725	0.425	183	37 (L)	28 (L)	25	37	0.0445	-2.865	0.987		IDW
H+A1 Field B- 2015	All geostatistical models were eliminated											IDW	
P Field B- 2015	All geostatistical models were eliminated											IDW	
													

Table 12 (continued)

Variables/ Fields	Models	C_0	C_1	Ra	%SDI	%ESDI	% $\gamma(1)$	%SMEI	ISI	ME*10 ²	SDME	Semivariogram	Best Inter- polator
pH CaCl ₂ Field B- 2015	Gaussian OLS	0.0624	0.0174	105	22 (L)	35 (L)	-59	22	0.0693	-0.891	0.271		OK
pH SMP Field B- 2015	All geostatistical models were eliminated											IDW	
SB Field B- 2015	Spherical WLS	1.360	0.861	167	39 (L)	32 (L)	18	39	0.0527	-5.516	1.398		IDW
V% Field B- 2015	All geostatistical models were removed											IDW	

C_0 nugget effect, C_1 partial sill, Ra range, %SDI spatial dependence index, %ESDI effective spatial dependence index, % $\gamma(1)$ first semivariance significance index, %SMEI slope of the model ends index, ISI interpolator selection index, ME mean error, SDME standard deviation of mean error, IDW inverse distance weighting, OK ordinary Kriging, OLS ordinary least squares, WLS weighted least squares

Classification of %SDI and ESDI: very low for %SDI/ESDI < 20%, low for 20 ≤ %SDI/ESDI < 40%, medium for 40 ≤ %SDI/ESDI < 60%, high for 60 ≤ %SDI/ESDI < 80%, and very high for %SDI/ESDI > 80%

Table 13 Result of selecting the best interpolator model for ordinary Kriging (OK) using interpolator selection index (ISI) for variables of fields A-2018, A-2019, and B-2015 using Method 3: The three criteria (%ESDI > 25%, % $\gamma(1)$ < 50%, and %SMEI > 20) are applied during geostatistics analysis

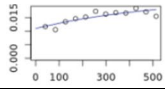
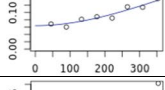
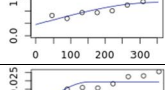
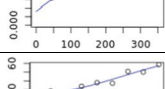
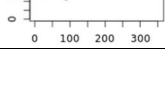
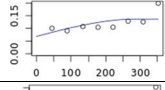
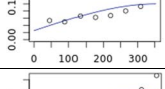
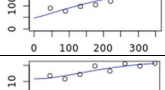
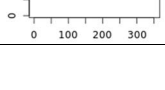
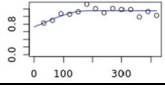
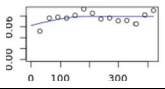
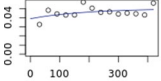
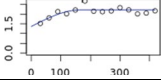
Variables/ Fields	Models	C ₀	C ₁	Ra	%SDI	%ESDI	% $\gamma(1)$	%SMEI	ISI	ME*10 ²	SDME	Semivariogram	Best Interpolator
K Field A- 2018	Exponential OLS	0.0109	0.0111	513	50 (M)	47 (M)	7	39	0.126	-0.028	0.109		IDW
A1 Field A- 2019	Matérn 2 WLS	0.064	0.372	353	85 (VH)	84 (VH)	1	52	0.1919	0.279	0.272		OK
H+A1 Field A- 2019	Spherical WLS	0.454	0.899	353	66 (H)	39 (L)	41	66	0.1872	1.131	0.907		OK
K Field A- 2019	Spherical WLS	0.00809	0.0163	154	67 (H)	35 (L)	48	67	0.1109	-0.091	0.137		IDW
m% Field A- 2019	Gaussian OLS	25.73	49.89	353	66 (H)	62 (H)	6	55	0.2179	6.602	5.508		OK
OM Field A- 2019	All geostatistical models were eliminated											IDW	
pH CaCl2 Field A- 2019	Spherical OLS	0.0682	0.0684	287	50 (M)	27 (L)	47	50	0.1248	-0.184	0.318		OK
pH SMP Field A- 2019	Spherical OLS	0.0253	0.0749	353	75 (H)	47 (M)	38	75	0.2185	-0.319	0.233		OK
V% Field A- 2019	Spherical WLS	46.7	106.7	353	70 (H)	41 (M)	40	70	0.2194	-10.318	9.467		OK
Clay Field A- 2019	Matérn 2 OLS	10.78	5.83	96	35 (L)	29 (L)	17	31	0.0452	-0.719	3.290		OK
Silt Field A- 2019	All geostatistical models were eliminated											IDW	
C Field B- 2015	All geostatistical models were eliminated											IDW	
Ca Field B- 2015	Spherical OLS	0.725	0.425	183	37 (L)	28 (L)	25	37	0.0445	-0.287	0.987		IDW
H+A1 Field B- 2015	All geostatistical models were eliminated											IDW	
P Field B- 2015	All geostatistical models were eliminated											IDW	
pH CaCl2 Field B- 2015	Spherical WLS	0.062	0.017	183	21 (L)	34 (L)	-62	21	0.0673	-0.084	0.271		OK

Table 13 (continued)

Variables/ Fields	Models	C_0	C_1	Ra	%SDI	%ESDI	$\% \gamma(1)$	%SMEI	ISI	$ME \cdot 10^2$	SDME	Semivariogram	Best Interpolator
pH SMP Field B- 2015	Exponential OLS	0.039	0.011	183	22 (L)	35 (L)	-57	21	0.0480	-0.035	0.217		IDW
SB Field B- 2015	Spherical WLS	1.360	0.861	167	39 (L)	32 (L)	18	39	0.0527	-5.516	1.398		IDW
V% Field B- 2015	All geostatistical models were eliminated											IDW	

C_0 nugget effect, C_1 partial sill, Ra range, %SDI spatial dependence index, %ESDI effective spatial dependence index, $(\% \gamma(1))$ first semivariance significance index, %SMEI slope of the model ends index, ISI interpolator selection index, ME mean error, $SDME$ standard deviation of mean error, IDW inverse distance weighting, OK ordinary Kriging, OLS ordinary least squares, WLS weighted least squares

Classification of %SDI and ESDI: very low for %SDI/ESDI < 20%, low for $20 \leq \%SDI/ESDI < 40\%$, medium for $40 \leq \%SDI/ESDI < 60\%$, high for $60 \leq \%SDI/ESDI < 80\%$, and very high for %SDI/ESDI > 80%

Table 14 Comparison of thematic maps created by OK using the semivariogram selected by each of three methods and IDW with its best interpolator

Var.	OK Sem. Meth 1	OK Sem. Meth 2	OK Sem. Meth 3	IDW	Var.	OK Sem. Meth 1	OK Sem. Meth 2	OK Sem. Meth 3	IDW
AI - Field A-2018					C - Field A-2018				
	Gau - WLS	Gau - WLS	Gau - WLS	IDW e5n5		Gau - OLS	Gau - OLS	Gau - OLS	IDW e1n7
	Reference	Reference	Reference	59.72%		Reference	Reference	Reference	2.38%
	0 0.56 1.19 1.82 2.45					173 193 213 234 254			
CRD	Reference	Reference	Reference	59.72%	Reference	Reference	Reference	2.38%	
Ca - Field A-2018					Cu - Field A-2018				
	Mat 1 - OLS	Mat 1 - OLS	Mat 1 - OLS	IDW e4.5n5		Gau - OLS	Gau - OLS	Gau - OLS	IDW e6n7
	Reference	Reference	Reference	8.23%		Reference	Reference	Reference	11.47%
	2.07 2.92 3.77 4.62 5.47					1.86 3.56 5.26 6.96 8.66			
CRD	Reference	Reference	Reference	8.23%	Reference	Reference	Reference	11.47%	
Fe - Field A-2018					H-AI - Field A-2018				
	Exp - WLS	Exp - WLS	Exp - WLS	IDW e2.5n12		Sph - OLS	Sph - OLS	Sph - OLS	IDW e5.5n10
	9.31%	9.31%	9.31%	Reference		7.36%	7.36%	7.36%	Reference
	5.08 11.8 18.6 25.4 32.1					3.97 6.24 8.52 10.79 13.06			
CRD	9.31%	9.31%	9.31%	Reference	7.36%	7.36%	7.36%	Reference	
K (cmol/dm ³) - Field A-2018					Mg - Field A-2018				
	Sph - OLS	Gau - OLS	Exp - OLS	IDW e3.5n7		Mat 2 - OLS	Mat 2 - OLS	Mat 2 - OLS	IDW e5.5n5
	7.47%	18.28%	15.04%	Reference		Reference	Reference	Reference	8.92%
	0.16 0.30 0.43 0.57 0.70					0.69 0.96 1.27 1.55 1.84			
CRD	7.47%	18.28%	15.04%	Reference	Reference	Reference	Reference	8.92%	
Mn - Field A-2018					OM - Field A-2018				
	Mat 2 - OLS	Mat 2 - OLS	Mat 2 - OLS	IDW e1.5n10		Gau - OLS	Gau - OLS	Gau - OLS	IDW e1n7
	9.08%	9.08%	9.08%	Reference		Reference	Reference	Reference	2.38%
	44.8 60.8 76.8 92.8 108.8					29.8 33.3 36.7 40.3 43.8			
CRD	9.08%	9.08%	9.08%	Reference	Reference	Reference	Reference	2.38%	
P - Field A-2018					pH CaCl2 - Field A-2018				
	Exp - WLS	Exp - WLS	Exp - WLS	IDW e3n9		Mat 2 - OLS	Mat 2 - OLS	Mat 2 - OLS	IDW e1n8
	Reference	Reference	Reference	11.27%		1.88%	1.88%	1.88%	Reference
	2.29 3.8 5.3 6.8 8.3					3.70 4.00 4.31 4.61 4.91			
CRD	Reference	Reference	Reference	11.27%	1.88%	1.88%	1.88%	Reference	
pH SMP - Field A-2018					SB - Field A-2018				
	Mat 2 - WLS	Mat 2 - WLS	Mat 2 - WLS	IDW e4.5n9		Exp - OLS	Exp - OLS	Exp - OLS	IDW e3.5n5
	2.43%	2.43%	2.43%	Reference		Reference	Reference	Reference	6.61%
	4.70 5.00 5.30 5.60 5.90					3.29 4.46 5.64 6.81 7.98			
CRD	2.43%	2.43%	2.43%	Reference	Reference	Reference	Reference	6.61%	

Table 14 (continued)

Var.	OK Sem. Meth 1	OK Sem. Meth 2	OK Sem. Meth 3	IDW	Var.	OK Sem. Meth 1	OK Sem. Meth 2	OK Sem. Meth 3	IDW
V% - Field A-2018					Zn - Field A-2018				
	Exp - OLS 2.40%	Exp - OLS 2.40%	Exp - OLS 2.40%	Reference		Gau - OLS 5.84%	Gau - OLS 5.84%	Gau - OLS 5.84%	IDW n2.5n11 Reference
Al (cmol _d dm ⁻³) - Field A-2019					Ca - Field A-2019				
	Sph - WLS 64.00%	Mat 2 - WLS Reference	Mat 2 - WLS Reference	IDW e1n9 30.98%		Mat 2 - WLS 5.39%	Mat 2 - WLS 5.39%	Mat 2 - WLS 5.39%	IDW e1n9 Reference
Cu - Field A-2019					Fe - Field A-2019				
	Mat 0.5 - WLS 5.58%	Mat 0.5 - WLS 5.58%	Mat 0.5 - WLS 5.58%	IDW e4.5n5 Reference		Gau - OLS Reference	Gau - OLS Reference	Gau - OLS Reference	IDW e1.5n6 7.38%
H-Al (cmol _d dm ⁻³) - Field A-2019					K (cmol _d dm ⁻³) - Field A-2019				
	Sph - OLS 6.02%	Exp - OLS 0.69%	Sph - WLS Reference	IDW e6n4 8.77%		Sph - OLS 12.92%	Gau - WLS 7.19%	Sph - WLS 7.68%	IDW e1n10 Reference
Mg - Field A-2019					Mn - Field A-2019				
	Gau - WLS Reference	Gau - WLS Reference	Gau - WLS Reference	IDW e1.5n6 11.28%		Exp - WLS Reference	Exp - WLS Reference	Exp - WLS Reference	IDW e6n9 9.53%
OM (g dm ⁻³) - Field A-2019		No geostatistical model selected	No geostatistical model selected		P - Field A-2019				
	Mat 2 - WLS 7.34%	-	-	IDW e1n4 Reference		Mat 2 - WLS Reference	Mat 2 - WLS Reference	Mat 2 - WLS Reference	IDW e1.5n9 14.43%
pH CaCl2 (Field A-2019)					pH SMP (Field A-2019)				
	Gau - OLS 0.79%	Mat 1 - OLS 0.39%	Sph - OLS Reference	IDW e6n8 3.84%		Sph - WLS 1.21%	Sph - OLS 0.35%	Sph - OLS Reference	IDW e6n4 1.81%

Table 14 (continued)

Var.	OK Sem. Meth 1	OK Sem. Meth 2	OK Sem. Meth 3	IDW	Var.	OK Sem. Meth 1	OK Sem. Meth 2	OK Sem. Meth 3	IDW
SB - Field A-2019					m% - Field A-2019				
	Mat 2 - WLS 4.74%	Mat 2 - WLS 4.74%	Mat 2 - WLS 4.74%	IDW e1n10 Reference		Sph - WLS 62.67%	Gau - OLS Reference	Gau - OLS Reference	IDW e1n9 25.67%
V% - Field A-2019					Zn - Field A-2019				
	Sph - WLS 5.71%	Sph - OLS 0.15%	Sph - WLS Reference	IDW e1n9 1.99%		Gau - OLS Reference	Gau - OLS Reference	Gau - OLS Reference	IDW e2.5n5 12.16%
Clay (%) - Field A-2019					Sand- Field A-2019				
	Gau - OLS 0.09%	Mat 2 - OLS Reference	Mat 2 - OLS Reference	IDW e1n5 1.16%		Gau - WLS Reference	Gau - WLS Reference	Gau - WLS Reference	IDW e1n4 6.24%
Silt (%) - Field A-2019		No geostatistical model selected	No geostatistical model selected		Al - Field B-2015				
	Gau - OLS 2.68%	-	-	IDW e1n10 Reference		Gau - WLS Reference	Gau - WLS Reference	Gau - WLS Reference	IDW e2.5n4 62.34%
C - Field B-2015		No geostatistical model selected	No geostatistical model selected		Ca - Field B-2015				
	Gau - WLS 1.83%	-	-	IDW e1n7 Reference		Exp - OLS 7.39%	Sph - OLS 4.54%	Sph - OLS 4.54%	IDW e1.5n7 Reference
Cu - Field B-2015					Fe - Field B-2015				
	Sph - OLS Reference	Sph - OLS Reference	Sph - OLS Reference	IDW e3n10 4.58%		Mat 1.5 - WLS 4.04%	Mat 1.5 - WLS 4.04%	Mat 1.5 - WLS 4.04%	IDW e1n5 Reference
H+Al (cmol/dm ³) - Field B-2015		No geostatistical model selected	No geostatistical model selected		K - Field B-2015				
	Mat 1.5 - WLS 7.65%	-	-	IDW e1.5n7 Reference		Sph - WLS Reference	Sph - WLS Reference	Sph - WLS Reference	IDW e2n6 9.05%

Table 14 (continued)

Var.	OK Sem. Meth 1	OK Sem. Meth 2	OK Sem. Meth 3	IDW	Var.	OK Sem. Meth 1	OK Sem. Meth 2	OK Sem. Meth 3	IDW
Mg - Field B-2015					Mn - Field B-2015				
	Gau - OLS 4.89%	Gau - OLS 4.89%	Gau - OLS 4.89%	IDW e1.5n6 Reference		Mat 2 - OLS 3.04%	Mat 2 - OLS 3.04%	Mat 2 - OLS 3.04%	IDW e1n4 Reference
CRD					CRD				
P (mg/dm ³) - Field B-2015		No geostatistical model selected	No geostatistical model selected		pH CaCl2 - Field B-2015				
	Exp - WLS 23.60%	-	-	IDW e2n7 Reference		Gau - WLS 0.28%	Gau - OLS 0.23%	Sph - WLS Reference	IDW e1n7 1.47%
CRD					CRD				
pH SMP - Field B-2015		No geostatistical model selected			SB - Field B-2015				
	Mat 2 - WLS 1.65%	-	Exp - OLS 1.30%	IDW e1.5n7 Reference		Gau - OLS 6.06%	Sph - WLS 2.73%	Sph - WLS 2.73%	IDW e1n7 Reference
CRD					CRD				
V% - Field B-2015		No geostatistical model selected	No geostatistical model selected		Zn - Field B-2015				
	Exp - WLS 2.47%	-	-	IDW e1n7 Reference		Exp - OLS 8.59%	Exp - OLS 8.59%	Exp - OLS 8.59%	IDW e2.5n12 Reference
CRD					CRD				

Var. variable, *OK Sem. Methods (1, 2 or 3)* means Ordinary Kriging using the semivariogram selected by method 1 (Only the best ISI), 2 (The three criteria are applied after geostatistics analysis), or 3 (The three criteria are applied during geostatistics analysis), *Sph* spherical, *Exp* exponential, *Gau* gaussian, *Mat* Matérn, *IDW e3.5n7* means inverse distance weighting with exponent 3.5 and 7 neighbors, *OLS* ordinary least squares, *WLS* weighted least squares, *CRD* coefficient of relative deviation, *Al* aluminum, *C* carbon, *Ca* calcium, *Cu* copper, *Fe* iron, *H + Al* potential acidity, *K* potassium, *Mg* magnesium, *Mn* manganese, *OM* organic matter, *P* phosphorus, *pH* the potential of hydrogen, *pH SMP* pH of buffer solution Shoemaker–McLean–Pratt, *SB* the sum of basis, *V%* base saturation, *Zn* zinc

Acknowledgements The authors would like to thank the Western Paraná State University (UNIOESTE), the Federal University of Technology of Paraná (UTFPR), the Coordination for the Upgrading of Higher Education Personnel (CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), the National Council for Scientific and Technological Development (CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico), the Itaipu Technological Park Foundation (FPTI, Fundação Parque Tecnológico Itaipu), and the Ministry of Agriculture, Livestock and Food Supply (MAPA, Ministério da Agricultura, Pecuária e Abastecimento) for funding this project.

Funding Funding was provided by CNPq, CAPES, and FPTI with studentships, UNIOESTE and UTFPR with studentships, analysis payment, and equipment purchase.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

- Aikes Junior, J., Souza, E. G., Bazzi, C. L., & Sobjak, R. (2021). *Thematic maps and management zones for precision agriculture*. Poncã.
- Amaral, L. R., & Justina, D. D. D. (2019). Spatial dependence degree and sampling neighborhood influence on interpolation process for fertilizer prescription maps. *Engenharia Agrícola*, 39, 85–95. <https://doi.org/10.1590/1809-4430-Eng.Agric.v39nep85-95/2019>
- Anselin, L. (1995). Local indicators of spatial association-LISA. *Geographical Analysis*, 27(2), 93–115.
- Bazzi, C. L., de Souza, E. G., & Betzek, N. M. (2015). *SDUM: Software para definição de unidades de manejo: teoria e prática*. PGEAGRI.
- Betzek, N. M., Souza, E. G., Bazzi, C. L., Schenatto, K., Gavioli, A., & Magalhães, P. S. G. (2019). Computational routines for the automatic selection of the best parameters used by interpolation methods to create thematic maps. *Computers and Electronics in Agriculture*, 157, 49–62. <https://doi.org/10.1016/j.compag.2018.12.004>.
- Bier, V. A., & Souza, E. G. (2017). Interpolation selection index for delineation of thematic maps. *Computers and Electronics in Agriculture*, 136(1), 202–209. <https://doi.org/10.1016/j.compag.2017.03.008>.
- Biondi, F., Myers, D. E., & Avery, C. C. (1994). Geostatistically modeling stem size and increment in an old-growth forest. *Canadian Journal of Forest Research*, 24(7), 1354–1368. <https://doi.org/10.1139/x94-176>.
- Borges, L. G., Bazzi, C. L., Souza, E. G., Magalhães, P. S. G., & Michelon, G. K. (2020). Web software to create thematic maps for precision agriculture. *Pesquisa Agropecuária Brasileira*. <https://doi.org/10.1590/S1678-3921.pab2020.v55.00735>.
- Cambardella, C. A., Moorman, T. B., Novak, J. M., Parkin, T. B., Karlen, D. L., Turv, R. F., et al. (1994). Field-scale variability of soil properties in central Iowa soil. *Soil Science Society of America Journal*, 58(5), 1501–1511. <https://doi.org/10.2136/sssaj1994.03615995005800050033x>.
- Coelho, E. C., Souza, E. G., Uribe-Opazo, M. A., & Pinheiro Neto, R. (2009). Influência da densidade amostral e do tipo de interpolador na elaboração de mapas temáticos (the influence of sample density and interpolation type on the elaboration of thematic maps). *Acta Scientiarum Agronomy*, 31(1), 165–174. <https://doi.org/10.4025/actasciagron.v31i1.6645>.
- Córdoba, M. A., Bruno, C. I., Costa, J. L., Peralta, N. R., & Balzarini, M. G. (2016). Protocol for multivariate homogeneous zone delineation in precision agriculture. *Biosystems Engineering*, 143, 95–107. <https://doi.org/10.1016/j.biosystemseng.2015.12.008>.
- Clark, I. (1979). *Practical geostatistics*. Applied Science Publishers.
- Cressie, N. A. C. (1993). *Statistics for spatial data*. Wiley.
- Dall'agnol, R. W., Michelon, G. K., Bazzi, C. L., Magalhães, P. S. G., Souza, E. G., Betzek, N. M., et al. (2020). Web applications for spatial analyses and thematic map generation. *Computers and Electronics in Agriculture*, 172, 105374. <https://doi.org/10.1016/j.compag.2020.105374>.
- Diggle, P. J., & Ribeiro Jr., P. J. (Eds.). (2007). *Model-based geostatistics*. Springer.
- Doerge, T. A. (2000). Management zone concepts. *Site-specific management guidelines*. Potash and Phosphate Institute. University South Dakota, Brookings. Retrieved July 28, 2021, from [http://www.ipni.net/publication/ssmg.nsf/0/C0D052F04A53E0BF852579E500761AE3/\\$FILE/SSMG-02.pdf](http://www.ipni.net/publication/ssmg.nsf/0/C0D052F04A53E0BF852579E500761AE3/$FILE/SSMG-02.pdf)
- Faraco, M. A., Uribe-Opazo, M. A., Silva, E. A. A., Johann, J. A., & Borssoi, J. (2008). Selection criteria of spatial variability models used in thematic maps of soil physical attributes and soybean yield. *Revista Brasileira de Ciência do Solo*, 32(2), 463–476. <https://doi.org/10.1590/S0100-06832008000200001>.
- Ferguson, R. B., & Hergert, G. W. (2009). *Soil sampling for precision agriculture*. University of Nebraska Extension. EC154.
- Fraser, B. T., & Congalton, R. G. (2019). Evaluating the effectiveness of unmanned Aerial Systems (UAS) for collecting thematic Map Accuracy Assessment Reference Data in New England forests. *Forests*, 10(1), 1–17. <https://doi.org/10.3390/f10010024>.
- Gojiya, K. M., Gontia, N. K., & Patel, K. C. (2018). Generation of thematic maps of a forest watershed using Remote sensing and GIS. *International Journal of Current Microbiology and Applied Sciences*, 7(12), 2952–2962. <https://doi.org/10.20546/ijemas.2018.712.337>.
- Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. Applied Geostatistics Series. Oxford University Press.
- Han, C., Wang, J., Zheng, M., Wang, E., Xia, J., Li, G., et al. (2016). New variogram modeling method using MGGP and SVR. *Earth Science Informatics*, 9, 197–213. <https://doi.org/10.1007/s12145-016-0251-9>.
- Isaaks, E. H., & Srivastava, R. M. (1989). *Applied geostatistics*. Oxford University Press.
- Journell, A. G., & Huibregts, C. J. (1978). *Mining geostatistics*. Academic Press.

- Konopatzki, M. R., Souza, E. G., Nóbrega, L. H., Uribe-Opazo, M. A., & Suszek, G. (2012). Spatial variability of yield and other parameters associated with pear trees. *Engenharia Agrícola*, 32(2), 381–392. <https://doi.org/10.1590/S0100-69162012000200018>.
- Lark, R. M. (2000). Estimating variograms of soil properties by the method-of-moments and maximum likelihood. *European Journal of Soil Science*, 51, 717–728. <https://doi.org/10.1046/j.1365-2389.2000.00345.x>.
- Li, Z., Zhang, X., Clarke, K. C., Liu, G., & Zhu, R. (2018). An automatic variogram modeling method with high reliability fitness and estimates. *Computers & Geosciences*, 120, 48–59. <https://doi.org/10.1016/j.cageo.2018.07.011>.
- Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58(8), 1246–1266. <https://doi.org/10.2113/gsecongeo.58.8.1246>.
- Michelon, G. K., Bazzi, C. L., Upadhyaya, S., Souza, E. G., Magalhães, P. S. G., Borges, L. F. (2019). Software AgDataBox-Map to precision agriculture management. *SoftwareX*, 10, 100320. <https://doi.org/10.1016/j.softx.2019.100320>
- Mueller, T. G., Pusuluri, N. B., Mathias, K. K., Cornelius, P. L., Barnhisel, R. I., & Shearer, S. A. (2004). Map quality for ordinary kriging and inverse distance weighted interpolation. *Soil Science Society of America Journal*, 68(6), 2042–2047. <https://doi.org/10.2136/sssaj2004.2042>.
- Oliver, M. A., & Webster, R. (2015). *Basic steps in geostatistics: The variogram and kriging*. Springer.
- Pimentel-Gomes, F. (2009). *Curso de estatística experimental (experimental statistics course)* (p. 15). FEALQ.
- Reza, S. K., Sarkar, D., Daruah, U., & Das, T. H. (2010). Evaluation and comparison of ordinary kriging and inverse distance weighting methods for prediction of spatial variability of some chemical parameters of Dhalai district. *Tripura Agropedology*, 20(1), 38–48.
- Ribeiro Jr., P. J., & Diggle, P. J. (2001). geoR: A package for geostatistical analysis. *R-NEWS*, 1(2), 14–18.
- Rodrigues, M. S., Alves, D. C., de Souza, V. C., de Melo, A. C., & do Nascimento Lima, A. M. (2018). Spatial interpolation techniques for site-specific irrigation management in a mango orchard. *Comunicata Scientiae*, 9(1), 93–101. <https://doi.org/10.14295/cs.v9i1.2645>
- Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM national conference (ACM '68)*. Association for Computing Machinery (pp. 517–524). <https://doi.org/10.1145/800186.810616>
- Souza, E. G., Bazzi, C. L., Khosla, R., Uribe-Opazo, M. A., & Reich, R. M. (2016). Interpolation type and data computation of crop yield maps is important for precision crop production. *Journal of Plant Nutrition*, 39(4), 531–538. <https://doi.org/10.1080/01904167.2015.1124893>.
- Uribe-Opazo, M. A., Borssoi, J. A., & Galea, M. (2012). Influence diagnostics in Gaussian spatial linear models. *Journal of Applied Statistics*, 39(3), 615–630. <https://doi.org/10.1080/02664763.2011.607802>.
- Webster, R., & Oliver, M. A. (1990). *Statistical methods in soil and land resource survey*. Oxford University Press.
- Whelan, B., & Taylor, J. (2013). *Precision agriculture for grain production systems*. CSIRO. <https://doi.org/10.1071/9780643107489>.
- Wollenhaupt, N. C., Wolkowski, R. P., & Clayton, M. K. (1994). Mapping soil test phosphorus and potassium for variable-rate fertilizer application. *Journal of Production Agriculture*, 7(4), 441–448. <https://doi.org/10.2134/jpa1994.0441>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.