

Doi: <http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v40n1p96-104/2020>**NUGGET EFFECT INFLUENCE ON SPATIAL VARIABILITY OF AGRICULTURAL DATA****Luciana P. C. Guedes^{1*}, Raquel T. Bach¹, Miguel A. Uribe-Opazo¹**^{1*}Corresponding author. Western Paraná State University/ Cascavel - PR, Brasil.E-mail: Luciana_pagliosa@hotmail.com | ORCID ID: <https://orcid.org/0000-0001-6631-2478>**KEYWORDS**

Geostatistics, Spatial Dependence Index (SDI), Relative Nugget Effect (RNE), Spatial Estimation.

ABSTRACT

Spatial variability description of soil chemical properties by thematic maps depends substantially on suitable geostatistical models. One of the parameters composing a geostatistical model is nugget effect. This study aimed to evaluate the simultaneous influence of nugget effect and sampling design on geostatistical model estimation and estimation of soil chemical properties at unsampled sites, considering simulated data. Our results will be used as scientific basis for spatial variability analyses of soil chemical properties in agricultural areas. Given the simulation results and agricultural data, we concluded that the high nugget effect values obtained here reduced spatial estimation efficiency. Moreover, a systematic sampling design promoted the least accurate estimates of geostatistical model and at non-sampled sites. Despite that, these nugget effect estimates should be kept in the analysis. However, further studies will be needed to investigate which factors are responsible for such high nugget effect values.

INTRODUCTION

Spatial analysis of a georeferenced variable using geostatistical models enables measuring the spatial dependence degree among samples within a determined area, thus describing its spatial dependence structure (Guedes et al., 2018). The spatial dependence analysis, mainly of soil chemical properties in farmlands, allows to know their values in subregions (management zones) within the area of interest, which, in turn, enables the application of agricultural inputs at specific points (Gazolla-Neto et al., 2016).

The spatial dependence structure of a certain georeferenced variable should be described considering a stochastic process, whose data are expressed by $Z(s_1)$, $Z(s_2)$, ..., $Z(s_n)$, which are known in n sites s_i ($i = 1, \dots, n$), being that $s_i = (x_i, y_i)^T$ is a two-dimensional vector. The georeferenced variable can be expressed by a Gaussian spatial linear model: $Z(s_i) = \mu(s_i) + \varepsilon(s_i)$ (Uribe-Opazo et al., 2012), in which $\mu(s_i) = \mu$ is the deterministic term, μ is a constant, and $\varepsilon(s_i)$ represents the stochastic term with mean zero, i.e., $E[\varepsilon(s_i)] = 0$; and variation between points in space separated by an Euclidean distance $h_{ij} = \|h\|$, so that $h = s_i - s_j$ ($i, j = 1, \dots, n$) is determined by a covariance function: $C(h_{ij}) = \text{cov}[\varepsilon(s_i), \varepsilon(s_j)] = \sigma_{ij}$, which depends only on h . Moreover, a covariance matrix is found from the covariance function, as follows: $\Sigma = [(\sigma_{ij})]$ (Dalposso et al., 2018; Guedes et al., 2011; 2018).

One of the functions used to describe a spatial dependence structure is referred to as a semivariance function $\gamma(h_{ij})$. It measures dissimilarity between values sampled at sites separated by a distance h_{ij} , for stationary and isotropic processes, and is associated with the covariance function by $\gamma(h_{ij}) = C(0) - C(h_{ij})$ (Uribe-Opazo et al., 2012).

In the literature, different theoretical models have been developed to define spatial dependence structure (semivariance function), and several methods have been used to estimate these models (Cressie, 2015; Monego et al., 2015; Cortés-D et al., 2016). The model estimated for semivariance function has the following parameters: range (a), partial sill (C_1), nugget effect (C_0), and sill ($C_0 + C_1$). Range is the longest distance between sampling sites spatially correlated within an area. Sill is the semivariance value when distance equals range, corresponding to the variance of a georeferenced variable (Kestring et al., 2015).

Semivariance function has a minimum distance (h_{min}), within which its semivariance value ($\gamma(h_{ij})$) is calculated. When it is high, the phenomenon has high variability over a small range of distances. In these situations, the semivariance function has a parameter called nugget effect (Peng & Wu, 2014; Cressie, 2015; Genton & Kleiber, 2015). Small-scale variability may be associated with features of the studied process and/ or

¹ Western Paraná State University/ Cascavel - PR, Brasil.

Area Editor: Teresa Cristina Tarlé Pissarra

Received in: 6-16-2016

Accepted in: 10-1-2019



measurement errors. It may also occur due to data heterogeneity or sampling size and scheme (BOSSEW et al., 2014; Seidel & Oliveira, 2014; Lark & Marchant, 2018; Wadoux et al., 2019).

Vallejos & Osorio (2014), Cressie (2015), Bassani et al. (2018) suggested a relationship between nugget effect and (a) features related to spatial prediction, (b) a structure that describes spatial dependence and (c) sample design. Chipeta et al. (2016) described that when nugget effect is not zero, a sampling design with closer pairs of sampling points should be considered to improve geostatistical model estimation and spatial prediction.

However, there is a gap about the implications of nugget effect and sample design on geostatistical model estimation and spatial prediction simultaneously (expressed by thematic maps). This is especially true if the evaluation considers the simultaneous influence of nugget effect and sampling design.

Thus, the goals of this study were: 1) to evaluate the influence of nugget effect on geostatistical model estimation and estimation at non-sampled sites of a georeferenced variable, using Monte Carlo simulated data and considering different sample designs (random, systematic, and lattice plus close pairs); 2) to analyze the spatial variability of the following soil chemical properties: carbon, calcium, magnesium, and pH, considering the sampling data of an area whose sampling design was a lattice plus close pairs.

MATERIAL AND METHODS

Simulated datasets originated from multivariate stochastic processes, assuming stationary variables, with an isotropic Gaussian linear model (Uribe-Opazo et al., 2012). Sampling designs with 100 sampling points arranged in a regular area with a maximum limit of coordinates equal to 100 m were considered. Three sample designs were simulated: systematic (lattice) 10 × 10 grid, random and lattice 19 × 19 grid added by 19 randomly chosen nearby points (lattice plus close pairs). Sample size and the latter sampling design were chosen due to the design used in the study area.

Twelve trials were considered for each sample design, totaling 36 sets of simulations. For each set of simulation were considered 100 simulations, totaling 3600 simulations for every trials. Each simulation attempt was made considering an exponential model for semivariate function, with fixed parameters for a practical range equal to 60 m and a sill of 10 m. Nugget effect was the parameter that varied between trials. Thus, 12 trials were performed considering the following nugget effect values: 0, 1, 2, 2.5, 3, 4, 5, 6, 7, 7.5, 8, and 9.

Semivariance function parameters were estimated for each simulation by maximum likelihood and respective asymptotic standard errors were calculated (De Bastiani et al., 2018). Moreover, the following measures that quantify the spatial dependence intensity were estimated: relative nugget effect (*RNE*) (Equation 1) and spatial dependence index (*SDI*) (Equation 2).

$$RNE = \left(\frac{C_0}{C_0 + C_1} \right) \cdot 100 \quad (1)$$

$$SDI = MF \cdot \left(\frac{C_1}{C_0 + C_1} \right) \cdot \left(\frac{a}{q \cdot MD} \right) \cdot 100 \quad (2)$$

in which: *MF* is the model factor and reflects the spatial dependence strength of each model (for exponential, spherical, and gaussian models, *MF* values are 0.317, 0.375, and 0.504 respectively); *C*₀ is the nugget effect; *C*₀ + *C*₁ is the sill; *a* is the practical range; and *q*·*MD* is the fraction (*q*) of the maximum distance (*MD*) between sample points. In this study, we assumed *q* as 50% of the maximum distance. *RNE* value was proposed by Cambardella et al. (1994), which describes the proportion of sill represented by the nugget effect. Yet, *SDI* was proposed by Seidel & Oliveira (2014) and includes a greater amount of information in its calculation compared to the *RNE* (nugget effect, sill, practical range, and semivariance function model).

Scale of values of *RNE* and *SDI* indexes are different from each other, as well as their interpretation. While *RNE* ranges from 0 to 100% for all semivariance function models, *SDI* depends on semivariance function model. According to Seidel & Oliveira (2014), for exponential, spherical, and Gaussian functions, *SDI* varies respectively from 0 to 31.7%, from 0 to 37.5%, and from 0 to 50.4%. Regardless of the model describing semivariance function, the closer the *SDI* to its maximum value, the greater the spatial dependence of the variable under study.

Considering the estimated geostatistical models, the values of the georeferenced variables at unsampled sites were estimated using the kriging method. Spatial estimation mean variance or kriging variance ($\overline{\sigma_0^2}$) was calculated since it is an estimation efficiency measure, wherein the lower its value, the better the spatial estimation efficiency (Cressie, 2015; Kleijnen & Mehdad, 2016).

Simulations using nugget effect values from 0 to 8 were compared to those with nugget effect equal to 9. This comparison encompassed the following measures: the sum of squared difference between spatial estimates and accuracy measure, known as overall accuracy (OA) (De Bastiani et al., 2012). Comparison methods were chosen to compare georeferenced spatial estimates with those generated with a lower degree of spatial dependence.

We also analyzed the spatial variability of a set of real data from a commercial grain area of 167.35 ha. The area is located in the city of Cascavel, western Paraná State, Brazil. The geographical coordinates are the following: 24.95° S latitude, 53.37° W longitude, and 650 m above sea level. Local soil is classified as Oxisol, with clayey texture and deep layers of good water storage capacity, porosity, and permeability (De Bastiani et al., 2012). Local climate is very wet and classified as mesothermal, *Cfa* (Köppen), with an average annual temperature of 21°C (Kestring et al., 2015).

A lattice plus close pairs sampling was performed, with a maximum distance of 141 m between sampling sites. In some sites, random samples were taken at shorter distances: 75 and 50 m between sites, resulting in a total of 102 sampling points. All samples were georeferenced and located with the aid of a signal reception device of Global Positioning System (GPS) GEOEXPLORE 3, in the UTM spatial coordinate system.

Soybeans have been grown under no-till system since 1994. We used data from the 2010/2011 crop season, which are related to the following soil chemical properties: carbon ($C - g\ dm^{-3}$), calcium ($Ca - cmol\ dm^{-3}$), magnesium ($Mg - cmol\ dm^{-3}$), and pH. The dataset of this study was acquired by routine chemical analysis, collecting a soil sample at each marked point and five subsamples of the 0.0 to 0.2 m depth range, close to the marked points, being mixed and placed in plastic bags of about 500 g, thus composing a representative sample of the portion. These samples were sent to the Laboratory of Soil Analysis of the Central Cooperative for Agricultural Research (COODETEC) for routine chemical analyses.

The best model was fit to the semivariance function for each variable under study, according to cross-validation criteria (Lu et al. 2012; Robinson et al., 2013). Asymptotic standard errors, *RNE* (Equation 1) and *SDI* (Equation 2),

were calculated for each model. Moreover, using the kriging method, thematic maps of spatial variability of variables were created for the area under study. Simulated data sets and statistical and geostatistical analyses were performed by R software (R Development Core Team, 2018), using the geoR module (Ribeiro Junior & Diggle, 2001).

RESULTS AND DISCUSSION

Simulated Data Analysis

Nugget effect estimation was the mostly influenced parameter by sample design changes in all simulations. On average, the worst results were obtained in systematic sampling, where the estimated values were more distant from the nominal value of this parameter (Table 1). Moreover, the standard error estimates of C_0 were higher (Figure 1-a) compared to those of random design and lattice plus close pairs (Figure 1-b and 1-c).

Nugget effect was overestimated in simulations using lower values of this parameter, but underestimated when higher values were used. The best estimates were achieved in random design as it was one of the sampling designs. Nugget effect estimates were on average close to the nominal value, showing less variability (Table 1) and lower standard errors (Figure 1-b).

TABLE 1. Descriptive statistics for estimates of nugget effect (C_0).

C_0 nominal value	Systematic		Random		Lattice Plus Close Pairs	
	Mean	CV (%)	Mean	CV (%)	Mean	CV (%)
0	0.297	225.670	0.068	210.730	0.254	202.170
1	5.873	69.060	0.761	74.390	0.770	125.620
2	4.531	82.690	1.600	53.850	1.396	99.400
2.5	4.472	79.270	2.028	49.120	1.724	91.300
3	4.225	78.970	2.455	46.200	2.108	82.880
4	3.546	82.080	3.368	41.270	2.820	73.150
5	2.747	91.110	4.202	41.970	3.525	67.370
6	2.162	98.960	4.961	42.380	4.369	61.430
7	1.581	113.460	5.887	40.130	5.026	59.940
7.5	1.303	125.200	6.159	41.120	5.335	60.950
8	1.058	137.860	6.589	41.850	5.606	62.170
9	0.645	167.020	7.443	44.580	6.662	53.050

CV: coefficient of variation (%).

Lattice plus close pairs was considered the second-best design in terms of nugget effect estimation, as it also showed on average nugget effect estimates closer to the nominal value (Table 1) and low standard error estimates (Figure 1-c). But unlike the previous one, this model showed a greater variability of estimates compared to those of the random design.

All sampling designs in this study generated similar range and sill estimates. In this sense, relevant differences in C_0 estimates influenced directly RNE and inversely SDI

index calculations. When considering RNE and SDI indexes and all simulations, the lowest spatial dependence was observed in systematic sampling if compared with random design and lattice plus close pairs. Estimations of RNE and SDI indexes closer to the nominal values were obtained by random design.

These results corroborate the conclusion of Kestring et al. (2015), Zhao et al. (2016), and Bussel et al. (2016), who claimed who stated about sample design and size influences on geostatistical model estimation.

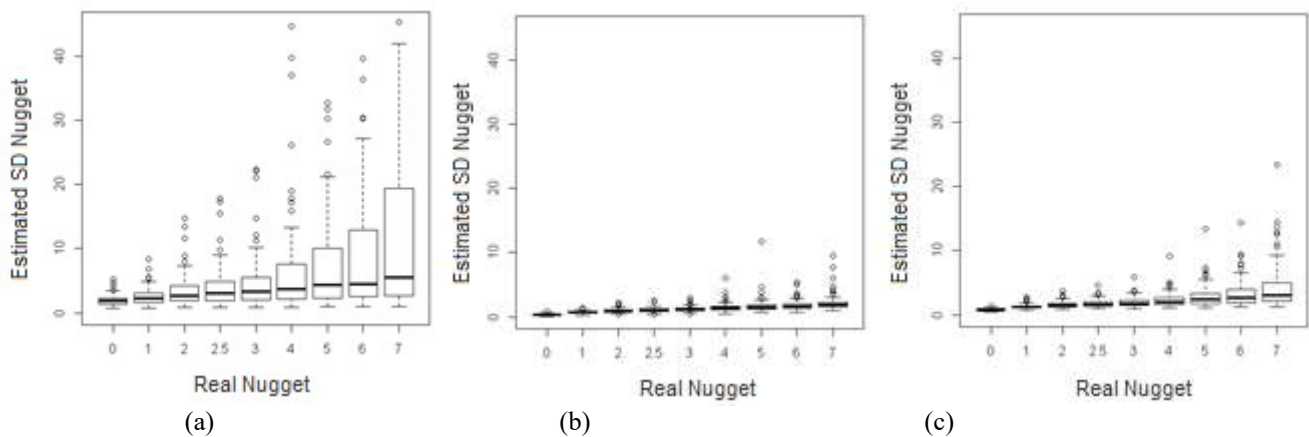


FIGURE 1. Box plot Graphs: estimates of standard deviation (or error) (SD) of nugget effect for each sampling design: (a) systematic, (b) random, and (c) lattice plus close pairs.

Table 2 also presents a descriptive summary of average kriging variance (σ_0^2) for a georeferenced variable at non-sampled sites, with different nugget effect values and sampling designs. Average kriging variance results were similar for all sampling designs and increased as nugget effect was raised.

According to Cressie (2015), georeferenced variables can be decomposed into two random terms: a second-order stationary process and a white-noise measurement process. In this case, when interpolation is done at an unsampled point, variance estimate exceeds stationary variance by the amount of white-noise measurement variance (corresponding to nugget effect variation) (Burgess & Webster, 2019). Therefore, the nugget effect and kriging variation are directly related. Average kriging variance shows how efficient a spatial estimation of unsampled sites was, thus, the smaller the spatial estimation the more efficient it is. These results (Table 2) showed that the higher the nugget effect (i.e., the lower the spatial dependence degree of a georeferenced variable), the lower the efficiency of spatial estimation by kriging.

Kriging spatial estimation results in unsampled sites for simulations with C_0 between 0 and 8 were compared to

those obtained for simulation with C_0 equal to 9 (Table 2 and Figure 2) by the sum of squared difference (SSD) (Table 2). In all sample designs, as the nugget effect increased, SSD between spatial estimates decreased.

This result indicates that the closer the nugget effect values are in a geostatistical model for spatial estimation, the more similar their results are in terms of estimation, regardless of the sampling design. The geostatistical model used for comparison ($C_0 = 9$) represents a georeferenced variable with pure nugget effect, that is, without spatial dependence. Thus, the results of SSD (Table 2) showed that the stronger the spatial dependence of a georeferenced variable (lower C_0 value), the greater the dissimilarity thereof with a georeferenced variable without spatial dependence, as far as spatial prediction is concerned.

Kriging equations depend on semivariance function, especially for nugget effect. Higher values of this parameter imply higher kriging weight for distant samples, which, in turn, produces a smoother thematic map (Bassani et al., 2018).

With C_0 changes, the systematic design showed less sensitivity to spatial estimation (lowest SSD (Table 2), underestimating nugget effect (Table 1)).

TABLE 2. Descriptive statistics for average kriging variance ($\overline{\sigma_0^2}$) and the sum of squared difference (SSD) between spatial estimations. CV is the coefficient of variation (%).

Nominal value of (C_0)	Estimated Measures	Systematic		Random		Lattice Plus Close Pairs	
		Mean	CV (%)	Mean	CV (%)	Mean	CV (%)
0	$\overline{\sigma_0^2}$	2.593	23.490	2.900	15.300	2.647	19.960
	SSD	11090	31.770	20110	43.160	20960	45.890
1	$\overline{\sigma_0^2}$	3.251	25.220	3.893	15.760	3.462	21.890
	SSD	9533	32.980	17560	45.540	18890	47.630
2	$\overline{\sigma_0^2}$	3.941	25.270	4.786	15.860	4.288	22.340
	SSD	7999	35.230	15140	48.170	16750	49.710
2.5	$\overline{\sigma_0^2}$	4.304	24.990	5.208	15.870	4.694	22.180
	SSD	7230	36.820	13990	49.620	15700	50.790
3	$\overline{\sigma_0^2}$	4.683	24.320	5.616	15.870	5.112	21.690
	SSD	6452	38.190	12890	51.170	14610	52.080
4	$\overline{\sigma_0^2}$	5.434	23.090	6.339	18.790	5.898	20.750
	SSD	4970	42.08	10770	54.710	12580	54.440
5	$\overline{\sigma_0^2}$	6.161	22.030	6.996	21.400	6.571	22.520
	SSD	3629	48.870	8778	58.950	10660	56.970
6	$\overline{\sigma_0^2}$	6.771	26.010	7.656	21.270	7.041	28.610
	SSD	2320	61.810	6978	63.240	8844	60.290
7	$\overline{\sigma_0^2}$	7.135	31.270	7.784	31.640	7.674	28.090
	SSD	1300	87.090	5326	68.010	7192	63.290
7.5	$\overline{\sigma_0^2}$	7.335	32.810	7.978	33.400	7.652	33.050
	SSD	905.700	109.580	4593	71.870	6373	65.920
8	$\overline{\sigma_0^2}$	7.160	40.890	7.580	45.270	7.328	42.580
	SSD	595.900	145.294	3925	73.030	5668	67.430
9	$\overline{\sigma_0^2}$	5.541	77.450	5.384	88.200	5.384	88.180

These results evidenced the importance of accurately modelling the nugget effect, considering its association with kriging estimation and sampling design. For high nugget effect values (concerning its sill), it is therefore recommended to consider a sample design with many short-distance sites (Chipeta et al., 2016), to minimize uncertainties in semivariance function parameters and kriging estimates (Wadoux et al., 2019).

Figure 2 shows the box plot graphs for each sampling design, with similarity measure OA by comparing spatial estimations, considering estimates from data simulated using nugget effect equals to 9 (as reference for mapping) and spatial estimation from data simulated using the other values of nugget effect (as a model maps).

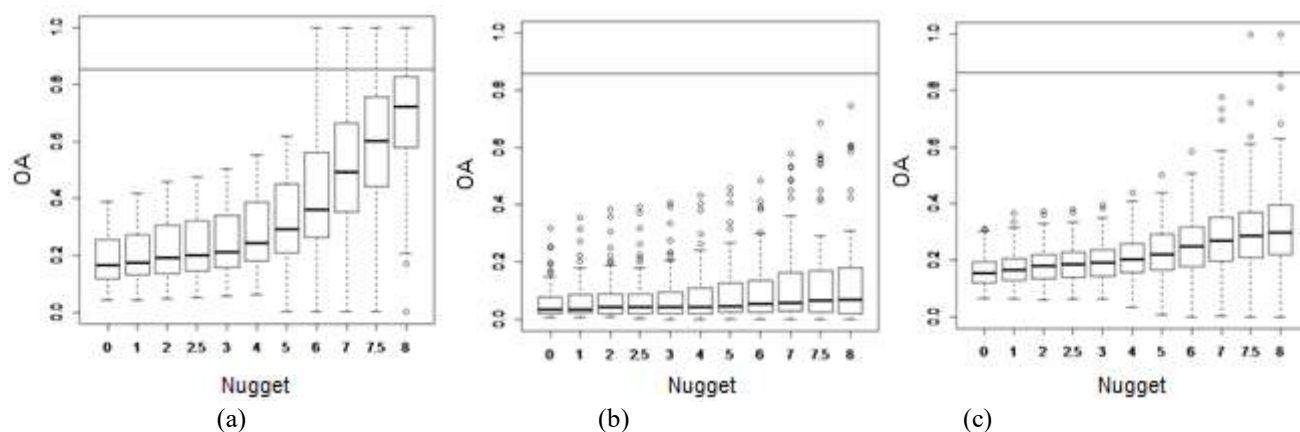


FIGURE 2. Box plot Graphs of overall accuracy (\widehat{OA}) for each sampling design: (a) systematic, (b) random, and (c) lattice plus close pairs. Dotted lines indicate ranges of minimum values of high similarity.

In both random (Figure 2-b) and lattice plus close pairs (Figures 2-c) designs, the increased nugget effect in the simulations generated a high percentage of simulations with low similarity between spatial estimates ($OA < 0.85$, De Bastiani et al., 2012). In considering $C_0 = \{3, 4, 5, 6, 7.5\}$ (moderate spatial dependence; Cambardella et al., 1994) and $C_0 = \{0, 1, 2, 2.5\}$ (strong spatial dependence; Cambardella et al., 1994), the spatial estimates of these sampling designs had low similarity in 100% of simulations with OA values.

For systematic sampling (Figures 2-a), the findings described above are valid for all simulations considering the nugget effect values that describe the georeferenced variable as having weak spatial dependence; and most nugget effect values that consider the georeferenced variable as having moderate spatial dependence.

The systematic sampling showed the highest measurements of accuracy compared with the others, especially for C_0 above 5. For $C_0 = 8$, which belongs to the same classification as spatial dependence intensity for the data simulated with nugget effect equal to 9, simulations increased (20%) with $OA \geq 0.85$, which indicates high

similarity between spatial estimations carried out at non-sampled sites (De Bastiani et al., 2012).

Based on the results, the systematic design had lower sensitivity to the spatial estimation process, with changes in nugget effect value, when compared with the random sampling and lattice plus close pairs. This is due to the poor quality of nugget effect estimation in systematic sampling (Table 1 and Figure 1), which hence produced a low quality in kriging estimation.

SPATIAL VARIABILITY ANALYSIS OF SOIL CHEMICAL PROPERTIES

Descriptive statistics for carbon (C), calcium (Ca), magnesium (Mg), and pH are given in Table 3. In analyzing these values, we can note that all parameters presented data with low dispersion and homogeneity. According to cross-validation criteria, the Gaussian model was the best-estimated model of the semivariance function for C, Ca, and Mg. As for pH, the best-estimated model was the exponential one. The spatial dependence radius for C, Ca, Mg, and pH were 254.90, 639.90, 685.47, and 300 m, respectively.

TABLE 3. Means and standard deviations (in brackets) of carbon, calcium, magnesium, and pH. Best-estimated model of the semivariance function and respective estimated parameter, estimation standard error (in brackets), spatial dependence index (SDI), and relative nugget effect (RNE).

Parameter	Mean	Model	Nugget (C_0)	Sill (C_1)	Range (a)	SDI	RNE
Carbon (C)	26.93 (3.25)	Gaussian	5.512 (1.063)	4.994 (1.993)	254.900 (64.207)	6.92%	52.47%
Calcium (Ca)	5.20 (1.37)	Gaussian	1.432 (0.214)	0.5859 (0.422)	639.900 (269.636)	10.60%	70.96%
Magnesium (Mg)	2.25 (0.69)	Gaussian	0.401 (0.059)	0.076 (0.066)	685.474 (416.130)	6.25%	84.03%
pH	5.10 (0.39)	Exponential	0.099 (0.027)	0.053 (0.033)	300.000 (264.550)	8.26%	64.87%

All parameters showed nugget effect values higher than sill ones. If we associate this with simulation results, we could notice that these high nugget effect values for all parameters generate loss of efficiency in spatial estimation by kriging. However, as reported by Webster & Oliver (2007), such values should not be disregarded because the model must be correctly estimated and incorporate the estimated nugget effect value. Furthermore, in analyzing spatial dependence degree, the parameters C, Ca, and pH are classified as with moderate ($25\% < RNE \leq 75\%$) while Mg as with weak ($RNE > 75\%$) (Cambardella et al., 1994).

Based on quartiles, the same analysis was performed for *SDI* since it makes up the strictest index to

evaluate degree of spatial dependence. This approach was used to rank C, Ca, and Mg within a weak spatial dependence and pH within a moderate degree (Seidel & Oliveira, 2014). When comparing the two indices, we found a difference for C and Ca, which is related to the range value and the estimated model.

Figure 3 displays the thematic maps of estimates for each factor. We observed that the thematic maps for carb2on (Figure 3-a) and pH (Figure 3-d) had less smoothing as for distribution of estimates in the area, whereas calcium (Figure 3-b) and magnesium (Figure 3-c) had more smoothed maps.

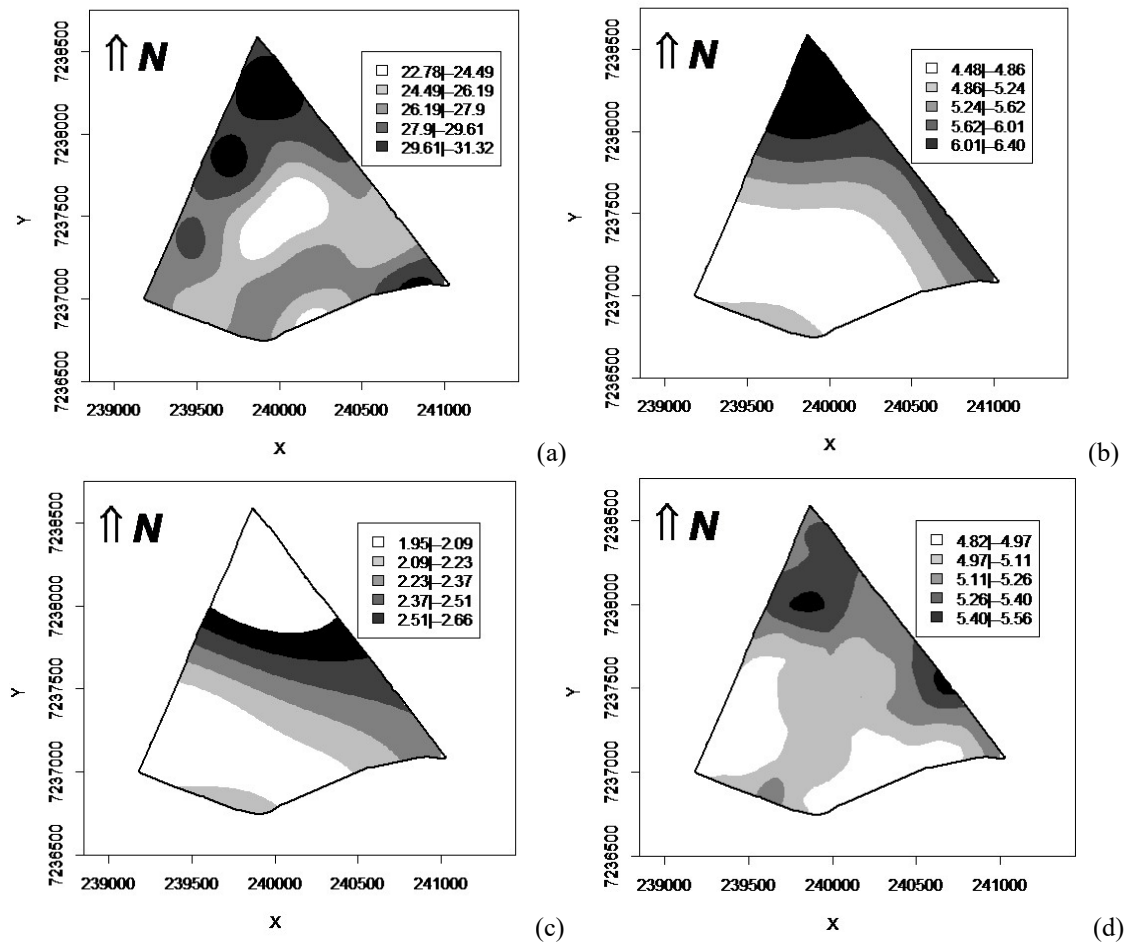


FIGURE 3. Thematic maps of: (a) carbon (C), (b) calcium (Ca), (c) magnesium (Mg), and (d) pH.

Calcium and magnesium showed weak spatial dependence and nugget effect values higher than those of sill (Table 3). Such finding emphasizes the influence of nugget effect on the spatial estimation of georeferenced parameters in non-sampled sites, and as a result, a greater thematic-map smoothing for high nugget effect values.

According to Webster & Oliver (2007) and Hofmann et al. (2010), increasing values of nugget effect provide an improved distribution of weights in spatial estimation, which hence generates smoother thematic maps. The nugget effect can be decreased by shortening gaps between samples, i.e., increasing sample density (Kestring et al., 2015); however, it is usually unfeasible for soil properties due to the costly requirements.

The spatial variability map for carbon (Figure 3-a) showed that lower carbon contents are located midwestern,

while high levels are mainly within the northern region. The thematic maps for calcium (Figure 3-b) and pH (Figure 3-d) presented reduced levels mainly within the southern region. Figure (3-c) presents the thematic map for magnesium, in which can be seen that the northern and southern regions have the lowest magnesium contents of the studied area.

CONCLUSIONS

The behavior of near-origin semivariograms described by the nugget effect strongly affects spatial estimation by kriging of georeferenced parameters at unsampled sites. Nugget effect has negative influence on stability of this type of modeling, that is, the higher the nugget effect, the lower the kriging efficiency of spatial

prediction. Nugget effect was also directly related to smoothed kriging estimates.

Systematic sampling exhibited the least accurate nugget effect estimation, the worst efficiency of spatial estimation, and the lowest degree of sensitivity to spatial changes as the nugget effect changes. Given the changes in nugget effect values, the best parameter estimates of the geostatistical model and estimates at unsampled sites occurred with the use of random design, followed by the lattice plus close pairs.

High nugget effect value was observed for all soil chemical properties. Low carbon contents were found in the midwestern region, while in the northern they were high. Calcium and pH presented the lowest values in the southern region, and the lowest magnesium levels were observed in the northern and southern regions.

ACKNOWLEDGEMENTS

The authors are grateful for the partial financial support by the Araucária Foundation of Paraná State (FA - Fundação Araucária), the Coordination of Higher-Education Personnel Improvement (CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) - Finance Code nº 001, and the National Council for Scientific and Technological Development (CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico).

REFERENCES

- Bossew P, Žunić ZS, Stojanovska Z, Tollefsen T, Carpentieri C, Veselinović N, Komatina S, Vaupotić J, Simović RD, Antignani S, Bochicchio F (2014) Geographical distribution of the annual mean radon concentrations in primary schools of Southern Serbia e application of geostatistical methods. *Journal of Environmental Radioactivity* 127:141-148. DOI: <https://doi.org/10.1016/j.jenvrad.2013.09.015>
- Bassani MAA, Costa JFCL, Guaglianoni WC, Rubio, RH (2018) Comparison between the indirect approach and kriging with samples of different support for estimation using samples of different length. *Stochastic Environmental Research and Risk Assessment* 32 (3): 785-797. DOI: <https://doi.org/10.1007/s00477-017-1398-8>
- Burgess TM, Webster R (2019) Optimal interpolation and isarithmic mapping of soil properties: I The semi-variogram and punctual kriging. *European Journal of soil science* 70 (1): 11-19. DOI: <https://doi.org/10.1111/ejss.12784>
- Bussel LGJV, Ewert F, Zhao G, Hoffmann H, Enders A, Wallach D, Asseng S, Baigorría GA, Basso B, Biernath C, Cammarano D, Chryssanthacopoulos J, Constantin J, Elliott J, Glotter M, Heinlein F, Kersebaum KC, Klein C, Nendel C, Priesack E, Raynal H, Romero CC, Rötter RP, Specka X, Tao F (2016) Spatial sampling of weather data for regional crop yield simulations. *Agricultural and Forest Meteorology* 220: 101–115. DOI: <https://doi.org/10.1016/j.agrformet.2016.01.014>
- Cambardella CA, Moorman TB, Parkin TB, Novack JM, Karlen DL, Turco RF, Knopka AE (1994) Field-scale variability of soil properties in Central Iowa Soils. *Soil Science Society America Journal* 58 (4): 1501-1511. DOI: <https://doi.org/10.2136/sssaj1994.03615995005800050033x>
- Chipeta MG, Terlouw DJ, Phiri KS, Diggle PJ (2016) Inhibitory geostatistical designs for spatial prediction taking account of uncertain covariance structure. *Environmetrics* 28(1) e2425. DOI: <https://doi.org/10.1002/env.2425>
- Cortès-D DL, Camacho-Tamayo JH, Giraldo R (2016) Spatial prediction of soil penetration resistance using functional geostatistics. *Scientia Agrícola* 73(5):455-461. DOI: <http://dx.doi.org/10.1590/0103-9016-2015-0113>
- Cressie NAC (2015) *Statistics for spatial data*. (Rev. Ed.), New York, John Wiley & Sons, 928 p.
- Dalposso GH, Uribe-Opazo MA, Johann JA, Galea M, De Bastiani F (2018) Gaussian spatial linear modelo of soybean yield using bootstrap methods. *Engenharia Agrícola* 38 (1): 110-116. DOI: <http://dx.doi.org/10.1590/1809-4430-eng.agric.v38n1p110-116/2018>
- De Bastiani F, Uribe-Opazo MA, Dalposso GH (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 (2): 393-404. DOI: <http://dx.doi.org/10.1590/S0100-69162012000200019>
- De Bastiani F, Uribe-Opazo MA, Galea M, Cysneiros AHMA (2018) Case-deletion diagnostics for spatial linear mixed models. *Spatial Statistics* 28: 284-303. DOI: <https://doi.org/10.1016/j.spasta.2018.07.007>
- Gazolla-Neto A, Fernandes MC, Vergara RO, Gadotti GI, Villela FA (2016) Spatial distribution of the chemical properties of the soil and of soybean yield in the field. *Revista Ciência Agronômica* 47 (2): 325-333. DOI: <http://dx.doi.org/10.5935/1806-6690.20160038>
- Genton MG, Kleiber W (2015) Cross-Covariance functions for multivariate geostatistics. *Statistical Science* 30 (2): 147-163. DOI: <http://dx.doi.org/10.1214/14-STS487>
- Guedes LPC, Ribeiro Jr PJ, Piedade SMDS, Uribe-Opazo MA (2011) Optimization of spatial sample configurations using hybrid genetic algorithm and simulated annealing. *Chilean Journal of Statistics* 2 (2): 39-50.
- Guedes LPC, Uribe-Opazo MA, Ribeiro Junior PJ, Dalposso GH (2018) Relationship between sample designs and geometric anisotropy in the preparation of thematic maps of chemical soil attributes. *Revista Engenharia Agrícola* 38 (2): 260-269. DOI: <http://dx.doi.org/10.1590/1809-4430-eng.agric.v38n2p260-269/2018>
- Hofmann T, Darsow A, Schafmeister MT (2010) Importance of the nugget effect in variography on modeling zinc leaching from a contaminated site using simulated annealing. *Journal of Hydrology* 389: 78 – 89. DOI: <https://doi.org/10.1016/j.jhydrol.2010.05.024>

- Kestring FBF, Guedes LPC, De Bastiani F, Uribe-OPazo MO (2015) Comparação de mapas temáticos de diferentes grades amostrais para a produtividade da soja. *Engenharia Agrícola* 35 (4): 733 – 743. DOI: <http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v35n4p733-743/2015>
- Klejnen JPC, Mehdad E (2016) Estimating the variance of the predictor in stochastic kriging. *Simulation Modelling Practice and Theory* 66: 166-173. DOI: <https://doi.org/10.1016/j.simpat.2016.03.008>
- Lark RM, Marchant BP (2018) How should a spatial-coverage sample design for a geostatistical soil survey be supplemented to support estimation of spatial covariance parameters? *Geoderma* 319: 89-99. DOI: <https://doi.org/10.1016/j.geoderma.2017.12.022>
- Lu A, Wang J, Qin X, Wang K, Han P, Zhang S (2012) Multivariate and geostatistical analyses of the spatial distribution and origin of heavy metals in the agricultural soils in Shunyi, Beijing, China. *Science of the Total Environment* 425: 66-74. DOI: [10.1016/j.scitotenv.2012.03.003](https://doi.org/10.1016/j.scitotenv.2012.03.003)
- Monego MD, Ribeiro Junior PJ, Ramos P (2015) Comparing the performance of geostatistical models with additional information from covariates for sewage plume characterization. *Environmental Science and Pollution Research* 27: 5850-5863. DOI: <https://doi.org/10.1007/s11356-014-3709-7>
- Peng CY, Wu CFJ (2014) On the choice of nugget in kriging modeling for deterministic computer experiments. *Journal of Computation and Graphics Statistics* 23 (1): 151-168. DOI: <https://doi.org/10.1080/10618600.2012.738961>
- R Development Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2018. Available: <http://www.R-project.org>.
- Ribeiro Jr PJ, Diggle PJ (2001) geoR: A package for geostatistical analysis. *R-NEWS* 1(2):15-18.
- Robinson DP, Lloyd CD, Mckinley JM (2013) Increasing the accuracy of nitrogen dioxide (NO₂) pollution mapping using geographically weighted regression (GWR) and geostatistics. *International Journal of Applied Earth and Geoinformation* 21: 374-383. DOI: <https://doi.org/10.1016/j.jag.2011.11.001>
- Seidel EJ, Oliveira MS (2014) Novo índice geoestatístico para a mensuração da dependência espacial. *Revista Brasileira Ciência do Solo* 38: 699-705. DOI: <http://dx.doi.org/10.1590/S0100-06832014000300002>
- Uribe-Opazo MA, Borssoi JA, Galea M (2012) Influence diagnostics in Gaussian spatial linear models. *Journal of Applied Statistics* 39 (3): 615-630. DOI: <http://dx.doi.org/10.1080/02664763.2011.607802>
- Webster R, Oliver MA (2007) *Geostatistics for environmental scientists*. Chichester, John Wiley and Sons, 315 p.
- Vallejos R, Osorio F (2014) Effective sample size of spatial process models. *Spatial Statistics* 9: 66-92. DOI: <https://doi.org/10.1016/j.spasta.2014.03.003>
- Wadoux AMJC, Marchant BP, Lark RM (2019) Efficient sampling for geostatistical surveys. *European Journal of soil science (Version of Record online)* 1-15. DOI: <https://doi.org/10.1111/ejss.12797>
- Zhao G, Hoffmann H, Yeluripati J, Xenia S, Nendel C, Coucheney E, Kuhnert M, Tao F, Constantin J, Raynal H, Teixeira E, Grosz B, Doro L, Kiese R, Eckersten H, Haas E, Cammarano D, Kassie B, Moriondo M, Trombi G, Bindi M, Biernath C, Heinlein F, Klein C, Priesack E, Lewan E, Kersebaum KC, Rötter R, Roggero PP, Wallach D, Asseng S, Siebert S, Gaiser T, Ewert F (2016) Evaluating the precision of eight spatial sampling schemes in estimating regional means of simulated yield for two crops. *Environmental Modelling & Software* 80:100–11. DOI: <https://doi.org/10.1016/j.envsoft.2016.02.022>