Mathematical Geology, Vol. 24, No. 7, 1992

Decomposed and Weighted Characteristic Analysis for the Quantitative Estimation of Mineral Resources[1]

Guocheng Pan[2] and DeVerle P. Harris[3]

Characteristic analysis has been employed as a means of selecting favorable targets for mineral exploration. This paper describes a new version of characteristic analysis that is designed to estimate mineral resources as well as delineating exploration targets. The new version, referred to as decomposed and weighted characteristic analysis, employs a weighting scheme for both samples and variables involved in the model to extract the useful information on tonnages and the order of importance of variables. To construct a decomposed model, an optimum cutting technique for the ordered quantity is developed. The model is demonstrated on a case study of pegmatitic Nb–Ta deposits in China.

KEY WORDS: decomposed and weighted characteristic analysis, mineral resources estimation, Nb–Ta deposit, optimum cutting.

INTRODUCTION

Characteristic analysis (also called decision modeling) was proposed by Botbol (1971). It was demonstrated as a tool for quantitatively describing how typical each of a number of geological characteristics, such as mineralogy, geochemistry, geophysics, etc., is of a set of mineral deposits (Harris, 1984). Since 1977, this statistical technique has been described as a means for selecting exploration targets (Botbol et al., 1978; McCammon et al., 1983; Pan and Wang, 1987). This methodology has been developed to the stage that it brings together three separate efforts: the analysis of geological characteristics, genetic modeling, and decision making.

One of the major advantages of characteristic analysis is its simplicity, rendering it easily understood by geologists. Another appealing feature of the method is that it consists of quantitative procedures by which different modes

[1]Received 28 February 1992; accepted 30 March 1992.
[2]NERCO Exploration Company, 500 NE Multnomah, Portland, Oregon 97232.
[3]Department of Mining and Geological Engineering, University of Arizona, Tucson, Arizona 85721.

of geological data pertinent to mineralization can be quantified, integrated, and analyzed.

Much progress has been made in recent years. Perhaps, the most noticeable development is the probabilistic description of association between characteristics (McCammon et al., 1983; Pan, 1985a). This development leads to a better understanding and interpretation of the kinds and strengths of relationships between variables. McCammon et al. (1983) proposed a so-called generalized characteristic analysis in which the established model is progressively expanded by adding to the statistical base the unknown cells that are most closely related to the initial control cells selected for model identification.

This paper presents a modification of characteristic analysis especially designed to make it more useful as a method for resource appraisal. This modified procedure is referred to as decomposed and weighted characteristic analysis (DWCA) and is demonstrated on a case study of Nb–Ta resources in pegmatite deposits in China.

CONVENTIONAL CHARACTERISTIC ANALYSIS

Like most other statistical models in mineral exploration, characteristic analysis is primarily based on observations or measurements taken in and around known ore deposits. Use of the model has been predicated upon the assumption that the favorability of the region of unknown mineral potential is a positive function of the similarity of its characteristics with those of the known deposits. Let $\mathbf{X} = (x_{ij})$ denote the matrix containing the data of n observations on m geological characteristics, where x_{ij} is either binary or ternary. When x_{ij} is binary, it is assigned one if characteristic j exists in cell i and zero otherwise. For ternary coding x_{ij} is one if the presence of characteristic j is favorable to cell i, negative one if the presence of characteristic j is unfavorable to cell i, and zero if the characteristic j is unknown or unevaluated.

The matching coefficients between pairs of m characteristics are defined by

$$\mathbf{S} = (s_{jk}) = n^{-1}\mathbf{X}^T\mathbf{X} \tag{1}$$

where $s_{jk} = n^{-1} \Sigma_{i=1}^{n} x_{ij}x_{ik}$, $j, k = 1, 2, \ldots, m$. Beginning with matrix $\mathbf{S}$, a weight for each of the m characteristics is usually calculated by one of the following two ways:

(1) *Square Root Method.* The weight for characteristic j is defined as

$$\hat{w}_j = \left[m^{-1} \sum_{k=1}^{m} s_{jk}^2 \right]^{1/2}, j = 1, 2, \ldots, m \tag{2}$$

(2) *Principal Component Method.* According to the theorem of spectrum decomposition for a matrix, the coefficient matrix $\mathbf{S}$ can be expressed as follows:

$$\mathbf{S} = \lambda_1 \mathbf{a}_1 \mathbf{a}_1^T + \ldots + \lambda_m \mathbf{a}_m \mathbf{a}_m^T \tag{3a}$$

where λ_j is the jth eigenvalue of $\mathbf{S}$ and $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m \geq 0$; $\mathbf{a}_j$ is the jth eigenvector of $\mathbf{S}$ associated with λ_j. Define

$$\phi_j = \lambda_j \Big/ \sum_k \lambda_k, \; j = 1, 2, \ldots, m \tag{3b}$$

Clearly, $1 \geq \phi_1 \geq \ldots \geq \phi_m \geq 0$, since $\mathbf{S}$ is generally a non-negative definite matrix. If ϕ_1 is large enough (e.g., 0.8), matrix $\mathbf{S}$ can be approximately expressed by its first eigenvector without losing much information, i.e.,

$$\mathbf{S} \approx \lambda_1 \mathbf{a}_1 \mathbf{a}_1^T \tag{3c}$$

Therefore, it is reasonable to define the weights for the m characteristics based upon the first principal component, i.e., $\hat{\mathbf{w}} = (a_1, a_2, \ldots, a_m)^T$. Consequently, a characteristic model can be constructed for measuring the favorability f of a given cell:

$$\hat{f} = \hat{\mathbf{w}}^T \mathbf{x} \tag{4}$$

where $\mathbf{x}$ is a vector containing observations of the m characteristics on the given cell.

If the weights are regularized, such that $\Sigma_j \hat{w}_j = 1$ and $\hat{\mathbf{w}}_j \geq 0$ ($j = 1, \ldots, m$), then values of f are valued in the interval [0, 1] for binary-transformed variables, or in the interval [−1, 1] for ternary-transformed variables. For a cell outside the model area, values of f close to 1 indicate high degrees of match with the model and hence are judged to be highly favorable; values of f close to −1 (for ternary variables) or 0 (for binary variables) indicate low degrees of match with the model and therefore are judged to be unfavorable; values of f close to zero (for ternary variables) indicate neither high degrees or low degrees of match, and consequently, are judged to be of undefined favorabilities.

OPTIMUM CUTTING TECHNIQUE FOR ORDERED GEOLOGICAL SEQUENCES

In mineral resources appraisal, it is important and useful to examine the ordered mineral resource values (such as tonnage of ore, average grade, number of mineral occurrences, etc.) with a set of standard sample points. These sample points are collectively referred to as the ordered variation sequence (OVS). In order to identify controlling properties (directions and magnitudes) of a quantitative variable to mineral resource potential in space, mineral resource descriptors should be correlated to patterns of characteristics in the OVS. Such associations allow one to examine a restricted set of combinations of features among known mineral deposits. Because of a great diversity of geological characteristics and complexity of mineral resource distributions in space it would be extremely difficult, if not impossible, to quantify the explanatory power of each attribute or attribute combination for all possible outcomes of mineral resource values. By using a cutting technique for ordered data, it is only necessary to investigate the discriminating power of each attribute or attribute combination for those ''critical'' mineral resource values.

Consider the subjective classification in which a known mineral deposit of a specific genetic type is placed into one of a set of categories, such as large, medium, and small. While reasonable and intuitive, such a procedure cannot generally give an optimum solution when the objective is resource prediction by geologic attributes (Pan and Harris, 1990). Classification should be based upon optimizing in some way the associations between geological attributes and an ordered variation sequence of mineral resources. Clearly, looking for the cutoffs in an OVS can be considered to fall within the scope of optimum cutting techniques. Optimum cutting techniques for ordered data so far have been constructed using the Fisher's criterion (called F-method) (see Fang and Pan, 1982), which determines the cutoffs in a sequence such that the sum of variances within each of the groups is minimized. However, in view of the particular features of mineral resources appraisal, the F-method possesses several important disadvantages:

(1) Data employed in mineral resource estimation are both qualitative and quantitative. Values assigned to qualitative attributes are only nominal. For example, the opposite values like 1 and -1 do not imply that one is greater than the other; instead, they can be equally important in the prediction of mineralization.

(2) The natural cutoffs should be determined on the basis of associations of geological characteristics, instead of considering only the contributions of variances separately as is done in the F-method. Allowance should be made for the possibility that different scales of mineral resources correspond to different subsets of characteristics.

(3) The cutting procedure should accommodate the so-called ''light-head and heavy-feet,'' i.e., few large deposits and many small deposits. Since the F-method is designed to cut a sequence evenly, i.e., cutoffs identified are such that the number of sample points in the groups tends to be equal (Fang and Pan, 1982). Cutting highly skewed distributions by the F-method may lead to absurd results.

Considering all of the shortcomings of the F-method, a new optimum cutting technique, called the P-method, has been developed by Pan (1985b). Assume that we have m geological characteristics and n sample points ranked to an order variation sequence in terms of some mineral resource descriptor (e.g., tonnage). Let $\mathbf{X} = (x_{pq})$ be the $n \times m$ data matrix, where x_{pq} s are binary values defined in an earlier section. Let $\{i, \ldots, j\}$ $(1 \leq i < j \leq n)$ denote the section of sample points from i to j in the variation sequence. Define

$$D(i, j) = m^{-1} \sum_{t=1}^{m} \left\{ \sum_{l=i}^{j-1} \sum_{k=l+1}^{j} [x_{lt} x_{kt} - (1 - x_{lt})(1 - x_{kt})] \right\}$$

$$= D_0(i, j) + \sum_{t=1}^{m} \sum_{l=1}^{j-1} \sum_{k=l+1}^{j} (x_{lt} + x_{kt}) \tag{5}$$

where $1 \le i < j \le n$, $D(j, j) = 0$, $1 \le j \le n$, and $D_0(i, j) = 2^{-1}(j - i)(j - i + 1)$. The quantity $D(i, j)$ is called the consistent degree in section $\{i, \ldots, j\}$, measuring the similarities among the sample points in section $\{i, \ldots, j\}$ when the characteristics are binary. Based upon this measure, a new cutting method (the P-method) is proposed in this section. The P-method described below maximizes the global consistency among the m characteristics within each group and minimizes the similarities between different groups.

Suppose that the ordered sequence is cut into g sections. The criterion is to search for cutting points $1 \le p_1 < p_2 < \ldots < p_{g-1} < n$ such that the following quantity is maximized:

$$Q_{E_g}(g, n) = D(1, p_1) + D(p_1 + 1, p_2) + \cdots + D(p_{g-1} + 1, n) \tag{6}$$

where $E_g = \{p_1, \ldots, p_g - 1\}$ is the set of cutoffs. The criterion (6) may also be expressed more concisely as

$$Q_{\hat{E}_g}(g, n) = \max_{E_g} \{Q_{E_g}(g, n)\}$$

where $\hat{E}_g$ is the set of optimum cutoffs. The algorithm for searching is described below.

First, calculate the matrix of consistent degree

$$\mathbf{D} = \begin{pmatrix} D(1, 1) & D(1, 2) & \cdots & D(1, n) \\ D(2, 1) & D(2, 2) & \cdots & D(2, n) \\ \vdots & \vdots & & \vdots \\ D(n, 1) & D(n, 2) & \cdots & D(n, n) \end{pmatrix}$$

Then, let

$$Q_{\hat{E}_1}(1, j) = D(1, j)$$

$$\hat{E}_1 = \{j\}, j = 1, 2, \cdots, n,$$

and

$$Q_{E_g}^{(i)}(g, j) = Q_{\hat{E}_{g-1}}(g - 1, i) + D(i + 1, j)$$

where $i = g - 1, g, \ldots, j - 1; j = g, g + 1, \ldots, n; g = 2, 3, \ldots, n$. Finally,

$$\begin{aligned} Q_{\hat{E}_g}(g, j) &= \max_{g-1 \leq i \leq j-1} [Q_{\hat{E}_{g-1}}(g - 1, i) + D(i + 1, j)] \\ &= \max_{g-1 \leq i \leq j-1} [Q_{\hat{E}_g}^{(i)}(g, j)] \end{aligned} \tag{7}$$

When $j = n$ in Eq. (7), $\hat{E}_g$ will be the set of optimum cutoffs. Accordingly, the ordered sequence of n sample units would be optimally cut into g groups: $\{1, \ldots, p_1\}, \{p_1 + 1, \ldots, p_2\}, \ldots, \{p_{g-1} + 1, \ldots, n\}$.

DECOMPOSED AND WEIGHTED CHARACTERISTIC ANALYSIS MODEL

More often than not, mineral resources are closely associated with a variety of related geological characteristics and their combinations. Therefore, quantifying mineral resource descriptors requires the detailed investigation of the relationships between geological characteristics. Such an idea is the basis of the new technique, decomposed and weighted characteristic analysis (Pan, 1985b). The DWCA model is constructed as follows.

(1) Construct matrix **X** by arranging the n samples into an ordered variation sequence of the selected resource feature or endowment descriptor. Use the P-method to cut the sequence into g groups. The number of g is appropriately determined on the basis of the $Q - g$ curve and the number of samples in each partitioned group.

(2) Construct decomposed characteristic model of variables that have been refined vis-à-vis the g groups obtained in the first step. The word "refine" indicates the further optimization of variable selection by auxiliary subjective or objective knowledge about relevant geological processes. Each characteristic is then given g weights by g decomposed models, that is, $\mathbf{w}_j^T = (w_{1j}, w_{2j}, \ldots, w_{gj}), j = 1, 2, \ldots, m$, where w_{kj} is generated from the ordinary characteristic analysis based only on the samples in the kth group.

(3) Calculate the average value of mineral resource descriptor (e.g., tonnage) in each group so that we have $\bar{\boldsymbol{v}}^T = (\bar{v}_1, \ldots, \bar{v}_g)$. Establish regression models:

$$E[\bar{\boldsymbol{v}}] = \beta_0 + \beta_1 \mathbf{w}_j, j = 1, 2, \ldots, m \tag{8}$$

(4) Reject, according to the results of regression analysis, two types of variable: (a) variables which are not significantly correlated with the mineral resource descriptor based upon some statistical criterion, e.g., t-statistics, (b) variables that are significantly correlated with the mineral resource descriptor, but the largest weights are too small, e.g., less that 0.1. Note that the justifi-

cation of a weight as "too small" is based upon the relative size of the coefficients. By deleting the rejected characteristics, a refined set of variables is obtained. Reassign values to those characteristics that are negatively associated with the mineral resource descriptor, i.e., assigning the negative characteristic a value 0 if it is present in a cell, and 1, otherwise. For convenience, the refined data set is still denoted by **X**.

(5) Construct a weighted characteristic model. Let

$$\mathbf{Y} = \mathbf{HXV}$$

where the weighting matrices $\mathbf{H} = \mathrm{diag}(h_1, h_2, \ldots, h_n)$ and $\mathbf{V} = \mathrm{diag}(v_1, v_2, \ldots, v_m)$ are appropriately selected on the basis of known samples and decomposed characteristic analysis. Thus, the matrix of match coefficients is given by

$$\begin{aligned}\mathbf{Z} &= n^{-1}\mathbf{Y}^T\mathbf{Y} \\ &= n^{-1}\mathbf{V}\mathbf{X}^T\mathbf{H}^2\mathbf{X}\mathbf{V} \qquad (9)\end{aligned}$$

(6) Beginning with matrix **Z**, the weight vector for m characteristics can be obtained in the same way as that of the conventional characteristic analysis (e.g., principal component method) described in an early section. The favorability Eq. (4) is then derived from the DWCA model.

(7) Establish the regression model based on the estimated favorability values in the control areas, i.e.,

$$E[Y] = \alpha_0 + \alpha_1 \hat{f}$$

where Y is the variable related to a mineral resource descriptor such as tonnage, number of occurrences, and average grade. It can also be a function of some mineral resource descriptors. For example, Y could be log-transformed metal tonnages, i.e., $Y = Ln(M)$, where M is the metal tonnage.

(8) Test the statistical significance of the regression models established above. If the regression models are acceptable, they are applied to the estimation of mineral resources in unknown sample units.

Care should be exercised in the selection of weighting matrices **H** and **V**, because an appropriate selection plays a crucial role in the estimation of mineral resources by the new model. Careful examination should also be made of the relationships between favorability f and mineral resource descriptors.

CASE STUDY

The DWCA model developed above is applied to pegmatitic niobite–tantalite deposits in southern China. A region of about 3500 km^2 has been extensively explored for Nb–Ta deposits. In the past decade, there have been more than 30 Nb–Ta deposits discovered and over 2000 pegmatite veins were recorded in this area. The pegmatitic Nb–Ta deposits occur in Precambrian metamorphic rocks and trend in the NE and SW directions. With the help of local geologists, data were collected on 40 geological and geochemical variables on 23 known pegmatitic Nb–Ta deposits. From the 2000+ pegmatitic veins, 67 groups were selected as promising units on the basis of the clustering features of the veins and their spatial distributions. In this case study, the grouped pegmatitic veins serve as the basic sampling units in place of the inter-grid areas used in previous studies.

Optimum Partitioning of Data Set

Of 40 variables, 27 are quantitatively valued and 13 are binary characteristics. With the aid of optimum discretization techniques (Pan, 1985b; Pan and Harris, 1990), the quantitative variables were transformed into binary characteristics. Then, the 23 known deposits were ranked into an ordered variation sequence to form the original data matrix $\mathbf{X}_{23\times 40}$ for the control region. Use of the P-method to cut the ordered variation sequence produced the results listed in Table 1, in which a datum is the sum (Q) of within-group similarities when there are L groups. For example, five ($L = 5$) optimally cut groups produce within-group similarities shown in Table 2, and the sum of these five similarities is 8.522. The relationship of Q to L is displayed in Fig. 1. Determination of

Table 1. Within-Group Similarity by P-Method

L	Q	L	Q	L	Q	L	Q
1	5.07	7	9.19	13	10.57	19	11.37
2	7.24	8	9.50	14	10.72	20	11.44
3	7.83	9	9.77	15	10.85	21	11.47
4	8.20	10	10.03	16	11.00	22	11.49
5	8.52	11	10.23	17	11.12	23	11.50
6	8.87	12	10.41	18	11.25	—	—

Table 2. Grouped Samples by P-Method

Group no.	Sample no.	Similarity
1	{1, 2, 3, 4, 5}	2.14
2	{6, 7, 8, 9, 10, 11}	1.95
3	{12, 13, 14, 15, 16, 17}	2.00
4	{18, 19, 20}	1.22
5	{21, 22, 23}	1.22

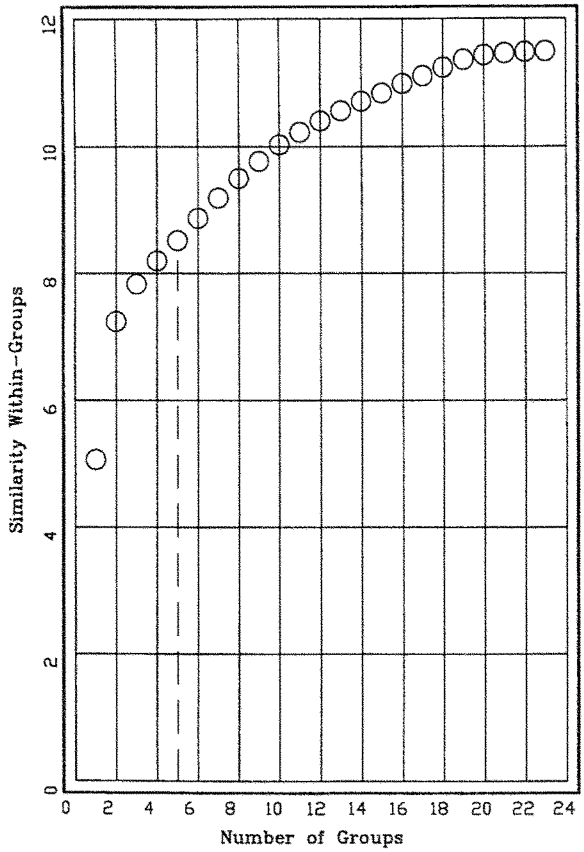

Fig. 1. Curve for the P-method.

the appropriate number of decomposed groups should consider both the features of the *Q–L* curve and the numbers of sample points within each of the groups. In this case, the sample was cut into five groups.

Decomposed Models and Variable Selection

On the basis of the partitioned groups of samples shown in Table 2, five decomposed characteristic models were established by the square root method, generating five weights for each of the 40 variables, which are shown in Table 3. A regression model between the average Nb + Ta metal tonnage in each group of samples and the weights from the decomposed models is estimated for each variable. A geologic characteristic is considered important if the regression

Table 3. Decomposed Characteristic Models

Variable	M1	M2	M3	M4	M5
1	0.215	0.347	1.000	0.415	1.000
2	1.000	0.739	0.000	0.726	0.000
3	1.000	1.000	0.398	0.375	0.000
4	1.000	1.000	0.720	1.000	0.000
5	1.000	0.869	0.404	0.415	0.000
6	0.818	0.347	0.265	0.415	0.000
7	0.818	0.711	0.905	1.000	0.725
8	0.000	0.708	0.630	0.726	0.439
9	0.625	0.211	0.000	0.000	0.459
10	0.632	0.340	0.775	0.000	0.000
11	0.811	0.739	0.852	0.728	0.439
12	0.632	0.211	0.243	0.000	0.000
13	0.233	0.866	0.390	0.726	0.725
14	0.818	0.359	0.443	0.415	0.000
15	1.000	0.681	0.630	0.726	0.000
16	1.000	1.000	0.759	1.000	0.619
17	0.625	0.866	0.596	1.000	0.419
18	0.233	0.211	0.000	0.000	0.000
19	0.625	0.866	0.596	1.000	0.419
20	1.000	1.000	0.672	0.000	0.784
21	1.000	0.540	0.408	0.000	0.459
22	0.818	0.398	0.243	0.428	0.784
23	0.632	0.375	0.443	0.415	0.459
24	0.000	0.353	0.183	0.000	0.000
25	1.000	0.519	0.243	0.415	0.000
26	0.000	0.576	0.243	0.656	0.419
27	0.233	0.576	0.516	0.726	0.439
28	0.000	0.661	0.390	0.000	0.439
29	0.415	0.409	0.609	0.726	0.784
30	0.215	0.000	0.183	0.000	0.000
31	0.233	0.576	0.265	0.428	0.784
32	0.619	0.000	0.000	0.428	0.725
33	0.439	0.211	0.527	0.000	0.419
34	0.000	1.000	0.443	0.726	0.662
35	0.233	0.366	0.422	0.428	0.784
36	1.000	0.211	0.000	0.000	0.000
37	0.811	0.000	0.483	0.359	0.000
38	1.000	0.519	0.000	0.415	0.000
39	1.000	0.519	0.000	0.000	0.000
40	1.000	0.211	0.375	0.000	0.000

model corresponding to this variable is statistically significant. Several features in Table 3 merit a comment.

1. Each variable shows different weights in different scales of mineral deposits. Some of these variables have weights which change approximately monotonically with respect to the size of ore deposits, while others do not. Most of the variables are positively correlated to the Nb + Ta metal tonnages, since their weights increase as the Nb + Ta tonnages in the different sample groups. Only a few samples are negatively correlated to the Nb + Ta tonnages, such as characteristics 1, 29, and 35.

2. Some variables having nearly constant weights across different decomposed models are not really useful in making quantitative estimation of mineral resources since they carry little information on variation in size of ore deposits. For instance, variable 16 has very large weights in every scale of ore deposits. Conversely, some variables with relatively small weights discriminate strongly between one or more magnitudes of deposit size, e.g., variable 18.

3. Global averages of weights in each of the five models display a very clear monotonic-changing pattern with respect to the size of ore deposits, i.e.,

$$\overline{w}_I > \overline{w}_{II} > \overline{w}_{III} > \overline{w}_{IV} > \overline{w}_V$$

where $\overline{w}_I$ is the mean of the estimated weight across all variables. This fact indicates that most of the useful variables are positive features, and that for different sizes of ore deposits the variables make different contributions to the model.

To summarize, the decomposed models show that (a) a set of variables should be refined under statistical criteria, (b) properties of elements in the set should be analyzed in detail, and (c) sample points with different scales of ore deposits should be treated differently in order to make a quantitative estimation of mineral resources.

In a consequent step, regression Eq. (8) was employed to select useful variables on the basis of data in Table 3 (note that numbers in Table 3 are the weights computed from the decomposed characteristic models for all five individual groups of samples; the "M1," etc., represent different models). Given the 10% significance level ($\alpha = 0.1$), we obtained the cutoff correlation coefficient $r_{0.1} = 0.67$. Of the 40 variables, 20 are statistically significant by means of this test, but only 19 of them are important, because variable 18 should be rejected, since the weights associated with this variable are all very small. The selected variables are listed in Table 4.

Because variables 1, 29, and 35 are negative characteristics, their values are reassigned as that described in an early section.

Weighted CA Models and Resource Estimation

The next step is the construction of weighted characteristic models on the basis of the decomposed models. The first task in this stage is to find appropriate weighting matrices **H** and **V**. It is intuitively reasonable to choose the following:

Table 4. The Refined Variable Set

New no.	Old no.	*R*	Name
1	1	−0.688	Wall rock I
2	2	0.687	Wall rock II
3	3	0.926	Pegamat. veins
4	4	0.925	Wein relations
5	5	0.966	Vein length
6	6	0.837	Max (L/W)
7	10	0.714	Fold features
8	11	0.739	Lense veins
9	12	0.903	Stratiform ore
10	14	0.860	Struc. number
11	15	0.839	Strong Nb Alt.
12	21	0.718	Ta_2O_5
13	25	0.895	Sn
14	29	−0.963	Ti
15	35	−0.901	Y
16	36	0.807	Sn∧Mn anomaly
17	38	0.799	Sn/Y∧Sn anomaly
18	39	0.886	Sn/La∧Sn anomaly
19	40	0.847	Sn/Y∧Mn anomaly

$$\mathbf{H}^2 = \text{Diag}(\ln M_1, \ln M_2, \ldots, \ln M_n) \tag{10}$$

where M_i ($i = 1, 2, \ldots, n$) are the sums of Nb and Ta metal quantities for known sampled units. And

$$\mathbf{V} = \text{Diag}(v_1, v_2, \ldots, v_m) \tag{11}$$

where $v_j = [(r_j^2 - r_{0.1}^2)/(1 - r_{0.1}^2)]^{1/2}, j = 1, 2, \ldots, m$ and r is the correlation coefficient between weights and log metal tonnages. These weights are listed in Tables 5 and 6.

Then, ten weighted characteristic analysis models are established, one model for each of the following numbers of control samples: 14, 15, 16, 17, 18, 19, 20, 21, 22, and 23. By comparing their results, two of them for the schemes of 17 and 22 sampled units are selected as the final models. For comparison purposes, the intimate probabilistic characteristic models (Pan, 1985a,b) were also estimated on the basis of the same control samples. Only three "best" models were used. The nomenclature for these models is given as follows: (1) SR17: model estimated by square root method based on 17 samples, (2) PP17: model estimated by principal component method based on the intimate probabilities and 17 samples, and (3) SQ22: model estimated by square root method based on 17 samples.

Table 5. Weights for Control Samples

No.	Value	No.	Value	No.	Value
1	7.21	9	3.42	17	1.43
2	7.20	10	3.25	18	0.85
3	5.78	11	2.51	19	0.80
4	5.29	12	2.26	20	0.42
5	5.06	13	2.00	21	0.21
6	4.89	14	1.95	22	0.15
7	4.32	15	1.87	23	0.07
8	3.59	16	1.80	—	—

Table 6. Weights for Characteristics

No.	Value	No.	Value	No.	Value
1	0.25	8	0.46	15	0.72
2	0.25	9	0.84	16	0.65
3	0.89	10	0.70	17	0.92
4	0.42	11	0.72	18	0.75
5	0.35	12	0.39	19	0.49
6	0.72	13	0.84	—	—
7	0.35	14	0.94	—	—

The estimated results for these models are collected in Table 7. Based on the control samples, the favorability value estimated from the three characteristic models are associated with the sum of Nb and Ta metal quantities by simple linear regression analysis:

$$\text{SR17: } \ln M = 1.272 + 5.350\hat{f} \tag{12a}$$

$$\text{PP17: } \ln M = 1.574 + 5.162\hat{f} \tag{12b}$$

$$\text{SR22: } \ln M = -0.479 + 7.548\hat{f} \tag{12c}$$

According to the F-test, all of the three regression models are highly significant at 1% level. Combination of the point estimate with its error produces an interval estimate, assuming that the quantity of Nb + Ta is lognormally distributed. These results are displayed in Fig. 2.

Using models (12a, b, and c), the metal quantities of Nb + Ta mineral resources in the selected 55 unknown pegmatitic veins were estimated. Table 8 shows the average estimates in order scales for the unknown veins.

Table 7. Estimated $\hat{f}$ in the Control Samples

Veins	SR17	PP17	SR22
1	1.000	1.000	1.000
2	1.000	1.000	1.000
3	0.940	0.931	0.951
4	0.710	0.710	0.751
5	0.727	0.668	0.766
6	0.649	0.550	0.550
7	0.468	0.449	0.696
8	0.324	0.249	0.553
9	0.162	0.159	0.424
10	0.280	0.266	0.294
11	0.022	0.002	0.400
12	0.050	0.002	0.174
13	0.160	0.007	0.200
14	0.246	0.169	0.297
15	0.352	0.285	0.362
16	0.097	0.002	0.460
17	0.001	0.015	0.238
18	—	—	0.157
19	—	—	0.210
20	—	—	0.001
21	—	—	0.084
22	—	—	0.003
R^2	0.945	0.954	0.941
F	126.1	152.8	155.1

COMPARISONS AND COMMENTS

In order to demonstrate the effects of decomposition and weighting as described in the preceding section, consider performing traditional characteristic analysis based upon the match matrix on each of the two subsets of 17 and 22 samples and computing the correlation of the resulting favorability indexes on quantity of metal (Nb + Ta). This was done separately for the square root (SQ) and principal components (PC) methods, and the results are provided in Table 9 under the column heading of CAMM. Similar analyses were performed for decomposed and weighted characteristic analysis, based upon the match matrix; these results are shown in Table 9 under the column heading of DWCA. Comparison of correlations for DWCA to CAMM emphasizes the great value of decomposition and weighting when the objective is the estimation of mineral resources. Table 9 also shows, for comparison purposes, the correlations of favorability to quantity of metal for traditional and decomposed characteristic analysis based upon the intimate probability matrix (Pan, 1985a) under the

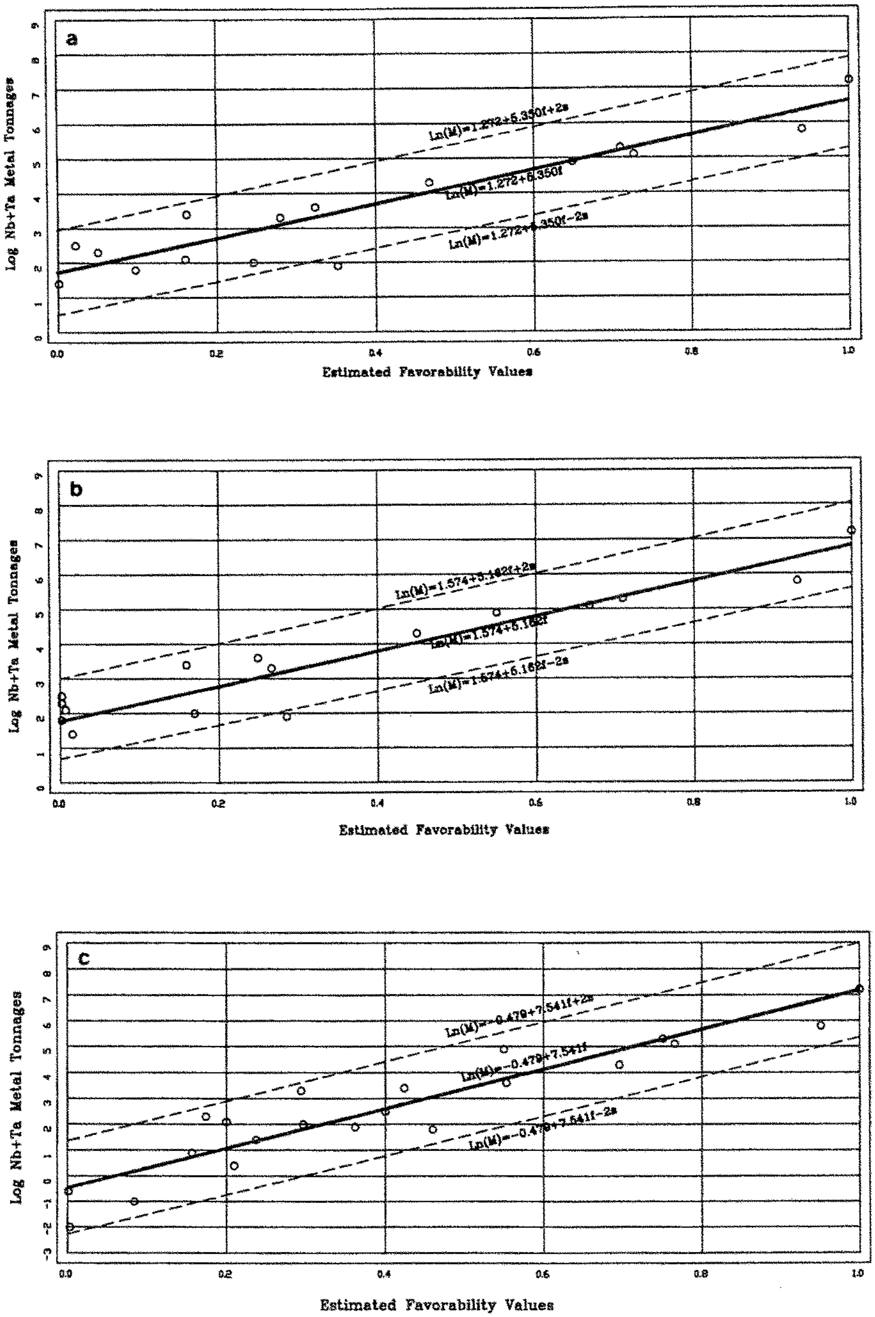

Fig. 2. Regression models based on the DWCA analysis: (a) SR17, (b) PP17, and (c) SR22.

Table 8. Predicted Nb + Ta Quantities in Unknown Veins

$\hat{f}$	Nb + Ta (tons)	Vein numbers
>0.75	>200	1
0.63–0.75	100–200	2
0.49–0.63	50–100	1
0.18–0.49	10–50	18
<0.18	<10	33

Table 9. Comparison of Several CA Models by Correlation of Favorability with Logarithm of Quantity of Metal Nb + Ta

Size	By	CAMM	CAIP	DPMM	DWCA
17	SQ	0.780	0.826	0.954	0.945
17	PC	0.699	0.868	0.914	0.911
22	SQ	0.823	0.879	0.901	0.941
22	PC	0.605	0.610	0.890	0.919

headings of CAIP and DPMM, respectively. These results clearly suggest that decomposed characteristic analyses (DPMM and DWCA) provide the best estimates, and for the 22-sample model, DWCA appears to be superior to DPMM. Several points merit a comment:

1. The new model, DWCA, created by decomposing and weighting, appears to be considerably improved over previous versions of characteristic analysis. The decomposition helps much in refining the useful variables and revealing the power of the variables in discriminating different sizes of mineral resources. Weighting enhances the contributions from different variables and sample units with different scales of mineral resources. The combination of the two kinds of weight increases considerably the potential power of the characteristic analysis for quantitative mineral resources appraisal.

2. Table 9 shows that the decomposition analysis is the most critical step in improving the quantitative estimation of mineral resources by characteristic analysis. The decomposition helps to refine the geologic characteristics by omitting unimportant variables. It should be noted, however, that decomposition can be done only when the control area contains a sufficiently large number of known samples. The P-method plays a central role in the partitioning of the entire data set, which contains the ordered samples in terms of known metal quantities. The P-method is optimum in the sense that the partitioned samples have the largest similarities within each group and have the smallest similarities between

groups. The best number of partitioned groups is selected on the basis of the pattern of within-group similarity with number of groups and the number of samples within each partitioned group. The choice of the group number may be somewhat subjective, when the similarity curve does not exhibit a unique inflection point.

3. The pre-weighting of the original data is also important in utilizing the different contribution of each individual sample or characteristic to the quantitative estimation of mineral resources. The choice of the weighting matrices for samples and variables in this study is generally appropriate for many other mineral resource estimations: metal tonnages are usually available in control areas. The decomposed characteristic models provide the weights for geologic variables based on simple regression analysis. Of course, other choices for the weights are also acceptable as long as the weights are reflective of the importance of different samples or variables. For example, the weights might be obtained from expert geologists who are knowledgeable on the geology and ore deposits of the study region.

4. The application of regression model to the decomposed and weighted characteristic analysis provides a means to quantitative estimation of mineral resources. Since the favorability value computed at each sampled unit (cell) not only characterizes the likelihood of the occurrence of mineral deposit, but also quantifies the magnitude of metal concentration, the regression model of the relationships of metal tonnage to favorability value established in the control region is applicable to the estimation of discoverable resources in unknown sampled units.

REFERENCES

Botbol, J. M., 1971, An Application of Characteristic Analysis to Mineral Exploration: Proc. 9th Ins. Sym. on Techniques for Decision-Making in the Mineral Industry, Special Vol. 12, Canadian Inst. of Mineral and Metallurgy, Montreal, Canada, p. 92–99.

Botbol, J. M., Sinding-Larsen, R., McCammon, R. B., and Gott, G. B., 1978, A Regionalized Multivariate Approach to Target Selection in Geochemical Exploration: Econ. Geol. v. 73, p. 534–546.

Fang, K., and Pan, E., 1982, Clustering Analysis, Geol. Publ. House, Beijing, 281 p. (in Chinese).

Harris, D. P., 1984, Mineral Resources Appraisal—Mineral Endowment, Resources, and Potential Supply: Concepts, Methods, and Cases: Oxford University Press, New York, 455 p.

McCammon, R. B., Botbol, J. M., Sinding-Larsen, R., and Bowen, R. W., 1983, Characteristic Analysis—1981: Final Program and a Possible Discovery: Math. Geol. v. 15, p. 59–83.

Pan, G. C., 1985a, Several Forms of the Intimate Degrees in Multidimensional Scaling and the Related Models of Mineral Resources Appraisal: A Special Selection for 27th Int. Geol. Cong., v. 6, p. 231–250 (in Chinese with English abstract).

Pan, G. C. 1985b, Quantitative Estimation of Mineral Resources with Case Studies on the Pegmatitic Nb–Ta Deposits in China: M.S. thesis, Changchun College of Geology, China, 141 p.

Pan, G. C., and Harris, D. P., 1990, Three Nonparametric Techniques for the Optimum Discretization of Quantitative Geological Variables: Math. Geol., v. 22, p. 699–722.

Pan, G. C., and Wang, Y., 1987, Weighted Characteristic Analysis and Its Application to Quantitative Estimation of Pegmatitic Nb–Ta Mineral Resources: Geol. Prospect., v. 23, p. 34–42 (in Chinese with English abstract).