SPATIAL PREDICTION OF TWO SOIL PROPERTIES USING TOPOGRAPHIC INFORMATION

D. P. KALIVAS*
D. P. TRIANTAKONSTANTIS
V. J. KOLLIAS

Agricultural University of Athens
Laboratory of soils and Agricultural Chemistry
75 Iera Odos, Botanikos, 11855 Athens, Greece

* to whom all correspondence should be addressed
tel: +30 - 210 - 5294095; fax: +30 - 210 - 5294092
e-mail: kalivas@aua.gr

ABSTRACT

The objective of this study was to determine whether the use of the co-regionalization of the distance-to-river topographic variable with the soil properties topsoil clay and sand can improve their mapping. The interpolation techniques: ordinary kriging, kriging combined with regression (two models) and heterotopic co-kriging were applied to data from 153 observation points. The two models of kriging combined with regression involve: (a) linear regression of the two soil variables with the distance-to-river variable on the 153 observation points followed by kriging and (b) summation of the kriged regression values and kriged regression residuals. For co-kriging 350 additional observations for the distance-to-river-variable were employed. The distance-to-river data were easily obtained from the map of the area which was stored in a Geographical Information System (GIS). The performances of the methods were evaluated and compared using the cross-validation method. The mean error of prediction indicates reasonably small bias of prediction for the two soil variables by almost all the methods. The mean square error showed that heterotopic co-kriging produced better estimates of the soil variables than kriging but there was a clear advantage in using the first model of kriging combined with linear regression technique. The second model of kriging combined with regression does not show any particular advantage over the other methods.

INTRODUCTION

Spatial interpolation is a procedure for estimating the values of a variable at unsampled locations. The interpolation techniques commonly used in earth sciences include linear regression, ordinary kriging and co-kriging (Laslett and McBratney; 1990; Leenaers *et al.*, 1990; Voltz and Webster, 1990; Kollias *et al.*, 1999).

Regionalized variable theory, which has been applied with apparent success to soil attribute data provides a summary of soil variability in the form of a semi-variogram and a predictive technique, kriging, for unobserved values (Journel and Huijbregts, 1978). Kriging is a geostatistical technique for optimal estimation and has been applied widely to soil properties. The advantage of representing the spatial variation of soil attributes geostatistically over other interpolation methods is

that an unbiased estimate (with a minimum known variance) can be made at any location of the study area (Burgess and Webster, 1980).

All interpolation techniques, geostatistical or otherwise, require a fairly dense network of sampling sites from which data are collected. However the high cost of collecting soil attribute data at many locations across landscape has created a need for interpolation methods that are using readily available secondary information such as topography attributes in order to improve the estimation of soil properties. Topography is a dominant control on earth surface processes and influence soil chemical and physical properties.

Co-kriging closely linked to kriging (Matheron, 1973) allows the use of a second variable in predicting values of the variable of interest. Co-kriging can be applied in situations where the examined soil variable is under-sampled and is more costly or difficult to sample than a topographic attribute which is closely related to it. Using co-kriging the spatial information of the topographic attribute is transferred to the soil property thus improving the quality of its estimates. Co-kriging could conceivably in considerable reduction of costs in achieving a comparable degree of accuracy by using fewer expensive variables and more inexpensive co-variables. The interpolation of an undersampled variable can be improved by using an auxiliary variable.

Models of prediction involving kriging combined with regression, were first proposed by Delhomme (1978; 1979) in hydrosciences domain. Ahmed and DeMarsily (1987) applied kriging combined with linear regression methodology to improve the mapping of transmitivity using specific capacity data. Recent studies involving kriging combined with regression models have revealed the validity of predicting a soil property from easy-to-measure morphological properties (McKenzie *et al.*, 1991; Moore *et al.*, 1993; McKenzie and Ryan, 1999). According to Odeh *et al.* (1994; 1995) and Knotters *et al.* (1995) kriging combined with regression methods may improve prediction performance in comparison to regression or kriging done separately.

Soil properties such as pH, texture and drainage are closely related to landscape position (Kreznor *et al.*, 1989) and these relationships have very often been used for their prediction at unsampled locations. Odeh *et al.*, (1994; 1995) have compared

several geostatistical methods for the spatial prediction of soil properties from landform attributes. River formed landscapes are reasonably similar the world over in terms of both landforms and processes. There is an interaction between geomorphological processes operative in river landscapes, the materials involved and ultimately pedological processes. Near present or former river beds, natural levees occur consisting of sandy sediments. Behind the levees lower lying areas occur formed of heavier-textured and impermeable silts and clays. A relationship exists between texture patterns and factors such as elevation relief and distance-to-river (Gerrard, 1981).

Leenaers *et al.* (1990) have employed elevation data for efficient mapping of zinc concentration in the soil of the Geul floodplains in the Southern Netherlands.

The objectives of this study are to investigate the benefits from the use of distance-to-river easily available data when estimating by geostatistical methods topsoil clay and sand soil properties on the floodplains of Acheloos river in Agrinio area of Western Greece. Also to assess whether heterotopic co-kriging or kriging combined with regression could make better use of the auxiliary variable.

MATERIALS AND METHODS
The soil data

An area of about 5700 ha (shown in Figure 1) which is located in the centre of Aitoloakarnania province in Western Greece was selected for this study. The average air temperature is 17.9°C, while the total annual precipitation is 1010 mm.

One data set with topsoil clay and sand content values and distance-to-river values was used for the purposes of this study. The data consisted from 153 observation points and was used for the mapping of soil properties. For co-kriging the sample points were supplemented with 350 additional points for which the distance-to-river value was estimated using a GIS system developed for the purposes of this study under the commercial GIS packages ARC/INFO, ARC VIEW. Together with the original sample points they formed a set of 503 points of which the distance-to-river values are known.

The descriptive statistics of the data set is given in Table 1 and include minimum, maximum, mean and standard deviation.
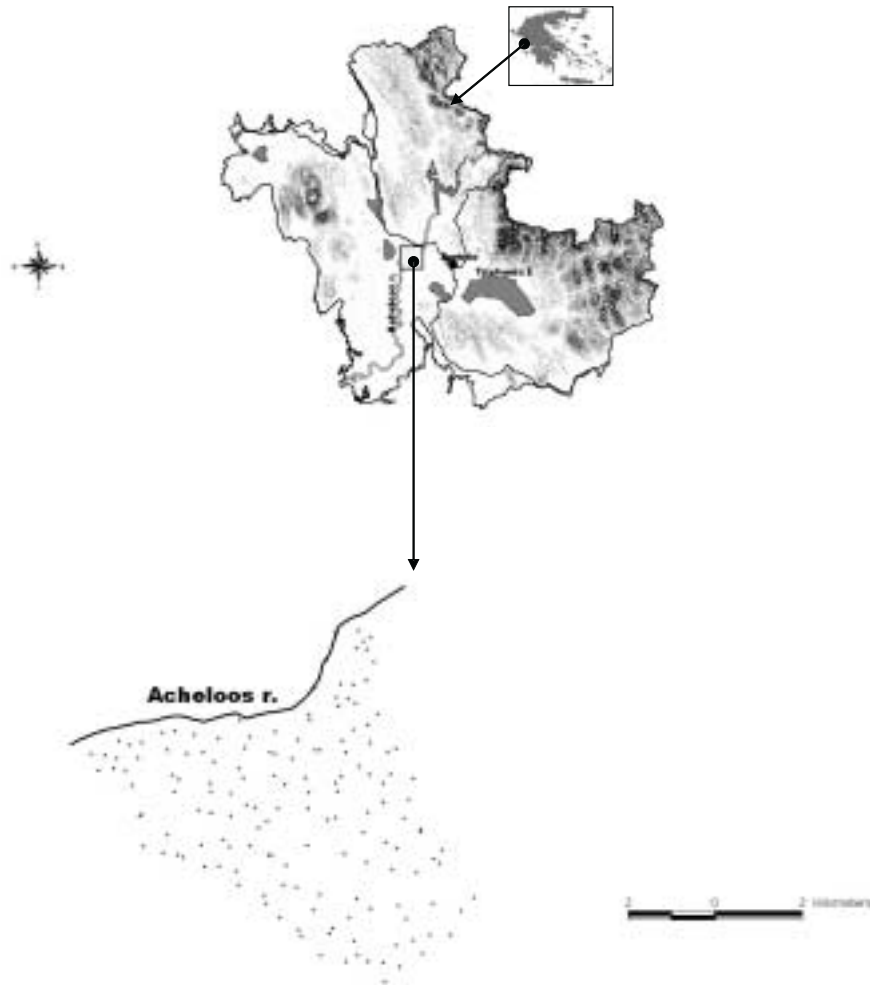
Figure 1.   The study area and the observation points

Table 1.   Statistics of the observation samples of the two soil and the distance-to-river variables

Soil variable	Minimum	Maximum	Mean	STD
Observation sample (153 points)				
%				
Clay 0 - 25 cm	16	47	32.8	7.71
Sand 0 - 25 cm	18	78	33.7	10.4
Observation sample (503 points)				
m				
Distance-to-river	40	6747	2803.2	1885.3

Because the distance-to-river data are very positively skewed they were normalized by computing logarithms to base e. The correlation coefficients of topsoil clay and sand against ln(distance-to-river) are 0.82 and 0.71 respectively.

Prediction methods

Since in-depth discussions about interpolation techniques are given by Journel and Huijbregts (1978), Isaaks and Srivastava (1989) and Burrough and McDonnell (1998) only an outline of the interpolation methods used will be given here.

The values of each soil variable at the observation points were used for the prediction of values at unknown points using the interpolation methods:

(a) ordinary kriging, (b) kriging combined with linear regression, (c) heterotopic co-kriging.

(i) Ordinary kriging (model ORD-KRG)
The spatial prediction of the values of a soil variable Z at an unsampled point x_0 is given by

$$\hat{Z}(\mathbf{x}_0) = \sum_{i=1}^{n} \lambda_i z(\mathbf{x}_i) \qquad (1)$$

where x denotes the set of spatial coordinates {x1, x2}, λ_i are the weights associated with the sampling points x_i and the ith observation point.

In kriging the weights are chosen so that the estimate $\hat{z}(x_0)$ of the true value $z(x_0)$ is unbiased and the prediction variance $\sigma^2(x_0)$ is minimized. That is:

$$E\left[\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)\right] = 0 \qquad (2)$$

and $\sigma^2(\mathbf{x}_0) = \text{var}\left[\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)\right] = \text{miminum}$ (3)

To ensure that the prediction is unbiased the weights placed on each neighbouring point must satisfy:

$$\sum_{i=1}^{n} \lambda_i = 1 \qquad (4)$$

(ii) Co-kriging (model CO-KRG)
Co-kriging is a geostatistical technique developed to improve the estimation of a variable using the information on other spatially correlated variables which are generally more densely sampled. The variables are called co-regionalized and are spatially dependent (McBratney and Webster, 1983).
We consider that U is an expensive to measure and therefore undersampled variable and V is a cheap to measure variable with more observations. If U and V are spatially and mutually correlated then it may be possible to use the spatial variation of V to help map of U (Burrough and MacDonnell, 1998). A co-kriged prediction is a weighted average in which the value of U at location x_0 is estimated as a linear weighted sum of co-variables V_k.

$$\hat{z}_U(\mathbf{x}_0) = \sum_{k=1}^{V} \sum_{i=1}^{n_V} \lambda_{ik} z(\mathbf{x}_{ik}) \text{ for all } V_k \qquad (5)$$

To avoid bias the weights λ_{ik} must sum as follows

$$\sum_{i=1}^{n_v} \lambda_{ik} = 1 \text{ for } U = V_k \text{ and}$$

$$\sum_{i=1}^{n_v} \lambda_{ik} = 0 \text{ for } U \neq V_k \qquad (6)$$

The first condition implies that there must be at least one observation of U for co-kriging to be possible

$$\sigma_U^2(\mathbf{x}_0) = E\left\{\left[z_U(\mathbf{x}_0) - \hat{z}_U(\mathbf{x}_0)\right]^2\right\} \qquad (7)$$

$$\sum_{j=1}^{V} \sum_{i=1}^{n_V} \lambda_{ij}\gamma_{ij}(\mathbf{x}_{ij}, \mathbf{x}_{gk}) + \varphi_k = \gamma_{UV}(\mathbf{x}_0, \mathbf{x}_{gk}) \qquad (8)$$

$$\gamma_{UV}(\mathbf{h}) =$$

$$= \frac{1}{2n(\mathbf{h})} \sum_{i=1}^{n(\mathbf{h})} \left[z_U(\mathbf{x}_i) - z_U(\mathbf{x}_i + \mathbf{h})\right]\left[z_V(\mathbf{x}_i) - z_V(\mathbf{x}_i + \mathbf{h})\right] \quad (9)$$

Co-kriging is the method having the best theoretical foundation, meaning that no assumptions are made on the nature of the correlation between the two variables. It exploits more fully the auxiliary information by directly incorporating the values of the auxiliary variable and measuring the degree of spatial association with the primary variable through the cross-semivariogram. The technique of co-kriging improves the estimation and reduces the variance of the estimation error, but at the same time is much more demanding than kriging in that n(n+1)/2 (n is the number of variables) direct and cross-semivariograms must be inferred and jointly modelled and a large co-kriging system must be solved. The calculation of the cross-semivariogram and the fitting of a theoretical model become very difficult, particularly when the two variables are not strongly correlated.

(iii) Kriging combined with regression
Linear regression analysis was used to predict topsoil clay and sand contents from the values of distance-to-river variable. The S-PLUS statistical package was used for the analysis. Linear regression models were generated to relate the soil variables to distance-to-river variable. The regression models are then used to predict the sand and clay

percentages to 153 sample locations where the values of sand, clay and the distance-to-river variables have been determined. If $Z_m(x)$ is a measured value at location x and $Z_p(x)$ the predicted value of the variable at the same location resulted from the regression then the residuals are obtained by the equation $R(x) = Z_m(x) - Z_p(x)$. Ordinary kriging was applied: (a) on the regressed values $Z_p(x)$ (model RG-KRG) and (b) on the regression residuals $R(x)$ (model RES-KRG). In the RES-KRG model the predicted values were calculated as the sum of the regressed values with the regression residuals at the same locations.

Semivariograms

The spatial variation of the soil properties within the examined area was quantified by semivariogram. The type of the theoretical model, which fitted best to the experimental variogram of each variable, was selected for further geostatistical analysis. The semivariograms of sand and clay variables (measured and regressed values) and of regression residuals for the two soil properties were computed using their values at the 153 sample points. The semivariogram of the distance-to-river variable was computed using the distance values at 503 points.

The semivariograms of the clay and sand content values and the cross-semivariograms of these variables with the distance-to-river variable were better described by a spherical model. This model shows a progressive decrease of spatial dependence of the variable when the distance decreases, which fades away altogether when the model reaches a sill.

A Gaussian model provided a good fit to the distance-to-river variable. The main feature of the Gaussian model is the parabolic shape at the origin. It expresses a smooth spatial variation of the variables. In this model the spatial dependence vanishes only at an infinite distance.

The semivariogram is a function having always positive or null values since it is a form of variance. The cross-semivariogram can take positive, negative or null values since it is a form of covariance. A cross-semivariogram will be positive when the values of the two variables (clay and distance-to-river) have the tendency to vary jointly. It will be negative when the values of the two variables (sand and distance-to-river) have tendency

to vary in opposite directions. It will be null when the two variables tend to vary independently.

Validation procedure

The performance of the four methods was evaluated using the cross-validation method, which involves comparison of predictions of clay and sand content at each data point with their corresponding observed values. Each of the 153 data points was removed in turn and estimates of the soil parameter for the removed point were made from the rest data points.

The effectiveness of each method was assessed by computing two indices from the measured and predicted values. The two indices used are: the mean error, *ME* and the mean square error, *MSE* and they are defined as:

$$ME = \frac{1}{m} \sum_{i=1}^{m} \left[z(\mathbf{x}_i) - \hat{z}(\mathbf{x}_i) \right] \qquad (10)$$

$$MSE = \frac{1}{m} \sum_{i=1}^{m} \left[z(\mathbf{x}_i) - \hat{z}(\mathbf{x}_i) \right]^2 \qquad (11)$$

The *ME* measures the bias of the prediction and should be close to zero for unbiased methods. It indicates whether the model is, on average, producing estimates that are overestimating or underestimating the observed values. The *MSE* measures the average precision of the prediction and should be as small as possible. The model that performs the best will be the one with the smallest *MSE*.

RESULTS AND DISCUSSION

Directional semivariograms were computed for the direction perpendicular to the river channel since short-range variation is observed at this direction due to the various types of deposits. In Table 2, the parameters of the fitted semivariograms for the two soil and the ln(distance-to-river) variables are listed. The parameters of the two cross-semivariograms clay - ln (distance-to-river) and sand - ln (distance-to-river) and the parameters of the fitted semivariograms for the regressed values of the two soil variables and for the regression residuals are also given in Table 2. The directional semivariograms, fitted with spherical and Gaussian models, indicated evidence of anisotropy, therefore anisotropy was assumed for all the kriging and co-kriging calculations. All the semivariograms and cross-semivari-

Table 2. The parameters of the fitted semivariograms

Soil variable	Model	Nugget	Sill	Effective range
Semivariograms				
Measured		——— $\%^2$ ———		— m —
Clay 0 - 25 cm	Spherical	1.6	63.2	4500
Sand 0 - 25 cm	Spherical	35.1	100.2	5400
		——— m^2 ———		— m —
Ln (distance)	Gaussian	0.19	2.64	8900
Predicted from				
Regression with ln(distance)		——— $\%^2$ ———		— m —
Clay 0 - 25 cm	Spherical	1.44	43.9	5100
Sand 0 - 25 cm	Spherical	3.84	59.2	5600
Regression residuals				
$R_{clay} = C_{laymeasured} - C_{laypredicted}$	Spherical	2.94	17.43	2418
$R_{sand} = Sand_{measured} - Sand_{predicted}$	Spherical	12.95	46.48	3192
Cross-semivariograms				
Sand-ln(distance)	Spherical	0.11	-8.45	5000
Clay-ln(distance)	Spherical	0.25	14.4	5800

ograms were computed and modelled using the geostatistical package VARIOWIN (Pannatier, 1996). The selection of the theoretical fitted models was based on the IGF index, and the visual inspection of the experimental semivariograms. The IGF index which was computed by the VARIOWIN is a standardized weighted average squared difference between the observed values and the fitted ones.

For the mapping of the two soil properties regular grids with cell size 50m x 50m were used. The estimated values of the two soil variables resulted from the four geostatistical methods are shown in Figure 2.

The results of comparing the methods: kriging, heterotopic co-kriging and kriging combined with regression (models RG-KRG and RES-KRG) are presented in Tables 3 and 4. In Table 3 the mean errors of the two soil properties for the four methods are given. The mean errors of the two variables are close to zero for all the interpolation methods. This indicates the unbiaseness of the methods. The differences between the methods are very small. However there are more important differences among mean square errors (Table 4).

The mean square errors given in Table 4 present the prediction performance of the four methods. The prediction performance is much wider than the differences in mean errors. This is because in calculating the mean errors the negative and positive bias of prediction tends to cancel each other. From Table 4 the following remarks can be made. Kriging gives the worst results. This is not surpris-

Table 3. Mean errors (*ME*) of the two soil variables according to four prediction methods

Soil variable	ORD-KRG	CO-KRG	RG-KRG	RES-KRG
Clay 0 - 25 cm	-0.028	-0.032	0.005	-0.041
Sand 0 - 25 cm	-0.063	-0.005	-0.007	-0.050

Table 4. Mean square errors (*MSE*) of the two soil variables according to four prediction methods

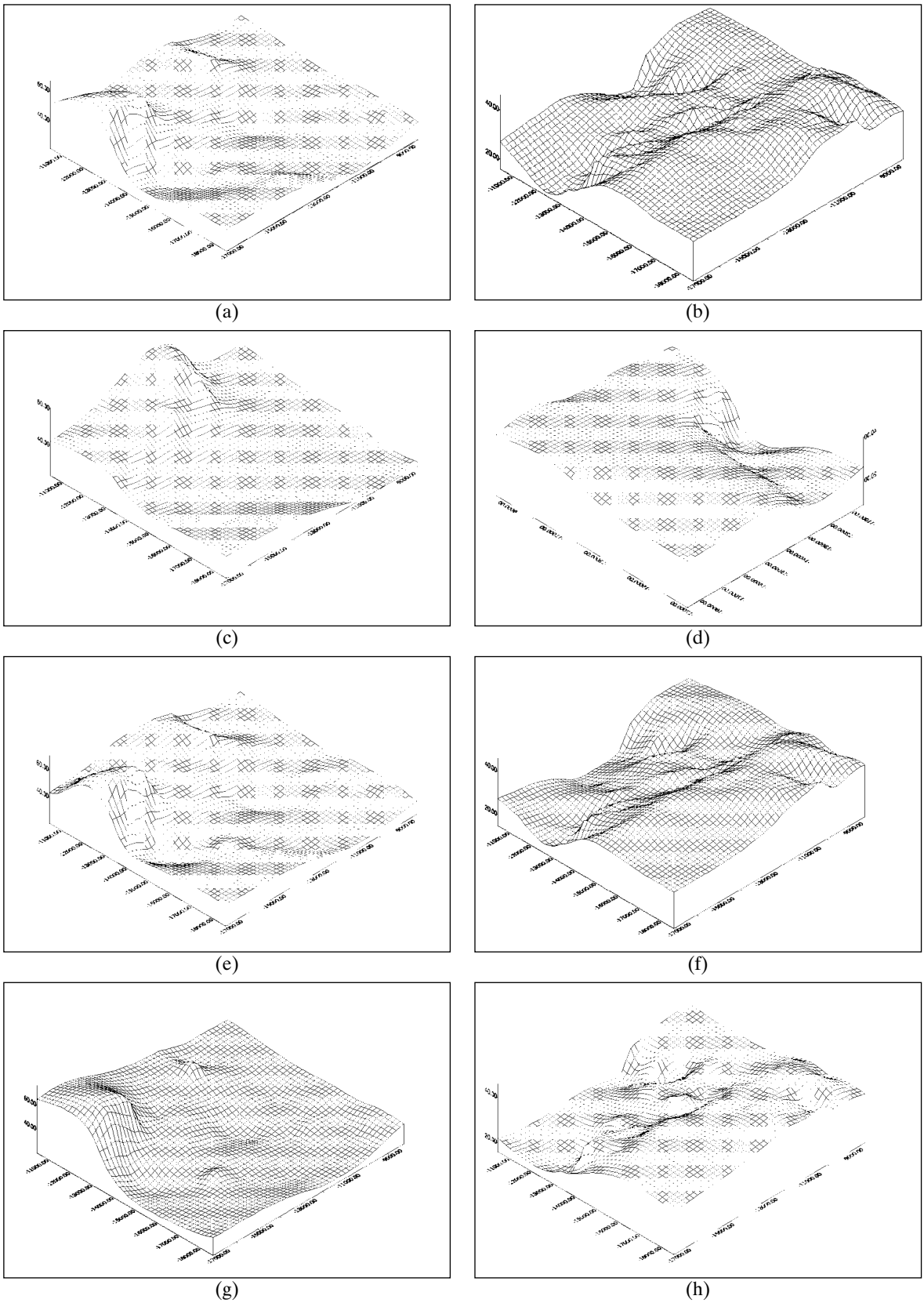| Soil variable	ORD-KRG	CO-KRG	RG-KRG	RES-KRG
Clay 0 - 25 cm	3.18	1.85	1.40	2.78
Sand 0 - 25 cm	12.20	9.45	3.51	12.10

Figure 2. Estimated values of topsoil sand (a, c, e, g) and clay (b, d, f, h) using ORD-KRG (a, b), RG-KRG (c, d), RES-KRG (e, f) and CO-KRG (g, h) models

ing because kriging does not use the covariation of the soil variables with the auxiliary variable. Moreover, according to Laslett and McBratney (1990) kriging methods are characterised by overfitting or under-smoothing of the data.

Heterotopic co-kriging gives good results for both soil variables because fully incorporates the nature and the spatial variability of the soil variables with the auxiliary variable. The co-kriging of clay performs better than the co-kriging of sand due, probably, to the stronger correlation of the clay with the distance-to-river variable.

Kriging combined with linear regression, RG-KRG model, which take uncertainties due to regression and measurements errors into consideration gives the best results for the two soil variables. On the contrary, the summation of the kriged regression values and the kriged residuals, RES-KRG model, does not reduce the mean square errors and performs worst than heterotopic co-kriging and the RG-KRG model for both soil variables.

Ahmed and De Marsily (1987) demonstrated that it is important for choosing between interpolation methods using auxiliary variables to know the strength of the relation between the target variable and the auxiliary variable and the absence or presence of a spatial structure of the residuals of this relation. They also stated that if the residuals are spatially correlated and the correlation coefficients between the variables are high then co-kriging performs better than kriging combined with regression. We get better results from kriging combined with linear regression for both soil variables than from co-kriging in spite that we calculated correlation coefficients 0.82 and - 0.71 between the two soil variables and the auxiliary variable and we have found that the residuals are spatially correlated (Table 2).

CONCLUSIONS

The use of co-regionalisation of the auxiliary variable distance-to-river definitely improved the mapping of the two soil variables topsoil clay and sand. The performance of geostatistical methods: heterotopic co-kriging and two models of kriging combined with regression was evaluated and compared with that of ordinary kriging.

In our study better use of the auxiliary variable in the interpolations of the two soil variables was made using kriging combined with linear regression while the summation of the kriged residuals with the kriged regression values does not give better estimates than co-kriging. Moreover, kriging combined with linear regression has the advantages over the co-kriging method that it uses simpler algorithms, it involves less variogram modelling and the system to be solved is less cumbersome than that of co-kriging. Finally, these result apply strictly only to the particular region. Nevertheless, the methodology should improve predictions in many other regions where soil variables are related to topography.

REFERENCES

Ahmed, S. and De Marsily, G. (1987), Comparison of geostatistical methods for estimating transmission data on transmitivity and specific capacity, *Water Resour. Res.* **23**, 1717-1737.

Burgess, T.M. and Webster, R. (1980), Optimal interpolation and isarithmic mapping of soil properties. I. The semi-variogram and punctual kriging, *Journal of Soil Sci.* **31**, 315-331.

Burrough, P. and MacDonnell, R.A. (1998), Principles of Geographical Information Systems, Oxford Univ. Press, New York. 333 p.

Delhomme, J.P. (1978), Kriging in hydrosciences, *Adv. Water Resour.* **1**, 251-266.

Delhomme, J.P. (1979), Spatial variability and uncertainty in groundwater flow parameters: a geostatistical approach, *Water Resour. Res.* **15**, 269-280.

Gerrard, A.J. (1981), Soils and Landforms, George Allen & Unwin Ltd, London.

Isaaks, E.H. and R.M. Srivastava (1989), An Introduction to Applied Geostatistics, Oxford Univ. Press, New York, 561 p.

Journel, A.G. and Huijbregts, C.J. (1978), Mining Geostatistics, Academic Press, London.

Knotters, M., Brus, D.J. and Oude Voshaar, J.H. (1995), A comparison of kriging combined with regression for spatial interpolation of horizon depth with censored observations, *Geoderma*, **67**, 227-246.

Kollias, V.J., Kalivas, D.P. and Yassoglou, N.J. (1999), Mapping the soil resources of a recent alluvial plain in Greece using fuzzy sets in a GIS environment, *European Journal of Soil Science*, **50**, 261-273.

Kreznor, W.R., Kenneth, R.O., Wayne, L.B. and Donald, L.J. (1989), Soil, landscape, and erosion relationships in a northwest Illinois watershed, *Soil Sci. Soc. Am. J.* **53**, 1763-1771.

Laslett, G.M. and McBratney, A.B. (1990), Further comparison of spatial methods for predicting soil pH, *Soil Sci. Soc. Am. J.* **54**, 1553-1558.

Leenaers, H., Okx, J.P. and Burrough, P.A. (1990), Employing elevation data for efficient mapping of soil pollution on floodplain, *Soil Use and Management*, **6**, 105-113.

Matheron, G. (1973), The intrinsic random functions and its applications, *Adv. Appl. Probabil.* **5**, 439-468.

McBratney, A.B. and Webster, R. (1983), Optimal interpolation and isarithmic mapping of soil properties: V. Co-regionalization and multiple sampling strategy, *Journal of Soil Science*, **34**, 137-162.

McKenzie, N.J., Smettem, K.R.J. and Ringrose-Voase, A.J. (1991), Evaluation of methods for inferring air and water properties of soils from field morphology, *Aust. J. Soil Res.* **29**, 587-602.

McKenzie, N.J. and Ryan, P.J. (1999), Spatial prediction of topsoil properties using enviromental correlation, *Geoderma*, **89**, 67-94.

Moore, I.D., Gessler, P.E., Nielsen, G.A. and Peterson, G.A. (1993), Soil attribute prediction using terrain analysis, *Soil Sci. Soc. Am. J.* **57**, 443-452.

Odeh, I.O.A., McBratney, A.B. and Chittleborough, D.J. (1994), Spatial prediction of soil properties from landform attributes derived from a digital elevation model, *Geoderma*, **63**, 197-215.

Odeh, I.O.A., McBratney, A.B. and Chittleborough, D.J. (1995), Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. *Geoderma*, **67**, 215-226.

Pannatier, Y. (1996), VARIOWIN: Software for Spatial Data Analysis in 2D, Springer-Verlag, New York.

Voltz, M. and Webster, R. (1990), A comparison of kriging, cubic splines and classification for predicting soil properties from sample information, *Journal of Soil Science*, **41**, 473-490.