Pergamon

Phys. Chem. Earth (A), Vol. 26, No. 11–12, pp. 885–891, 2001

1464-1895/01/$ - see front matter

PII: S1464-1895(01)00137-5

The Application of Fuzzy C-Means Cluster Analysis and Non-Linear Mapping to a Soil Data Set for the Detection of Polluted Sites

M. Hanesch[1], R. Scholger[1] and M. J. Dekkers[2]

[1]Institute of Geophysics, University of Leoben, Peter-Tunner-Str. 25-27, A-8700 Leoben, Austria
[2]Paleomagnetic Laboratory 'Fort Hoofddijk', Faculty of Earth Sciences, Utrecht University, Budapestlaan 17, 3584 CD Utrecht, The Netherlands

Received 14 July 2000; revised 7 February 2001; accepted 22 March 2001

Abstract. It is important to map the distribution of pollutants and to trace their sources to assess potential environmental hazard. The present work concerns the application of multivariate statistical methods to a soil data base from the province of Styria (Austria) to delineate polluted areas and to distinguish between different types of pollution. The soil data base comprised pedological, geochemical and geological data and was extended by magnetic susceptibility measurements to further test the suitability of magnetic susceptibility as a tracer for pollution. Topsoil data from 521 locations were analysed by fuzzy c-means cluster analysis and non-linear mapping. Robust cluster solutions grouped the database according to the geological background and the land use at the sampling sites. The extraction of information on heavy metal pollution appeared to be possible by analysing the geological units separately and reducing the variables to those indicative for the pollution. The link between magnetic susceptibility and the heavy metal content, which was too complex to be described by bivariate statistics, was revealed by the multivariate methods. © 2001 Elsevier Science Ltd. All rights reserved

1 Introduction

The possibility to monitor environmental pollution by rock magnetic methods has been extensively studied in recent years (Petrovský and Ellwood, 1999). Links between heavy metals and magnetic minerals were already postulated in the 1980's (Thompson and Oldfield, 1986). In some later studies, significant correlations between heavy metal concentrations and low-field magnetic susceptibility were found (e.g. Scholger, 1998). Magnetic iron oxides are able to adsorb heavy metals (Georgeaud, 1999).

In many industrial regions magnetic particles as well as heavy metals primarily originate from iron and steel production, as well as coal combustion in power plants. Higher pollution levels are generally found in urban areas (Flanders, 1994; Strzyszcz et al., 1996). Another important source for magnetic particles associated with heavy metals is road traffic (Hoffmann et al., 1999). Measurements of magnetic susceptibility are comparatively inexpensive and can be carried out rapidly. They provide an ideal first base to map and delineate polluted areas, restricting the more time-consuming and expensive geochemical methods to those regions where a more thorough investigation is desired.

Although some studies indicated a relation between heavy metal pollution and magnetic susceptibility, significant correlations were not always found. Relations between magnetic susceptibility and pollution source are complex, particularly in the case of multiple pollution sources. Even when maps seem to show a link between pollution and magnetic susceptibility, tracing the actual source(s) is not always possible with bivariate statistical techniques (Kapička et al., 1999). Multivariate techniques are more apt to account for complicated links between pollutants and magnetic parameters. For a soil database from Estonia, principal component analysis proved to be useful to identify these links (Bityukova et al., 1999).

Another way to discriminate polluted and unpolluted areas is the recognition of groups in the data set by fuzzy c-means cluster analysis (FCM) and non-linear mapping (NLM). These methods have been applied successfully in geochemistry (e.g. Vriend et al., 1988) and have recently also been used to link rock-magnetic parameters to the geochemical environment (Dekkers et al., 1994), to magnetically characterize ocean sediments (Schmidt et al., 1999) and to identify connections between climatic influences and magnetic parameters (Kruiver et al., 1999). FCM is 'unsupervised': no a-priori knowledge concerning any grouping is assumed. In the present contribution, FCM and NLM are used to analyse a database combining low-field magnetic susceptibility data with pedological, geochemical, and geological parameters. It will be shown that pollution effects are expressed only as a 'second-order type phenomena' in the data.

Correspondence to: M. Hanesch

2 Data

The soil samples were taken in Styria, a province (16387 km^2) in southeastern Austria. A regular grid of 4 by 4 km over the whole province ensured a systematic investigation covering the total area. Large forest areas were not sampled. At regions which were estimated to be polluted, some additional samples were taken. Each site is given by a circle with a radius of 10 m around the calculated grid point. In the four main directions – north, east, south, west – samples are taken out of defined layers: 0 – 20, 20 – 50, 50 – 70 cm for agricultural soils, 0 – 5, 5 – 20, 20 – 50 cm for other sites. The four samples of each layer were dried and mixed to obtain representative data. Sampling and geochemical analyses were performed by the Styrian Agricultural Laboratory. The data kindly made available for this study comprised the following variables: geology at the sampling site, geographical coordinates, soil type, land use, pH value, P_2O_5, K_2O, exchangeable bases (Ca, Mg, Na, K), clay content, humus, fluorine, EDTA-extractable manganese and iron, and heavy metals extracted by aqua regia (Pb, Cr, Ni, Co, Mo, Cd, Hg, As, Cu, Zn). The analytical methods are described by Krainer (1998).

The dried soils are stored in the soil archive of the Styrian Agricultural Laboratory in plastic cylinders (9.5 cm diameter, 13.5 cm height), which can be measured without any sample manipulation with a Bartington MS2C loop sensor for cores (125 mm; operating frequency 565 Hz). Low-field volume susceptibility of the samples was measured using range 1 of the MS2 meter with a resolution of $1*10^{-5}$ SI and a measuring range of $99*10^{-3}$ SI. The raw data were corrected to account for sample-diameters and -lengths different from the standard dimensions. We established a specific calibration factor for the sample containers by comparing the results from the loop sensor with mass specific measurements on 134 sieved sub samples. The correlation coefficient between the volume specific and the mass specific measurement was 0.96, justifying the use of the faster MS2C measurements.

For the purpose of the statistical analyses, a topsoil value was defined as the value in the layer 0 – 20 cm. A weighted average of the uppermost two horizons was calculated to obtain a topsoil value for grassland and Alpine pasture sites. Data ranges of magnetic susceptibility and the most important heavy metal concentrations in the topsoil layer, for the samples used in this study, are given in Table 1.

Table 1. Minimum and maximum values for the magnetic susceptibility and the most important heavy metals measured for the samples used within this investigation. The corrected susceptibility values are given as mass specific susceptibility.

	susceptibility (10^{-8} m^3/kg)	Pb (mg/kg)	Cr (mg/kg)	Ni (mg/kg)
minimum	0.2	8	5	1
maximum	132.5	254	324	420
	Co (mg/kg)	Cd (mg/kg)	Cu (mg/kg)	
minimum	1	0.1	3	
maximum	41	2.0	135	

3 Methods

As a first step in the analysis, the distribution of the variables and the correlations between them were analysed. Heavy metal concentrations and susceptibility show nearly lognormal distributions. Therefore, the logarithmic values of these variables were used as input for the multivariate statistical analyses. Outliers were removed from the dataset to obtain robust solutions. Outliers were defined as values deviating more than 3 times the standard deviation from the mean logarithmic value.

Two multivariate methods were used in combination to find subsets in the soil dataset: fuzzy c-means cluster analysis and non-linear mapping. In cluster analysis, a set of data points (or cases) is split into a given number of groups (or clusters). In the conventional hierarchical approach each case is completely allocated to a single cluster, thus assuming well-defined boundaries between clusters. The fuzzy approach used here calculates the similarity of a case to all clusters. This similarity is expressed by a membership value that varies between 0 and 1 for each case. All membership values for one case sum up to 1. Thus, gradual changes between the clusters can be described. Soil composition depends on many geological and environmental processes; gradual transitions between different soil types and a wide range of pollution stages are to be expected. The fuzzy approach traces these gradual changes and is therefore more appropriate than the conventional approach. The algorithm was first published by Bezdek et al. (1984). For the representation of the data points on maps, the cases are assigned to the cluster with the highest membership value if the ratio of the second highest membership value to the highest value is smaller than 0.75. Otherwise, the case is classified as 'intermediate'.

In non-linear mapping, a multidimensional scaling method, the distance between the data points in the multidimensional variable space is approached as far as possible by a distance matrix in a 2-dimensional space (Sammon, 1969). The graphical display of the 2-dimensional solution illustrates the relations between the samples in the multidimensional parameter space. Labelling of the samples in this 2-dimensional map according to the cluster affiliation calculated before provides us with a rapid quality control of the cluster solution. When the cases of the same cluster coherently group on the non-linear map, it is probable that the clustering is meaningful (compare Figs. 4 and 5).

The number of clusters to be calculated must be specified in advance. Since the most appropriate number is not a priori known, solutions are calculated for cluster numbers between 2 and 9. The best solution is chosen by comparing the non-linear maps as well as by calculating the partition coefficient F and the classification entropy H (Bezdek et al., 1984) which indicate the robustness of the grouping from a mathematical viewpoint (Table 2). The best number of clusters in this sense is given by the lowest H and highest F value. Furthermore, the robustness of the solution is tested by running the algorithm several times starting from differ-

Table 2. Calculation of the validity functionals and their limiting values. N: number of samples, c: number of clusters, u_{ki}: membership value for the k^{th} sample in the i^{th} cluster. For the comparison of F and H for different numbers of clusters, the values of F and H are scaled with regard to the interval of possible values.

Validity functional	Formula	Limits
Partition coefficient F	$\sum_{k=1}^{N}\sum_{i=1}^{c}[(u_{ki})^2/N]$	$1/c \le F \le 1$
Classification entropy H	$\sum_{k=1}^{N}\sum_{i=1}^{c}[u_{ki} * \ln(u_{ki})/N]$	$0 \le H \le \ln(c)$

ent samples. Only if a solution is stable in different runs and if a sound explanation of the clusters can be found by further analysing the memberships and the location of the cluster centers, the solution is accepted.

4 Results and interpretation

The first part of the analysis consisted of including all numeric variables (23), i.e. all variables except geology, land use and soil type. The validity functionals for the different numbers of clusters are given in Table 3. They indicate that the three cluster solution is the best solution. Accordingly, a sound explanation of the three clusters could be found. The memberships reflect the three main geological units of the region (Fig. 1 and Fig. 2). There is a clear distinction among the gneiss and schist areas (Cluster 1, □), the sediments (Cluster 2, x), and the Paleozoic rocks and greywacke (Cluster 3, ▲). Sediments in the vicinity of the third group of rocks are assigned to cluster 3 rather than to cluster 2 containing the sediments in the basin of the southeast. The different land use in the basin might be one of the causes for this observation. Land use is in part related to the clustering because land use is dependent on geology in this mountainous area.

The same cluster analysis without susceptibility produces a similar result. There are some points which cross the threshold between intermediate points and a cluster assignment, but none of the cases changes from one cluster to another cluster. There is a significant difference between the susceptibility values of the geological units (assessed by using the Mann-Whitney test for the equality of the median values), but susceptibility is only one among 23 variables. pH-value, nutrients and some trace elements also differ significantly between the soils formed on the different geological units.

To avoid the influence of the geological background, the clustering was repeated for the samples from one geological unit only. The sedimentary basin is formed by Tertiary and Quaternary sediments. The Tertiary sediments were chosen for this experiment because they are a sufficiently large group (122 samples) and most of them had a high membership to the second cluster. Thus, a background as homogeneous as possible was created for the samples in this second

Table 3. Validity functionals for the cluster solutions for all data points using all variables. The scaled values for F and H are given (see Table 2). From the mathematical viewpoint, the three-cluster solution is the best solution as it has the lowest H and the highest F value. It also has a physical meaning since it reflects the geological units of the region (compare Figs. 1 and 2).

N	F	H
2	0.108	0.772
3	0.174	0.692
4	0.145	0.720
5	0.124	0.746

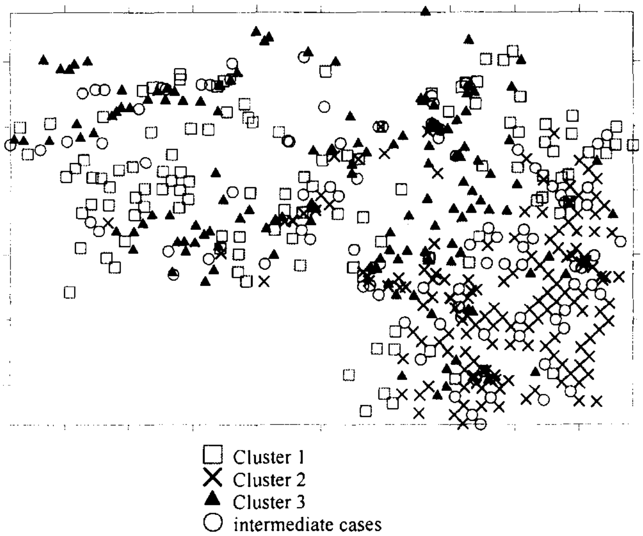

Fig. 1. Prevailing cluster membership for the geographical location of the samples (top of picture points north). The distance between two ticks is 20 km. 23 variables were used for 521 data points.

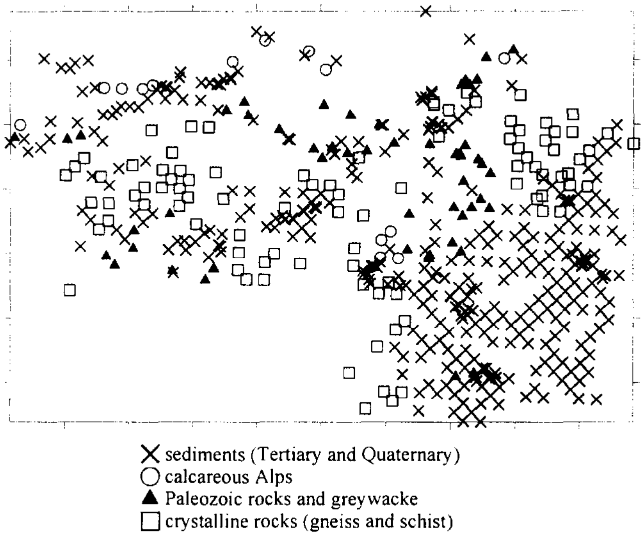

Fig. 2. Geological units at the sample locations. The same points as in Fig. 1 are shown.

experiment. The same 23 variables were used. Again, a three-cluster model appeared to be the best solution from a

Table 4. Validity functionals for the cluster solutions for the Tertiary sediments using all variables. The highest F and lowest H is calculated for the three-cluster solution which is also a good solution according to the non-linear map (Fig. 4).

N	F	H
2	0.109	0.770
3	0.140	0.741
4	0.109	0.783
5	0.094	0.803

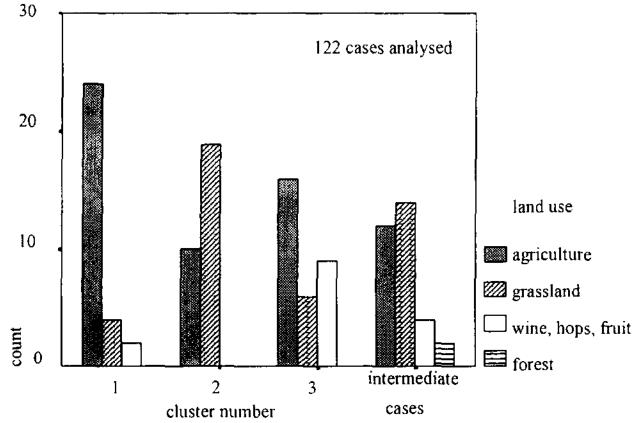

Fig. 3. Fuzzy c-means cluster analysis for the Tertiary sediments. Distribution of the land use in the clusters. The grouping is determined by the intensity of land use. Cluster 1 contains mainly agricultural soils, cluster 2 mainly grassland and cluster 3 contains the most intensively used soils: almost all the wine, hops and fruit cultivation is found there.

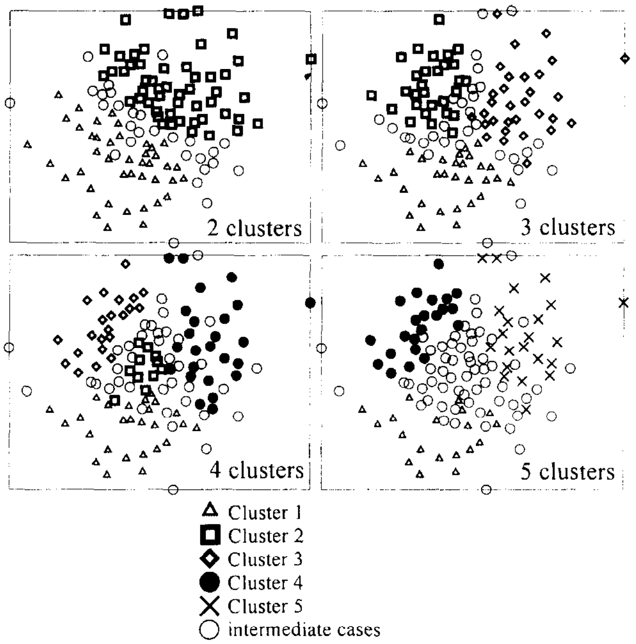

Fig. 4. Plot of the cluster solutions for the Tertiary sediments (all variables) in the nonlinear map. The three-cluster model shows only a few intermediate cases whereas the five-cluster model classifies a large number of cases as intermediate. After defuzzification, no samples would be allocated to two out of the five clusters, indicating that this five-cluster model is not robust.

mathematical viewpoint. The values of the validity functionals are given in Table 4. Comparison of the cluster memberships with land use shows that the clusters clearly reflect the intensity of land use (Fig. 3). Susceptibility does not differ significantly for the land uses; pH, humus content and nutrients should be responsible for this grouping. The copper content plays an important role in distinguishing the wine cultures.

These first two clustering experiments showed that the variables indicating geology and land use are dominating. Also, a further subdivision of the three-cluster model was not possible, because many samples (~31 % of the samples for the four-cluster solution and ~48 % for the five-cluster solution for the Tertiary sediments) had to be classified as intermediate cases (Fig. 4). Apparently, using all variables leads to overinformation: the samples are insufficiently distinguishable from one another.

4.1 Effects of pollution

To extract the information on pollution contained in the data set, the number of variables has to be reduced. Analysis so far has indicated that in the Styrian setting, bedrock and land use are dominant factors. Because our prime interest is the relation between heavy metals and susceptibility, susceptibility was selected along with some heavy metals which are known to be geogenic in the study area (Cr, Ni, Co, Cu) and some which are typically anthropogenic (Pb, Cd) as pointed out in the soil survey report of the Styrian Agricultural Laboratory (Kreiner, 1998). Regional normal values of metal concentrations were defined during this survey by analysing subsoils. Lead and cadmium were the only trace elements which exceeded these normal values in more than 30 % of the topsoils. The same result was obtained by analysing the difference between topsoil and subsoil values: as a first approximation, anthropogenic input was inferred when this difference exceeded two times the error of the analysis. This occurred for lead in 68 % of the soils, for cadmium even in 86 % of the soils. On the other hand, less than 15 % of the soils showed this behaviour concerning Cr, Cu, Ni and Co.

With this set of seven variables (susceptibility, Cr, Cu, Ni, Co, Pb, Cd), there appeared to be much less intermediate cases (less than 15 % for cluster numbers between 2 and 5, see Fig. 5). Based on the validity functionals (Table 5), the two-cluster solution should be selected, although solutions with more than two clusters seem acceptable as well. The picture becomes obscured for six clusters where no satisfactory solution was achieved in 250 iterations. With a higher number of clusters increasingly subtle features of the data set are visualized. The two-cluster solution distinguishes between samples with high values for all variables in one cluster ('polluted cluster') and samples with lower values in the other cluster ('unpolluted cluster'). In the

Table 5. Validity functionals for the cluster solutions for the Tertiary sediments using the reduced set of variables: magnetic susceptibility, Pb, Cr, Ni, Co, Cu, Cd. According to the F- and H-values, the two-cluster solution should be chosen.

N	F	H
2	0.499	0.334
3	0.455	0.360
4	0.436	0.369
5	0.417	0.389

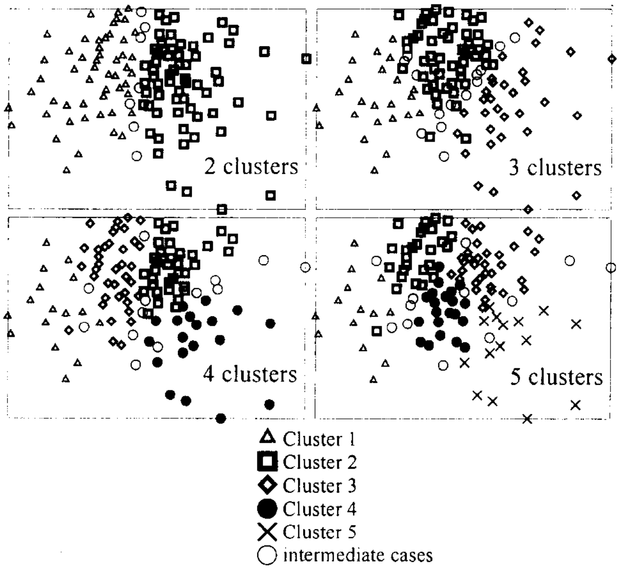

Fig. 5. Plot of the cluster solutions for the Tertiary sediments (reduced set of variables) in the non-linear map. There are less intermediate cases than in the solutions with all variables (compare Fig. 4).

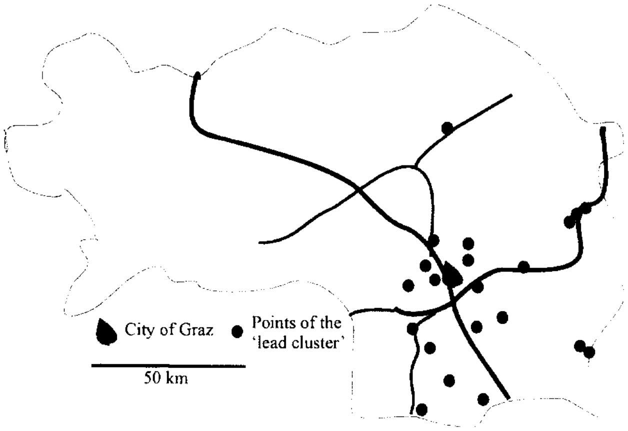

Fig. 6. Schematic drawing of the study area with the main roads of the region. The points mark the locations of the samples which were attributed to a cluster characterized by high values of lead in the five-cluster model for the Tertiary sediments with the reduced variable set: Cr, Ni, Co, Cu, Cd, Pb, magnetic susceptibility.

three-cluster solution, a cluster appears with relatively high values for lead and for cadmium and lower values for the other heavy metals ('cadmium cluster'). The cadmium in these samples is envisaged to originate from mica schist.

In the four-cluster solution, an additional cluster with relatively high values of the geogenic variables (Cr, Cu, Ni, Co) and moderate values of the anthropogenic variables (Pb, Cd) is separated from the 'unpolluted cluster'. It is interpreted as an anthropogenically unpolluted cluster with some geogenic heavy metals ('geogenic cluster'). In the five-cluster solution, a new cluster appears which has the second highest lead values and the second highest susceptibility values ('lead cluster') whereas the values of the other heavy metals are relatively low. It is mainly formed by cases which separate from the 'geogenic cluster' but the 'polluted cluster' also contributes to it. The highest lead and susceptibility values are still allocated to the 'polluted cluster' where values for all variables are high (likely industrial contamination). It has to be noted that no real hierarchy exists: though new clusters may be formed mainly by points of one of the old clusters, usually there are cases from other clusters which contribute to it.

It is now possible to distinguish between polluted and unpolluted soils. As an example, Figure 6 shows the samples of the 'lead cluster' of the five-cluster solution, printed on a road map of Styria. They group around the city of Graz and along the main roads. These lead contaminations are presumably caused by the traffic. There are, however, some points as well where the high lead values are probably geogenic in origin. Anthropogenic and geogenic contamination are mixed in this cluster.

The two clusters which are estimated to be polluted (cluster 4: 'lead cluster' and cluster 5: 'polluted cluster') have significantly higher susceptibility values as is demonstrated by the boxplots of the logarithmic susceptibility values of the five clusters (Fig. 7).

The influence of susceptibility on the cluster assignment is higher when using only seven variables instead of 23. However, only nine out of the 122 cases changed from one cluster to another when susceptibility was excluded from the analysis. Also, the general trend in cluster formation remains the same. The potential use of susceptibility as a pollution indicator requires the solution with susceptibility as a variable. To define areas where further investigations might be worthwhile, this procedure seems to be warranted.

The definition of the groups would have been impossible by using bivariate statistics only. Figure 8 shows the scatterplots for the logarithmic values of susceptibility versus the logarithmic values of lead and chromium. From cluster 2 to 5, susceptibility values rise with increasing lead contents (Fig. 8a). This accounts for their correlation within the whole data set (correlation coefficient between lead and susceptibility: 0.311). Within the individual clusters, the data points are basically uncorrelated. The correlation between lead and susceptibility within the 'lead cluster' is insignificant. Cluster 1 is not distinguished by the lead content.

Other variables have to be analysed to find the reasons for the discrimination of cluster 1. For example, it clearly separates in the plot of susceptibility versus chromium (Fig. 8b). The representation of the data points in real two-dimensional planes (Fig. 8) appears to be less distinctive than the non-linear map (Fig. 5). In the scatterplots the

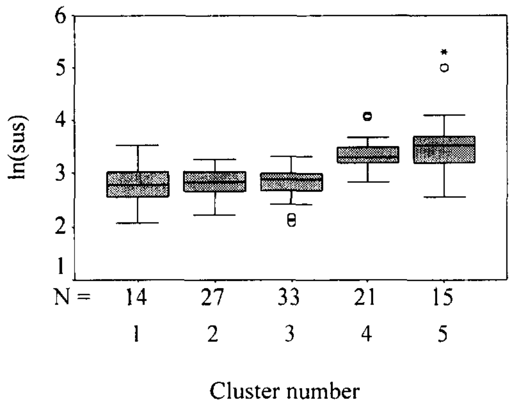

Fig.7. Boxplot of susceptibility for the clusters built by using the reduced set of variables on soils formed on Tertiary sediments. The boxes contain 50 % of the values in the cluster. The line in the box indicates the median value. N is the number of cases in the cluster. The two clusters (No. 4 and 5) which are estimated to be affected by pollution, have significantly higher susceptibility values than the other clusters.

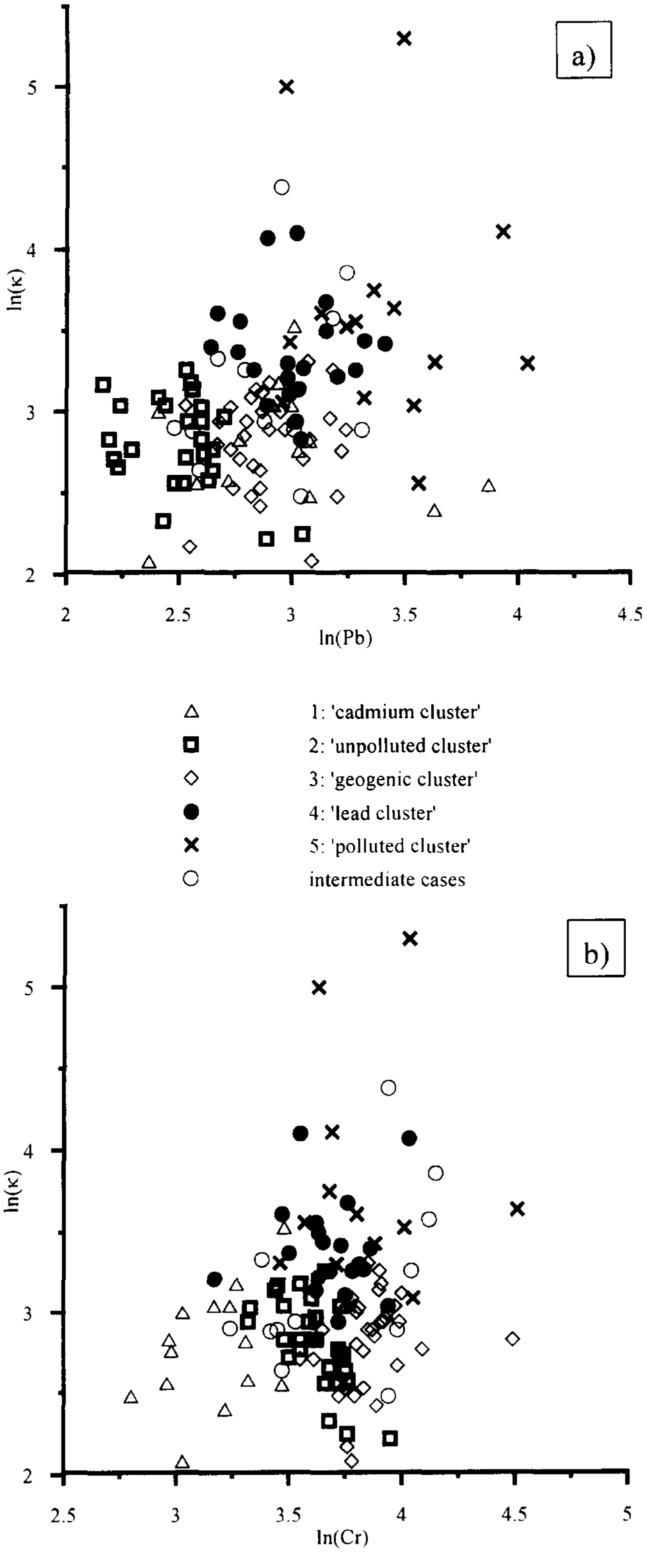

Fig. 8. Fuzzy cluster analysis for Tertiary sediments with the reduced set of variables. a) Points of the clusters in the ln(Pb)-ln(sus) space. From cluster 2 to 5, susceptibility and lead values tend to increase. Lead obviously plays no role in distinguishing cluster 1. b) Points of the clusters in the ln(Cr)-ln(sus) space. Here, cluster 1 separates quite clearly from the other clusters. Finding groups in the data set by analysing the bivariate plots would be much more difficult than by multivariate methods as the points are not as clearly distinguished as in the nonlinear map (Fig. 5).

points are mixed and it would be necessary to analyse many of these plots at the same time to get an idea about possible groups. This clearly shows the merit of multivariate analysis by taking several variables simultaneously into account to arrive at a consistent pattern.

5 Conclusions and implications

Fuzzy c-means cluster analysis and non-linear mapping on the Styrian soil data set resulted in a grouping according to the main geological units of the study area when all topsoil variables at all sites were used. Splitting of the samples according to their geological unit and redoing the analysis resulted in a grouping according to land use. This proves that the results of the analysis have a physical meaning and that the algorithm works successfully. But for solutions with higher numbers of clusters, almost all data points were classified as intermediate. The distinction between the cases is insufficient if too many variables are used.

This difficulty could be circumvented by choosing a smaller set of variables, in the case of the present study indicative of pollution. Fuzzy c-means cluster analysis for one geological unit with this reduced variable set returned interpretable information on pollution, setting the stage for an interpretation of the pollution sources. Geogenic and anthropogenic contributions appear to be mixed.

Like in previous studies, susceptibility turns out to be related to heavy metal content. The relations appear to be complex and influenced by many parameters, like the geological background, land use, and various pollution sources. On the one hand, these multiple influences prevent a straightforward application of magnetic susceptibility on its own to distinguish between various pollution sources. On the other hand, however, most pollution sources appear to produce magnetically susceptible materials as a 'byproduct'. Therefore, susceptibility measurements are a rapid tool to define those areas which warrant further investigation. Multivariate analysis, including the appropriate variables, can then be used to distinguish between different sources of pollution.

Two main points have to be taken into account when using the combination of fuzzy c-means cluster analysis and non-linear mapping for the interpretation of soil data sets: 1) If only topsoil values are used, geological background quickly dominates the outcome of the analysis; when more subtle information is desired, the data set should be divided into subsets with the same geological background. 2) Too many variables may veil the differences between the samples. The set of variables must be reduced according to the aim of the analysis. When these constraints are taken into account, the method is a powerful tool to find meaningful groupings in soil data sets.

Information on potentially hazardous levels of heavy metals in soils has undoubtedly implications for agriculture. Research into how far they are incorporated in the cultivated plants, independent of the origin of the pollutants, may be required. If the analysis is to be used to define anthropogenically polluted areas, it will be necessary to optimise the investigation by finding a method to better take the background into account. By including subsoil samples in the analysis it may be possible to differentiate meaningfully between airborne heavy metals and those originating from the bedrock.

Acknowledgements. This study was carried out in the frame of the MAGNET project (EU FMRX980247). The Styrian Agricultural Laboratory provided the soil data. The help and advice of Dr. W. Krainer is gratefully acknowledged. Two anonymous PCE reviewers are thanked for their helpful comments.

References

Bezdek, C.J., Ehrlich, R., and Full, W., FCM: the fuzzy c-means clustering algorithm, *Comput. Geosci.*, *10*, 191-203, 1984.

Bityukova, L., Scholger, R., and Birke, M., Magnetic susceptibility as indicator of environmental pollution of soils in Tallinn, *Phys.Chem. Earth (A).*, *24*, 829-835.1999.

Dekkers, M.J., Langereis, C.G., Vriend, S.P., van Santvoort, P.J.M., and de Lange, G.J., Fuzzy c-means cluster analysis of early diagenetic effects on natural remanent magnetisation acquisition in a 1.1 Myr piston core from the Central Mediterranean, *Phys. Earth Planet. Inter.*, *85*, 155-171, 1994.

Flanders, P.J., Collection, measurement, and analysis of airborne magnetic particulates from pollution in the environment, *J. Appl. Phys.*, *75*, 5931-5936, 1994.

Georgeaud, V., *Adsorption des métaux lourds sur lex oxydes magnétiques, vers des procédés magnétiques de décontamination et d'expertise des sites pollués*, PhD thesis, Université de Droit, d'Economie et des Sciences d'Aix-Marseille, 1999.

Hoffmann, V., Knab, M., and Appel, E., Magnetic susceptibility mapping of roadside pollution, *J. Geochem. Explor.*, *66*, 313-326, 1999.

Kapička, A., Petrovský, E., Ustjak, S., and Macháčková, K., Proxy mapping of fly-ash pollution of soils around a coal-burning power plant: a case study in the Czech Republic, *J. Geochem. Explor.*, *66*, 291-297, 1999.

Krainer, W., *Bodenschutzbericht 1998, Bodenzustandsinventur der Steiermark*, Landwirtschaftliches Versuchszentrum Steiermark, 1998.

Kruiver, P.P., Kok, Y.S., Dekkers, M.J., Langereis, C.G., and Laj, C., A pseudo-Thellier relative palaeointensity record, and rock magnetic and geochemical parameters in relation to climate during the last 276 kyr in the Azores region, *Geophys. J. Int.*, *136*, 757-770, 1999.

Petrovský, E. and Ellwood, B.B., Magnetic monitoring of air-, land-, and water-pollution. In: Maher, B.A., Thompson, R. (Eds.), *Quaternary Climates, Environments and Magnetism*, Cambridge University press, 1999.

Sammon, J.W., A non-linear mapping for data structure analysis, *IEEE Trans. Comput. C18*, 401-409, 1969.

Schmidt, A.M., von Dobeneck, T., and Bleil, U., Magnetic characterization of Holocene sedimentation in the South Atlantic, *Paleoceanography*, *14*, 465-481, 1999.

Scholger, R., Heavy metal pollution monitoring by magnetic susceptibility measurements applied to sediments of the river Mur (Styria, Austria), *Europ. J. Environ. Engineering Geophys.*, *3*, 25-37, 1998.

Strzyszcz, Z., Magiera, T., and Heller, F., The influence of industrial immissions on the magnetic susceptibility of soils in upper Silesia, *Stud. Geoph. Geod.*, *40*, 276-286, 1996.

Thomspon, R. and Oldfield, F., *Environmental Magnetism*, Allen and Unwin, London, 1986.

Vriend, S.P., van Gaans, P.F.M., Middelburg, J., and de Nijs, A., The application of fuzzy c-means analysis and non-linear mapping to geochemical datasets: examples from Portugal, *Appl. Geochem.*, *3*, 213-224, 1988.