GIS-based Stochastic Approach For Mapping Soil Vulnerability

L. PÁSZTOR, J. SZABÓ and T. NÉMETH

Research Institute for Soil Science and Agricultural Chemistry (RISSAC) of the
Hungarian Academy of Sciences, Budapest

The increase of soil degradation at an alarming rate all over the world requires the modelling and quantification of the regional extent and severity of these processes (FAO, 1983; VÁRALLYAY, 1991). The various regional scale soil information systems and the inherent techniques of Geographical Information Science provide a unique basis for studying environmental degradation by modelling changes in soil characteristics, e. g. mapping the vulnerability of soils to degradation or pollution (BATJES, 1994; PÁSZTOR et al., 1998a). The compiled maps may increase awareness on the potential nature, severity and extent of soil degradation at regional scales, and also permit the identification of environmental conflict areas (PÁSZTOR et al., 1998b).

The rapid evolution in computer sciences and informatics from the early sixties made possible the development of a new technique for facilitating the spatial modelling and analysis of various environmental factors, based on computers (MAGUIRE, 1991). Geographical Information Systems (GIS) are designed for storing, querying, analyzing as well as for displaying data and/or information with spatial characteristics. The essence of GIS technology is the mutual treatment of geometrical description and attribute data belonging to the spatial objects. The Spatial Information Theory summarizes the general, application-independent rules of Geographical Information Systems developed for various purposes. Geographical Information Systems represent both the techniques realizing the spatially manageable storage of spatial data and the extended opportunities of spatial analysis of the stored data (SCHOLTEN, 1995). Recently, this latter is getting more and more attention. An interesting feature of recent studies is that they scarcely go beyond the classic univariate statistical methods – ignoring the parallel development in spatial and/or multivariate statistics –, however the studied objects are represented spatially and are characterized by multiple information in these systems. In our work we have tried to pass on, using an approach featured by the introduction of stochastic models (LINHART & ZUCCHINI, 1986) and the application of various multivariate statistical analysis methods (LEBART et al. 1984). There are some evidences in recent works (e. g. COX & SNELL, 1981) that the automatizable multivariate descriptive data processing methods can be applied efficiently to the analysis and – jointly with tools of information theory – modelling of various environmental processes (MURTAGH & HECK, 1987; PÁSZTOR & CSILLAG, 1995).

In our work a stochastic approach is presented for the evaluation of the vulnerability of land and susceptibility of soils to various soil degradation processes. The method was first applied for mapping the N-leaching hazard in Hungary at a scale of 1:1M (NÉMETH et al., 1998). For testing the applicability of the procedure, two further degradation processes were evaluated using the method, and the results were compared to those of deterministic (in our context as opposed to stochastic) approach, that is the map of „Susceptibility of Hungarian soils to acidification" (VÁRALLYAY et al., 1989) and „Vulnerability of Hungarian soils to physical degradation" (VÁRALLYAY & LESZTÁK, 1989) respectively. According to the experimented efficiency, the application of our procedure is highly recommended in land degradation mapping.

The Method

An approach has been developed for the evaluation of the vulnerability of land and/or susceptibility of soils to various degradation processes. According to the basic assumption of the method, land can be characterized by values of a nominal variable in compliance with its vulnerability to a given degradation factor. Evaluation of this variable over an area results in the vulnerability map of the region.

Actually, a semideterministic-semistochastic technique is applied. The relevant, influencing factors are identified. Following the harmonization of input data, the maps of various factors are intersected. Units of the resulted map are elements of a multidimensional factor space. According to the deterministic approach, the severity of a given vulnerability feature can be characterized by some categories, but generally there is no obvious way for determining the classes. According to the stochastic approach, the elements of the factor space can be clustered, however the number, shape and location of these groups are a priori unknown. A sequence of non-hierarchical clustering, as different statistical models, may provide partitions of the map units. For choosing an optimum and parsimonious model over the set of these competing models, information theoretic criteria are suggested to be used. The resulted optimum classification provides the vulnerability categories which are defined on the map units, thus providing the vulnerability map. The steps of the method are (Figure1):

1. Identification of the relevant, influencing factors.
2. Data harmonization (interpolation, scale adjustment, generalization, etc.).
3. Overlay of digital maps.
4. Sequence of non-hierarchical clustering.
5. Information theory based determination of optimum classification.
6. Interpretation of classes, definition of vulnerability categories.

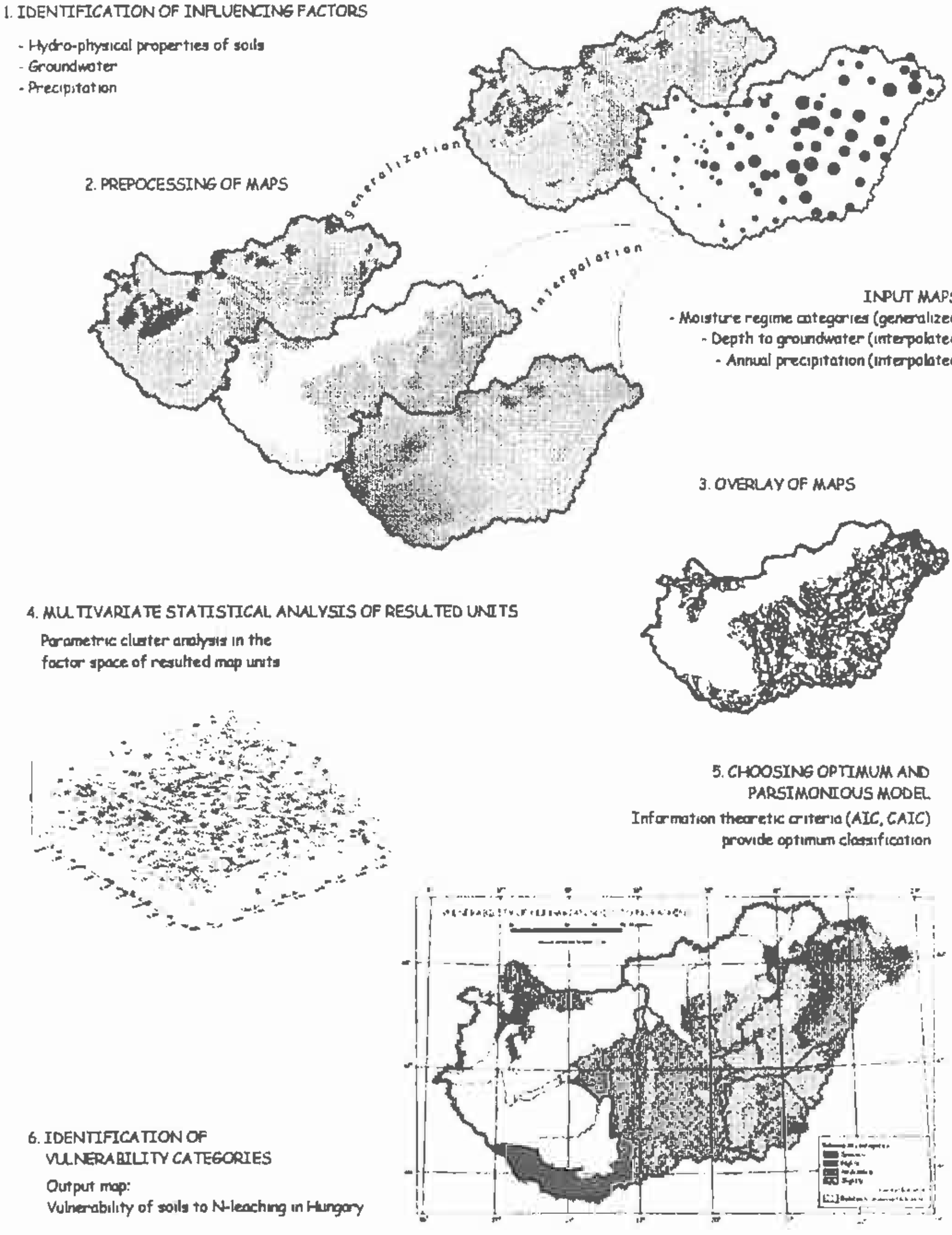


Figure 1

The steps of the method used for mapping vulnerability of soils to N leaching

The output of the algorithm is a digital map showing spatial extension of vulnerability categories. The first application of the method was the 1:1M scale mapping of the N leaching hazard in Hungary (Figure 2).

Mapping the Vulnerability of Soils to Nitrate Leaching in Hungary

In agricultural practice the application of nitrogen to enhance crop yields is necessary in most countries. The improper use of fertilizer-N might play a significant role in the nitrate contamination of subsurface waters which are the main drinking water supplies. The proportion of nitrogen present in mineral forms is greatly affected by climatic and soil conditions, by soil microbial population, and by land use. The ratio between organic and inorganic forms of nitrogen in soil can only be modified slightly over the years. However, under certain environmental conditions (dry climate, negative water balance, overfertilization) a great part of the surplus nitrogen can accumulate in the soil profile, leaving the root zone of various crops, in the form of nitrate after the growing season (NÉMETH, 1993a,b; KOVÁCS et al., 1995), even when land is cropped annually. The integration of knowledge related to the environmental conditions of a certain area with soil, water, and crop management practices helps to prevent the simultaneity of unfavourable processes leading to nitrate leaching, thus water resources may be protected from nitrate pollution of agricultural origin. It is of increasing importance that such an approach be applied in Hungarian crop production. As the great spatial variability of soil forming factors is clearly reflected by the heterogeneous soil cover in Hungary, the differentiation of categories within soil types is strongly needed for agricultural practices.

According to our assumption, various regions of the country can be characterized by a single categorical value in compliance with the vulnerability of their soils to nitrate leaching. A further assumption is that this classification may be determined by three geographical parameters. From now on we follow the steps of the method, presenting its application as it was carried out.

Identification of the relevant, influencing factors. – For the determination of the N leaching hazard under various geographical conditions three influencing factors were considered relevant and available. These are: annual precipitation, groundwater table characteristics and hydrophysical properties of soils. Information on two of them (amount of yearly precipitation, and moisture regime categories of soils) are available within AGROTOPO, the agrotopographical digital database compiled in RISSAC HAS (SZABÓ & PÁSZTOR, 1994), which is composed of territorial information on soils (9 parameters) in a scale of 1:100,000 and meteorological conditions (13 parameters) in a scale of 1:1,000,000 for the whole country. Hungarian soils were classified into nine main soil water management categories according to their hydrophysical properties (VÁRALLYAY et al., 1980). The categories characterize infiltration rate, permeability, hydraulic conductivity, field capacity and water retention features of Hungarian soils. Annual average precipitation data are collected by the National Meteorological Institute and were registered by meteorological stations for the period of 1951-1980. The groundwater depth map is based on data collected in the frame of the groundwater observation well network of the Scientific Research Centre for Water Resources, registered for the period of 1961-1980. The scale of this map is 1:1,000,000 and covers non-mountainous regions of the country.

Data harmonization. – Once the input data were identified their adjustment had to be carried out because of various reasons. (i) Originally groundwater and precipitation data were collected in points, then interpolated, which resulted in isolines. However, our procedure requires polygon features as input. Thus these maps were restructured by transforming contour information into polygon information. (ii) The scale of these two maps and that of soil water management categories map were greatly dissimilar. Due to this fact the information content of the latter was reduced, that is this map was also generalized to a scale of 1:1,000,000. (iii) Finally, as the map of groundwater depth is constrained to non-mountainous regions, the other two maps had to be restricted to this extent, too. The digital form of available data highly facilitated the above detailed work due to its executebility in GIS environment.

Overlay of digital maps. – Having the accurate database as a set of topologically constructed, associated coverages, the geographic analysis was performed. To complete our objectives, the maps of various factors, as different layers, were intersected. Feature attributes from all coverages were joined, that is a polygon of the resulted map was characterized by three attributes as opposed to the single attributes of the original maps.

It is obvious that the inherent errors of both analogue and digitized maps result in shifts of arcs. Thus even the common geographical boundaries deviate from each other in the different maps. As a consequence, in the course of overlay of maps displaying correlated information sliver (that is small and/or elongated) polygons emerge. The number of this kind of polygons can be significant. Since they both a priori distort the results and cause overcomputation, they were eliminated. The criterion of elimination was set up based on the distribution of the area of resulted polygons. The sharp peak of the distribution was cut by defining a threshold value under which polygons were eliminated.

Sequence of non-hierarchical clustering. – From mathematical/statistical point of view units of the resulted map are elements of a multidimensional factor space. As the resulted units are closed polygons, which are also characterized by some geometrical features (area, perimeter, etc.), the grade of the factor space is more than three, however, its three dimensional sub-space (defined by the relevant attribute data) was of interest in our context. The statistical behaviour of the nearly 500 units in this three dimensional sub-space was then studied. Geometrical properties were also involved in the further analysis. Since the number of units does not necessarily reflect their extent, in the computations, evidently, their areas were used as weights.

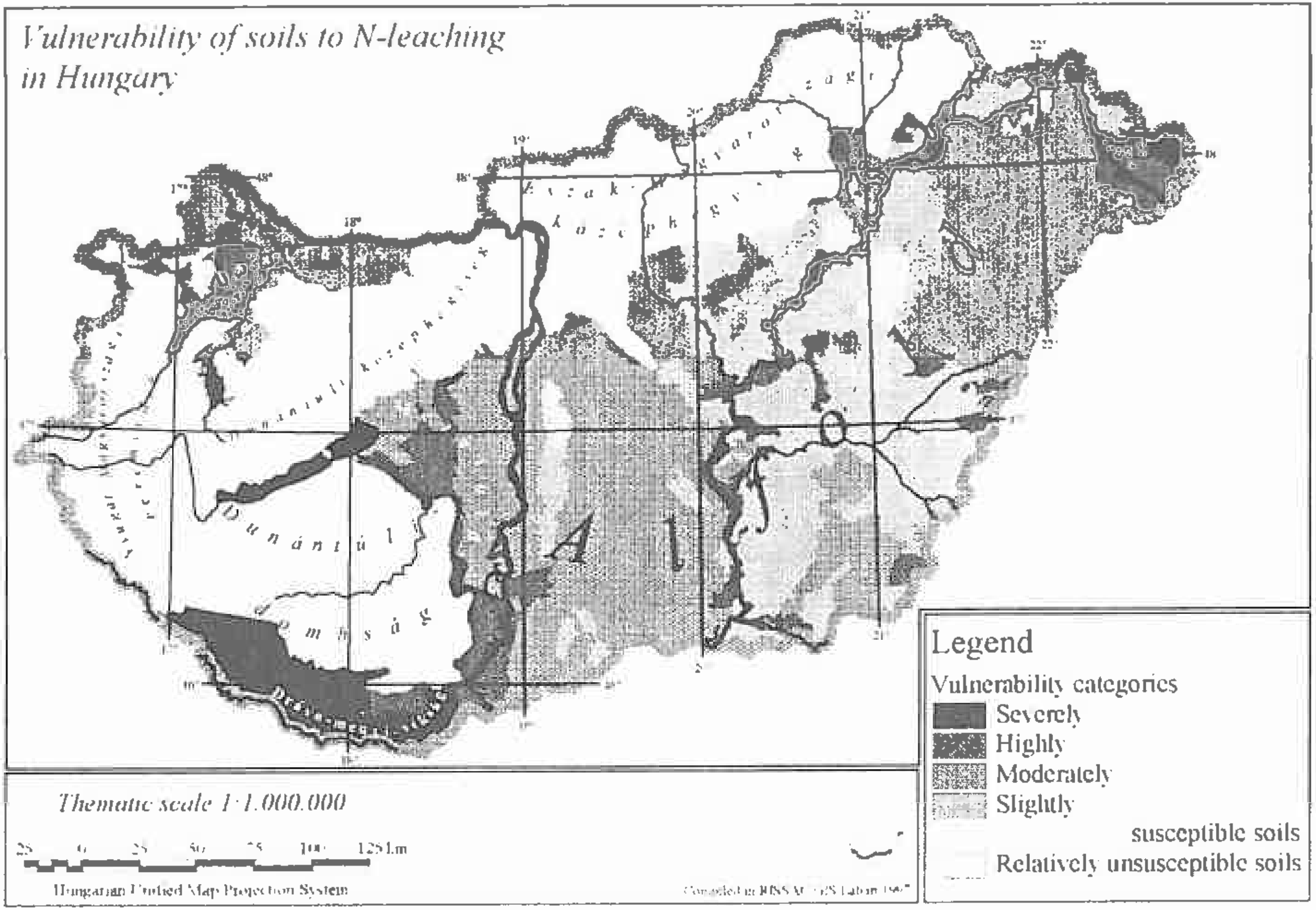
Information theory based determination of optimum classification. – Applying pure clustering techniques where there is no a priori rule to define the optimum number of groups, one needs some measure of the reality of the groups found by the partitioning algorithm. Finding the extreme value of an information theoretic criterion can provide the best fitting model and the best partition of the sample. Many model selection procedures may be found in literature. Most of them take the form of a penalized likelihood, where a penalty term is added to the log-likelihood in order to compromise between the goodness-of-fit and the number of parameters. The ancestor of these models was developed by AKAIKE (1972a) and we also turned to it, since it provides a versatile procedure for statistical model identification. One of the most desirable properties of AIC is that (as it penalizes for large degrees of freedom) it tends to adopt simpler models and achieves a principle of parsimony. As AIC is basically an estimator of the risk of a model selection, it should be minimized to select among the alternative possibilities, that is the smaller AIC is, the better the classification of the points into groups is. Its estimate can be computed by,

$$AIC(estimated)=n\ln(R)+2p,$$

where n is the number of objects to be grouped, p is the number of estimated parameters and R is the residual sum of squares of deviation from the fitted model (AKAIKE, 1974).

Interpretation of classes, definition of vulnerability categories. – The estimated AIC function for our dataset showed two local minima at 5 and 12 categories, respectively. The 12 class solution in our scale provides too detailed thematic resolution, however a possible spatial zooming-in in the future may also require this thematic detailedness. Consequently, merely the 5 class solution was further studied. The identification of vulnerability categories was facilitated by displaying the resulted categories on the units of the intersected map. The different categories showed well recognizable patterns. Analyzing their geographical distribution and extent, it was also possible to rank the resulted categories into a one-parameter sequence from severe hazard to the case of no hazard (Figure 2).

Severely susceptible soils are located on calcareous alluvial deposits where annual precipitation is relatively and groundwater level is actually high. They are characterized by a shallow humous layer but relatively deep and coarse textured pedon, pH is mostly neutral, they have carbonate from the top, their texture class is mainly sand and loam.


Vulnerability of soils to N-leaching in Hungary
Legend
Vulnerability categories
Severely
Highly
Moderately
Slightly
susceptible soils
Relatively unsusceptible soils
Thematic scale 1:1.000.000
25 0 25 50 75 100 125 km
Hungarian Unified Map Projection System
Compiled in RISSAC GIS Lab in 1997
Északi középhegység
Dunántúli középhegység
Dunántúl
dombság
Alföld
16°
17°
18°
19°
20°
21°
22°
46°
47°
48°

Highly susceptible soils are located on alluvial plains, with relatively low (below 600 mm/year) annual precipitation and high groundwater level. The pH, carbonate content and particle size distribution of these soils depend on the flooding material.

Moderately susceptible soils are located on chernozem areas and plains covered by coarse textured soils. In spite of good drainage conditions these poor and humous sandy soils are moderately susceptibility due to the low annual precipitation and deep groundwater level.

Slightly susceptible soils are located on poorly drained hydromorphic soils. They either occupy flat areas or depressions, where the groundwater level is relatively high, soil texture tends to be heavier, the amount of annual precipitation is the lowest in the country.

Relatively unsusceptible soils are located on salt affected landscapes; on heavy textured meadows of floodplains and peats; in poorly drained boggy and swampy depressions. In spite of the different properties of these soils, they are commonly featured by poor drainage conditions.

A 12 class solution of the procedure, by all means, might require more detailed, e. g. multi-level or multi-furcated explanation.

Testing the Method

Soils represent a considerable part of the natural resources of Hungary, consequently their rational utilization has particular importance. As an outcome, in Hungary there is a long tradition of soil mapping, especially in connection with different land degradation processes (VÁRALLYAY, 1989). Based on the agro-topographical map series (analogue progenitor of AGROTOPO database; VÁRALLYAY et al., 1985), a couple of maps were prepared for the whole country in the scale of 1:100,000 or 1:500,000. Two of them are: susceptibility of soils to acidification and susceptibility of soils to physical degradation (structure destruction, compaction). As the AGROTOPO database contains all the necessary information that was used in the compilation of these traditional vulnerability maps, it seemed obvious to subject their determining factors to our procedure.

The eight categories of the map of susceptibility of soils to physical degradation were determined using six soil factors (namely: genetic soil type, soil reaction and carbonate status, soil texture, organic matter content, rootable depth, hydrophysical properties). The present procedure also provided eight categories as optimum solution. After the cross-identification of the stochastic and deterministic classes spatial coincidence could be determined, which resulted in a value of 76%.

In the case of acidification six categories were defined, based on the various combinations of six soil factors (namely: soil reaction and carbonate status, soil texture, organic matter content, rootable depth, hydrophysical properties, parent material). Interpretation of the results of our procedure did not prove to be as straightforward as in the former case. Realizing that one of the soil factors (namely soil reaction and carbonate status) directly defines three of the susceptibility categories, this factor and the respective three categories were excluded and the analysis was repeated. However, neither this way three category solution resulted by our method, but optimum classification was experienced at ten classes. Cross-tabulation and spatial analysis of the result allowed to merge these ten categories into three. Spatial coincidence of these three merged stochastic and the respective deterministic categories proved to be 81%. Consequently, the ten stochastic categories can be treated as refined subclasses of the (in this way rather raw) deterministic classes.

Conclusion

The presented procedure provides an application-sensitive approach to the general problem of evaluating the vulnerability of various environmental elements to different degradation processes. The method relies upon the up-to-date tools of GIS, multivariate statistics and information theory. In the case of mapping the N leaching hazard the compiled map was easily interpretable, providing a straightforward characterization of the vulnerability of Hungarian soils to the leaching of nitrate accumulated after the growing season. A comparison of two traditional vulnerability maps with maps focused on the same degradation processes but prepared according to our procedure also demonstrated the applicability of the introduced approach. Consequently, our procedure can be proposed for application in various fields with similar conditions.

Acknowledgements

The present work was partly funded by the Hungarian National Scientific Research Fund (OTKA) (Grant No. T021275 and F021232).

References

AKAIKE, H., 1972a. Information theory and an extension of maximum likelihood principle. In: Proc. 2nd International Symp. on Information Theory. (Eds.: PETROV, B. N. & CSÁKI, F.) 267-281. Akadémiai Kiadó. Budapest.

AKAIKE, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control. AC-19. 716-723.

BATJES, N. H., 1994. Development of soil databases for global environmental modelling In:. Trans. 15th World Congress of Soil Science. 6a. 40-57. Acapulco.

COX, D. R. & SNELL, E. J., 1981. Applied Statistics. University Press. Cambridge.

FAO, 1983. Guidelines for the Control of Soil Degradation. UNEP-FAO. Rome.

KOVÁCS, G. J., RITCHIE, J. T. & NÉMETH, T., 1995. Testing simulation models for assessment of crop production and nitrate leaching in Hungary. Agricultural Systems. 49. (4) 385-397.

LEBART, L., MORINEAU, A. & WARWICK, K. M., 1984. Multivariate Descriptive Statistical Analysis. Wiley. New York.

LINHART, H. & ZUCCHINI, W., 1986. Model Selection. Wiley. New York.

MAGUIRE, D. J., 1991. Geographical Information Systems: Principles and Application. Longman. London

MURTAGH, F. & HECK, A., 1987. Multivariate Data Analysis. Reidel. Dordrecht.

NÉMETH, T. 1993a. Effect of N fertilization on the nitrate-N content of soil profiles in long-term experiment. Agrokémia és Talajtan. **42.** 115-120.

NÉMETH, T., 1993b. Fertiliser recommendations - Environmental aspects. Zeszyty Prob. Post. Nauk Roln. **400.** 95-104.

NÉMETH, T., PÁSZTOR, L. & SZABÓ, J., 1998. Stochastic modelling of N-leaching using GIS and multivariate statistical methods. In: Proc. of IAWQ 19th Biennial Conference. (In press)

PÁSZTOR, L. & CSILLAG, F., 1995. Reduction of high resolution spectra: application to characterization of salinity status of soils. In: Sensors and Environmental Applications of Remote Sensing. (Ed.: ASKNE, J.) 393-397. Balkema. Rotterdam.

PÁSZTOR, L. et al., 1998a. Land degradation mapping in Hungary. In: Proc. of MERA (MARS and Environmental Related Applications) Project: 1994-1996 Results Conference. (In press)

PÁSZTOR, L. et al., 1998b. HunSOTER as a tool for evaluation of vulnerability of soils to various soil degradation processes in Hungary. In: The Soil as a Strategic Resource: Degradation Processes and Conservation Measures. ESSC Edition. (In press)

SCHOLTEN, H., 1995. Geographical Information Systems, spatial data analysis and spatial modelling, A perspective in European context. In: Proc. of 10th European ARC/INFO User Conference. Prague. 2-14

SZABÓ, J. & PÁSZTOR, L., 1994. The agro-ecological database of Hungary and the possibilities of its application in environmental studies. (In Hungarian). In: National Environmental Conference. Siófok. 156-163.

VÁRALLYAY, GY., 1989. Soil degradation processes and their control in Hungary. Land Degradation and Rehabilitation. **1.** 171-188.

VÁRALLYAY, GY., 1991. Soil vulnerability mapping in Hungary. In: Proc. Int. Workshop on „Mapping of soil and terrain vulnerability to specified chemical compounds in Europe at a scale of 1:5 M" (Wageningen, 20-23 March, 1991). 83-89.

VÁRALLYAY, GY. & LESZTÁK, M., 1989. Susceptibility of soils to physical degradation in Hungary. Soil Technology. **3.** 289-298.

VÁRALLYAY, GY., RÉDLY, M. & MURÁNYI, A., 1989. Map of the susceptibility of soils to acidification in Hungary. In: Ecological Impact of Acidification. (Ed.: SZABOLCS, I.). Proc. Symp. Environmental Threats to Forest and Other Natural Ecosystems. Oulu, Finland, Nov. 1-4, 1988. 79-94. Budapest.

VÁRALLYAY, GY. et al., 1980. Hydrophysical properties of Hungarian soils and the map of their categories in the scale of 1:100,000. (In Hungarian). Agrokémia és Talajtan. **29.** 77-108.

VÁRALLYAY, GY. et al., 1985. Soil factors determining the agro-ecological potential of Hungary. Agrokémia és Talajtan. **34.** (Suppl.) 90-94.