The use of different sampling grids in determining the variability of soil physical attributes of Oxisol

Eduardo Leonel Bottega*, Daniel Marçal de Queiroz, Francisco de Assis de Carvalho Pinto, Nerilson Terra Santos

Universidade Federal de Viçosa, Viçosa, MG, Brasil
*Autor correspondente, e-mail: bottega.elib@gmail.com

Abstract

This study aimed to analyze the influence of different sampling grids in determining the spatial variability of physical attributes of Oxisol. It was used to study an area of approximately 90 hectares, where soil was sampled at depth from 0 to 0.20 meters, using a grid of 2 points per hectare (G1). Each soil sample was composed of four subsamples and obtained using a soil sampler Dutch type. The samples were sent to the laboratory for granulometric analysis. From the initial grid, the area was divided into sampling cells of 2.9 (G2) and 4.7 hectares (G3), and assigned a coordinate value representative of the center of each cell. Classical statistical and geostatistical methods were used to characterize the data and to model the spatial dependence. Spatial dependence was detected for all physical variables of the soil, regardless of the sampling grid used. The utilization of sampling grid of 1 point for each 2.9 hectares, and the sampling cell characterized by 12 subsamples, showed itself capable of detecting the spatial variability of the physical attributes of the soil, guaranteeing reliability in the estimates, even reducing the quantity of points when compared to the densest grid.

Keywords: precision agriculture, geostatistics, soil, spatial variability

O uso de diferentes grids de amostragem para determinar a variabilidade dos atributos físicos do solo de Latossolo

Resumo

O presente estudo teve por objetivo analisar a influência de diferentes grades amostrais na determinação da variabilidade espacial dos atributos físicos de um Latossolo Vermelho distroférrico. Utilizou-se para o estudo uma área de aproximadamente 90 hectares, onde se amostrou solo na profundidade de 0 – 0,20 metros, utilizando grade de 2 pontos por hectare (G1). Cada amostra de solo foi composta de quatro subamostras e obtida utilizando um trado tipo holandês. As amostras foram encaminhadas ao laboratório para realização da análise granulométrica. A partir da grade inicial, a área foi dividida em células amostrais de 2,9 (G2) e 4,7 hectares (G3), sendo atribuído um valor de coordenada representativo do centro de cada célula. Métodos estatísticos clássicos e geostatísticos foram empregados para caracterizar os dados e modelar a dependência espacial. Foi detectada dependência espacial para todas as variáveis físicas do solo, independentemente da grade amostral utilizada. A modelagem da dependência espacial dos atributos físicos do solo utilizando a grade amostral de 2 pontos por hectare foi a que apresentou, de forma geral, os melhores parâmetros de ajuste para validação cruzada. A utilização de grade amostral de 1 ponto para cada 2,9 hectares, sendo a célula amostral caracterizada por 12 subamostras, mostrou-se capaz de detectar a variabilidade espacial dos atributos físicos do solo, garantindo confiabilidade nas estimativas mesmo reduzindo a quantidade de pontos quando comparada a grade mais densa.

Palavras-chave: agricultura de precisão, geoestatística, solo, variabilidade espacial
Introduction

An important factor to be considered in the planning of agricultural production is the soil initial condition. Advanced analysis techniques have been employed to quantify and characterize the physical and chemical attributes. These analyses can be considered as the main cost component in the characterization of the productive area. In this context, soil sampling becomes an important factor. The number of samples used to characterize certain area directly influences in the cost with analyses, and in the ability to express their real, physical and chemical condition. Using a large number of soil samples, the accuracy of its characterization will be high; however, the cost will also be high. The reverse is also true.

In studies of spatial variability, an important technique, known as geostatistics, is widely used. This technique emerged through studies by Krig for mines reserves estimates. For their application, it is necessary to know the data variance and the distance between observations. In agriculture, geostatistics is widely used in the soil science field as an important tool used to characterize the spatial variability of physical and chemical properties of soil.

Corrêa et al. (2009) emphasize that until recently, agronomic area researchers studied the soil attributes variability through the classical statistics, which assumes that the observations of a given attribute are independent of each other, regardless of its location in the area. In this case, the experiments were conducted to minimize the spatial variability impact, ignoring the fact that observations can be spatially dependent.

Li et al. (2002) emphasize that both spatial variability and temporal variability of soil attributes should be incorporated in procedures and technologies applied to agriculture. Amirinejad et al. (2011), studying the mapping and evaluation of the variation of soil physical health, stress the importance of a good sampling strategy for the characterization and monitoring of the soil quality variability.

In Brazil, precision agriculture has been restricting itself to the application of variable-rate fertilizer. This technology has resulted in gains for the producer. However, fertilizers recommendations have been performed from soil analysis obtained by sampling grids that do not always accurately detect the spatial variability of the attributes analyzed.

Li et al. (2007) point out that an optimal sampling system, in any study, should provide an estimate with lower sampling cost, while representing the existing variability. Sampling must be sufficient to detect the spatial variability of soil attributes; otherwise, a denser sampling grid (larger amount of points) should be deployed. When using denser grid, it is possible to know which variable presents the lowest range value.

Vašát et al. (2010) stress the need for new sampling methodologies that optimize this process, not just for one soil variable, but also for several. The sampling grid should correspond to diverse requirements. Firstly, the number and spatial distribution of sampling points should ensure a minimum precision for estimates in non-sampled locations. Secondly, the optimization technique should be numerically feasible.

Based on the foregoing, it is emphasized the importance of the choice of the grid used for sampling of soil attributes. The ideal sampling grid is one that is able to characterize the spatial variability of the attributes of the field of production with a minimum number of points, ensuring reliability in the estimates. This study sought to evaluate the influence of using different sampling grids in characterization of the spatial variability of soil physical attributes in an agricultural area located in the Brazilian Savannah.

Material and Methods

Data collection were performed on a farm located in the city of Sidrolândia, Mato Grosso do Sul, UTM zone 21 south, with coordinates 702879.040 m east and 7673084.461 m north, in datum SIRGAS2000. This property has a total area of 2,491.07 hectares destined to agriculture. The average altitude, compared to sea level, is 490 meters. The relief is considered slightly undulating. The predominant soil is classified as Oxisol (Soil Survey Staff, 2006). On the farm, it is cultivated soybean (Glycine max), corn (Zea mays) and cotton (Gossypium hirsutum L.), in crop rotation system, performed through no-tillage. For the
study, it was used an area of approximately 90 hectares.

For the mapping of the physical attributes of soil, grids with a density of 2 points per hectare (G1), 1 point for each 2.9 hectares (G2) and 1 point for each 4.7 (G3) were used. The grids G2 and G3 were obtained by dividing the study area in sampling cells of 2.9 and 4.7 hectares. In the middle of each sampling cell was created a representative point of the cell. In Figure 1, the different sampling grids used for the study are showed.

The sampling points were georeferenced using a Topographic GPS reception apparatus of centimetric accuracy with post-processed differential correction. For differential correction, data in the base of the Brazilian Network for Continuous Monitoring (RBMC) of the Brazilian Institute of Geography and Statistics (IBGE), located in the city of Campo Grande / MS, were used. It was used for corrections the datum SIRGAS 2000. The correction was performed using the computer program GNSS Solutions®, provided by GPS receiver manufacturer.

To the characterization of the physical attributes of the study area, soil was sampled in 181 points, using the sampling grid G1 as a reference (Figure 1). It was collected, at each point, a soil sample consisting of four other single samples, representative of the soil layer of 0.0-0.20 meters depth. The single samples were collected in a radius of 3 meters of the georeferenced point using a soil sampler Dutch type. The four single samples were homogenized for withdrawal of approximately 300 g of soil. The composite samples were placed in plastic bags, identified and sent to the laboratory to determine the textural composition (total sand, silt and clay).

In order to obtain the values of physical attributes of soil, representative of the points of the sampling grid G2, it was calculated the mean values of 3 points of grid G1, located within each sampling cell of grid G2. The points were selected randomly and considered representative of sampling cell where they were located. The same procedure was used to obtain the values of soil physical attributes of each point of grid G3. Thus, each soil sample representative of the points of the grids G2 and G3 was composed of 12 single samples.

The data of the physical attributes of soil were submitted to exploratory analysis to verify the presence of discrepant values (Libardi et al., 1996). In this analysis, the critical limit for discrepant values is defined from the interquartile dispersion (DQ). The upper limit was defined by \((Q3 + 1.5 \times DQ)\) and the lower limit by \((Q1 - 1.5 \times DQ)\), where \(Q1\) and \(Q3\) are the first and third quartile, respectively.

After discrepant analysis, the data were analyzed using descriptive statistics, calculating the mean, median, minimum value, maximum value, coefficient of variation, lower quartile, upper quartile, standard deviation, and coefficient of asymmetry and kurtosis, thus seeking to characterize the distribution of data. Normality was tested by Shapiro-Wilk’s test (p <0.05). The spatial dependence of each variable was assessed by the variograms adjustments, presupposing the stationarity of the intrinsic hypothesis, defined by Equation 1.

\[
\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2
\]  (1)
where \( \hat{\gamma}(h) \) = Semivariance in function of separation distance (h) between pairs of points; \( h \) = separation distance between pairs of points, (m); \( N(h) \) = number of experimental pairs of observations \( Z(x_i) \) and \( Z(x_i + h) \) separated by a distance h.

The experimental semivariance is represented in the graph \( \hat{\gamma}(h) \) versus h. It was adjusted the model that best represented the relationship between \( \hat{\gamma}(h) \) and h, and then, being able to determine the parameters: nugget effect (C₀), plateau (C₀ + C) and range (A). The spatial dependence index (SDI) was determined and classified according Zimback (2001), using Equation 2, and thus, assuming the following intervals: low spatial dependence for SDI < 25%, moderate for 25% < SDI < 75% and strong for SDI >75%.

\[
IDE = \left( \frac{C}{C_0 + C} \right) \times 100
\]  

(2)

The degree of correlation between the maps of each variable, prepared after analysis of the different sampling grids was evaluated by calculating the Pearson correlation coefficient. It was selected 36 points randomly, common to all maps produced, regardless of the sampling grid used in the analysis. Montgomery & Runger (2009), emphasize that the number of points used in correlation analysis influences in a biased manner the hypothesis test, and the greater the number of points, the greater the chance of rejection of the nullity hypothesis. In Figure 2, it is presented the distribution map of the 36 points used in the analysis of correlation between the maps of the soil attributes, produced from different grades studied.

Results

It is presented, in Table 1, descriptive statistical analysis of soil physical properties for the different sampling grids studied. It was not detected the presence of discrepant values among the data used for the analysis. It is observed that, by using the sampling grid G2, the data tended to normality by the Shapiro-Wilk’s test at 5% of probability. All variables analyzed showed values of measures of close central tendency (mean and median), indicating that the data tend to symmetrical distribution.

The coefficients of variation (CV %) ranged from 3 to 16. The coefficient of variation, regardless of the sampling grid used was classified as medium (12 < CV% < 60) for the variables: sand and silt. The variable clay, independent of the sampling grid used for analysis, showed a low coefficient of variation (CV% <12), according to the classification proposed by Warrick & Nielsen (1980). Similar results were observed by Valente (2010) in studies performed in a Red-Yellow Oxisol in mountainous area cultivated with coffee.

The adjustment parameters of theoretical models of semivariance of soil physical attributes, for the different sampling grids studied, are presented in Table 2. All variables, regardless of the sampling grid used, showed spatial dependence. The Gaussian model showed the best adjustment for the soil physical variables, regardless of the sampling grid used, except for the variable clay, adjusted to the spherical model when is used to analyze the sampling grid G3.
The minimum distance between points on the sampling grids G1, G2 and G3, were, respectively, 49.97, 103.42 and 170.77 meters. The highest range (1498 meters) was observed for the variable sand. The lowest range (711 meters) was observed for the variable silt. Both ranges were observed when using the grid sampling G3.

The variable clay presented SDI classified as moderate (25% < SDI < 75%) when using the sampling grid G1 in the analysis. The other variables, independent of the sampling grid used, were classified as variables of strong spatial dependence (SDI > 75%), according to the classification proposed by Zimback (2001).

The highest values of nugget effect were observed for the adjustments using sampling grid G1. All coefficient of determination showed high values. The highest value observed was 0.99, for the modeling of the spatial dependence of the variable silt, using sampling grid G2. The lowest coefficient of determination (0.85) was observed for the variogram adjustment, in the modeling of spatial dependence of the variable clay, using sampling grid G3.

The parameters of cross-validation of theoretical models of semivariance are presented in Table 3. The best estimates of values of soil physical attributes were observed for the sampling grid G2. This grid was the one that presented estimates with the highest values of determination coefficient. Less accurate estimates were observed from the analyses using the sampling grid G1.
Soil and Water

Table 3. Cross-validation parameters of semivariance theoretical models

<table>
<thead>
<tr>
<th>Variables*</th>
<th>Regression coefficient</th>
<th>Y Intercept</th>
<th>Square error (SE)</th>
<th>Prediction square error</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>2 points for hectare (G1)</strong></td>
<td><strong>1 point for 2.9 hectares (G2)</strong></td>
<td><strong>1 point for 4.7 hectares (G3)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sand</td>
<td>1.05</td>
<td>-0.62</td>
<td>0.05</td>
<td>1.13</td>
<td>0.70</td>
</tr>
<tr>
<td>Silt</td>
<td>1.01</td>
<td>-0.32</td>
<td>0.06</td>
<td>2.34</td>
<td>0.61</td>
</tr>
<tr>
<td>Clay</td>
<td>0.99</td>
<td>0.42</td>
<td>0.11</td>
<td>2.24</td>
<td>0.30</td>
</tr>
<tr>
<td>Sand</td>
<td>0.88</td>
<td>1.42</td>
<td>0.07</td>
<td>0.60</td>
<td>0.86</td>
</tr>
<tr>
<td>Silt</td>
<td>0.93</td>
<td>1.63</td>
<td>0.08</td>
<td>1.32</td>
<td>0.85</td>
</tr>
<tr>
<td>Clay</td>
<td>0.90</td>
<td>6.60</td>
<td>0.15</td>
<td>1.25</td>
<td>0.57</td>
</tr>
<tr>
<td>Sand</td>
<td>1.07</td>
<td>-0.85</td>
<td>0.13</td>
<td>0.87</td>
<td>0.79</td>
</tr>
<tr>
<td>Silt</td>
<td>1.02</td>
<td>-0.38</td>
<td>0.10</td>
<td>1.50</td>
<td>0.82</td>
</tr>
<tr>
<td>Clay</td>
<td>0.93</td>
<td>4.37</td>
<td>0.21</td>
<td>1.73</td>
<td>0.54</td>
</tr>
</tbody>
</table>

* [dag kg$^{-1}$]; $^1$ Determination coefficient

Figure 3 shows the maps of spatial variability of soil attributes. The maps were obtained through interpolation, using the method of ordinary kriging. The interpolation process was carried out taking into account the parameters adjusted for the physical variables of the soil. It was observed that the maps presented loss of detailing with the reduction of the number of representative grid points of each sampling grid studied.

Figure 3. Spatial variability maps of soil physical properties for the sampling grids: 2 points for hectare (G1), 1 point for 2.9 hectares (G2) and 1 point for 4.7 hectares (G3).
The result of Pearson correlation analysis, between the maps of each physical attribute of the soil prepared from the use of different sampling grids, is presented in Table 4. All correlations presented significance at 1% probability level. The highest Pearson correlation coefficients between the maps prepared from the grid G1 were observed for comparison, using maps prepared with the grid G2. The lowest correlation coefficient (0.72) was observed for the correlation analysis between the maps of silt, prepared from the sampling grids G1 and G3. The highest correlation coefficient (0.99) was observed between the maps of clay, prepared using sampling grids G1 and G2.

**Table 4. Pearson correlation coefficient between the maps made from different sampling grids**

<table>
<thead>
<tr>
<th>Variables</th>
<th>G1 x G2</th>
<th>G1 x G3</th>
<th>G2 x G3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sand (dag kg⁻¹)</td>
<td>0.97**</td>
<td>0.97**</td>
<td>0.93**</td>
</tr>
<tr>
<td>Silt (dag kg⁻¹)</td>
<td>0.96**</td>
<td>0.72**</td>
<td>0.79**</td>
</tr>
<tr>
<td>Clay (dag kg⁻¹)</td>
<td>0.99**</td>
<td>0.92**</td>
<td>0.90**</td>
</tr>
</tbody>
</table>

** Significance at 1% probability

**Discussion**

Not all sampling grids that were used presented normality in the distribution of data variables. In classical statistics, it is required that the basic assumptions of normality of errors, homogeneity of variances and independence of errors be considered for their efficient application; when they are not considered it will probably lead to inferences that lack confidence and precision (Guimarães, 2000). However, Cressie (1993) emphasizes that geostatistics does not require normality of the data to be applied; it is only advisable that the distribution does not present very long tails.

The variogram adjustment to the spherical model for the variable clay, using the grid G3 in the analysis, indicates loss of efficiency of the sampling grid of the spatial detection variability of this variable. The spherical model represents a low spatial continuity, unlike the Gaussian model, representative of the extremely continuous phenomena. Isaaks & Srivastava (1989), emphasize that the Gaussian model represents smooth variations over small distances of observation.

All physical variables of the soil presented range values much higher than to the shortest distance between points of each sampling grid used. Range is important for the interpretation of the semivariograms, to indicate the distance to where the sample points are correlated among themselves, i.e., points located in an area whose radius is the range, are more similar than those are that are separated by greater distances. According to Corá et al. (2004), estimates made with ordinary kriging interpolation, using greater range values, tend to be more reliable, presenting maps that better represent reality. When considering as distance between sampling points half of the range value, it is guaranteed the detection of the spatial variability of the attribute under study without losing precision in the estimates, because the spatial continuity of the variable is maintained (Carvalho et al., 2002).

In addition to the range value, other parameters of the variogram can assist in choosing the model that best represents the spatial variability of the variable analyzed. It can be cited as examples the number of pairs of points that compose the representative points of the semivariogram based on distance, the relationship value between the nugget effect and plateau, and the value of coefficient of determination of the theoretical model adjusted to empirical semivariance of the data. Journel & Huijbregts (1978), point out that, to ensure the reliability of the theoretical model adjusted, it is important to observe if the points that compose the semivariogram are representative of the variance between at least 30 pairs of points. Guimarães (2004) points out that the smaller the proportion of nugget effect to the plateau of the variogram, the greater the continuity of the phenomenon, the smaller the variance of the estimate and the greater confidence that one can have in the estimate. The value...
of the coefficient of correlation can serve as a parameter in the choice of theoretical model that best represents the semivariance of data; however, the analysis of cross-validation is necessary to ensure the accuracy of the model chosen.

According to Montes et al. (2005), the idea of cross-validation is to validate the ability of the model adjusted of the semivariogram uncertainty associated with the uncertain of the attribute non-sampled. For this, one withdraws the sampled value and obtains the estimate by kriging, using the values of neighboring points. This process is performed for all the sampled points. At the end, for each point, there will exist the true value and the estimated value, and therefore, the estimation error.

The inferiority of the coefficient of determination, observed for cross-validation using sampling grid G1, rather than the coefficient of determination of the other grids G2 and G3, indicates increased estimation error. Soares (2006) reports that several factors may contribute to the increase of the estimation error, highlighting the high nugget effect values regarding the total variance. The nugget effect reflects the non-explained variability according to the distance of the sampling used, as local variations, errors in analysis, sampling errors and other. It is observed that, in adjustment of the variograms, using for analysis the sampling grid G1, the nugget effect values were higher than those observed for adjustment using the others sampling grids.

The loss of detail observed by comparing the maps of Figure 3, is the effect of attenuation in estimating extreme values, which generally leads to an underestimation of the proportion of values above average, and an overestimation of the proportion of values below the average (Soares, 2006). The quantity of observations to be used in the kriging process should not be too high because it results in interpolated values very close to, or correlated to, the nearest point. However, the quantity of points should not be too small because it smoothes excessively the interpolated value, resulting in loss of the result sought (Andriotti, 2003).

In order to improve the sampling process, it is necessary the choice of the size of the sampling cell to be used in the characterization of the spatial variability of soil attributes. It is suggested, before starting the sampling process, the discussion with the producer, about which size of sampling cell is economically viable to be managed, i.e., which size of area this producer is able to handle differently. This decision implies in costs with analysis and precision in the detailing of the spatial variability of soil attributes.

After choosing the sampling cell to be used to the characterization of the spatial variability of soil attributes, it should be recognized the quantity of single samples capable to characterize the cell chosen. Based on these questionings, new studies are suggested regarding the minimum number of single samples able to characterize different sizes of sampling cell.

Conclusions

The utilization of sampling grid of 1 point for each 2.9 hectares, and the sampling cell characterized by 12 subsamples, showed it capable to detect the spatial variability of soil physical attributes, guaranteeing reliability in the estimates, and even reducing the quantity of points, when compared the densest grid.

Acknowledgements

The Scientific and Technological Development National Council (CNPq) and Foundation for Research Support of the Minas Gerais State (FAPEMIG) by the financial support.

References


Corá, J.E., Araujo, A.V., Pereira, G.T., Beraldo, J.M.G. 2004. Variabilidade espacial de atributos
do solo para adoção do sistema de agricultura
de precisão na cultura de cana-de-açúcar.
Revista Brasileira de Ciência do Solo 28: 1013-
1021.

Corrêa, A.N., Tavares, M.H.F., Uribe-Opazo, M.A.
2009. Variabilidade espacial de atributos físicos
do solo e seus efeitos sobre a produtividade do

Wiley & Sons, New York, United States. 900 p.

Inc. Golden, Colorado, United States. CD-ROM.

atributos de um Latossolo Vermelho Amarelo-
escuero, textura argilosa, da região do cerrado,
submetido ao plantio direto e ao plantio
convencional. 85f. (Tese de Doutorado) –
Universidade Estadual de Campinas, Campinas,
Brasil.

aplicada. Universidade Federal de Uberlândia,
Uberlândia, Brasil. 78 p.

to applied geostatistics. Oxford Univerosit, New
York, USA. 561 p.

geostatistics. Academic Press, Waltham, USA.
600 p.

Li, H., Lascano, R.J., Booker, J., Wilson, L.T., Bronson,
heterogeneity: water and nitrogen use in cotton.
Soil Science Society of America Journal 66: 585-
595.

Li, Y., Shi, Z., Li, F. 2007. Delineation of Site-Specific
Management Zones Based on Temporal and
Spatial Variability of Soil Electrical Conductivity.
Soil Science Society of China. Pedosphere
17: 156-164.

gravimétrica de um solo hidromórfico. Revista
Brasileira de Ciência do Solo 20: 1-12.

A geostatistical approach to cork production
sampling estimation in Quercus suber forests.
Canadian Journal of Forest Research 35: 2787-
2796.

Montgomery, D.C., Runger, G.C. 2009. Estatística
aplicada e probabilidade para engenheiros.

Robertson, G.P. 2008. GS*: Geostatistics for
the Environmental Sciences. Gamma Design
Software, Plainwell , Michigan, USA. 179 p.

Soares, A. 2006. Geoestatística para as ciências
da terra e do ambiente. IST Press, Lisboa, Portugal.
214 p.


Valente, D.S.M. 2010. Desenvolvimento de um
sistema de apoio à decisão para definir zonas
de manejo em cafeicultura de precisão. 122f.
(Tese de Doutorado) – Universidade Federal de
Viçosa, Viçosa, Brasil.

Sampling design optimization for multivariate soil

Warrick, A.W., Nielsen, D.R. 1980. Spatial variability
of soil physical properties in the field. In: Hillel, D.
(ed.) Applications of soil physics. Academic Press,
New York, USA. 385 p.

químicos de solos para fins de mapeamento da
fertilidade do solo. 114f. (Tese de Doutorado) –
Universidade Estadual Paulista, Botucatu, Brasil.