EXPLORING THE MULTIVARIATE SPATIAL STRUCTURE OF SOIL ACIDITY DATA

Manzione RL, Câmara G, Monteiro AMV, Zimback CRL, Druck S

INPE-National Institute for Space Research. Av.dos Astronautas, 1758. C.P. 515, 12245-970 S. J. dos Campos, SP – Brazil. e-mail: manzione@dpi.inpe.br

SUMMARY

Soil acidity is a frequent problem in tropical soils, which affects major agricultural crops. The main effect of high acidity is the availability of Aluminum (Al) on soil, which is toxical to plants. On an attempt to explore the multivariate spatial structure of soil acidity data, the variables Calcium (Ca), Magnesium (Mg), active acidity (pH) and potential acidity (H+Al) were measured by 204 soil samples collected 0-20 depth from a 60 m regular grid on a corn/soybeans crop at Araguari-MG, Brazil. They were chemically analyzed as well. Under scope of Linear Model of Corregionalization (LMC), direct and cross variograms were modeled with a nugget effect and a spherical model at 451,28 m range, analyzing corregionalization structure range among soil acidity variables, in parallel with Principal Component Analysis (PCA). This analysis captured acidity phenomena structure in micro scale (nugget effect) with the influence divided between Ca and Mg elements and pH of this soil, given by an intense agronomical handling which was submitted to the soil over the years. On a long range the phenomenon was explained by H+Al, given by the geological soil influence (pedological formation). The eigenvalues of the structures modeled given by PCA show that nugget effect is an important part of the model. The model proposed estimated by cokriging was filtered, once interested on long range variability. The filtering process removes high frequencies signaled by nugget effect, resulting in a more continuous map of the acidity phenomenon on this soil.

INTRODUCTION

Aluminum (Al) represents a serious problem in natural acid soils and in soils acidified by human activity. The main effect of high soil acidity is the availability of Al on soil, which is toxical to plants. Much attention has been paid to the determination of the content of labile aluminum forms on soil, and pH is generally the most important factor controlling aluminum solubility and mobility. However, simple practices like lime application can solve that question, making Al unavailable to roots, besides providing Calcium (Ca) and Magnesium (Mg) to plants. A lot of factors interfere in this process, by soil handling (human factor) or by soil type (pedological factor). A soil sampling can help on this investigation. Farmers are interested not only in spatial variation of soil properties but also on the sources of variation. If human activities are cause of deficiencies or excesses of some element, farmers should modify their farming management. Optimum benefits on profitability and environment protection depend on how well agronomic practices are fitted to variable soil conditions. It is critical to characterize soil with precision, both quantitatively and spatially.

Soil variability is the result of both natural processes and management practices, acting at different spatial and temporal scales. Natural variability results from complex geological and pedological processes. Soil forming factors, such as parent material, biota, climate, and topography, explain most of general variability; however management practices may also affects soil variability significantly. Some factors which govern soil variations have a short range action, whereas others operate at longer distances. As a consequence, soil variables are expected to be correlated in a way that is scale-dependent (Castrignanò et al. 2000). Kriging is accepted by many soil scientists as an appropriate technique for estimating the values of soil properties at unsampled sites and over larger blocks of land. For properties such as nutrient concentration and lime requirement, the maps may be used to determine how much fertilizer or lime to add to the soil for a particular crop or farming regime. Coregionalization analysis is more revealing than univariate geostatistical analysis to model the scale-dependent correlation structure of soil properties (Castrignanò et al. 2000). This requires a particular statistical approach that combines classical factor analysis for describing the correlation structure of multivariate data sets, with geostatistics, to take into account the regionalized nature of the variables. The method called Factorial Kriging Analysis (FKA) was developed by Matheron (1982) and enable to modeling the correlations among the soil physical and chemical properties at each of the different spatial scales. This work had as objective study the spatial variability of soil acidity at a precision farming field in central Brazil and search plausible explanations of their distributions among different scales.

DESCRIPTION OF PROPOSED METODOLOGY

FACTORIAL KRIGING ANALYSIS AND LINEAR MODEL OF CORREGIONALIZATION

Factorial Kriging Analysis is a geostatistical method for analyzing multivariate spatial data set. The theory underlying FKA has been described and evaluated in earlier papers (Goulard and Voltz 1992, Goovaerts 1992, Castrignanò et al. 2000, Lark and Papritz 2003) and textbooks (Wackernagel 1995, Goovaerts 1997). Here, we will describe only the basic steps, as the following:

Modeling the coregionalization of the variables using linear model of coregionalization (LMC)

The most commonly used model is the linear model of coregionalization (LMC). Journel and Huijbregts (1978) consider all the studied variables to be generated by the same independent physical processes acting at N_s different spatial scales. The $p(p+1)/2$ experimental direct and cross-variograms are modelled as linear combinations of the same set of N_s basic variogram functions $g^u(\boldsymbol{h})$:

$$\gamma_{ij}(\mathbf{h}) = \sum_{u=1}^{Ns} b_{ij}^u g^u(\mathbf{h}) \qquad i, j = 1,..., p \tag{1}$$

where $\gamma_{ij}(\mathbf{h})$ is the cross-variogram model between variables Z_i and Z_j; u denotes the spatial scale; $\boldsymbol{h}$ is a distance vector, the lag, and b_{ij}^u are the coefficients of the function $g^u(\boldsymbol{h})$. Using the matrix notation, the LCM can be rewritten as:

$$\boldsymbol{\Gamma}(\mathbf{h}) = \sum_{u=1}^{Ns} \mathbf{B}^u g^u(\mathbf{h}) \tag{2}$$

where $\boldsymbol{\Gamma}(\boldsymbol{h})$ is a p x p symmetric matrix whose diagonal and off-diagonal elements are the direct and cross-variogram values, respectively, for a given lag $\boldsymbol{h}$; $\mathbf{B}^u$ is a p x p symmetric matrix of coefficients b_{ij}^u, called the coregionalization matrix.

In the LMC, the spatial behaviour of the variables is supposed to be resulting from superimposition of different independent processes working at different spatial scales. As regards soil science, these processes may include soil forming factors (parent material, topography, vegetation, climate, and time), farming practices (tillage, crop rotations, and fertilization), and erosion. These processes may affect the behaviour of experimental variograms, which can then be modeled by a set of functions $g^u(\mathbf{h})$. The choice of number and characteristics (model, sill, and range) of the functions $g^u(\mathbf{h})$ is quite delicate and can be made easier by a good experience of the studied phenomena.

Fitting of the LMC can be performed by weighted least-squares approximation under the constraint of positive semi-definiteness of the $\mathbf{B}^u$. The best model was chosen, as suggested by Goulard and Voltz (1992), by comparing the goodness of fit for several combinations of functions of $g^u(\mathbf{h})$ with different ranges in terms of the weighted residual sum of squares. Lark and Papritz (2003) suggested simulated annealing as a method for automated fitting of variogram functions to auto and cross-variogram estimates, minimizing a weighted sum of squares between the observed and modeled variograms.

Analyzing the correlation structure between variables applying principal component analysis (PCA)

Classical multivariate methods may be applied to matrices $\mathbf{V}$, $\mathbf{\Gamma(h)}$ and $\mathbf{B}^u$. Among techniques described by Wackernagel et al. (1989), principal components analysis is commonly chosen to decompose the coregionalization matrices $\mathbf{B}^u$ into matrices $\mathbf{A}^u$ and to summarize relation between variables. New orthogonal variables, the principal components, which are linear combinations of the original variables with coefficients contained in the eigenvectors of the matrix are found (Goovaerts 1992). Each principal component explains a percentage of total variance that corresponds to the ratio of the associated eigenvalue to the trace of the matrix. The first few components generally account for most of the variance and so are the most informative.

The classical variance–covariance matrix is a combination of coregionalization matrices related to different spatial scales. However, when correlations between variables change with the spatial scale, the classical approach of the variance–covariance matrix does not appropriately describe such correlations, whereas the FKA considers correlations between variables at each spatial scale by means of the coregionalization matrices $\mathbf{B}^u$. As shown by Goovaerts (1992), classical principal components are combinations of the main relationships between variables at different spatial scales, whereas the geostatistical approach increases the explanatory power of factorial analysis, because it allows distinction among factors according to their spatial scale of variation.

Estimating and mapping the regionalized factors and variables

The LCM allows the estimation of the regionalized factors by cokriging. Information on the derivation of the cokriging system is available in Wackernagel (1995). The scores of the regionalized factors can be mapped to show the behaviour and relationships among variables at different spatial scales more clearly.

The behaviour and relationships between variables at different spatial scale can be illustrated by mapping the regionalized variable $Z_i(\mathbf{x})$, the spatial components $Z_i^u(\mathbf{x})$, and the regionalized factors $Y_v^u(\mathbf{x})$. In each case the estimation is performed by cokriging. The estimator is written:

$$CK^*(\mathbf{x_0}) = \sum_{i=1}^{p} \sum_{a=1}^{n} \lambda_{ai} Z_i(\mathbf{x_a}) \tag{3}$$

where p is the number of variables, n is the number of data points (surrounding the point $\mathbf{x_0}$) used in the estimation and λ_{ai} is the weight assigned with the value of the ith variable at the ath location, $Z_i(\mathbf{x_a})$. Minimizing the estimation variance under the constraint of unbiasedness leads to a system of $p(p+1)$ linear equations, the solution of which provides the weights λ_{ai}. The three cokriging system and further information on the derivation of these systems is available in Wackernagel (1995) and Goovaerts (1997).

FILTERING COMPONENTS

The model is the combination of a drift part and a covariance part. The exactitude property of the Kriging estimator creates discontinuities (peaks) at data location (Goovaerts 1997). Factorial Kriging Analysis can also be applied in order to filter out some covariance or drift components of the model. If the model is made of several nested basic structures, you may consider that the phenomenon is a combination of several components, at different scales, which could be filtered out during the kriging process.

One possibility to remove such discontinuities at data location is filtering the nugget (high frequency) component. By filtering out the appropriate components of the covariance part in the model, could estimate only the short range (high frequency) components or the long range (low frequency) components, instead of the entire phenomenon. Moreover, is possible also allows filtering out components of the drift part of the model, introduced on the model or given by universality condition. Filtering the universality drift term consists in removing the contribution of the mean (either local or global depending on your neighborhood choice), similarly to filtering the X and Y drift that consists in removing the linear trend (Geovariances 2003). The estimation of the error should be unbiased, whatever the unknown drift might be. In addiction to being an unbiasedness constraint, the presuppose of the universality conditions is to filter out the unknown drift components in the random function (Olea 1991).

CASE STUDY

The area studied covers 71,79 ha of a rural property at Araguari-MG, Brazil, coordinates 18^{o} 40’ south latitude and 48^{o} 15’ west longitude, in which corn and soybeans are cultivated under rotation system. Details about soil type, climate, vegetation and topography are present in Manzione (2002). Georeferrenced samples were taken allow a known 60 m square grid, performing a total of 204 soil samples collected 0-20 depth (Figure 1). Physical and chemical parameters were measured on soil samples and analyzed according to Embrapa (1997): sand, clay, silt, active acidity (pH) and potential acidity (H+Al), organic matter (OM), aluminum (Al), calcium (Ca), magnesium (Mg), potassium (K), phosphorus (P), bases saturation (V%), sum of bases (SB) and cation exchange capacity (CEC). Ca, Mg, pH and H+Al was chosen for their associated behavior under handling practices, like liming or from original soil acidity, like parental material.

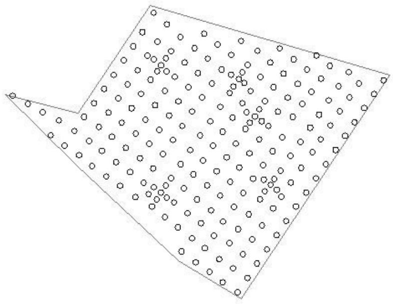

Fig. 1. Sampling scheme on precision farming field at Araguari (MG) – Brazil.

Data were standardized in order to have a fare variance range between the chosen variables. FKA was performed on ISATIS software (Geovariances 2003). Direct and cross variograms were adjusted mixing automatic and interactive model fit available on ISATIS, as well filtering procedures on linear coregionalization model after estimation by cokriging. pH variable was chosen to discuss filtering procedure taken into account their relation with Aluminum availability on soil. For the authors this consists on an important layer on agro-ecological GIS projects. Operations with vector and raster results and map visualizations were performed on SPRING GIS (Câmara et al. 1996).

RESULTS AND DISCUSSION

CORREGIONALIZATION ANALYSIS

Classical statistics and exploratory data analysis of all measure variables on soil sample and chemical and physical analysis are present in Manzione (2002). Under scope of LMC and supposing second order stationary, direct and cross variograms were modeled with a nugget effect and a spherical model at 451,28 m range, analyzing coregionalization structure range among soil acidity variables. A first visual inspection of the variograms and cross-variograms suggested the presence of three basic components at different spatial scales. The first observed structure was pure nugget effect due to measurement errors and micro-variation within the smallest sampling interval 60 m. The second structure reflected a transitive process of plot-size range, approximately of 100 m. The third structure seemed unbounded at the scale of study; it could be represented either by an unbounded variogram or by a transitive model with a longer range, approximately from 400 m. In this case PCA analysis reveal that this structure didn't have great contribution to coregionalization model, and only nugget effect and a spherical model at 451,28 m remains at the model. At Table 1 direct (diagonal) and cross variograns parameters expressed as nugget and sill values of modeled structure here proposed. These values correspond to the b^{u}_{ij} matrix for each basic structure. Adjusted direct and cross variograms are showed on Figures 2 and 3, respectively.

Table 1. Variance-Covariance matrix for each spatial scale.

Nugget effect	Ca	H+Al	Mg	pH
Ca	0,5908	-	-	-
H+Al	-0,3164	0,3610	-	-
Mg	0,4963	-0,2929	0,5119	-
pH	0,5096	-0,4092	0,4489	0,5421
Spherical model	Ca	H+Al	Mg	pH
Ca	0,4820	-	-	-
H+Al	-0,3391	0,7013	-	-
Mg	0,3443	-0,3642	0,4544	-
pH	0,3955	-0,3632	0,3549	0,4890

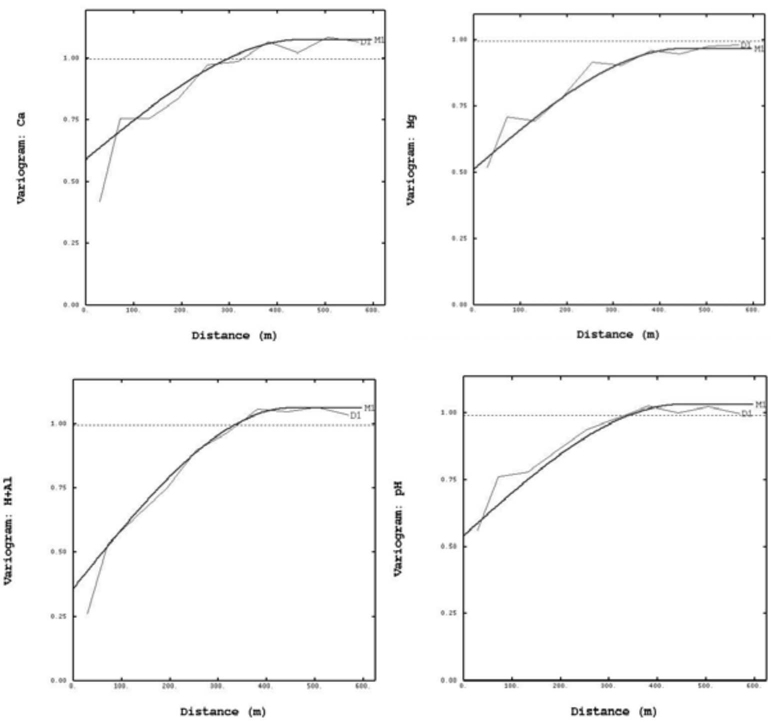

Fig. 2. Direct variograms adjusted for Ca, Mg, ph, H+Al.

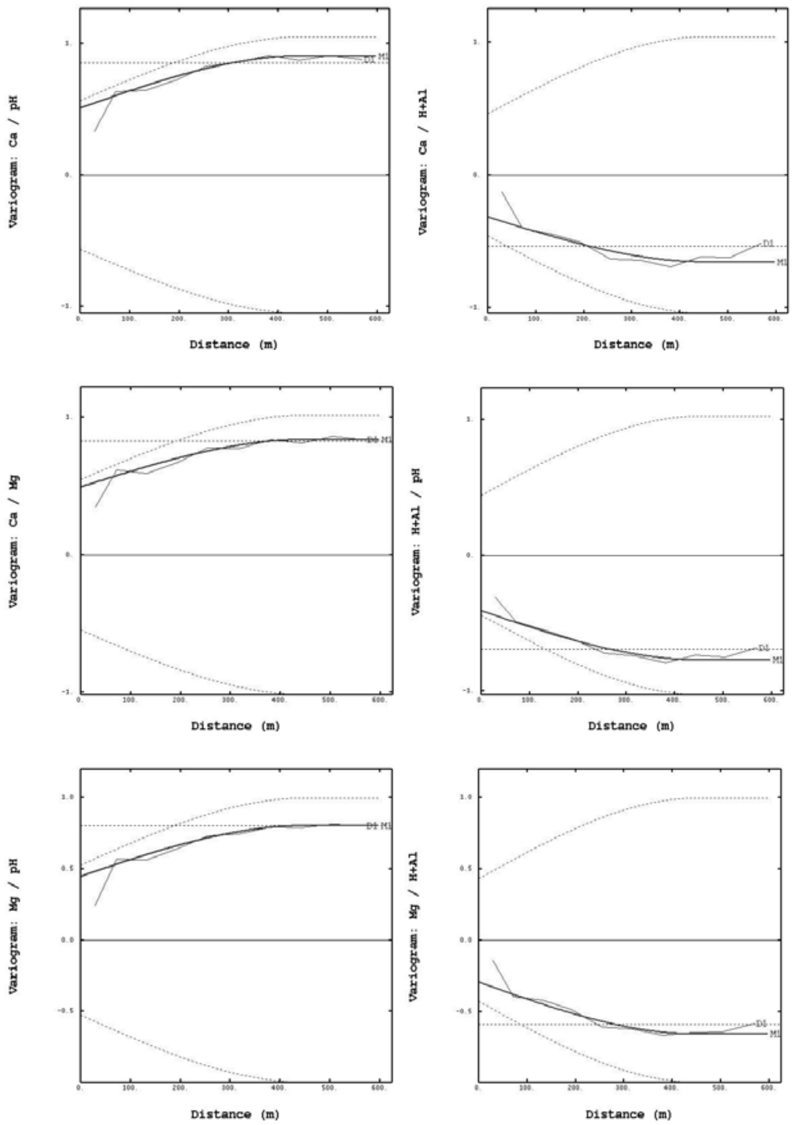

Fig. 3. Cross variograms adjusted for bivariate relationships between Ca, Mg, ph, H+Al.

The product-moment correlation coefficient did not reveal the real relationships among the variables, since it averages out distinct changes in the correlation structures occurring at different spatial scales (Castrignanò et al. 2000). The values of b^u_{ij} were used to calculate the structural correlation coefficients for each basic structure adjusted (Goovaerts 1992, Borůvka and Kozák 2001), and are presented at Table 2. Most of the calculated structural correlation coefficients were higher than correlation coefficients at Table 1, which means that a large part of the relationship between variables was spatial-based. Given by example pH and H+Al structural correlation coefficients, they were -0,925 at first structure and -0,6202 at second structure. This decrease of correlation could be explained by both variables have a similar covariance function at nugget effect and at spherical model H+Al had a stronger covariance function than pH.

Table 2. Structural correlation matrices for each spatial scale.

Nugget effect	Ca	H+Al	Mg	pH
Ca	1	-	-	-
H+Al	-0,6851	1	-	-
Mg	0,9025	-0,6814	1	-
pH	0,9004	-0,9250	0,8522	1
Spherical model	Ca	H+Al	Mg	pH
Ca	1	-	-	-
H+Al	-0,5832	1	-	-
Mg	0,7357	-0,6452	1	-
pH	0,8146	-0,6202	0,7529	1

REGIONALIZED FACTORS AND VARIABLES

Geostatiscal analysis combined with PCA analysis allowed discrimination of soil sources of variability at each scale modeled. The factors resulting from PCA analysis represents a synthesis of the information provided by the individual variables. First two eigenvectors and eigenvalues are at Table 3.

Table 3. First two eigen values and eigen vectors given by PCA analyzes.

		Eigen vector 1	Eigen vector 2
Nugget effect	Ca	0,5511	-0,4216
	H+Al	-0,3866	-0,7254
	Mg	0,5023	-0,4406
	pH	0,5426	0,3192
	Eigen values	1,7668 (88,08%)	0,1793 (8,94%)
Spherical model 451,28 m	Ca	0,4788	-0,4241
	H+Al	-0,5571	-0,8176
	Mg	0,4661	-0,1716
	pH	0,4931	-0,3496
	Eigen values	1,6189 (76,12%)	0,2936 (13,81%)

This analysis captured acidity phenomena structure in micro scale (nugget effect) with the influence divided between Ca and Mg elements and pH of this soil, given by an intense agronomical handling which soil was submitted over the years. The first two factors at short range explained more than 97% of the overall variation at this spatial scale. On a long range the phenomenon was explained by H+Al, given by the geological soil influence (pedological formation). The first two factors of the spherical model structure explained more than 90% of the overall variation at this spatial scale. The eigenvalues of the structures modeled given by PCA confirm that nugget effect is an important part of the model, and that the sample scheme didn't captures small variations presented on this soil.

The spatial interrelations among the variables, as described above by coregionalization matrices for each structure, can be clearly displayed in the plots of correlations corresponding to different spatial scales (Goovaerts 1992, Castrignanò et al. 2000). At nugget effect structure, pH and H+Al had a strong correlation, which was no so important at the second structure, where pH had strong correlations with Ca and Mg. The spatial continuity of H+Al had a strong influence at second structure, and made pH establish a cluster with Ca and Mg values. Figure 4 presents the pair of coordinates of each variable, determined by the pair of correlation coefficients between your spatial component and the first two regionalized factors.

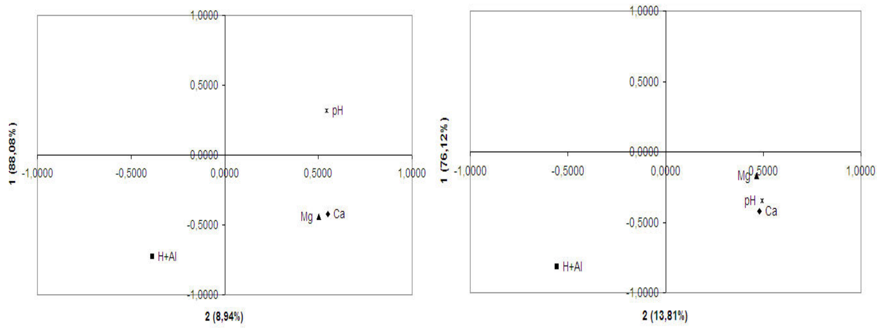

Fig. 4. Principal components results for nugget effect and spherical model at 451,28 m.

FILTERING SPATIAL SCALES

The model proposed estimated by cokriging was filtered, once interested on long range variability. This procedure allow visualize only interest features, removing discontinues individually, enhancing long range variability of the phenomenon, and showing the drift contained on the model. pH results of filtering the nugget effect structure, the spherical model structure and the drift part of cokriging maps are present on Figure 5.

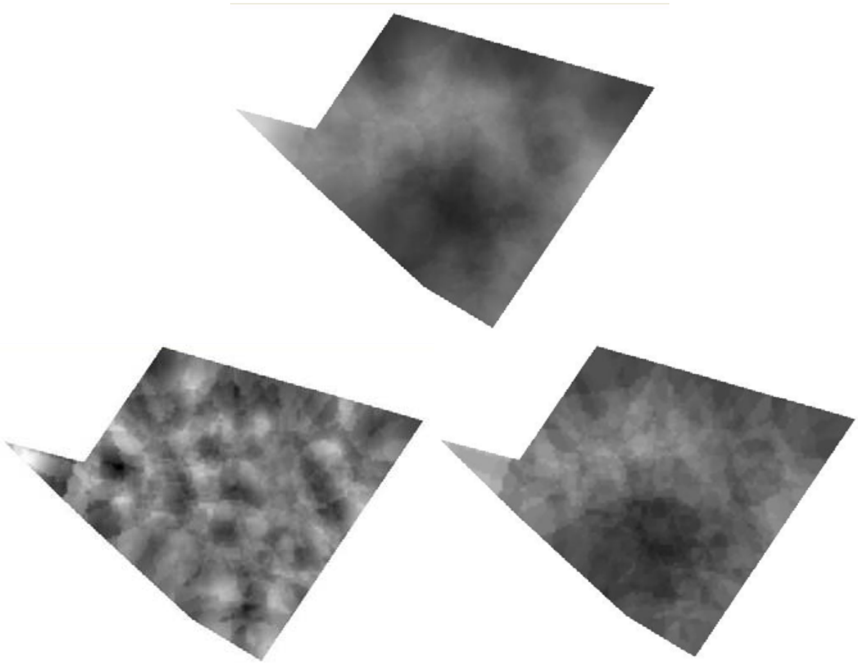

Fig. 5. pH maps estimated by cokriging with nugget effect, drift and spherical model filtered, respectively.

Sources of intense handling affect intensely the variation on topsoil (0-20 cm). The soil over the studied area has been subjected for intense soil operations over the years. Practices like tillage and plough have a nominal range at soil relative of used implement (plow, harrow, subsoilers). It causes topsoil desegregation, inversion and pulverization of the soil sod and variations which a grid of 60 square meters didn't capture, resulting in a nugget effect which has an important influence in the model. The filtering procedure treated this problem removing high frequencies signaled by nugget effect, and resulting in a more continuous map of the acidity phenomenon on this soil. This procedure could treat problems at sample scheme, removing the nugget effect, and enhancing spatial continuity of the acidity phenomena. Without the drift part, the presence of local variation is clearly, confirming the choice to remove the nugget and consequently give a global mean effect on the final map.

CONCLUSIONS

Multivariate geostatistical analysis has allowed separating the different sources of spatial variation at different scales. This method has offered ways for formulating hypotheses about the probable sources of soil variation, resulting in a better understanding of the variations within the study area would allow more sophisticated land management to be developed. FKA approach has been demonstrated to be a highly effective method to separate the area into homogenous zones in order to optimize their management. Spatial distribution of soil properties in this context is influenced by many factors and the geostatistical approach is a useful tool to understand the dynamics of addition/removal of soil attributes. Filtering procedure enhance spatial scales, separating their effects on final map results and removing noise at high frequencies. That is a possible way to treat problems like discontinuities resulted by sample schemes or nugget effects. These techniques have great potential to explore phenomena over chrono and toposequences, landscapes, watersheds and drainage basins.

REFERENCES

Borůvka L, Kozák J (2001) Geostatistical investigation of a reclaimed dumpsite soil with emphasis on aluminum. Soil & Tillage Research 59, pp 115-126

Câmara G, Souza RCM, Freitas UM, Garrido J (1996) Spring: Integrating remote sensing and GIS by object-oriented data modelling. Computers & Graphics 20, pp 395-403

Castrignanò A, Giugliarini L, Risaliti R, Martinelli N (2000) Study of spatial relationships among some soil physico-chemical properties of a field in central Italy using multivariate geoestatistics. Geoderma 97, pp 39-60

Embrapa (1997) Manual de métodos de análise de solo. 2 edn. Centro Nacional de Pesquisa de Soja, Rio de Janeiro

Geovariances (2003) ISATIS Software Manual. Geovariances, Avon

Goovaerts P (1992) Factorial kriging analysis: a useful tool for exploring the structure of multivariate spatial soil information. Journal of Soil Science 43, pp 597-619

Goovaerts P (1997) Geostatistics for natural resources evaluation. Oxford University Press New York

Goulard M, Voltz M (1992) Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix. Mathematical Geology 24, pp 269-286

Journel AG, Huijbregts CJ (1978) Mining geostatistics. Academic Press London

Lark RM, Papritz A (2003) Fitting a linear model of coregionalization for soil properties using simulated annealing. Geoderma 115, pp 245-260

Manzione RL (2002) Variabilidade especial de atributos químicos do solo em Araguari-MG. Dissertação (Mestrado em Agronomia/Energia na Agricultura) Universidade Estadual Paulista, Botucatu

Matheron G (1982) Pour une analyse krigeante des donnees regionalisees. Report Nº732. Centre de Geostatistique, Fontainebleau

Olea RA (1991) Geostatistical glossary and multilingual dictionary. Oxford University Press New York

Wackernagel H (1995) Multivariate geoestatistics – An introduction with applications. Springer-Velag Berlim

Wackernagel H, Petitgas P, Touffait Y (1989) Overview of methods for corregionalization analysis. In: Armstrong M (ed) Geostatistics, vol 1. Kluver Academic Publishers, Dordrecht, pp 409-420