SIMILARITY ANALYSIS USING RANK IN TILL GEOCHEMISTRY

JARI MÄKINEN

MÄKINEN, JARI, 1991: Similarity analysis using rank in till geochemistry. *Bull. Geol. Soc. Finland 63, Part 1*, 49—57

A similarity analysis method based on Spearman's rank correlation coefficient (ρ) was developed aiming to indirectly measure the similarity between the geochemistry of till and bedrock. The study area lies in Central Finland, comprising c. 4800 km², and it includes Pyhäsalmi Säviä zone characterized by copper-zinc-mineralizations. The data consists of 1099 till samples with 11 variables; Al, Ba, Co, Cr, Cu, Li, Mn, Sc, Ti, V and Zn. Subset of 83 cases, representing Kiuruvesi schist belt area, was selected from the till data for the model. The model was formed from the median values of the subset. Model based on the till data consists of information on glacial and bedrock geology, and usage of different similarity measures consequently emphasize one or the other geological factors in the till. The till data covering the whole study area was compared to the model with an aid of similarity analysis, using the Euclidean distance, the cosine theta and the rank correlation coefficient as a measure.

The results indicate that Euclidean distance reflects mainly information connected to the glacial geology and similarity pattern produced by cosine theta or rank correlation coefficient correlates very well with major bedrock structures. Similarity analysis of the till data based on rank reflected the bedrock structures better than using cosine theta as a measure. The rank correlation coefficient is nonparametric and consequently less insensitive to anomalous element concentrations than the common parametric cosine theta measure. Hence the rank correlation is readily applicable to geochemical data, which invariably include anomalous values.

Key words: geochemistry, till, bedrock, statistical analysis, correlation coefficient, Central Finland, Finland.

Jari Mäkinen: Geological Survey of Finland, Pb Box 1237, SF-70701 Kuopio, Finland.

Introduction

Till, being produced from bedrock by glacial processes, carries information about bedrock and glacial geology. The proportions of factors determining the geochemical composition of till vary for area to area, which makes it difficult to interpret geochemical till mapping data.

Though several factors affect the element concentrations of the fine fraction of till (< 0.064 mm, material in this study), the variations in concentrations have some general properties such as the dilution of high element concentrations as transport distance increases (Salminen & Hartikainen 1985). Because of the dilution effect, statistical methods describing the geochemical characteristics of till should omit the secondary (glacial) information while emphasizing the primary (lithological) information.

The study area is part of the Raahe-Ladoga ore belt and located in Central Finland (Fig. 1). The dominant feature of element concentrations in till
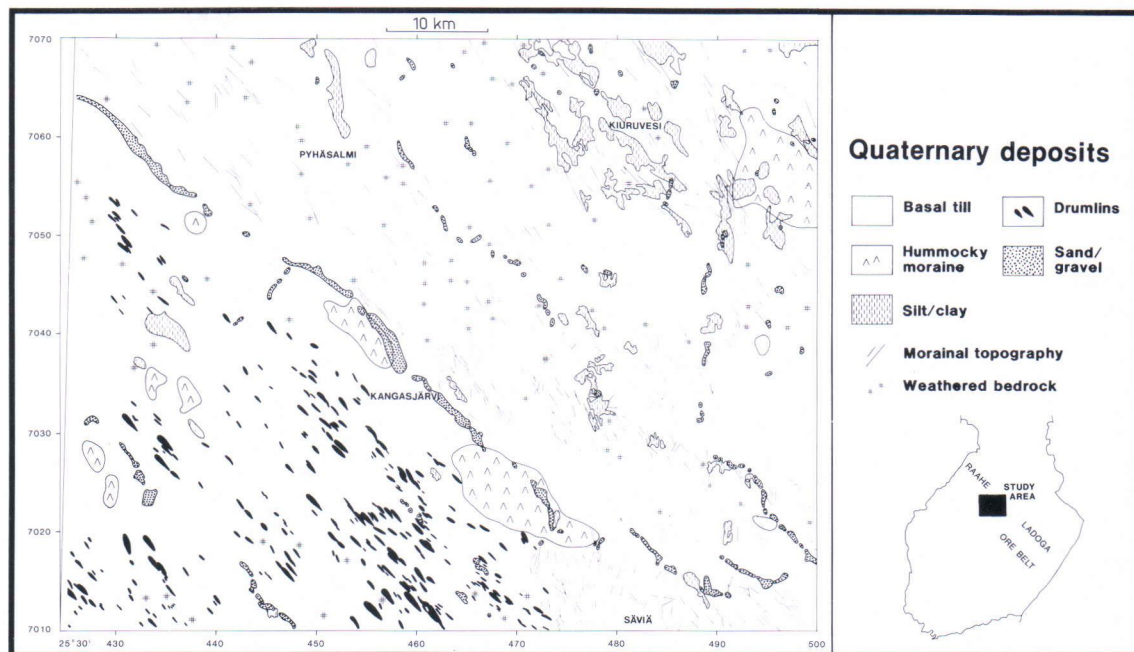
4

Fig. 1. The Quaternary deposits of the study area. Modified after Salonen et al. (1988).

is a sharp areal bimodality conforming with both the ore belt and glacial geology (Figs. 1 and 2). In the NE-part of the area glacial erosion has been weak and drumlins are rare. On the contrary glacial erosion was stronger in the SE-part of the area, glacial transport distance is longer and drumlins are more common as in the NE-part.

The bedrock of the area includes metallogenic provinces characterized by copper-nickel- and copper-zinc-mineralizations (Tiainen 1988). Cordierite-bearing supracrustal rocks including copper-zinc-mineralizations extend from Pyhäsalmi to Säviä. An outline of the structure in the bedrock is presented on the magnetic greytone map (Fig. 7).

Selecting the statistical method

The aim of the study was to measure the similarity between geochemical till data and the model, and to describe the areal similarity distribution of till samples. The model was formed from subset of the till data and depending on the similarity measure — Euclidean distance, cosine theta or Spearman's rank correlation coefficient (ρ) etc. — the results of the similarity analysis reflect the glacial geology and lithological factors to different degrees.

When the multivariate similarity between the cases and the model is measured with Euclidean distance, cases within the same multielement concentration level assume the greatest similarity values; for example, similarity is high between cases rich in zinc but decreases as the difference between the zinc contents increases (cf. Fig. 2). Therefore the Euclidean distance reflects changes of levels but not ratios.

The formula of Euclidean distance d for cases X and Y is;

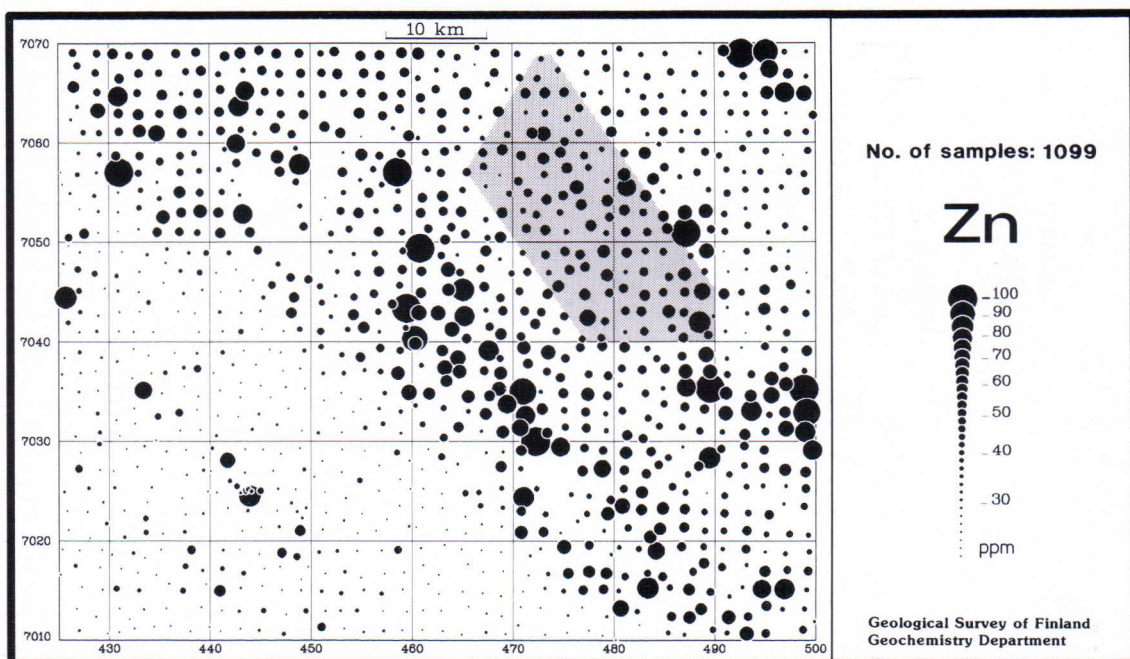$$d = \left[\sum_{i=1}^{N} (Y_i - X_i)^p \right]^{1/p} \tag{1}$$

Fig. 2. Zinc concentration of fine fraction of till in study area. Hatched area shows the target of the samples selected for the model in Kiuruvesi schist belt. Data from Geological Survey of Finland Geochemistry Department.

where N is number of variables (i) and p is constant.

Variation of multielement concentration level of till depends both on the bedrock lithology and glacial dilution. The basic assumption in this study, concerning with geochemical properties of the till, is that glacial dilution generally has more influence on the multielement concentration level than on the multielement ratios. The assumption that the dilution for most of the cases is similar for several elements is a considerable simplification, yet it is valid as will be seen later. On account of it, the Euclidean distance is not an appropriate measuring unit for the similarity analysis if the aim is to reflect lithology of bedrock by means of till samples.

In order to eliminate the variation caused by glacial dilution we can examine the geochemical characteristics of the cases by means of ratios (cf. Salminen 1980). The ratios of the element concentrations are independent of the proportional variation in concentration levels, and thus it is possible to compare the geochemical characteristics of the different background levels.

A similarity measure comparable to the ratio of the element concentrations is the cosine of the angle between the case vectors (Le Maitre 1982, Tontti et al. 1981).

The formula of the cosine theta is;

$$\cos \theta = \sum_{i=1}^{N} X_i Y_i \Big/ \left[\sum_{i=1}^{N} X_i^2 \sum_{i=1}^{N} Y_i^2 \right]^{1/2} \qquad (2)$$

where N is number of variables (i) for cases X and Y.

However cosine theta is sensitive to anomalous values, because squares of the values in the denominator dominate the value of the quotient. The examination of anomalous cases is important in the geochemical study, and thus the cases with exceptional element concentrations should not be removed from the data. Preferably a statistical method should be selected that does not
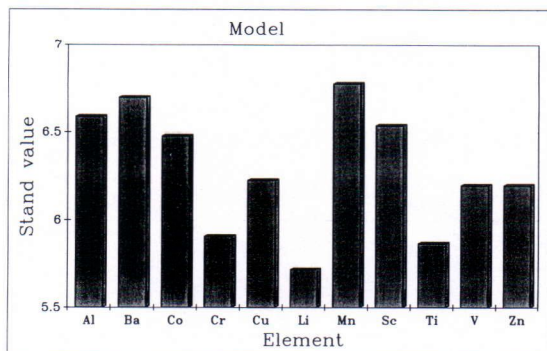
Fig. 3a. Median values of the element concentrations for the model computed from the till samples from the Kiuruvesi area. Values are computed from the standardized data.
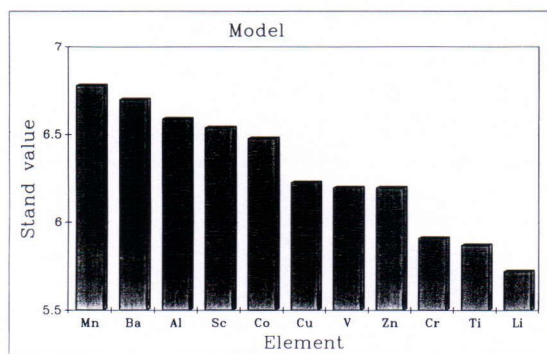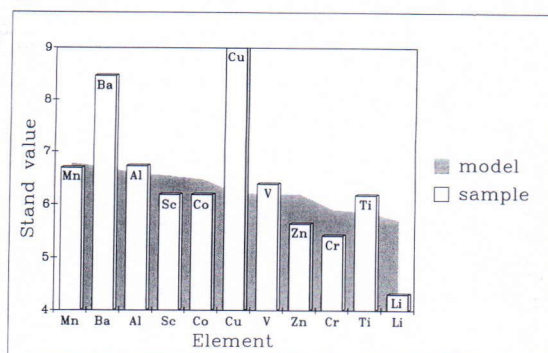

Fig. 4a. Element spectrum of the model is marked as a grey background and the element spectrum of the case as columns.
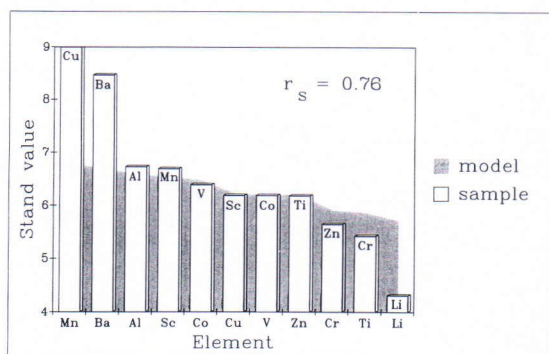

Fig. 3b. Ranked element spectrum of the model.

Fig. 4b. Ranked element spectrum of the model. Spearman's rank correlation coefficient between the model and the case is 0.76.

cause anomalous cases to distort the results, but presents them in proper relationship to the background data.

As the aim was to measure similarities with element ratios, the Spearman's rank correlation coefficient (ρ) was selected as a measuring unit. For the ordinal scale measurement, ρ is less sensitive to anomalous element concentrations than cosine theta or the product moment correlation coefficient. Statistical methods using ordinal level measure are more robust than methods using interval or ratio level measure, because a ordinal level measure reflects the order of the values but not the magnitude of the values. For example increasing copper-content for sample in Fig. 4b will

not change the order of elements and result of the ordinal level statistics remain constant.

The formula of Spearman's rank correlation coefficient is

$$\rho = 1 - \frac{6 \times \sum\limits_{i=1}^{N} D_i^2}{N(N^2-1)} \tag{3}$$

where N is number of cases (i) and D_i = (rank of X_i — rank of Y_i). X and Y are variables.

The similarity analysis yields different results, depending on the measurement formula, and thus the similarity pattern will differ from case to case. Inferences based on geological knowledge of the area were made on the validity of the similarity maps produced by different methods.

Characteristics of the data and execution of the procedure

The geochemical data set covered the 1:100 000 map sheets of Pyhäjärvi (3321), Pihtipudas (3312), Kiuruvesi (3323) and Pielavesi (3314). The study area is thus the same as in the MM-project in Geological Survey of Finland (Tenhola 1988). The original data set consists of the concentrations of 25 elements in the fine fraction of till (< 0.064 mm) analysed with the ICP method after aqua regia leach. For statistical data processing, the 11 elements were selected: Al, Ba, Co, Cr, Cu, Li, Mn, Sc, Ti, V and Zn. The study area sample grid was one case/4 km².

In the study area the most striking geochemical feature of till is the sharp areal bimodality of the element concentrations (cf. Fig. 2 and Tenhola 1988). The boundary between the two different areas conforms with differences in glacial geology (Figs. 1 and 2). According to preliminary investigations the proportion of clay-size material in till is higher in the northeastern than in the southwestern part of the area (M-L. Räisänen *pers. comm.*). Further the geochemical boundary conforms with the Raahe-Ladoga orebelt.

The first stage in data processing was to standardize the variables. The value 5 was arbitrarily selected as the median. Standardization was necessary in order to calculate ρ, because standardized values are comparable among themselves. Besides, the calculation of the cosine theta values requires non-negative values.

In similarity analysis, cases of the data are compared with one or several models, which can be defined from external material or from data under consideration. In this case, one model was selected from the data available.

For the model, 83 cases were selected from the till data lying upon Kiuruvesi schist belt, and the number of cases was considered adequate to represent the till in the model area. Hence, the data for the model represents geochemical characteristics of the Kiuruvesi schist belt shown by till samples. The model was based on the median values of selected variables (Fig. 3a), and similarities between the model and each cases were calculated (Fig. 5). If the method omits all the glacial information, the similarity values of the cases should reflect the similarity of the bedrock to the Kiuruvesi schist belt underlying the till sampled.

The procedure for calculating the ρ is presented graphically in Figs. 3—4. First the element spectrum of the model is ranked (Figs. 3a—3b) and the rank of the element spectrum of a case is compared with the model by ρ (Figs. 4a—4b). The ρ values are then plotted as symbols on the map (Fig. 5). The smoothed and interpolated similarity values are presented as a coloured map together with indications of copper-zinc mineralizations (Fig. 6).

A similarity analysis based on the Kiuruvesi model was also performed using cosine theta as a measurement unit. The similarity pattern is expressed as symbols in Fig. 8.

Results and discussion

The difference between the patterns of similarity and zinc values in the till samples is striking (Figs. 2 and 6). The bimodal areal distribution reflected by the zinc concentrations is a basic feature of till as almost all the elements have a similar pattern (cf. Tenhola 1988). The cause of the bimodality is not unambiquous because it conforms areally both with glacial geology and the Raahe-Ladoga ore belt.

The bimodal areal distribution of the element concentrations in till does not seem to conform with bedrock structures reflected by magnetic greytone map (cf. Figs. 2 and 7). In contrast, the compatibility between the similarity and geophysical greytone patterns is much better (Figs. 6 and 7). One possibility is that glacial factors and/or the ore belt accounts for the variation in the element concentration levels and the local bedrock lithology for the element ratios. Comparison of Figs. 2 and 6 supports the assumption
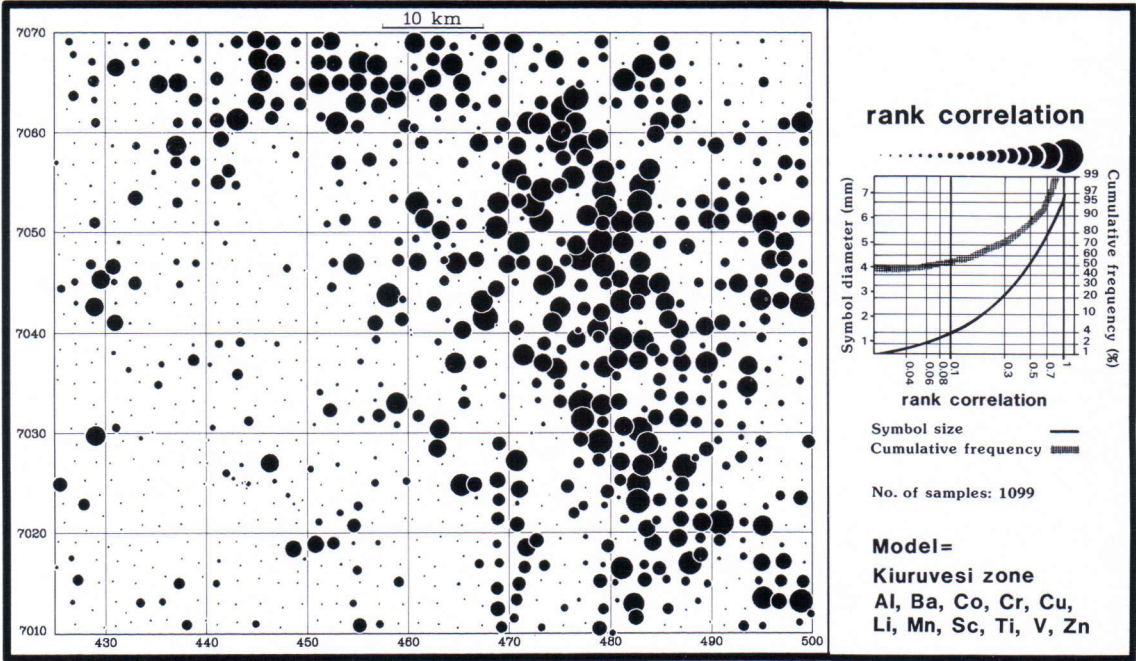
Fig. 5. Similarity values between cases and the model using Spearman's rank correlation coefficient.
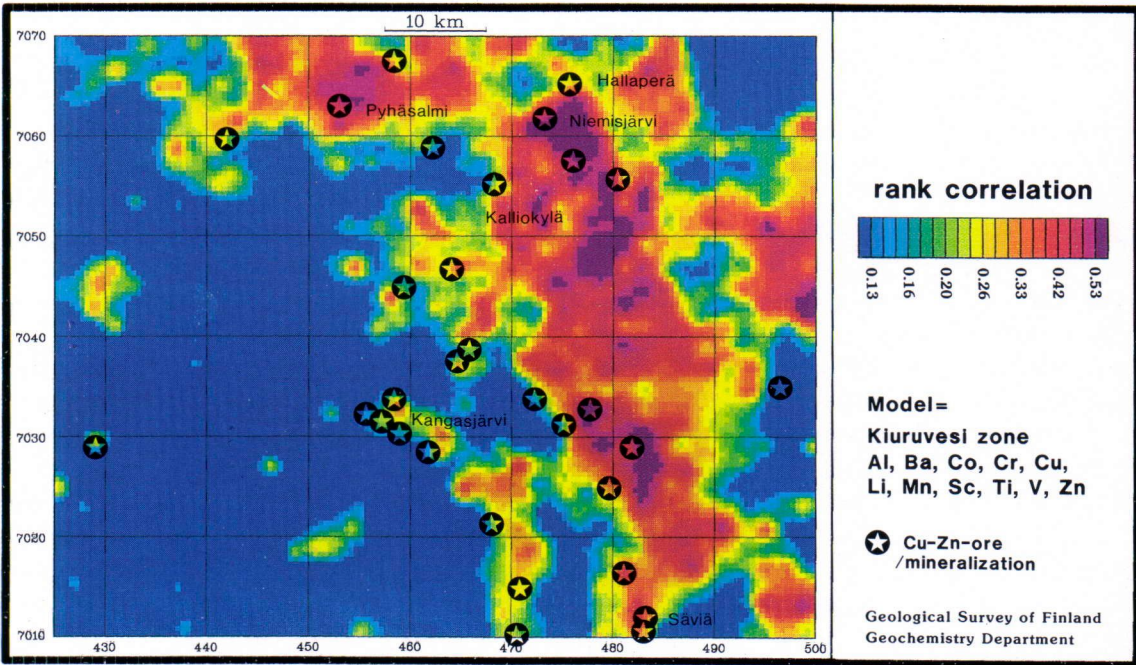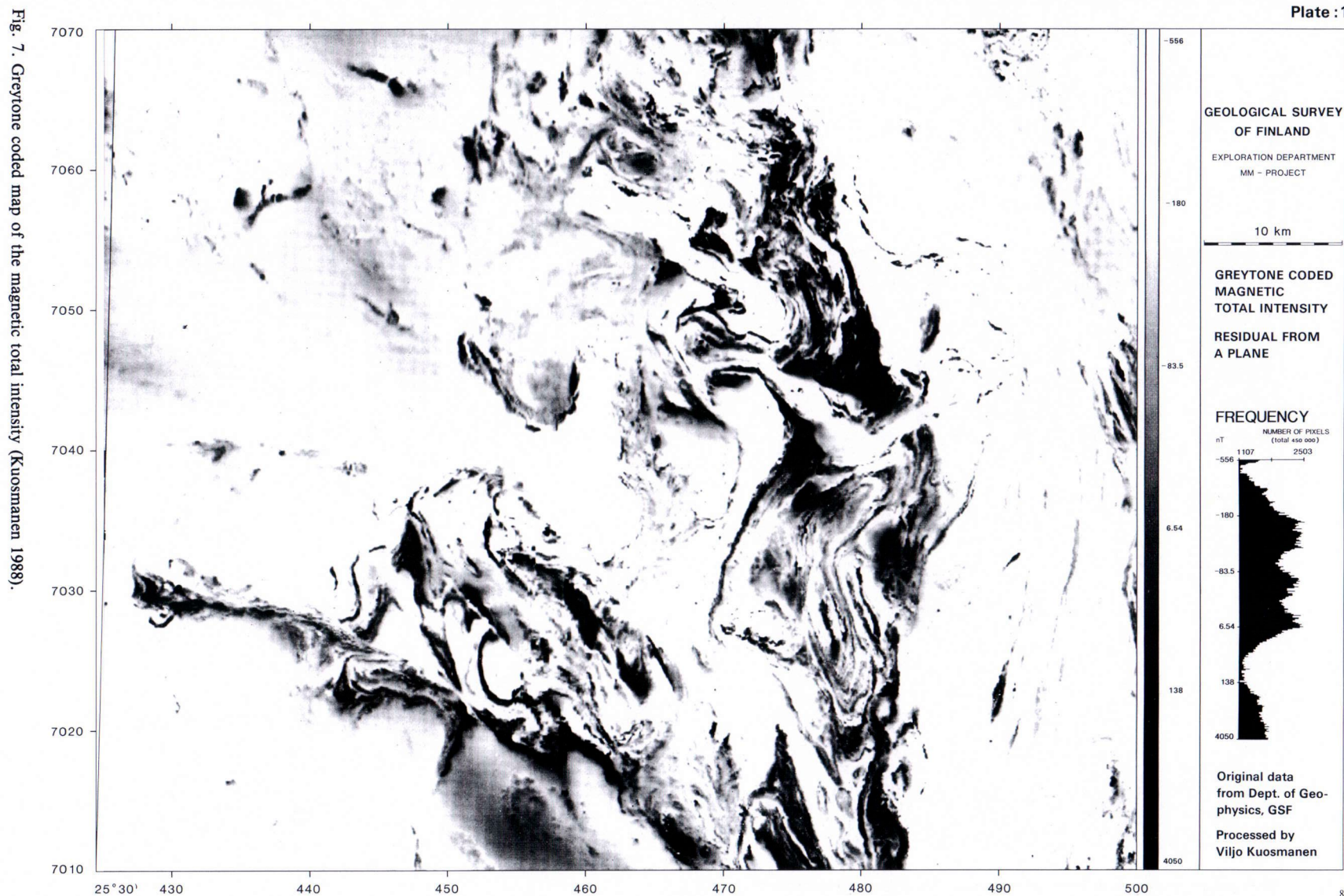

Fig. 6. Pattern of similarity values as a smoothed, interpolated and coloured map.

Fig. 7. Greytone coded map of the magnetic total intensity (Kuosmanen 1988).

Plate : 1

GEOLOGICAL SURVEY
OF FINLAND

EXPLORATION DEPARTMENT
MM – PROJECT

10 km

GREYTONE CODED
MAGNETIC
TOTAL INTENSITY

RESIDUAL FROM
A PLANE

FREQUENCY

NUMBER OF PIXELS
(total 450 000)
nT
1107   2503

Original data
from Dept. of Geo-
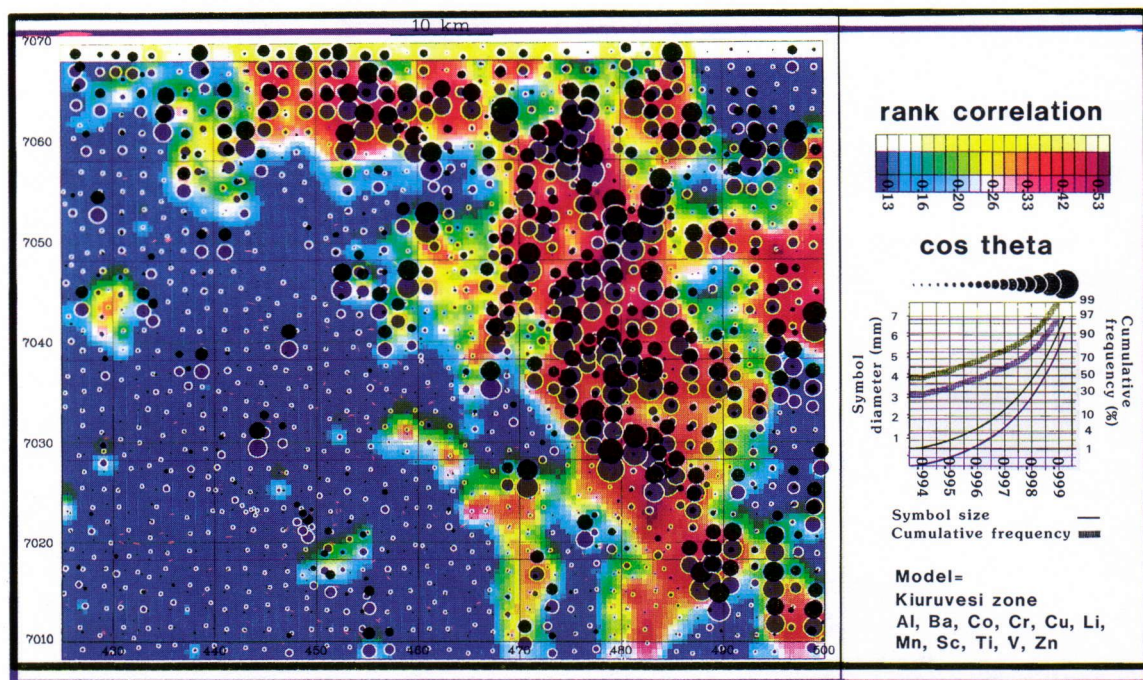physics, GSF

Processed by
Viljo Kuosmanen

Fig. 8. The pattern of similarity values computed by Spearman's rank correlation coefficient is presented as coloured background and cosine theta similarity values as symbols.

that element ratios in till are similar within the certain bedrock area (eg. from Kiuruvesi to Säviä) altough multielement concentration level will change within different parts of the area on account of glacial geology.

The uniformity of the Pyhäsalmi Säviä zone and the NE-SW-trending lineament on the northern flank of the Kiuruvesi schist belt may be mentioned as compatibilities between the similarity pattern and geological structures. The relative rise in the Ba and Mn contents in the model area may account for the high Ba values of pyroxene granitoids (Lahti and Hölttä 1987) and the high Mn-values of altered rocks (Glumoff & Nikkarinen 1991, Pajunen 1988).

Cosine theta and rank correlation measuring methods were also compared. The model was identical in each experiment, and the conclusion were based on the patterns of Fig. 8. The similarity pattern for rank correlation is expressed as colour background and that for cosine theta as

symbols. On average, the two measures reflect very well major bedrock structures, but the most conspicuous difference between the patterns lies within the Säviä and Kangasjärvi subareas. As each subarea has a similar lithology as the model area, the rank correlation pattern is more compatible than the cosine theta pattern. Methodologically the robust measure is more applicable to geochemical data as the parametric measure and robust statistical applications should therefore be applied more to the geochemical studies in future.

References

Glumoff, S. & Nikkarinen, M., 1991. Moreenin raskasmineraalifraktion geokemiallinen kartoitus Pielaveden — Pyhäsalmen alueella. Geol. Surv. Finland, Rep. Invest, in press.

Kuosmanen, V., 1988. Exploration target selection by integration of geodata using statistical and image processing techniques: an example from Central Finland. Part II, (Atlas). Geol. Surv. Finland, Rep. Invest. 84.

Lahti, S. I. & Hölttä, P., 1987. The emplacement, mineralogy and geochemistry of the Vaaraslahti pyroxene granitoid stock in Eastern Finland. IGCP 217, Abstracts, Proterozoic Geochemistry, Lund June 3—6 1987.

Le Maitre, R. W., 1982. Numerical petrology. Elsevier, 281 p.

Pajunen, M., 1988. Tectono-metamorphic evolution of the Hallaperä pyrrhotite-pyrite ore deposit, Central Finland. In Korsman K. (Ed.). Tectono-metamorphic evolution of the Raahe-Ladoga zone. Geol. Surv. Finland, Bull. 343, 51—76.

Salminen, R., 1980. On the geochemistry of copper in the Quaternary deposits in the Kiihtelysvaara area, North Karelia, Finland. Geol. Surv. Finland, Bull. 309, 48 p.

Salminen, R. & Hartikainen, A., 1985. Glacial transport of till and its influence on interpretation of geochemical results in North Karelia, Finland. Geol. Surv. Finland, Bull. 335, 48 p.

Salonen, V.-P., Tenhola, M. & Murtoniemi, S., 1988. Quaternary geology. In G. Gaál (Ed). Exploration target selection by integration of geodata using statistical and image processing techniques: an example from Central Finland. Part I (Text). Geol. Surv. Finland, Rep. Invest. 80, 60—64.

Tenhola, M., 1988. Interpretation of geochemical data. In G. Gaál (Ed). Exploration target selection by integration of geodata using statistical and image processing techniques: an example from Central Finland. Part I (Text). Geol. Surv. Finland, Rep. Invest. 80, 95—101.

Tiainen, M., 1988. Precambrian geology. In G. Gaál (Ed). Exploration target selection by integration of geodata using statistical and image processing techniques: an example from Central Finland. Part I (Text). Geol. Surv. Finland, Rep. Invest. 80, 48—60.

Tontti, M., Koistinen, E. & Seppänen, H., 1981. Vihannin Zn-Cu-malmivyöhykkeen geomatemaattinen arviointi. Summary: Geomathematical evaluation of the Vihanti Zn-Cu-ore zone. Geol. Surv. Finland, Rep. Invest. 54, 58 p.