ELSEVIER

Journal of Geochemical Exploration 72 (2001) 47–58

JOURNAL OF GEOCHEMICAL EXPLORATION

www.elsevier.nl/locate/jgeoexp

Multidimensional statistical technique for detection of low contrast geochemical anomalies

M. Karger[a], S. Sandomirsky[b,*]

[a]*Isotop Ltd, Ashdod, Israel*
[b]*Physical Optics Corporation, Torrance, CA USA*

Dedicated to the memory of Prof. A.P. Solovov.

Received 3 May 1999; accepted 16 October 2000

Abstract

In this article we propose an advanced technique for detecting low contrast geochemical anomalies using a set of features. There are three principal elements in this technique: (1) a statistical measure of the contrast of the anomaly, denoted as τ; (2) selection of a background population; and (3) reduction of the dimensionality of the feature space. In the frame of the model, which describes the statistical distribution of geochemical background as a multidimensional normal distribution of logarithms of concentrations, the index, τ, is a powerful test statistic for the hypothesis of abnormality of an observation. Maps of τ anomalies can be rigorously interpreted on the basis of statistical inferences. Under all equal conditions this technique allows the detection of geochemical anomalies with at least the same contrast (if the chemical elements in a background population are correlated, then even the better) as using selective extractions of metals from soil or other techniques for data processing. The advantages of the proposed technique are demonstrated both theoretically and on examples of rare-metal and copper–nickel mineral deposits.

Keywords: anomaly contrast; background; correlation; buried mineral deposit

1. Introduction

At present, there is general agreement that many mineral deposits that occur on the surface are almost completely exhausted. Therefore, mineral prospecting and exploration in the last decade has concentrated on discovering buried and concealed mineral deposits. One of the pertinent features of these deposits is that, on the surface, they exhibit as low contrast geochemical anomalies. That is, differences between anomalies for each element and background concentrations are statistically insignificant.

In this paper we propose an advanced technique for detecting multicomponent geochemical anomalies using methods of multivariate statistics and multidimensional heuristic methods. The main ideas of the proposed technique are congruent with of those presented by Garrett (1989a) for computer processing of multielement geochemical data.

They also comply with the general tendency exhibited during the 1980s and 1990s toward invoking the multivariate methods into exploration geochemistry. This tendency no doubt will continue in the future, and exploration geochemistry will continue to be ‘a

* Corresponding author. 24 Mann Street, Irvine, CA 92612, USA. Tel.: +1-310-530-1416.

E-mail address: ssandomir@earthlink.net (S. Sandomirsky).

0375-6742/01/$ - see front matter
PII: S0375-6742(00)00162-X

center of multivariate data analysis' (Garrett, 1989a, p. 591).

Three principal constituents form the basis of the proposed technique: (1) a statistical measure of the contrast of the anomaly, denoted as τ; (2) selection of a background population; and (3) reduction of the dimensionality of the feature space. The quantity, τ, is a function of the generalized or Mahalanobis distance between an individual observation and a centroid of the background population in the feature space. Mahalanobis distance, which was comprehensively described by Garrett (1989c), has been repeatedly used as a measure of multidimensional distinctions, mainly in the context of classification of two or more geochemical populations. As a measure of background-to-observation distinctions that is a multidimensional analog of a univariate normal standard deviation, the Mahalanobis distance was considered by Campbell (1983) and Garrett (1989b,c) in their research articles.

The measure τ is a normal approximation of Mahalanobis distance and, as such, is almost independent of dimensionality. With some insignificant variations it can be handled like the standardized difference between the individual observation and the background mean.

The problem of selection of a background population has been discussed through the entire history of geochemical exploration (Solovov, 1959; Hawkes and Webb, 1964). As the term, 'background' is broad and indefinite, the problem is unlikely to possess a full-blown solution. We now hold the view that feasible solutions can be achieved by combining the two following methods: (1) heuristic determination of the multidimensional data pattern with the use of 'space-structure' techniques, such as principal component, factor, and cluster analyses (Bellehumeur et al., 1994; Kramar, 1995; Grunsky and Smee, 1999); and (2) the statistical rejection of outlying observations (Garrett, 1989b,c). Accordingly, our selection of a background population is developed as a two-step procedure.

Reduction of dimensionality of a feature space is a common problem of multidimensional classification. A review of relevant research could be found in the paper by Garrett (1989a). There are two formal reasons for its necessity (Aivazian et al., 1989). First, if the number of variables (elements) is less than the number of observations, the background covariance matrix becomes degenerate. Second, along with increasing dimensionality for the finite number of observations when all the other factors are equal, the Mahalanobis distance decreases, passes through a minimum, and then increases infinitely. A set of most significant elements, known a priori from a theory or experience often solves the problem, particularly in two opposite cases of small scale, regional prospecting and very-large-scale, detailed exploration (Beus and Grigorian, 1977; Garrett et al., 1980). However, as is shown below, the a priori significant sets are not necessarily the best for the selection of *low-contrast* anomalies, significantly better results can be achieved with the sets selected automatically. The algorithms described here use a stepwise variable selection procedure similar to that of Beauchamp et al. (1980) and select variables that are significant with respect to maximizing the τ-anomaly contrast either simultaneously with the anomaly detection or in the predefined observations.

This article is arranged as follows: In Section 2 we briefly review the problem of detection of geochemical anomalies. An index τ and alternative indexes of contrast for multicomponent anomalies are described. Theoretical advantages of index τ, as compared with alternative indexes, are proven. In Section 3, the technique of multicomponent anomaly detection is explained. It consists of three procedures: (1) Selecting the background population; (2) estimating parameters of the multidimensional geochemical background; and (3) selecting the set of features that provides the highest-contrast detection of anomalies. In Section 4 we arrange a competition between index τ and alternative indexes using geochemical data from two buried mineral deposits. The advantages of index τ, which is capable of enriching the practice of multicomponent anomalies detection, are demonstrated.

2. Measures of contrast of anomalies

2.1. Index of contrast of monocomponent anomaly

The representation of geochemical fields as 'normally' distributed backgrounds that are complicated with anomalies is a popular assumption in

state-of-the-art geochemical exploration. Concentrations or logarithms of concentrations of constituents of samples taken from a background area are considered as independent quantities with equal and normal distribution. Abnormal observations are those that significantly differ from an average background concentration. This representation is the foundation for a conventional method of detecting monocomponent anomalies that is known as the 'rule of three sigma'. As a measure of abnormality of an observation, x, these methods utilize a quantity

$$g = \frac{(x - \bar{x}_b)}{s_b}, \quad (1)$$

where x is a logarithm of the concentration of the chemical element and $\bar{x}_b$ and s_b^2 are estimates of mean value and variance, respectively, of a background distribution.

Geometric interpretation of the contrast of anomaly, g, is a distance between points x and $\bar{x}_b$ that is measured by units of standard deviation, s_b, on a scale of logarithms of concentrations. If $\bar{x}_b$ and s_b^2 are obtained over the sample of size, n_b, in the background area, g follows the Student distribution with $n_b - 1$ degrees of freedom. For large values of n_b, the distribution of Eq. (1) differs slightly from the normal distribution with zero mean and unit variance. Below, the fact that x belongs to the normal distribution with parameters, μ, σ^2, we will denote as $x \in N(\mu, \sigma^2)$.

The procedures that operate with the index, g, can be viewed as statistical tests of the hypothesis that x belongs to a given normally distributed population. Also, a criterion in the form, $|g| \geq g^*$ (where g^* is a threshold), is the most powerful criterion for testing this hypothesis. This classic approach has the following features: (1) it is based on a model that reflects reasonable assumptions on how substances are dispersed through the lithosphere; (2) it utilizes a quantitative measure of abnormality; and (3) it allows rigorous (in the context of the model) estimates of differences between observations.

We acknowledge that small-scale geochemical exploration performing over large areas deals with a very complicated mixture of different populations. In this case any accomplished statistical procedure might fail. However, in the case of large-scale exploration, which is usually conducted over areas of moderate size, geochemical data are relatively more uniform, and some additional gain from a classic approach could be extracted.

The ensemble of the features mentioned above provides an exploration geochemist with an accomplished instrument based on a metrological foundation. This, from a metrological point of view, puts geochemical exploration at the same level as geophysical exploration. Of course, a population of independent quantities with equal and (log)normal distribution does not represent an ideal model for background distribution. However, this model can be physically interpreted because these distributions, along with the Pareto distribution, are the only ones explained by physical processes. As the academician Andrei Kolmogorov told M. Karger in a private talk: "A poor model is better than none if we want to have a productive discussion".

The technique for detection of multicomponent anomalies, which is described below, holds all these features.

2.2. Index of contrast of multicomponent anomaly

Let $X = (x_1, x_2, \ldots, x_p)$ be a p-dimensional observation. If it is taken from a normally distributed background population with parameters, $\mu_b = (\mu_1, \ldots \mu_p)$ (vector of mean values) and Σ_b (covariance matrix), then $X \in N(\mu_b, \Sigma_b)$.

The multidimensional analog of the quantity yielded by Eq. (1) is the generalized Mahalanobis distance:

$$D^2 = (X - \bar{X}_b) S_b^{-1} (X - \overline{X_b})^T, \quad (2)$$

where $\overline{X_b} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p)$ is a vector of estimates of mean values, and S_b is an estimate of a covariance matrix. This is a distance between points X and $\overline{X_b}$ in the p-dimensional feature space that is measured by the units of standard deviation along a straight line that connects X and $\overline{X_b}$. If the features in the background population are independent, then Eq. (2) takes the maximal value, $D^2 = \sum g_i^2$. In all other cases $D^2 < \sum g_i^2$.

If $D^2 = 0$, then

$$T^2 = D^2 n_b (n_b - p)[(n_b^2 - 1)p]^{-1} \quad (3)$$

has F-distribution with p and $n_b - p$ degrees of freedom (Kendall and Stuart, 1968). With this basis, a criterion for testing the hypothesis that X is taken

from a background population can be established. However, because each background sample (e.g. each pair of number, n_b, p) requires the individual value of F-distribution this criterion can be inconvenient for mass data processing, especially, if geochemical data have to be integrated with other data, e.g. with digital maps.

Normal approximation of Eq. (2) proposed by Kelley (1948) avoids this drawback:

$$\tau = (\varphi_{n_b-p}T^{2/3})\left(\frac{1}{p} + \frac{T^{4/3}}{n_b - p}\right)^{-1/2}, \quad p > 3,$$
$$\tau^* = \tau + 0.08\tau^5(n_b - p)^{-3}, \quad p \le 3, \tag{4}$$

where $\varphi_k = 3/\sqrt{2} - 2/(3k)$, $k = p, n_b - p$.

If the observation, X, is taken from a background population, the distribution of quantities in Eq. (4) is very close to the normal distribution $N(0,1)$. Owing to this feature, the same criteria of abnormality that are applied for analysis of the fields of g values could be applied for analysis of the fields of τ values. The only difference is in interpretation of negative anomalies: negative τ values correspond to observations that are close to a background mean.

2.3. Other indexes of multidimensional contrast

Many researchers (Beus and Grigorian, 1977; Garrett et al., 1980) overcome the obstacles arising in detecting multicomponent geochemical anomalies using indexes of the form of

$$\Gamma = \sum_1^p f(g_i), \tag{5}$$

where $f(g_i)$ is a monotonous function of g_i.

In particular, Russian geochemists have employed the following 'multiplicative' and 'additive' indexes:

$$\sum_1^p g_i^2 = \Gamma_1; \quad \sum_1^p g_i = \Gamma_2; \quad \prod_1^p \exp(x_i) = \nu. \tag{6}$$

It is seen that in Eq. (6) different features have equal weights in sums and products. This means that these indexes are appropriate only in the case when the features are mutually independent in the background population.

Let compare the properties of quantities from Eqs. (4) and (6) in the case of two features that are strongly correlated with each other:

$$X = \begin{pmatrix} 1.75 \\ 1.75 \end{pmatrix}, \quad \overline{X_b} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$
$$S_b = \begin{pmatrix} 1 & 0.99 \\ 0.99 & 1 \end{pmatrix}, \quad n_b = 100. \tag{7}$$

Taking into account that if the features are independent, then for the background population we have,

$$\frac{n_b + 1}{n_b}\Gamma_1 \in \chi^2_p; \quad \frac{1}{\sqrt{p}}\Gamma_2 \in N(0,1),$$

and we receive for Eq. (7), $\Gamma_1(n_b + 1)/n_b = 6.06$ and $\Gamma_2 p^{-1/2} = 2.47$. Comparing these numbers with the 95% critical values of distributions' chi-square and $N(0,1)$ (5.99 and 1.96, respectively) we find that the 'observation' X is recognized as abnormal from the point of view of indexes Γ_1 and Γ_2. In reality one feature in Eq. (7) actually duplicates the other that provides $|g| < 1.96$. This observation cannot be recognized as abnormal, because it is in accordance with the value of $\tau^* = 1.24$.

It follows from the above consideration that, if the features are correlated in the background population, the index, τ, has advantages over the indices of Eq. (6). These advantages are not only theoretical. The index, τ, provides real benefits in practice. This is illustrated by the examples presented in Section 4.

3. Technique for detecting anomalies of τ

3.1. Step 1: preliminary selection of background observations

The problem of detecting geochemical anomalies has two specific features. First, we do not know a priori whether a territory under investigation contains any anomalies. Second, if an anomaly does exits, we do not know the position of the boundaries between the anomaly and the surrounding sites that are not abnormal. This is an ill-posed problem, which people solve 'unconsciously' in 2D or 3D space. If we study a spatial distribution of one feature at the flanks of a well-developed deposit, we can pinpoint with high confidence the sites without signs of mineralization. The samples collected away from the deposit we

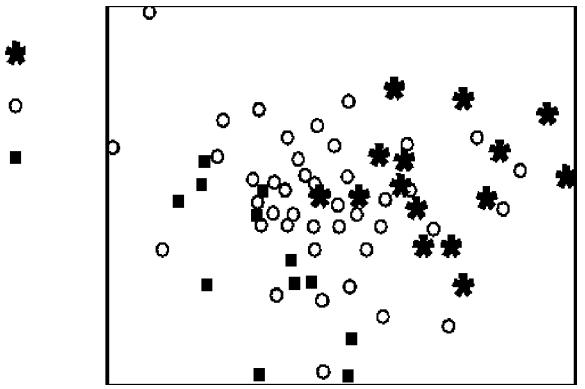

Fig. 1. Pattern of multidimensional sample of Pechenga deposit (the result of nonlinear mapping from 18-dimensional feature space). Asterisks — observations from the contour of projection of ore zone onto the land surface; circles — observations located apart from this contour; solid squares — observations composing the background sample.

recognize as representing a background population and 'unconsciously' detect anomalies. In the other case, when multicomponent anomalies have to be detected in a poorly investigated territory, we need some kind of tool to create an opportunity to represent a set of multidimensional observations in a form convenient for visual analysis. This can be achieved by reducing the original dimension of this set down to 2 without significant distortion of a sample pattern in the feature space.

Linear projection onto the principal components is the classic method for the reduction of dimensionality (Kendall and Stuart, 1968). This transformation linearly projects the sample onto the principal axes of the ellipsoid of dispersion. Projection from p-dimensional space onto the first l components has the following unique features: (1) among all linear projections it minimizes the sums of squares of lengths of perpendiculars, reducing points in p-dimensional space down to l-dimensional subspace; and (2) it preserves in the best manner the angles between points and distances from them to the center of the population (Kendall and Stuart, 1968).

In practice, it is known that if the sum of variances of the first two principal components is more than 80% of the total variance, then the projection onto the plane of these principal components properly reflects the multidimensional pattern of a sample. If this quantity is less than 60%, then this projection is inappropriate for our purposes. Projections of this exact type, those that do not contain enough information are commonly encountered in geological studies (Howarth, 1973).

The algorithm of nonlinear mapping described below (Sammon, 1969) belongs to the family of multidimensional scaling algorithms. Through slight distortion of mutual arrangement of points in a population, the algorithm reflects the main features of a complex multidimensional pattern.

Let

$$x_i = (x_{i1}, \ldots, x_{ip})^{\mathrm{T}}, \quad i = \overline{1,n}, \tag{8}$$

be a set of points in p-dimensional space. The points in Eq. (8) have to be mapped into the points on a plane. This is represented by

$$y_i = (y_{i1}, y_{i2})^{\mathrm{T}}, \quad i = \overline{1,n}. \tag{9}$$

The mapping is performed by the algorithm:

$$\arg\min(\Delta), \quad \Delta = \sum_i \sum_j (d^*_{ij} - d_{ij})^2 / d^*_{ij} / \sum_i \sum_j d^*_{ij}, \tag{10}$$

where d^*_{ij}, $i > j = \overline{1,n}$, is the distance (d) between points, measured by any metric; d_{ij} is the Euclidean distance between points as determined by Eq. (9).

The method of rapid descent is used to reach the minimum value of Eq. (10). We used projection on the first of the principal components as an initial approximation for Eq. (9) to avoid a local-minimum trap. The number of examples (Sammon, 1969; Howarth, 1973; Howarth et al., 1977; Karger and Krupnov, 1980; Nason and Sibson, 1991) shows that the algorithm, Eq. (10), provides results that are at least not worse than the projections onto the principal components.

Fig. 1 depicts the procedure for selection of background observations for the Pechenga copper–nickel deposit (see Section 4.3). Because of the exploration challenge, we were trying to achieve, in this case, the main feature of a sample pattern that we were interested in was an elongation of a cloud of points that was caused by the variation of observations due to a transition from the sites uninfluenced by mineralization to the contour of the ore zone. Therefore, the background sample was composed of randomly selected 'waste' observations.

3.2. Step 2: estimation of parameters of a multidimensional background

It frequently happens that geochemical observations occur beyond the limit of sensitivity of a given analytical technique. There are a number of methods for estimating the mean, variance, and covariance (correlation) over truncated samples. However, if the estimates of covariance are obtained separately for each pair of features, then the combined covariance matrix in general will not be positively defined. An algorithm described below is close to the idea of a 'bootstrap' as well as to one of the algorithms mentioned in the classical work of Afifi and Elashoff (1969).

Let us assume that in a correlation matrix, R, there are some elements that were obtained over truncated samples. Let $r_{ij}, i \neq j$ be a correlation coefficient between ith and jth features, which has maximal absolute value among other coefficients.

$$x_{1i}, \ldots \overset{\circ}{x}_{ki}, \ldots, \overset{\circ}{x}_{mi}, \ldots, x_{ni}, \quad x_{1j}, \ldots x_{kj}, \ldots, \overset{\circ}{x}_{mj}, \ldots, x_{nj}, \tag{11}$$

are the samples for ith and jth features; circles marks those values, which are beyond a limit of sensitivity.

Let us take two cases of truncation in Eq. (11): (1) only one of two features has a truncated value (kth observation); (2) both features have truncated values (mth observation). If r_{ij} significantly differs from zero, then in Case (1) the truncated value is replaced by its most probable value:

$$x_{ki} = \bar{x}_i + (x_{kj} - \bar{x}_j)(s_{ij}/s_{jj}), \tag{12}$$

where s_{ij}/s_{jj} is a regression coefficient. In Case (2), a random number taken from a corresponding truncated population, which has normal distribution, replaces one of the truncated values, either x_{mi} or x_{mj}. The other truncated value then is restored in accordance with Eq. (12). After all truncated values in Eq. (11) are restored, the same procedure has to be performed for an other correlation coefficient, the absolute value of which is the next after the maximum. This process of regressive restoration is performed for all significant correlation coefficients. For insignificant correlation coefficients, all truncated values are replaced with random numbers. Finally, the estimations for vector $\bar{X}$ and matrix S are calculated in the usual way.

If R contains both significant and insignificant correlation coefficients and p is not too small, then this algorithm provides unbiased estimates for trace and determinant matrix R. Our simulation shows that reasonable results can be achieved at $p > 5, n_b > 50$. This agrees well with the results of Haitovski (1968), who tested similar algorithms.

We will not discuss in depth other routine procedures that could be carried out at this step. We will mention only the procedure of rejecting the outlying observations, which, in particular, completes the selection of the background population. This procedure uses a computerized version of the algorithm that was put forward by Garrett (1989c).

3.3. Selection of significant features

This section describes two modes of an algorithm for selecting features that are significant for maximizing the contrast of detected anomalies. In the first mode, the selection is performed simultaneously with the detection of anomalies (*wandering selection*). In the second mode, a set of abnormal observations is defined a priori and selection of significant features is performed with the goal of maximizing contrast just in these predefined observations (*pursued selection*).

3.3.1. Wandering selection of features

Let us assume that concentrations of p chemical elements (features) have been measured in samples collected over some area that is geochemically homogeneous. In the course of detection of τ anomalies, $n^{(p)}$ samples were recognized as abnormal. Their average contrast will be equal to

$$\overline{\tau_i^p} = \frac{\sum_1^{n^{(p)}} \tau_j}{n^{(p)}}. \tag{13}$$

Then, eliminating sequentially one feature at a time, we find $n_i^{(p-1)}$ abnormal samples for each $(p-1)$-dimensional set of features and calculate their average contrast, $\overline{\tau_i^{(p-1)}}$. The optimal subset is the one that satisfies the following conditions:

$$\overline{\tau^{(p-1)}} \equiv \overline{\tau_i^{(p-1)}} \geq \tau^{(p)} \quad \text{and} \quad n^{(p-1)} \equiv n_i^{(p-1)} \geq n^{(p)}. \tag{14}$$

From this set, again we eliminate sequentially one feature, and for each $(p-2)$-dimensional set we find abnormal samples and calculate their average contrast, $\overline{\tau_i^{(p-2)}}$. We also stop at that set, which satisfies conditions similar to those of Eq. (14). These procedures are repeated until the conditions of Eq. (14) are not satisfied.

3.3.1. Pursued selection of features

When abnormal observations are known a priori, the algorithm provides maximum average contrast, τ, in the given observations. In this case, the sequence of steps, which is described above, will be performed with exception of the second inequality in Eq. (14).

4. Comparison of indexes of contrast of multicomponent anomalies for selected mineral deposits

In this section we investigate how efficient the proposed technique is in practice using two mineral deposits as examples. One of them is a rare metal deposit and the other is a copper–nickel deposit. Both deposits are located on the Kola peninsula (Russia) and are concealed under moraine overburden. The thickness of the moraine deposits is 20–40 m at the rare metal deposit and 10–20 m at the copper–nickel deposit. The humus layer of soil was sampled in both cases and the <1 mm fraction was analyzed by two procedures: (1) extraction of mobile salt constituents with an acetate buffer followed by determination for the sum of metals (Nesvizhskaya et al., 1975); and (2) a direct current (DC) arc emission spectrometric analysis for 32 chemical elements.

The competition between indices of contrast, Eqs. (4) and (6), was performed on the basis of the results obtained with the DC arc emission spectrometric analysis. Two criteria were applied for evaluation in the competition: (1) the contrast of anomalies; and (2) the accuracy of matching the contours of the anomalies. Contours of an ore zone on the paleosurface were matched with the contours of anomalies obtained by acetate extracts. To eliminate bias in evaluating the indices of efficiency of anomaly detection, which are described in Section 4.1, the evaluations were undertaken by an impartial referee.

4.1. Indexes of efficiency of anomalies detection

Let us assume that among the samples, which were collected over some area, $n_A(\alpha)$ samples were recognized as abnormal at a significance level of a%. Some number, $n_A(\alpha)_{Ore}$, of them fall inside the contour of an ore zone, and $n_A(\alpha)_{Extr}$ samples match anomalies in aqueous extracts. The quantities

$$\eta(\alpha)_{Extr} = \frac{n_A(\alpha)_{Extr}}{n_A(\alpha)}, \quad \eta(\alpha)_{Ore} = \frac{n_A(\alpha)_{Ore}}{n_A(\alpha)} \tag{15}$$

are the measures indicating how precisely the anomalies match these contours. Lets assume that the values of any index of contrast for a set of samples are mutually independent, and that sampling spots are randomly distributed over the ore zone and over the contours of the anomalies in the aqueous extracts. This assumption is not too violent from the 'point of view' of the ore zone, which is the common origin of all the contours that are considered. The assumption realizes a binomial distribution of the overlap, η, of these contours, and approximate boundaries of $(1-\epsilon)100\%$ confidence interval for η can be inferred:

$$\Delta\eta \approx [n_A\eta + 0.5l^2 \pm l\sqrt{n_A(1-\eta)\eta + 0.25l^2}](n_A + l^2)^{-1} \tag{16}$$

where l is a $(1-\epsilon)100\%$ quantile of the distribution $N(0,1)$.

It is convenient to compare quantities in Eq. (15) which are acquired for different indexes of contrast of anomalies, using interval estimates from Eq. (16). It is useful to complete the interval estimates with a test detecting outlying values of η. Let $H=(\eta)$ be an accuracy of matching for several indexes of contrast, which are used for calculation of an average

$$\bar{\eta} = \sum_H n_A\eta\left(\sum_H n_A\eta\right)^{-1}.$$

The statistics,

$$\eta^* = (\eta - \bar{\eta})n_A[n_A(1-\eta)\bar{\eta}]^{-1/2}, \tag{17}$$

are standardized deviations of η from the mean $\bar{\eta}$. If these values are statistically indistinguishable, then for all η it is true that $\eta^* \in N(0,1)$, approximately. If one of these values differs from others, then for that

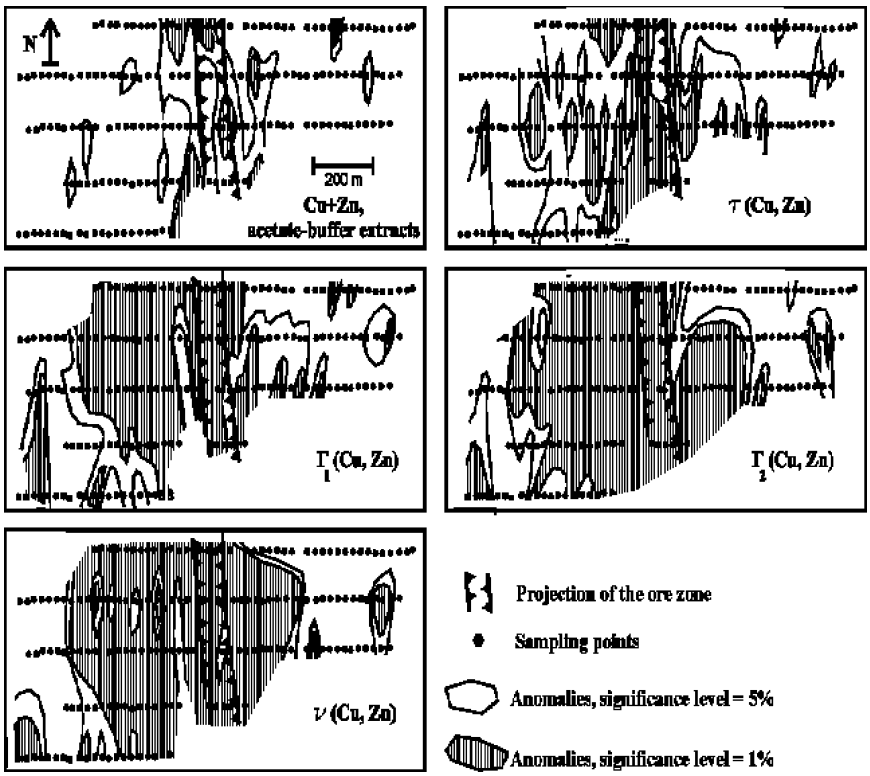

Fig. 2. Rare metal deposit. Anomalies of Cu + Zn in extracts and anomalies of τ, Γ_1, Γ_2, and ν by concentrations of Cu and Zn in bulk samples.

value, $\eta^* \gg 0$. Below, the parameter, η^*, will be used with the same indices as in Eq. (17).

4.2. Rare metal mineral deposit

As a historical note, due to the security system operating in the Soviet Union in the 1980s, the authors were prohibited from knowing the name of the deposit for which they had been provided samples and data. The rare metal tin–beryllium deposit is located in the Kola peninsula (Russia). It is confined to the granite massive, which was intruded through sedimentary-metamorphic rocks of the middle-Proterozoic age. The vein-like, steep-dipping orebodies are located among skarn limestones containing disseminated mineralization. The orebodies are several meters thick and several hundred meters long along the strike. The investigation site is confined to the western part of the deposit. The humus layer of soil was sampled along the grid 100 m × 20 m. A sample of $<$1 mm was analyzed by both DC arc emission spectrometric analysis (bulk fraction) and by the dithizone method (Irving, 1980) for the sum of Cu and Zn concentrations (acetate extract).

The local multicomponent geochemical background population was selected beyond the known ore zones (Fig. 1) using a nonlinear mapping technique described in Section 3.1. Despite the significant thickness of the transported overburden, high concentrations of ore elements were detected both in the acetate extracts and in the bulk samples. That is why, in this study, we consider anomalies of τ, Γ_1, Γ_2, ν, which were calculated for Cu and Zn concentrations in the bulk samples.

Fig. 2 shows anomalies that have been detected at two significant levels: $\alpha = 5\%$ and $\alpha = 1\%$.[1] Visual comparison of these maps reveals that the contour of the τ anomaly looks more ragged than the contours of

[1] To avoid overloading of plots not all confidence intervals are shown. An asymmetry of confidence intervals is caused by errors of normal approximation, which was used in Eq. (13).

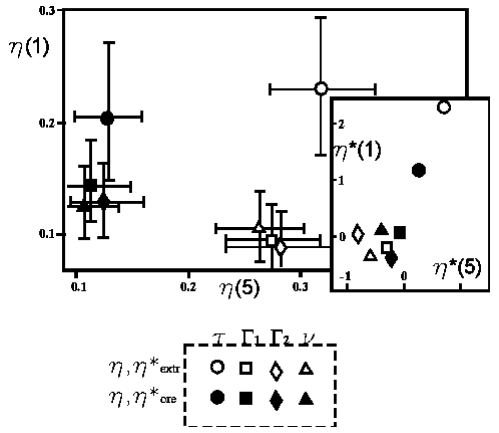

Fig. 3. Rare metal deposit. Accuracy of matching anomalies to contour of ore zone and to contour of anomalies in extracts. Solid signs — $\eta(\cdot)_{ore}$, $\eta^{*>}(\cdot)_{ore}$; empty signs — $\eta(\cdot)_{extr}$, $\eta^{*}(\cdot)_{extr}$; circles — τ, squares — Γ_1, triangles — Γ_2, diamonds — ν.

the Γ_1, Γ_2, and ν anomalies, and that the τ anomaly looks more similar to the contour of the sum of the metals Cu and Zn in acetate extracts. The area of greatest contrast and the most extended area of τ anomalies closely match the contour of ore zone projected onto the land surface. At the same time, the wide, smooth anomalies of Γ_1, Γ_2, and ν look like they were obtained as the result of less-detailed mapping.

Fig. 3 quantitatively confirms these visual impressions. At the confidence level, $\alpha = 5\%$, the differences between compared values of η, η^* can be considered statistically insignificant. However, at the confidence level, $\alpha = 1\%$, the values of η^* (Eq. (1)) for τ (circles in Fig. 2) stand apart from the cluster of all other points on the plot. This means that high values of τ match more precisely the given contours than the indexes Γ_1, Γ_2, and ν. In other words, with respect to all other indexes, index τ first of all demonstrates the closest matching to the field of concentration obtained from the acetate extracts, and, secondly, it provides anomalies with maximum contrast over the contour of the ore zone.

4.3. The Cu–Ni Pechenga deposit

The Cu–Ni mineral deposits of Kola Peninsula are thoroughly investigated (Smirnov, 1977). The Cu–Ni mineralization is associated with the basic and ultrabasic intrusions; most of them are located among sedimentary rocks of the upper parts of the Pechengskaya series. Ore mineralization is confined to exo and endocontacts of big differentiate massifs. Mineral composition includes chalcopyrite, pentlandite, and pyrrhotite, contaminated with sphalerite, galena, rutile, graphite, and native copper.

The investigation site is confined to the central part of the Pechengskaya mineral zone. The humus layer of soil was sampled along a grid 100 m × 20 m. The <1 mm fraction was analyzed as described above. The sum of Cu + Ni + Co concentrations was determined in acetate extract.

Concentration of the sum of metals in this case is 10 times less than in the previous example. The relative area of anomalies delineated at a 1% confidence level is also small (Fig. 5). The indexes, Γ_1, Γ_2, and ν, calculated by Cu, Ni, and Co concentrations in balk samples, also provide low contrast anomalies. That is why, in this example, we were forced to expand the spectrum of features for anomalies detection.

We considered five sets of features. They are denoted by letters from ‘A’ to ‘E’ (Table 1). The first two conventional sets represent the lists of typical elements: the wide set of elements, ‘A,’ is usually utilized in the course of prospecting for mineral deposits at the Kola peninsula; the narrow set of elements, ‘B,’ is typical for Cu–Ni mineralization.

Sets ‘C’, ‘D’, and ‘E’ were selected automatically by algorithms. This is described above in Section 2.3. The selection was conducted under the most favorable conditions for each competing index of contrast. To obtain set ‘C’ we applied the pursued selection, maximizing the contrast of anomalies Γ_1 (Γ_2, and ν, respectively) in the contour of ore zone. We applied the same pursued selection maximizing the contrast of τ anomalies to obtain the set ‘D’. The last set, ‘E,’ was obtained using wandering selection maximizing the contrast of τ anomalies.

Fig. 4 shows the contours of anomalies at the 5% and 1% confidence levels. Fig. 5 demonstrates the degree of matching various indexes of contrast calculated for different sets of features to the contour of the ore zone. It is clearly seen that index τ is superior. It provides maximum values of $\eta(\alpha)_{Ore}$ for all α and for all sets of features. Significantly bigger values of $\eta(\alpha)_{Ore}$ are taken at $\alpha = 1$ for empirically selected sets ‘D’ and ‘E’. The indexes, Γ_1, Γ_2, and ν

Table 1
Sets of features

Set	Description	Chemical elements included
A	Wide set of typical elements	Sr, Ba, Ti, Mn, Cr, V, Ni, Co, Cu, Zn, Pb, Sn, Mo, Zr, Ca, Mg, Al, Fe
B	Narrow set of typical elements	Cr, Ni, Co, Cu, Zn, Pb, Mo
C	Features selected for Γ_1 by the pursued selection	Sr, Mn, Cr, Ni, Co, Ca, Mg
D	Features selected for τ by pursued selection	Sr, Ba, Ti, Mn, Ni, Cu, Mo, Ca, Mg
E	Features selected for τ by wandering selection	Sr, Ba, Ti, Ni, Co, Cu, Sn, Zr, Ca

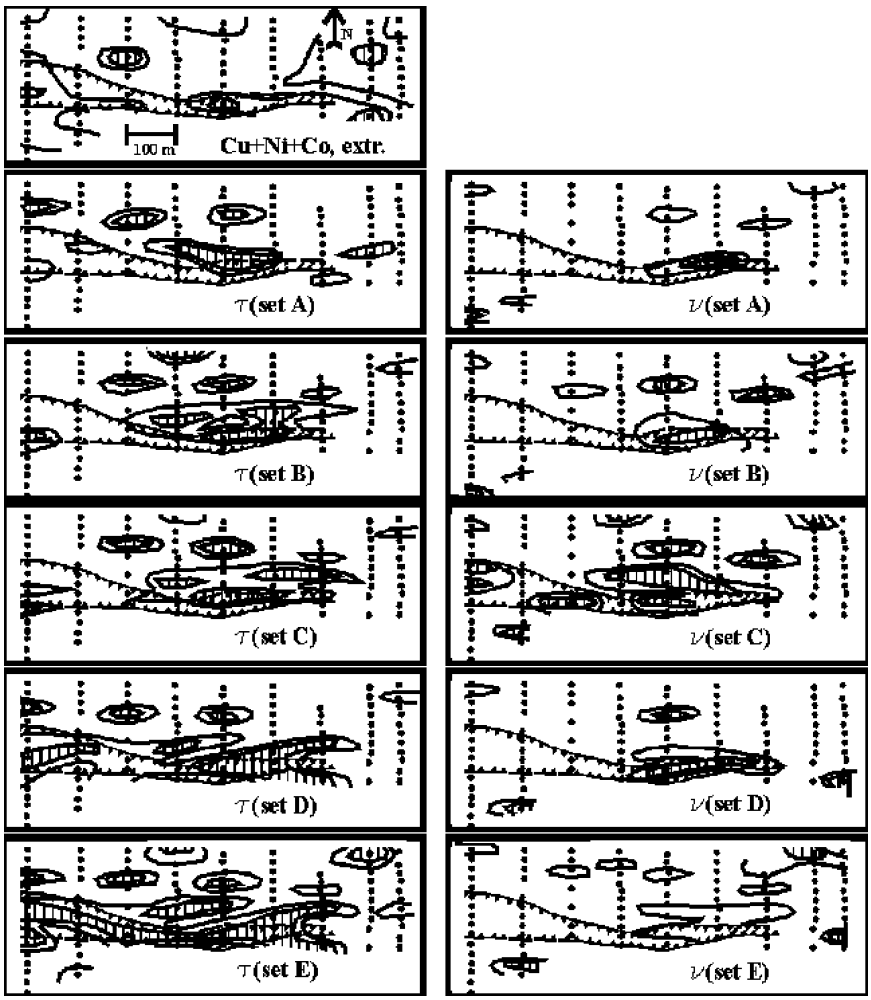

Fig. 4. Anomalies of Cu + Ni + Co in extracts and anomalies of τ, Γ_1, Γ_2 and ν by concentrations in bulk samples. A–E — labels of sets of features (see Table 1).

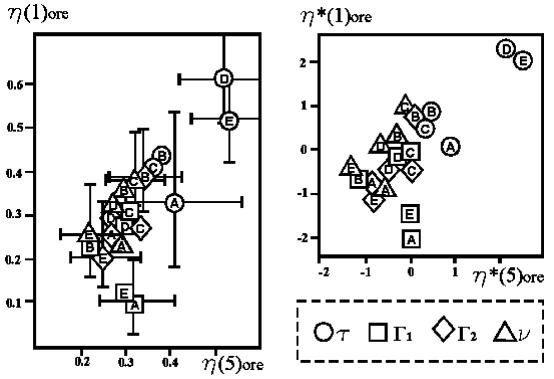

Fig. 5. The Pechenga deposit. Accuracy of matching anomalies to contour of ore zone. A–E — labels of sets of features; circles — τ, squares — Γ_1, triangles — Γ_2, diamonds — ν.

(Eq. (6)), provide the best opportunities for prospecting when they are calculated for sets 'A' and 'B'. In these cases, indexes (Eq. (6)) are comparable with worse variants of index τ (sets 'A', 'B', and 'C').

The differences between $\eta(5)_{extr}$, and $\eta^*(5)_{extr}$ (Fig. 6) are not demonstrated clearly. We can say that τ anomalies more precisely match to anomalies in acetate extracts if they are calculated using the feature set, 'E,' selected by wandering selection. In this case $\eta^*(5)_{extr} = 1.65$.

These examples confirm the advantages of τ with respect to other indexes of contrast. If selection of significant features is properly performed, then the index τ is an effective tool for delineating anomalies of mobile salt constituents in soil and for localization of lode mineralization on a paleosurface as well.

5. Conclusions

We conclude that it is reasonable to detect multicomponent geochemical anomalies using the statistical measure of contrast, τ. It provides maximum contrast for detection of anomalies. Other indexes of contrast, which ignore the correlation between features in the background population, represent geochemical fields in a very generalized form, similar to that obtained by a small-scale mapping. The advantages of index τ provide the maximum benefit in prospecting with low-cost methods of soil analysis for buried and concealed deposits. In the case presented a combination of index τ and the selection of significant features permits the detection of low contrast anomalies with at least the same success as utilizing more expensive acetate extracts of mobile salt forms.

Acknowledgements

The authors deeply respect the memory of professor A.P. Solovov, who was one of the founders of exploration geochemistry, and whose

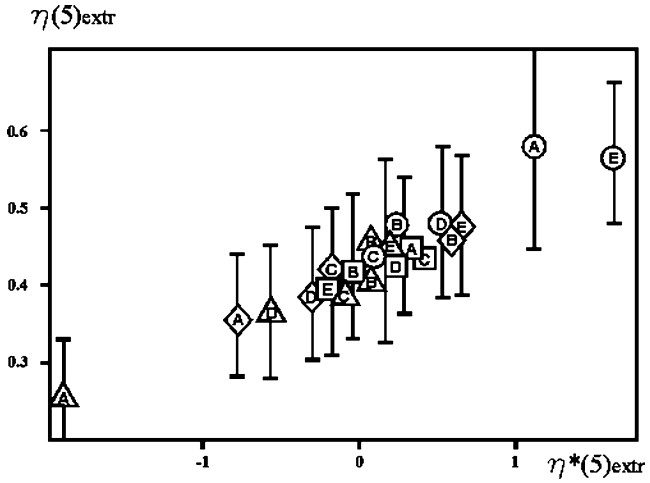

Fig. 6. Pechenga deposit. Accuracy of matching anomalies at 5% confidence level to the contours of anomalies in extracts. A–E — labels of the sets of features; circles — τ, squares — Γ_1, triangles — Γ_2, diamonds — ν.

ideas encouraged this work. The authors have the honor to be his disciples. We also gratefully acknowledge professor S.V. Grigorian for discussions and useful recommendations.

References

Aivazian, S.A., Buchstaber, B.M., Enukov, I.S., Meshalkin, L.D., 1989. Applied Statistics: Classification and recduction of Dimensionality, (Prikladnaya Statistika: Klassifikatsiya i Snizhenie Razmernosti). Moscow, Finansy I Statistika Publication, 607pp. (in Russian).

Afifi, A.A., Elashoff, R.M., 1969. Missing observations in multivariate statistics. J. Am. Stat. Assoc. (325), 337–358.

Beauchamp, J.J., Begovich, C.L., Kane, V.E., Wolf, D.A., 1980. Application of discriminant analysis and generalized distance measure to uranium exploration. J. Math. Geol. 12 (6), 539–558.

Bellehumeur, C., Marcotte, D., Jebrak, M., 1994. Multi-element relationships and spatial structures of regional geochemical data from stream sediments, southwestern Quebec, Canada. J. Geochem. Explor. 51, 11–35.

Beus, A.A., Grigorian, S.V., 1977. Geochemical Exploration Methods for Mineral Deposits Ill. Applied Publication, 287pp.

Campbell, N.A., 1983. Statistical treatment for geochemical data. In: Smith, R.E. (Ed.). Geochemical Exploration in Deeply Weathered Terrain. CSIRO Institute of Energy and Earth Resources, Wembley, Western Australia, pp. 181–184.

Garrett, R.G., 1989. The role of computers in exploration geochemistry. In: Garland, G.D. (Ed.), Proceedings of Exploration'87. Third Decennial International Conference on Geophysical and Geochemical Exploration for Minerals and Groundwater, Special Volume 3, pp. 586–608.

Garrett, R.G., 1989. A robust multivariate allocation procedure with application to geochemical data. In: Agterberg, F.P., Bonham-Carter, G.F. (Eds.), Statistical Applications in the Earth Sciences, Geological Survey of Canada, Paper 89-9, pp. 309–318.

Garrett, R.G., 1989c. The chi-square plot a tool for multivariate outlier recognition. J. Geochem. Explor. 32, 319–341.

Garrett, R.G., Kane, V.E., Zeigler, R.K., 1980. The management and analysis of regional geochemical data. J. Geochem. Explor. 13, 115–152.

Grunsky, E.C., Smee, B.W., 1999. The differentiation of soil types and mineralization from multi-element geochemistry methods and digital topography. J. Geochem. Explor. 67, 287–299.

Haitovski, Y., 1968. Missing observations and the estimations of the parameters of the normal distribution. J. R. Stat. Assoc. Ser. B 30 (1), 67–82.

Hawkes, H.E., Webb, J.S., 1964. Geochemistry in Mineral Exploration. Harper and Row, New York (415pp.).

Howarth, R.J., 1973. Preliminary assessment of a nonlinear mapping algorithm in a geological context. J. Math. Geol. 5 (1), 39–58.

Howarth, R.J., Cronan, D.S., Glashy, G.P., 1977. Nonlinear mapping of regional geochemical variability of manganese nodules in Pacific ocean. Inst. Min. Metall Trans. B 87, 4–8.

Irving, H.M.N.H., 1980. Crit. Rev. Anal. Chem. 8 (4), 321–405.

Karger, M.D., Krupnov, I.I., 1980. Visualization of multidimensional data for classification of mud waters of the Black Sea (Vizualizatsiya mnogomernykh dannykh pri klassifikatsii ilovykh vod Chernogo morya). Vestnik MGU Ser. Geologiya , 2, 82–86 (in Russian).

Kelley, T.L., 1948. The Kelley Statistical Tables, Cambridge, MA.

Kendall, M.J, Stuart, A., 1968. The Advanced Theory of Statistics. Design and Analysis, and Time-Series. 2nd ed., vol.3. Charles Griffin, London.

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

Nason, G.P., Sibson, R., 1991. Using projection pursuit in multispectral image analysis. Computing science and statistics. Proceedings of the 23rd Symposium on the Interface, pp. 579–582.

Nesvizhskaya, N.I., Sayet, Yu. Ye., Grigorian, S.V., 1975. Prospecting of overburden sulfide deposits using overlaid haloes, (Poiski perekrytykh sul'fidnykh mestorozhdenii po nalozhennym oreolam,). Metod. Rekomendatsii, Moscow, IMGRE (in Russian).

Sammon, J.W., 1969. A nonlinear mapping for data structure analysis. IEEE Trans. Comput. C 18 (5), 401–409.

Smirnov, V.I. (Ed.), 1977. 1st ed.. Ore Deposits of the USSR, vol. 3. Pitman, London.

Solovov, A.P., (1959). Theory and practice of metallometric surveys (Teoria i Praktika Metallometricheskikh S'emok). Alma-Ata, 318pp. (in Russian).