

Reduction of Sample Size in the Soil Physical-Chemical Attributes Using the Multivariate Effective Sample Size

Letícia Ellen Dal Canton (Corresponding author)

Western Paraná State University – UNIOESTE

Rua Universitária, 1619, Cascavel, Paraná, 85819-170, Brazil

Tel: +55 (45) 3220-7320 E-mail: leticiacanton@hotmail.com

Luciana Pagliosa Carvalho Guedes

Western Paraná State University – UNIOESTE

Rua Universitária, 1619, Cascavel, Paraná, 85819-170, Brazil

Tel: +55 (45) 3220-7320 E-mail: luciana.guedes@unioeste.br

Miguel Angel Uribe-Opazo

Western Paraná State University – UNIOESTE

Rua Universitária, 1619, Cascavel, Paraná, 85819-170, Brazil

Tel: +55 (45) 3220-3228 E-mail: miguel.opazo@unioeste.br

Received: August 4, 2020 Accepted: September 29, 2020 Published: January 25, 2021

doi:10.5296/jas.v9i1.17473

URL: <https://doi.org/10.5296/jas.v9i1.17473>

Abstract

Financial investment with collection and laboratory analysis of soil samples is an important factor to be considered when mapping agricultural areas with soybean planting. One of the alternatives is to use the spatial autocorrelation between the sample points to reduce the number of elements sampled, thus restricting the collection of redundant information. This work aimed to reduce the sample size of this agricultural area, composed of 102 sample points, and use it to analyze the spatial dependence of soil macro- and micro- nutrients, as well as the soil penetration resistance. The agricultural area used in this study has 167.35 ha,

cultivated with soybean, which the soil is Red Dystroferic Latosol, and the sampling design has used in this agricultural area is the lattice plus close pairs. The reduction of the sample size was made by the multivariate effective sample size (ESS_{multi}) methodology. The studies with the simulation data and the soil attributes showed an inverse relationship between the practical range and the estimated value of the univariate effective sample size. With the calculation of ESS_{multi} , the sample configuration was reduced to 53 points. The Overall Accuracy and Tau concordance index showed differences between the thematic maps elaborated with the original and reduced sampling designs. However, the analysis of the variance inflation factor and the standard error of the spatial dependence parameters showed efficient results with the resized sample size.

Keywords: geostatistics, simulation, spatial dependence, redundant information, variance inflation factor

1. Introduction

Throughout economic cycles, the Brazilian agribusiness has shown to be fundamental to the country's development, in addition to ensuring a prominent position capable of influencing the international market. Currently, soy (*Glycine max* (L.) Merrill) is the main agricultural product in Brazil and, in recent years, it has shown continuous growth in production and in planted area (Balbinot Junior et al., 2017). Agricultural expansion around the planted area is encouraged by the favorable climate and topography and by competitive hectare prices in some states in the North and Northeast of the country (Bolfe et al., 2016).

In the last decade, the area planted with soybeans in Brazil went from 23.4 million hectares to 36.8 million hectares, an increase of 57.2% (CONAB, 2020), whereas production increased from 68.9 million tons to 124.2 million tons of this oilseed, representing an increase of 80.2% (CONAB, 2020). Paraná ranks among the states with the highest soybean productivity in the country and contributed approximately 17% of the national planted area of this crop, as well as 18% of the national production of this grain between the 2009/10 and 2019/20 crop years (CONAB, 2020). These prove the importance of the commodity in the national and state context.

The chemical and physical properties of the soil can affect root growth and the production of sensitive crops, such as soybeans (Debiasi et al., 2013). Thus, knowledge about the spatial variability of soil attributes is essential for the future development of soybeans and should be considered in crop management (Sanches et al., 2019).

The theory of regionalized variables is considered as the basis of geostatistics, in which, from a set of georeferenced sampling elements, the spatial dependence structure is analyzed and interpolated thematic maps are constructed that represent the spatial variability of physical-chemical attributes of the soil (Cressie, 2015). However, the efficiency of geostatistical techniques in the spatial characterization of soil properties is directly conditioned by the quality of the soil sampling performed (Cherubin et al., 2014).

Sampling planning mainly involves establishing the number of sampling points collected and the sampling configuration, and it must be efficient in describing the spatial variability of soil

properties in the study area (Benedetti et al., 2015). In the study of spatial data there are several sample configurations in the literature, such as simple random, stratified, regular (square, triangular, hexagonal), lattice plus close pairs, and lattice plus in fill (Diggle and Lophaven, 2006). The sample configuration must consider mainly the shape of the area, the relief characteristics, and the available resources, which are mainly influenced by the costs of laboratory analyses (Gao et al., 2016).

The costs of collecting and analyzing samples of soil attributes have led to the development of many studies within the scope of sample resizing, aiming to reduce sampling costs, considering a minimal loss of information in spatial prediction. Among these, we can mention from the optimization algorithms (Guedes et al., 2016; Wadoux et al., 2017; Maltauro et al., 2019) to the use of the Effective Sample Size (*ESS*) (Acosta et al., 2018; Acosta and Vallejos, 2018). The calculation of the effective sample size considers the effect of spatial autocorrelation between the sampled points collected. To calculate the value of the effective sample size, a previous sample survey of the physical-chemical attributes of the soil in the desired agricultural area is necessary (Griffith, 2005; Vallejos and Osorio, 2014).

Studies and applications of effective sample size have been developed, especially with a univariate approach and space-time models (Griffith, 2008; Acosta et al., 2016). However, in the agricultural context, several soil attributes are generally considered in the study of spatial variability, making it impossible to obtain a different sample resizing for each variable.

Therefore, the objective of this work was to redefine the number of sample points using the effective multivariate sample size (*ESS_{multi}*). This proposal makes it possible to obtain a single sample size, considering the information on the spatial dependence structure of soil physical-chemical attributes collected in an agricultural area.

2. Material and Methods

Two studies were carried out: the first considering simulated data, and the second using data of physical-chemical attributes of the soil obtained in an agricultural area. The simulation study aimed to reproduce a list of possibilities present in the real data, in addition to adding practical and theoretical knowledge about sample resizing in the soil attributes with spatial dependence structure.

2.1 Description of the simulations

We considered 14 georeferenced variables with different spatial dependence structures (Figure 1 – B). Given the Gaussian linear spatial model, stationary and isotropic (Uribe-Opazo et al., 2012), with fixed mean ($\mu = 5$), we performed 100 simulations for each of the 14 variables using a Monte Carlo experiment (Cressie, 2015). In these simulations, we fixed the same configuration (lattice plus close pairs), as well the same sample size (102 sampling points), from the commercial agricultural area used in the experimental data (Figure 1 – C). We also fixed the exponential semivariance model (Diggle and Ribeiro Jr., 2007), with the partial sill ($\phi_2 = 1$), as well the other values of the parameters that define

the spatial dependence structure of the simulated variables are described in Figure 1–B. The scheme presented in Figure 1 – A presents other features of the simulations.

After estimating the parameter vector $\theta_i = (\mu, \varphi_1, \varphi_2, \varphi_3)^T$ by the maximum likelihood method for each simulated variable, the value of the univariate effective sample size (ESS_{uni}) was estimated according to Equation 1 (Vallejos and Osorio, 2014).

$$\widehat{ESS}_{uni} = \mathbf{1}^T \mathbf{V}(\widehat{\boldsymbol{\varphi}})^{-1} \mathbf{1}, \quad (1)$$

$$v(\widehat{\boldsymbol{\varphi}})_{ij} = \begin{cases} 1, & \text{if } i = j = 1, \dots, n \\ \frac{\widehat{\varphi}_2 r_{ij}}{\widehat{\varphi}_1 + \widehat{\varphi}_2}, & \text{if } i \neq j = 1, \dots, n, \end{cases} \quad (2)$$

where n is the number of simulated sampling points in the original grid ($n \geq 1$); $\mathbf{1}$ is an $n \times 1$ unit vector; $\mathbf{V}(\widehat{\boldsymbol{\varphi}}) = [(v(\widehat{\boldsymbol{\varphi}})_{ij})]$ is an $n \times n$ estimated spatial correlation matrix of the sample points, where the estimated spatial correlation between the i -th and the j -th sampling points are given by Equation 2; $\widehat{\varphi}_1$, $\widehat{\varphi}_2$ are the estimated values of the nugget effect and partial sill parameters, respectively; and r_{ij} are the elements of the $\mathbf{R}(\varphi_3)$ matrix, whose calculation depends on the geostatistical model and on the Euclidean distance between observations (Uribe-Opazo et al., 2012).

Also, we created four scenarios (S_1, S_2, S_3, S_4) (Figure 1 – B), where each one has a different number of variables and corresponds to a subset of the set composed of the 14 variables (Figure 1 – B). The first scenario (S_1) aggregates variables with moderate to strong intensity of spatial dependence (defined by the relative nugget effect (%), $\widehat{RNE} = 100(\widehat{\varphi}_1 / (\widehat{\varphi}_1 + \widehat{\varphi}_2))$) and intermediate spatial dependence radius (α). The second and third scenarios (S_2 and S_3 , respectively) have a strong spatial dependence. However, S_2 only has high range values, whereas in S_3 , the range varies between intermediate and high. Finally, the fourth scenario (S_4) brings together all the variables.

For these scenarios, we estimated the value of the multivariate effective sample size (ESS_{multi}) considering the methodological scheme presented in Figure 1 – A and the proposal developed by Vallejos & Osorio (2014) (Equation 3):

$$\widehat{ESS}_{multi} = \frac{\mathbf{1}_{(1)}^T \left((\mathbf{A}_d \boldsymbol{\Phi}_d \otimes \mathbf{I}_n) \mathbf{W}_d^{1/2} \right)^T (\mathbf{C} \otimes \mathbf{I}) \left((\mathbf{A}_d \boldsymbol{\Phi}_d \otimes \mathbf{I}_n) \mathbf{W}_d^{1/2} \right) \mathbf{1}_{(1)}}{\mathbf{1}_{(2)}^T \mathbf{A}_d \boldsymbol{\Phi}_d \mathbf{C} \boldsymbol{\Phi}_d \mathbf{A}_d \mathbf{1}_{(2)}}, \quad (3)$$

where n is the number of sampling points in the original grid ($n \geq 1$); \otimes denotes the Kronecker product; \mathbf{C} is a $p \times p$ linear correlation matrix between the p simulated variables; \mathbf{A}_d is a $p \times p$ diagonal matrix with the coefficient of the standardized eigenvectors of matrix \mathbf{C} ; $\mathbf{1}_{(1)}$ and $\mathbf{1}_{(2)}$ are $p \times 1$ and $np \times 1$ unit vectors, respectively; Φ_d is a $p \times p$ diagonal matrix with the standard deviation of the m -th variable (σ_m), $m = 1, \dots, p$; \mathbf{I}_n is the $n \times n$ identity matrix; and \mathbf{W}_d is an $np \times np$ block diagonal matrix, where each element of the diagonal is an $n \times n$ matrix, given by $\mathbf{V}_m(\hat{\Phi})$, that representing the spatial correlation matrix of the m -th variable, $m = 1, \dots, p$.

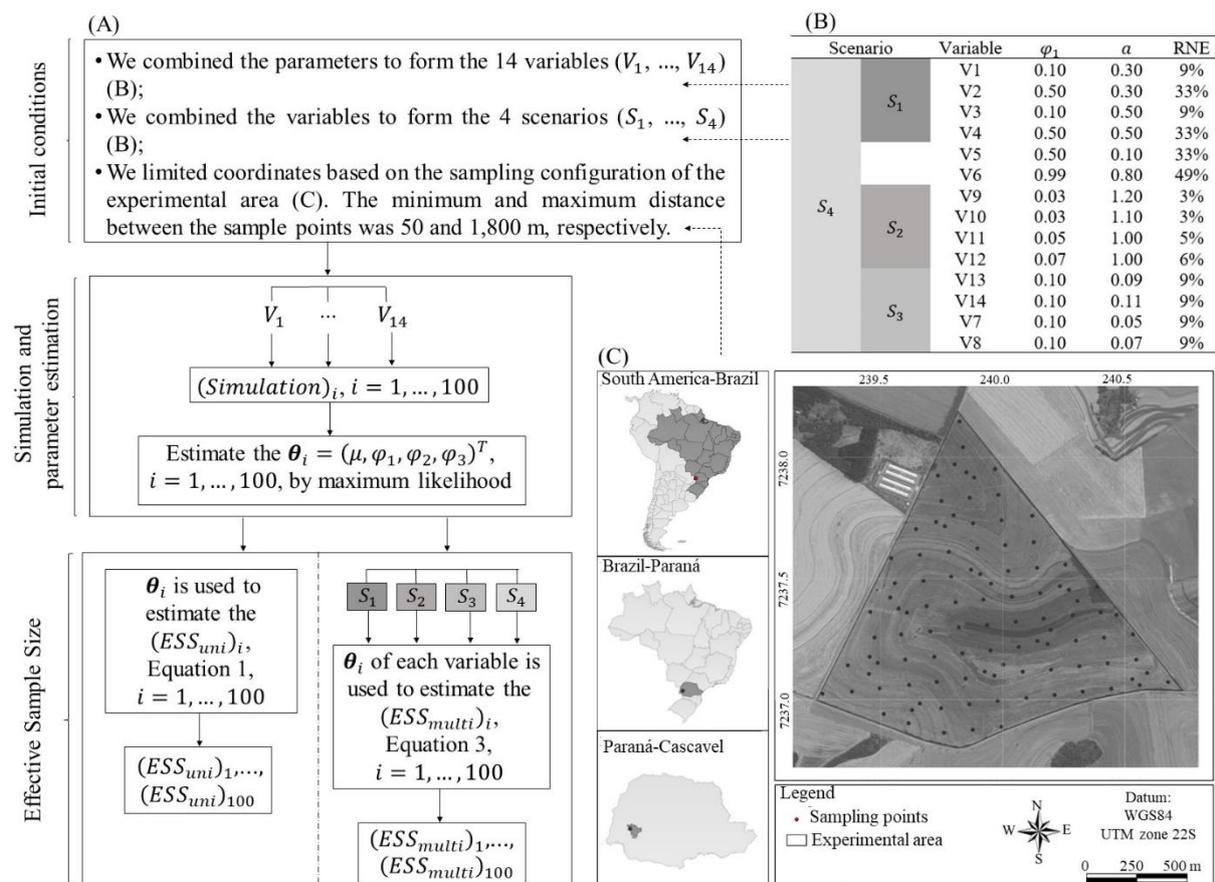


Figure 1. (A) Methodological scheme of the simulation study, (B) Values of the parameters that define the spatial dependence structure of the simulated variables. The gray scale shows the grouping of variables in the formation of each scenario, (C) Experimental area with the

location of the points sampled. μ : mean, φ_1 : nugget effect, φ_2 : partial sill, φ_3 : function of

the range, α : practical range (in kilometers), RNE: Relative Nugget Effect (in %),

ESS_{uni} : Univariate effective sample size, ESS_{multi} : Multivariate effective sample size.

2.2 Description of the Experimental Data

Data was collected in the 2015/2016 crop year in a commercial area with 167.35 hectares cultivated with soybean, where direct planting has been carried out since 1994 (Figure 1 – C). The area is located in the municipality of Cascavel, in Western Paraná, Brazil, with approximate geographical latitude and longitude coordinates of 24.95° South and 53.37° West, and 650 meters of average altitude. The soil is Red Dystroferic Latosol with a clay texture and has an average slope of approximately 4%, categorized as smooth undulating. The climate of the region is temperate mesothermic and superhumid, type Cfa climate (Koeppen), with a mean annual temperature of 21°C.

We used a lattice plus close pairs sampling design, with 102 sampling points. This design contained a regular grid (with minimum distance between points equal to 141 meters), to which we added 19 sample points. These added locations presented smaller distances with some points of the regular grid (50 and 75 m). The sample was georeferenced and located with the aid of a signal receiving apparatus with a Geoexplore 3 (Trimble®) Global Positioning System (GPS) set up for the Universal Transverse Mercator (UTM) coordinate system.

The data set we used contained the following: as physical attribute, the soil penetration resistance (in Mpa) at depths of 0-10 cm (SPR₁), 11-20 cm (SPR₂), 21-30 cm (SPR₃), and 31-40 cm (SPR₄); as chemical attributes, the following soil macro- and micro- nutrients: calcium (Ca, cmol dm⁻³), carbon (C, g dm⁻³), copper (Cu, mg dm⁻³), manganese (Mn, cmol dm⁻³), and zinc (Zn, mg dm⁻³). The experimental data of this physical and chemical attributes belongs to the database of the Laboratory of Spatial Statistics of the Western Paraná State University - UNIOESTE, Cascavel. The analysis of these soil attributes is important because the imbalance of their quantities in the soil can alter the growth and development phases of the plant, thus affecting the grain and, consequently, the soybean productivity (Taiz et al., 2017). Moreover, to better understand the nutritional characteristics of the soil, it is important to combine samples of macro- and micro- nutrients and physical attributes, such as soil penetration resistance (SPR), which is related to the soil compaction. Compacted soils tend to hinder the availability of nutrients and water to the plant, which also interferes in the development of the plant (Colombi and Keller, 2019).

Considering for each point the collection of five replications, the soil samples were obtained from 0 to 20 cm deep in the vicinity of the points, which were mixed and placed in plastic bags, with approximately 500 g, for the sample composition homogeneous and representative of the parcel. The values of micronutrients were extracted by the Mehlich-1 method, the carbon by Walkley-Black, and the calcium by KCl 1 mol L⁻¹. The determination of the soil

penetration resistance was measured by the penetrometer, as follows: for each sampling point, we performed three readings per centimeter, from 0 to 40 cm, covering the four depths considered (0-10 cm, 11-20 cm, 21-30 cm, and 31-40 cm). The data obtained was transformed in MPa, and the value of the soil resistance penetration at each depth consisted of the arithmetic mean of the three measurements.

Considering the original sample design, we performed the exploratory and geostatistical analyses of each physical-chemical attribute of the soil, as detailed in Figure 2. Details on the geostatistical methodologies used in this research are in Cressie (2015).

We estimated the values of the univariate effective sample size (ESS_{uni} , Equation 1) and multivariate effective sample size (ESS_{multi} , Equation 3) (Figure 2) by the same methodology applied in the simulated data (Figure 1 – A). The effective sample size methodology uses the Fisher's information matrix, which only considers the information on the spatial dependence structure, that is, it does not consider any information on spatial prediction in its sample resizing process (Vallejos and Osorio, 2014).

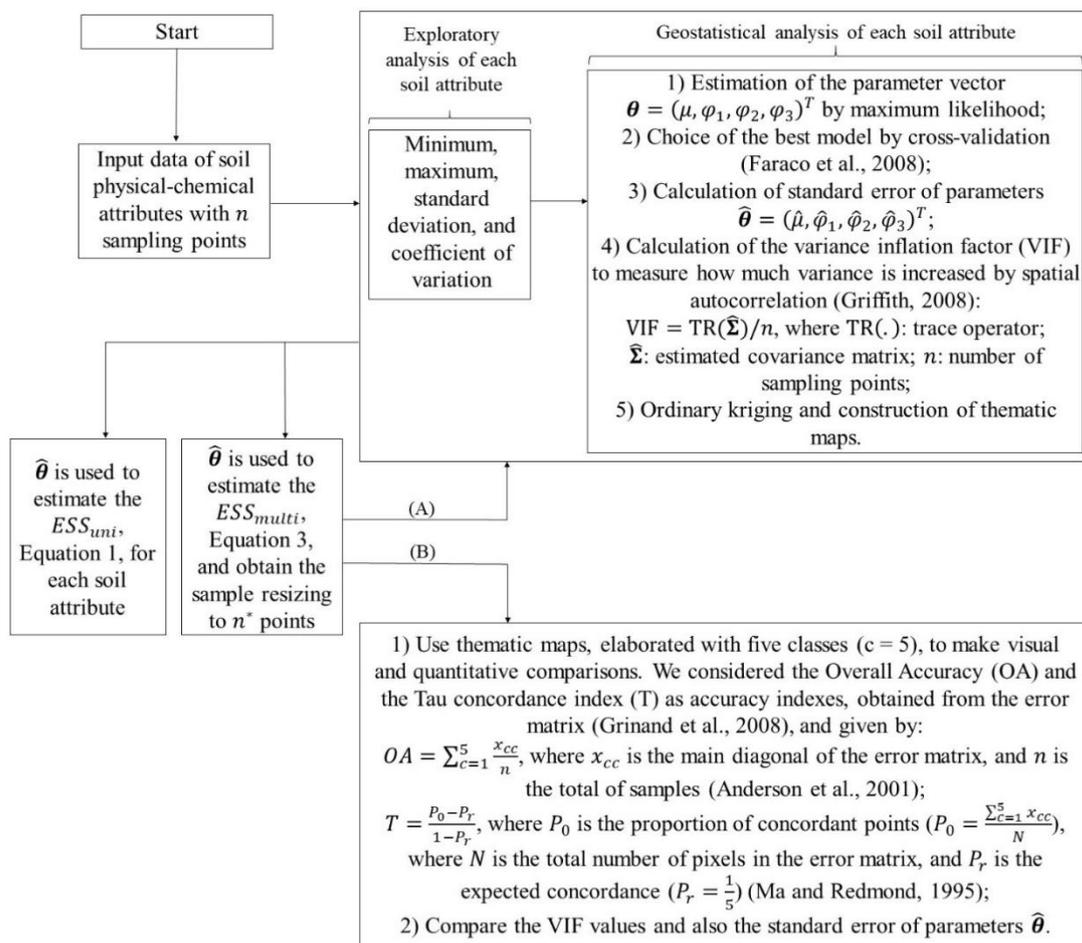


Figure 2. Methodological scheme used in experimental data. μ : mean, φ_1 : nugget effect,

φ_2 : partial sill, φ_3 : function of the range, ESS_{uni} : Univariate effective sample size,

ESS_{multi} : Multivariate effective sample size. (A) Redo the exploratory and geostatistical analyses with the resized sample configuration, (B) Compare the original and reduced sampling configurations

To verify the suitability of the reduced sample size provided by ESS_{multi} , a new sample configuration was selected from among the 102 points of the original sample grid, maintaining the same sampling design. For this reduced sample size, the exploratory and geostatistical analyses were also performed (Figure 2 – A).

Finally, we compared the results obtained between the two sample configurations (original and reduced), using the methodologies presented in Figure 2 – B.

The simulations and the statistical and geostatistical analysis were prepared in the R software (R Development Core Team, 2020), using the geoR package (Ribeiro Jr and Diggle, 2001). We developed a computational routine in the R software (R Development Core Team, 2020) using the geoR (Ribeiro Jr and Diggle, 2001) and matrixcalc (Novomestky, 2012) packages to estimate the univariate (ESS_{uni}) and multivariate (ESS_{multi}) effective sample size.

3. Results

3.1 Simulation Studies

The estimated values of the univariate effective sample size (ESS_{uni}) evidenced that the simulated variables divided into three groups (Figure 3 – A): variables V5, V7, V8, V13, and V14 made up the first group, the second group consisted of variables V1, V2, V3, V4, and V6, and variables V9, V10, V11, and V12 constituted the third group.

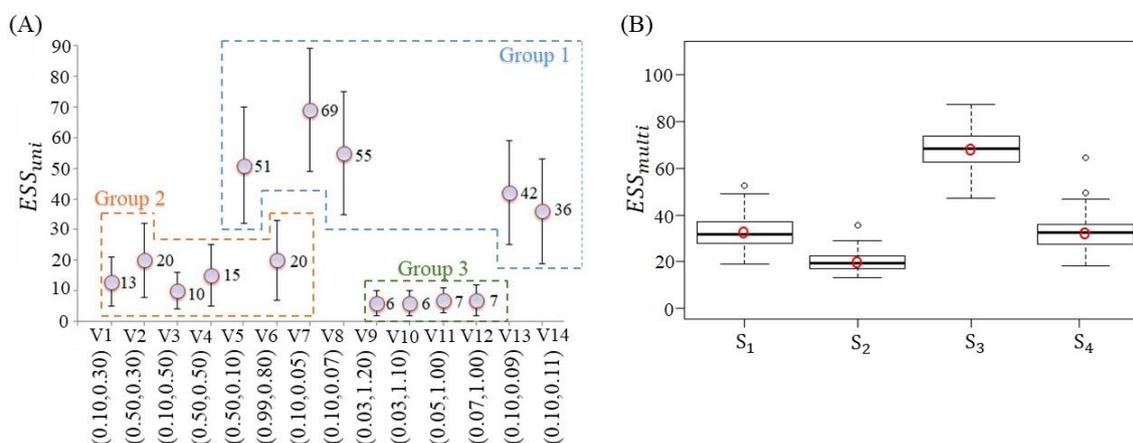


Figure 3. (A) Mean and Standard Deviation of the estimated values of the univariate effective

sample size (ESS_{uni}) for each variable (V_1, \dots, V_{14}). The nugget effect and practical range parameters are shown in parentheses, (B) Boxplot of the estimated values of the multivariate effective sample size (ESS_{multi}) in each scenario (S_1, \dots, S_4). The symbol “O” represents the estimated mean value of ESS_{multi}

Regarding the estimated values of the effective multivariate sample size (ESS_{multi}), there was a similarity between scenarios S_1 and S_4 (Figure 3 – B). These scenarios exhibited intermediate estimated mean values for ESS_{multi} (33 sample points), and were composed of different numbers of variables (4 and 14 variables, respectively), which indicated that the number of variables used did not influence the estimated of ESS_{multi} . Most of the variables had intermediate to high values of the spatial dependence radius (from 300 to 1,200 m), which possibly contributed to the similarity of scenarios S_1 and S_4 in relation to the estimated value of ESS_{multi} .

Scenario S_2 only aggregated variables with high values for the spatial dependence radius (between 1,000 and 1,200 m) and was different from the other scenarios because it presented the lowest estimated mean value of ESS_{multi} (20 sample points) (Figure 3 – B). Scenario S_3 diverged from the other ones because it showed the highest estimated mean value of ESS_{multi} (68 sample points) (Figure 3 – B). In this scenario, the variables had the lowest values of practical ranges, between 50 and 110 m, these being low when compared to the maximum distance in the area (approximately 1,800 m).

3.2 Application of the Methodology in Physical-Chemical Attributes of the Soil

When estimating the effective sample size of each physical-chemical attribute of the soil, four groups were identified (Figure 4 – A). These groups were obtained by simultaneous analyzing the estimated values of the spatial dependence radius (a), the intensity of spatial dependence (RNE), and the ESS_{uni} of each attribute.

In the identified first group, the attributes exhibited high intensity of spatial dependence (\widehat{RNE} between 4% and 30%, respectively; Cambardella et al., 1984) and low values for the spatial dependence radius (between 100 and 116 m), making this group have the highest estimated values of ESS_{uni} (approximately 90 sample points) (Figure 4 – A). This group was composed of the SPRs at depths of 11-20 and 21-30 cm.

The identified second and the third groups exhibited estimated ESS_{uni} values between 22 and 50 sample points (Figure 4 – A). These two groups contained the following attributes: carbon (C), manganese (Mn), and zinc (Zn) (group 2); and calcium (Ca) and SPRs at depths of 0-10 cm and 31-40 cm (group 3). The estimated values of the spatial dependence radius were close (between 220 and 390 m) for all attributes of these groups, except for carbon.

Thus, the intensity of spatial dependence, i.e., the estimated value of the RNE, influenced the difference in the estimated value of ESS_{uni} between these groups. That is because, in the second group, the attributes showed less intensity of spatial dependence (higher \widehat{RNE}) compared to the third group (Figure 4 – A). Isolated from the other attributes, the copper (Cu) content in the soil presented the lowest estimated values of ESS_{uni} (11 sample points) and exhibited the highest estimated value of practical range (538.7 m) (Figure 4 – A).

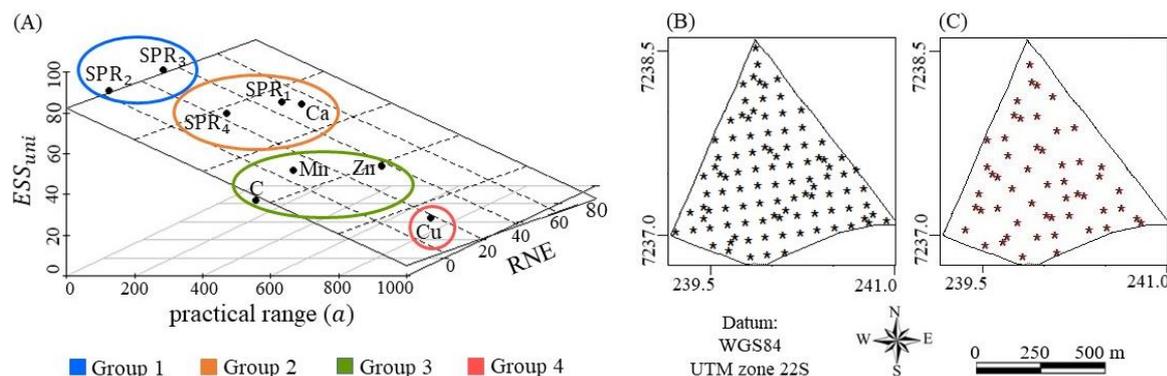


Figure 4. (A) Estimated values of the univariate effective sample size (ESS_{uni}) as a function of the practical range and of the Relative Nugget Effect (RNE), where C: Carbon, Ca: Calcium, Cu: Copper, Mn: Manganese, Zn: Zinc, and SPR₁, SPR₂, SPR₃, SPR₄: Soil Penetration Resistance at depths of 0-10, 11-20, 21-30, and 31-40 cm, respectively. (B) Original sampling design with 102 sample points. (C) Reduced sampling design with 53 sample points

For sample resizing, we estimated the value of the multivariate effective sample size (ESS_{multi}), obtaining 53 sample points. Thus, a new sample configuration with 53 points (Figure 4 – C), chosen randomly from the 102 sample points of the original grid (Figure 4 – B), was selected for the study of the spatial dependence of the physical-chemical attributes of the soil.

We observed that the reduction in the sample size did not influence dispersion, given by the coefficient of variation, or the mean value of the physical-chemical attributes of the soil (Table 1). The chemical analysis of the soil showed high mean levels for most of the macro- and micro-nutrients of the soil. Also, the most superficial depth layers of soil (0 to 20 cm) presented some limitations to root growth (Canarache, 1991).

Table 1. Descriptive statistics of the physical-chemical attributes of the soil in the agricultural area with soybean production under study, considering the original ($n = 102$) and reduced ($n^* = 53$) sample configurations.

	Minimum		Mean		Maximum		SD		CV	
	<i>n</i>	<i>n*</i>								
Ca	2.51	2.51	5.51	5.49	11.43	8.60	1.33	1.29	24.12	23.56
C	23.80	25.20	32.01	32.47	39.34	39.06	3.39	3.08	10.58	9.48
Cu	1.51	1.51	3.82	3.66	5.97	5.97	0.91	0.91	23.79	24.82
Mn	44.48	44.48	86.41	87.58	150.00	150.00	22.18	23.83	25.66	27.21
Zn	1.85	1.85	4.97	4.72	13.84	8.52	2.03	1.49	40.92	31.63
SPR ₁	0.75	1.48	3.12	3.03	6.67	6.07	0.97	0.91	31.22	30.03
SPR ₂	1.78	1.78	3.01	3.08	5.53	5.53	0.62	0.66	20.70	21.48
SPR ₃	1.50	1.57	2.03	2.04	3.86	2.85	0.33	0.28	16.27	13.81
SPR ₄	1.48	1.56	2.08	2.08	3.95	3.27	0.37	0.34	18.01	16.23

SD: Standard Deviation, CV=100(SD/Mean): Coefficient of Variation (%), Ca: Calcium (cmol dm^{-3}), C: Carbon (g dm^{-3}), Cu: Copper (mg dm^{-3}), Mn: Manganese (mg dm^{-3}), Zn: Zinc (mg dm^{-3}), SPR₁, SPR₂, SPR₃, SPR₄: Soil Penetration Resistance at depths of 0-10, 11-20, 21-30, and 31-40 cm, respectively (MPa).

In the geostatistical analysis, we verified that the spatial dependence structure depends only on the distance separating the locations observed, and does not differ with the direction. Therefore, the spatial dependence structure can be considered isotropic for all the physical-chemical attributes of the soil (Guedes et al., 2013).

Comparing the original and reduced sampling designs, we found that the chemical attributes of the soil showed greater variation in the estimated values of the parameters of the spatial dependence structure than the physical ones (Table 2).

For most soil attributes, the standard error value of the parameters was that of the spatial dependence structure increased after sample resizing (Table 2). Also, the estimated standard error values display the same magnitude as the estimated parameter (Table 2).

All the soil attributes (except zinc) showed a reduction in the estimated value of the spatial dependence radius, having, on average, a practical range 27% smaller with the reduced sample configuration (Table 2). However, even with a smaller radius of spatial dependence, most of the soil attributes showed a higher intensity of spatial dependence (*RNE*) (Table 2).

Finally, there was a lower value of the variance inflation factor (VIF) (Griffith, 2008) in 80% of the physical-chemical attributes of the soil when the reduced sample configuration was used (Table 2).

Table 2. Estimated values of the parameters and their respective standard errors (in parentheses) that define the spatial dependence structure of the physical-chemical attributes of the soil with

the original ($n = 102$) and reduced ($n^* = 53$) sample configurations

			$\hat{\mu}$	$\hat{\varphi}_1$	$\hat{\varphi}_2$	$\hat{\varphi}_3$	$\hat{\alpha}$	\widehat{RNE}	VIF
Ca	Gaus	n	5.53	1.29	0.48	0.16	0.284	72.88	1.77
			(0.18)	(0.36)	(0.36)	(0.08)			
		n^*	5.47	0.55	1.15	0.12	0.204	32.35	1.70
			(0.21)	(0.43)	(0.55)	(0.04)			
C	Gaus	n	31.86	8.02	3.39	0.31	0.539	70.28	11.41
			(0.70)	(1.36)	(1.80)	(0.11)			
		n^*	32.39	8.25	1.09	0.30	0.526	88.33	9.34
			(0.55)	(2.17)	(1.74)	(0.36)			
Cu	Exp	n	4.03	0.34	0.52	0.28	0.855	39.53	0.86
			(0.27)	(0.14)	(0.22)	(0.19)			
		n^*	3.69	0.11	0.71	0.11	0.322	13.41	0.82
			(0.18)	(0.37)	(0.42)	(0.07)			
Mn	Gaus	n	86.78	268.79	226.14	0.21	0.367	54.31	494.93
			(4.16)	(59.89)	(89.74)	(0.06)			
		n^*	87.77	319.45	236.98	0.18	0.308	57.41	556.43
			(4.45)	(140.85)	(161.97)	(0.08)			
Zn	Gaus	n	5.10	1.59	3.05	0.22	0.368	34.27	4.64
			(0.45)	(0.38)	(0.95)	(0.04)			
		n^*	4.72	1.58	0.62	0.23	0.404	71.82	2.20
			(0.28)	(0.52)	(0.54)	(0.14)			
SPR ₁	Gaus	n	3.11	0.64	0.29	0.15	0.255	68.82	0.93

		(0.13)	(0.21)	(0.22)	(0.07)			
	n^*	3.01	0.49	0.30	0.14	0.235	62.02	0.79
		(0.14)	(0.25)	(0.27)	(0.08)			
	n	3.00	0.01	0.36	0.03	0.101	2.70	0.37
		(0.07)	(1.16)	(1.17)	(0.06)			
SPR ₂	Exp	n^*	3.07	0.00	0.43	0.03	0.087	0.00
			(0.09)	(3.86)	(3.86)	(0.15)		
	n	2.02	0.03	0.07	0.07	0.116	30.00	0.10
		(0.04)	(0.08)	(0.08)	(0.07)			
SPR ₃	Gaus	n^*	2.03	0.00	0.08	0.05	0.080	0.00
			(0.04)	(0.39)	(0.39)	(0.11)		
	n	2.07	0.06	0.07	0.07	0.221	46.15	0.13
		(0.05)	(0.10)	(0.10)	(0.07)			
SPR ₄	Exp	n^*	2.07	0.00	0.11	0.02	0.071	0.00
			(0.04)	(5.27)	(5.27)	(0.37)		

Ca: calcium (cmol dm⁻³), C: carbon (g dm⁻³), Cu: copper (mg dm⁻³), Mn: manganese (mg dm⁻³), Zn: zinc (mg dm⁻³), SPR₁, SPR₂, SPR₃, SPR₄: soil penetration resistance at depths of 0-10, 11-20, 21-30, and 31-40 cm, respectively (MPa), $\hat{\mu}$: mean, $\hat{\varphi}_1$: nugget effect, $\hat{\varphi}_2$:

partial sill, $\hat{\varphi}_3$: function of the range, $\hat{\alpha}$: practical range (kilometers); \widehat{RNE} : Relative Nugget Effect (%); VIF: Variance Inflation Factor. Geostatistical models: Gaus: gaussian, Exp: exponential.

With the original sampling configuration, there was a sample every 1.7 hectares. Considering the sample resizing obtained and the size of the experimental agricultural area, approximately one sample would be collected every 3 hectares. As a result, both the visual analysis of the thematic maps of the soil attributes, as well as the quantitative results, signaled some differences between the original and reduced sampling grids.

For Soil Penetration Resistance (SPR), medium to high accuracy was found between maps with the original and reduced sample configurations, at depths of 11-20 cm, 21-30 cm, and 31-40 cm ($T \geq 67\%$; Ma and Redmond, 1995; De Bastiani et al., 2012) (Figure 5 – G, H, and I). For the chemical attributes of the soil, lower accuracy indexes were found ($T < 67\%$; Ma and Redmond, 1995) (Figure 5 – A, B, C, D, and E).

However, we observed that, even with a 40% reduction in the sample size, the thematic maps

obtained with the resized sample configuration maintained the pattern of spatial variability described by the maps with the original sample size. This can be observed through the points that represent the spatial distribution of the contents of the macro- and micro- nutrients in the soil, as well as at a depth of 0-10 cm of the SPR, according to each class of the thematic map (Figure 5 – A, B, C, D, E, and F). The pattern was also maintained at depth layers between 11 and 40 cm of the SPR (Figure 5 – G, H, and I), where the maps presented circular regions around the sample points.

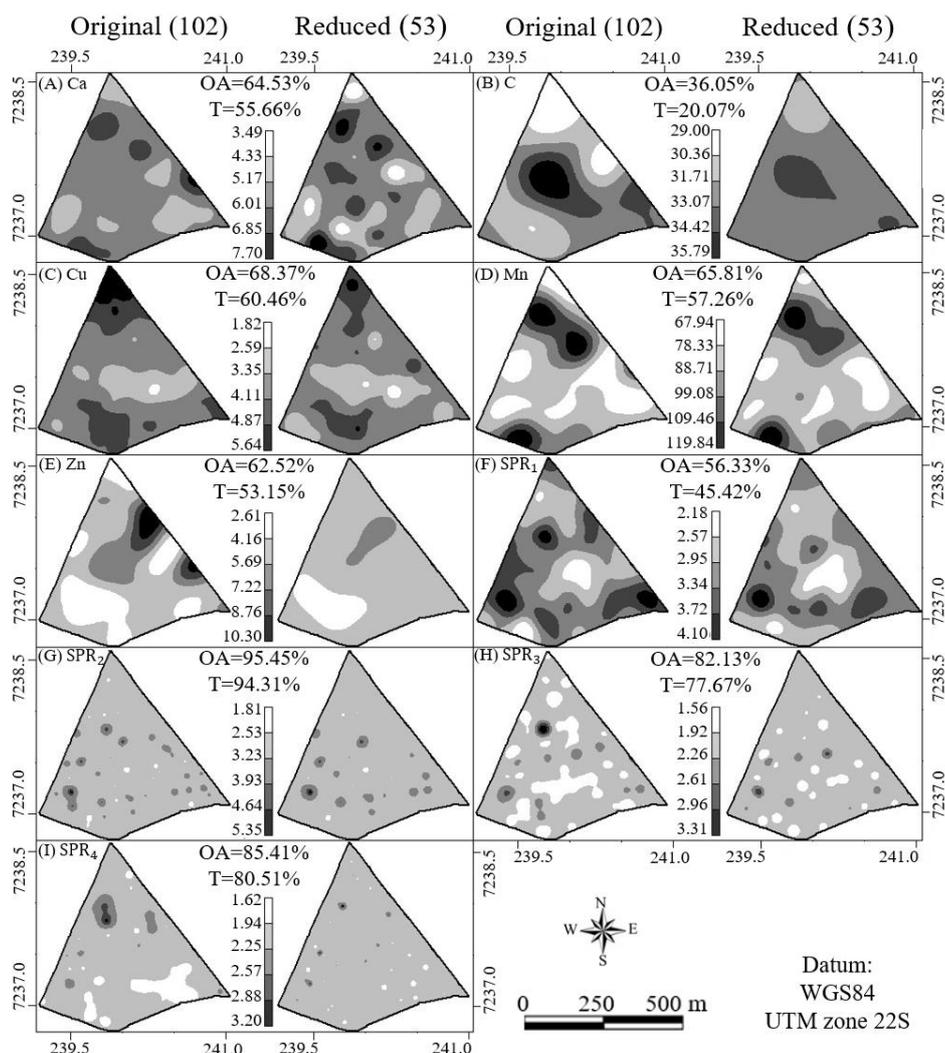


Figure 5. Thematic maps of: (A) calcium (cmol dm^{-3}), (B) carbon (g dm^{-3}), (C) copper (mg dm^{-3}), (D) manganese (mg dm^{-3}), (E) zinc (mg dm^{-3}) and soil penetration resistance at depths of (F) 0-10 cm, (G) 11-20 cm, (H) 21-30 cm, (I) 31-40 cm (MPa), considering the original and reduced sampling designs. Estimated values of OA (Overall Accuracy) and T (Tau concordance index)

4. Discussion

4.1 Analysis of Simulations Studies

Comparing these three groups formed by the simulated variables, we observed an inverse

relationship between the radius of spatial dependence and the estimated values of ESS_{uni} . In other words, the greater the radius of spatial dependence, the lower the estimated value of ESS_{uni} , thus the greater the reduction in the number of sample points. Even considering another sample configuration, results similar to those obtained were found by Vallejos & Osorio (2014). Moreover, when fixing the value of the spatial dependence radius and varying the value of the nugget effect, and consequently changing the intensity of spatial dependence, no relevant changes were verified in the interval given by the (mean \pm Standard Deviation) of the estimated value of ESS_{uni} (Figure 3 – A).

The spatial dependence radius also proved to be a parameter that inversely influences the estimated values of the univariate effective sample size (ESS_{uni}) and of the multivariate effective sample size (ESS_{multi}).

4.2 Analysis of the Physical-Chemical Attributes of the Soil

The results showed that the greater the radius of spatial dependence, the lower the estimated value of the effective univariate sample size (ESS_{uni}). This result corroborates those obtained in the simulation studies of the present study. As well as in Griffith (2005) and in Vallejos & Osorio (2014), who used different geostatistical models to explore the spatial variability of the biomass, arsenic, and lead index in the soil and to resize the sample size, and found the same inverse relationship between the radius of spatial dependence and ESS_{uni} . Furthermore, when the practical reach values were close, the intensity of the spatial dependence had a direct influence, that is, the higher the RNE, the greater the estimated value of ESS_{uni} .

The reduction of the sample size to 53 points represents a reduction of approximately 47% of the original sample size. This result corroborates with studies carried out by Griffith (2005, 2008) in which, considering soil macro- and micro- nutrients and calculating the univariate effective sample size, sample reductions were obtained which varied between 30% and 48%. Also, this new sample size is similar to those used by Pelissari et al. (2014) and by Siqueira et al. (2014), who considered between 46 and 60 sample points for the analysis of the spatial variability of physical-chemical attributes in experimental areas, respectively; with a smaller dimension (less than 100 ha) and a similar size (150 ha) compared to this research.

Comparing the original and reduced sampling designs (Table 2), the greater variation in the values of the spatial dependence parameters of the chemical attributes in relation to the physical ones can be explained by the agricultural practices of soil and crop management, as well as by the natural variability of the soil. Both factors contribute to the chemical attributes having greater spatial variability than the physical ones, therefore being more sensitive to the reduction in the number of sample elements (Jacob and Klutle, 1956).

The increase in the standard error of the parameters after sample resizing indicates that the reduction in the number of sample points influenced the spatial dependence structure of the physical-chemical attributes of the soil (Table 2). Moreover, regarding the standard error values following the magnitude of the estimated parameter, this characteristic is also

perceived in the results of Schemmer et al. (2017) and Fagundes et al. (2018), who used in their studies both the Gaussian linear spatial model (as well as this study), and the Slash and t-Student linear spatial models, applied to variables related to soil and plants.

Although the results indicated that sample resizing provided changes in the spatial dependence structure of the physical-chemical attributes of the soil; the lower values of the VIF (Table 2) showed that the soil attributes were less affected by the spatial autocorrelation, after the reduction in the number of sample points. This suggests that the original sample configuration presented sample elements with redundant spatial information.

The lower accuracy in the maps of the chemical attributes of the soil in relation to the physical ones (Figure 5) agrees with several studies about spatial variability carried out with chemical attributes of the soil using different sample sizes (from 9 to 164 samples) and sample densities (from 25x25 to 173x173), in different Brazilian states, therefore susceptible to different climates and soil types, and with other crops besides soy (corn and oats) (Cherubin et al. 2014, 2015; Kestring et al., 2015; Guarçoni et al., 2017). This is because the chemical attributes of the soil are more sensitive than the physical ones (therefore, than the SPR), concerning the description of spatial variability with a reduced number of samples (Jacob and Klutle, 1956). Furthermore, the chemical attributes of the soil are often corrected by the application of inputs, during and between harvests.

Regarding the circular regions around the sample points (Figure 5 – G, H, and I), this is a phenomenon known as ‘bull eyes effect’ (Menezes et al., 2016) and is justified by the small dependence radius value in these depth layers of SPR in the sample grids, both the original and the reduced ones (Table 2).

The thematic maps with the reduced configuration maintained the standard in spatial variability (Figure 5); this implies that it is possible to continue carrying out soil management and correction in the necessary places, however using a smaller number of samples collected and without having equipment with a high cost of acquisition and maintenance, such as machinery with a harvest monitor.

Therefore, this work showed that, with a single reduction in the number of sample points obtained by proposing the effective multivariate sample size (ESS_{multi}), it was possible to characterize the spatial dependence of the agricultural area. Also, the methodology implemented showed that there was collection of spatially redundant information in the area, which in future collections would imply unnecessary costs to the production process. We used physical-chemical attributes of the soil with different spatial dependence structures and obtained a considerable resizing of the sample, from 102 to 53 sample elements. In the practice, reducing the number of samples by approximately 50% implies a reduction in costs by the same proportion, as it will reduce the demand for time, labor, and laboratory analysis.

Acknowledgement

The authors gratefully acknowledge the funding received from the Coordination for the Improvement of Higher Education Personnel (CAPES), Financing Code 001, and from the

National Council for Scientific and Technological Development (CNPq). The authors would also like to thank the Spatial Statistics Laboratory - UNIOESTE - Cascavel, PR, Brazil.

References

- Acosta, J., & Vallejos, R. (2018). Effective sample size for spatial regression models. *Electronic Journal of Statistics*, 12, 3147-3180. <https://doi.org/10.1214/18-EJS1460>
- Acosta, J., Osorio, F., & Vallejos, R. (2016). Effective sample size for line transect sampling models with an application to marine macroalgae. *Journal of Agricultural, Biological and Environmental Statistics*, 21(3), 407-425. <https://doi.org/10.1007/s13253-016-0252-7>
- Acosta, J., Vallejos, R., & Griffith, D. (2018). On the effective geographic sample size. *Journal of Statistical Computation and Simulation*, 88(10), 1958-1975. <https://doi.org/10.1080/00949655.2018.1428977>
- Anderson, J. R., Hardy, E. E., Roach, J. T., & Witmer, R. E. (2001). *A land use and land cover classification system for use with remote sensor data*. U.S. Washington (DC): Government Print Office. p. 28.
- Balbinot Junior, A. A., Hiraçuri, M. H., Franchini, J. C., Debiasi, H., & Ribeiro, R. H. (2017). Análise da área, produção e produtividade da soja no Brasil em duas décadas (1997-2016). *Boletim de Pesquisa e Desenvolvimento*, n. 11. Londrina, PR: Embrapa Soja. [Online] Available: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/156652/1/Boletim-de-PD-11.pdf> (July 31, 2020).
- Benedetti, R., Piersimoni, F., & Postiglione, P. (2015). *Sampling spatial units for agricultural surveys*. Berlin: Springer. p. 343. <https://doi.org/10.1007/978-3-662-46008-5>
- Bolfe, É. L., Victória, D. D. C., Contini, E., Bayma-Silva, G., Spinelli-Araujo, L., & Gomes, D. (2016). Matopiba em crescimento agrícola: aspectos territoriais e socioeconômicos. *Revista de Política Agrícola*, 25(4), 38-62.
- Cambardella, C. A., Moorman, T. B., Novak, J. M., Parkin, T. B., Karlen, D. L., Turco, R. F., & Konopka, A. E. (1994). Field-scale variability of soil properties in Central Iowa Soils. *Soil Science Society of America Journal*, 58(5), 1501-1511. <https://doi.org/10.2136/sssaj1994.03615995005800050033x>
- Canarache, A. (1991). Factors and indices regarding excessive compactness of agricultural soils. *Soil & Tillage Research*, 19(2-3), 145-164. [https://doi.org/10.1016/0167-1987\(91\)90083-A](https://doi.org/10.1016/0167-1987(91)90083-A)
- Cherubin, M. R., Santi, A. L., Eitelwein, M. T., Amado, T. J. C., Simon, D. H., & Damian, J. M. (2015). Dimensão da malha amostral para caracterização da variabilidade espacial de fósforo e potássio em Latossolo Vermelho. *Pesquisa Agropecuária Brasileira*, 50(2), 168-177. <https://doi.org/10.1590/S0100-204X2015000200009>
- Cherubin, M. R., Santi, A. L., Eitelwein, M. T., Menegol, D. R., Da Ros, C. O., Pias, O. H. C.,

& Berghetti, J. (2014). Eficiência de malhas amostrais utilizadas na caracterização da variabilidade espacial de fósforo e potássio. *Ciência Rural*, 44(3), 425-432. <https://doi.org/10.1590/S0103-84782014000300007>

Colombi, T., & Keller, T. (2019). Developing strategies to recover crop productivity after soil compaction - A plant eco-physiological perspective. *Soil & Tillage Research*, 191, 156-161. <https://doi.org/10.1016/j.still.2019.04.008>

CONAB, Companhia Nacional de Abastecimento (2020). Grãos – Série Histórica. [Online] Available: <https://portaldeinformacoes.conab.gov.br/safra-serie-historica-dashboard> (April 16, 2020).

Cressie, N. A. C. (2015). *Statistics for Spatial Data*. (rev. ed.). New York: Wiley. p. 928.

De Bastiani, F., Uribe-Opazo, M. A., & Dalposso, G. H. (2012). Comparison of maps of spatial variability of soil resistance to penetration constructed with and without covariables using a spatial linear model. *Engenharia Agrícola*, 32, 393-404. <https://doi.org/10.1590/S0100-69162012000200019>

Debiasi, H., Franchini, J. C., Conte, O., Balbinot Junior, A. A., Torres, E., Saraiva, O. F., & Oliveira, M. C. N. (2013). Sistemas de preparo do solo: trinta anos de pesquisas na Embrapa Soja. Documentos, n. 342. Londrina, PR: Embrapa Soja. [Online] Available: <https://www.infoteca.cnptia.embrapa.br/bitstream/doc/970601/1/Doc342OL.pdf> (July 31, 2020).

Diggle, P., & Lophaven, S. (2006). Bayesian geostatistical design. *Scandinavian Journal of Statistics*, 33(1), 53-64. <https://doi.org/10.1111/j.1467-9469.2005.00469.x>

Diggle, P., & Ribeiro Jr, P. J. (2007). *Model-Based Geostatistics*. Lancaster: Springer. p. 228. <https://doi.org/10.1007/978-0-387-48536-2>

Fagundes, R. S., Uribe-Opazo, M. A., Guedes, L. P. C., & Galea, M. (2018). Slash spatial linear modeling: soybean yield variability as a function of soil chemical properties. *Revista Brasileira de Ciência do Solo*, 42, e0170030. <https://doi.org/10.1590/18069657rbcs20170030>

Faraco, M. A., Uribe-Opazo, M. A., Silva, E. A. A., Johann, J. A., & Borssoi, J. (2008). Seleção de modelos de variabilidade espacial para elaboração de mapas temáticos de atributos físicos do solo e produtividade da soja. *Revista Brasileira de Ciência do Solo*, 32(2), 463-476. <https://doi.org/10.1590/S0100-06832008000200001>

Gao, X., Zhao, X., Wu, P., Brocca, L., & Zhang, B. (2016). Effects of large gullies on catchment-scale soil moisture spatial behaviors: A case study on the Loess Plateau of China. *Geoderma*, 261, 1-10. <https://doi.org/10.1016/j.geoderma.2015.07.001>

Griffith, D. A. (2005). Effective geographic sample size in the presence of spatial autocorrelation. *Annals of the American Association of Geographers*, 95(4), 740-760. <https://doi.org/10.1111/j.1467-8306.2005.00484.x>

Griffith, D. A. (2008). Geographic sampling of urban soils for contaminant mapping: how

many samples and from where. *Environmental Geochemistry and Health*, 30(6), 495-509. <https://doi.org/10.1007/s10653-008-9186-5>

Grinand, C., Arrouays, D., Laroche, B., & Martin, M. P. (2008). Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. *Geoderma*, 143, 180-190. <https://doi.org/10.1016/j.geoderma.2007.11.004>

Guarçoni, A., Alvarez, V. V. H., & Sobreira, F. M. (2017). Fundamentação teórica dos sistemas de amostragem de solo de acordo com a variabilidade de características químicas. *Terra Latinoamericana*, 35(4), 343-352. <https://doi.org/10.28940/terra.v35i4.172>

Guedes, L. P. C., Ribeiro Jr, P. J., Uribe-Opazo, M. A., & De Bastiani, F. (2016). Soybean yield maps using regular and optimized sample with different configurations by simulated annealing. *Engenharia Agrícola*, 36, 114-125. <http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v36n1p114-125/2016>

Guedes, L. P. C., Uribe-Opazo, M. A., & Ribeiro Jr, P. J. (2013). Influence of incorporating geometric anisotropy on the construction of thematic maps of simulated data and chemical attributes of soil. *Chilean Journal of Agricultural Research*, 73, 414-423. <https://doi.org/10.4067/S0718-58392013000400013>

Jacob, W. C., & Klute, A. (1956). Sampling soils for physical and chemical properties. *Soil Science Society of America Journal*, 20(2), 170-172. <https://doi.org/10.2136/sssaj1956.03615995002000020008x>

Kestring, F., Guedes, L. P. C., De Bastiani, F., & Uribe-Opazo, M. A. (2015). Thematic maps comparison of different sampling grids for soybean productivity. *Engenharia Agrícola*, 35, 733-743. <https://doi.org/10.1590/1809-4430-Eng.Agric.v35n4p733-743/2015>

Ma, Z., & Redmond, R. L. (1995). Tau coefficients for accuracy assessment of classification of remote sensing data. *Photogrammetric Engineering and Remote Sensing*, 61(4), 453-439.

Maltauro, T. C., Guedes, L. P. C., & Uribe-Opazo, M. A. (2019). Reduction of sample size in the analysis of spatial variability of nonstationary soil chemical attributes. *Engenharia Agrícola*, 39, 56-65. <https://doi.org/10.1590/1809-4430-eng.agric.v39nep56-65/2019>

Menezes, M. D, Silva, S. H. G., Mello, C. R., Owens, P. R., & Curi, N. (2016). Spatial prediction of soil properties in two contrasting physiographic regions in Brazil. *Scientia Agrícola*, 73(3), 274-285. <https://doi.org/10.1590/0103-9016-2015-0071>

Novomestky, F. (2012). matrixcalc: collection of functions for matrix calculations. Version 3.3.1. New York: R Package Manual. [Online] Available: <https://cran.r-project.org/web/packages/matrixcalc/matrixcalc.pdf> (March 18, 2020).

Pelissari, A. L., Caldeira, S. F., & Santos, V. S. (2014). Variabilidade espacial dos atributos químicos do solo em povoamento de *Tectona grandis*. *Cerne*, 20(3), 377-384. <https://doi.org/10.1590/01047760201420031437>

- R Development Core Team. (2020). R: A language and environment for statistical computing. Version 4.0.0. Vienna, Austria: R Foundation for Statistical Computing, ISBN 3-90005107-0. [Online] Available: <https://www.r-project.org/> (March 16, 2020).
- Ribeiro Jr, P. J., & Diggle, P. J. (2001). geoR: a package for geostatistical analysis. *R News*, 1, 15-18.
- Sanches, G. M., Magalhães, P. S. G., & Franco, H. C. J. (2019). Site-specific assessment of spatial and temporal variability of sugarcane yield related to soil attributes. *Geoderma*, 334, 90-98. <https://doi.org/10.1016/j.geoderma.2018.07.051>
- Schemmer, R. C., Uribe-Opazo, M. A., Galea, M., & Assumpção, R. A. B. (2017). Spatial variability of soybean yield through a reparametrized t-Student model. *Engenharia Agrícola*, 37, 760-770. <https://doi.org/10.1590/1809-4430-eng.agric.v37n4p760-770/2017>
- Siqueira, D. S., Marques Jr, J., Pereira, G. T., Barbosa, R. S., Teixeira, D. B., & Peluco, R. G. (2014). Sampling density and proportion for the characterization of the variability of Oxisol attributes on different materials. *Geoderma*, 232, 172-182. <https://doi.org/10.1016/j.geoderma.2014.04.037>
- Taiz, L., Zeiger, E., Møller, I. M., & Murphy, A. (2017). *Fisiologia e desenvolvimento vegetal*. (6th ed.). São Paulo: Artmed Editora. p. 888.
- Uribe-Opazo, M. A., Borsoi, J. A., & Galea, M. (2012). Influence diagnostics in Gaussian spatial linear models. *Journal of Applied Statistics*, 39, 615-630. <https://doi.org/10.1080/02664763.2011.607802>
- Vallejos, R., & Osorio, F. (2014). Effective sample size of spatial process models. *Spatial Statistics*, 9, 66-92. <https://doi.org/10.1016/j.spasta.2014.03.003>
- Wadoux, A. M. J. C., Brus, D. J., Rico-Ramirez, M. A., & Heuvelink, G. B. M. (2017). Sampling design optimization for rainfall prediction using a non-stationary geostatistical model. *Advances in Water Resources*, 107, 126-138. <https://doi.org/10.1016/j.advwatres.2017.06.005>

Copyright Disclaimer

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).