

USING MULTIVARIATE GEOSTATISTICS FOR DESCRIBING SPATIAL RELATIONSHIPS AMONG SOME SOIL PROPERTIES

Castrignanò A.¹, Cherubini C.², Giasi C.I.², Castore M.², Di Mucci G.², Molinari M.³

¹CRA-Istituto Agronomico Sperimentale, via Celso Ulpiani 5, 70125 Bari, Italy

²Politecnico di Bari, Ingegneria Civile ed Ambientale, via Orabona 4, 70125 Bari, Italy

³Eni S.p.a., Refining and Marketing, via Laurentina 449, 00142 Roma, Italy

1. Introduction

The soils represent critical environment at the interface among rock, air and water; they can then be centre of various chemicals, deriving from several human activities (industry, agriculture, transports, etc), but, at the same time, be the living pollution sources for superficial and deep waters, for organisms, sediments and oceans. Optimum benefits on profitability and environment protection depend on how well land use and remediation practices are fitted to variable soil conditions.

The crucial problem is to characterize soil with precision, both quantitatively and spatially (Castrignanò et al., 2000), because soil variability is the result of both natural processes and management practices, acting at different spatial and temporal scales. For this reason it is necessary to use adequate techniques of analysis, capable to put in evidence important spatial relationships and identify those factors that control the variability of geochemical data.

Multivariate geostatistics uses information coming from the relationships among variables in order to improve estimation precision and to disclose the different causes of variation working at different spatial scales. Some of the several factors that govern soil variations are likely to 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.

The main objective of this paper is to study the scale-dependent correlation structure of some soil variables, supposing that it can reflect the different sources of variability. This requires a particular statistical approach that combines classical Principal Component Analysis, to describe the correlation structure of multivariate data sets, with geostatistics, to take into account the regionalized nature of the variables. We then applied a method, called Factorial Kriging Analysis (FKA) and originally developed by Matheron (1982), to study the correlations among some soil properties at the different spatial scales.

2. Material and Methods

2.1. Multivariate geostatistical approach

Multivariate spatial data set can be analyzed by FKA, a relatively recent geostatistical method developed by Matheron (1982). The theory underlying FKA has been described in several publications (Goovaerts and Webster, 1994; Castrignanò et al., 2000b; Wackernagel, 2003); here we will describe only the most salient points. The approach consists of decomposing the set of original second-order random stationary variables $\{Z_i(\mathbf{x}); i = 1, \dots, n\}$ into a set of reciprocally orthogonal regionalised factors $\{Y_v^u(\mathbf{x}); v = 1, \dots, n; u = 1, \dots, N_S\}$ where N_S is the

number of spatial scales, through transformation coefficients a_{iv}^u , combining the spatial with the multivariate decomposition:

$$Z_i(\mathbf{x}) = \sum_{u=1}^{N_S} \sum_{v=1}^n a_{iv}^u Y_v^u(\mathbf{x})$$

The three basic steps of FKA are the following:

- 1) modelling the coregionalization of a set of variables, using the so called Linear Model of Coregionalization (LMC);
- 2) analysing the correlation structure between the variables, at the different spatial scales by applying Principal Component Analysis (PCA);
- 3) cokriging specific factors at characteristic scales and mapping them.

2.1.1. Linear Model of Coregionalization

The LMC, developed by Journel e Huijbregts (1978), assumes all the studied variables are the result of the same independent processes, acting at different spatial scales u . The $n(n+1)/2$ simple and cross semivariograms of the n variables are modelled by a linear combination of N_S standardized semivariograms to unit sill $g^u(\mathbf{h})$. Using the matrix notation, the LMC can be written as:

$$\Gamma(\mathbf{h}) = \sum_{u=1}^{N_S} \mathbf{B}^u \mathbf{g}^u(\mathbf{h})$$

where $\Gamma(\mathbf{h}) = [\gamma_{ij}(\mathbf{h})]$ is a symmetric matrix of order $n \times n$, whose diagonal and non diagonal elements represent simple and cross semivariograms, respectively for lag h ; $\mathbf{B}^u = [b_{ij}^u]$ is called coregionalization matrix and it is a symmetric positive semi-definite matrix of order $n \times n$ with real elements b_{ij}^u at a specific spatial scale u . The model is authorized if the mathematical functions $g^u(\mathbf{h})$ are authorized semivariogram models.

In the linear model of coregionalization the spatial behaviour of the variables is supposed resulting from superimposition of different independent processes working at different spatial scales. These processes may affect the behaviour of experimental semivariograms, which can then be modelled by a set of functions $g^u(\mathbf{h})$. The choice of number and characteristics (model, sill, range) of the functions $g^u(\mathbf{h})$ is quite delicate and can be made easier by a good experience of the studied phenomena (Chiles and Guillen, 1984). Fitting of LMC is performed by weighed least-squares approximation under the constraint of positive semi-definiteness of the \mathbf{B}^u , using the iterative procedures developed by Goulard (1989). 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 sum of squares.

2.1.2. Regionalized Principal Component Analysis

Regionalized Principal Component Analysis consists of decomposing each coregionalization matrix \mathbf{B}^u into two other diagonal matrices: the matrix of eigenvectors and the diagonal matrix of eigenvalues for each spatial scale u through the matrix \mathbf{A}^u of order $n \times n$ of the transformation coefficients a_{iv}^u (Wackernagel, 2003). The transformation coefficients a_{iv}^u in the matrix \mathbf{A}^u correspond to the covariances between the original variables $Z_i(x)$ and the regionalized factors $Y_v^u(\mathbf{x})$.

2.1.3. Mapping multivariate spatial information

The behaviour and relationships among variables at different spatial scales can be displayed by interpolating the regionalized factors $Y_v''(\mathbf{x})$ using cokriging and mapping them (Castrignanò et al., 2000a). The cokriging system in FKA has been widely described by Wackernagel (2003).

2.2. Sampling and measurements

In the year 2003, the spatial variability of some soil properties was studied in an industrial area of 300 ha located in Taranto (Apulia Region, southern Italy). A monitoring net composed by 184 boreholes was placed on the site and in each point a sample was collected at 1 m depth. These samples were analyzed in laboratory in order to evaluate different soil properties. The final database consisted of 184 samples and 16 variables, which were the following: Be, Cd, Va, Zn, total Cr, Hg, Ni, Pb, Cu, Cation Exchange Capacity (CEC), Organic Carbon, Fraction (%) of soil particle size from 2 mm to 20 mm, Fraction < 2 mm, light hydrocarbons, humidity at 105°C, pH. It is important to underline that the soil matrix did not result to be contaminated because no sample value overcame the critical threshold imposed for each compound by the Italian decree D.M.471/99.

3. Results and discussion

3.1. Exploratory analysis

First of all we determined the descriptive statistics of all variables, as reported in Table 1.

Table 1. Descriptive statistics of the soil properties

Variable	Mean	Min	Max	Standard Deviation	Skewness	Kurtosis
Be (mg Kg ⁻¹)	1.14	0.05	9.72	1.39	2.74	13.10
Cd (mg Kg ⁻¹)	0.06	0.00*	0.51	0.05	3.58	26.61
CEC	24.2	8.90	45.60	7.83	0.29	2.39
	9					
Organic Carbon	0.25	0.01	1.24	0.21	1.72	6.64
total Cr (mg Kg ⁻¹)	7.99	0.8	46.10	8.28	2.73	11.23
Fraction 2mm - 20mm (%)	21.3	0.00*	69.50	17.35	0.71	2.89
	0					
Fraction < 2mm (%)	74.8	0.00*	100	22.07	-1.37	5.27
	9					
light hydrocarbons (mg Kg ⁻¹)	4.69	0.10	308	33.25	8.22	70.94
Hg (mg Kg ⁻¹)	0.02	0.00*	0.44	0.04	6.70	60.07
Ni (mg Kg ⁻¹)	7.77	0.70	67.50	10.75	3.15	14.05
Pb (mg Kg ⁻¹)	6.10	0.22	34.40	5.62	2.33	9.38
Cu (mg Kg ⁻¹)	7.17	0.60	138	12.33	7.17	70.87
Va (mg Kg ⁻¹)	11.4	1.10	38.10	6.00	1.21	5.08
	9					
Zn (mg Kg ⁻¹)	16.2	1.60	154	17.61	3.78	25.76
	6					
pH	8.28	7.25	11.55	0.52	2.25	13.60

*detection limit

From the inspection of the table, we can notice high shifts of skewness and kurtosis from 0 and 3, respectively, which are the characteristic values of normal distribution. Therefore, the variables generally exhibit non symmetric distributions, with long tails and several outliers. The variables were then normalised and standardised to 0 mean and unit variance.

The visual inspection of the variogram maps, (not shown) did not reveal any significant anisotropy in chemicals distribution, therefore an isotropic model of variogram was assumed.

3.2. Coregionalization analysis

Before performing a coregionalization analysis to separate the different sources of variation, we decided to select a smaller number of the most relevant variables, in order to save computer time and make easier the interpretation of the results. As to the selection, the following five variables total Cr, Ni, Pb, Va and organic carbon were analysed, because they appeared to be the more spatially structured and correlated variables. A LMC was fitted to the set of the 15 direct and cross-variograms including 3 basic spatial structures: 1) a nugget effect; 2) a spherical model with range=249.58 m ; 3) an exponential model with range=1300.00 m. Most of direct and cross-variograms appeared well spatially structured and for some pairs of variables (Ni-Cr, Va-Cr) the spatial cross-correlation was very strong, close to the maximum corresponding to intrinsic correlation. The inspection of fig.1 shows also that the goodness of fitting was generally quite high with the exception of some mismatch at the origin, quite probably because of the presence of outliers. The goodness of fitting was also tested by cross-validation, calculating mean error and reduced variance (variance of standardised error), which were close to 0 and 1, respectively (not reported). These results mean that the estimates were unbiased and the estimation variance reproduced experimental variance accurately. The cokriging maps of the estimated values of organic carbon, total chromium, nickel, lead and vanadium contents are reported in fig 2 (a,b,c,d,e). For any estimated value, the cokriging has allowed also to calculate the variance of the estimation error associated to it, giving a measure of the reliability of the estimation (not shown). Once LMC was estimated, cokriging was applied to the transformed data to obtain the estimates which were then back-transformed to express them in the original variable.

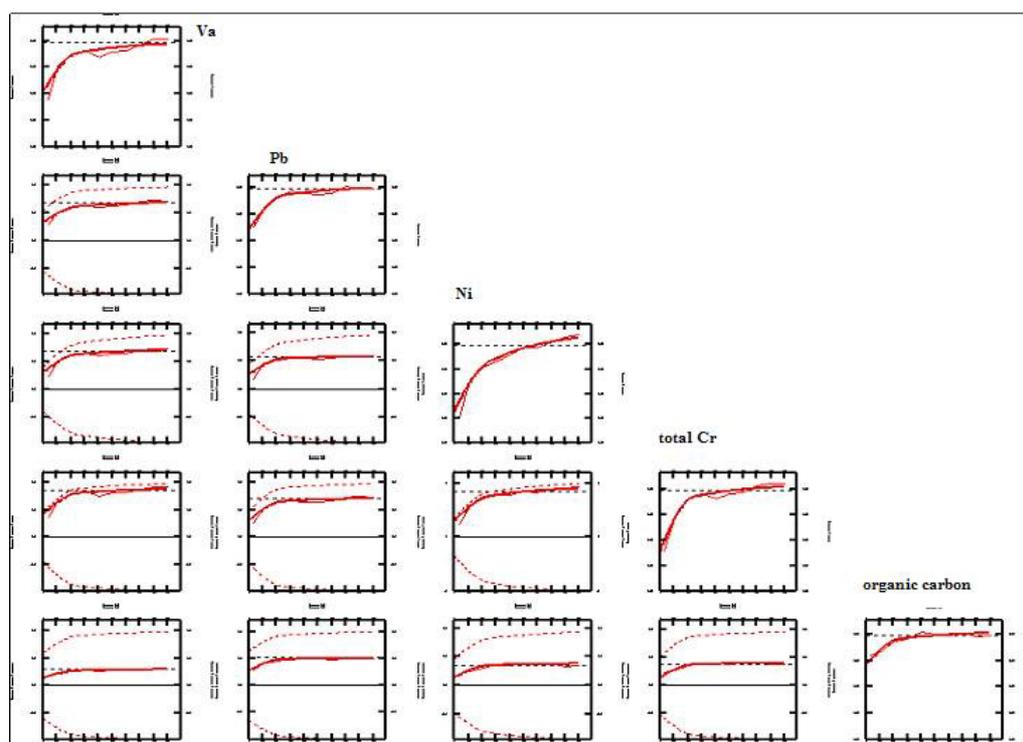


Figure 1. Experimental direct and cross semivariograms (fine line) with the fitted linear coregionalization model (bold line); the dashed line in cross-variogram represents the maximum correlation between the two variables.

The obtained maps put into relevance how total chromium and nickel on one hand and vanadium and lead on the other show similar spatial distributions, whereas organic carbon looks quite different from all the other variables. At this point, we wanted to inquire more deeply into the different sources of variation working in the study area.

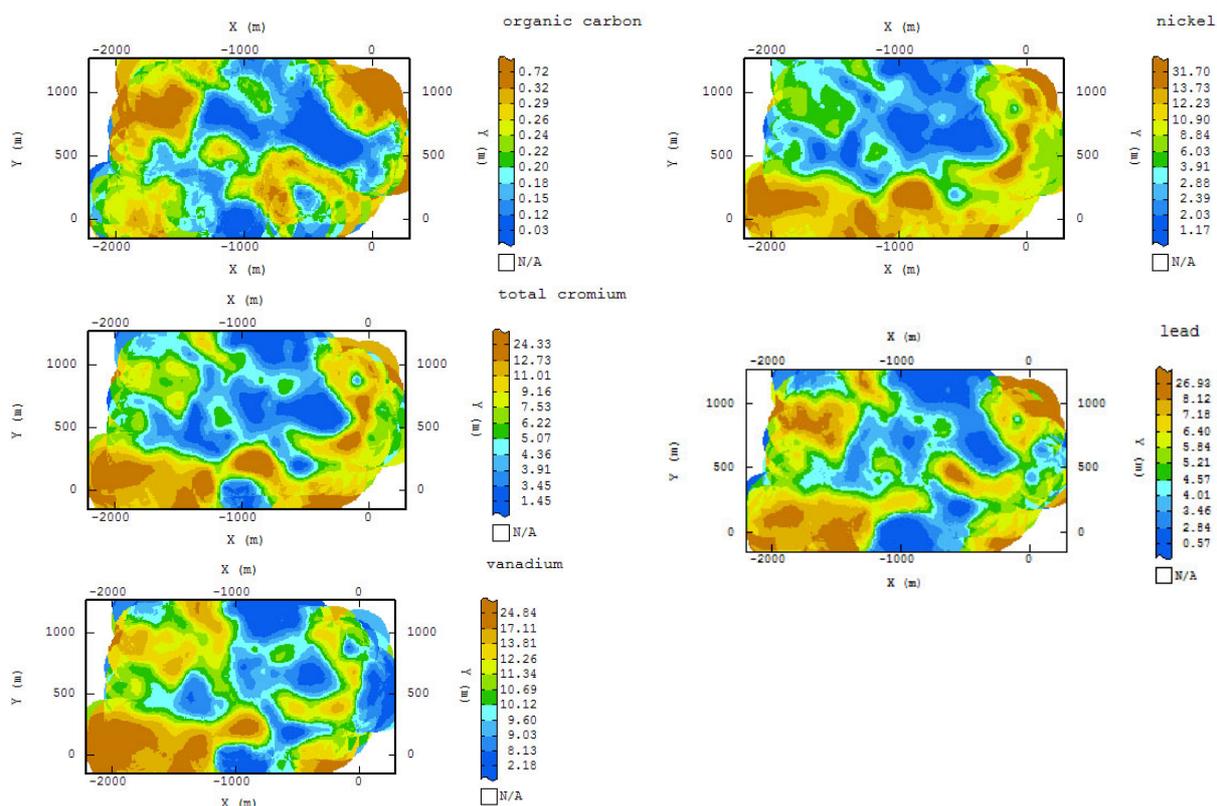


Fig. 2 : Cokriging maps of OC, Va, total Cr, Ni and Pb

The Factorial Kriging

FKA has allowed to isolate the first 2 regionalized factors that, at the cost of an acceptable loss of information, have given a synthetic description of the process in study at the different selected spatial scales. Passing over the nugget effect, because mostly affected by measurement error, we will concentrate on the short-range and long-range components of the first two regionalized factors.

S1 : Nugget effect

Coregionalization matrix :

	Carbon	Cr	Ni	Pb	Va
Carbon	0.6968	0.1378	0.1396	0.2679	0.1378
Cr	0.1378	0.3919	0.2922	0.2970	0.4083
Ni	0.1396	0.2922	0.3188	0.2506	0.2939
Pb	0.2679	0.2970	0.2506	0.5980	0.3268
Va	0.1378	0.4083	0.2939	0.3268	0.5147

Eigen vectors matrix:

	Carbon	Cr	Ni	Pb	Va	Eigen Val.	Var. Perc.
Factor 1	0.3778	0.4488	0.3744	0.5173	0.4981	1.5472	61.39
Factor 2	0.8806	-0.2750	-0.1706	0.0488	-0.3427	0.5880	23.33
Factor 3	0.2804	0.2539	0.1988	-0.8542	0.2963	0.2501	9.92
Factor 4	-0.0561	0.0000	0.8272	-0.0197	-0.5587	0.1049	4.16
Factor 5	-0.0017	-0.8115	0.3270	0.0023	0.4842	0.0300	1.19

S2 : Spherical - Range = 249.58m

Coregionalization matrix :

	Carbon	Cr	Ni	Pb	Va
Carbonio	0.2041	0.2119	0.1699	0.2024	0.1037
Cr	0.2119	0.4340	0.3258	0.3054	0.3181
Ni	0.1699	0.3258	0.2556	0.2478	0.2476
Pb	0.2024	0.3054	0.2478	0.2597	0.2151
Va	0.1037	0.3181	0.2476	0.2151	0.2805

Eigen vectors matrix:

	Carbon	Cr	Ni	Pb	Va	Eigen Val.	Var. Perc.
Factor 1	0.3105	0.5755	0.4479	0.4370	0.4252	1.2688	88.49
Factor 2	0.7476	-0.1462	-0.0580	0.2844	-0.5793	0.1461	10.19
Factor 3	0.0806	0.7487	-0.1752	-0.5401	-0.3325	0.0189	1.32
Factor 4	-0.0733	-0.1630	0.8743	-0.3330	-0.3045	0.0000	0.00
Factor 5	-0.5769	0.2457	0.0303	0.5705	-0.5295	0.0000	0.00

S3 : Exponential - Scale = 1300 m

Coregionalization matrix :

	Carbon	Cr	Ni	Pb	Va
Carbon	0.1267	0.0650	0.0804	0.0348	0.0574
Cr	0.0650	0.2254	0.3278	0.1192	0.1717
Ni	0.0804	0.3278	0.5623	0.0879	0.1747
Pb	0.0348	0.1192	0.0879	0.1538	0.1695
Va	0.0574	0.1717	0.1747	0.1695	0.1996

Eigen vectors matrix:

	Carbon	Cr	Ni	Pb	Va	Eigen Val.	Var. Perc.
Factor 1	0.1525	0.4944	0.7316	0.2496	0.3670	0.9182	72.42
Factor 2	0.1092	0.0347	-0.5379	0.6223	0.5570	0.2423	19.11
Factor 3	-0.9796	0.0491	0.0748	0.1591	0.0836	0.1067	8.42
Factor 4	0.0430	-0.8609	0.4089	0.2589	0.1508	0.0006	0.05
Factor 5	-0.0572	-0.1036	-0.0508	-0.6769	0.7248	0.0000	0.00

Table 2. Linear Model of Coregionalization with reported for each spatial scale (S): the coregionalization matrix, the eigen vector matrix, the corresponding eigen values and the percentage of variance explained by them.

In table 2 for each spatial scale are reported:

- 1) the variance-covariance (coregionalization) matrix;
- 2) the eigen vector matrix;
- 3) the eigen values which represent the variances of the corresponding eigenvectors;
- 4) the percentage of variance explained by each eigen vector.

The first two factors explain most variance both at short and long range (98.68%, and 91.51%, respectively).The short-range component of the first factor (F1) explains 88.49% and is mostly correlated with chromium (0.5755) and in smaller measure with the other variables, whereas the long-range component of F1 explains the 72.42% of the variance and is mainly correlated to nickel (0.7316) and in smaller measure to chromium (0.4944).

The second factor F2 at short range explains the 10.19% of variance and is strongly correlated with organic carbon (0.7476) whereas at long range explains the 19.11% of the variance and is positively correlated with lead (0.6223) and vanadium (0.5570) and less with the others (negatively with nickel).

The above results lead to think that organic carbon is more linked to intrinsic factors of the soil; it doesn't influence the mobility and the distribution of the examined inorganic chemicals, with whom it doesn't look to be much spatially correlated

The distributions of the two factors both at short and at long range looks as “pepper and salt” type with a high component of erraticity. This puts into evidence that the points of emission for the examined inorganic chemicals *are not concentrated*: more precisely we could assert that in the study case we are in presence of *more than one points of emission*, jointly working and being ascribed to causes of anthropic origin. The analysis of the cycle of production and of the activities carried out in the areas, disclosed by the maps of the regionalized factors, will allow to individuate the cause or the causes of the presence of chemicals. The probable sources of Cr variation operate at both short and long distances, whereas Ni acts rather at longer ranges, like lead and vanadium. However, lead and vanadium perform differently from Ni but quite similarly between them.

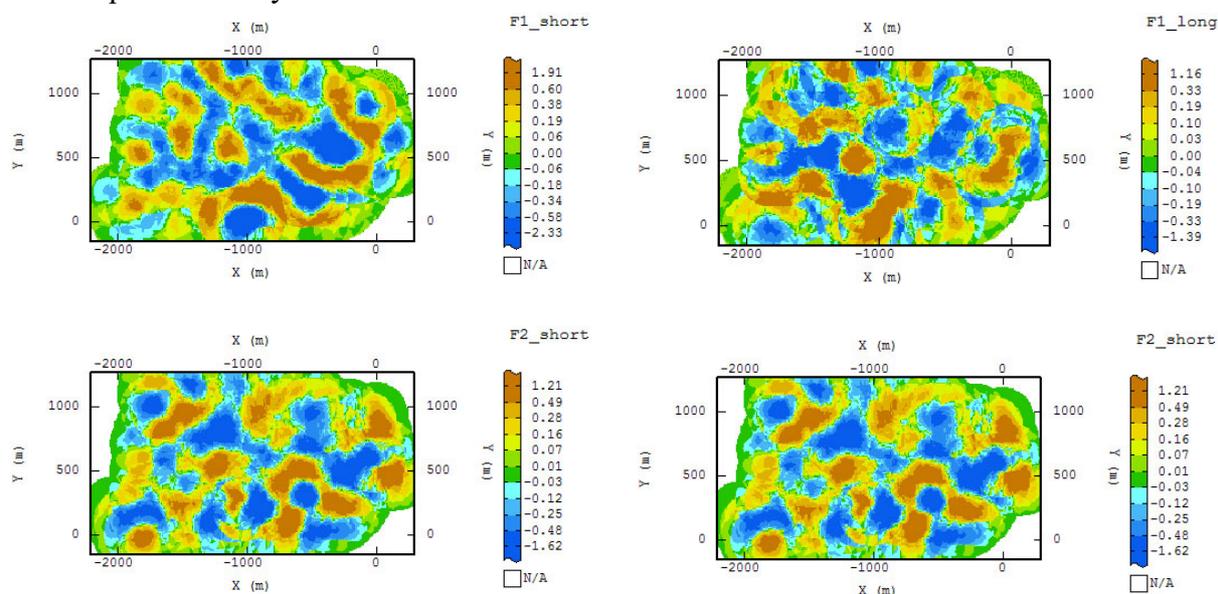


Fig 3. Cokriged maps of the short-range (a) and long-range (b) spatial components of the first two principal components (F1,F2).

4. Conclusions

Spatial variability of some soil components, measured in an industrial area of southern Italy, is the result of superimposed processes acting at different spatial scales.

The study shows that the points of emission for the examined inorganic chemicals *are not concentrated*: more precisely we could assert that in the study case we are in *presence of more points of emission evenly distributed*. The lacking of large structures of spatial dependence means quite probably that the origin of soil variability may mostly be ascribed to human activities. It needs to emphasize however that the high value of the nugget shown by the experimental variograms of Va and Pb suggests to refine the mesh of investigation.

Intensifying the sampling might so allow local variation to be adequately taken into account in designing monitoring net and in planning land recovery.

References

- Castrignanò, A., Giugliarini, L., Risaliti, R., Martinelli, N., 2000a. Study of spatial relationships among some soil physico-chemical properties of a field in central Italy using multivariate geostatistics, *Geoderma* 97, 39-60.
- Castrignanò, A., Goovaerts, P., Lulli, L. Bragato, G. 2000b A geostatistical approach to estimate probability of *Tuber melanosporum* in relation to some soil properties. *Geoderma* 98 (2000) 95-113
- Chiles, J.P., Guillen, A., 1984. Variogrammes et krigeages pour gravimétrie et le magnétisme. *Sciences de la Terre, Série Informatique* 20, pp. 455-468
- Goovaerts, P., Webster, R., 1994. Scale dependent correlation between topsoil copper and cobalt concentrations in scotland. *Eur. J. Soil Sci.* 45, 79-95
- Goulard, M., Voltz, M., 1992. Linear coregionalization model : tools for estimation and choice of cross-variogram matrix. *Math. Geol.* 24 (3), 269-286
- Journel, A.G., Huijbregts, C.J., 1978. *Mining Geostatistics*. Academic Press, London, 600 pp.
- Matheron, G., 1982. Pour une analyse krigeante des données régionalisées in Report 732. Centre de Geostatistique, Fontainebleau
- Wackernagel, H. *Multivariate Geostatistics An Introduction with Applications*, 2003. Third Edition Springer Verlag