Multivariate Geostatistical Inventory of Sodicity Hazard in the Hungarian Puszta

M. VAN MEIRVENNE [1], P. DE GROOTE [1], M. KERTESZ [2], T. TOTH [2], G. HOFMAN [1]

1. Dept. Soil Management and Soil Care, University Gent, Coupure 653, 9000 Gent, Belgium.
2. Research Institute for Soil Science and Agricultural Chemistry, H. Ottó ut 15, 1022 Budapest, Hungary.

Abstract

The Hortobágy National Park, located in the east of Hungary, contains the last remnants of the typical Puszta ecosystem of the Great Hungarian Plain. It is characterized by a semi-natural vegetation growing on sodic soils. In order to preserve the original vegetation, the Park authorities wish to have better information about the salinization of the area and its relationship with vegetation.

The basis of the soil and vegetation sampling was a panchromatic SPOT image. This image was subdivided by a quadtree subdivision resulting in 256 quadtree leaves of which the centres have been visited in the field. Sampling included characterisation of the vegetation type and soil samples at 3 depths (0-10, 10-20 and 20-30 cm). The soil samples were analysed for water saturation, pH, electric conductivity and sodium content measured on a saturated soil paste.

Univariate, multivariate (principal component analysis) and regionalized multivariate (Factorial Kriging Analysis, FKA) techniques were used to analyse the relationship between soil variables, or scores on the first principal axis, and vegetation types. FKA was found to perform best and was used to map a "sodicity index".

Résumé

Le Parc National de Hortobágy est situé à l'est de la Hongrie. Il contient les derniers vestiges de l'écosystème typique de "Puszta" de la Grande Plaine Hongroise. Cet écosystème est caractérisé par une végétation semi-naturelle sur des sols sodiques. Afin d'essayer de protéger cette végétation originelle, les autorités du Parc désiraient avoir des informations plus complètes sur la salinisation du périmètre ainsi que sur la relation avec la végétation.

L'échantillonnage du sol et de la végétation était basé sur une image panchromatique SPOT. Cette image était subdivisée en un algorithme *quadtree*, résultant dans 256 feuilles de *quadtree*. Les centres de ces feuilles ont été observés sur le terrain. L'échantillonnage comprend la caractérisation de la composition botanique et des échantillons du sol à 3 profondeurs (0-10, 10-20 et 20-30 cm). Les échantillons du sol ont été analysés pour la saturation en eau, pH, conductivité électrique et la teneur en sodium, tous déterminés sur des pâtes de sol saturé.

Des techniques univariable, multivariable (analyse en composantes principales) et multivariable régionalisée (Analyse Factorielle Krigeante, AKF) ont été utilisées pour analyser la relation entre les variables du sol, ou les "scores" sur le premier axe, avec le type de végétation. L'AKF donnait les meilleurs résultats et a été utilisée finalement pour élaborer une carte d'un "index de sodicité".

Introduction

Assessing the hazard of soil sodicity is a difficult task. Mostly this is done by sampling the soil and by measuring one or some properties of it. One often used property is the soil pH as an expression of the alkalinity. Also, a combined parameter like the sodium adsorption ratio (SAR) has been used to quantify the problem (RICHARDS, 1954). However, since plant growth is adversely affected in salt-affected soils (ABROL *et al.*, 1988), the botanical composition of the land is probably the best indicator of its salt-induced hazards. This is especially true for natural or semi-natural vegetations. Based on reported and personal experience, we characterised a solonized semi-natural vegetation in terms of response to soil sodicity. Then we tried to find uni- and multivariate indices expressing the sodicity hazard of the soil, as reflected by its vegetation, based on analysed soil attributes. Our final aim was to map this hazard using soil properties only.

Study area and vegetation categories

The study area is located in the Hortobágy National Park situated in East Hungary. It contains the largest remains of the salt-affected landscape of the Great Hungarian Plain, known as the Puszta (Fig. 1). The area is extensively grazed, so its vegetation can be considered to be semi-natural. The park authorities wish to have a map indicating the sodicity hazard in order to manage the area and evaluate its dynamics.

Of this region, the soils, vegetation and their interrelationships have been intensively studied for many years (e.g. MAGYAR, 1928; BODROGKÖZY, 1965 ; RAJKAI *et al.*, 1988). Based on BODROGKÖZY's work we defined 6 categories of semi-natural vegetations, each belonging to a distinct position within a topographic catena. The definitions of these vegetation categories and the most dominant associated soil types are described in table 1.

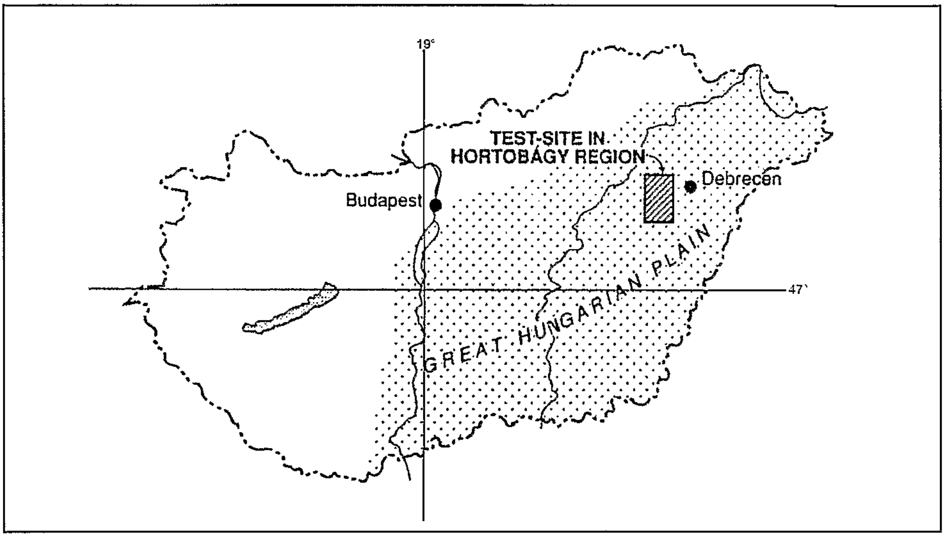

Figure 1. Localisation of the Hortobágy region.

Table 1. Catena of vegetation categories and related soils.

Vegetation Category	Habitat type	Genetic soil type	Soil Taxonomy Great group
Anthropomorphic	Arable land and rangeland	Chernozem	Haplustoll
Steppe	Rangeland	Meadow chernozem with sodic subsoil	Haplustoll
Achilleo-Festucetum	Alkali rangeland	Deep solonetz	Natrustoll
Artemisia-Festucetum	Alkali rangeland	Shallow to crusty solonetz	Natrustalf to Natraquept
Meadows	Slightly alkali meadow	Meadow soil with sodic subsoil	Haplaquoll
Wetlands	Slightly alkali wetlands and temporary lakes	Peaty meadow soil with sodic subsoil	Haplaquoll

It will be noticed that the most severely alkali-affected soils are found at the intermediate catena positions. The top categories (Anthropomorphic and Steppe) are the highest located areas, so the salt containing groundwater has only a limited impact on the vegetation and land use. The low-lying and wettest categories (Meadows and

Wetlands) have less salt-accumulation near the soil surface due to dilution and removal by excess water during the wet season.

Sampling

As a basis for the sampling design of both vegetation and soil, a panchromatic SPOT satellite image with 10 m nominal resolution was used (Fig. 2). The image was taken on September 28, 1990, when the study area was very dry, so when the vegetation cover reflected closely the hazards imposed by salts. The image covered an area of 5120 by 5120 m, which was taken as the study area. The image was subdivided according to a quadtree algorithm, using a modified version of the Kullback-divergence as criterion. Details of this procedure can be found in CSILLAG *et al.* (1995) and KERTESZ *et al.* (1995). This resulted in 256 quadtree leaves of which the centres were the sampling sites. So the sampling scheme reflected the local heterogeneity of the soil surface features and vegetation cover. Direct classification of the image in terms of sodicity hazard was not possible due to the similarity in reflection of completely different features. Further details about the evaluation of this procedure are given by KERTESZ and TOTH (1994).

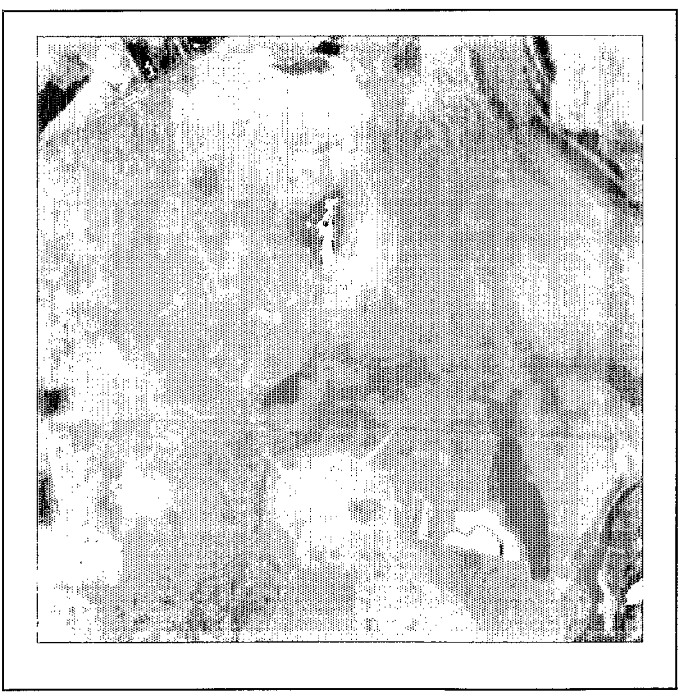

Figure 2. Panchromatic SPOT image of 512 x 512 pixels with a nominal resolution of 10 m, defining the study area.

At each sampling location, soil samples were taken over 3 depth-intervals of 10 cm (0-10, 10-20 and 20-30 cm). Of 64 randomly chosen locations, the vegetation category (VC) was determined (within a circle of 15 m radius) This field work was carried out during the summer of 1991. Every soil sample was analysed for water saturation percentage (SP), pH, electric conductivity (EC) and sodium content, all measured on a saturated soil paste. Additionally, the pH was determined in a standard 1/2.5 soil water solution, here indicated as pH_w. The sodium concentration ($mol.l^{-1}$) was expressed as its negative logarithm (pNa). Some descriptive statistics of these properties (except SP), averaged over the three depths (0-30 cm), are given in table 2. It can be noticed that the EC is strongly positively skewed. Therefore, this variable was logarithmically transformed (ln EC) in order to approximate a normal distribution (g_1 and g_2 approaching zero and three respectively).

Table 2. Descriptive statistics of the determined soil properties, averaged over the three sampling depths (m = mean, s = standard deviation, g_1 = skewness and g_2 = kurtosis).

	M	s	Min.	Max.	g_1	g_2
pH	7.17	0.89	5.13	9.48	-0.27	2.41
EC (dS/m)	1.35	1.03	0.2	6.72	1.66	6.99
ln EC	-0.01	0.32	-0.70	0.81	0.004	2.27
pNa	1.98	0.66	1.03	3.75	0.99	2.88

It can be concluded that the average pH is slightly alkaline and that at average the total salt content (measured by the EC) is relatively low. So the major limitation is created by the sodium content (average concentration 10 $mmol_{[Na+]}.l^{-1}$ measured in the saturation paste). Therefore, plants growing on these soils suffer mainly from the chemical and physical limitations imposed by sodicity.

Univariate relationships

First we tried to relate directly measurable soil properties with the VC's observed at the same locations. Therefore we grouped the locations belonging to the same VC and calculated the means of these properties. To express the variability of these mean values we calculated the standard error of the mean, as $s_e = (s^2/n)^2$ with s the standard deviation of the property of the n observations belonging to that group. Fig. 3 shows the result for pH (0-30 cm), with the mean value symbolised by a dot and enveloped by the 68% confidence interval ($\pm 1 \times s_e$). It can be observed that the Artemisia-Festucetum (Art.-Fest.) category has on average the highest pH, as could be expected by its description (Table 1). However, large overlaps exist for the confidence intervals, suggesting that these differences are not significant at 68% probability level. Similar results were obtained for the other soil variables. Therefore, we concluded that single properties averaged over the topsoil are too heterogeneous to be useful to map the sodicity hazard.

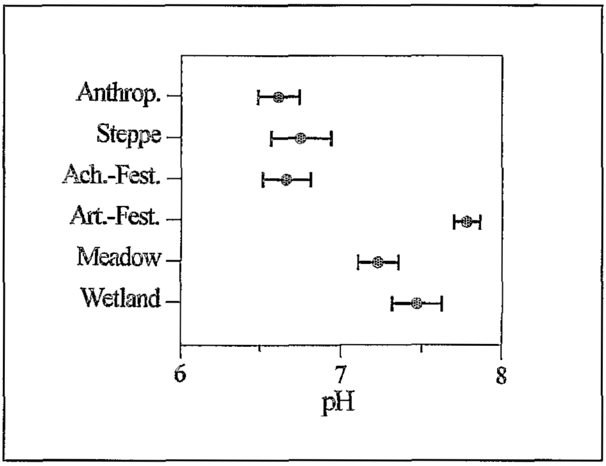

Figure 3. Mean and ± 1 x s_e of pH grouped into vegetation categories.

Multivariate relationship

Next we used a principal component analysis (PCA) (DAVIS, 1986 ; JOHNSON and WICHERN, 1992) to create a multivariate index of sodicity. Therefore we treated the soil analyses (SP, pH, pH_w, ln EC and pNa) separately per sampled layer (0-10, 10-20 and 20-30 cm), yielding 15 different variables. We standardized all variables to a common scale. The first principal component (PC) explained 60% of the total variance within the dataset, the second PC 14,5% and the third 11.1%. Fig. 4 shows the factor loadings projected on the first two PC's.

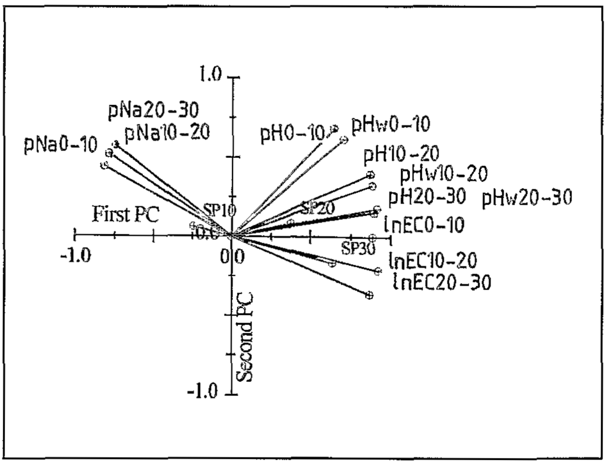

Figure 4. Factor loadings projected on the first two principal components.

It can be concluded that SP does not contribute much to both components, whereas almost all other variables dominate the first PC. So this PC could be interpret as indicating soil sodicity. Moreover, a strong similarity was found for the same variables measured at different depths.

Of each observation point, the scores on the first PC were calculated and averaged within the vegetation categories. Also the standard error of the mean was calculated. The result can be seen in figure 5.

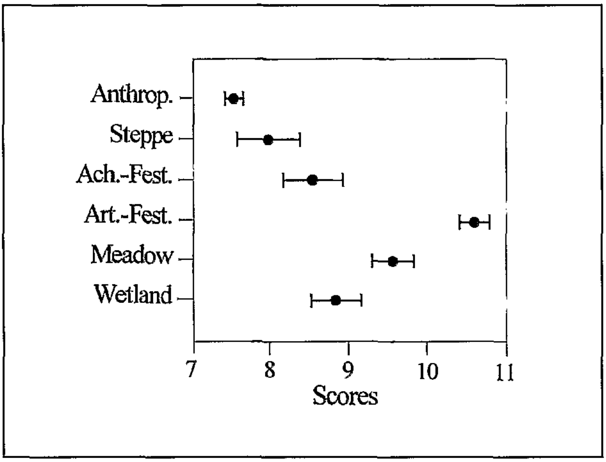

Figure 5. Mean and ±1 x s_e of the scores of the observation points on the first principal component grouped into vegetation categories.

Compared to figure 3, an improvement in the ability to distinguish the different vegetation categories was made possible by this multivariate index. Most neighbouring categories are significantly different. However, some overlaps remain and could hinder the use of this index as indicator of the sodicity hazard.

Scale-dependent multivariate relationship

Geostatistics

In the previous analyses, no use was made of the spatial interrelationships of the observations. Soil properties of a given location were directly confronted with the vegetation growing at that place. However, it is well known that soil processes could act on different scales and influence vegetation at different resolutions. To characterise this, an analysis of the structure of the spatial variability of the parameters characterizing soil forming processes is required. For this aim, geostatistical concepts have been shown to excel (JOURNEL and HUIJBREGTS, 1978 ; WEBSTER and OLIVER, 1990). As an example, figure 6 shows the variogram of ln EC (20-30 cm). It shows that this variable has a spatial dependency up to a dimension of 550 m (called the range). Beyond this separation distance, observations are no longer correlated and thus unusable for spatial prediction, at least not when a reduction in the prediction precision is wanted. The y-intercept (nugget-effect) indicates variability at close distance (the closest sampling interval was 80 m) and/or measurement and sampling errors.

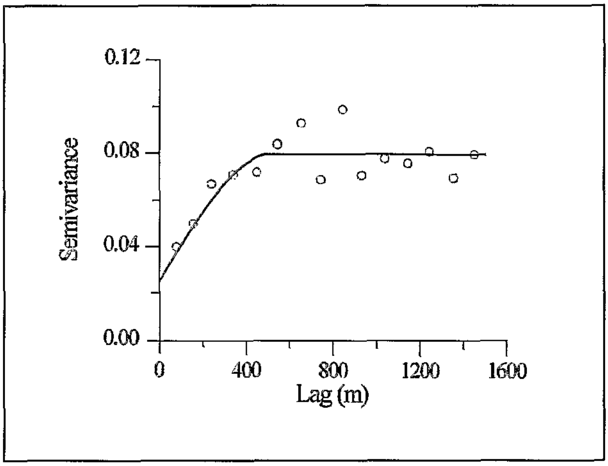

Figure 6. Variogram of ln EC (20-30 cm).

For each variable, the variogram was calculated and it could be used for mapping these variables by kriging. All these variograms showed a range between 320 m (pH 10-20 cm) and 630 m (ln EC 20-30 cm), with most of them between 500 and 550 m. However, since the relationship between these variables and the vegetation categories was unsatisfying, we did not use one of these soil variables solely for mapping the sodicity hazard.

Factor Kriging Analysis

As mentioned before, sampling and measurement errors produce short distance noise, mostly random in nature, while a.o. geological factors generally act at much larger scales. It has been shown by GOOVAERTS and WEBSTER (1994) that by filtering this short range noise and some short distance sources of variability, the ability to correlate the Co and Cu content of soils with the parent material was substantially improved. They used a relatively new approach called Factorial Kriging Analysis (FKA) (MATHERON, 1982). This approach combines the ability of multivariate techniques to analyse the correlation-structure of a data set with the strengths of kriging in terms of spatial prediction. Details on the theoretical background of FKA can be found in GOOVAERTS (1992), GOOVAERTS *et al.* (1993) and GOOVAERTS and WEBSTER (1994).

Simplified, FKA consists out of three major steps : (i) modelling the coregionalization of the multivariate data set and calculation of the structural correlation coefficients for each spatial level considered, (ii) a PCA of the matrix of structural correlation coefficients, resulting in the regionalized factors acting at each spatial level, and (iii) mapping the scores of the regionalized factors of a specified spatial level.

First, a linear model of coregionalization (LMC) was fit to the simple and cross variograms of all variables considered. This was done using the iteration procedure developed by GOULARD (1989). Two spatial levels were considered : one at 0 m (nugget effect) and a second at 550 m. The same variables as before, plus elevation, were considered. Next, for each level the structural correlation coefficients were calculated according to (GOOVAERTS, 1992):

$$\rho_{uv}^{k} = \frac{b_{uv}^{k}}{\sqrt{b_{uu}^{k} b_{vv}^{k}}}$$

with ρ_{uv}^{k} the structural correlation coefficient between variables u and v of the k-th spatial level and b_{uu}^{k} (and b_{vv}^{k}) the coefficient of the simple variogram of variable u (and v) of the k-th level and b_{uv}^{k} the coefficient of the cross-variogram between variables u and v of the k-th level. Table 3 summarises the change of these structural correlation coefficients of the variables (averaged over the 0-30 cm depth) between the first spatial level (0 m) and the second (550 m).

Table 3. Change of the structural correlation coefficients by going from the first spatial level (0 m) to the second (550 m), averaged per variable (positive changes are for positive correlation coefficients, negative changes for negative coefficients).

	SP	ln EC	pH	pNa
SP	0.21			
ln EC	0.06	0.32		
pH	-0.42	0.19	0.21	
pNa	-0.21	-0.72	-0.28	0.59

Table 3 shows that for most variables there is no important change in the correlation structure by moving from the first level to the second, except for two:

- ln EC and pNa, indicating that their correlation increases importantly by increasing the spatial scale;

- pNa and pNa, indicating that the correlation between pNa measured at different depths increases by changing the spatial scale.

This result indicates that it might be worthwhile to filter out the first, local, random component in order to clarify the interrelationships between these variables.

A PCA of the matrix of structural correlation coefficients of the second spatial level (550 m) showed that the first regionalized factor covered 52.1% of the total variance, whereas the second factor covered only 17.5%. Figure 7 shows the factor loadings projected on these first two axes.

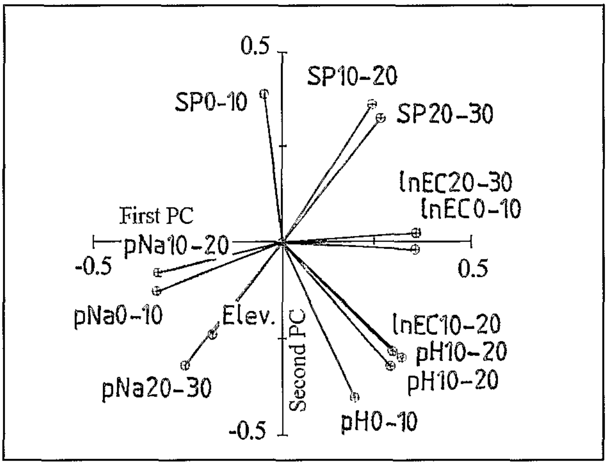

Figure 7. Factor loadings of the first two regionalized factors at lag 550 m.

It can be observed that the first factor is strongly dominated by ln EC at depths 0-10 and 20-30 cm and pNa at depths 0-10 and 10-20 cm. All other variables, except SP 0-10 cm, contributed to both factors in almost equal amounts.

Scale-dependent index of sodicity hazard

The scores on the first regionalized factor of the second spatial level of lag 550 m, were grouped according to the vegetation category to which the observation locations belonged. The result is given in figure 8.

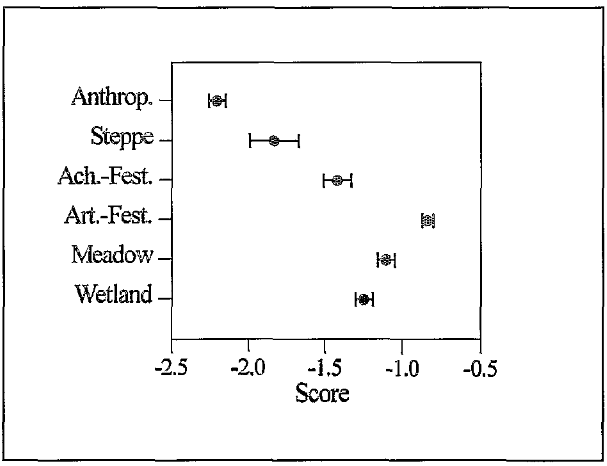

Figure 8. Mean and ±1 x s_e of the scores of the observation points on the first regionalized factor of the second spatial level of lag 550 m grouped into vegetation categories.

A very clear distinction between the vegetation categories was obtained. No overlap of the 68% confidence intervals occurred, suggesting that the differences are significant at this level of probability. So we concluded that this multivariate index of which the local random component was filtered, proved to be the best indicator of the sodicity hazard within the study area.

Map of sodicity hazard

Block-kriging was used to produce a map of the scores on the first regionalized factor of the second spatial level of lag 550 m (Fig. 9). We used blocks of 80 x 80 m (finest sampling resolution). Since the legend of this map is rather arbitrary, we interpret it as varying between slightly sodic and strongly sodic. According to our field experience, this map is a satisfactory reflection of the sodicity hazard of the area.

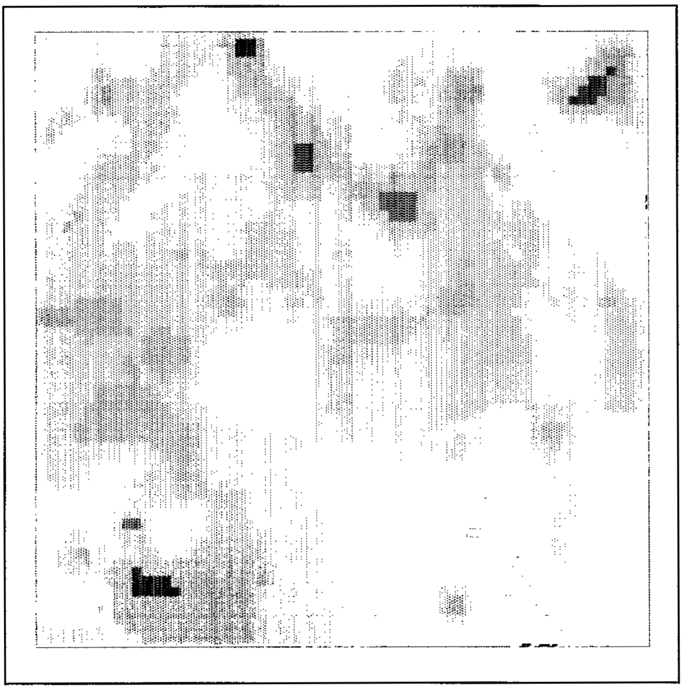

Figure 9. Sodicity hazard map of the study area (continuous grey scale with white = no sodicity hazard, light grey = slight sodicity hazard and dark grey = strong sodicity hazard - scale and orientation are identical to Fig. 2).

Conclusions

Based on our postulation that the natural vegetation is an integrated reflection of the growing environment, we tried to obtain an index of the sodicity hazard of our study area. So we linked observed botanical categories with univariate and multivariate measures created out of measured soil properties.

Univariate soil properties (like the pH, EC and Na-content) showed to be unsatisfactory, large overlaps remained between the categories. The first PC of a PCA improved matters, but the most successful was FKA. The latter includes information of the geographical location in the analysis, allowing to filter out local and random sources of variability. Our approach is also usable in areas where the natural vegetation is strongly disturbed, and thus no longer usable as indicator of the sodicity status of the soil.

This research illustrates the need for using more intensively multivariate techniques in combination with geostatistical concepts. We believe that FKA is promising since it combines both.

Acknowledgements

This research was supported by the Program of Science and Technology, U.S. Agency for International Development (Grant no. DHR-5600-G-00-1055-00), and by the National Science Research Foundation, Hungary (OTKA, Grant no. F-5400). Support of the Bilateral Exchange Program Flanders-Hungary improved this international co-operation considerably. We thank sincerely R. WEBSTER and P. GOOVAERTS for fitting a linear model of coregionalization to our data and providing us with advice concerning the Factorial Kriging Analysis.

References

ABROL I.P., YADAV J.S.P., MASSOUD F.I. (1988). "Salt-affected soils and their management", *FAO soils bulletin* 39, 131 p.

BODROGKÖZY Gy. (1965). "Ecology of the halophilic vegetation of the Pannonicum. II. Correlation between alkali plant communities and genetic soil classification in the Northern Hortobágy". *Acta Botanica Hungarica*, 11 : 1-51.

CSILLAG F., KERTESZ M., KUMMERT A. (1995). "Sampling and mapping two-dimensional surfaces adjusted to the local variability of ancillary data". Accepted *by Int. J. of Geographical Information Systems.*

DAVIS J.C. (1986). *Statistics and data analysis in geology*. J. Wiley & sons, 646 p.

GOOVAERTS P. (1992). "Factorial Kriging Analysis : a useful tool for exploring the structure of multivariate spatial information". *Journal of Soil Science* 43 : 597-619.

GOOVAERTS P., SONNET P., NAVAREE A. (1993). "Factorial kriging analysis of springwater contents in the Dyle river basin, Belgium". *Water Resources Research* 29 : 2115-2125

GOOVAERTS P., WEBSTER R. (1994). "Scale-dependent correlation between topsoil copper and cobalt concentrations in Scotland". *European Journal of Soil Science*, 45 : 79-95.

GOULARD M. (1989). "Inference of a coregionalization model", In : *Geostatistics*, Vol. 1 (Ed. M. Armstrong), Kluwer Academic Publishers, pp. 397-408.

JOHNSON R.A., WICHERN D.W. (1992). *Applied multivariate statisical analysis.* Prentice-Hall International, 642 p.

JOURNEL A.G., HUIJBREGTS Ch.J. (1978). *Mining geostatistics.* Academic Press Inc., 599 p.

KERTESZ M., CSILLAG F., KUMMERT A. (1995). "Optimal tiling of heterogeneous images". Accepted by *Int. J. Remote Sensing.*

KERTESZ M., TÖTH T. (1994). "Soil survey based on sampling scheme adjusted to local heterogeneity", *Trans. 15th World Congr. Soil Sci. ISSC*, Mexico. Vol. 6b : 312-313.

MAGYAR P. (1928). "Data on the coenotaxonomical and geobotanical conditions of Hortobágy", *Erdészeti Kisérletek*, 30 : 26-63 (in Hungarian).

MATHERON G. (1982). "Pour une Analyse Krigeante de données régionalisées", Note N-732, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau.

RAJKAI K., MARCHAND D., OERTLI J.J. (1988). "Study of the spatial variability of soil properties on alkali soils". *Proceedings of the international symposium on solonetz soils, problems, properties and utilization.* Osijek, 15-20 June 1988, pp. 150-155.

RICHARDS L.A. (ed.), (1954). "Diagnosis and improvements of saline and alkali soils", *USDA Agricultural handbook* no. 60, 160 p.

WEBSTER R., OLIVER M.A. (1990). *Statistical methods in soil and land resource survey*, Spatial Information Systems, Oxford University Press, Oxford, 316 p.