Journal of Geochemical Exploration 71 (2000) 73–82

JOURNAL OF GEOCHEMICAL EXPLORATION

www.elsevier.nl/locate/jgeoexp

# Application of fuzzy clusters to quantify lithological background concentrations in stream-sediment geochemistry

G. Rantitsch*

*Institut für Geowissenschaften, Montanuniversität Leoben, A-8700 Leoben, Austria*

Received 9 May 1999; accepted 15 May 2000

## Abstract

Fuzzy c-means clustering of multivariate stream-sediment geochemistry in a geologically complex area of the Eastern Alps (Austria) allows to separate four dominant associations of lithologies. Weak geochemical contrasts between the involved lithotypes minimize the assignment variances by using a fuzzy coefficient of 1.3. The residual component of the geochemical signal, which is not explained by cluster membership, is interpreted as concentration exceeding the geogene induced background. Therefore, it is possible to specify the background concentration for all lithotypes. Elemental enrichment is visualized by mapping the Geoaccumulation index of Müller [1979. Schwermetalle in den Sedimenten des Rheines-Veränderungen seit 1971. Umschau 79, 778–785]. This technique detects anomalous (maybe contaminated) areas within a geochemical landscape.

*Keywords*: Eastern Alps; stream-sediments; geochemical anomalies; fuzzy logic

## 1. Introduction

The chemical composition of stream-sediments provides information on the lithological composition of the drainage basin and on the presence of contaminants and mineral deposits. However, in geologically complex environments a major problem in geochemical mapping is to distinguish between these sources. This has been mostly done by using samples uneffected by anthropogenic processes (e.g. Singh et al., 1997). Also the use of statistical techniques allowed the separation of anthropogenic from natural anomalies (e.g. Selinus and Esbensen, 1995; Zhang and Selinus, 1998).

In this paper multivariate statistical techniques are applied to separate geochemical anomalies from background concentrations. Using this approach, the complex system, which produces the geochemical signal in the sediment, is transformed to a conceptual model in which variations in the geochemical signature are mainly related to variations in the lithological composition of the catchment area. Uncertainty in the database arises from stochasticity in the conceptual model parameters (Type II uncertainty of Mann, 1993). This type of uncertainty can be modeled by the integration of fuzzy logic with classical multivariate statistical methods. Kramar (1995) has applied this procedure to recognize geochemical anomalies by the application of fuzzy-c-means clustering (FCMC; Bezdek, 1981; Bezdek et al., 1984). FCMC models the mixture of components derived from different sources (bedrock lithology, industrial and anthropogenic contamination, weathering, mineral deposits), which is inherent in the geochemical signature of the sediment.

---

* Tel.: +43-3842-402-795; fax: +43-3842-402-9902.
*E-mail address:* rantit@unileoben.ac.at (G. Rantitsch).

PII: S0375-6742(00)00143-6

Stream sediment data of a geologically complex area within the Eastern Alps are used to determine natural background concentrations by the application of fuzzy clustering as an unsupervised classification technique. The objective of this study is to demonstrate the usefulness of fuzzy clustering to quantify lithological background concentrations in stream-sediment geochemistry. The obtained results are used to investigate a geochemical mapping technique which allows to draw conclusions on the element enrichment relative to the background concentration (Müller, 1979).

## 2. Study area

The study area covers an area of approx. 780 $km^2$ (Fig. 1). The central part of the study area belongs to the Gurktal Nappe Complex (GNC) of the Eastern Alps, involving very low to low grade metamorphic metasediments, volcanics and carbonates of Ordovician to early Carboniferous age, and late Carboniferous very low grade metamorphic shales and conglomerates (Königstuhl Formation). Medium grade metamorphic rocks (gneisses, amphibolites, micaschists, marbles) of the Austroalpine Crystalline Complex underlying the GNC are exposed in the western and southern part of the study area (see Pistotnik, 1980; Neubauer and Sassi, 1993; Schönlaub and Heinisch, 1993 for further information).

## 3. Data base

Multielement data (35 elements) of the −80 mesh (0.18 mm) fraction of 738 samples of stream sediments reported in the “Geochemical Atlas of the Republic of Austria” (Thalmann et al., 1989) have been used in this study. They were analyzed by optical spectrometry and induced coupled plasma emission, wavelength-dispersive X-ray fluorescence spectrometry, and atomic absorption spectrometry. Arsenic was analyzed by the semi-quantitative method of Gutzeit (Thalmann et al., 1989). The element specific analytical procedures are given in the latter paper.

In a topographic map of the study area (at a 1:25 000 scale), the hydrological catchment area was digitized manually for each sample location. This area is taken as the zone of influence of the sample. Samples are organized in a hierarchic structure, where the catchment area of samples that are downstream include the catchment area of sample locations that are upstream. A hand-digitized geological map of the western part of the study area (Fig. 1; Pistotnik, 1996) is used to evaluate the spatial association between bedrock lithology and stream-sediment geochemistry. The formations of this map were reclassified into 14 lithotypes. Both the map of the catchment areas and the map of the bedrock lithology were transformed to a raster map with a spatial resolution of 30m × 30m.

## 4. Data processing

Traditional classification techniques assign each sample to one specific (“hard”) cluster, whereas in FCMC a sample is allowed to belong to more than one cluster. The assignment probability of a sample to a specific cluster is described in FCMC by a membership value ($u$) ranging between 0 and 1. The membership values of each sample sum to 1 (Bezdek, 1981; Bezdek et al., 1984).

The algorithm used in this paper is described in Bezdek (1981), Bezdek et al. (1984) and Kramar (1995). The log normalized data set is partitioned by means of c-means clustering into a predetermined number of hard clusters. From the obtained membership values new cluster centroids are iteratively calculated until the cluster centroids remain stable. In this calculation the degree of fuzziness is controlled by the fuzzy coefficient m. A fuzzy coefficient of $m = 1$ results in hard clustering and, with higher fuzzy coefficient ($m \rightarrow \infty$), the samples are assigned equally to all clusters. To achieve a better cluster separation only samples with a assignment probability above 20% were used to calculate the new cluster centroids (Kramar, 1995). The assignment variance and the assignment error can be calculated using the original data, the membership values and the cluster centroids. In former studies it was demonstrated that a fuzzy coefficient between 1.3 and 3 minimizes the assignment variance (Vriend et al., 1988; Kramar, 1995).

To avoid a systematic bias, samples of a catchment area which include one or more samples that are upstream from the sampling site are not included in FCMC. 420 samples remain after this data reduction.

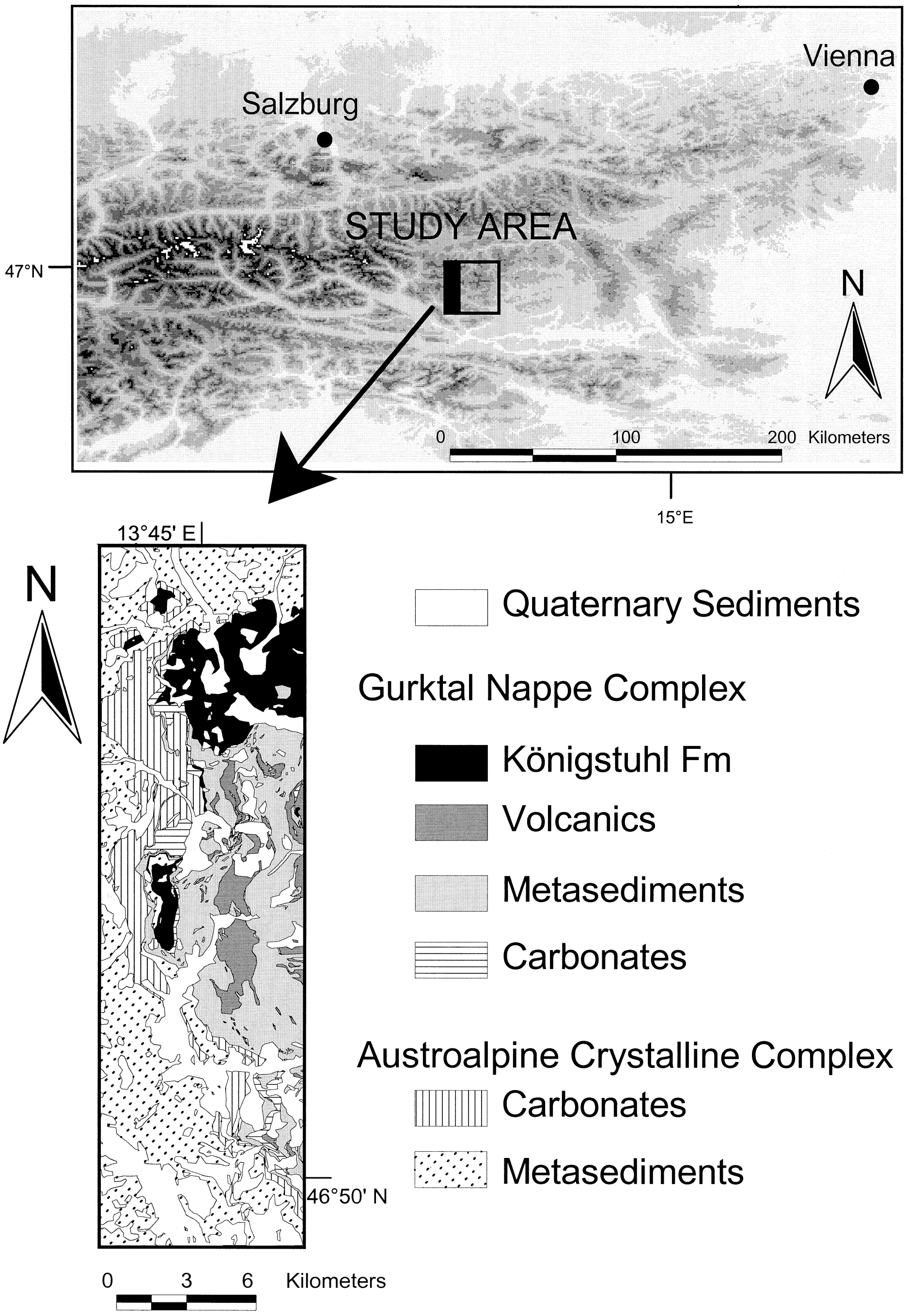

Fig. 1. Location of the study area within the Eastern Alps. In a geological map of the western part of the study area (western margin of the Austroalpine Gurktal Nappe Complex) both, the bedrock lithology (redrawn from Pistotnik, 1996) and stream-sediment geochemistry is known. The lithological variation within this subset is representative for the entire study area.

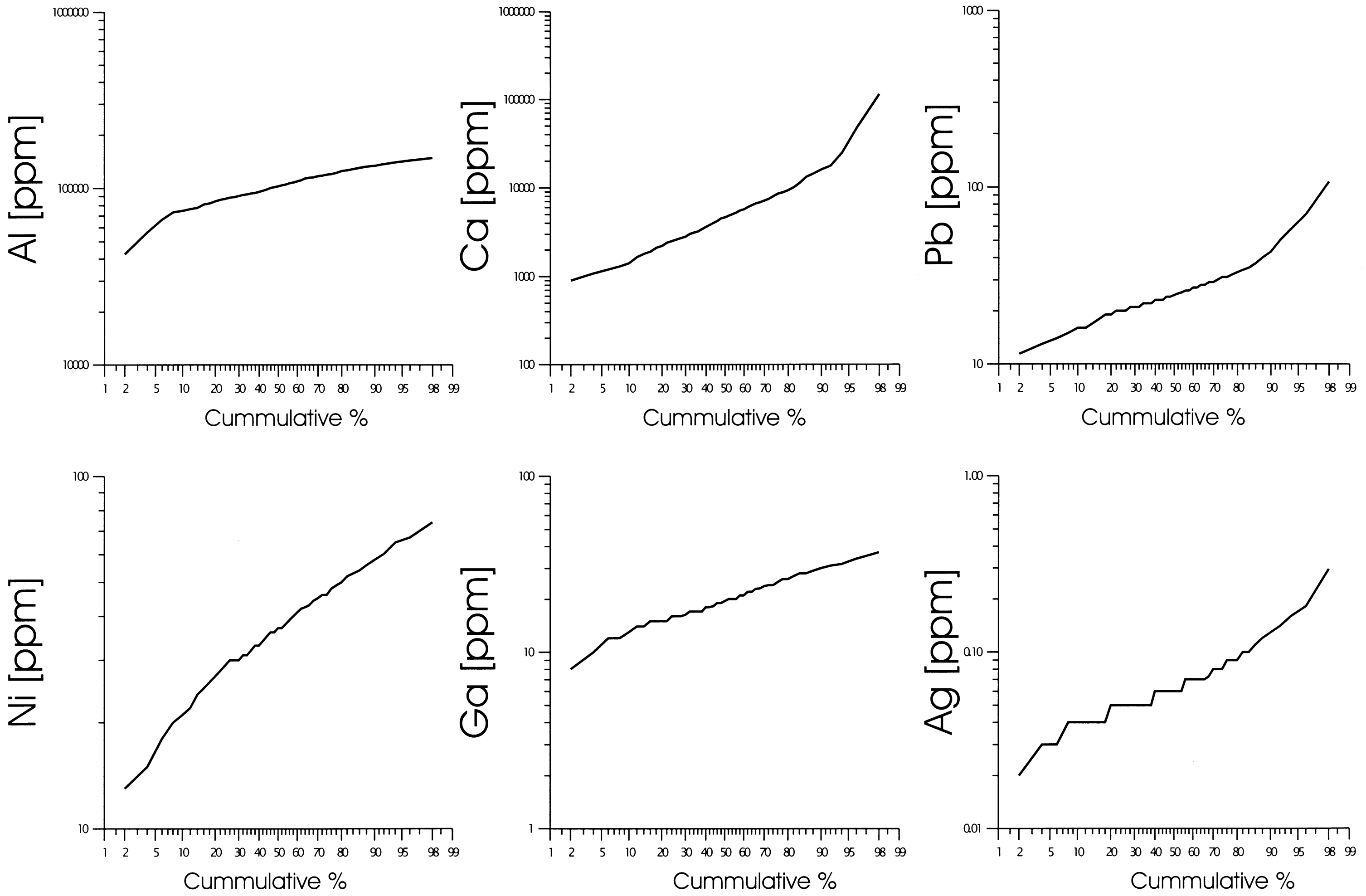

Fig. 2. Cumulative probability plots of representative elements within the study area (Fig. 1).

From a spatial overlay of the lithological map with the membership maps of the distinct clusters, the spatial association between cluster membership and lithology is examined using an area cross-tabulation. Based on this tabulation the cluster centroids are interpreted as indicators of a distinct lithotype.

The FCMC algorithm calculates, for each element and sample, a residual which is not explained by cluster membership. Subtracting this value from the measured elemental concentration, the lithologically induced background concentration is defined. The residual proportion of the measured concentration is interpreted as the result of contamination by artificial sources (possible anthropogenic sources), presence of mineral deposits and/or analytical errors.

The anomalies can be expressed by using the "Geoaccumulation index" ($I_{\mathrm{geo}}$) proposed by Müller (1979) for the quantification of heavy metal accumulation in sediments. $I_{\mathrm{geo}}$ describes the relationship between the measured elemental concentration ($C_n$) and the background concentration ($B_n$) as

$$I_{\mathrm{geo}} = \log_2(C_n/1.5B_n).$$

The calculated indices are classified into seven enrichment classes ranging from class 0 (unpolluted) to class 6 (highly polluted). The advantage of this index is the explicit consideration of the geochemical background concentration when a measured concentration is evaluated with respect to a contamination. It focuses the view towards potential contamination, not towards a lithologically induced concentration maximum.

## 5. Results

As reflected by the straight lines of cumulative probability plots (see Fig. 2), log transformation of the data results in near normal distributions of elemental concentrations which are representative for the lithological variation within the study area. Ga is an indicator for the presence of metasediments within the catchment area, and Ni is related to volcanics (Schroll in Thalmann et al., 1989). A bimodal lognormal distribution is seen in some data distributions (e.g. Al, Ca, Pb, see Fig. 2). These elements characterize catchments which are dominated by carbonates as bedrock lithology. Using the transformed data, the data set is meaningfully classified into four clusters. Their cluster centroids are shown in Fig. 3. The assignment variances are minimized with a fuzzy coefficient of 1.3 (Fig. 4).

Harff and Davis (1990) proposed regionalized classification to make probabilistic statements about a specific discriminant function, using geostatistical techniques. Geostatistical interpolation assumes a homogeneity of the underlying random process. Because of the heterogeneous lithological composition within the catchment area of the samples, this assumption is not appropriate here. Therefore, the membership values of the samples are assigned to the corresponding catchment area (Fig. 5).

FCMC extracts four lithotypes dominating the lithologically induced signal in the stream sediment. They are characterized by metasediments and volcanics in cluster 1, by metasediments and metamorphic rocks in cluster 2, by the Königstuhl Formation in cluster 3 and by dolomite in cluster 4 (Table 1). Quaternary sediments overlying the Paleozoic rocks are exposed in the entire study area, therefore, they contribute significantly to all clusters. The group of clastic metasediments contributes to two clusters (cluster 1 and 2). In cluster 1 there is more influence of volcanic material and in cluster 2 there is more influence of metamorphic rocks (Table 1).

It was demonstrated by Kramar (1995) that elemental residuals can be interpreted as a measure of contamination. This is also demonstrated by the data of this paper: if the cluster membership values of Fig. 5 are compared with the lithological units of Fig. 1, a clear spatial correlation between cluster assignment and lithology is observable (see also Table 1). Therefore, a specific lithological unit is predictable from cluster membership and the residual is the proportion of the geochemical signal which does not reflect the exact lithological composition of the catchment. The nature of this residual (contamination, misclassification and/or analytical error) can be examined by a careful investigation of their spatial structure and their spatial correlation with known locations of contaminations (see below).

The results of this approach are presented by two ways. Firstly, it is possible to specify "global" background concentrations without more or less arbitrary a priori assumptions. Secondly, mapping of

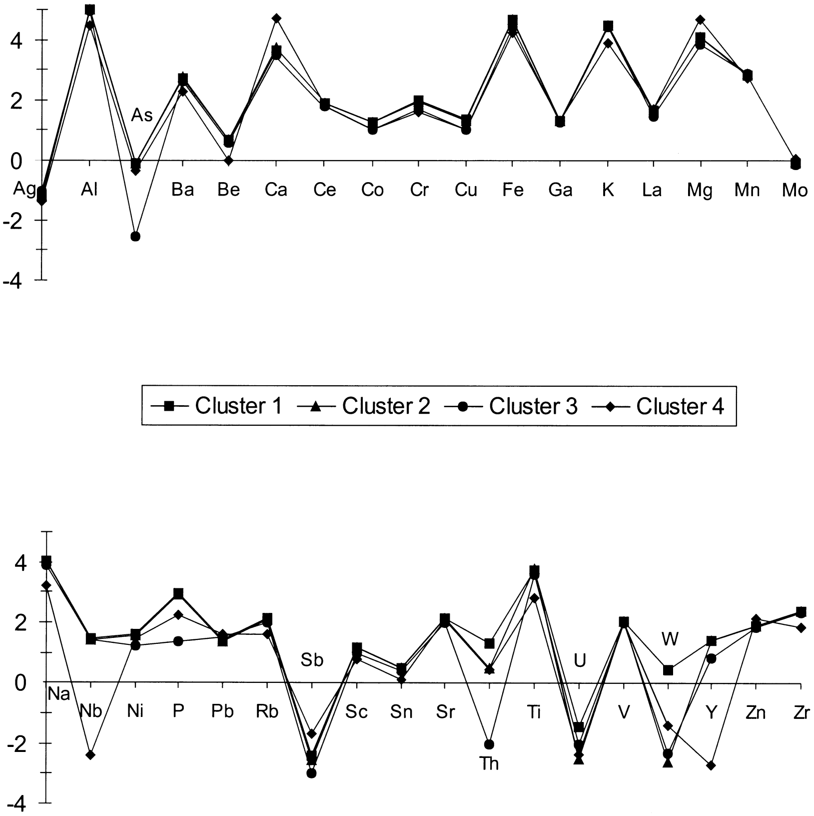

Fig. 3. Normalized element pattern of the four cluster centroids applying fuzzy c-means clustering with a fuzzy coefficient of 1.3.

the clustering results gives a view of the geochemical landscape which is focused towards anomalous areas:

- Interpreting the elemental residuals in the context of contamination, samples with a minimum cluster membership for the four clusters of 0.75 (assignment to a specific lithotype-cluster with a minimum probability of 75%) were used to calculate mean (lithologically induced) background concentrations representative for the different lithotypes (Table 2). Because of their very similar lithological composition (see above), samples of cluster 1 and 2 are

Table 1
Lithological composition (in areal percentages) of the extracted clusters derived from a cross tabulation between the lithotypes and the areas of the dominating cluster membership within the western part of the study area (see Fig. 1) (1 = Quaternary sediments, 2 = limestone, 3 = dolomite, 4 = shales, 5 = red beds, 6 = Königstuhl Formation, 7 = orthogneiss, 8 = marble, 9 = volcanics, 10 = iron-rich dolomite, 11 = phyllite, 12 = micaschist, 13 = metavolcanics, 14 = paragneiss)

| Cluster | Lithotype | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 1 | 22.8 | 1.3 | 7.1 | 0.0 | 0.0 | 6.2 | 2.2 | 2.0 | 9.9 | 2.2 | 31.4 | 5.1 | 0.1 | 6.0 |
| 2 | 31.8 | 0.2 | 6.5 | 0.0 | 0.0 | 5.2 | 2.8 | 0.5 | 4.4 | 0.8 | 21.5 | 12.9 | 0.3 | 9.3 |
| 3 | 26.8 | 0.0 | 2.2 | 0.0 | 0.0 | 44.3 | 2.5 | 0.0 | 0.0 | 0.0 | 0.1 | 10.3 | 0.0 | 12.2 |
| 4 | 16.3 | 10.8 | 58.7 | 0.4 | 0.5 | 0.8 | 2.2 | 0.2 | 0.0 | 0.0 | 3.2 | 2.2 | 0.0 | 1.2 |

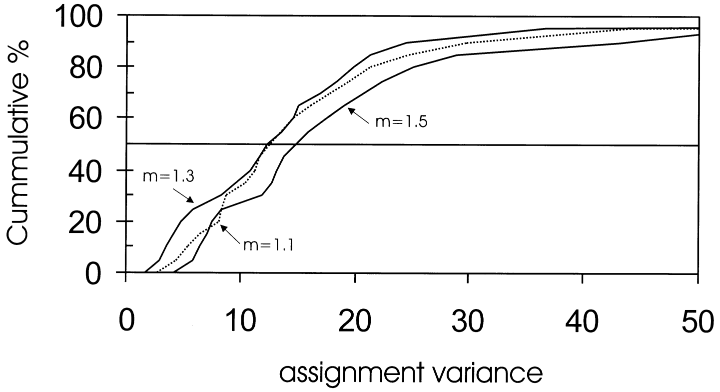

Fig. 4. Cumulative frequency of assignment variances for four clusters and fuzzy coefficients *m* between 1.1 and 1.5. A minimum in the assignment variances is achieved using a fuzzy coefficient of 1.3.

included in one common lithotype. The chosen threshold value of 0.75 is an optimum obtained by an evaluation of several experiments. A lower value dilutes the differences between the lithotypes, and a higher value results in an insufficient number of samples.

- Fig. 6 shows three examples for mapping of elemental enrichment (As, Pb, W) by using the Geoaccumulation index of Müller (1979). In these maps the location of known base metal mineralizations are indicated by point symbols. With these maps it is demonstrated that mineralizations does not influence the univariate geochemical landscape (Fig. 6).

## 6. Conclusions

Fuzzy c-means-clustering (FCMC) of geochemical

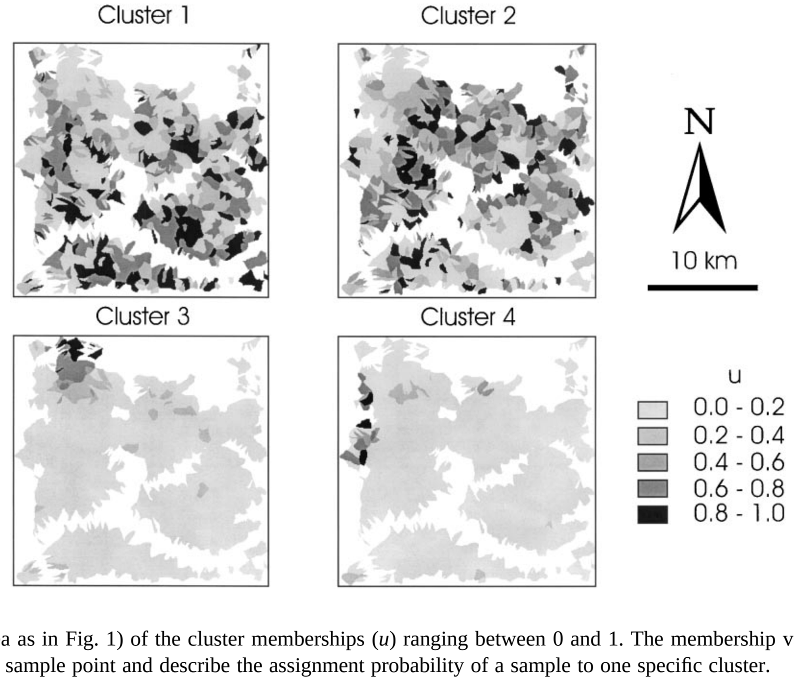

Fig. 5. Maps (same area as in Fig. 1) of the cluster memberships (*u*) ranging between 0 and 1. The membership values are assigned to the catchment area of each sample point and describe the assignment probability of a sample to one specific cluster.

Table 2
Lithologically induced background concentrations (ppm) derived from fuzzy c-means clustering at the western margin of the Austroalpine Gurktal Nappe Complex (*n* = number of observations, *m* = mean value, *s* = standard deviation)

| | Cluster 1 and 2 | | | Cluster 3 | | | Cluster 4 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Metasediments | | | Königstuhl Fm | | | Dolomite | | |
| | *n* | *m* | *s* | *n* | *m* | *s* | *n* | *m* | *s* |
| Ag | 235 | 0.07 | 0.04 | 13 | 0.09 | 0.00 | 9 | 0.04 | 0.00 |
| Al | | 106220 | 20870 | | 85155 | 2065 | | 38325 | 2556 |
| As | | 9 | 12 | | < 2 | | | < 2 | |
| Ba | | 569 | 161 | | 434 | 8 | | 229 | 13 |
| Be | | 5 | 1 | | 4 | 0 | | 1 | 0 |
| Ca | | 7640 | 10030 | | 3622 | 421 | | 31773 | 4643 |
| Ce | | 84 | 22 | | 61 | 1 | | 63 | 1 |
| Co | | 20 | 7 | | 12 | 0 | | 12 | 0 |
| Cr | | 103 | 32 | | 55 | 2 | | 46 | 2 |
| Cu | | 23 | 7 | | 11 | 1 | | 12 | 0 |
| Fe | | 47157 | 10642 | | 28426 | 716 | | 20340 | 949 |
| Ga | | 21 | 6 | | 18 | 0 | | 21 | 0 |
| K | | 30366 | 6998 | | 25374 | 908 | | 9494 | 709 |
| La | | 45 | 15 | | 31 | 1 | | 46 | 1 |
| Mg | | 13997 | 6701 | | 8263 | 757 | | 35809 | 3065 |
| Mn | | 740 | 299 | | 692 | 6 | | 551 | 9 |
| Mo | | 0.95 | 0.53 | | 0.76 | 0.02 | | 1.14 | 0.02 |
| Na | | 13056 | 4939 | | 7645 | 252 | | 2393 | 239 |
| Nb | | 33 | 12 | | 19 | 4 | | < 5 | |
| Ni | | 42 | 12 | | 20 | 1 | | 30 | 0 |
| P | | 1256 | 422 | | 44 | 13 | | 214 | 11 |
| Pb | | 29 | 24 | | 32 | 0 | | 36 | 1 |
| Rb | | 147 | 42 | | 99 | 2 | | 53 | 3 |
| Sb | | < 2 | | | < 2 | | | < 2 | |
| Sc | | 16 | 5 | | 10 | 0 | | 7 | 0 |
| Sn | | 4 | 2 | | 3 | 0 | | 2 | 0 |
| Sr | | 149 | 43 | | 104 | 3 | | 122 | 0 |
| Th | | 30 | 10 | | < 10 | | | < 10 | |
| Ti | | 6797 | 3320 | | 3751 | 138 | | 979 | 115 |
| U | | < 5 | | | < 5 | | | < 5 | |
| V | | 119 | 32 | | 106 | 1 | | 106 | 0 |
| W | | 4 | 6 | | 0.01 | 0.00 | | 0.04 | 0.01 |
| Y | | 28 | 6 | | < 10 | 1 | | < 10 | |
| Zn | | 87 | 46 | | 75 | 2 | | 127 | 4 |
| Zr | | 261 | 53 | | 220 | 7 | | 92 | 6 |

data from stream sediments models stochastic parameters, which are inherent in a simplified model of elemental accumulation due to erosion of bedrock lithology.

FCMC extracts lithotypes which dominate within a specific study area. In the study area these lithotypes are composed either of one distinct lithology (i.e. dolomite in cluster 4) or of a composite lithology (i.e. diverse metasediments in cluster 1 and 2). In addition, a cluster is extracted (cluster 4) which is related to the northern outcropping area of the Königstuhl Formation (compare Fig. 1 with Fig. 5). Because a significant membership to cluster 3 is not outlined in the southern outcropping area, a difference in the geochemical composition between these areas is seen by the application of FCMC. The residual component of the geochemical signal, which is not explained by cluster membership, is interpreted as

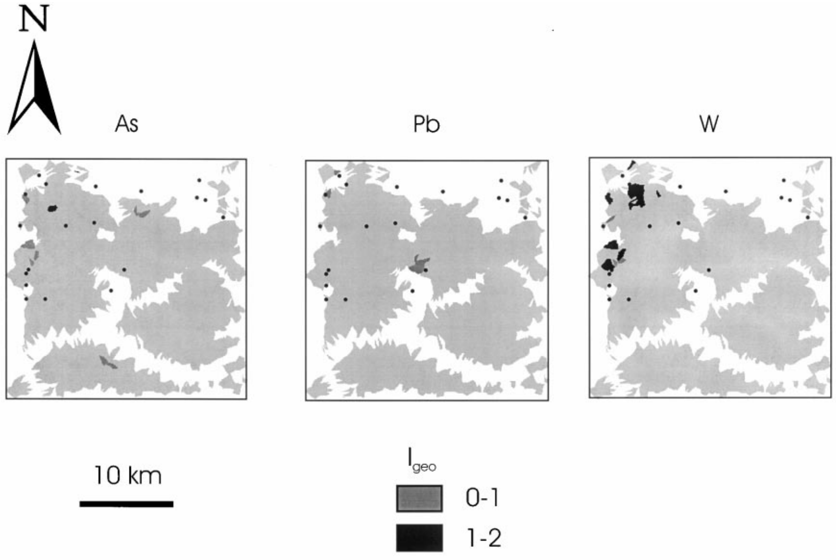

Fig. 6. Maps (same area as in Fig. 1) of the Geoaccumulation index ($I_{geo}$) for some elements (As, Pb, W). The locations of base metal mineralizations (Weber, 1997) are indicated by point symbols.

concentration exceeding the background. Consequently, it is possible to specify the background matrix factor for all dominant lithotypes. Based on these specific background concentrations, a mapping technique is proposed which visualizes the enrichment of elements in relation to the geogene induced concentration within each distinct catchment area. Following this approach it is obvious that there is no spatial association between geochemical anomalies and the presence of base metal mineralizations within the study area. Because of the small size of the mineralized occurrences (Weber, 1997), the metal content of mineralized rocks is possibly diluted during transport and mixing of the sediment away from the source and the effects of these mineralizations are indistinguishable from the observed natural background concentrations. Consequently, it is supposed that anthropogenic contamination and/or analytically errors are responsible for the observed geochemical anomalies.

In geochemical exploration and environmental research, fuzzy c-means-clustering provides an objective and effective way to specify lithologically induced background concentrations in the geochemical composition of stream sediments. By using this technique, a misinterpretation of anomalous patches, which are only related to the lithological composition within the drainage area of the samples, is avoided. The nature of the anomalies (industrial and anthropogenic contamination, presence of mineralizations, analytical error) can be explained by mapping of the Geoaccumulation index of Müller (1979).

## Acknowledgements

The Geological Survey of Austria and L. Weber are acknowledged for providing the database. The author wants to thank E. Grunsky, C. Mann and R. Swennen for a critical reading of the manuscript.

## References

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York (pp. 1–256).

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

Harff, J., Davis, J.C., 1990. Regionalization in geology by multivariate classification. Math. Geol. 22, 573–588.

Kramar, U., 1995. Application of limited fuzzy clusters to anomaly recognition in complex geological environments. J. Geochem. Explor. 55, 81–92.

Mann, C.J., 1993. Uncertainty in Geology. In: Davis, J.C., Herzfield, U.C. (Eds.), Computers in geology—25 years of progress. Int. Assoc. Math. Geol. Studies in Math. Geol., 5, 241–254.

Müller, G., 1979. Schwermetalle in den Sedimenten des Rheines-Veränderungen seit 1971. Umschau 79, 778–785.

Neubauer, F., Sassi, F.P., 1993. The Austro-Alpine quartzphyllites and related Paleozoic formations. In: Raumer, J.F., Neubauer, F. (Eds.). Pre-Mesozoic geology in the Alps, Springer, Wien, pp. 423–439.

Pistotnik, J., 1980. Die westlichen Gurktaler Alpen (Nockgebiet). In: Oberhauser, R. (Ed.). Der geologische Aufbau Österreichs, Springer, Wien, pp. 358–363.

Pistotnik, J. (Ed.), 1996. Geologische Karte der Republik Österreich. Blatt 183 Radenthein. Geologische Bundesanstalt, Wien.

Schönlaub, H.P., Heinisch, H., 1993. The classic fossiliferous units of the Eastern and Southern Alps. In: Raumer, J.F., Neubauer, F. (Eds.). Pre-Mesozoic geology in the Alps, Springer, Wien, pp. 395–422.

Selinus, O.S., Esbensen, K., 1995. Seperating anthropogenic from natural anomalies in environmental geochemistry. J. Geochem. Explor. 55, 55–66.

Singh, M., Ansari, A.A., Müller, G., Singh, I.B., 1997. Heavy metals in freshly deposited sediments of the Gomati River (a tributary to the Ganga River): effects of human activities. Environ. Geol. 29, 246–252.

Thalmann, F., Schermann, O., Schroll, E., Hausberger, G., 1989. Geochemischer Atlas der Republik Österreich 1: 1,000.000. Textteil. Geologische Bundesanstalt, Wien.

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

Weber, L. (Ed.), 1997. Handbuch der Lagerstätten der Erze, Industrieminerale und Energierohstoffe Österreichs. Arch. f. Lagerst. forsch. Geol. B.-A., 19, 607.

Zhang, Ch., Selinus, O., 1998. Statistics and GIS in environmental geochemistry—some problems and solutions. J. Geochem. Explor. 64, 339–354.