Australian Journal of Crop Science AJCS

AJCS 17(1):20-27 (2023) ISSN:1835-2707
doi: 10.21475/ajcs.23.17.01.p3714

Bivariate spatial correlation between soil attributes and soybean productivity in an agricultural area with Dystroferric Red Latosol

Letícia Ellen Dal' Canton*, Tamara Cantú Maltauro, Luciana Pagliosa Carvalho Guedes, Miguel Angel Uribe-Opazo, Elizabeth Giron Cima

Western Paraná State University (UNIOESTE), Cascavel, Paraná, Brazil

***Corresponding author: leticiacanton@hotmail.com**

Abstract

In agricultural soils with low cation exchange capacity, it is essential to analyze the bivariate spatial correlation of soybean productivity and organic matter with the soil chemical attributes. Using bivariate spatial correlation makes it possible to identify patterns and behaviors that suggest a spatial association between two soil attributes, thus enabling better soil management and more efficient use of resources. The main objective of this study was to analyze bivariate spatial correlation considering variables with different spatial dependence structures. The bivariate Lee index was also calculated for this purpose. To model and describe the spatial pattern of two spatially correlated variables, the Bivariate Gaussian Common Component Model was used. In addition to calculating the bivariate spatial correlation of soil chemical attributes with soybean productivity and organic matter, the Lee index was also calculated for pairs of simulated variables with different weight matrices and geographic distance functions. It was observed that the greater the common practical range, the higher the Lee index value, indicating a higher bivariate spatial correlation. Furthermore, shorter distances between neighboring point pairs caused higher Lee index values. The distance function to calculate the distance between the point pairs was more relevant than the weight matrix in estimating the spatial dependence radius and the Lee index value. Soybean productivity showed a direct spatial correlation with the sum of bases, as well as with the calcium and magnesium contents. Organic matter had a direct spatial correlation with the sum of bases and an inverse one with the phosphorus content.

Keywords: BGCCM; correlogram; cross-semivariogram; geostatistics; Lee's index.
Abbreviations: BGCCM_Bivariate Gaussian Common Component Model; Ca_calcium; CEC_cation exchange capacity; ED_Euclidean distance; IED_inverse of the Euclidean distance; Mg_magnesium; OM_organic matter; P_phosphorus; Prod_soybean productivity; SB_ sum of bases.

Introduction

The current challenge faced by the agricultural sector is to maintain an increasing pace of production growth, preferably without expanding the planted area, which implies a rise in productivity. One of the possibilities for increasing agricultural productivity is using technological advances in order to improve knowledge on the nutritional requirements of each crop, providing proper use of inputs in the agricultural property (Deiss et al., 2020).

A number of studies show the importance of investigating the existence of a relationship between the amount of soil nutrients and soybean productivity, aiming to establish better management of the production of this agricultural commodity (Malvezi et al., 2019; Deus et al., 2020). This is because excess or lack of macro- and micro-nutrients in the soil can alter the growth and development phases of the plant, thus affecting the grains and, consequently, soybean productivity (Mengel and Kirkby, 2001; Malavolta, 2006; Barbosa et al., 2016). In agricultural soils with low cation exchange capacity (CEC), organic matter plays a relevant role since, in adequate amounts, it improves the physical and chemical conditions of the soil, in addition to assisting in availability of nutrients to the plants, contributing to increased fertility (Cunha et al., 2015; Siqueira-Neto et al., 2021).

Thus, it becomes indispensable to evaluate soybean productivity and organic matter in relation to the soil chemical attributes, using spatial statistics that simultaneously considers the information related to the value and geographical location of the variables. Using bivariate spatial correlation, it is possible to identify patterns and behaviors that suggest a spatial association between two soil attributes, thus enabling decision-making (Cima et al., 2018).

Among the studies published in the literature that use bivariate spatial correlation, the use of two main expressions is observed (Matkan et al., 2013; Cima et al., 2018). One of them is described by Lee (2001) and by Anselin et al. (2002) in a very similar way, while the other is presented by Almeida (2013). In Almeida's proposal (2013), only the variance of one of the variables is considered in the denominator of the ratio that expresses this measure, so that one attribute must be a covariate of the other. Thus, changing the order in which the variables are selected directly influences the value of the bivariate spatial correlation. The other proposal, obtained by Lee (2001) and by Anselin et al. (2002), uses the variance of both variables, so that the order in which the attributes are taken does not affect the value of the spatial correlation between them.

However, when compared to the proposals by Anselin et al. (2002) and Almeida (2013), which are called bivariate Moran index, the methodology developed by Lee (2001) is little explored in the literature. Performing a systematic mapping of the literature in the main databases (ScienceDirect, Scopus, Web of Science, and Wiley) over more than a decade and excluding duplicates, a total of 229 scientific papers were obtained that used the bivariate Moran index, against only 25 that employed the Lee index. Only one study was found with the Lee index in the agricultural context (Gaso et al., 2021), which only used this metric to validate a soybean yield prediction model. Thus, it was not an exploratory study on such index.

Therefore, the objectives of this paper were as follows: a) to analyze tests considering pairs of simulated variables with different values for the common practical range and to explore the Lee index to calculate bivariate spatial correlation; b) to calculate the Lee index exploring different spatial weighting matrices and metrics for calculation of the distance between point pairs; c) to calculate the bivariate spatial correlation using the Lee index for different attribute pairs: organic matter (OM) and phosphorus (P), OM and sum of bases (SB), soybean productivity (Prod) and calcium (Ca), Prod and magnesium (Mg), and Prod and SB. For data adjustment, the Bivariate Gaussian Common Component Model (BGCCM) and the cross-semivariogram were used, which their results had compared.

Results

Simulated data using weight matrix W and the Euclidean distance function

The BGCCM spatial model presented the highest estimated mean values for the common practical range (a_0), as well as the smallest dispersion (Standard Deviation-SD) of the estimated values of this parameter, when compared to the estimates obtained for the cross-semivariogram (a) (Table 1). It was observed that, as the common practical range simulated by the exponential model (a_0=3φ_0) increased (T1: 375, T2: 525, and T3: 825 m), the mean values of the practical range estimated by the cross-semivariogram (a) and by BGCCM (a_0) also increased and were very close to the simulated practical ranges (Table 1), indicating that both estimates for the spatial dependence radius were satisfactory. It was also verified that, in the last test (T3), the mean values estimated for the practical ranges of the cross-semivariogram and the bivariate model were very similar to the cutoff point of the simulated area (875 m), given the purpose of this test.

For the nine pairs of simulated variables, considering all tests, the Lee index values varied between 0.19 and 0.23 (T1; Figure 1S-a), between 0.26 and 0.33 (T2; Figure 1S-a,b), and between 0.37and 0.46 (T3; Figure 1S-c). Thus, there was a positive bivariate spatial correlation in neighboring regions (Lee, 2001). In addition to that, as the simulated ranges increased, the values on which the Lee index ranged also increased.

For T1 and T2, the spatial dependence radius varied between 305 m and 380 m (Figure 1S-a) and between 505 m and 780 m (Figure 1S-b), approaching the mean values estimated for the practical ranges of the cross-semivariogram (375.56 m and 510.86 m; Table 1) and of the bivariate model (385.41 m and 539.49 m; Table 1) in these tests. For T3, the spatial dependence radius was from 1,105 m (Figure 1S-c), therefore being higher than the cutoff point of the simulated area. Consequently, there was overestimation in the spatial dependence radiuses observed in the correlograms, in relation to the estimated mean values for the practical ranges of the cross-semivariogram and of the bivariate model (Table 1).

Lee's index in data simulated with different spatial weighting matrices and distance metrics

The correlograms were statistically significant for all the pairs of simulated variables (Figure 1S-d and Figure 2S), regardless of the spatial weighting matrix and of the metric used. There was a positive spatial bivariate correlation in neighboring locations for every pair of simulated variables.

Considering spatial weighting matrix $\boldsymbol{W}$ and the IED, the nine pairs of variables presented simulated Lee index values varying between 0.20 and 0.25, as well as a spatial dependence radius between 330 m and 505 m (Figure 1S-d). By comparing these results with test T1 (Table 1; Figure 1S-a), in which the simulations were carried out changing only the metric (ED), it was verified that the values of the spatial dependence radius were smaller using the ED (from 305 m to 380 m), although the Lee index values were similar (from 0.19 to 0.23).

Using weight matrix $\boldsymbol{C}$ and the ED, the Lee index values varied from 0.13 to 0.23, with a spatial dependence radius between 230 m and 430 m (Figure 2S-a). When compared to T1, in which the altered weight matrix (for $\boldsymbol{W}$), a similarity was verified in the maximum value obtained for the spatial dependence radius (380 m) (Figure 1S-a and Figure 2S-a). However, most of the Lee index maximum values were lower and with greater amplitude in the range of their values.

Finally, with $\boldsymbol{C}$ and the IED, Lee index values between 0.18 and 0.27 were obtained, as well as a spatial dependence radius between 255 m and 530 m (Figure 2S-b). Comparing weight matrices $\boldsymbol{C}$ and $\boldsymbol{W}$ and maintaining the IED as the metric, it was verified that the Lee index maximum values were similar (Figure 1S-d and Figure 2S-b). However, using weight matrix $\boldsymbol{C}$, the spatial dependence radius presented greater amplitude ($\boldsymbol{W}$: between 330 m and 505 m).

To perform the Lee index method for the 63 distance classes between neighbors of the simulated data, considering 99 permutations per class, a machine with an Intel® Core™ i5-8265U CPU @ 1.60GHz, 64-bit OS and 8 GB of installed physical memory (RAM) was used. Computational time was a relevant factor. This is because, for the ED in both weight matrices, computational time was similar, between 120 and 140 seconds. Using the IED, although weight matrices $\boldsymbol{C}$ and $\boldsymbol{W}$ were similar (760 and 810 seconds), computational time was nearly seven times higher.

Methodology application to the soil attributes and soybean productivity

Considering the soil chemical attributes for Paraná, on average, the mean levels of Ca (4.03 $cmol_c\ dm^{-3}$), P (19.53 $mg\ dm^{-3}$), Mg (1.73 $cmol_c\ dm^{-3}$), OM (24.93 $g\ dm^{-3}$), and SB (6.05 $cmol_c\ dm^{-3}$) are considered high or very high (SBCS-NEPAR, 2017); while the mean of Prod (3.12 $t\ ha^{-1}$) in the 2016-2017 harvest year was below the mean values for Paraná (3.72 $t\ ha^{-1}$) and Brazil (3.36 $t\ ha^{-1}$) (CONAB, 2017). Through the coefficient of variation (CV), the exploratory analysis showed that dispersion of the soil attributes, as well as of productivity, varies between medium (10%≤CV≤20%), high (20%<CV≤30%) and heterogeneous (CV>30 %) (Pimentel Gomes and Garcia, 2002) (Table 2).

In the cross-semivariogram, the Prod covariates were better adjusted by the Gaussian and exponential models (Table 3); while the OM covariates had their data set adjusted by the Matérn model with k=2.5 (Table 3). All attribute pairs showed strong spatial dependence (Cambardella et al., 1994). It can also be observed that the sill (φ_1+φ_2) of the OM and P pair is negative (Table 3). As in a cross-semivariogram, the sill approaches the value of the covariance between both variables; this indicates that the correlation between OM and P is inverse (Righetto, 2012).

For the BGCCM, the OM covariates were better adjusted by the Gaussian and exponential models; most of the Prod covariates were better adjusted by means of the Matérn model with k=2.5 (Table 4). Among the adjusted models, Matérn with k=2.5 is the one that had the highest constant multiplied to the range function, approximately $5.92\varphi_t$, $t = 0,1,2$ (Diggle and Ribeiro Jr., 2007). Consequently, the spatial dependence radiuses were high (above 1,000 m; Table 4). The range functions both of the common component (φ_0) and for each variable (φ_1, φ_2) were similar to each other in the attribute pairs (Table 4). The dispersion parameters ($\sigma_{01}, \sigma_1, \sigma_{02}, \sigma_2$) were low for most of the attribute pairs (Table 4).

A higher Lee index value was obtained for all attribute pairs in a module, in the shortest distance between neighbors (140 m) (Figure 2). After this distance, the attribute pairs that had positive bivariate spatial correlations showed a sharp drop for the Lee index, presenting a downward trend for the correlogram, which is gradually attenuated with increases in the distance, until stabilizing close to zero. The OM-P correlogram presented a negative Lee index, which indicates a negative bivariate spatial correlation (Figure 2-a). Thus, the correlogram's behavior was inverse, increasing with increases in the distances between neighbors, until stabilizing close to zero.

Considering the spatial dependence radius identified in the correlograms (Figure 2), the attribute pairs with higher Lee index values, in a module, presented the largest radiuses; where the highest values of the spatial dependence radiuses and Lee index by pair were as follows: OM-SB (465 m; L: 0.36), Prod-SB (440 m; L: 0.35), and Prod-Mg (415 m; L: 0.34). On the other hand, the Prod-Ca and OM-P pairs had the same spatial dependence radius (315 m) and presented lower values for the Lee index: 0.26 and -0.27, respectively.

Discussion

Considering the simulated data, in all the tests conducted with **W** and the ED, the Lee index presented a downward trend due to the distance between the neighbors, approaching zero as the distance increases (Figure 1S-a, b, c). Thus, the highest Lee index values were obtained with the smallest distances between the neighbors. The same result was observed in the papers by Liu et al. (2013) and by Costa and Scalon (2015), which used the univariate Moran index to analyze the autocorrelation of different attributes.

The reduction was more pronounced in the first class of distances (between 200 m and 510 m) and attenuated in the second class (between 510 m and 820 m), from which it begins to stabilize and approach to zero. In a correlogram analysis, the decrease over increasing distances between the neighbors, until stabilization of its curve, indicates the stationary character of the stochastic process. In addition to that, the increase in the simulated practical range implied slower decreases in the correlogram; this can be observed mainly by comparing T1 and T2 (Figure 1S-a, b) to T3 (Figure 1S-c). The highest Lee index values (between 0.37 and 0.46) and the highest spatial dependence radiuses (above 1,105 m) were obtained in T3.

As the simulated practical range increased, there was also a larger difference between the values of the spatial dependence radius in the correlogram (Figure 1S-a, b, c) and the mean values of the spatial dependence radius estimated by the cross-semivariogram (Table 1). Such difference was smaller for T1 and T2, and higher for T3. This can be explained by the difference between the methodologies of the correlogram and of the cross-semivariogram (Liu et al., 2013).

Comparing the IED and the ED, maintaining $\boldsymbol{C}$ fixed, larger spatial dependence radiuses were obtained using the IED (Figure 2S). However, in both cases the difference between the smallest and the largest maximum Lee index value was approximately 0.10 and the values of the bivariate spatial correlation were higher with the IED (Figure 2S). The conclusions regarding the spatial dependence radius and the bivariate spatial correlation values considering the IED were also obtained by fixing $\boldsymbol{W}$. In general, the correlogram's behavior considering the IED was different from the ED in both weight matrices used since, in larger distances, the Lee index value did not approach zero.

In the data corresponding to the soil attributes and soybean productivity, as the values of the spatial dependence radiuses (a and a_0) were influenced by choice of the model, the comparisons were made in relation to the range functions, contrasting those obtained in the cross-semivariogram (φ_3) with those estimated by the bivariate model (φ_0). The range function values were higher for the BGCCM for all the attribute pairs (Tables 3 and 4). This fact was also observed in the simulation studies conducted in this paper, as well as by Cantu (2015), considering other pairs of soil attributes analyzed in the same experimental area in the 2010-2011 and 2013-2014 harvest years.

Regarding the correlogram's behavior being decreasing and tending to zero with an increase in the distance between neighbors, this has also been observed in the tests simulated in this study (Figure 1S-a, b, c). However, unlike the simulation, the pairs formed by Prod and OM with their covariates presented correlograms that remained overlapped to the upper limit of the envelope graph until completely entering in the distance identified as spatial dependence radius.

The Lee index values showed a negative bivariate spatial correlation between OM and P and a positive one for the other pairs (Figure 2), corroborating with what the estimated values of the sill indicated in the cross-semivariogram, which was only negative for this pair (Table 3). The bivariate spatial correlation between all pairs indicated that the closer the neighbors (smaller distance), the stronger the spatial correlation bivariate (positive or negative) (Dalchiavon et al., 2017).

The positive spatial correlation between OM and SB is expected since, in tropical soils, the OM content is critical to raising the soil's CEC; therefore, soils with higher CEC tend to have higher SB values, that is, greater retention of cations in the topsoil (Ramos et al., 2018). SB represents the sum of exchangeable cations (Ca^{2+}, Mg^{2+}, and K^+) in the soil; consequently, the positive spatial correlation between Prod

Table 1. Estimated values for the practical range (a) of the cross-semivariogram and for the common practical range (a_0) of the BGCCM.

	Cross-semivariogram			BGCCM		
	T1	T2	T3	T1	T2	T3
S1	378.75	577.32	856.83	391.32	533.55	851.70
S2	372.81	485.70	804.06	385.56	541.80	845.13
S3	332.28	571.98	816.96	378.15	532.38	842.82
S4	383.34	392.19	740.40	400.80	548.22	853.14
S5	413.04	537.27	812.79	389.10	528.93	845.34
S6	435.99	519.78	834.15	381.33	570.66	827.94
S7	434.31	522.27	905.10	378.36	535.80	842.13
S8	387.99	590.34	770.07	380.07	532.26	836.10
S9	334.86	400.86	755.52	383.97	531.78	846.24
$\bar{\mathbf{X}}$	375.56	510.86	810.65	385.41	539.49	843.39
SD	60.26	72.62	51.62	7.37	13.11	7.69

T1, T2, T3: tests 1 to 3; S1, . . . , S9: the nine pairs of simulated variables; SD: standard deviation; $\bar{X}$: mean. The practical ranges are expressed in meters.

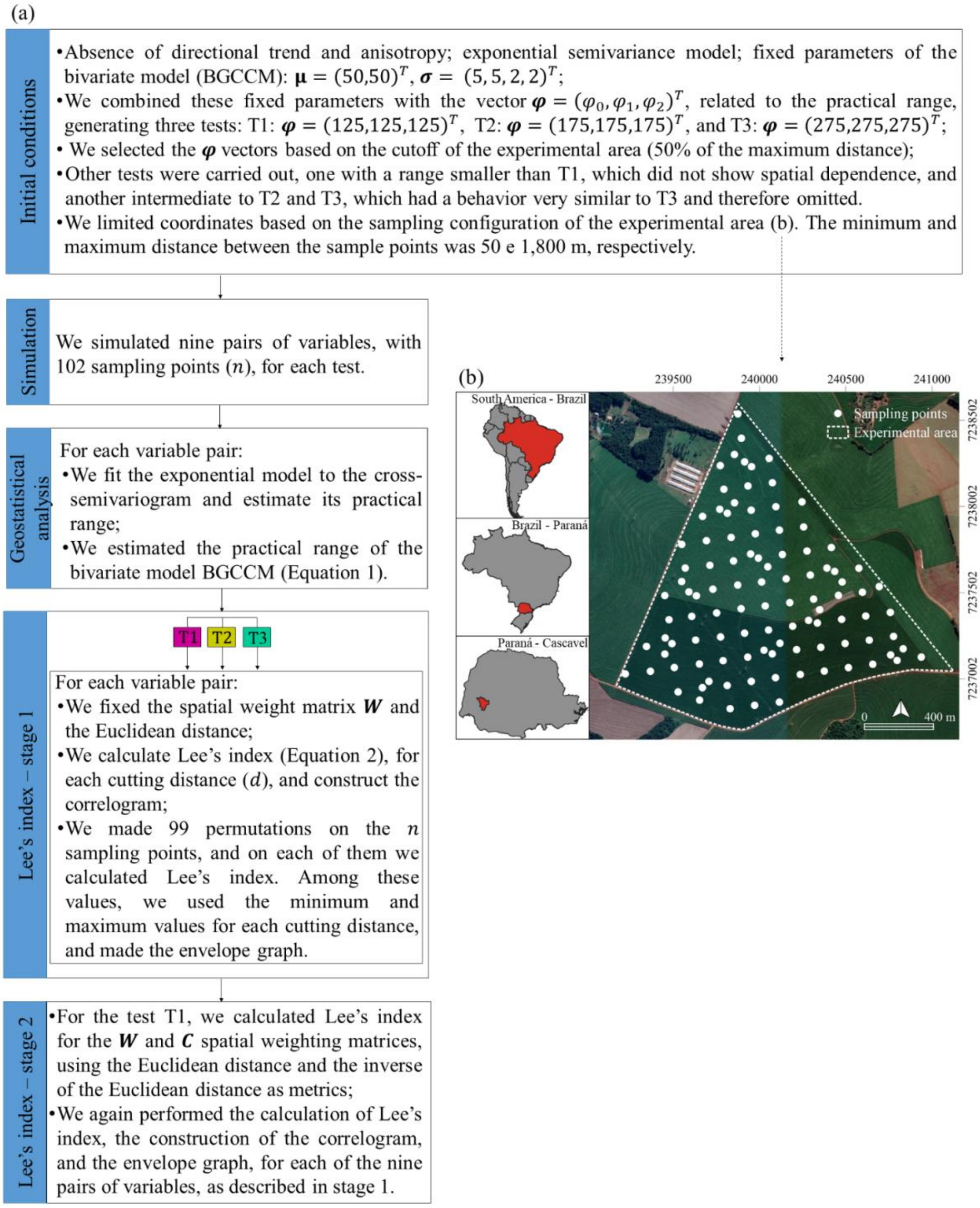

Fig 1. (a) Methodological scheme used in the simulation studies, (b) Experimental area with the locations of the sampling points in UTM coordinates. $\boldsymbol{\mu}$: vector of means; $\boldsymbol{\sigma}$: dispersion vector; $\boldsymbol{\varphi}$: range vector; $T1, T2, T3$: tests 1 to 3; $\boldsymbol{W}$ and $\boldsymbol{C}$: spatial weighting matrices.

Table 2. Descriptive statistics for the attributes: calcium (Ca, $cmol_c\ dm^{-3}$), phosphorus (P, $mg\ dm^{-3}$), magnesium (Mg, $cmol_c\ dm^{-3}$), organic matter (OM, $g\ dm^{-3}$), soybean productivity (Prod, $t\ ha^{-1}$), and sum of bases (SB, $cmol_c\ dm^{-3}$).

Attributes	Minimum	Maximum	Mean	SD	CV
Ca	1.40	6.00	4.03	0.85	21.27
P	4.62	56.48	19.53	10.69	54.74
Mg	0.40	4.20	1.73	0.73	42.09
OM	13.40	89.80	42.14	10.51	24.93
Prod	1.51	4.20	3.12	0.54	17.33
SB	2.55	9.65	6.05	1.38	22.82

SD: Standard Deviation; CV=$100\frac{SD}{Mean}$: coefficient of variation (%).

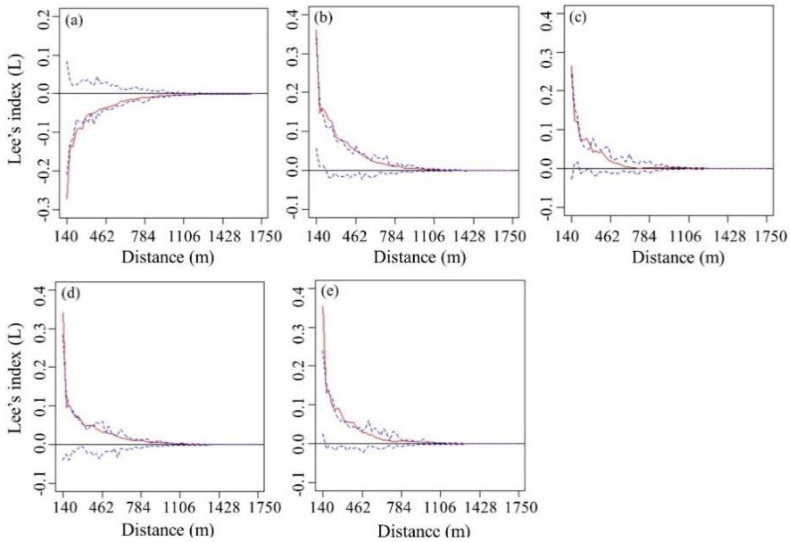

Fig 2. Correlogram (red) and envelope graph (blue) for the pairs of attributes (a) OM-P, (b) OM-SB, (c) Prod-Ca, (d) Prod-Mg, and (e) Prod-SB, considering weight matrix $\boldsymbol{W}$ and Euclidean distance.

Table 3. Estimated values for the parameters of the cross-semivariogram of the best geostatistical model for the attributes: calcium (Ca, $cmol_c\ dm^{-3}$), phosphorus (P, $mg\ dm^{-3}$), magnesium (Mg, $cmol_c\ dm^{-3}$), organic matter (OM, $g\ dm^{-3}$), soybean productivity (Prod, $t\ ha^{-1}$), and sum of bases (SB, $cmol_c\ dm^{-3}$).

Attributes	Geostatistical models	$\hat{\varphi}_1$	$\hat{\varphi}_2$	$\hat{\varphi}_3$	$\hat{a}$	$\widehat{RNE}$
OM and P	Matérn k=2.5	-0.0041	-0.0337	178.43	1056.30	10.84
OM and SB	Matérn k=2.5	0.1819	5.1839	61.19	362.24	3.38
Prod and Ca	Gaussian	0.0125	0.1454	159.85	276.86	7.90
Prod and Mg	Exponential	-0.0576	0.1857	161.62	484.86	23.67
Prod and SB	Exponential	-0.1616	0.5268	106.55	319.65	23.47

$\hat{\varphi}_1$, $\hat{\varphi}_2$, $\hat{\varphi}_3$, $\hat{a}$: the estimated values of the nugget effect, partial sill, range function, and practical range (meters) parameters; RNE = $100\frac{\varphi_1}{\varphi_1+\varphi_2}$: Relative Nugget Effect (%).

Table 4. Estimated values for the parameters of the BGCCM of the best geostatistical model for the attributes: calcium (Ca, $cmol_c\ dm^{-3}$), phosphorus (P, $mg\ dm^{-3}$), magnesium (Mg, $cmol_c\ dm^{-3}$), organic matter (OM, $g\ dm^{-3}$), soybean productivity (Prod, $t\ ha^{-1}$), and sum of bases (SB, $cmol_c\ dm^{-3}$).

	Attributes	OM - P	OM - SB	Prod - Ca	Prod - Mg	Prod - SB
	Geostatistical models	Gaussian	Exponential	Matérn k=2.5	Gaussian	Matérn k=2.5
Parameters	$\widehat{\mu_1}$	45.81	43.95	-0.96	3.16	3.50
	$\widehat{\mu_2}$	17.52	6.14	5.95	1.94	6.40
	$\widehat{\sigma_{01}}$	$-1.30\cdot10^{-11}$	3.14	$1.54\cdot10^{-13}$	$2.12\cdot10^{-14}$	$-5.65\cdot10^{-14}$
	$\widehat{\sigma_1}$	$-1.90\cdot10^{-10}$	$1.91\cdot10^{2}$	$3.46\cdot10^{-11}$	$5.84\cdot10^{-12}$	$-3.36\cdot10^{-12}$
	$\widehat{\sigma_{02}}$	$-1.55\cdot10^{-13}$	1.87	$5.70\cdot10^{-14}$	$1.98\cdot10^{-12}$	$-3.62\cdot10^{-12}$
	$\widehat{\sigma_2}$	$-1.91\cdot10^{-10}$	$5.78\cdot10^{-7}$	$3.18\cdot10^{-12}$	$5.92\cdot10^{-12}$	$-2.20\cdot10^{-11}$
	$\widehat{\varphi_0}$	179.41	79.43	179.53	190.95	186.60
	$\widehat{a_0}$	310.53	237.97	1,062.60	330.50	1,104.45
	$\widehat{\varphi_1}$	270.46	269.28	267.82	259.49	275.55
	$\widehat{a_1}$	468.13	806.70	1,585.16	449.13	1,630.90
	$\widehat{\varphi_2}$	275.66	296.02	267.84	247.56	269.10
	$\widehat{a_2}$	477.12	886.81	1,585.25	428.28	1,592.71

$\widehat{\mu_k}$, $\widehat{\sigma_k}$, $\widehat{\varphi_k}$, $\widehat{a_k}$: the estimated values of the mean, dispersion, range and practical range parameters, respectively, associated with the k-th variable ($k = 1,2$). $\widehat{\sigma_{0k}}$, $\hat{\varphi}_0$, $\widehat{a_0}$: the estimated values of the dispersion, range, and practical range parameters, associated with the common random field.

and SB is also reflected in the positive correlations between Prod and the Ca and Mg contents. The inverse spatial correlation between OM and P can be explained by the fact that, in places with more OM, the P content adsorbed in the clay fraction of the latosol is lower (Fink et al., 2016). Therefore, in regions with more OM, there was greater P availability in the soil solution and, as P is an element of low mobility and that is rarely lost by leaching (Maggi et al., 2011), it was possibly lost due to erosion or absorbed by the plants in larger amounts, resulting in neighborhoods with lower P contents in the soil.

Materials and methods

Description of the study area and of the soil attributes

The data were collected during the 2016-2017 agricultural year in a commercial grain production area of 167.35 ha cultivated with soybean, where no-till has been carried out since 1994 (Figure 1b). The area is located in the municipality of Cascavel, western region of Paraná, Brazil, with approximate geographic coordinates of 24.95º South for latitude and 53.37º West for longitude, with 650 m of mean altitude. The soil is classified as Dystroferric Red Latosol, the regional climate is mesothermic and super humid temperate, climatic-type Cfa (Köppen).

The soil attributes used in this study were calcium content (Ca, $cmol_c\ dm^{-3}$); phosphorus content (P, $mg\ dm^{-3}$); magnesium content (Mg, $cmol_c\ dm^{-3}$); organic matter content (OM, $g\ dm^{-3}$); sum of bases (SB, $cmol_c\ dm^{-3}$), which represents the sum of exchangeable cations (Ca^{2+}, Mg^{2+}, and K^{+}) in the soil; and soybean productivity ($t\ ha^{-1}$). Soybean productivity (Prod) with the calcium (Ca), magnesium (Mg) and sum of bases (SB) attributes was considered as pairs of variables, in addition to organic matter (OM) with the phosphorus (P) and sum of bases (SB) attributes. Such pairs were chosen because they have an agronomically interesting association due to their contribution to plant growth and development and, consequently, to grains (Mengel and Kirkby, 2001; Malavolta, 2006; Dalchiavon et al., 2017).

Bivariate Gaussian Model

In the cases where there is statistical evidence of spatial correlation between two attributes, the spatial pattern of these variables can be modeled and described considering a bivariate Gaussian spatial model (Fonseca, 2008). The Bivariate Gaussian Common Component Model (BGCCM) (Diggle and Ribeiro Jr., 2007) was employed in this study.

In the BGCCM proposal, there are two random Gaussian fields that can be modeled as follows (Fonseca, 2008):

$$\begin{cases} Y_1 = \mu_1 + \sigma_{01} S_0 + \sigma_1 S_1 \\ Y_2 = \mu_2 + \sigma_{02} S_0 + \sigma_2 S_2 \end{cases}, \quad (1)$$

where μ_1, μ_2 are the mean values of the Y_1 and Y_2 variables, respectively; $\boldsymbol{\sigma}^* = (\sigma_{01}, \sigma_1, \sigma_{02}, \sigma_2)^T$ is the vector of dispersion parameters of the bivariate geostatistical model; and S_0, S_1, and S_2 are mutually independent Gaussian random fields. Random field S_0 is common to the Y_1 and Y_2 variables, while S_1 are S_2 individually associated with each variable (Righetto, 2012). Thus, BGCCM presents a covariance structure built from three correlation functions valid for S_0, S_1 and S_2, which will be denoted by ρ_0, ρ_1 and ρ_2, respectively.

Let us suppose that $Y_k(\boldsymbol{s}_m)$ and $Y_k(\boldsymbol{s}_l)$ are Y_k attribute observations measured in locations $\boldsymbol{s}_m$ and $\boldsymbol{s}_l$, which are separated by an Euclidean distance of $h = h_{ml}$, with $m, l = 1, \ldots, n_k$ and $k = 1,2$. In this way, there is $\boldsymbol{Y} = (Y_1, Y_2)^T$, in which vector $\boldsymbol{Y}$ has an n-varied Gaussian distribution, with $n = n_1 + n_2$, n_1 and n_2 being the sample sizes of Y_1 and Y_2, respectively. In this paper, it was considered that $n_1 = n_2$. In addition to that, $\boldsymbol{Y}$ has mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ and positive definite covariance matrix $\boldsymbol{\Sigma}_Y$, given by (Fonseca, 2008):

$$\boldsymbol{\Sigma}_Y = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{1,2} \\ \boldsymbol{\Sigma}_{1,2}^T & \boldsymbol{\Sigma}_2 \end{pmatrix},$$

where $\boldsymbol{\Sigma}_k$ is the $n_k \times n_k$ covariance matrix of the Y_k variable, $k = 1,2$; $\boldsymbol{\Sigma}_{1,2}$ is the $n_1 \times n_2$ matrix with the cross-covariances between the Y_1 and Y_2 variables; the elements of the $\boldsymbol{\Sigma}_Y$ covariance matrix are given by $C(h) = \sigma_{0k}^2\, \rho_0(h) + \sigma_k^2 \rho_k(h)$, where C represents the covariance function in relation to distance h, and ρ_0, ρ_k are the correlation functions in S_0 and S_k, respectively, for $k = 1,2$.

Hence, the probability distribution of vector $\boldsymbol{Y}$ depends on parameter vector estimation $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\varphi})^T$, in which vector $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ is associated with the mean, vector $\boldsymbol{\sigma} = (\sigma_{01}, \sigma_1, \sigma_{02}, \sigma_2)^T$ is the dispersion vector, and vector $\boldsymbol{\varphi} = (\varphi_0, \varphi_1, \varphi_2)^T$ is associated with range function φ_t, linked to the geostatistical model chosen for the ρ_t spatial correlation function, where $t = 0,1,2$ is related to random field S_t. According to Righetto (2012), the estimation of the $\boldsymbol{\theta}$ parameters follows the same criteria of univariate geostatistical techniques. Thus, the maximum likelihood method for parameter estimation was used in this paper (Cressie, 2015).

Spatial weighting matrix and bivariate spatial correlation

A spatial weight matrix is an n x n square matrix, where the w_{ij} space weights represent the connection degree between the regions according to some proximity criterion, showing the influence of location j on location i, $i, j = 1, \ldots, n$. The weight matrix presents a kind of weighting of the influence that the locations exert on each other (Almeida, 2013).

To establish the connection degree expressed in the spatial weight matrices, the geographic distance was considered, which in turn depends on a metric. In this paper, the $\boldsymbol{W}$ and $\boldsymbol{C}$ spatial weighting matrices were used, both of dimension n x n, where n is the number of observations, and the Euclidean Distance (ED) and the inverse of the Euclidean Distance (IED) were employed as metrics.

Matrix $\boldsymbol{W} = [(w_{ij})]$ is standardized by row, so that the sum of the weights of each row equals 1. For the ED, $w_{ij} = 1/v_i$, where v_i is the total number of neighbors in the i-th row considering the j columns that have neighbors. For the locations that have no neighbors, $w_{ij} = 0$. Locations i and j are considered neighbors whose distance is less than the cutoff distance. The cutoff distance varied from 200 m to 1,750 m for the simulated data and from 140 m to 1,750 m for the real data. For the IED, $w_{ij} = \frac{1}{\sum_{j=1}^{n} 1/d_j} \frac{1}{d_{ij}}$, where $\sum_{j=1}^{n} 1/d_j$ represents the sum of the inverse of the distances between neighbors in the i-th line considering the n columns, and $1/d_{ij}$ is the inverse of the distance of the element belonging to line i and column j, $i \neq j = 1, \ldots, n$ (Bivand and Wong, 2018), for $i = j$, $w_{ij} = 0$.

Matrix $\boldsymbol{C} = [(c_{ij})]$ is globally standardized, making the sum of the weights to be n. For the ED, $c_{ij} = 1/v$, where v is the total number of neighbors considering the n points; while for the IED, $c_{ij} = \frac{1}{\sum_{i,j=1}^{n} 1/d_{ij}} \frac{1}{d_{ij}}$, where $\sum_{i,j=1}^{n} 1/d_{ij}$ is the sum of the inverse of the distances considering all neighbors (Bivand et al., 2013), for $i = j$, $c_{ij} = 0$.

Using a spatial weighting matrix, it is possible to establish the spatial correlation between two georeferenced variables in n sampling points. The expression for calculating bivariate spatial correlation, developed by Lee (2001), is given by:

$$L = \frac{n}{\sum_{i=1}^{n}(\sum_{j=1}^{n} w_{ij})^2} \frac{\sum_{i=1}^{n}[(\sum_{j=1}^{n} w_{ij}(x_i - \bar{x}))(\sum_{j=1}^{n} w_{ij}(y_j - \bar{y}))]}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

(2)

where n is the number of georeferenced sampling points; x_i, x_j and y_i, y_j are the values of the X and Y attributes in the i,j-th observations, respectively, $i,j = 1,\dots,n$; $\bar{x}$ and $\bar{y}$ are the mean values of the X and Y attributes in the area under study; and w_{ij} is an element of the spatial weight matrix.

Description of the simulations

The simulation study aimed at reproducing situations that could occur in different experimental agricultural areas with any attribute pairs, thus adding practical-theoretical knowledge about the bivariate spatial correlation between soil attributes with a spatial dependence structure.

Three tests were elaborated $(\mathrm{T1}, \mathrm{T2}, \mathrm{T3})$ considering different spatial dependence structures (Figure 1a), and based on BGCCM. The first objective of these tests was to analyze the Lee index variation in relation to the increase in the values of the simulated range parameters ($\boldsymbol{\varphi}$).

The second objective consisted in evaluating performance of the Lee index considering different spatial weighting matrices, combined with different metrics for calculating the distances between observations. T3 was discarded for this purpose because the variables had radiuses greater than the cutoff point in the previous stage (matrix **W** with ED). Due to the similarity observed between T1 and T2 in the first stage regarding the Lee index values, tests were performed for both tests with matrices **W** and **C**. However, at T2, the spatial dependence radius extrapolated the cutoff point. Thus, only T1 was considered in the second stage of this study.

For each test, nine simulations were generated, with two variables, containing 100 sampling points each. A Monte Carlo experiment was used for this, from the Cholesky decomposition of covariance matrix $\boldsymbol{\Sigma}_{\mathrm{Y}}$ (Cressie, 2015). Following the scheme in Figure 1a, for each of the nine pairs of simulated variables from each test, the practical range was estimated considering BGCCM, as well as from construction of the cross-semivariogram, in which the practical range represents the maximum distance of spatial dependence between two variables (Cressie, 2015). Subsequently, calculation of the bivariate spatial correlation, using the Lee index, was subdivided into two stages that correspond to both of the aforementioned objectives.

In both stages, the correlogram was designed to investigate the behavior of Lee's bivariate spatial correlation. To analyze the statistical significance of the values that make up the correlogram, a simulated envelope graph was used, which is generated from permutations of the simulated data in the sampling grid coordinates. Thus, to obtain the Lee index value, the geographic coordinates are kept unchanged, and the n values of the variable pair are permuted. According to Diggle and Ribeiro Junior (2007), the principle of performing permutations is to try to break the spatial dependence structure of the data, generating a type of independent data.

We indicated that the correlogram is not statistically significant if the line corresponding to the Lee index values is completely contained between the upper and lower limits of the simulated envelope. Otherwise, if the line is above the upper limit for some cutoff distance d, the correlogram is considered statistically significant in favor of positive spatial dependence; if the line is below the lower limit for some distance, the correlogram is statistically significant but indicates a negative spatial dependence between the variables (Diggle and Ribeiro Jr., 2007; Costa and Scalon, 2015). The distance at which the correlogram enters the envelope is the spatial dependence radius of the variable pair in question (Costa and Scalon, 2015). The spatial dependence radius was compared to the practical ranges obtained by fitting the cross-semivariogram and using the bivariate model.

Description of the real data analysis and computational resources

Lattice plus close pairs sampling was used in the agricultural area, consisting of 102 sampling points, which comprises both a regular grid with a minimum distance of 141 m between the points, as well as 19 locations that were randomly added to the regular grid and which have smaller distances with some observations of it (from 50 m and 75 m) (Figure 1b). The samples were located and georeferenced by a Global Navigation Satellite System (GNSS) signal receiver device using a spatial UTM (Universal Transverse Mercator) coordinate system.

The soil attributes used in this study were spatially dependent, isotropic and without any directional trend. The same spatial statistical analyses were performed for the real data set (estimation of the BGCCM spatial model, cross-semivariogram, and Lee index calculation) as in the simulated data. In real data sets, geostatistical models were adjusted for each pair of variables (Diggle and Ribeiro Jr., 2007) for which the quality of the estimates obtained were evaluated using the cross-validation criteria (Faraco et al., 2008).

The routines for calculating the Lee index and for other statistical and geostatistical analyses were developed using the R software (R Development Core Team, 2022), considering the geoR (Ribeiro Jr. and Diggle, 2001) and spdep (Bivand, 2020) packages.

Conclusion

When comparing the different spatial weighting matrices in the simulated data analysis, a similarity was verified between the Lee index values. However, the IED and the ED presented different results, with larger spatial dependence radiuses and higher Lee index values achieved using the IED, which also required more computational time. Therefore, the metric to be considered to calculate the distance between the point pairs was more relevant than the weight matrix in estimating the spatial dependence radius and the Lee index value.

Also in the simulated tests, as the simulated common range increased, the mean values of the practical range estimated in the cross-semivariogram and by BGCCM also increased and were satisfactory, as they approached the simulated practical ranges. In addition to that, the bivariate spatial correlation value calculated by means of the Lee index was higher as the simulated practical range increased. The same relationship between Lee index values and spatial dependence radius was verified for the attribute pairs analyzed in the real data, so that the practical studies corroborate with the simulated ones.

Soybean productivity presented a positive spatial correlation with the sum of bases and with the calcium and magnesium contents, indicating that in the regions with the highest soybean productivity there was greater availability of these attributes. In addition, organic matter had a positive spatial correlation with the sum of bases and a negative one with phosphorus, respectively due to the CEC of the region's soil and to absorption by the plants or erosion.

Acknowledgements

The authors are grateful for the funding received from the Coordination for the Improvement of Higher Education Personnel (CAPES), Financing Code 001, National Council for Scientific and Technological Development (CNPq), and the Spatial Statistics Laboratory - UNIOESTE.

References

Anselin L, Syabri I, Smirnov O (2002) Visualizing multivariate spatial correlation with dynamically linked windows. In: Anselin L, Rey S (eds) New tools for spatial data analysis: proceedings of the specialist meeting. Center for Spatially Integrated Social Science (CSISS), Santa Barbara.

Almeida E (2013) Econometria Espacial Aplicada, 2nd edn. Alinea, Campinas.

Barbosa JM, Rezende CFA, Leandro WM, Ratke RF, Flores RA, Silva AR (2016) Effects of micronutrients application on soybean yield. Aust J Crop Sci. 10(8):1092-1097.

Bivand RS, Pebesma EJ, Gómez-Rubio V (2013) Applied spatial data analysis with R, 2nd edn. Springer, New York.

Bivand RS, Wong DW (2018) Comparing implementations of global and local indicators of spatial association. Test. 27:716-748.

Bivand RS (2020) spdep: Spatial Dependence: Weighting Schemes, Statistics and Models. R package version 1.1.5. http://CRAN.R-project.org/package=spdep

Cambardella CA, Moorman TB, Novak JM, Parkin TB, Karlen DL, Turco RF, Konopka AE (1994) Field-scale variability of soil properties in Central Iowa Soils. Soil Sci Soc Am J. 58:1501-1511.

Cantu, JG (2015) Influência de configurações amostrais na qualidade de estimação espacial sob

o uso de modelos espaciais bivariados (Mestrado em Engenharia Agrícola). Universidade Estadual do Oeste do Paraná, Cascavel.

Cima EG, Uribe-Opazo MA, Johann JA, Rocha Jr. WFD, Dalposso GH (2018) Analysis of spatial autocorrelation of grain production and agricultural storage in Parana. Eng Agríc. 38:395-402.

CONAB (2017) Companhia Nacional de Abastecimento. Acompanhamento da Safra Brasileira de Grãos. CONAB, Brasília. Available in https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos. Access in 08/2017.

Costa FM, Scalon JD (2015) Análise da estrutura de dependência espacial em dados de áreas. Rev Bras Biom. 33(2):197-215.

Cressie NAC (2015) Statistics for Spatial Data, rev. edn. John Wiley & Sons, New York.

Cunha TJF, Mendes MAS, Giongo V (2015) Matéria orgânica do solo. In: Nunes RR, Rezende MOO (eds) Recurso solo: propriedades e usos. Cubo, São Carlos.

Dalchiavon FC, Rodrigues AR, Lima ES, Lovera LH, Montanari R (2017) Variabilidade espacial de atributos químicos do solo cultivado com soja sob plantio direto. Rev Cienc Agrovet. 16(2):144-154.

Deiss L, Kleina GB, Moraes A, Franzluebbers AJ, Motta AC, Dieckow J, Sandini IE, Anghinoni I, Carvalho PC (2020) Soil chemical properties under no-tillage as affected by agricultural trophic complexity. Eur J Soil Sci. 71(6):1090-1105.

Deus ACF, Büll LT, Guppy CN, Santos SDMC, Moreira LLQ (2020) Effects of lime and steel slag application on soil fertility and soybean yield under a no till-system. Soil Tillage Res. 196:104422.

Diggle PJ, Ribeiro Jr. PJ (2007) Model based geostatistics. Springer, New York.

Faraco MA, Uribe-Opazo MA, Silva EAA, Johann J, Borssoi JA (2008) Seleção de modelos de variabilidade espacial para elaboração de mapas temáticos de atributos físicos do solo e produtividade da soja. Rev Bras Cien Solo. 32(2):463-476.

Fink JR, Inda AV, Tiecher T, Barrón V (2016) Iron oxides and organic matter on soil phosphorus availability. Cienc e agrotecnologia. 40(4):369-379.

Fonseca BHF (2008) Um estudo sobre estimação e predição em modelos geoestatísticos bivariados (Mestrado em Estatística e Experimentação Agronômica). Universidade de São Paulo, Escola Superior de Agricultura "Luiz de Queiroz", São Paulo.

Gaso DV, Wit A, Berger AG, Kooistra L (2021) Predicting within-field soybean yield variability by coupling Sentinel-2 leaf area index with a crop growth model. Agric For Meteorol. 308-309:108553.

Lee SI (2001) Developing a bivariate spatial association measure: an integration of Pearson's r and Moran's I. J Geogr Syst. 3:369-385.

Liu Q, Xie WJ, Xia JB (2013) Using semivariogram and Moran's I techniques to evaluate spatial distribution of soil micronutrients. Commun Soil Sci Plant Anal. 44(7):1182-1192.

Maggi CF, Freitas PS, Sampaio SC, Dieter J (2011) Lixiviação de nutrientes em solo cultivado com aplicação de água residuária de suinocultura. Rev Bras Eng Agríc Ambient. 15(2):170-177.

Malavolta E (2006) Manual de nutrição de plantas. Editora Agronomica Ceres, São Paulo.

Malvezi KED, Zanão Jr. LA, Guimaraes EC, Vieira SR, Pereira N (2019) Soil chemical attributes variability under tillage and no-tillage in a long-term experiment in southern Brazil. Biosci J. 35(2):467-476.

Matkan AA, Shahri M, Mirzaie M (2013) Bivariate Moran's I and LISA to explore the crash risky locations in urban areas. N-Aerus. 14:1-12.

Mengel K, Kirkby E (2001) Principles of plant nutrition, 5th edn. Kluwer Academic Publishers, Boston.

Pimentel Gomes F, Garcia CH (2002) Estatística Aplicada a Experimentos Agronômicos e Florestais. FEALQ, Piracicaba.

R Development Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Retrieved from https://www.R-project.org/

Ramos FT, Dores EFC, Weber OLS, Beber DC, Campelo Jr. JH, Maia JCS (2018) Soil organic matter doubles the cation exchange capacity of tropical soil under no-till farming in Brazil. J Sci Food Agric. 98(9):3595-3602.

Ribeiro Jr. PJ, Diggle PJ (2001) geoR: a package for geostatistical analysis. R-NEWS. 1:15-18.

Righetto AJ (2012) Avaliação de modelos geoestatísticos multivariados (Mestrado em Estatística e Experimentação Agronômica). Universidade de São Paulo, Escola Superior de Agricultura "Luiz de Queiroz", São Paulo.

SBCS - NEPAR (2017) Sociedade Brasileira de Ciência do Solo - Núcleo Estadual Paraná. Manual de adubação e calagem para o estado do Paraná. SBCS-NEPAR, Curitiba. Available in https://sbcs-nepar.org.br/produto/manual-de-adubacao-e-calagem-para-o-estado-do-parana-2a-edicao/. Access in 09/2017.

Siqueira-Neto M, Popin GV, Piccolo MC, Corbeels M, Scopel E, Camargo PB, Bernoux M (2021) Impacts of land use and cropland management on soil organic matter and greenhouse gas emissions in the Brazilian Cerrado. Eur J Soil Sci. 72(3):1431-1446.