Correlating Soil Porosity and Respective Geological Unit in Paraíba do Sul Valley, Brazil - A Geostatistical Methodology Proposal

P.I.M. Camarinha, S.J.C. Simões, G.P. Bernardes, I.C.B. Trannin

Abstract. This manuscript aims proposing a methodology for correlating soil porosity to the respective geological units using geostatistical analysis techniques, including interpolation data by kriging. The site studied was in Lorena municipality, Paraíba do Sul Valley, southeastern Brazil. Specifically all studies were carried out within an area of 12 km^2 located at Santa Edwirges farm. The database comprehended 41 soil samples taken at different geological and geomorphologic units at three different depths: surface, 50 cm and 100 cm depth. The geostatistical analyses results were correlated to a geological mapping specifically elaborated for the site. This mapping accounts for two different geological formations and a geological contact characterized by a shearing zone. The results indicate the existence of a significant relationship between the soil porosity and the respective geological units. The studies revealed that the residual soils from weathered granitic rocks tend to have higher porosities than the residual soils from weathered biotite gneiss rocks, while the soil porosity within the shearing zone is relatively un-sensitive to the respective geological formation. The spatial patterns observed were efficient to evaluate the relationship between the soil porosity, geology unit and the and geomorphology showing a good potential for correlating with others soil properties such as hydraulic conductivity, soil water retention curves and erosion potentials.
Keywords: geostatistic, sampling, error prediction, porosity map.

1. Introduction

The knowledge of the physical properties of the soil contributes to improve our understanding of their mechanical and hydraulic properties. However, depending on the size of the area of interest and the proposed objectives, awareness of these physical properties, in plot field, can be very expensive and time consuming. In different research domains such as agronomy, mining, hydrology and river basin planning, where large portion of land has to be analyzed, the plot field rarely contributes to satisfactory results, requiring assess to spatial variability of soil properties. In geotechnical engineering, mapping surveys of areas susceptible to mass movement have provided to be an important tool to reduce - or avoid - the potential losses from natural hazards, besides to providing important information for decision-makers regarding the land use occupation. Therefore, spatial information can provide a wider view of the environment studies and allow correlation with other properties or attributes related to physical landscape elements.

A practical limitation found in studies of soil properties is that, generally, the amount of the good-quality field datasets are scarce for a particular area of study. In the same way, when one wants to integrate data from other landscape elements for example, a geological map, the difficulty becomes even greater in areas of small dimensions, due to the lack of thematic maps produced in appropriate scale.

For these reasons, it is important to develop low-costs methodologies that provide tools for spatial representation of a certain property. Among the available options, the spatial data interpolation methods are commonly used to meet these needs. They are capable of predicting the spatialization of a random variable for large or small areas based on punctual observations (sampled points).

Burrough (1998) states that, when there are enough data, most interpolation methods produce similar values. However, in the case of sparse data, such methods have limitations in the representation of spatial variability, since they do not consider the location of the samples and then ignore the continuity of the phenomenon. Krige (1951) initiated studies seeking to understand the spatial variability of concentration of gold, considering the spatial location of each sample and its interference in neighboring occurrences.

Pedro Ivo Mioni Camarinha, Engenheiro Civil, MSc Student, Faculdade de Engenharia de Guaratinguetá, Universidade Estadual Paulista "Júlio de Mesquita Filho", Guaratinguetá, SP, Brazil. e-mail: pedruivo@hotmail.com.
Silvio Jorge Coelho Simões, Geólogo, PhD, Associate Professor, Departamento de Engenharia Civil, Universidade Estadual Paulista "Júlio de Mesquita Filho", Guaratinguetá, SP, Brazil. e-mail: simoes@feg.unesp.br.
George de Paula Bernardes, Engenheiro Civil, PhD, Assistant Professor, Departamento de Engenharia Civil, Universidade Estadual Paulista "Júlio de Mesquita Filho", Guaratinguetá, SP, Brazil. e-mail: gpb@feg.unesp.br.
Isabel Cristina de Barros Trannin, Engenheira Agrônoma, PhD, Assistant Professor, Professor do Departamento de Engenharia Civil, Universidade Estadual Paulista "Júlio de Mesquita Filho", Guaratinguetá, SP, Brazil. e-mail: isatrannin@feg.unesp.br.
Submitted on February 16, 2011; Final Acceptance on November 21, 2012; Discussion open until April 30, 2013.

Based on studies of Krige (1951), Matheron (1963, 1971) proposed the theory of geostatistics, known as the *Theory of Regionalized Variables*, which gives the theoretical basis for the application of kriging. According to Vieira (2000), kriging has the ability to produce better estimates among the interpolation methods because it is based on two premises: non-bias estimator and minimum variance estimates.

Several studies were then developed using kriging in order to describe the spatial distribution of soil properties. Vieira *et al.* (1981), for example, investigated the spatial dependence of the infiltration rate of water on Yolo (CA) clay loam. In their study, the spatial variability of 1280 field-measured infiltration rate was studied using geostatistical concepts. Depending on the correct selection of samples, it was observed that a minimum of 128 samples was enough to obtain nearly the same information as with 1280 samples.

Gomes *et al.* (2007) made use of kriging techniques to know the spatial distribution of soil density, concentration of organic matter and soil texture in order to find relations with land use occupation. The study was carried out with 165 point samples distributed over an area of 47 km^2, located at Ribeirão Marcela Basin, MG, Brazil. The authors found that some types of agricultural soil use causes direct interferences in the spatial distribution of soil attributes. Thus, they emphasized the importance of the use of kriging to find critical areas regarding the soil management and, also, to provide important information for proper planning of the land use.

Fernandes da Silva *et al.* (2007) analyzed some geotechnical properties such as grain size distribution fractions and plasticity index, in Ubatuba area, north coast of São Paulo State, Brazil. The purpose of the study was to find spatial patterns of these properties that allow the estimation of geotechnical behavior of soils. The authors employed three interpolation methods (*Nearest-neighbor interpolation, Weighted mean* and *Kriging*) using 73 point samples in order to find the better technique to estimate the spatialization of the geotechnical properties. The results suggested that kriging is a better model to be use in regionalization of the parameters.

From the previous studies presented, kriging technique has proved to be an important tool for spatial representation of soil properties. However, kriging and any other spatial interpolation method aggregate errors in their estimates, which is often overlooked in geostatistical studies. In some cases, the error associated with interpolation method can be so high that causes large discrepancies between estimates and observed reality.

It is important to know the error associated with the prediction in order to evaluate the results obtained. Besides, to validate and to refine the methodology applied it is also important to seek additional spatial information regarding the investigated area. This can be done, for example, correlating the spatial of a certain soil property with other field spatial information previously known, such as geology, geomorphology and pedology.

Based on the fundaments proposed from Matheron (1963, 1971), this study investigate the existence of spatial dependence of the variable porosity (η) for an area of 12 km^2 in Lorena municipality, located in the Paraíba do Sul Basin, São Paulo State. These analyses were done for three different depths: surface (samples taken between 0 and 20 cm); 50 cm and 100 cm respectively. The study presents a methodology for correlating soil porosity to the respective geological unit using geostatistical analysis techniques, including interpolation data by kriging and the error involved.

2. Materials and Methods

2.1. Study area

The study area comprises the region of Santa Edwiges Farm, which is inserted in the region upper Taboão stream watershed, located in the Paraíba do Sul Valley, Southeast of Brazil (Fig. 1).

Despite its small size (12 km^2), this investigated area was chosen because it reflects the diversity found in geology and geomorphology in the region, which characterizes the transition between the extensive plains of the Paraíba do Sul Valley and the coastal mountain chain of Serra do Mar.

Therefore, the Santa Edwiges farm is entire inserted into a geological context formed by crystalline rocks of pre-Cambrian age (>500 million years). The map of Fig. 2 shows the various geological units encountered within the study area: a) metamorphic rocks (schists, gneisses and migmatites) of the Embu Complex (Hasui & Sadowski, 1976; Carneiro *et al.*, 1978; Landim, 1984); b) igneous rocks (in most cases are of granitic composition) of the Quebra-Cangalha Suite (Landim, 1984); c) high deformation bands (milonites rocks); d) unconsolidated sediments located in the floodplain of small streams.

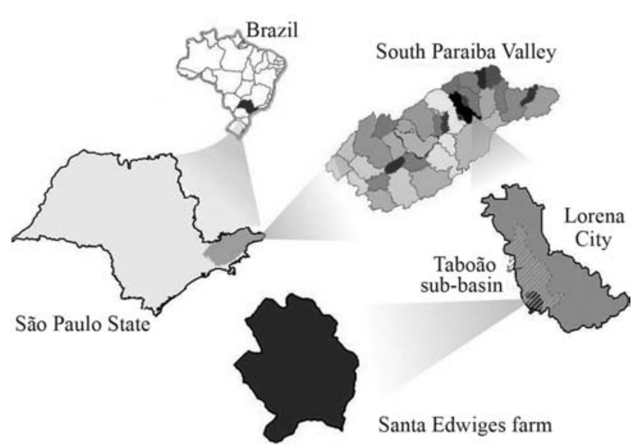


Figure 1 - Location of study area.

Based on Fig. 2 one can see that the northern portion of the area consists of the *Embu Complex*. This unit consists of metamorphic rocks (gneisses) having in its composition minerals more easily weathered such as mica and feldspar. Thus, the soils from these rocks are usually fine-grained soils where predominate clay minerals favoring the development of a more impervious and more homogeneous soil. The soils are usually thicker and have a reddish color because of the presence of iron-rich minerals such as biotite.

The unit *Granitoid Quebra-Cangalha* occurs in the southern area. It is composed predominantly of white to grey leucocratic granites. Rocks of this unit are composed of minerals more resistant to alteration such as quartz and feldspar. Due to the presence of these minerals, soils are predominantly whitish, have a sandy-clay constitution and with a significant presence of mineral fractions of coarser material, such as silt. The coarser texture of these soils and the absence of vegetation, favors the occurrence of erosion processes in advanced stages, such as ridges and ravines.

Milonite rocks account for about 10% of the total area studied. These areas were subjected to intense tectonic tensions in ductile conditions, that is, at depths greater than 10 km (Ramsay, 1980). For this reason, they have a well developed foliation and the presence of well-structured and fine-grained minerals like mica and chlorite as a result of processes of retro metamorphism due to percolation of fluids in shear zones. While on the surface, the intense foliation of these rocks facilitates the processes of weathering and the formation of soils with a high proportion of clay.

The unit *Fluvial Terrace* and the unit *Unconsolidated Sediments* are associated with a fluvial plain Ribeirão Taboão and its main tributaries. In this area are identified paleo-terraces with pelitic sediment composition (silt and clay) and, secondarily, sand and angular pebbles of quartz and feldspar. Dark-colored sediments are also observed indicating the presence of rich organic soil.

Regionally, this area is inserted in the geomorphological unit of Mid Plateau of the Valley of *Paraíba do Sul* which was described by Ponçano *et al.* (1981) in the Geomorphological Map of the State of Sao Paulo. The diversity of geological substrate as described above is directly responsible for the wide variation of reliefs and soils found in the Santa Edwiges farm.

On the scale adopted (1:10.000), it was possible to identify, on the basis of morphometric elements of terrain (hypsometric and slope) three distinct geomorphological units: Ridge Escarpments, Mountain with Moderate to Gentle Hillslopes and Smooth Convex Surface. These units can be observed through a digital elevation model (DEM), developed by Lima (2005) (Fig. 3). The description of these units is as follows.

Ridge Escarpments - located in the northern portion of the study area is characterized by relief with steep slopes (>30%) and large amplitudes of altitude (>300 m). Characteristics of this unit include deep narrow valleys and high drainage density. The soils associated with Ridge Escarpments are, predominantly, young residual soil and saprolite.

Mountain with moderate to gentle hillslopes - occupies the central and northern area of study. They feature rounded shapes with medium slopes (>15%) lower than elongated ridge escarpments and amplitudes of altitude ranging between 100 and 300 m. The valleys are more extensive with the presence of alveoli and medium drainage density. The profiles of weathering in this region vary widely, with sites presenting large thicknesses of transported soil.

Smooth Convex Surface - represents the transition between both units above-mentioned situated in the central portion of the study area. It corresponds the areas of very

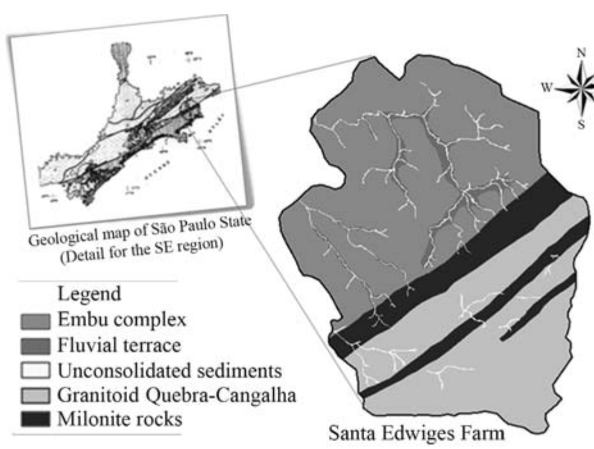


Figure 2 - Geological Map of Fazenda Santa Edwiges, drawn to scale 1:10.000, adapted from Rodrigues & Milanezi (2005).

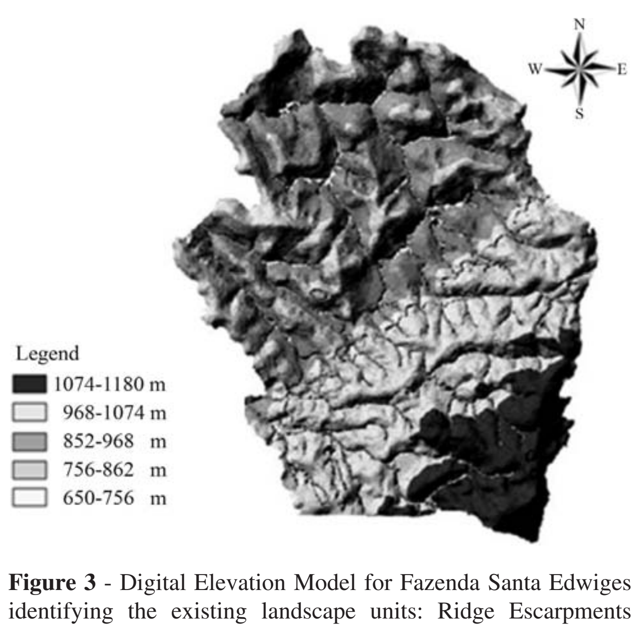


Figure 3 - Digital Elevation Model for Fazenda Santa Edwiges identifying the existing landscape units: Ridge Escarpments (black), Mountain with Moderate to Gentle Hillslopes (light gray) and Smooth Convex Surface (dark gray). Adapted from: Lima (2005).

low slope (<5%) with low-altimetric variations (<50 m) usually related to alluvial plains. In this region predominates thicker layers of mature residual soil.

2.2. Geostatistics

The geostatistics is able to provide estimates in a context governed by a natural phenomenon with distribution in space. It assumes that the values of variables are autocorrelated spatially Landim *et al.* (2002), , such that samples close together in space are more alike than those that are further apart, and is based on the *Theory of Regionalized Variables*, proposed by Matheron (1963).

Geostatistics uses the variogram as one of its primary tools (sometimes called semivariogram) to measure the spatial variability of a regionalized variable, and provides the input parameters for the spatial interpolation of kriging (Krige, 1951; Webster & Oliver, 1993.)

2.2.1. Variograms

The variogram is used to investigate the relationship of the distribution of variable ($z(x)$) in space. This tool is able to measure the degree of spatial dependence between samples over a specific support. The expected squared difference between paired data values {$z(x)$ and $z(x + \mathbf{h})$} to the lag distance **h**, are simply used for its construction assuming stationarity in increments, (Landim, 2006).

To obtain an estimate of the parameters a theoretical semivariogram model is used to define the weights of the kriging function. One can formulate an estimator for the semivariogram which may be calculated thus:

$$2\hat{\gamma}(\mathbf{h}) = \frac{1}{|N(h)|}\sum_{i=1}^{N(h)}[z(x_i) - z(x_j)]^2, \quad h \in \Re^d \qquad (1)$$

where $Y(x_i)$ represents the value of the data at location x_i; h is the displacement between the data pairs; and $H(h)$ is the number of such data pairs in the region, which is given by:

$$N(\mathbf{h}) \equiv \{(x_i, x_j) : x_i - x_j = \mathrm{h}; i, j = 1,..., n\} \qquad (2)$$

When there is spatial dependence, usually, the closest two measures are more alike than two others that are further apart, allowing $\gamma(\mathbf{h})$ to increase as the distance h increases too. However, from a certain distance, it will not find related values with $z(\mathbf{h})$ because the spatial correlation between the samples ceases to exist (Goovaerts, 1997; Landim & Struraro, 2002; Gumiaux *et al.*, 2003). The semivariogram point where the data present no spatial dependence, maintained around the same semi-variance (y axis) and where it is established a straight line in the graph, called the "*sill*" (C). The distance from the origin (x and y coordinates equals zero) to the *sill*, is called the "*range*" (a), which represents the radius of influence of sampling points on its neighborhood, indicated by the distance at which the variance stabilizes (Fig. 4).

By definition, $\gamma(0)$ should be zero, but in practice it is noticed that there are cases where as h approaches zero, $y(\mathbf{h})$ approaches a positive value called "*nugget effect*" or "*nugget*" (C_0). This parameter demonstrates the discontinuity of semivariogram for distances smaller than the smallest distance observed among samples. The nugget effect is the value of semi-variance for the distance zero and represents the component of spatial variability that can not be related to a specific cause, that is, random variability (Camargo, 1997), or also to be linked to errors in measurement.

2.2.2. Kriging

The kriging method uses information from the theoretical variogram model to find the optimal weights to be associated with points with known values (sampled points) which will estimate the unknown points. In this respect, it is understood as a series of techniques of regression analysis that seeks to minimize the estimated variance from a previous model, which takes into account stochastic dependence among the data distributed in space (Matheron, 1971; Isaaks & Srivastava, 1989).

The difference between kriging and other methods of interpolation is the way the weights are distributed in the different samples. For traditional methods, such as *Simple Linear Interpolation*, all samples have weights equal to $1/N$ (N being the total number of samples). In the Inverse Distance Weighting (IDW), the weights given to samples are related to the inverse of the distance that separates the estimated to the observed values. In the case of kriging, it is the weighted mobile average of values observed in the neighborhood where the closest neighbors have more weight and, the neighbors further apart, have increasingly smaller weights, zero and even negative values (Cressie, 1993; Ribeiro Junior, 1995).

Moreover, the kriging provides unbiased estimates and minimum variance. Unbiased estimates indicate that,

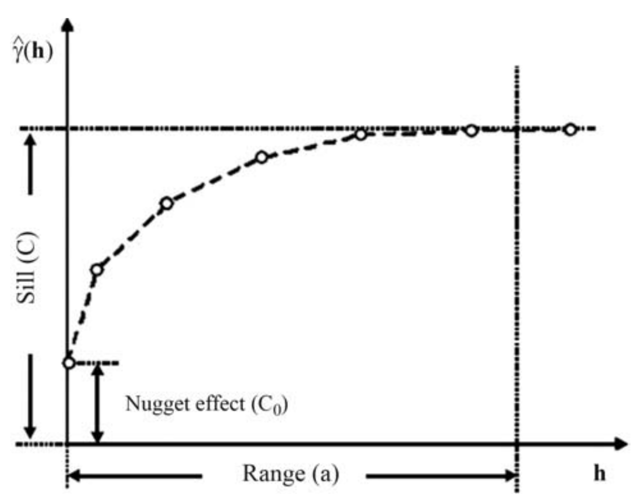


Figure 4 - Parameters of the semivariogram. Adapted from Camargo, 1997.

on average, the difference between the estimated and observed values, through a sampling, for the same point must be zero; and minimum variance means that these estimators have the smallest variance among all unbiased estimators (Camargo, 1997).

2.3. Sampling

One of the main questions in studies involving interpolation methods refers to the amount of samples needed to obtain representative results. Generally, the size of the study area is taken as reference. For geotechnical maps in detail scale of 1:10.000, Matula & Pasek (1984-apud Zuquette; Gandolfi, 2004) suggest minimum sampling between 10 to 25. For this scale, Zuquette (1987) recommend 15 sampling points at minimum with a distance between points egual to 258 m. However, some authors, like Webster & Oliver (1993), assert that there is no a specified minimum number of samples for to realize geostatistical studies and emphasize the importance of the findings being complemented by technical knowledge or information of areas similar to the study area.

Due to the wide diversity of landscape of the study region, set in rugged terrain, a specific number of samples were not previously determined. The study sought another path, where the database was developed in several stages to achieve desired quality, guided from observation of the spatial distribution of the variable analyzed and also based on the error attributed to the process. The sampling process also sought to include all geological units in order to search possible correlations between the pattern of spatial porosity and the Geologic Map shown in Fig. 2.

2.4. Geostatistics procedures

The procedures adopted for the application of geostatistics approach permit the evaluation, at each stage, the quality of data and the partial results obtained and seeking new solutions to improve the final results. Figure 5 summarizes the step-by-step procedures used to prepare the map for predicting porosity and the different options for certain situations (Camarinha *et al.*, 2008). The steps were organized in three main stages: 1st stage - identifying and preparing data (from step 1 to 5), 2nd stage - analysis of data (step 6 to 9) and 3rd stage - optimization of the process (step 10 to 13).

2.4.1. Definition and preparation of data

After the definition of the study area, *Santa Edwiges* farm, the total porosity was chosen as the variable of soil being analyzed by the present proposals due to two factors: a) easy in the sampling process and laboratory testing and b) association with the infiltration process and water movement in soil.

The next step was the establishment of the database, which originally were constituted of 30 georeferenced samples, in continuation of the field works of Domingos (2004). Each sample is constituted of the following parameters: total porosity, natural and dry specific gravity, moisture content, void ratio and specific gravity of the grains evaluated through the classical soil mechanics approach at

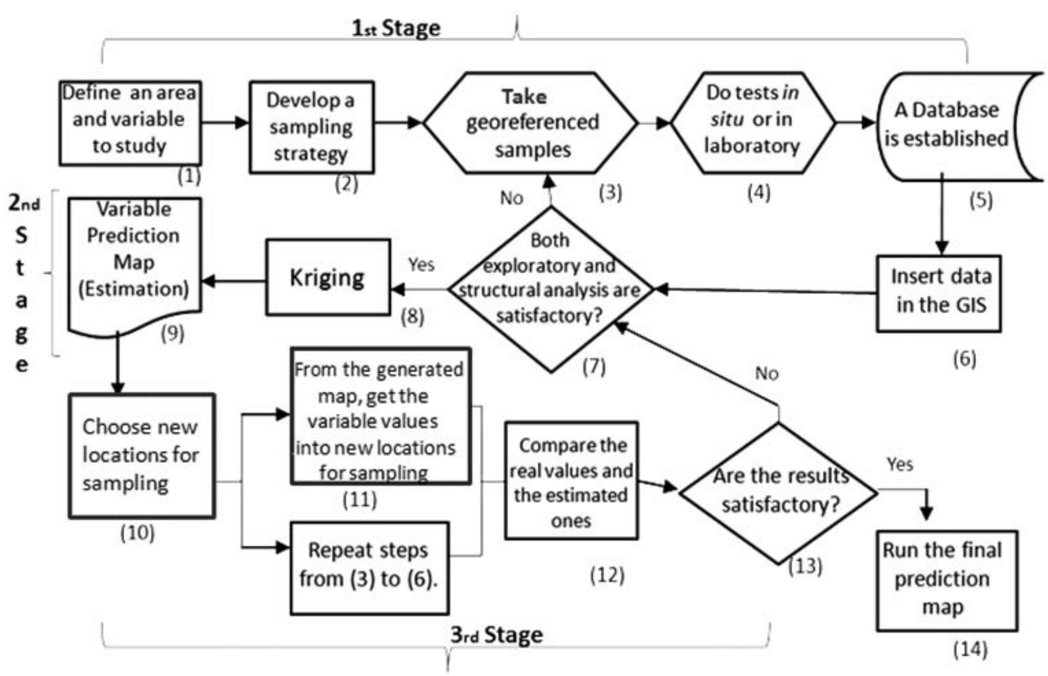


Figure 5 - Schematic chart proposed for the procedures adopted in the preparation of a map to predict desirable quality (Camarinha *et al.*, 2008).

the Laboratory of Soil Mechanics, UNESP/FEG. All these parameters were obtained at 3 different depths: surface (samples taken between 0 and 20 cm), 50 and 100 cm.

2.4.2. Data analysis

In this study, the establishment of a georeferenced database was made by inserting new points, one by one, using a Geographic Information System (ArcGIS® 9.2 and 9.3 versions) available at the Laboratory of GeoSpatial Analysis (LAGE). Then, the values of various physical indexes known were assigned, among them the porosity, the parameter chosen for this study.

From these data, it was aimed to elaborate a partial map to predict the special distribution of soils porosity and also the error associated with the estimates. This error map is prepared analogously to the prediction of any variable. However, instead of using for each point the known value of the variable, it is used all design points on prediction in order to estimate the values . This error exists due to the fact that kriging considers the influence of neighboring samples to estimate the values in any location; even if there is already a collected sample in that place. Thus, the estimated value is not necessarily equal to the actual value of the field sample. Both maps are made using the extension Geostatistical Analyst, ArcGIS® program. The theoretical framework used in the routine of the program is presented in 0.

It is important to observe that in the steps preceding the drafting of the final prediction map (step 14, Fig. 5), some statistical analysis (step 7) such as verification of the histogram, cross-validation, trend analysis and so on, will not be carried out in detail. This is done because the methodology could become impractical. Then, every time it becomes necessary to draw up a map of preliminary prediction (step 9). Nevertheless, attention will be focused just on the obtaining variograms that are able to verify spatial dependence of the study variable.

2.4.3. Process optimization

The quality of dataset collected is evaluated from the spatial and exploratory analysis (histograms, verification of clusters etc.). Depending on such assessments, one can verify whether the dataset provides sufficient conditions for the generation of reliable maps prediction (Houlding, 2000). If this is not checked, it is possible to use dataset to get information that could guide the development of the next steps of the research. Then, it is possible to devise better strategies from sampling and then avoid the difficulties encountered previously.

Before heading to the final steps which involve the geostatistics, it is necessary that the dataset show a desirable quality in accordance with the reality observed in the field. This assessment is made by comparing the real values of the variable, taken in the field, and the preliminaries values estimated by kriging. Only after achieving the desired quality, further studies will be directed to deep analysis in order to provide the final maps.

The analyses of this step make possible to verify the prediction error distribution, the existence and representativeness of the variogram and, besides, whether the actual result is consistent with that expected for the variable studied (Camarinha et.al, 2008). For all the preliminary steps were generated experimental semivariograms. However, in this paper, only the semivariograms for the final stage were presented.

3. Results and Discussions

3.1. Procedure for the prediction of the porosity maps and the associated error

The analysis started with 30 existed samples collected and tested at the Laboratory of Soil Mechanics, UNESP/FEG. Figure 6 shows the location of each sample within the Santa Edwiges Farm. Its spatial position and the porosity values for each depth (samples taken between 0 and 20 cm, 50 cm and 100 cm) are presented in the table.

Based on this initial database, the first analysis was carried out to produce a map with the prediction of the surface porosity and the associated error of the estimate (Fig. 7). The figure presents the first map generated, in which the darker lines and surfaces represent where there is higher prediction error and higher porosity, respectively.

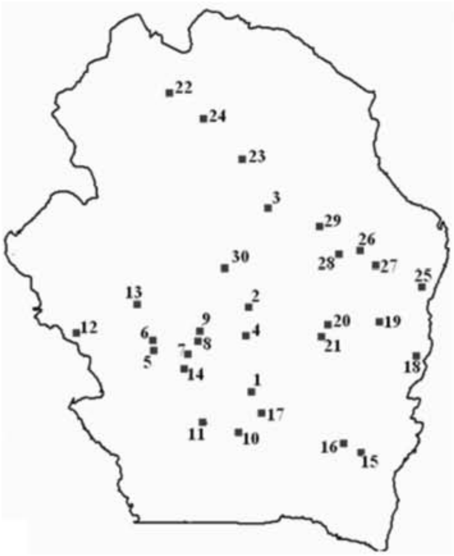
In this first preliminary map, shown in the Fig. 7(A), the regions near the physical boundary of the study area, especially in the north portion, identify areas with greater uncertainty, for this reason, a new set of soil sampling were carried out in this area as illustrated in Fig. 7(B). These samples also allowed the verification whether the number of samples is representative regarding the different geologic units within the study area.

From the results obtained with these new samples, a comparison was made between the values of porosity estimated by kriging (Fig. 7A) with the actual values determined from laboratory tests (Table 1).

The results shows that only two points (34 and 35), show the real values quite discrepant from those estimated by the kriging method. Considering the size of the study area (12 km^2) and the low number of samples at this stage, this initial comparison of the data reflects a substantial quality of the interpolation model used.

Although the quality of these preliminary results were acceptable, it was produced another map with the prediction error using the new dataset with 38 sampled points. Maps 1 and 2 in Fig. 8 represent the distribution of the error for both cases analyzed and indicate the region of new soil investigation.

From map 2 in Fig. 8 it is still possible to observe the existence of a region with high error at the northwest portion of the area, which means that the predicted porosity has low accuracy. Based on this observation, it was sought to



	Coordenadas		η		
Ponto	X	Y	0	50	100
1	490663	7470340	0,51	0,46	0,51
2	490651	7471193	0,42	0,40	0,41
3	490839	7472196	0,50	0,52	0,43
4	490638	7470914	0,49	0,40	0,46
5	489677	7470758	0,47	0,37	0,36
6	489664	7470842	0,50	0,52	0,50
7	490034	7470719	0,45	0,55	0,45
8	490144	7470868	0,54	0,52	0,53
9	490142	7470946	0,51	0,44	0,46
10	490541	7469925	0,46	0,49	0,44
11	490193	7470021	0,46	0,46	0,45
12	488892	7470923	0,60	0,58	0,53
13	489506	7471225	0,50	0,49	0,40
14	489983	7470585	0,42	0,45	0,42
15	491789	7469721	0,60	0,46	0,48

	Coordenadas		η		
Ponto	X	Y	0	50	100
16	491596	7469830	0,52	0,50	0,49
17	490767	7470114	0,53	0,46	0,43
18	492348	7470712	0,58	0,51	0,56
19	491953	7471043	0,61	0,46	0,53
20	491452	7471027	0,47	0,48	0,51
21	491376	7470901	0,51	0,45	0,44
22	489854	7473350	0,47	0,48	0,50
23	490568	7472693	0,50	0,44	0,50
24	490185	7473099	0,47	0,59	0,57
25	492399	7471405	0,49	0,50	0,56
26	491764	7471779	0,51	0,47	0,50
27	491940	7471615	0,52	0,44	0,48
28	491550	7471725	0,50	0,43	0,43
29	491368	7472015	0,54	0,50	0,45
30	490395	7471591	0,49	0,45	0,53

Figure 6 - Location and attributes of the first 30 samples.

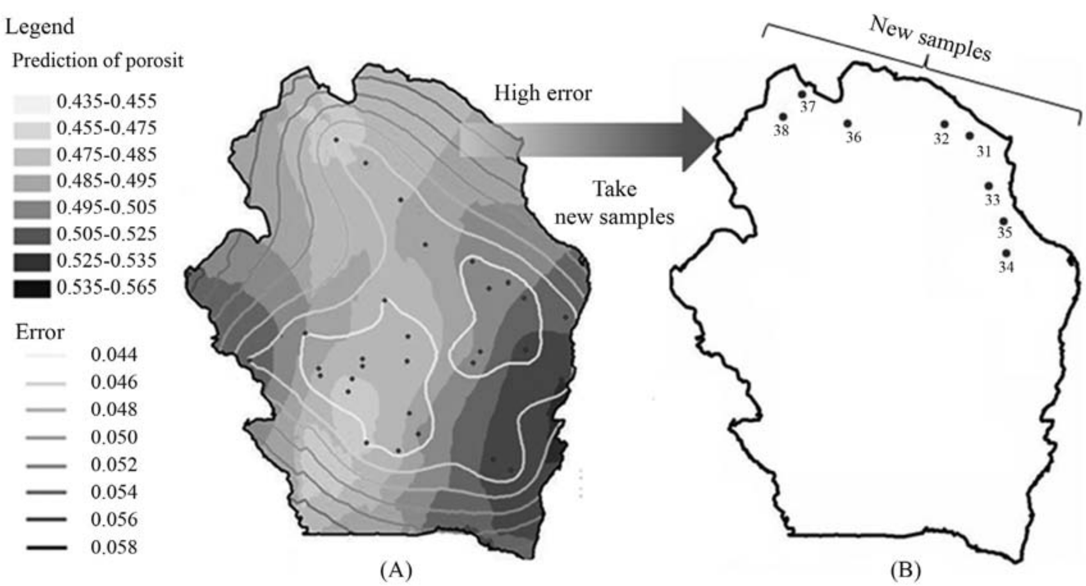


Figure 7 - (A) Prediction of surface porosity and the distribution of error associated; (B) the new locations of soil sampling.

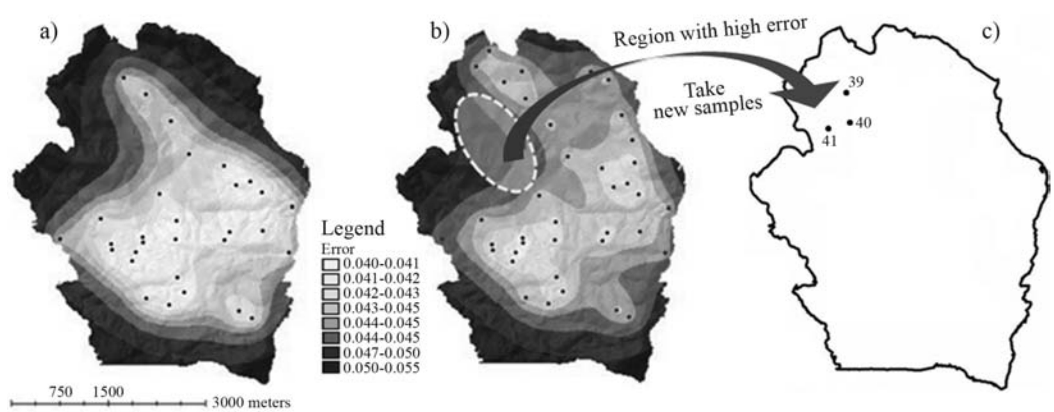


Figure 8 - Maps of error associated with the prediction of surface porosity: (A) database with 30 samples; (B) database with 38 samples. Map (C), location of the 3 new soil samples.

Table 1 - Comparison between the values of soil porosity obtained by laboratory analysis (Real value) and estimated by kriging (Estimated value).

Sample	Estimated value	Real value	Real error	Estimated error
31	0.490	0.474	0.016	0.033
32	0.495	0.520	0.025	0.048
33	0.495	0.465	0.030	0.061
34	0.510	0.414	0.096	0.188
35	0.505	0.420	0.085	0.169
36	0.482	0.489	0.007	0.015
37	0.475	0.459	0.016	0.034
38	0.492	0.477	0.015	0.031

collect three new samples in this region, indicated by the samples 39, 40 and 41 located at the right on the map in Fig. 8.

At this stage of the analysis, the sampling procedure was completed with 41 samples to create the final map. Their spatial coordinates and values of porosity for the three depths analyzed are shown in Table 2. No sampling was made at the extreme southern and western parts of the study area because access difficulties (steep slopes and dense forest).

Before carrying out the final kriging process the quality of data was evaluated in order to determine whether they present a good spatial dependence and characteristics that are consistent with the basic principles of geostatistics.

3.2. Final maps of porosity

To create the final maps with the predictions of the porosity for the three different depths, it is necessary to investigate whether the distribution of the porosity variable allows the application of geostatistic methods (structural and exploratory analysis). After this step, it was analyzed if the results of the predicted porosity were consistent with the values from the literature that's correlate porosity regarding geological units.

3.2.1. Histograms

To examine the quality of database, some authors proposed the use of histograms to verify the kind of data distribution (Folegatti, 1996; Houlding, 2000). Folegatti (1996) states that for variogram adjustments, the normal distribution of data is just desirable but not a necessary prerequisite. If the distribution found of data is not normal, but is reasonably symmetrical, it is possible to accept the assumptions necessary for the construction of the semivariogram.

The traditional histogram only shows the frequency of some bands of variable values, for example points having porosity between 0.41 and 0.43, but the distance between samples is ignored. Therefore, although this is not a

Table 2 - Final dataset with 41 sampling points and their attributes.

Sample	Coordinates in UTM		Porosity (η)		
	X	Y	Surface	50 cm	100 cm
1	490663	7470340	0.51	0.46	0.51
2	490651	7471193	0.42	0.4	0.41
3	490839	7472196	0.5	0.52	0.43
4	490638	7470914	0.49	0.4	0.46
5	489677	7470758	0.47	0.37	0.36
6	489664	7470842	0.5	0.52	0.5
7	490034	7470719	0.45	0.55	0.45
8	490144	7470868	0.54	0.52	0.53
9	490142	7470946	0.51	0.44	0.46
10	490541	7469925	0.46	0.49	0.44
11	490193	7470021	0.46	0.46	0.45
12	488892	7470923	0.6	0.58	0.53
13	489506	7471225	0.5	0.49	0.4
14	489983	7470585	0.42	0.45	0.42
15	491789	7469721	0.6	0.46	0.48
16	491596	7469830	0.52	0.5	0.49
17	490767	7470114	0.53	0.46	0.43
18	492348	7470712	0.58	0.51	0.56
19	491953	7471043	0.61	0.46	0.53
20	491452	7471027	0.47	0.48	0.51
21	491376	7470901	0.51	0.45	0.44
22	489854	7473350	0.47	0.48	0.5
23	490568	7472693	0.5	0.44	0.5
24	490185	7473099	0.47	0.59	0.57
25	492399	7471405	0.49	0.5	0.56
26	491764	7471779	0.51	0.47	0.5
27	491940	7471615	0.52	0.44	0.48
28	491550	7471725	0.5	0.43	0.43
29	491368	7472015	0.54	0.5	0.45
30	490395	7471591	0.49	0.45	0.53
31	491471	7473389	0.47	0.48	0.48
32	491194	7473517	0.52	0.49	0.39
33	491673	7472839	0.47	0.56	0.51
34	491873	7472103	0.41	0.44	0.43
35	491839	7472453	0.42	0.38	0.42
36	490129	7473522	0.49	0.49	0.57
37	489625	7473842	0.46	0.42	0.53
38	489422	7473598	0.48	0.51	0.46
40	489707	7472633	0.45	0.51	0.55
39	489670	7473159	0.48	0.54	0.51
41	489361	7472539	0.51	0.52	0.5

common practice, it was established relations between the intervals of higher frequency of the histogram and the location of samples which is composed (Fig. 9). This analysis can help identify whether the most similar samples are located close or far apart.

From Fig. 9, it is observed that both results, on the surface as well as at 100 cm depth are consistent with one of the main principles of the geostatiscal approach that state that samples taken closer together are more similar on average (ESRI, 2001; Nicol *et al.*, 2003). For the depth of 50 cm, similar samples are scattered throughout the region and are far from each other.

3.2.2. Variograms

The analysis was performed aiming the identification of the semivariogram model that best fits the distribution of points. This step is undertaken by using the ArcGIS® software which seeks to establish an approximate semivariogram via the observed distribution. Figure 10 represents this check and, when the parameters that define them are indentified, they are highlighted: (a) range, (C) sill and (Co) nugget effects. Therefore, the variogram was fitted by a spherical theoretical curve with a nugget and used for subsequent analysis.

It is possible to observe at Fig. 10 that the data collected on the surface (Fig. 10-a) .and at 100 cm depth (Fig. 10-c) showed a spatial distribution which allowed the adjustment of the semivariogram model. On other hand, the dispertion of the data for the depth of 50 cm (Fig. 10-b) made impossible to adjust any theoretical variogram. For this reason, it was impossible to create prediction maps from kriging method for 50 cm depth with the risk of generating inconsistent results. Therefore, the further investigation was carried out for these two depth, in order words, the kriging analyses were applies to predict surface and at 100 cm depth porosities.

One of the most important steps is to find an theoretical variogram bases on experimental variogram (circular, spherical, gaussian, among others), which best approximates to the observed distribution. From the theoretical variogram chosen, the interpretation of the spatial correlation structure using krigings inferences is made (Camargo *et al.*, 2004; Silva, 2005). Using the *Geostatistical Wizard* tool - which makes up the *Geostatistical Analyst* extension of ArcGIS, it was analyzed the variability in both depths (superficial and 100 cm depth). Thus, it was obtained the variograms presented in Fig. 11 for both depths using the Spherical model as a model for the theoretical variogram.

3.2.3. Search neighborhood settings

It is common in practice to specify *search neighborhood* that limits the number and the configuration of the

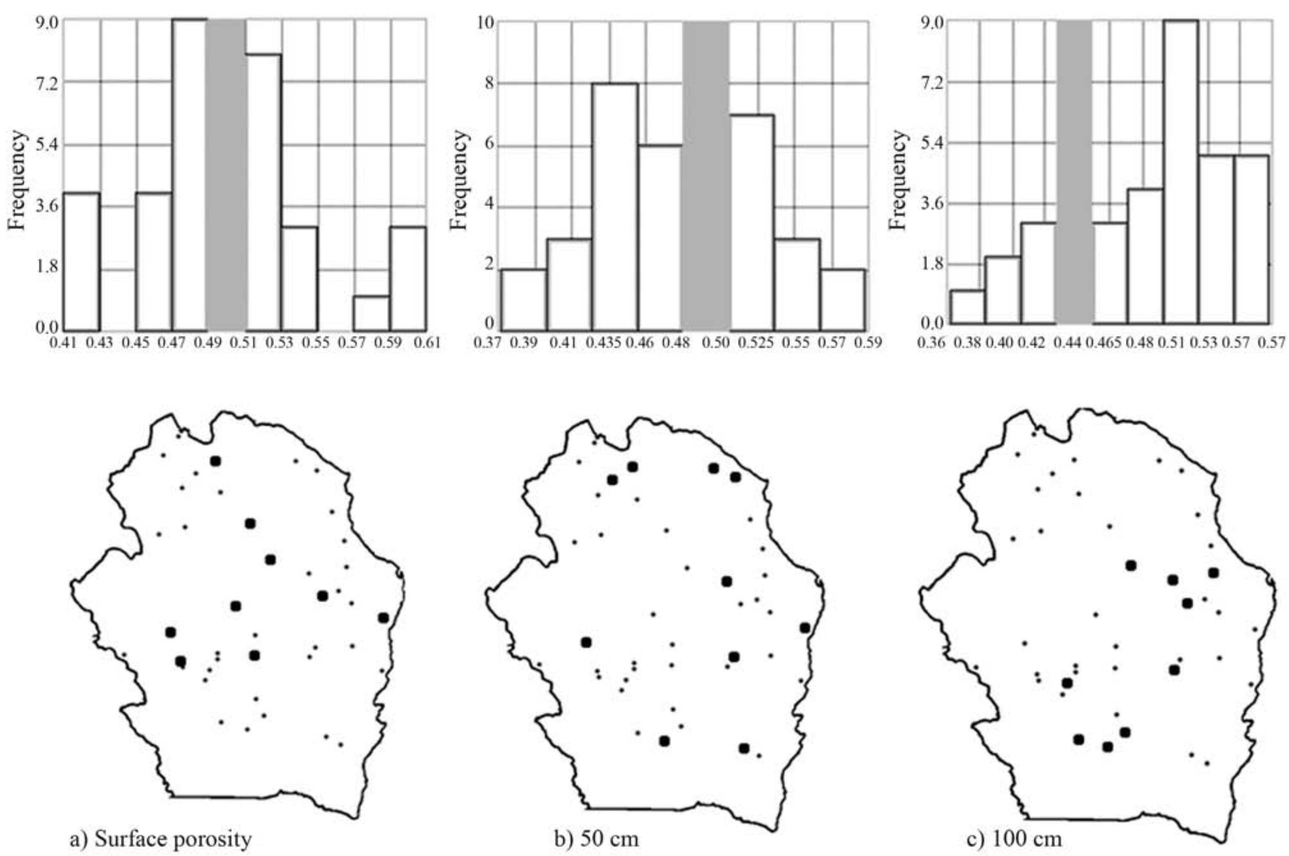


Figure 9 - Histograms of the spatial distribution - a) surface, b) 50 cm, c) 100 cm.

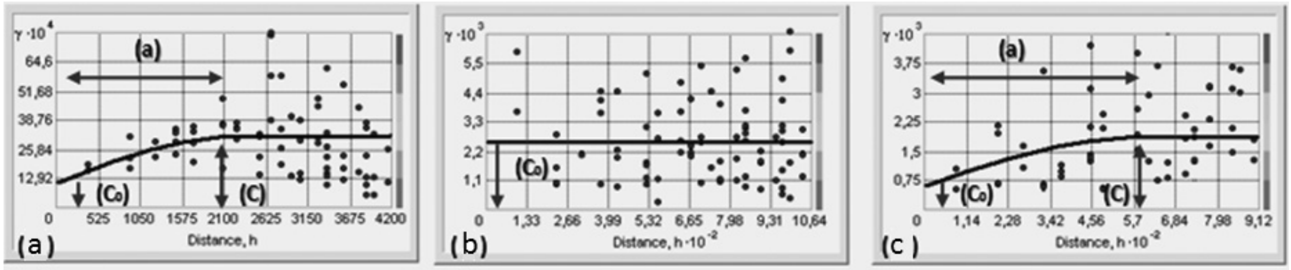


Figure 10 - Variograms for the depths: (a) surface, (b) at 50 cm and (c) at100 cm.

points that will be used in the predictions. There are two controlling mechanisms to limit the points used: specifying the shape of neighborhood and establishing constraints on the points within and outside the shape (ESRI, 2001).

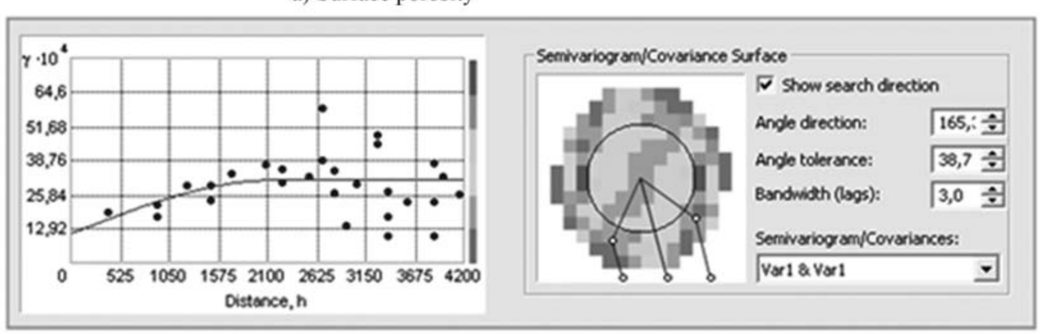
From the Geostatistical Wizard tool, it is possible to choose the number of surrounding samples that will influence the estimation for a non-sampled location (Neighbors to Include), that is, this number of samples will be included inside the *search neighborhood*. This settings is carried out by the user, generally by trial and error, verifying the number which provides the highest root-mean-square (RMS) and which gives value closer to unity.

For this settings, it is common to divide the *search neighborhood* in sectors because it contributes to a more homogeneous inclusion of the neighbors samples. For the present work, the *search neighborhood* was divided in four sectors in order to always use samples from different portions of the area. It was set to included three samples per sector, always including at least two samples, because in some locations of the *search neighborhood* can have insufficient samples, requiring its extrapolation (see Fig. 12).

The last adjustments made before the generation of the final maps is the definition of the geometric shape of the *search neighborhood.* Generally, the parameter *range* shapes the *search neighborhood*, which acquires the shape of a circle with radius exactly equals to it. However, it is possible to modify these settings if there are explanations for realize such changes.

For this present research, it was not used a circular shape. A factor of anisotropy for providing an ellipse shape was intentionally used and rotated to 40º. This settings facilitates the use of samples with a higher degree of similarity, reducing the probability of samples with different properties influencing the estimates and, thus, improves the quality of result. These settings were done in accordance

a) Surface porosity



b) Porosity at 100 cm

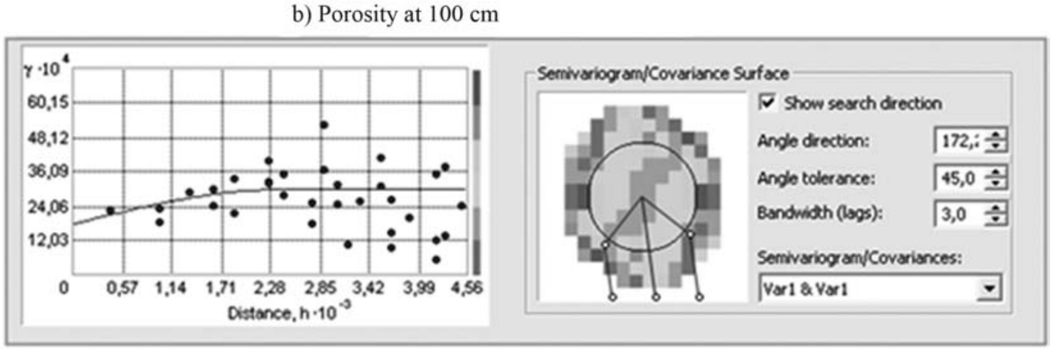


Figure 11 - Directional variograms obtained: a) surface porosity and b) porosity at 100 cm.

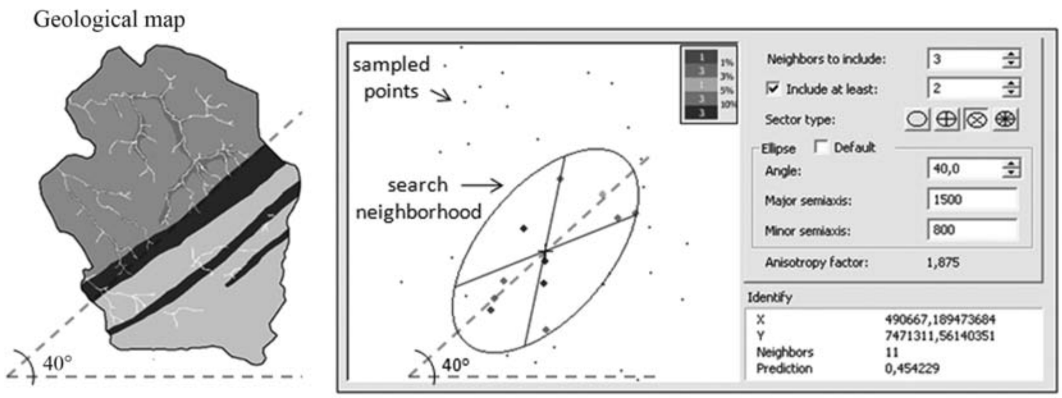


Figure 12 - The Geological Map and the adjustments done to *Search Neighborhood's* shape.

with the Geological Map (Fig. 2), which can be observed a shear zone with same orientation.

Figure 12 shows the ellipse shape of the *search neighborhood* for the surface porosity data, with 1500 m and 800 m for major and minor semi-axis, respectively. The value of 1500 is nearly the same value of the *range* parameter observed in directional variograms (Fig.11).

The same criteria were used for the settings used to porosity data at the 100 cm depth. The only modification was the value adopted by major and minor semi-axis, which was set to 1200 and 600 m, respectively. This reduction was necessary because higher semi-axis values have implied in lower RMS and the values predicted to porosity have presented values nearly equal throughout the area.

3.2.4. Porosity prediction by kriging and associated error

The next step of the calculation is the application of kriging with the established parameters such as, the semivariogram and respective direction, the size of the area search, number of neighbors included. The software ArcGIS® is now able to run the mathematical algorithm to calculate the spatial value of the porosity as presented in Fig. 13 for the two investigated depth.

The kriging method also allows the knowledge of the distribution of the error associated with estimates as explained in section 2.4.2. The quality of the maps for prediction of porosity presented in Fig. 13 depends essentially on the distribution of the error existing in the methodology. Figure 14 shows this distribution error for the two depths investigated. For the two maps, the darker areas indicates higher error which means that the predicted porosity is less reliable at this regions and, on the other hand, lighter areas indicate good quality of the prediction.

3.3. Discussion

For geotechnical mapping based on geostatistical analysis, the number of sampling must be sufficient. Iincreasing the number of samples will result in a better representativeness of the variable on the proposed model, reducing the spread of the error. Unfortunately, there are some site characteristics, topography, vegetation, which field works become restricted.

In order to avoid insufficient data or excessive field activities, the proposed methodology permits the identification of areas with some peculiarities which is indicated by the error values. This behavior was observed in points 34 and 35, where the porosities are different from the samples that make up your neighborhood, requiring the increase of sampling process in this region for statistical analysis to become more consistent. This specific area can also characterize a particular geological condition, which must be considered individually.

The geostatistics analyses made from parameters collected at surface and at 100 cm depth gave the spatial dependence which allowed the adjustment of the semivariogram model, allowing the use of kriging of the spatial porosity predictions. This analysis was not possible for the depth of 50 cm due to the scatter of the field data. In this case, a more detail field investigation is required to understand the reason of such variability. Depending on the geomorphology, the variability can be attributed also to the thickness of this intermediate layer and detail investigation must be carried out in small areas.

In general, Fig. 14 indicates that the data collected on the surface have better representativeness (lower error) when compared with 100 cm depth. For the surface data, there is a larger amount of samples that have similar values and are spatially close together. This fact represents a greater homogeneity of the surface porosity throughout the study area.

The analysis for 100 cm depth showed lower porosity when compared with the surface porosity. This fact can be explained by the interference of the rooting system of the plants at shallow depth which contributes to more porous soils.

Comparing the predicted maps of porosity with the geological map developed in scale 1:10.000 (Fig. 2), it is possible to make some considerations regarding the results of the geoestatistic analyses with the geological units.

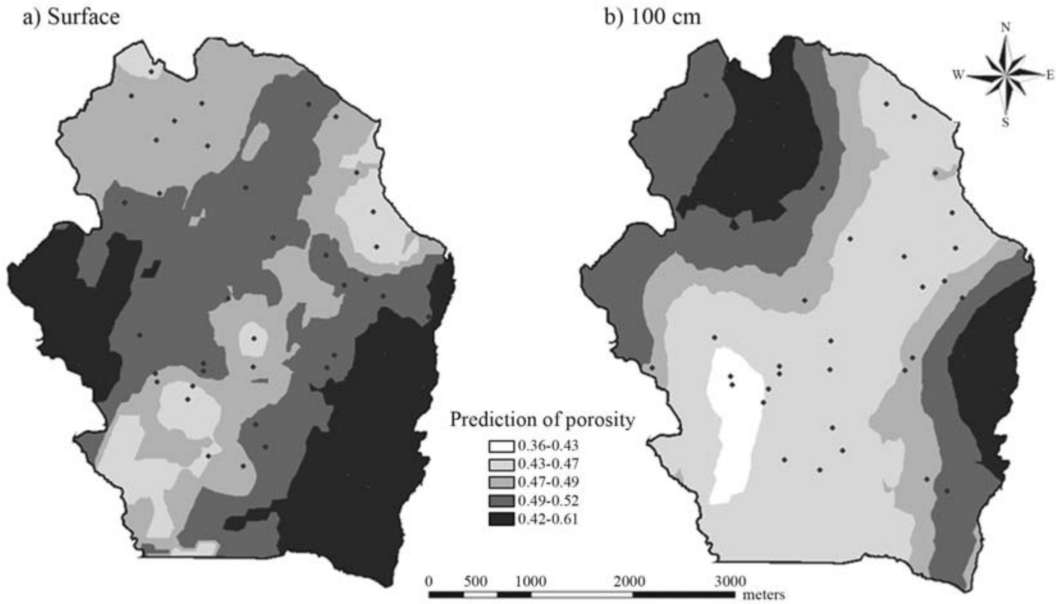


Figure 13 - The final prediction maps generated for the surface porosity (a) and at 100 cm depth (b).

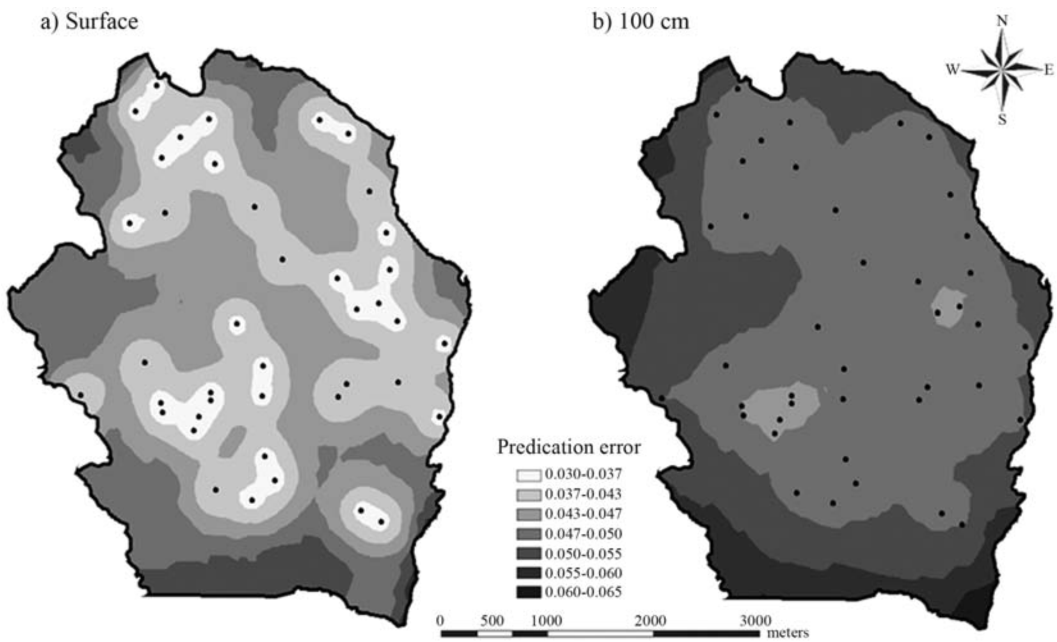


Figure 14 - Maps of the distribution of error at (a) surface and (b) 100 cm depth.

The geological unit "Granitoids Quebra-Cangalha" is mainly composed of leuco-granite, which occupies the entire southeast study area. The soils from this geological unit have, mostly large amount of coarse material, sand and silt, characteristic of young residual soils, little thick layer, heterogeneous and that would be classified as Cambisols in Pedology. According to the literature, it is expected that these soils have porosity in order of 55% at the surface and reducing with depth. Observing the porosity maps in Fig. 13, it can be observed that the measured and predicted spatial porosity varies from 53% at the surface to 43% at 100 cm depth. This difference in porosity with increasing depth is one of the factors responsible for the physical conditions that impose restriction to water infiltration process and can induce erosion processes at Santa Edwiges farm (Santos, 2007).

The other geologic unit, is consisted of mylonitic rocks, in which is inserted a shear zone with NE-SW direction, Fig. 2. Due to the rocks formation derived from retrometamorphism and having many foliation plans, the weathered processes are likely to produce changes in surface and to produce thin soils with a predominant clay fraction. In the other hand, the susceptibility of erosion of these rocks often causes them to be associated with relatively low relief. In fact, this is what is observed in the study area, where the mylonitic zones occupy areas of relatively low height compared to the other geological units such as Granitoids Quebra-Cangalha and the Embu Complex. In the area cropped by the shear zone, the porosity does not show significant variation depending on the depth, varying from 38% to 48% in both cases.

The Embu Complex has a wide variety of rock types, but in the northern portion of the farm there is a predominance of gneissic rocks rich in biotite. The soils of this unit are well developed, mature residual soils with clay texture, thick and homogeneous, which would be classified as Latossolos. The correlation between the sequence of gneissic rocks and predicted maps of porosity indicates an increase of porosity with depth, from 45% to 53%, particularly at the northern end of the study area. This increase might be related to conditions of good drainage and water circulation within these soils, is also an indication that they do not have a high degree of compaction to a meter deep. However, in the eastern portion of the Embu Complex, there was a reverse situation, where the values of porosity were higher at the surface (55%). However, this specific situation should be better assessed by being in a region where the error associated with the estimates was high (Fig. 14), thus requiring a larger number of samples.

These considerations suggest a relationship between the spatial patterns of porosity and units of the geological map presented at scale 1:10.000. With increasing depth, porosity increases in the area of granite rocks, reduces in the area of gneissic rocks rich in biotite and remains relatively unchanged in the fields of rocks strongly deformed (shear zones). In the western portion of the study area, where large errors were found associated with scarcity of samples, the model was not efficient for this analysis, requiring a larger amount of data for future characterization of this area.

4. Final Considerations

This manuscript presents a methodology for correlating soil porosity to the respective geological units using geostatistical analysis techniques, including interpolation data by kriging. Therefore, the comparison between a porosity prediction map - based on geostatistical approach - and a geological map constitutes a different way of comparing these two parameters. Furthermore, the "step by step" method presented in this paper (Fig. 2), in which each sampling procedure was used as a basis for the next step, may be considered a useful method to produce better prediction map for geotechnical parameters.

The proposed methodology permits the identification of areas with some peculiarities, which are indicated by the error values. This analysis helps to avoid insufficient data or excessive field activities. The behavior observed in points 34 and 35, where the porosities are different from the neighborhood, indicates the necessity of extra soil sampling in order to evaluate if it was a lack of data for geostatistic analysis or a specific area that must analyzed separately.

For two depths the geostatistics analyses using kriging were possible because the data shows spatial dependence, verified by the semivariogram constructions. This analysis was not possible for the depth of 50 cm due to the dispersion of the field data. Depending on the geology and the geomorphology, the variability can be attributed also to the thickness of this intermediate layer and detail investigation must be carried out in small areas.

It can be noticed that the database used can still be complemented in order to minimize the prediction error and provide a better quality of the results. Even with these matters, the study confirmed the existence of spatial patterns between spatialization of the soil porosity and geology units, which represents a potentiality for correlating with others soil properties, such as hydraulic conductivity, soil water retention curves and erosion potentials.

References

Burrough, P.A. (1998) Principles of Geographical Information Systems for Land Resources Assessment. Clarendon Press, Oxford, 194 pp.

Camargo, E.C.G. (1997) Desenvolvimento, Implementação e Teste de Procedimentos Geoestatísticos no Sistema de Processamento de Informações Georreferenciadas. M.Sc. Dissertation, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 123 pp (In Portuguese).

Camargo, E.; Druck, S. & Câmara, G. (2004) Análise espacial de superfícies. In: Druck, S.; Carvalho, M.S.; Câmara, G. & Monteiro, A.V.M. (eds) Análise Espacial de Dados Geográficos. Brasília, EMBRAPA, 209 pp. Available at http://www.dpi.inpe.br/gilberto/livro/analise/cap3-superficies.pdf.

Camarinha, P.I.M. (2008) Análise geoestatística da variável porosidade - Uma busca por correlações com parâmetros do meio físico e propriedades hidráulicas do solo. Undergraduated Research Report. Universidade Estadual Paulista, Guaratinguetá, 90 pp (In Portuguese).

Camarinha, P.I.M.; Simões S.J.C. & Bernardes, G.P. (2008) Variabilidade espacial de propriedades físicas dos solos em áreas de relevos acidentados – O médio Vale do Paraíba do Sul, Sudeste do Brasil. Proc. Congresso Luso-Brasileiro de Geotecnia. Coimbra, Portugal, pp. 193-200 (In Portuguese).

Carneiro, C.D.R.; Hasui,Y. & Giancursi, F.D. (1978) Estruturas da Bacia do Taubaté na região de São José dos Campos (Structures of the Taubaté Basin region of

São José dos Campos region). Proc. XXIX Congresso Brasileiro de Geology, Ouro Preto, v. 4, p. 247-256.

Cressie, N.A.C. (1993) Statistics for Spatial Data. John Wiley, New York, 900 pp.

Domingos, L.A. (2005) Determination of some geotechnical characteristics of superficial soil in the sub-basin of Ribeirão Taboão, Lorena, São Paulo, as a basis to study the characteristics of soil infiltration. Report on Scientific Initiation, São Paulo State University, Guaratinguetá, São Paulo.

ESRI (2001) ArcGIS Geostatistical Analyst: Statistical Tools for Data Exploration, Modeling, and Advanced Surface Generation. ESRI, Redlands, 19 pp. http://www.esri.com/library/whitepapers/pdfs/geostat.pdf and downloaded on May 14th, 2010.

Folegatti, M.V. (1996) Temporal stability and spatial variability of moisture and of water storage in silty soil. Thesis. Agricultural College "Luiz Queiroz", São Paulo State University (USP), Piracicaba, São Paulo, 96 pp.

Goovaerts, P. (1999) Geostatistics in soil science: State-of-the-art and perspectives. Geoderma, v. 89:1-2, pp. 1-45.

Gomes, N.M.; Faria, M.A.; Silva, A.M.; Mello, C.R. & Viola, M.R. (2007) Variabilidade espacial de atributos físicos do solo associados ao uso e ocupação da paisagem. Revista Brasileira de Engenharia Agrícola e Ambiental, v. 11:4, p. 427-435 (In Portuguese).

Gumiaux C.; Gapais D. & Brun J.P. (2003) Geostatistics applied to best-fit interpolation of orientation data. Tectonophysics, v. 376:3-4, p. 241-259.

Hasui, Y. & Sadowski, G. (1976) Evolução geológica do precambriano na região sudeste de São Paulo. Revista Brasileira de Geociências, São Paulo, v. 6, p. 187-200 (In Portuguese)

Houlding, S.W. (2000) Practical Geostatistics: Modeling and Spatial Analysis. Berlin, Springer, 179 pp.

Isaaks, E.H. & Srivastava, R.M. (1990) An Introduction to Applied Geostatistics. Oxford University Press, New York, 592 pp.

Krige, D.G. (1951) A statistical approach to some basic mine evaluation problems on the Witwatersrand. Johanesburg Chemistry Metallurgy Mining Society South African, Johanesburg, v. 52, p. 151-163.

Landim, P.M.B. (1984) Mapa Geológico do Estado de São Paulo, Escala 1:50.000. IGCE/UNESP, São Paulo (In Portuguese).

Landim, P.M.B. (2003) Análise Estatística de Dados Deológicos. Universidade Estadual Paulista, Rio Claro, 2ed, 254 pp (In Portuguese).

Landim, P.M.B. (2006) Análise Estatística Espacial de Dados Geológicos Multivariados. Universidade Estadual Paulista, Rio Claro, 158 pp (In Portuguese).

Landim, P.M.B.; Sturaro, J.R. & Monteiro, R.C. (2002) Krigagem ordinária para situações com tendência regionalizada. DGA, IGCE, UNESP, Rio Claro, 13 pp. (In Portuguese). Available at http://www.rc.unesp.br/igce/aplicada/didaticos/landim/tkrigagem.pdf and downloaded on 25th January 2010.

Lima, B.A. (2005) Desenvolvimento de um mapa geomorfológico como subsídio a um programa de plantio de eucalipto - Fazenda Santa Edwirges, Lorena. Proc. XIII Simpósio Internacional de Iniciação Científica da USP, São Carlos (CD-ROM - In Portuguese).

Matheron, G. (1963) Principles of geostatistics. Economic Geology, v. 58, pp. 1246-1266.

Matheron, G. (1971) The theory of regionalized variables and its applications. Paris, Les Cahiers du Centre de Morphologie Mathematique in Fontainebleu, 211 pp.

Nicol, G.W.; Glover, A. & Prosser, J. (2003) Spatial analysis of archaeal community structure in grassland soil. Applied and Environmental Microbiology, v. 5:3, p. 152-162.

Ramsay, J. (1980) Shear zone geometry: A review. Journal of Structural Geology, v. 2:1-2, pp. 83-89.

Ribeiro Junior, P.J. (1995) Métodos Geoestatísticos no Estudo da Variablidade Espacial de Propriedades Físicas do Solo. M.Sc. Dissertation, São Paulo State University, São Paulo, 99 pp (In Portuguese).

Santos, C.M.Z.; Bernardes, G. P.; Simões, S.J.C. & Automare, G.B.B. (2007) Metodológica para a caracterização de áreas com potencial de erosão. Proc. III Simpósio sobre solos tropicais e processos erosivos no Centro-Oeste, Cuiabá, v. 1 Art. 38.

Silva, P.C.F. da; Mendes, R.M.; Nery, T.D. & SILVA, I.C. (2007) Análise espacial de propriedades geotécnicas de solos do município de Ubatuba a partir de modelagem determinística local e geoestatística. Proc. XI Simpósio Brasileiro de Cartografia Geotécnica e Geoambiental, Uberlandia (CD-Rom).

Silva, W.M. (2005) Análise da precipitação com base em Krigagem - Bacia do Rio Paraíba do Sul. Undergraduated Research Report. Universidade Estadual Paulista, Guaratinguetá, 87 pp (In Portuguese).

Vieira, S.R.; Nielsen, D.R.; Biggar, J.W. (1981) Spatial variability of field-measured infiltration rate. Soil Science Society of America Journal, v. 45, p.1040-1048.

Webster, R.; Oliver, M.A. (1993) How large a sample is needed to estimate the regional variogram adequately? Soares, A. (ed) Geostatistics Troia '92. Kluwer, Dordrecht, pp. 155-166.

Zuquette, L.V.; Gandolfi, N. (2004) Cartografia Geotécnica (Geotechnical Mapping). Oficina de Textos, São Paulo.

Zuquette, L.V. (1987) Análise Crítica sobre Cartografia Geotécnica e Proposta Metodológica para as Condições Brasileiras. Ph.D. Thesis, Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.