

## COMPUTER-BASED TECHNIQUES IN THE COMPILATION, MAPPING AND INTERPRETATION OF EXPLORATION GEOCHEMICAL DATA

**R.J. Howarth**

Applied Geochemistry Research Group, Imperial College of Science and Technology, London, England

**L. Martin**

Computer Applications and Systems Engineering, Rexdale, Ontario

*Howarth, R.J. and Martin, L., Computer-based techniques in the compilation, mapping and interpretation of exploration geochemical data; in Geophysics and Geochemistry in the Search for Metallic Ores; Peter J. Hood, editor; Geological Survey of Canada, Economic Geology Report 31, p. 545-574, 1979.*

### Abstract

Computer methods used for planning, quality control, presentation, and interpretation of exploration geochemical data may be broadly classified into purely numerical methods and mapping, in which sample location is considered. This paper reviews both current industrial and governmental practice, and research techniques of potential application.

Correct organization of field sampling, and laboratory quality control will minimize bias in the data. The majority of industrial surveys are still conducted on a single-element interpretation basis. National geological survey agencies have a growing awareness of the utility of regional multi-element studies for resource appraisal, and data-base management and retrieval systems greatly assist operation of these larger exploration programs. Data treatment is consequently mainly limited to derivation of the mean, standard deviation, etc. and histograms, although computer-based data transformation and analysis of frequency distributions can improve definition of thresholds and class boundaries for mapping. Geochemical maps usually consist of graded symbols at the sample locations, although contour-like maps are preferable for display of the regional geochemical patterns and are often based on moving average smoothing. Proper attention to cartographic design can greatly improve the geochemist's acquisition of information from maps. Multivariate statistical analysis is mainly used to obtain inter-element correlations, to correct for possible Mn/Fe scavenging, and (less frequently) for determination of element associations by principal components or factor analysis, and discriminant or cluster analysis.

While in some cases the lack of suitable computer hardware has limited the application of these methods, in others the restraining factor is the lack of appreciation of their power and flexibility when used intelligently. Geochemists are now gaining a better understanding of the range of statistical and mapping techniques available to them and powerful, relatively inexpensive, minicomputers are becoming more widely available. It is to be hoped that the standard of quantitative analysis of geochemical data will improve in consequence.

### Résumé

Les méthodes de l'informatique utilisées pour la planification, le contrôle de la qualité, la présentation et l'interprétation des données d'exploration géochimique peuvent, d'une manière générale, regrouper les méthodes purement numériques et la cartographie, qui tient compte de la localisation des échantillons. Cette étude examine les pratiques gouvernementales et industrielles courantes et les techniques d'application potentielle de la recherche.

Une organisation convenable de l'échantillonnage sur le terrain et du contrôle de la qualité en laboratoire, permettra de minimiser les erreurs au niveau des données. La majorité des levés industriels sont encore effectués d'après une base d'interprétation qui ne tient compte que d'un seul élément. Les levés géologiques à l'échelle nationale tiennent de plus en plus compte de l'utilité des études régionales portant sur plusieurs éléments, pour l'appréciation des ressources; d'autre part, les systèmes d'extraction et de gestion des fichiers de données aident beaucoup au fonctionnement de ces plus importants programmes d'exploration. En conséquence, le traitement des données vise surtout à l'obtention de moyennes, d'écart types, etc., de même que d'histogrammes, même si la transformation des données et l'analyse des distributions de fréquences à l'aide d'ordinateurs peut améliorer la définition des seuils et des limites de classes pour fins de cartographie. Sur les cartes géochimiques, les points d'échantillonnage sont indiqués par des symboles tramés; cependant, les cartes de type topographique sont mieux adaptées à la représentation des profils géochimiques régionaux et se fondent souvent sur le lissage des moyennes mobiles. Le fait de consacrer l'attention qui convient à la conception cartographique peut de beaucoup améliorer, pour le géochimiste, l'acquisition de renseignements à partir de cartes. L'analyse à plusieurs variables est surtout utilisée pour l'obtention de corrélations entre éléments, pour compenser la contamination Mn/Fe possible et (moins souvent) pour la détermination d'associations d'éléments à l'aide de leurs principales composantes ou analyse factorielle, et pour l'analyse discriminante ou de dispersion.

Tandis que dans certains cas le manque de matériel informatique approprié a limité l'application de ces méthodes, dans d'autres le facteur restrictif est le manque d'appréciation de leur efficacité et de leur flexibilité lorsqu'utilisées intelligemment. Les géochimistes comprennent actuellement de mieux en mieux la gamme des techniques statistiques et cartographiques disponibles, et des mini-ordinateurs puissants et relativement peu coûteux deviennent de plus en plus accessibles. Il est à espérer que la qualité de l'analyse quantitative des données géochimiques en sera en conséquence améliorée.

## INTRODUCTION

We aim in this paper to review current practice in the acquisition of exploration geochemical data and methods for its subsequent treatment by numerical and mapping techniques as an aid to interpretation. Since it is not possible within the scope of this paper to pursue all the topics raised in equal depth, an extensive bibliography has been included. The review has been divided into four main parts: (1) quality control and methodology in both the field and the laboratory; (2) requirements of the geochemical map, its compilation and presentation; and statistical interpretation of the data using (3) univariate, and (4) multivariate methods.

## QUALITY CONTROL AND METHODOLOGY

### *Quality Control of Laboratory Analyses*

One of the fundamental aspects of any geochemical exploration program is the adequate supervision of laboratory results. It is surprising how many organizations are content to send samples for analyses (either to within-house or contractor laboratories) with little or no attempt to monitor analytical precision or accuracy. With luck, analytical errors may be reflected by spurious patterns on the map which lead to identification of the erroneous data (e.g. Lockhart, 1976) but this is by no means a certainty. Various authors have drawn attention to the importance of randomizing samples prior to submission to the laboratory in order to convert systematic to random errors, and to facilitate recognition of an abnormal situation (Miesch, 1964, 1967b, 1971; Plant, 1973; Plant et al., 1975; Howarth, 1977). It may be argued that it is not convenient during a rapid reconnaissance to randomize all samples prior to analysis. In a long term program, such as that practised by a national survey, it may be possible, however, to randomize all samples for one map sheet at a time. Randomization of samples (and preferably standards and duplicates also) at the batch level will certainly be more beneficial than not doing so. One should always attempt to define both within- and between-batch error levels, since the latter is often the larger effect. It should be realized that careful choice of suitable standards (Allcott and Lakin, 1975; Hill, 1975a, b; Plant et al., 1975) and their submission at random locations in the analytical stream is an integral part of the control process. It may be argued that randomization could lead to sample renumbering errors when passing through the laboratory. Use of pre-randomized number sequences in the field which are subsequently sorted into sequential order for laboratory submission (Plant, 1973) have been found very satisfactory (Plant et al., 1975) and with adequate supervision of personnel no problems have arisen with more complex laboratory-based schemes (e.g. Howarth, 1977). The unequivocal identification of laboratory-induced error in maps from unrandomized samples (e.g. analyzed in traverse order) is not possible in many cases; even if it is suspected, later subtraction of such bias may be impossible since the contribution of local spatial variation will not be known exactly. Whether the additional time and care involved in sample randomization can be justified in a rapid exploration survey on the grounds of cost is a different question, but from the statistical and interpretational viewpoint it is undeniably highly desirable.

Because of the difficulty of obtaining suitable standards in situations where a multiplicity of rock types or stream sediments may be analyzed together (aspects to be considered are: amount of standard; lithological and mineralogical similarity to samples; adequate concentration range for elements of interest; long-term segregation of bulk standard, etc.), attention has turned increasingly to the use of duplicate

analyses as an additional control method (Garrett, 1969, 1973a; James, 1970; Michie, 1973; Plant et al., 1975; Howarth and Thompson, 1976; Miesch, 1976b; Thompson and Howarth, 1976, 1978). Recent simulation studies (Howarth and Thompson, 1976) have shown that many of the standard laboratory practices in recording results (such as rounding up to integer ppm values; quoting negative machine readings, normally resulting from a statistical estimation of a near-zero concentration, as zero; or values below some presumed detection limit as 0.5 of this limit) prior to determining accuracy and precision will lead to an over-optimistic bias in the estimation of these parameters. The practice of quotation of within-batch precision rather than between-batch values also leads to optimistically biased results. However, any indication of analytical precision is better than the more usual situation of none.

A rapid method of control based on duplicate analyses and suitable for long production runs, which leads to a measure of total laboratory error variability (i.e. analytical error plus subsampling error) in terms of precision as a linear function of concentration has been described elsewhere (Thompson and Howarth, 1976, 1978). This model is in contrast with the usual assumptions of either constant absolute error (standard deviation) or constant relative error (coefficient of variation) and appears to conform well with most laboratory situations. Analysis of variance techniques are still in most common use for estimating and partitioning field and laboratory sources of error and are discussed below. Other types of control chart methods based on estimates of analytical precision are also in use (e.g. Hill, 1975b).

In many cases large-scale exploration programs are concerned with routine analysis of less than five elements. However, the increasing availability of rapid multi-channel analytical techniques, such as plasma-source emission spectrography, means that problems of efficient multi-element quality control must be solved for upwards of 20 elements at a time. This is a challenging situation and adequate criteria for batch rejection and re-analysis on a multi-element basis are currently being investigated in a number of laboratories.

### *Geochemical Sampling*

There is now increasing awareness of the effect of sampling strategies in geochemical exploration, and this has placed emphasis on sampling media for which an individual sample acts as an indicator for a large geographical area (although the samples are taken from a single location). For example, stream sediments, lake sediments, stream or lake waters, and tills may all derive from relatively large "catchment" areas, whereas soil and rock samples generally are representative of more localized areas.

While it is difficult to attach actual costs to particular types of survey method, this has been done in some cases. For example, Cameron and Hornbrook (1976) illustrate cost functions in relation to regional lake geochemical sampling. Undoubtedly any sampling medium with an inbuilt vector property will prove advantageous in a reconnaissance situation. Aspects of field sampling methodology have been discussed by Miesch, 1964, 1967b, 1971, 1976a, b; Garrett, 1969, 1973a, 1977; DeGeoffroy and Wu, 1970; Dahlberg, 1971; Howarth and Lowenstein, 1971, 1972; Kayser and Parry, 1971; Plant, 1971; Smith, 1971; Bolviken and Sinding-Larsen, 1973; Hodgson, 1973; Connor et al., 1974; Sharp and Jones, 1975; Sinclair, 1975; Cameron and Hornbrook, 1976; Hawkes, 1976; Chork, 1977; and David, 1977. In many of these cases analysis of variance methods have been used to partition field

and laboratory sources of error (e.g. Garrett, 1969, 1973a, b; Howarth and Lowenstein, 1971, 1972; Plant, 1971; Bolviken and Sinding-Larsen, 1973; Michie, 1973; Plant et al., 1975; Cameron and Hornbrook, 1976; Miesch, 1976a, b; Sinding-Larsen, 1977; and Chork, 1977). However, it has been suggested recently (I. Clark, pers. comm., 1977) on the basis of mine sampling data that the basic assumptions on which classical analysis of variance rests may be sufficiently violated in practice for the results of methods of assessing field variability using this approach to be misleading. This appears to be partly related to spatial distribution of the data as well as to nonconstancy of variance with concentration, although it is not yet certain just how important these effects may be in exploration geochemical data.

**Choice of Sample Spacing**

If prior knowledge exists of the likely target size, shape, and orientation, then exact solutions may be obtained for optimum grid sampling techniques. These are summarized by Sinclair (1975) in a geochemical context. For example, if an elliptical target with major to minor axes of a:b is anticipated, then the grid should be oriented parallel to the axes of the ellipse with spacings of  $\sqrt{2a}$  and  $\sqrt{2b}$  respectively. Singer (1972) gives a computer program for calculating the exact probability of locating an elliptical or circular target with either a square, rectangular or hexagonal grid. More recently Garrett (1977) has illustrated the use of computer simulation methods to determine the detection probabilities for similar targets with irregularly spaced sampling points, such as those more usually encountered in lake or stream sediment sampling plans. His results reinforce the consideration of appropriate choice of sample spacing. For example, an elliptical target with a 1:2 size ratio and major axis length of 5 miles would have an 0.95 probability of detection with a sample density of 1/5 square miles, this falls to 0.50 at 1/10 square miles. However, for a smaller target of the same shape but major axis length of 4 miles the probabilities are approximately 0.70 and 0.20, and with a major axis length of 2 miles they have fallen to below 0.01 in either case. As Sinclair (1975) pointed out, it is the knowledge of target size (which in most exploration programs will be the size of the secondary dispersion geochemical halo) which is unfortunately so difficult to obtain. Hawkes and Webb (1962) suggested 500 feet to 10 miles for drainage and areal soil anomalies, and 5 to 500 feet for localized soil and biogeochemical anomalies as maximum dimensions.

As a result of this difficulty, a recent development has been the interest in selection of optimum sample spacing based upon the spatial properties of the data. This is a reflection of the inherent correlation between values at adjacent sample locations, which will increase as the intersample distance falls, and conversely, at extreme distances the concentrations become uncorrelatable. Hodgson (1973), Dijkstra and Kubik (1975), Sinclair (1975) and Dijkstra (1976), have discussed this from the point of view of autocorrelation.

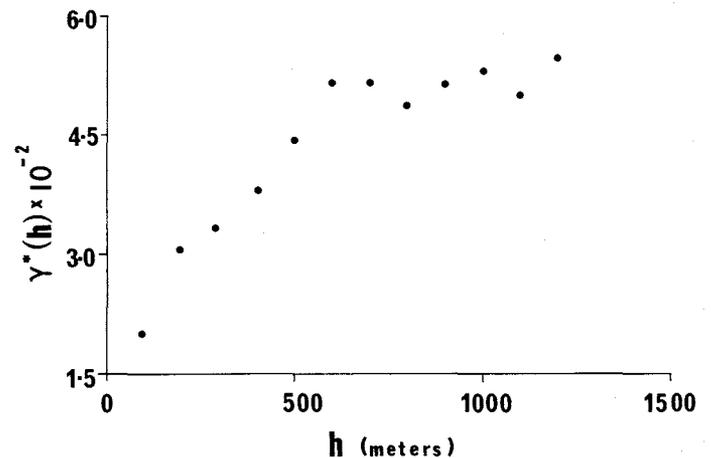
An alternative approach is the application of the geostatistical theory of Professor G. Matheron, founder of the Centre de Morphologie Mathématique, Fontainebleau<sup>1</sup>. His Theory of Regionalized Variables (1957, 1962, 1963) has been successfully applied in the mining industry to ore reserve calculations and optimization of drillhole locations (see for example David and Dagbert, 1975; Journel, 1975; Guarascio et al., 1976; David, 1977; and Alldredge and Alldredge, 1978). In exploration geochemistry, a variety of

investigations are in progress (Sinclair, 1975; Crossant, 1977; S.A.M. Earle and R. Sinding-Larsen, pers. comm., 1977) concerning its utility for the optimization of sample spacing in soil traverses or grids and are showing encouraging results.

Following Matheron's treatment, if  $z_{x_i}$ , the observed element concentration at a location  $x_i$ , is one possible realization of a certain random function  $Z(x_i)$  representing all possible concentration values which could be obtained at  $x_i$ , and  $z_{x_{i+h}}$  is the observed concentration at a point a distance  $h$  away from the first, then the intrinsic hypothesis is that the differences in concentration between all pairs of points separated by distance  $h$  (e.g.  $z_{x_i} - z_{x_{i+h}}$  and  $z_{x_j} - z_{x_{j+h}}$ ) are considered as different realizations of the same random increment  $\Delta(h) = Z(x) - Z(x+h)$ , and are independent of the location of  $x$  and  $x+h$ . This condition of stationarity applies, from the intrinsic hypothesis, to both the mean and variance of  $\Delta(h)$ . If it is also assumed that the mean is zero, then this behaviour may be studied experimentally by the semi-variogram:

$$\gamma^*(h) = \frac{1}{2n} \sum_{i=1}^n [z(x_i) - z(x_{i+h})]^2$$

where  $n$  is the number of pairs of points a distance  $h$  apart. This function is suitable even in the presence of gaps in the data, but (as a rule of thumb) requires preferably at least 50 points in a traverse. Different shapes of semi-variograms may be obtained depending on the nature of the underlying spatial behaviour of the phenomenon (see, for example David, 1977). The most commonly observed pattern with soil data is for  $\gamma^*(h)$  to rise fairly rapidly from a small initial value at the shortest sample spacing used (corresponding to all sources of variation at a scale smaller than that of the sampling interval) until it becomes approximately constant for large values of  $h$ . Figure 24.1 shows an example taken from a study in the Mendip Hills lead-zinc area of southwest England by the Applied Geochemistry Research Group (AGRG) of Imperial College, London. The rising portion of the curve allows an estimation to be made of the range of influence of the element, that is the distance above which the concentration values will be statistically independent.



**Figure 24.1.** Semi-variogram for log(Zn) in B-horizon soil samples over Carboniferous Limestone Series of Mendip Hills, Somerset, England, 70 samples taken at 100 m intervals along traverse. Range of influence is approximately 750 m (Information supplied by S.A.M. Earle, AGRG, 1977).

<sup>1</sup> Where 'geostatistical' is used in this paper the term refers specifically to Matheron's theory rather than to the general usage of statistics in the earth sciences.

Semi-variograms could be made in various directions (e.g. along and across strike) to assess the variability of range with direction, since on average the values of the regionalized variables will become independent at distances exceeding the range. Setting the sample spacing in a subsequent investigation as approximately equal to the range is a useful criterion, but the optimum solutions will vary with the type of semi-variogram. If multi-element data are available, the interval may be set with regard to the overall ranges found for the various elements. Sinclair (1975) pointed out that from the prospecting point of view the critical sampling interval would need to be determined from both background and anomalous areas (and there may be a shortage of data for the latter) since the range for a background region will probably not be related to that from an anomalous area, and geological information on likely target size and shape would still be required. Even if the geostatistical approach is not yet a practical evaluation tool in prospecting, widespread application should give a valuable insight into the ranges of different elements in particular environments, thus leading to a better understanding of the fundamentals of the dispersion phenomena with which we are concerned.

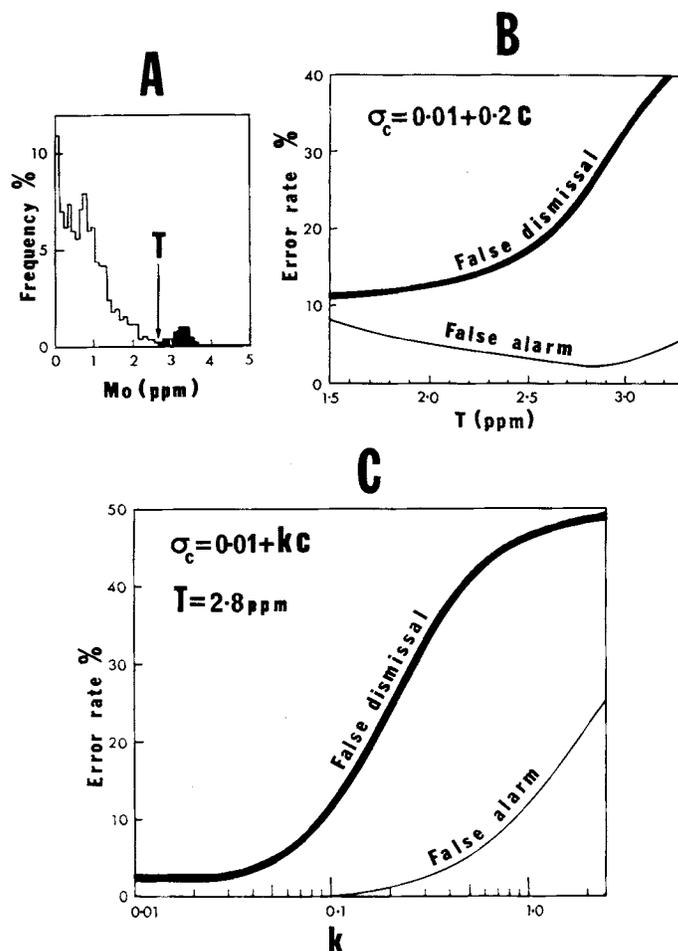
Sharp and Jones (1975) have devised a topologically based method for detailed sampling of a drainage basin stream network using the results of sequential (on-site) analyses which may well be useful for more detailed geochemical exploration, say in the follow-up stage of a regional program. Regional stream sediment surveys are generally planned on the basis of the size of the feature which it is required to delineate.

In all cases the decision whether to operate a phased reconnaissance or a single high density sampling program will depend ultimately on time and cost criteria. The latter approach will in many cases lead to gross oversampling to determine potential target areas, whereas a phased program may extend over more than one field season. Miesch (1976a, b) has devised elegant optimization techniques, based on analysis of variance, to distinguish between very broad scale units pre-defined on a geological or geographical basis, but the resulting sampling densities appear to be rather low for exploration geochemical purposes.

### The Role of Sampling and Analytical Error

It is often assumed in planning a survey that a small target (e.g. geochemical halo) will be successfully recognized with one sample. Naturally, there will be the possibility of additional geochemical, geological or geobotanical information which may assist, but in many instances initial recognition will depend on the presence of an "anomalous" concentration value. Decisions affecting identification of such values will depend not only on the natural geochemical environment (many known orebodies are not distinguished geochemically at the surface), but also on what transforms (if any) are to be applied to the data (see for example the discussion in Govett et al., 1975), and on the variance of the geochemical determinations.

A general linear model for increasing analytical plus sampling variance with element concentration seems to be applicable in a wide variety of laboratory situations (Thompson and Howarth, 1976). It is therefore quite possible as a consequence of this that the observed concentrations from target locations on occasion could be below the threshold set by the investigator to define "anomalous" samples and vice versa. If the form of the increase of variance (which could include field sampling error if duplicates were taken at this level) with concentration is known, then it is possible to predict the probability that background samples will be wrongly classed as anomalous (false alarm), and anomalous samples wrongly classed as



- A) A priori frequency distribution for Mo from primary reconnaissance stream sediment survey, showing 'anomalous' population above threshold, T.
- B) Expected error rates for misclassification of background samples (false alarm) or anomalous samples (false dismissal) in a subsequent survey, assuming a constant analytical system equation, as a function of T.
- C) Expected error rates assuming constant threshold (T) with worsening analytical precision (k in the system equation).

**Figure 24.2.** Example of effect of analytical precision on expected accuracy of geochemical target recognition. For full explanation see text.

background (false dismissal). The false dismissal probability may be obtained by integration of the expected probability distributions for all observed concentrations corresponding to samples whose true concentrations lie above the defined threshold, and the false alarm probability is obtained in a similar manner from samples below this threshold. A normal distribution of errors is assumed in this model.

Consider an example in which an orientation survey for molybdenum in stream sediments has yielded a bimodal frequency distribution (Fig. 24.2A). Analyses have been determined to 0.1 ppm, and the total error variability of the system (including both analytical and subsampling errors) has in this case been established as being adequately represented by the equation  $\sigma_c = 0.01 + 0.2C$ , where  $\sigma_c$  is the standard deviation for the normal error distribution about a true concentration C. Let us further assume that in the area

under investigation it is probable that the maximum concentration which could contribute significantly to the anomalous population in a subsequent reconnaissance survey is expected to be 5 ppm (i.e. we exclude the rare events with concentrations above this from the integration). Figure 24.2B shows the resulting false dismissal and false alarm error rates to be expected subsequently if the threshold for defining the "anomalous" concentrations is changed from 1.5 to 3.3 ppm. Without the unacceptably high false alarm rate from the background population (necessitating many check analyses or follow-up surveys), a significant proportion (24 per cent) of the anomalies could go undetected even were the threshold to be placed at the optimum boundary of 2.8 ppm (Fig. 24.2A). The false alarm and false dismissal rates for the same data are plotted in Figure 24.2C as the slope ( $k$ ) in the equation  $\sigma_c = 0.01 + kC$  is increased from 0.01 to 1.0 (i.e. the errors increase) given a constant threshold of 2.8 ppm. This figure dramatizes the need for laboratory quality control as a prerequisite, but also demonstrates that only when the magnitude of total field plus laboratory error is known can one gain a reliable knowledge of the likelihood of anomaly detection among the population of all sites sampled in an area. This does not, of course, include loss of targets which went undetected because no sample fell within their range of influence. There will always be a danger that subtle anomalies (e.g. related to deeply buried mineralization) may be missed if they are in the "background" range. This chance is minimized if the data are plotted in map form, but careful attention must naturally be given to the selection of class intervals. Even then, no method will guarantee complete success.

## Discussion

It should be apparent that one of the major problems in cost-effective geochemical exploration is the adequate definition of the geochemical target in terms of size, shape, orientation and likelihood of correct recognition on the basis of the chemical analysis of the sample. Long-term acquisition of reliable data from geostatistical studies should assist in definition of the spatial behaviour of the elements of interest, both in background and anomalous areas.

If one wishes to find as many targets of fixed size as possible in a given area from a single field sampling exercise, then there is no recourse but to put in blanket coverage at a suitable density. For a fixed number of samples the density should be made as low and even as possible over the whole area of interest (other factors permitting). However, a phased approach based on an initial lower density regional survey to identify areas more suitable for later detailed examination would be generally preferable. Rapid low-cost analytical techniques are generally considered suitable in the early stages of such work.

A sampling medium which represents a large geographical area from a point-source sample (e.g. lake or stream sediment) will be of assistance in areal reconnaissance of suitable terrain. Where soil or rock sampling has to be undertaken, the price of more localized sample representivity will inevitably be a larger number of samples. Systematic control of field and laboratory practice should help to ensure that a target will stand the best chance of recognition. It should not be overlooked that the increasing availability of multi-element information on a routine basis could well enhance the probability of recognition of a potential target (if only because the joint probability of false dismissals would be lowered).

## Data Storage and Retrieval

The computer forms an ideal medium for the storage of geochemical data and related field observations. Several commercial organizations are known to be investing a great deal of effort in building up data storage and retrieval facilities, since the ease with which data base interrogation may be undertaken is felt to justify the installation of such a facility even without a requirement for sophisticated statistical processing of the information.

Gordon and Hutchison (1974), Hutchison (1975) and Merriam (1976, 1977) include papers describing a wide variety of computer-based systems for storage and retrieval of geological data. Major files of exploration geochemical data are held in computer-compatible form by national or state geological surveys, academic and industrial organizations in many countries. It is becoming more common for field data collected in the course of a geochemical survey to be entered directly on forms designed as keypunch documents for subsequent computer entry (e.g. U.S. Geological Survey, 1969; Garrett, 1974a; Plant et al., 1975; Baucom et al., 1977).

The general philosophy involved in such systems may be gleaned from the papers included in Gordon and Hutchison (1974), Hutchison (1975) and Merriam (1976, 1977). A variety of generalized data base management systems may be bought, or leased, from commercial manufacturers thus ensuring long-term maintenance and updating. However, several organizations with sufficient in-house facilities have preferred to develop systems specifically for geological data. Systems such as GAS (Garrett, 1974b), GRASP (Bowen and Botbol, 1975), GEXEC (Plant et al., 1975; Jeffery and Gill, 1976), SIGMI (Kremer et al., 1976), NDMS (Ferguson et al., 1977) and System 2000 (Davidson and Moore, 1978) are typical of those currently being used for geochemical data files. A wide variety also exist tailored to the requirements of the field geologist or igneous petrologist, with which we are not concerned here. There are those who would argue that data base management packages supplied by major software companies or the computer manufacturers are preferable to home developed packages, on the grounds of reliability, standards of documentation and, perhaps of greatest importance, long term maintenance and update of the system. Against this must be set the fact that such packages are seldom tailored for the geochemist's requirement, particularly if a single system embracing data management, statistical treatment and graphics facilities is required.

In the long run, it is the considerable ease with which statistical data and simple graphical displays and maps can be obtained, using integrated data base management facilities to prepare the requisite information for management decision-making, which will justify the cost and effort required to create and maintain this type of facility. For many of the larger national and industrial organizations, this justification has now been amply proven.

## GEOCHEMICAL MAP COMPILATION

Location is the key element of exploration data. Because the eye is well adapted to recognition of patterns in spatial data, maps have the capacity to present exploration information with great impact, and the spatial component becomes an integral part of the compilation and interpretational process. Comparison of geochemical data with topographic, geological or geophysical information is rendered a much easier task when all are in map form at the same scale, and the geochemical map has thus become a useful and familiar tool to the explorationist.

### Types of Geochemical Map

Map usage can be divided into three broad categories generally related to scale: (a) the regional atlas, at scales upwards of 1:250 000; (b) reconnaissance maps, at scales between 1:25 000 and 1:250 000; and (c) local detail, generally prepared at the follow-up stage at scales between 1:100 and 1:25 000. Within each class variety exists in both the type of data, and the way in which it is presented.

### Regional Geochemical Atlases

While not specifically exploration oriented, this type of atlas provides the broadest view of large portions or the totality of a country. Typically, sampling densities will be well below 1/km<sup>2</sup>, over an area in excess of 10 000 km<sup>2</sup>, and information for 12 to 30 elements will be included, usually presented as moving-average smoothed maps. To the explorationist such maps can provide useful information by defining geochemical 'provinces' within which more intensive exploration surveys can be framed.

In general, three main lines of approach have emerged: firstly, extrapolation of the results of detailed studies in type localities considered to be representative of large areas; secondly, very low-density multi-stage (hierarchical) sampling of large units considered to be relatively distinct; and thirdly, low-density sampling of the entire survey area, maintaining where possible a uniform coverage.

The first two methods have generally been applied to broad units, often related to soil type, and while of value to environmental geochemistry (Tourtelot and Miesch, 1975), have little application from the mineral exploration standpoint.

The third approach, however, can provide a rapid cost-effective tool for the identification of broad regions within which conventional mineral exploration surveys might be applied, and in this context is analogous and strictly complementary to national geological and soil maps and to satellite imagery. Generally speaking, samples representative of a catchment zone (such as lake sediment or stream sediment) have been preferred for this type of survey.

Garrett and Nichol (1967) and Armour-Brown and Nichol (1970) demonstrated the ability of stream sediment sampling, at densities sparser than one sample per 180 km<sup>2</sup>, to identify major mineralized belts in Sierra Leone and Zambia. The regional lake geochemistry sampling program currently being undertaken by the Geological Survey of Canada (Hornbrook and Garrett, 1976; Garrett, 1977) at a mean density of one sample per 13 km<sup>2</sup> has successfully identified a number of major geochemical anomalies, in some cases related to known mineralization, in the 421 000 km<sup>2</sup> covered so far. A regional geochemical atlas of Uganda (Geological Survey of Uganda, 1973) has been completed, based on cuttings from wells bored for water, and stream sediment sampling where suitable drainage exists, at overall mean densities of about one sample per 70 and 260 km<sup>2</sup> respectively. Regional geochemical atlases of Northern Ireland (Webb et al., 1973) and England and Wales (Webb et al., 1978) based on stream sediment sampling at a mean density of one sample per 2.8 km<sup>2</sup> have shown many features of interest from the geological and environmental points of view, in addition to the major mineralized districts.

An extensive hydrogeochemical and stream-sediment reconnaissance of the 7 827 617 km<sup>2</sup> of the conterminous United States and Alaska was initiated in 1975 as part of the U.S. Department of Energy National Uranium Resource Evaluation (NURE) program, with the aim to obtain uranium data for stream and lake water, groundwater and stream and lake sediments. It is hoped to complete the 621 NTMS 1°x2° quadrangles, with average sampling densities of the order

1/13 km<sup>2</sup> and 1/25 km<sup>2</sup> for stream sediment and groundwater samples respectively, by 1983. Currently, some 60 quadrangles have been sampled and are being analyzed for up to 40 elements (Bendix Field Engineering Corp., 1976, 1977, 1978). This project will undoubtedly provide a source of multi-element regional geochemical atlas material of immense interest.

Clearly, the compilation of national multi-element geochemical atlases is a responsibility which lies with government. In the context of resource appraisal (particularly in the Third World), the diversity of requirements for mineral exploration, agriculture, public health, land-use and pollution emphasize the necessity for a multidisciplinary approach aimed at establishing the best methods of acquiring and presenting regional geochemical information, and for research to develop interpretational criteria to meet the differing interests of all the eventual users (Webb and Howarth, in press). In many parts of the world even a preliminary broad-scale coverage will take a considerable time to accomplish. In the meantime, reconnaissance geochemical surveys of very extensive areas have been, and are being, undertaken by national and international agencies and industry.

### Reconnaissance Geochemical Maps

Surveys in this class are mainly undertaken by government agencies and industry. The objectives are to identify areas of economic interest and select specific targets for more detailed follow-up investigation. Physical and climatic conditions, and (in some cases) the degree of man-made interference determine the kind of sampling media used. The more common sampling media include stream sediment, lake sediment, and soil, till and rock where appropriate. Lake, stream and spring or well waters are less commonly used. Sampling densities generally range between 2/km<sup>2</sup> and 1/25 km<sup>2</sup>. Semi-airborne geochemical surveying techniques (such as the Barringer Research Surtrace method based on sampling the surface microlayer from the air) hold considerable promise for the future.

The most valuable and commonly produced geochemical map types show firstly, the sample location and secondly, the element analytical value at the sample site. These are being supplemented with increasing frequency by various types of 'derived' maps which include those in which some kind of geochemical surface is fitted to the data, either to show broad-scale trends or deviations from them, the results of statistical or numerical manipulation of the data, or maps of multi-element combinations.

While many industrial reconnaissance scale surveys are still restricted to, say, 3 to 5 elements, in recent years there has been an increasing tendency for surveys undertaken by government agencies to cover a wider range of elements (e.g. Plant and Rhind, 1974; Wolf, 1974, 1975; Bendix Field Engineering Corp., 1976, 1977, 1978; Larsson, 1976; Cockburn, 1977; Baucom et al., 1977; Nichols et al., 1977).

### Local Geochemical Maps

The results of an intensive detailed survey, often with sampling densities of 10 to over 100/km<sup>2</sup>, can be presented in any of the ways described under reconnaissance. Soil samples are most commonly used for this type of survey, and since the area is usually small can be considered relatively homogeneous, so that contouring the element concentrations is often used.

### Map-Plotting Equipment

There are now a wide variety of possible media on which maps may be produced. Criteria for their choice include: the purpose of the survey (e.g. government or industrial use); ease of availability; size of output, print quality and ease of reproduction; physical stability of the final product; the duration for which the map will be required; and special characteristics such as colour plotting capability.

Pen plotters are probably the most widely used output medium for maps. The older drum plotters often suffered from paper stretch and inadequate inking of symbols, but have been much improved recently. Flat-bed plotters can attain very high cartographic quality and can cope with large map sheets. They usually have the ability to scribe high quality masters on to specially stable material for subsequent reproduction, and many have the capability of optical projection of symbols on to stable photographic film. Black and white line symbols are used generally, although some plotters have a colour pen capability. Unless optical plotting is used, large areas of tonal symbols may be tedious to produce. The lineprinter is one of the cheapest and most widely available media and is commonly used for posting data values or plotting tonal maps using overprinting. Special printer chains or coloured print ribbons can be used if required. Large maps can be made by sticking parallel strips of output together, but paper stretch may be a problem. Electrostatic or thermal printers produce similar output, although more flexibility in character generation may be achieved (allowing square symbols and a wide range of tonal symbols) but again paper stability may be a problem.

Computer-generated microfilm is becoming increasingly available. The cost per 35 mm frame is generally low, although high quality photographic treatment is required to preserve accuracy in the final enlargement. Most modern microfilm plotters have very good line definition, some with a variable line intensity capability. Colour microfilm plotters are still uncommon, although effective colour-separated masters for coloured maps may be generated on black-and-white film (Schweitzer, 1973; Mancey and Howarth, 1978).

Plotters using a laser beam to produce repeatable variable intensity images (with 16 or more grey levels) on sheets of dimensionally stable photographic film, are now being used extensively for production of seismic cross-sections and remote sensing imagery. They are all capable of producing larger original images than microfilm, and are well suited to the generation of colour separated masters for subsequent colour printing (e.g. Webb et al., 1978), although some devices can now plot directly onto colour film.

A new type of colour plotter using jets of cyan, magenta and yellow ink to draw directly onto paper was originally developed in Scandinavia (Smeds, 1976) and is now coming into general use. Early problems with ink colour and stability have now been overcome, and recent examples of maps produced with this type of plotter are of very good quality (J.G. Knudsen, pers. comm., 1978).

Exploration geochemists concerned with the rapid examination of geochemical map data as an aid to decision-making, rather than elaborate final copy, have at their disposal an enormous range of interactive computer display terminals, with which the ease of changing parameters (such as: class intervals, the number of contours and their levels, user-controlled values determining the type of contouring or moving average calculation to be performed) and redisplaying the result immediately is maximized.

Alphanumeric terminals can display text or a number of special characters (with up to 7 colours on some terminals) and may be used for lineprinter-like output. Graphic

terminals, which can display accurate line drawings, or in special cases many grey levels or colour, are available in a wide variety of costs and capabilities. The two basic types are (1) storage tube displays, in which a large quantity of line information is drawn on a special long-persistence phosphor, and to change the image the whole screen must be replotted; and (2) "refresh" displays in which the image is rapidly and continuously redrawn, as with a television picture, which gives a greater flexibility for particular types of graphic operations such as real-time rotation of figures in two or three dimensions. This last facility is particularly useful for the inspection of three-dimensional views of geochemical surfaces (Botbol, 1977).

While ease of map production, time and cost may be the ultimate criteria to the geochemist, it is inescapable that as the complexity of the data increases the visual design aspects of the display play an increasingly important part in helping the geochemist to extract the maximum of information from the map.

### Geochemical Basemap Requirements

Before examining the various types of geochemical map we consider the requirements for the basemap on which the geochemical data are to be plotted. The relevant factors are contents, scale and map projection, and reproduction process.

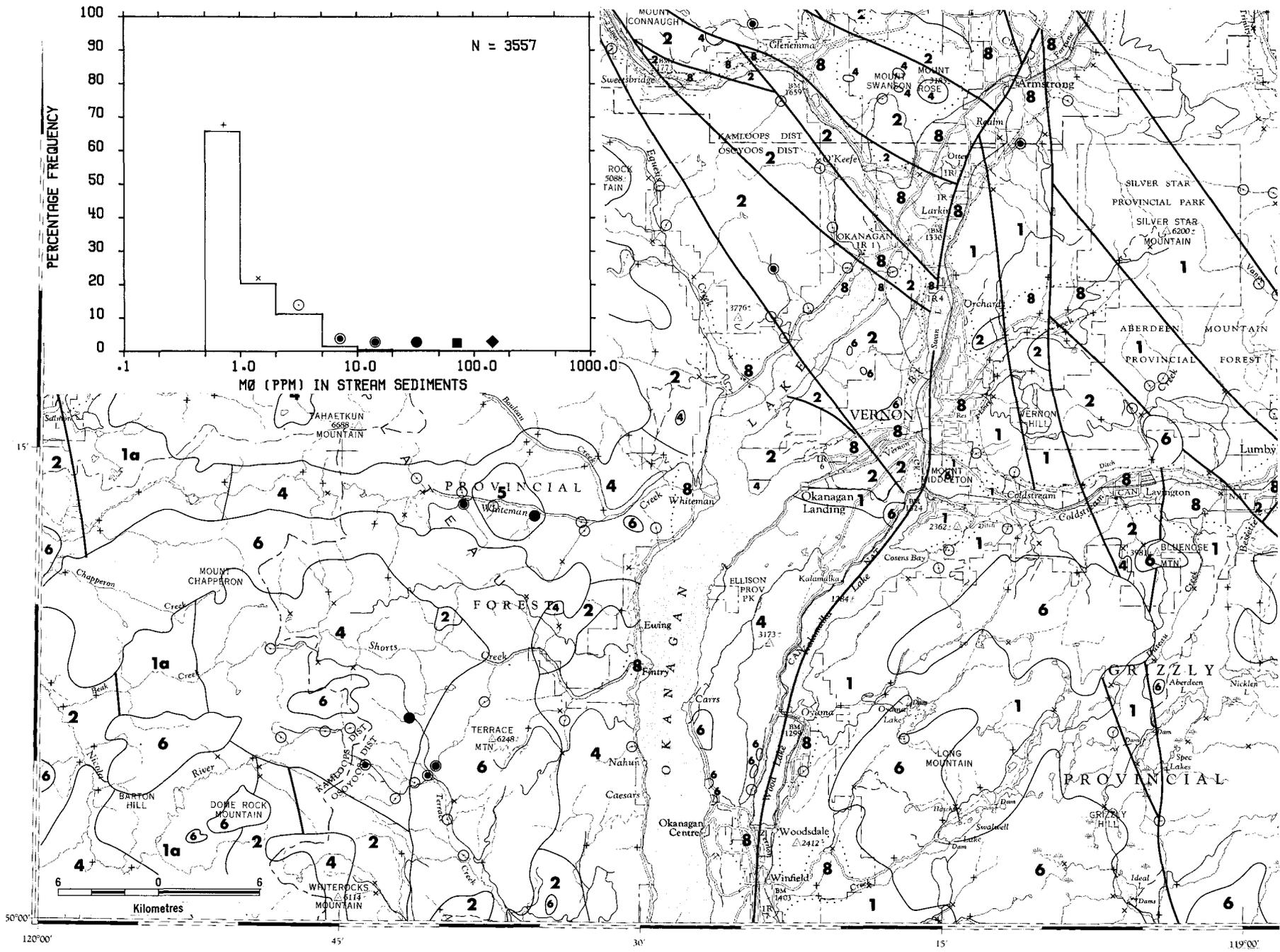
The main function of the basemap is to provide ready and adequate information so that a sample site can be located geographically. Where available, geological and possibly geophysical information may be included since it will be relevant to the interpretation as well as the geochemistry. The use of overlays may be helpful, as an overly complex basemap may be self-defeating when it comes to recognition of the geochemical patterns, particularly at the regional scale. The topographic information in the basemap may be derived from maps published by government agencies (where available, there are still many parts of the world not mapped at scales below 1:75 000), or from air photograph mosaics. It is usually necessary to enhance, by tracing or selective reproduction, the features relevant to the survey.

The map scale will depend on the area of the survey, the density of sampling and the kind of data to be included. A good choice will result in a small number of maps of convenient size on which the relevant information can be shown with clarity. Hawkes and Webb (1962) suggested that for a map in which individual sites are to be shown, a scale such that sample points are usually about on-quarter inch apart results in a pleasingly legible map. A confused overlapping of symbols should be avoided at all costs.

The choice of a particular reproduction process depends on the availability of plotting and reproduction equipment, cost and user requirements, and may range from dyeline reproduction to photo offset-lithography or other methods in government survey work. It should be borne in mind however, that unless the basemap and subsequently reproduced geochemical maps are on dimensionally stable materials significant stretching may occur in a relatively short time. The addition of the computer-generated geochemical data to the basemap can be done by plotting on to it directly, by plotting on a blank medium and then jointly reproducing the two to generate the final map, or by use of transparent overlays.

### Sample Location Map

This is produced during the survey by recording the position and corresponding sample identification on a topographic or other working field map, or air photograph.



**Figure 24.3.** Point-source measured value map with graded symbols for molybdenum in stream sediments. Note the geological information on the basemap. (Geological Survey of Canada - British Columbia Ministry of Mines and Petroleum Resources joint Open File 410, NTS sheet 82 L, Vernon, B.C.)

Experience has shown that human recording of location co-ordinates in the field is prone to significant error (e.g. inadvertently reversing digits in the numbers). Automatic recording of site co-ordinates directly from the field document can conveniently be carried out by placing it on an X-Y digitizer (of which several types are now available), the operator moves a cursor from point to point, entering the location co-ordinates automatically. Relevant topographical or geological information may also be encoded at the same time for later computer-aided drafting (U.S. Geological Survey, 1969). The computer can subsequently be used to convert relative site positions in terms of the digitizer table co-ordinates to UTM or other grid co-ordinates and to make any appropriate transformations for the required map projection.

Computer-assisted drafting of the sample location map will often save time and will certainly improve accuracy in the final result.

### **Geochemical Map Presentation Techniques**

In this section we review a variety of different methods used to make geochemical maps, all of which may be implemented by computer plotting, and the computer has widespread applications as a tool to aid the selection of appropriate class-intervals, even if the maps are subsequently drawn by hand. While we illustrate the methods with reference to elemental concentration values, they are equally applicable to the mapping of associated information such as pH, stream characteristics, overburden thickness, etc.

A recent survey of methods of selection for class-intervals (Evans, 1977) found at least 16 have been used in practice. Of these, the use of: 'natural breaks' in the frequency distribution; class width defined as a proportion of the overall standard deviation; arithmetic or geometric progressions of class width; and percentile classes (in which each class contains a known percentage of the data) appear to be the choices most often used by the geochemist. (Class selection is returned to under Univariate Statistics, below.) It should not be overlooked that there is usually an unstated assumption that every map should permit its users to unambiguously identify the class to which each individual symbol belongs. The upper limit to the number of distinct shadings or sizes of symbol which the human eye can distinguish is generally taken as seven or eight, and would only exceptionally exceed ten. If the map is to be used simply as a spatial data bank, then writing the actual values onto the map is clearly desirable. However, if it is to be used to study the overall spatial pattern of the map, to assess contrasts between different places (e.g. look for geochemical 'anomalies') or to compare the patterns of different element distributions then some kind of symbolism is desirable. Reducing the number of classes may achieve simplification at the expense of local contrasts.

Values may be shown on a map by numbers, symbols, symbol-number combinations, and contour-like patterns. We consider first, methods related to showing the value at a discrete site (point-symbol maps), and secondly maps emphasizing the distribution of regional or background patterns of variation, such as contour or trend-surface maps.

### **Point-Symbol Geochemical Maps**

#### *Sample Value Map*

The simplest type of map is one in which the element concentration is written at the collection site as a numerical value. This avoids possible bias on the part of the mapmaker by not putting the data into symbolic classes for presentation,

but unless the scale is carefully chosen to avoid overwriting of concentration values, and the values themselves are colour coded into broad classes, it may be extremely difficult to make out patterns without subsequent contouring by hand (with all its attendant difficulties, see below).

#### *Symbol Map*

A widely used method is the designation of the sampling location by a symbol. By varying in shape, size and internal detail, the symbol can serve the dual role of indicating both site and element value. A convenient set of symbols can thus be used in a map, each symbol denoting a class or a range of element values. The choice of classes can be based on the results of numerical evaluation of the data and the symbols will then indicate some level of significance. This can be a useful aid for anomaly identification.

Symbol-only maps have the advantage of clarity and minimum overprinting (Fig. 24.3). They are most useful where element values have a narrow dynamic range so that a symbol will represent a value within a few units of significance and analytical resolution. A wide variety of black and white symbols may be easily drawn using a pen or light-spot projection flat-bed plotter (e.g. Garrett, 1973b), or microfilm (e.g. Baucom et al., 1977).

Proportional symbols have been used in a variety of ways. A proportional width 'worn diagram' method has been used for manually produced maps of stream sediment surveys, whereby the stream is enclosed by parallel lines whose separation is proportional to the element value along the sampled section of the stream (e.g. Gregory and Tooms, 1969). It is however, very laborious, of limited value and unsuited to automated drafting. Proportional line lengths are being used for reconnaissance maps of Scotland (Plant and Rhind, 1974; Experimental Cartography Unit, 1978). Dickinson (1973) in discussion of this type of symbol noted that while it is easy to estimate visually (either unaided, or with the help of some kind of scale) and is thus useful to determine concentration at a given location, "the linear nature of the symbol makes it very difficult to accommodate a range of quantities of any great extent . . . as its size increases the visual 'weight' of the bar becomes more and more detached from the actual locality it is supposed to symbolize and the message of the map becomes rather vague from a distributional point of view". Experimental tests (Rhind et al., 1973) confirmed the difficulty of assessment of background values with this type of symbol, although it may be useful for giving an idea of local variability (J. Plant, pers. comm., 1978). Provided overlap of adjacent symbols is not disadvantageous, proportional circles or squares can better encompass a wide dynamic range than linear symbols, area being proportional to concentration. Errors in estimation are generally lower using proportional squares (Dickinson, 1973), although circular symbols appear to have been used more frequently in geochemical studies (e.g. Dall'Aglio, 1971; Barringer, 1977).

The use of graded (shaded) symbols seems to have been much more widespread than that of all types of directly proportional symbols (e.g. Webb et al., 1964; Garrett, 1973b; Hornbrook and Garrett, 1976; Baucom et al., 1977; Cockburn, 1977). Colour-coded dots have occasionally been used (e.g. Nichol et al., 1970), each colour being based on a black and white master produced from microfilm (Nichol et al., 1966).

The limitations of the completely symbolic map become apparent with wide dynamic ranges which must be covered either by a very large number of symbols, thereby losing simplicity and convenience of use, or by having symbols represent classes of values too wide for adequate resolution.

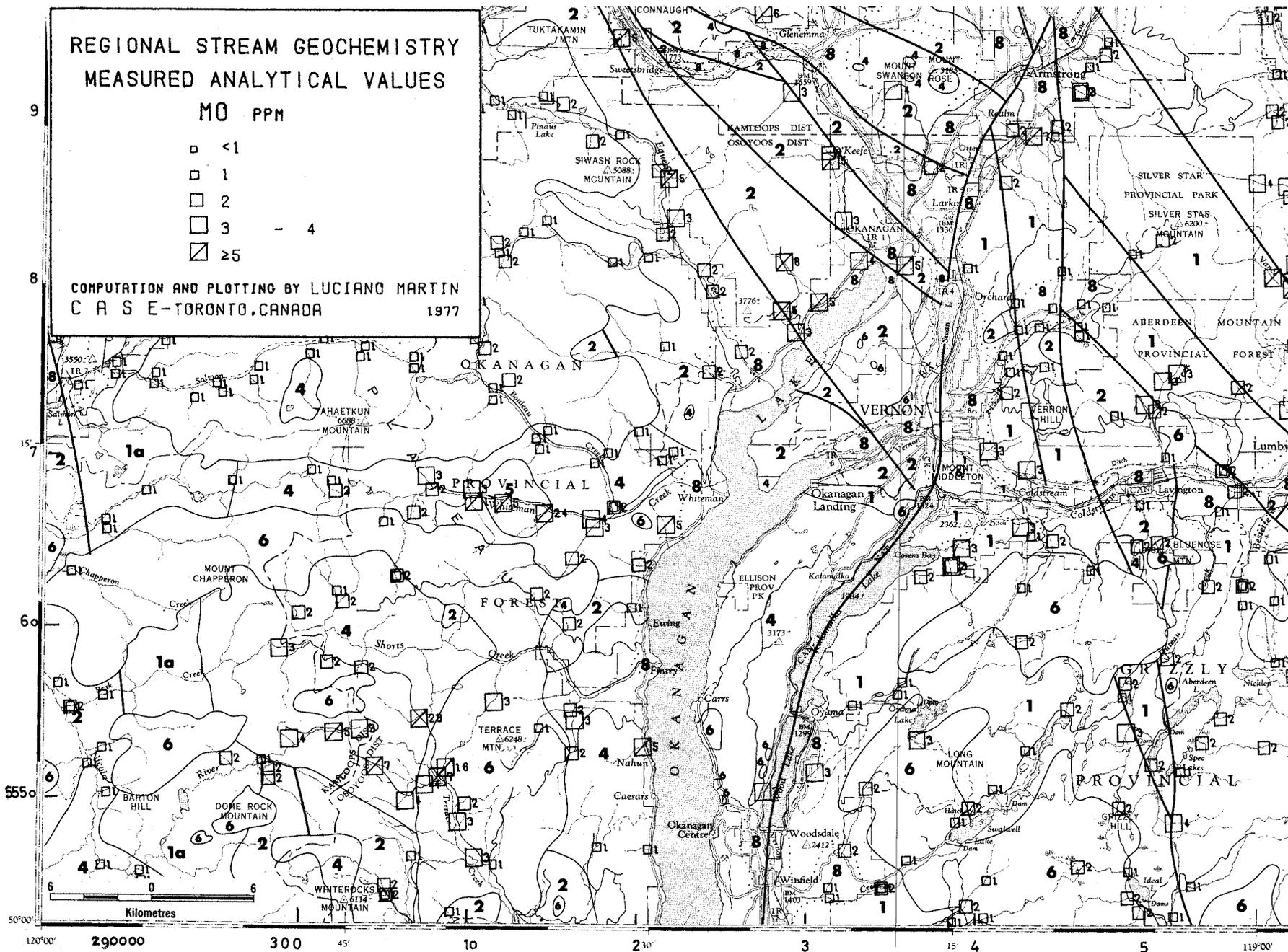


Figure 24.4. Measured value map showing graded symbols together with actual values for molybdenum in stream sediments. Map is portion of GSC Open File 410, NTS sheet 82 L, Vernon, B.C.

### Symbol-number Map

By explicitly printing the measured element value next to a proportional or graded location symbol (Fig. 24.4), the geochemist has a convenient graphic representation of the element values, yet is not limited to accepting the interpretation of significance inherent in the classes chosen for the map. While the added information in such maps may create some problems of legibility and overprinting, these can be minimized by choosing a suitable scale, and shifting the position of the numbers where necessary. The advantages of this kind of presentation usually greatly outweigh the minor drawbacks, and it is widely used by many exploration geochemists.

### Background-related Symbol Map

While qualitative considerations based on experience, intuition and general knowledge of the area have a significant role, a quantitative evaluation of the geochemical information should be very useful. The geochemist will usually start this evaluation by applying a basic statistical treatment to the total data set. From inspection of the frequency distribution histogram, the cumulative curve, or knowledge of the mean and standard deviation, 'background' and 'threshold' limits may be defined, and the element classes selected accordingly. The samples could now be divided into statistically anomalous and 'background' classes.

While this selection process will satisfy the statistician, we must remember that our targets are spatial geological anomalies, not just statistical. Since the population consisted of all the samples, it is clear that for statistical significance to be geologically relevant the data should have one uniform background population. Reconnaissance surveys usually cover several lithologies, each providing a distinct geochemical background. This creates the possibility that samples which would be anomalous within their local environment will not be statistically anomalous in the total population. Conversely, a value which appears anomalous in the total population might be geologically insignificant if the sample is from an area of high local background.

The background-related symbol map (which Martin terms the 'justified measured value map') is similar to the symbol-number map described above, as it includes the actual element value and a grading symbol. However, whereas in the earlier map the grading was based on overall statistical parameters, the symbol now reflects the significance of a sample within a group with similar background. To prepare this type of map one must first know the background determining factors (such as underlying lithologic units) and use these to divide the data into groups of compatible samples. Statistical parameters are then computed for each group and are used to produce the justified grading by one of the standard methods of class selection. The computer is invaluable for this type of work, making use of integrated data base manipulation and statistical tools.

### Maps Emphasizing the Regional Geochemical Distribution

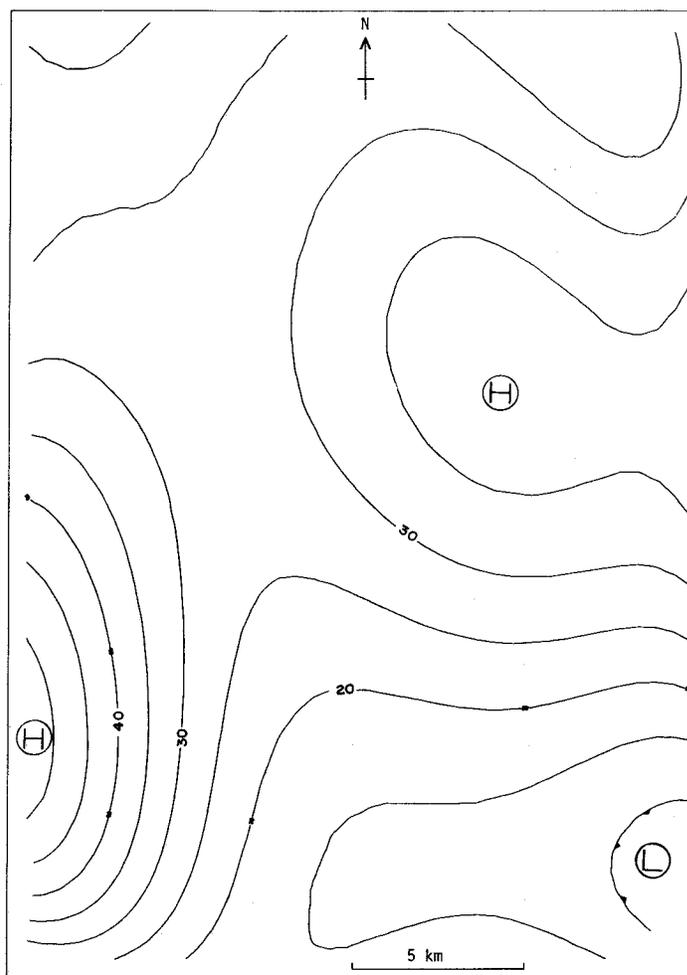
The emphasis so far has been placed on individual sample locations. However, for further interpretation it may be useful to take into account variations in the topography of the regional geochemical surface of element concentration values. Before considering suitable methods to do this, it is essential to realize that for the extrapolation from the data points to a more general surface to be valid, the area under consideration must be considered homogeneous, and the density of the sample points must be sufficiently representative of a continuous surface. In reconnaissance

surveys these prerequisites are seldom met and therefore in most cases contouring of untreated concentration values may not be justified; it may however be possible at a local scale, or at a broad regional scale.

Following Rhind (1975) we may divide surface-fitting methods in to two broad classes: firstly, global methods in which a generalized "trend surface" represented by a polynomial equation, is fitted to the entire data set; and secondly, local methods which include contouring in general, with moving average and kriging being regarded as different types of interpolation functions.

### Trend Surfaces

Very broad-scale regional variation is approximated by fitting a surface of simple geometric form, normally a polynomial equation in two dimensions involving powers and cross-products of the (x,y) spatial co-ordinates and (z) the element concentration values. The 'degree' of the fitted surface corresponds to the highest power term used. Trend-surfaces are usually 1st (linear), up to 5th (quintic) degree (Fig. 24.5), although higher powers have occasionally been used. So-called Fourier surfaces are similar and involve sine and cosine terms but do not seem to have been used for exploration geochemistry problems. Trend-surface analysis requires a certain amount of care both with programming (e.g. Agterberg and Chung, 1975) and interpretation, as the



**Figure 24.5.** Derived regional map: 5th degree polynomial trend surface used to approximate the regional component; 5 ppm contour interval.

orientation of the surface may be sensitive to the number of data points and their disposition (Howarth, 1967; Norcliffe, 1969; Chayes, 1970a; Doveton and Parsley, 1970; Shaw, 1977).

The element value at a location is assumed to be a product of both the geological environment at large (the regional component, represented by the trend-surface) and local variations imposed on this surface (residuals). Trend-surface analysis separates these two components in an arbitrary fashion. The residual value (difference between the measured concentration and the regional surface value) may be used as an indicator of anomalous mineralization, and mapped (e.g. in symbol-number form), class selection being based on statistical analysis of the residual values. Examples of trend analysis in exploration geochemistry include Connor and Miesch, 1964; Nackowski et al., 1967; Nichol and Webb, 1967; Cameron, 1968; Bosman et al., 1971; Brabec and White, 1971; and Lepeltier, 1977.

Garrett (1974b) concluded, based on both his and E.M. Cameron's experience, that "the method has a limited application to either very broad regional surveys, or very detailed local surveys, and in either case the order of the surfaces should not exceed quartic at the maximum. Bearing in mind the heterogeneity of individual rock-units and whole survey areas we are not convinced that a polynomial model is appropriate as we do not have, in the vast majority of cases, a sufficient knowledge of the true pattern of areal variation and the included discontinuities".

#### Contour and Related Maps

One of the commonest methods of representation of the regional component of the geochemical surface is by contouring. Contour maps may be produced manually or by computer techniques. It is a common fallacy that all humans can produce "good" contour maps, because one has only to give the same point source data to a number of persons to contour to find out how much subjective variation can be introduced.

The steps involved in computer contouring from a given set of data points all involve some subjective choices of methodology but once the parameters for a particular method have been chosen, subsequent contouring is completely objective and reproducible providing the sampling density is adequate. The initial choice is the size of the square mesh on to which the data values are to be interpolated; related to this is the method used to choose the number and disposition of the data points surrounding each grid node (nearest  $n$ -neighbours, quadrant or octant search) on which the node value will be based, and the distance weighting assigned to each point (inverse squared distance is commonly used). There are a large number of possible algorithms upon which the interpolation is based, typically between three and ten points per node are commonly used.

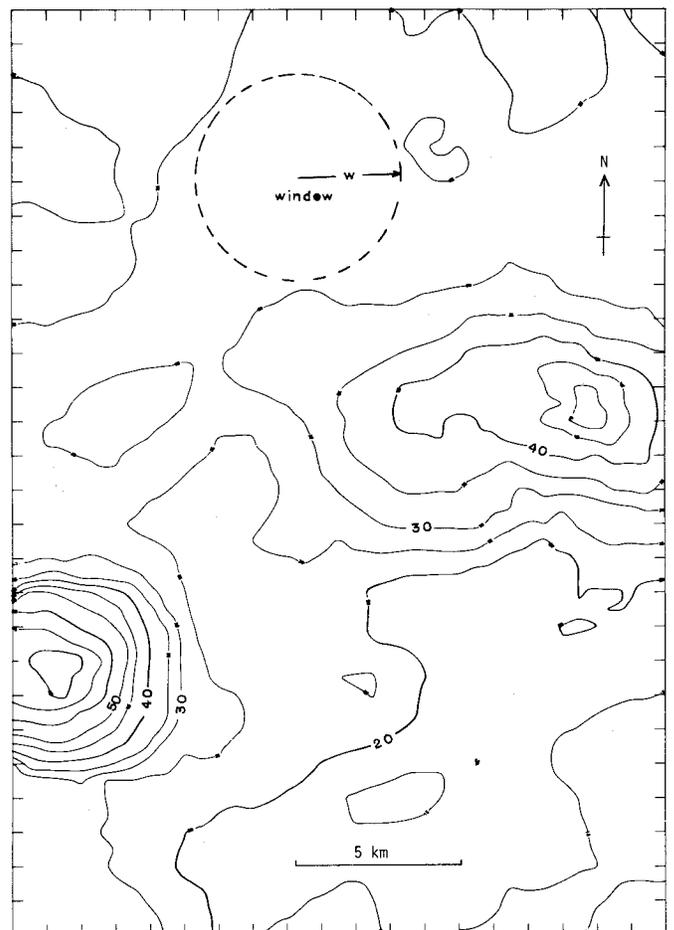
Most contour interpolation techniques are exact at the data points (although this is not necessarily true of methods based on polynomial interpolation), but it is quite possible for spurious maxima and minima to be generated in intermediate areas as artifacts of the methods used (although these may be of small magnitude). There are few comparative studies of the results of different contouring algorithms (Walden, 1972) and no general guidelines have yet emerged about techniques best suited to geochemical data.

The geostatistical Theory of Regionalized Variables (mentioned earlier in the context of sampling) provides in kriging or universal kriging a method yielding an optimal exact linear interpolator (the kriging estimators match the true values at the locations of the data points), which is also unbiased, and optimum in that the estimation variance is minimum (see Huijbregts and Matheron, 1971; Olea, 1975;

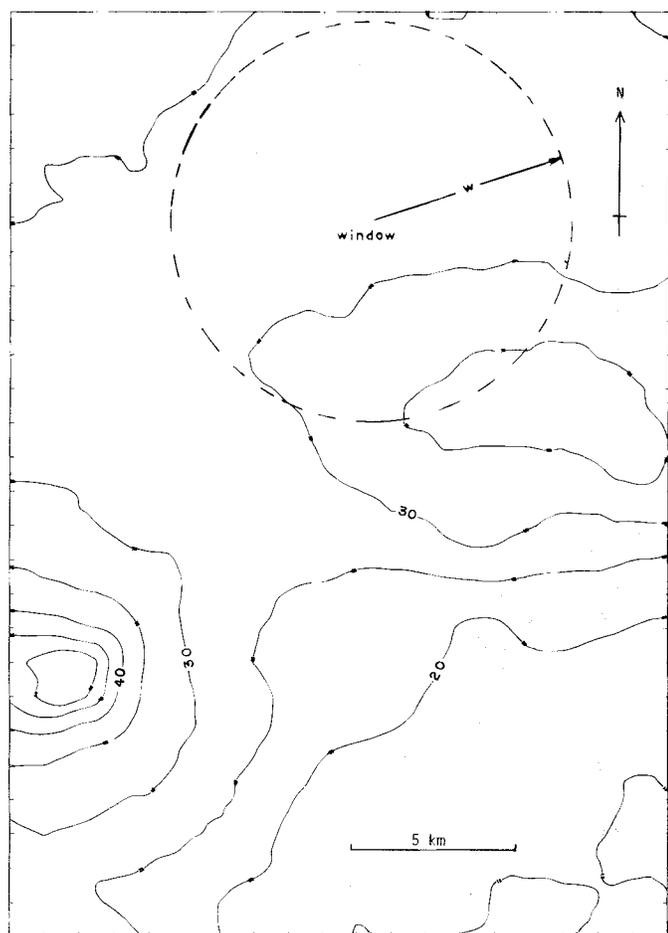
Guarascio et al., 1976; David, 1977). The weights used in the kriging estimator reflect the spatial autocorrelation within the surface and the distribution of the data points. The estimation variance is a superior measure of probable error in the map compared with the more usual variance of residuals from a fitted surface, or variance of the data values within a cell.

This approach has so far been mainly applied to problems in ore reserve analysis and has yet to be explored as a tool for geochemical mapping.

Because of the noisy nature of much geochemical data (field and analytical variability), moving average methods have long been applied to extraction of the local background variations at broad-scale regional or more detailed scales, density of the data points permitting. A square, or circular, window is passed across the map area and the average value of all data points within it is assigned to the grid node (or, more rarely, to the centroid of the data point locations). Variables include window size, degree of overlap (if any) between successive window positions, and the weighting function used. The resulting map is usually contoured (Nichol et al., 1966; DeGeoffroy et al., 1968; Wolf, 1974, 1975; Hornbrook and Garrett, 1976; see Figs. 24.6 and 24.7). An alternative approach has been to present the map in a cell form, using small square or rectangular cells (the latter corresponding to the lineprinter print positions). This avoids the computational overheads of contour-threading, while still presenting the spatial patterns of variation satisfactorily



**Figure 24.6.** Derived regional map: Weighted moving average using 3 km radius window applied to the same original data used to obtain Figure 24.5, 5 ppm contour interval.



**Figure 24.7.** Derived regional map: As Figure 24.6 but with 6 km radius window.

(Howarth, 1971a; Howarth and Lowenstein, 1972, 1974; Webb et al., 1973). More recently, the use of recursive line-by-line moving average methods has enabled the processing of data sets in the order of 50 000 samples with minimal use of computer storage (Webb et al., 1968; Fig. 24.8).

For interpretational effectiveness the regional contours and the residual values are often plotted on the same map (see Fig. 24.9). Among the more significant advantages of this method is the dual use of the regional component as a flexible, continuously variable background estimator, and as a direct aid to geologic mapping. The full realization of the potential benefits depends on the choice of a computational technique and parameters for the regional component which optimize the sampling density and best reflect the underlying geological features of the survey area.

While computer-contoured maps can look aesthetically pleasing, too often the geochemist is unaware of the method used to produce them. One should always be prepared to discuss the exact basis of the interpolation and contouring techniques used with the mapmaker. Facilities such as contour annotation, smooth contours, suppression of contours where they bunch too closely together, a variety of line styles, and blanking out of contours in areas of inadequate control are available in some contouring packages, but sophisticated graphics are computationally more expensive to produce.

Alternative methods for enhancing anomalous patterns using mapping techniques have been in experimental use at AGRG, London, for some time and include high-pass, picture frame and Kolmogorov-Smirnov filters (Howarth, 1974). Current work with regional geochemical data from the United Kingdom, Canada and the United States is yielding encouraging results.

### Multi-element Geochemical Maps

In many cases the interaction of more than one element may be of relevance to the interpretation of the geochemical map. A simple example would be the mapping of elemental ratios to manganese in stream sediments from areas where manganese coprecipitation is known to take place. Multi-arm 'wind rose' symbols, in which arm length is proportional to concentration and each element has a specific arm orientation, have been used (Rhind et al., 1973; Plant and Rhind, 1974; Experimental Cartography Unit, 1978) for point-symbol maps at a reconnaissance scale. Colour-mixing techniques for three elements at a time (in which each element is represented by either a cyan, yellow or magenta map, with an intensity proportional to concentration; the resulting colour mixtures of the three maps when superimposed show relative compositions) have been successfully used at a regional scale for maps produced by lineprinter (Howarth and Lowenstein, 1976), laser plotter (Lowenstein and Howarth, 1973; Webb et al., 1978), and microfilm (Mancey and Howarth, 1978). A related technique for radiometric data has been described by Lindén and Åkerblom (1976).

These are all essentially cartographic devices which attempt to use graphical, rather than statistical, means to optimize presentation of multi-element data for comprehension by the geochemist. They can be regarded as further aids to interpretation and would generally be thought of as complementary to the original single-element maps.

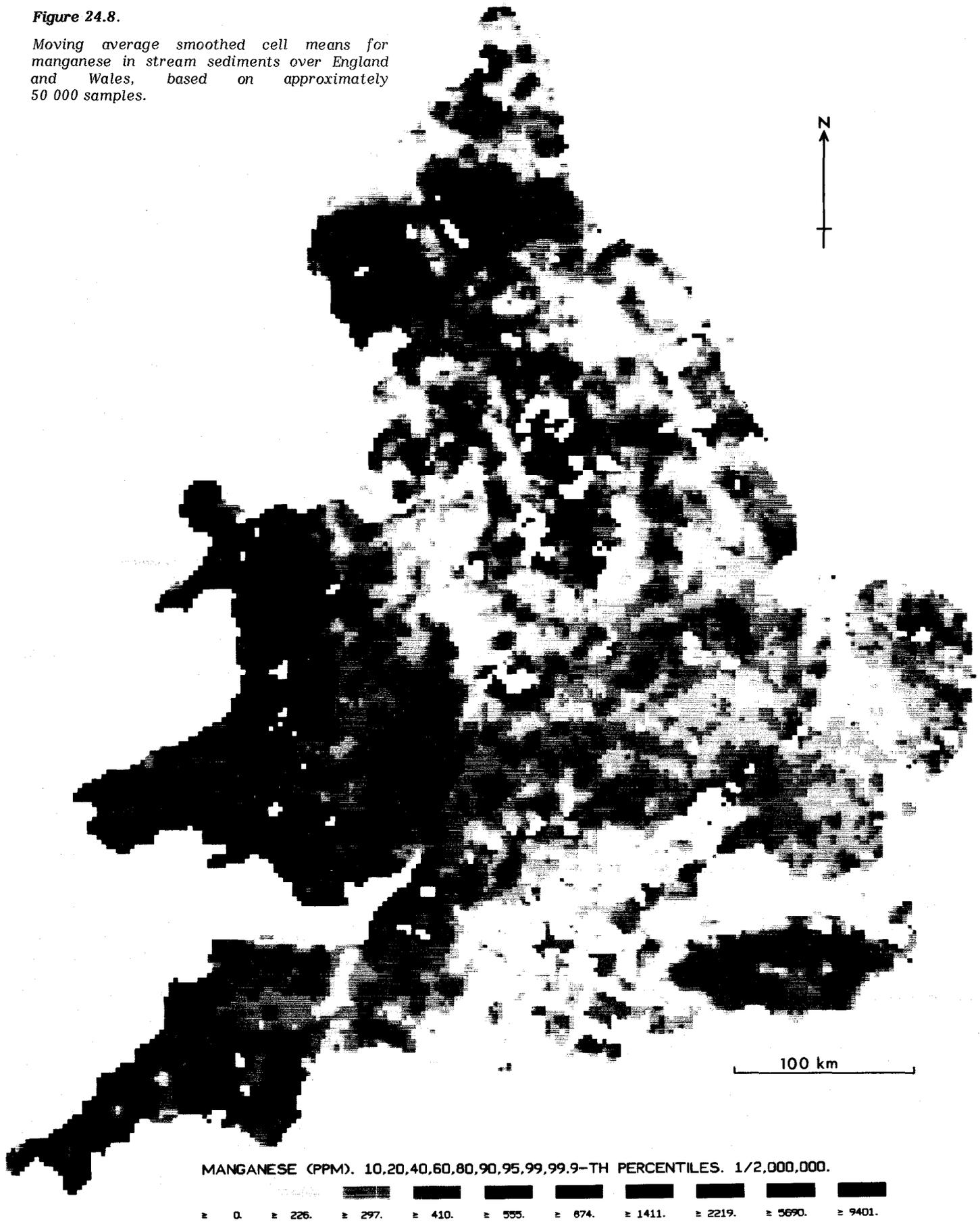
Statistical methods which have been used in conjunction with multi-element interpretation include multivariate regression and the mapping of factor scores, and are considered later.

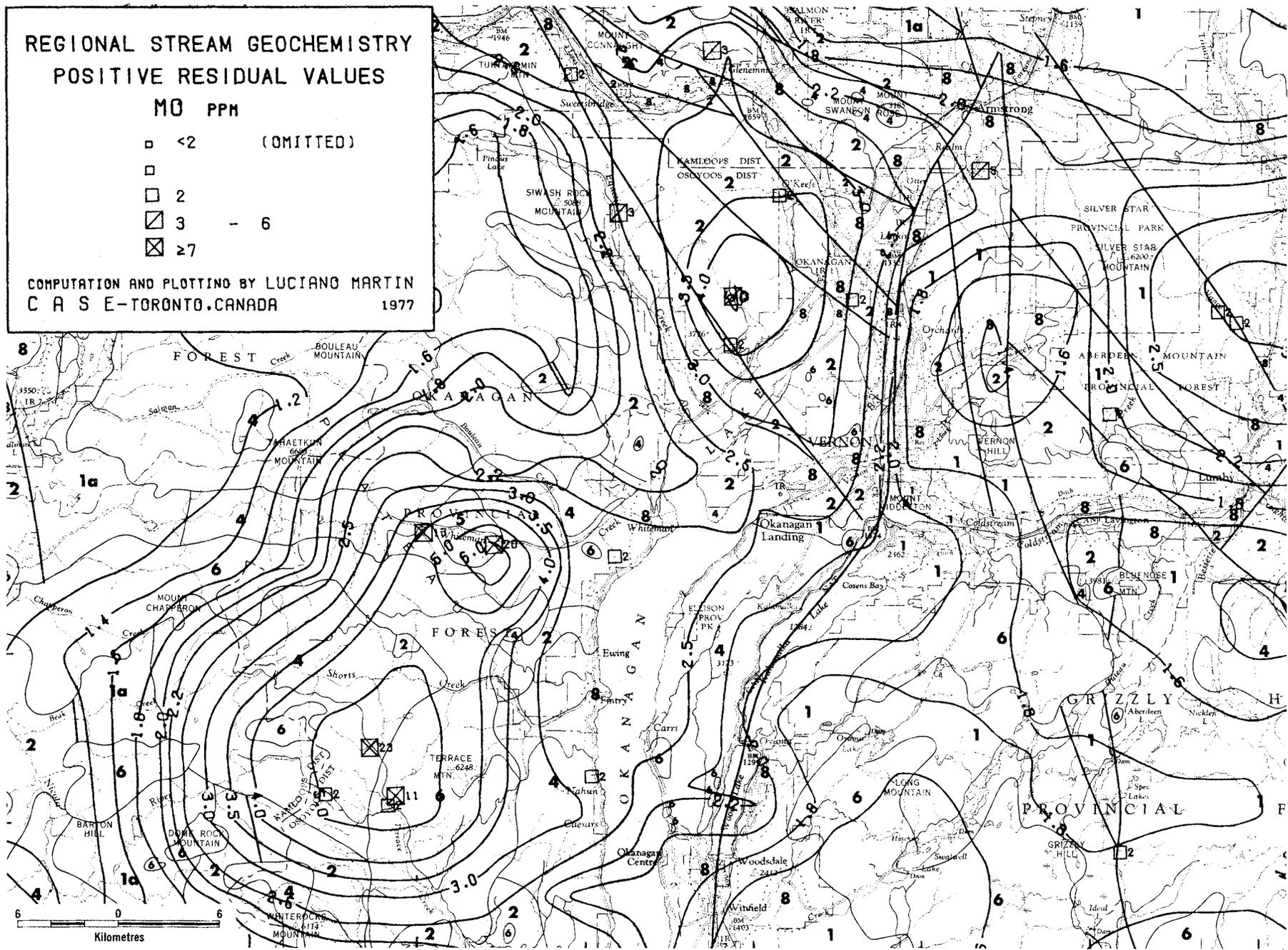
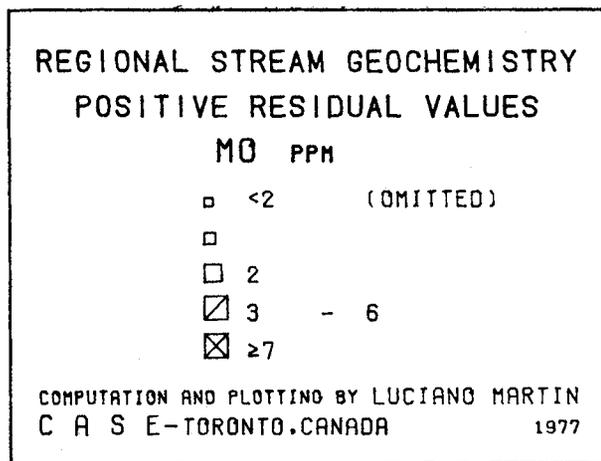
### UNIVARIATE STATISTICS

It is apparent from information we have obtained from many countries that the majority of geochemical exploration programs encountered in industry tend to be considered on a single-element basis, from the points of view of both data description and interpretation. There are indications that national geological surveys, research institutions and universities in the more advanced countries are applying multivariate techniques, but these have not become widespread in industry. The reasons for this are probably twofold: (1) the lack of routine analysis for, say, 5 to 30 elements means that the problem of interpretation of complex multi-element interrelationships is partly avoided (although the investigator may thereby be missing helpful information); and (2) the lack of adequate computer facilities for geochemists in many countries makes effective use of more than the most elementary statistical techniques impossible. This situation can be expected to change rapidly during the next few years as computers are now becoming very much cheaper, as well as more powerful, in the mini- and micro-computer market. A third problem is the time and personnel involved in preparation of the data for entry to the computer (Lepeltier, 1977) which usually necessitates writing data on coding forms for punching of cards or, more recently, direct key-to-disc or terminal entry systems. Data-logging devices are ideally required for ease of transfer of multi-element data for analytical systems with a throughput of several hundred samples per day.

**Figure 24.8.**

*Moving average smoothed cell means for manganese in stream sediments over England and Wales, based on approximately 50 000 samples.*





**Figure 24.9.** Derived regional map: Contoured weighted moving average map for molybdenum in stream sediments. Significant positive anomalies shown by graded symbols together with residual value. Portion of GSC Open File 410, NTS sheet 82 L, Vernon, B.C.

It is arguable that for the interpretation of only a few elements, the computer offers little advantage over a roomful of people collating data manually and plotting it on maps. If one is content with single "one-off" exercises with inherent human error, this is probably a fair statement. However, once the data have been encoded, the computer affords the flexibility to repeat investigations rapidly with changed parameters. This is equally true of the study of a frequency distribution, where choice of class-width can be critical, as of more complex situations such as decomposition of frequency distributions, multivariate statistics, or computer mapping (discussed earlier). The computer however has been used successfully in many projects in exploration geochemistry for the relatively unsophisticated tasks of converting instrumental output to concentrations, and generating histograms prior to the preparation of maps (even if these are plotted by hand). As mentioned earlier, one of the initial time-consuming tasks is the preparation of the data for the computer.

### Elementary Statistics

Miesch (1967a, 1967b) reviews the fundamental univariate aspects of geochemical error and estimation of abundance. However, one of the practical problems which can arise in calculation of even elementary descriptive statistics such as the mean and variance, is their accurate calculation without storage of all the values in the computer at one time, and on a machine in which the number of binary "bits" representing the concentration value is relatively small (it can commonly vary from word-lengths of 8 to 60 bits for single precision arithmetic on different computers). Ling (1974) reviewed possible algorithms for the computation of sample means and variances and recommended two pass methods using  $M_1 = \sum x_i/n$  followed by  $M_2 = M_1 + \sum(x_i - M_1)/n$  and  $S_2 = \sum(x_i - M_1)^2 - \frac{1}{n}[\sum(x_i - M_1)]^2/n$ , where  $n$  is the number of samples. The variance is then  $S_2/(n-1)$ . For discussion of one-pass algorithms (which have to be chosen to some extent with regard to the data characteristics) the reader is referred to Ling's paper. A similar approach can be used to build up matrices of cross-products in the multivariate case. If equal class widths are used for histogram computation, the frequency counts may be built up without storing all the values of the element in core at one time, and algorithms for cumulative frequency plots on the lineprinter are available (e.g. May, 1972). Burch and Parsons (1976) recently proposed a very simple new test for normality based on use of the cumulative curve.

### Threshold Selection

The major interest in univariate statistical analysis has focussed on the problems of anomaly threshold determination and the decomposition of polymodal frequency distributions. Criteria based on a cut-off point of the arithmetic mean plus 2.0 standard deviations (Hawkes and Webb, 1962; Lepeltier, 1969), or less commonly utilizing the "geometric deviation" (calculated as for the standard deviation but differencing with regard to the geometric mean) or empirical criteria, are being abandoned by several institutions in favour of either percentile cut-offs based on the actual frequency distribution (i.e. the concentration values exceeded by known percentages of the data), or the results of frequency distribution splitting (see below). The appropriate percentiles to use will depend not only on the form of the frequency distribution, but also on the size of the data set (e.g. the 95th percentile for 1000 samples would include 50 points, but the equivalent would be 500 points for 10 000 samples, which might be undesirably large in spatial terms in a final map). A percentile-based criterion however does have the advantage of exact comparability from one element to another irrespective of the forms of their frequency distributions. An inconvenience

is that the threshold would be different for data from different areas, whereas a purely arbitrary criterion can remain fixed during the lifetime of a continuing project. The penalty of the latter is usually insensitivity to differing regional background values. The interrelationship of threshold and sampling-plus-analytical error has been discussed earlier. More complex geochemical criteria have also been advanced for anomaly recognition (e.g. Dubov, 1973; Cachau-Hereillat, 1975).

### Splitting Frequency Distributions

Techniques for the decomposition of mixed frequency distributions are now fairly well known, and a useful survey has been given by Clark (1976). Probably the best known graphical methods in geochemistry are those of Lepeltier (1969) and Sinclair (1974, 1975) which are based on analysis of the cumulative distribution of concentration values. The splitting of the cumulative distribution into straight line segments by Lepeltier (1969) implies the relatively unusual situation of mutually truncated populations whereas the fitting of a smooth-curve, used by Sinclair and others, assumes mutually overlapping populations, which is probably more realistic. Bolviken (1971), Parslow (1974) and Sinclair (1975) illustrate families of smooth cumulative curves which may be encountered with two mixed normal distributions, and McCammon (1977) gives a Fortran subroutine for calculating percentiles of such distributions. The graphical approach is straightforward to apply, but may be somewhat subjective (it can be difficult to pick the inflection point(s) by eye), and may give misleading results for small populations.

Numerical methods are well suited to computer evaluation. Clark and Garnett (1974) described estimation of parameter values by a nonlinear least-squares technique, which minimizes the sum of squared differences between the observed and expected frequencies up to each class end-point, based on the cumulative normal distribution. A Fortran program (Clark, 1977) has been successfully implemented on a mini-computer and is well suited for interactive use; it has given excellent results with minimum chi-square values in comparative goodness-of-fit tests against other methods (Clark, 1976).

The temptation when splitting multi-modal distributions by either graphical or numerical techniques must be to fit a new distribution corresponding to each small irregularity in the cumulative curve. Here one must have regard to the confidence belt around the observed distribution (e.g. fitting bounds on the basis of the Kolmogorov-Smirnov statistic, cf. Stevens, 1974), and it is reasonable to suppose that in the context of a geochemical map the populations might also be geographically distinct (O. Celenk, pers. comm. 1972; I. Clark, pers. comm. 1976-7; Govett et al., 1976). Examples of the decomposition of frequency distributions in exploration geochemistry are cited by Brabec and White, 1971; Lenz, 1972; Bolviken and Sinding-Larsen, 1973; Sinclair, 1974; Montgomery et al., 1975; Sinding-Larsen, 1977; and Lepeltier, 1977, and graphical methods are known to be in wide use among mining companies.

Techniques also exist for dealing with censored frequency distributions (i.e. in which the number of determinations falling below detection limit is known, but their exact values are not), both graphically on the basis of the cumulative curve (Sinclair, 1976) and from the point of view of estimation of mean and standard deviation corrected for the degree of censoring (Cohen, 1959, 1961, also cited in Miesch, 1967b; Selvin, 1976) or the computer method of Hubaux and Smiriga-Snoeck (1964). Such corrections can be of great importance when it is required to compare abundance estimates from two or more data sets in which a significant proportion of censored data occur.

### Data Transformation

A problem of continuing interest in the exploration context is how helpful it is to apply logtransformation to "pseudo-lognormal" distributions. Such distributions are commonly positively skewed and have a high kurtosis, thus superficially resembling a lognormal distribution, and are often composed (in exploration geochemical data at least) of aggregates of smaller populations over a large geographical region. Recent papers (Link and Koch, 1975; Govett et al., 1975) and ensuing correspondence (Chapman, 1976b) have suggested that logtransform may not always be appropriate for estimation of the mean, owing to introduction of a (generally small) bias; and that logtransformation will often enhance the contribution of background with deleterious effect on the recognition of high concentration values particularly in a multivariate situation. Miesch (1977) suggests that a small bias in the estimate of the mean may be preferable to a very wrong answer because the method of estimation is inefficient, and also that logtransformation is desirable to reduce dependency of error variance on the magnitudes of the means (a common situation in analytical data, as discussed earlier): this is particularly important for analysis of variance and it enables one to examine proportional differences in the data. It is clear from Chapman's reply (1977) that, for him at least, this controversy hinges on the recognition of anomalous populations and the interpretability of the results gained from logtransformed data. Many would agree with Miesch (1977) that although "there are many situations in geochemistry where a log transformation of the data is inappropriate, . . . log transformations are helpful far more often than not".

A more general nonlinear power transform exists for which the lognormal distribution is a special case (Box and Cox, 1964):

$$\left. \begin{aligned} z &= (x^\lambda - 1)/\lambda, \quad \lambda \neq 0 \\ z &= \log_e x, \quad \lambda = 0 \end{aligned} \right\} \quad x > 0$$

The value of  $\lambda$  being chosen to minimize departures of either skewness, or skewness and kurtosis, from that of the normal distribution. Continuing investigations at AGRG, London, on a variety of data sets suggest that this provides a most useful alternative to the usual logtransformation. Initial results are encouraging and a variety of computer methods exist to optimize  $\lambda$  (Howarth and Earle, 1979). This transform is further discussed in the next section.

### MULTIVARIATE STATISTICS

In contrast to univariate analysis, in which the computer is often performing functions that could be undertaken manually (although once the geochemical data have been encoded, it is generally much faster and more accurate than the human), multivariate statistical analysis allows us to investigate complex inter-element relationships that would be prohibitively time-consuming to undertake by hand, even if it were humanly possible.

Our recent survey of computer usage among mining companies, national geological surveys, research institutions and universities suggests that multivariate statistics are being little used among exploration geochemists as a whole,

but where adequate computer facilities exist regression and correlation analysis are the most common applications; principal components or factor analysis, discriminant analysis, and cluster analysis being relatively little used. Davis (1973) gives a very readable introduction to these techniques.

In most institutions reliance is generally placed on existing statistical packages for general use. Slys (1973) gave an interesting survey of the contents of various standard packages for multivariate data analysis with comparative time and cost evaluations, and showed some major performance differences among them. Anyone considering which package to use would be well advised to consult his paper, although comparative evaluations of accuracy could have usefully been included. Those which seem to be used most often by exploration geochemists are BMD/BMDP (Dixon, 1975), SPSS (Nie et al., 1975) and, in earlier work, SSP (IBM Corp.). A vast range of special programs are also available (particularly for factor and principal components analysis, discriminant analysis, and cluster analysis).

It is surprising how often, when re-running older "standard" computer programs with the more powerful debugging compilers now available, erroneous code appears to have been released on the unsuspecting! Users are cautioned particularly to ensure that they have reliable algorithms for operations such as matrix inversion and determination of eigenvalues<sup>1</sup> since these can make a significant difference to the result of a calculation. However, programs tailored for the handling of the very large data sets commonly encountered in exploration geochemistry have generally to be developed by the users themselves.

### Correlation

Correlation is a measure of the extent to which one element varies either sympathetically or antipathetically with another. The familiar Pearson linear correlation coefficient (Pearson, 1899) is easily calculated (e.g. Davis, 1973) and can be very informative. It was shown in the discussion of univariate statistics how the necessary cross-products of deviations from the mean could be accumulated without having to store a large data-matrix in the computer. However, as was pointed out by Howarth (1973a) and Chapman (1976a), the correlation matrix may be seriously affected by: erroneous or "wild" data values; the presence of more than one geochemical population in the data; very common sample types which may overwhelm the characteristics of less usual samples; scaling of the data, e.g. logtransformation, will affect the correlation coefficients obtained; and testing the significance of the Pearson coefficient requires that the elements are approximately normally distributed. The number of samples needed to reliably estimate the multivariate correlation matrix should preferably be at least 10 times the number of elements involved (Howarth, 1973a; Trochimczyk and Chayes, 1977). Alternatives such as the non-parametric Spearman (1904) and Kendall (1938) rank correlation coefficients are more complex to compute (see, for example, Siegel, 1956) and may be too time-consuming to use in practice (Chapman, 1978).

In many cases a matrix of Pearson correlation coefficients between all element pairs will give much useful information on the structure of inter-element relationships

<sup>1</sup> The NAG (NAG Central Office, Oxford OX2 6NN, England) and IMSL (IMSL Inc., Houston, Texas 77036) program libraries are examples of software in which particular attention has been paid to numerical accuracy.

for little computational effort. However, it is usually wise to check abnormally high or low correlations, by making scatter plots of the data, to guard against spurious effects. Chapman (1976a) contains a cautionary example in which the correlation between Cu and Mo in a set of 211 stream-sediment samples was reduced from 0.938 to 0.173 following the removal of one spurious sample, logtransformation of the original data (including the "wild" sample) yielded a correlation value of 0.226. This suggests that comparative evaluation of untransformed and logtransformed data may have its benefits. Rank correlation coefficients of the original data would also have shown lack of correlation.

A problem which still remains unresolved is concerned with determining the significance of correlation coefficients in data which add to a constant sum (e.g. 100 per cent). Such data were termed "closed" by Chayes (1960) who first pointed out that spurious correlations could be introduced into data by the effects of closure (e.g. by the percentaging operation), often showing up as apparent strong negative correlations. Chayes and Kruskal (1966) devised a statistical test to determine whether a given correlation coefficient derived from such data might be nonzero. Miesch (1969) suggested that while their criterion might be used to test the hypothesis that none of the variables in a specific geochemical problem are correlated in excess of the correlation attributed simply to the effects of closure, if this hypothesis is rejected it is still impossible to tell which specific pairs of variables are truly correlated geochemically. Chayes (1970b, 1971) rebutted this, but a recent re-examination of the problem (Kork, 1977) has confirmed that identification of significant correlations is essentially impossible even if the number of variables and sample size are large.

Very often, careful analysis of a correlation matrix will give sufficient information about element associations to form hypotheses, but is not capable of the more sophisticated decomposition possible with principal components or factor analysis techniques.

### Regression

Linear regression is generally used, in a predictive sense, to fit an equation of the form:

$$Y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_m x_m$$

where Y is the dependent element and  $x_1$  to  $x_m$  are the independent elements (Davis, 1973). For example, it may be required to regress zinc as a function of manganese to assess the effect of possible manganese scavenging. In general, error can only be associated with one variable; usually the X values are assumed error free and the regression line is fitted so that the sum of squared deviations of the Y values from the line is minimized. The situation becomes more complex if both X and Y values are assumed to have independent errors associated with them, weighted regression techniques must be used, or the "reduced major axis" (e.g. Chapman, 1976a) can be fitted. If the errors in X and Y are not independent, correct fitting techniques will be highly complex. For a discussion of the related theory of "functional analysis" the reader is referred to Mark and Church (1977).

Practical applications in exploration geochemistry have almost entirely been concerned with correction of element concentration values for supposed manganese and iron scavenging. In many cases so-called stepwise multiple regression techniques have been used which fit the regression equation only to those elements which contribute significantly (in a statistical sense) to the observed variation in the dependent variable, omitting those which do not from the equation. A useful survey of such methods appears in

Berk (1978). Applications illustrating regression on organic content, and/or iron and manganese for stream sediments include: Rose et al., 1970; Rose and Suhr, 1971; Brundin and Nairis, 1972; Austria and Chork, 1976; and Larsson, 1976. It should be noted that the strength of the regression coefficients are not consistent from one study to another and can vary significantly between areas in the same region of similar environment and geology (O. Selinus, pers. comm., 1977). Computation of a global regression equation over a very large area may be unwise for similar reasons to those discussed earlier for correlation coefficients.

Other applications of regression in exploration geochemistry include: DeGeoffroy et al., 1968; Dahlberg, 1969; DeGeoffroy and Wignall, 1970; Rose et al., 1970; Hesp, 1971; Chatupta and Fletcher, 1972; Dall'Aglio and Giggli, 1972; Pelet and de Jekhowsky, 1972; Timperley et al., 1972; Dall'Aglio and Tonani, 1973; Smith and Webber, 1973; Culbert, 1976; and Rose and Keith, 1976. In some cases nonlinear functions may be more appropriate. Cameron and Hornbrook (1976) illustrate a spectacular quadratic relationship between mean U and L.O.I. for lake sediments in the Canadian Shield.

### Factor Analysis and Related Techniques

"Factor" analysis was one of the first multivariate analytical methods to become widely used among geologists and geochemists. While it has been helpful in many cases, misrepresentation of its capabilities in the early years probably served more to bring computer applications into disrepute in exploration geochemistry than to promote the power of the computer as an aid to problem solving.

For those who require an in-depth survey of the subject, the recent textbook by Jöreskog et al. (1976) is set in a geological context, and a straightforward introduction will be found in Davis (1973). The intention underlying the use of principal components or factor analysis in exploration geochemistry has been generally to separate the element associations inherent in the structure of the correlation matrix into a number of groups of elements which together account for the greater part of the observed variability of the original data. The intention being to represent the large number of elements in the original data by a smaller number of "factors", each of which is a linear function (transformation) of the element concentrations, thus giving a greater efficiency in terms of information compression over the original data, and hopefully also gaining something in interpretability. Techniques taking the correlation (or covariance) matrix as a starting point are generally referred to as R-mode.

In principal components analysis (PCA), the inter-element covariance matrix for m elements is decomposed into an orthogonal (uncorrelated) set of new variables which are linear combinations of the original element values. The eigenvectors and related eigenvalues are extracted in such a way that the first eigenvector (which may be regarded as the coefficients of the first linear transform equation) accounts for as much as possible of the total variability of the data, the second for as much as possible of the residual variance, and so on; m eigenvectors are needed to account for all of the variability of the original data. The eigenvectors will not necessarily be of unit length when based on the covariance matrix. However, since in many cases the measurement scales are different, it is common practice to treat the data as standardized (weighting all the variables equally in terms of measurement scale by subtracting the mean from each observation and dividing by the standard deviation). The covariance matrix calculated from standardized data is identical with the correlation matrix, and in the majority of geological PCA studies the correlation matrix is used as the

basis. In order to ensure reliable estimation of the multivariate structure of the data, the correlation matrix should be based on a data set with preferably at least 10 times as many samples as variables (Howarth, 1973a; Glasby et al., 1974; and Trochimczyk and Chayes, 1977).

In factor analysis, the correlations among the  $m$  elements are assumed to be accountable by a model in which the variance for any element is distributed between a number of "common" factors (with at least two elements significantly associated with it), and a remaining number of "unique" factors (with only one significantly associated element). The remaining variance being accounted for by an error term. The correlation matrix is the starting point and the resulting (normalized) eigenvectors are of equal unit length. It is hypothesized that the number of common factors ( $k$ ) is much less than  $m$ , and in the psychological studies from which the factor analysis approach originated,  $k$  was generally postulated a priori. A subjective element enters factor analysis at this stage. There are no statistically based tests which are universally agreed to define the significant number of common factors. The usual criteria are: a certain cumulative percentage of the overall variance is accounted for by  $k$  factors; those factors with corresponding eigenvalues exceeding 1.0 are considered to be "significant"; a plot of eigenvalues versus factor number shows a "significant" break in slope (Cattell's scree test; Cattell, 1966); factors associated "significantly" with more than one element are retained; factors which make sense from the point of view of geochemical interpretability are retained. In order to obtain more interpretable results, the original orthogonal linear transform may be rotated in a geometrical sense (although it is performed algebraically) until the coefficients for the elements are maximized in contrast, that is to say high ones tend to  $\pm 1.0$  and low ones to zero, in order to obtain a "simple structure". The varimax rotation method (Kaiser, 1959) is most commonly used although others have also been investigated (e.g. Chapman, 1978). Some investigators have also dispensed with representing the data variability in terms of a set of uncorrelated transformed variables, preferring so-called "oblique" solutions in which the transformed variables are themselves correlated (usually obtained by promax rotation, e.g. Chapman, 1978). There is some disagreement among investigators whether this is justifiable in geochemical terms, quite apart from the difference it may make to the interpretability of the data.

The end-product of the transform equations, when element values for the actual samples have been substituted in, are generally known as the "factor scores". Trochimczyk and Chayes (1978) recommend that if the components have been obtained from the covariance matrix, scores should be based on unstandardized data; and if components have been based on the correlation matrix, then the scores should be based on the standardized element values to ensure that the resultant score values are uncorrelated.

Thus we see that PCA gives a clearly based mathematical solution, and in terms of data compression, a set of PCA linear transform equations truncated at the  $k$ -th eigenvector is well defined. In contrast, the various factor analytic approaches are more subjective in choice of possible methods, and in difficulties of deciding how many factors should be retained (unless there is, unusually, clear a priori evidence about how many common factors might be expected to operate on the data) since the aim is to determine the best estimate of the number of common factors operating on the observed variates whether it be large or small. It should be understood by the geochemist that the interpretation of factors as the underlying causes giving rise to the observed element behaviour is not implicit in the model. Separation into "background" and "mineralization" element groups may occur in some cases, but factor analysis is not in itself

capable of sorting such associations out of the data in all cases. The appearance of such groups merely reflects the structure of the covariance (correlation) matrix which is subject to the same guidelines as those discussed earlier under correlation and regression. Occasionally rank correlation coefficients have been used in the initial matrix (R.P. Chapman, N.J. Marshall, pers. comms., 1975, 1976), although the use of a linear transform based on an essentially nonlinear coefficient poses some theoretical problems. Applications of R-mode PCA or factor analysis in exploration geochemistry include: Garrett and Nichol, 1969; Saager and Esselaar, 1969; Nairis, 1971; Nichol, 1971; Wennervita et al., 1971; Brundin and Nairis, 1972; Rose, 1972; Summerhayes, 1972; Garrett, 1973b; Obial and James, 1973; Smith and Webber, 1973; Glasby et al., 1974; Saager and Sinclair, 1974; Wennervita and Papunen, 1974; Shikawa et al., 1975; Ek and Elmliid, 1976; Chapman, 1978; and Mancey and Howarth, 1978.

A related technique, known as Q-mode factor analysis (Davis, 1973) attempts to find samples of extreme compositional type, and to represent all other samples in terms of mixtures of these "end-members", beginning with a matrix of intersample (cosine theta) similarity coefficients which is decomposed by factor analytic methods. In general, the results seem to have been more easily interpretable in petrologically oriented cases where the concept of end-member mixtures has genuine significance, than in studies based on stream-sediment data. Cluster analysis methods rather than Q-mode factor analysis, are now tending to be used for problems where it is required to assign the samples to similar groups. Examples of Q-mode factor analysis in exploration geochemistry include: Nichol and Webb, 1967; Garrett and Nichol, 1969; Armour-Brown and Nichol, 1970; Dawson and Sinclair, 1974; Saager and Sinclair, 1974; Chapman, 1975, 1978; and Culbert, 1976.

Two related techniques based upon the R-mode approach, but which allow both samples and element associations to be shown on the same diagram as an aid to interpretation, have recently been applied in exploration geochemistry. These are correspondence analysis (David and Dagbert, 1975; Teil, 1975) and the eigenvector biplot display (Sinding-Larsen, 1975); full details will be found in the references cited. These techniques will certainly find application to exploration geochemistry in the future, in accord with the wider usage of the R-mode techniques.

The effect of data transformation on the structure of the correlation matrix is illustrated by preliminary results from a study of 895 stream sediment samples from a regional geochemical survey of Denbighshire, UK, originally carried out by Nichol et al. (1970). PCA was performed on the basis of standardized results for Co, Cr, Cu, Fe<sub>2</sub>O<sub>3</sub>, Ga, Mo, Mn, Ni, Pb, Ti, V and Zn (all ppm except Fe<sub>2</sub>O<sub>3</sub>, per cent) from untransformed data, logtransformed data, and data subjected to the nonlinear Box and Cox (1964) transform (discussed earlier in the section on Univariate Statistics). The data include contributions from various background lithologies as well as minor areas of mineralization and contamination in the vicinity of smelting and mining operations.

It is apparent from Figure 24.10 that the effect of a global logtransformation is to induce negative skewness for all of the elements, although this is not as severe as the original positive skewness; kurtosis is also reduced. For some elements (Ga, Ti and V) which were approximately normally distributed originally, the effect of logtransformation has been particularly unhelpful. The Box-Cox transform has improved the approach to a normal distribution in all cases.

Comparison of the eigenvalues (Table 24.1) and unrotated PCA loadings (Fig. 24.11) for the three solutions indicates that the effect of the logtransform could (in this example) be to induce rather high positive correlations

Table 24.1

Eigenvalues (E.V.) and cumulative variance (Cum. Var.) for 895 stream sediment samples from regional geochemical survey of Denbighshire (U.K.) showing effects of nonlinear data transformation on correlation matrix

	Untransformed		Logtransformed		Box-Cox (1964)*	
	E.V.	Cum. Var. %	E.V.	Cum. Var. %	E.V.	Cum. Var. %
1	4.067	33.9	6.125	51.0	5.374	44.8
2	1.779	48.7	1.940	67.2	2.132	62.5
3	1.589	61.9	0.895	74.7	1.024	71.1
4	1.258	72.4	0.706	80.6	0.748	77.3
5	0.907	80.0	0.586	85.4	0.633	85.6
6	0.591	84.9	0.418	88.9	0.462	86.4
7	0.541	89.4	0.376	92.1	0.443	90.1
8	0.374	92.5	0.301	94.6	0.360	93.1
9	0.292	95.0	0.225	96.4	0.275	95.4
10	0.222	96.8	0.194	98.1	0.224	97.3
11	0.206	98.5	0.124	99.1	0.189	98.9
12	0.174	100.0	0.109	100.0	0.135	100.0

\* $(x^\lambda - 1)/\lambda$

between some of the element pairs. Following the Box-Cox transform the eigenvectors and unrotated PCA loadings (Fig. 24.11) show a tendency to have the high coefficients move towards  $\pm 1.0$ , and the low coefficients towards zero compared with the logtransform, although the general structure in terms of element associations remains the same. This suggests that alternatives to the logtransform could be very helpful in increasing the interpretability of multivariate data analysis in exploration geochemistry. A wider investigation is currently in progress at AGRG and is showing interesting results on a number of data sets.

The application of PCA-based factor analysis to a large regional data set, based on 21 432 moving average cell means for 22 elements from the Wolfson Geochemical Atlas of England and Wales (Webb et al., 1978) covering approximately 50 000 mi<sup>2</sup>, has shown in preliminary results that although known areas of mineralization form a very small proportion of the whole, relevant mineralization associations can be identified despite expected difficulties associated with the size of the data set and multiplicity of background types associated with regional lithological changes. The initial factor score maps have been colour-combined to further increase their information content (Mancey and Howarth, 1978), and are providing much useful information.

### Cluster Analysis

The aim of cluster analysis techniques is to group the samples together without a priori knowledge, on the basis of similarity in terms of their elemental compositions, in order to separate areas of similar background composition, and identify samples of unusual composition. Generally the Euclidean distance measure is used:

$$d_{ij} = \left[ \sum_{q=1}^m (x_{iq} - x_{jq})^2 \right]^{1/2}$$

where  $x_{iq}$  and  $x_{jq}$  are the  $q$ -th measurements on the  $i$ -th and  $j$ -th samples respectively,  $q=1, \dots, m$ . To increase computational speed the related metric:

$$d_{ij} = \sum_{q=1}^m |x_{iq} - x_{jq}|$$

known as the City block or "Manhattan" distance, may also be used.

PCA is often used as a suitable decorrelation preprocessor to cluster analysis, to transform the original data into a set in which the new variables are genuinely uncorrelated (and hence orthogonal for the calculation of  $d$ ). It should be remembered that if the last few components are not retained in this case, then information characterizing the less common groups of samples may have been sacrificed, and it is probably safer to retain all components resulting from the PCA transform initially.

A detailed review of the numerous possible methods is not appropriate here (see, for example, Everitt, 1974, or Duda and Hart, 1973). It is sufficient to point out that the applicability of many techniques to very large sets of geochemical data is limited by the necessity to keep track of all possible pairwise distances within the data set, which will soon build up a computationally impossible number of items. In order to overcome this difficulty samples are either: 1) grouped into classes if their composition is within a threshold distance of an existing class centre, represented by its mean composition; or 2) a random subset of the original data is clustered using a preferred method and these clusters are subsequently taken as training information for discriminant analysis (discussed below) which classifies all the remaining samples in to one of these groups. Either method is reasonably fast to implement and can be tailored to existing computer requirements. Techniques of the former type generally aim to obtain compact classes, in the sense that between-group variance exceeds within-group variance and it is possible in some methods (e.g. ISODATA, Ball and Hall, 1966) to split or merge the clusters as the calculation

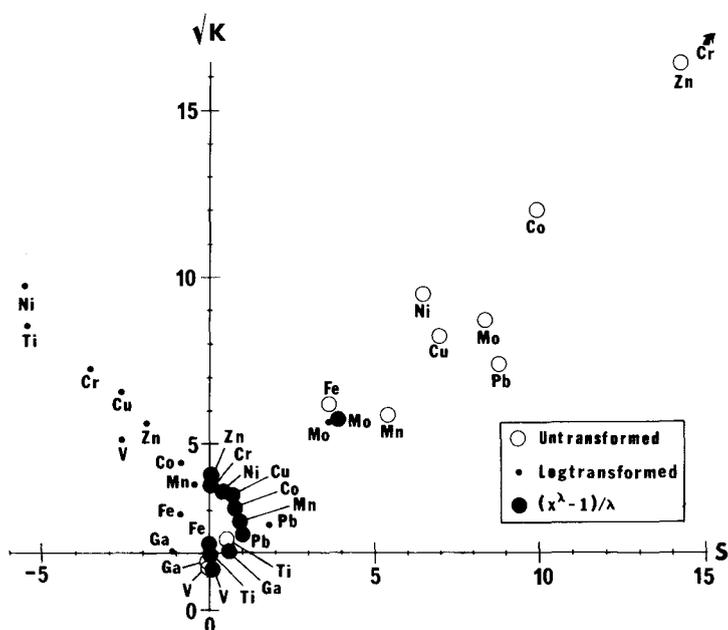


Figure 24.10. Effect of data transformations on symmetry (skewness, S) and peakedness (kurtosis, K) of frequency distributions for 12 elements in 895 stream sediment samples from a regional survey of Denbighshire, England. Origin of the axes placed at expected values for these parameters for a Normal distribution; note that ordinate is scaled as  $\sqrt{K}$ .

proceeds in order to optimize the clustering criterion. While they may be somewhat artificial as regards "naturalness" of the resultant classification they have the advantage of speed. More complex techniques in current use include the tree (dendrogram) methods (Obial, 1970; Lenthall, 1972; Hesp, 1973; Hesp and Rigby, 1973; Howarth, 1973a; Obial and James, 1973; and Glasby et al., 1974) yielding a hierarchical structure which may be better suited to interpretational requirements, or attempt to obtain a completely natural classification based on optimal two-dimensional representation of the multi-dimensional data structure by methods of the nonlinear mapping type (Garrett, 1973b; Howarth, 1973a, b). Sinding-Larsen (1975) implemented a technique similar to ISODATA following a PCA type of transform, and Duval (1976) used a distance threshold grouping in combination with Q-mode factor analysis.

Detailed studies based on approximately 900 regional stream sediment samples from Denbighshire, UK, on the basis of 12 elements, using hierarchical cluster analysis and nonlinear mapping (Castillo-Muñoz, 1973) and ISODATA (Crisp, 1974), suggest that scaling effects can be important in all three techniques. "Anomalous" samples of unusual composition could be detected by all the methods used and in the untransformed data these were generally clearly identified. In general however, the groups based on logtransformed data were comparable for the three methods and gave more interpretable results. ISODATA was very much the fastest, taking approximately 300 seconds on the CDC 6400 central processor to process the entire data set. Nonlinear mapping in contrast took approximately 1200 seconds for 500 samples.

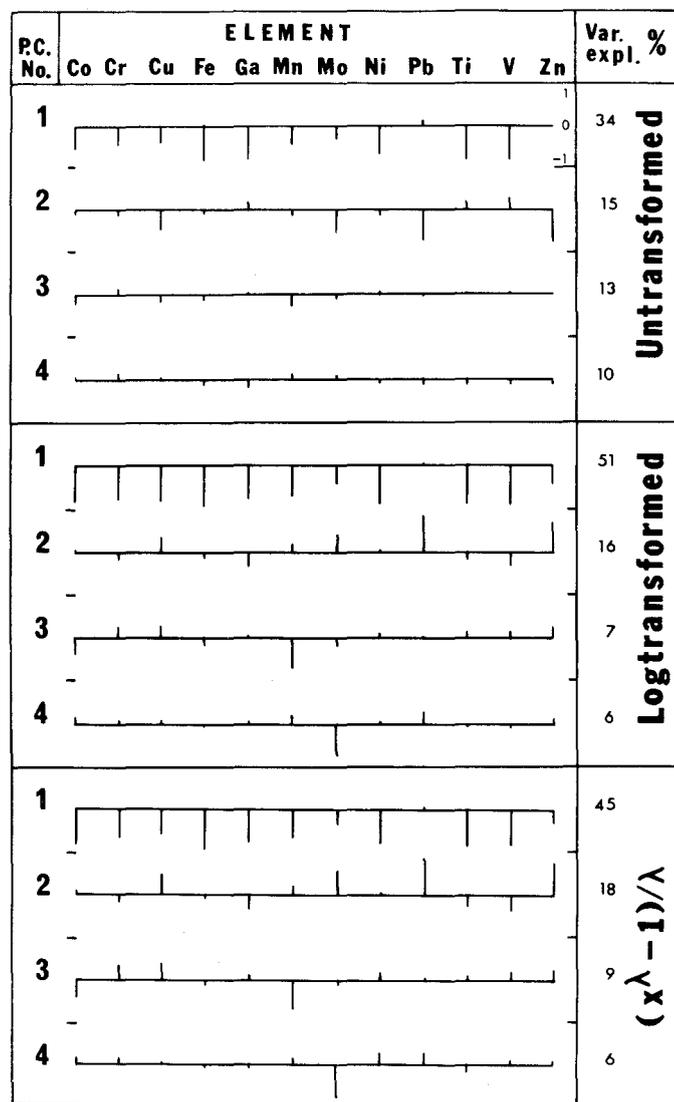


Figure 24.11. Effect of data transformations on the loadings of the first four unrotated principal components for the data of Figure 24.10. Line length indicates strength of loading (between -1 and +1); percentage of overall variance accounted for by each component also shown.

These results suggest that methods based on group mean compositions are more suitable for processing large sets of exploration data where rapid division into broad groups is required, although a number of rather critical parameters must be empirically set by the user to control the clustering process. The group mean compositions could always be reclustered by another method if it is required to verify the nature of the groupings. It must be borne in mind that with any clustering method there is no well-defined criterion for the number of clusters present, and the final interpretation is somewhat subjective.

### Discriminant Analysis

In contrast to cluster analysis, discriminant analysis aims to classify samples of unknown type into one of a number of pre-defined groups (e.g. mineralized or barren) on the basis of their elemental compositions. In all methods, irrespective of the nature of the decision rule (discriminant function) used, the starting point is to have an established "training set" of samples representing each group. This should be as large as possible for each class, to ensure adequate statistical representivity (the rule of thumb for the number of samples per class of at least 10 times the number of variables also applies here for most reliable performance, although as few as 3 times the number of variables is adequate if training samples are hard to obtain). Ideally, once the decision criteria have been determined, an independent "test set" of samples should be classified on the basis of the established decision rules to find out how well they perform before applying them to known data. A detailed discussion will be found in Howarth (1973a), in a geochemical context, or Duda and Hart (1973).

Although extensively used in other branches of the earth sciences and in igneous petrology, discriminant functions have not been widely used in exploration geochemical problems to date. The reason for this is almost certainly the practical difficulties associated with obtaining adequate data for formulating the training sets. The majority of applications have used Fisher (1936) linear discriminant functions (Wignall, 1969, 1970; Dahlberg, 1970; Cameron et al., 1971; Govett and Pantazis, 1971; Lenthall, 1972; Rose, 1972; Govett, 1973; Pantazis and Govett, 1973; Shiikawa et al., 1974; Whitehead and Govett, 1974; Chapman, 1975, 1978; Govett et al., 1975; and Culbert, 1976) and because of the inherent simplicity of the method (Davis, 1973), and availability of suitable programs in packages such as BMDP (Dixon, 1975), and SPSS (Nie et al., 1975), it continues to be widely used in industry. An alternative method, which does not depend on the multivariate normal assumptions inherent in the Fisher linear discriminant is the "empirical discriminant function" (Howarth, 1973a, c) which is a form of potential function (Duda and Hart, 1973). This has proven very successful in regional geochemical survey classification for multi-class applications (Howarth, 1971b, 1972, 1973a; Castillo-Muñoz and Howarth, 1976) and is being used increasingly for exploration problems (Gustavsson and Bjorklund, 1976; C.Y. Chork, pers. comm., 1976; N. Gustavsson, pers. comm., 1977). Occasionally discrimination problems can be solved graphically by a suitable choice of elements (Govett, 1972; Joyce and Clema, 1974; Bull and Mazzuchelli, 1975).

It is known that there is an increasing interest among mining companies in techniques for selecting the "best" subset of elements from a set of multi-element analyses to distinguish between mineralized and unmineralized bedrock, with the aim of obtaining reliable discriminant functions. A variety of methods are available for this type of problem (Howarth, 1973a, c; Duda and Hart, 1973). Once the discriminant functions have been evaluated subsequent classification is extremely rapid, even when several thousand samples are screened. It is for this reason that combination of discriminant analysis following cluster analysis of a subset of a large data set (the groups of samples obtained becoming the training set for the discriminant function) is a very powerful technique for dealing with a large data set.

A somewhat related technique, known as Characteristic Analysis, has been developed by Botbol et al. (1978) as an extension of a method originally devised by him for the study of mineral assemblages. Geochemical and other information, such as geology or geophysical survey data, is expressed in binary (0/1) form. A value 1 indicates favourability for exploration in a local spatial context, and is defined on the

basis that the measured value for a variable in a map cell is higher than its surrounding neighbours. Thus second derivatives of geological, geochemical or geophysical parameters can be treated in a similar manner. Training sets are established for known mineralized areas and comparison of the characteristics of the matrix of joint "favourable" occurrences for the training set with the coded binary data for the unknown cells allows their similarity to the training set to be determined, thus establishing the regional classification. Results from the Coeur d'Alene district, Idaho (Botbol et al., 1978) are very encouraging.

### Discussion

The multivariate techniques described in this section are complex and are best performed on a digital computer because of necessity for speed and accuracy, irrespective of the problems associated with the large data sets produced in exploration geochemistry. The majority of the techniques can generally be programmed for storage requirements of the order of  $m^2/2$  for  $m$  elements, but (as discussed above) cluster analysis methods generally require storage of the order of  $n^2/2$  intersample similarity values for  $n$  samples and are therefore more difficult to carry out without very large computer facilities, unless methods storing only the group mean compositions are used.

However, taken as a whole, the multivariate approach offers an immensely more powerful means of data analysis than univariate techniques because of the ability to study the interaction of the various elements. It is often extremely difficult to reliably predict the joint behaviour of a number of elements from a study of their univariate characteristics

As instrumental facilities increase the ease with which multi-element analyses become routinely available at low cost, investigators will be forced to turn to multivariate techniques in order to make optimum use of the information at their disposal. The published case histories cited in this survey show that in most instances multivariate analysis provided greater insight into the geochemical behaviour of the multi-element survey. The basic dilemma for the exploration geochemist remains as discussed by Lepeltier (1977), that the application of powerful computer methods to the data resulting from exploration projects often awaits collation of the total data set, and may then come too late to aid the decision-making (the time of which is often dictated by political or economic factors outside the geochemist's control). In the long term, however, a better understanding of the geochemical relationships governing observed geochemical behaviour, the discovery of common element associations, and interrelationship with geophysical data (e.g. Culbert, 1976) etc. must be beneficial to prediction of criteria for the likelihood of occurrence of a geochemical target.

At the present state-of-the-art of geochemical data interpretation it is probably true that what we need is not a widespread development of new methods of analysis, but adaption of existing techniques to the small-computer environment and the modification of methods for the treatment of large data sets. It is also highly desirable that standard data sets be made available on which comparative evaluations of existing methods may be made (the basis for this is currently being established by the Centre de Recherche Pétrographique et Géochimique, Nancy). Release of "old" exploration geochemical data by mining companies for research purposes, once confidentiality restrictions can be lifted, would be very helpful for this type of study. A growing impetus will probably come from national geological surveys, since the establishment of multi-purpose multi-element regional geochemical surveys is now being seen as increasingly desirable in the context of resource evaluation

and environmental monitoring. It may be confidently predicted that with the gradual spread of multi-element analysis and availability of powerful microprocessors, together with an increasingly numerate education of geologists and geochemists, the routine application of multivariate analysis in exploration as an aid to general understanding of geochemical behaviour will become established.

Perhaps with the spread of minicomputers and the ability to transmit data from remote sites to a central computer by short-wave radio (Botbol, 1975, 1977), integration of field operations with computer processing for short-term turn-around of results will become easier. It is unfortunate that in many cases the power and flexibility of the computer for mapping or statistical analysis are not being utilized until it is too late.

#### ACKNOWLEDGMENTS

We would like to thank the many individuals and organizations who responded to our requests for information on current practice in the application of computers for exploration geochemical data processing. Our thanks also go to the referees for their helpful comments on the manuscript, and for encouragement to track down some references from the early history of statistics. Miss J. Nielsen and Mrs. J. O'Donnell are also thanked for typing versions of this manuscript.

#### SELECTED BIBLIOGRAPHY

- Agterberg, F.P. and Chung, C.F.  
1975: A computer program for polynomial trend surface analysis; *Geol. Surv. Can., Paper 75-21*, 51 p.
- Allcott, G.H. and Lakin, H.W.  
1975: The homogeneity of six geochemical exploration reference samples; in *Geochemical Exploration 1974*, I.L. Elliot and W.K. Fletcher (ed.), Elsevier Publ. Co., Amsterdam, p. 659-681.
- Allredge, J.R. and Allredge, N.G.  
1978: *Geostatistics: a bibliography*; *Int. Stat. Rev.*, v. 46, p. 77-88.
- Armour-Brown, A. and Nichol, I.  
1970: Regional geochemical reconnaissance and the location of metallogenic provinces; *Econ. Geol.*, v. 65, p. 312-330.
- Austria, V. and Chork, C.Y.  
1976: A study of the application of regression analysis for trace element data from stream sediment in New Brunswick; *J. Geochem. Explor.*, v. 6, p. 211-232.
- Ball, G.H. and Hall, D.J.  
1966: ISODATA, an interactive method of multivariate data analysis and pattern classification; *IEEE Int. Communications Conf.*, p. 116-117.
- Barringer, A.R.  
1977: A multi-disciplinary approach to airborne mineral exploration; Unpub. ms. (presented at Univ. California, Berkeley, 22 April 1977), Barringer Research Ltd., Rexdale, 64 p.
- Baucom, E.I., Price, V., and Ferguson, R.B.  
1977: Savannah River Laboratory hydrogeochemical and stream sediment reconnaissance. Preliminary raw data release Winston-Salem 1° x 2° NTMS area. North Carolina, Virginia, Tennessee; E.I. DuPont de Nemours Co., Savannah River Laboratory, Aiken, S. Carolina, DPST-77-146-1/GJBX-66(77), 98 p.
- Bendix Field Engineering Corporation  
1976: Annual NURE Report 1976; Bendix Field Engineering Corporation, Grand Junction, Colorado, Rept. GJBX-11(77), 75 p.  
1977: Symposium on hydrogeochemical and stream-sediment reconnaissance for uranium in the United States; Bendix Field Engineering Corporation, Grand Junction, Colorado, Rept. GJBX-77(77), 468 p.  
1978: NURE 1977. National Uranium Resource Evaluation annual activity report, May 1978; Bendix Field Engineering Corporation, Grand Junction, Colorado, Rept. GJBX-11(78)R, 91 p.
- Berk, K.N.  
1978: Comparing subset regression procedure; *Technometrics*, v. 20, p. 1-6.
- Bolviken, B.  
1971: A statistical approach to the problem of interpretation in geochemical prospecting; *Can. Inst. Min. Met., Spec. Vol. 11*, p. 564-567.
- Bolviken, B. and Sinding-Larsen, R.  
1973: Total error and other criteria in the interpretation of stream-sediment data; in *Geochemical Exploration 1972*, M.J. Jones (ed.), *Inst. Min. Met., London*, p. 285-295.
- Bosman, E.R., Eckhart, D., and Kubik, K.  
1971: Application of automatic data processing to problems of exploration geochemistry; *Geol. Mijnb.*, v. 50, p. 768-770.
- Botbol, J.M.  
1975: Field site data processing. A high frequency radio communication link between field camp and computer; *Am. Inst. Min. Eng., Preprint 75-L-18*, Annu. Mtg., New York, Feb. 1975, 16 p.  
1977: The use of the storage cathode ray tube (SCRT) in geochemical exploration; *Sci. Terre, Ser. Inform. No. 9*, p. 93-103.
- 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.
- Bowen, R.W. and Botbol, J.M.  
1975: The Geologic Retrieval and Synopsis Program (GRASP); *U.S. Geol. Surv., Prof. Paper 966*, 87 p.
- Box, G.E.P. and Cox, D.R.  
1964: An analysis of transformations; *J. Roy. Statist. Soc., Ser. B*, v. 26, p. 211-243.
- Brabec, D. and White, W.H.  
1971: Distribution of copper and zinc in rocks of the Guichon Creek batholith, British Columbia; *Can. Inst. Min. Met., Spec. Vol. 11*, p. 291-297.
- Brundin, N.H. and Nairis, B.  
1972: Alternative sample types in regional geochemical prospecting; *J. Geochem. Explor.*, v. 1, p. 7-46.
- Bull, A.J. and Mazzucchelli, R.H.  
1975: Application of discriminant analysis to the geochemical evaluation of gossans; in *Geochemical Exploration 1974*, I.L. Elliott and W.K. Fletcher (ed.), Elsevier Publ. Co., Amsterdam, p. 219-226.
- Burch, C.R. and Parsons, I.T.  
1976: "Squeeze" significance tests; *Appl. Stat.*, v. 25, p. 287-291.

- Cachau-Hereillat, F.  
1975: Towards quantitative utilization of geochemical exploration: The threshold and its estimation in soil surveys; in *Geochemical Exploration 1974*; I.L. Elliott and W.K. Fletcher (ed.), Elsevier Publ. Co., Amsterdam, p. 183-189.
- Cameron, E.M.  
1968: A geochemical profile of the Swan Hills Reef; *Can. J. Earth Sci.*, v. 5, p. 287-309.
- Cameron, E.M. and Hornbrook, E.H.W.  
1976: Current approaches to geochemical reconnaissance for uranium in the Canadian Shield; in *Exploration for Uranium Ore Deposits*, IAEA, Vienna, p. 241-266.
- Cameron, E.M., Siddeley, G., and Durham, C.C.  
1971: Distribution of ore elements in rocks for evaluating ore potential: Nickel, copper, cobalt and sulphur in ultramafic rocks of the Canadian Shield; *Can. Inst. Min. Met., Spec. Vol. 11*, p. 298-314.
- Castillo-Muñoz, R.  
1973: Application of discriminant and cluster analysis to regional geochemical surveys; Unpubl. Ph.D. thesis, London University, 258 p.
- Castillo-Muñoz, R. and Howarth, R.J.  
1976: Application of the empirical discriminant function to regional geochemical data from the United Kingdom; *Bull. Geol. Soc. Amer.*, v. 87, p. 1567-1581.
- Cattell, R.B.  
1966: Handbook of multivariate experimental psychology; Rand McNally, Chicago, p. 174-243.
- Chapman, R.P.  
1975: Data processing requirements and visual representation for stream sediment exploration geochemical surveys; *J. Geochem. Explor.*, v. 4, p. 409-423.  
1976a: Limitations of correlation and regression analysis in geochemical exploration; *Inst. Min. Met. Trans.*, London, Sect. B, v. 85, p. 279-283.  
1976b: Some consequences of applying lognormal theory to pseudolognormal distributions; *Math. Geol.*, v. 8, p. 209-214.  
1977: Logtransformations in exploration geochemistry; *Math. Geol.*, v. 9, p. 194-198.  
1978: Evaluation of some statistical methods of interpreting geochemical drainage data from New Brunswick; *Math. Geol.*, v. 10, p. 195-224.
- Chatupa, J. and Fletcher, K.  
1972: Application of regression analysis to the study of background variations in trace metal content of stream sediments; *Econ. Geol.*, v. 67, p. 978-980.
- Chayes, F.  
1960: On correlation between variables of constant sum; *J. Geophys. Res.*, v. 65, p. 4185-4193.  
1970a: On deciding whether trend surfaces of progressively higher order are meaningful; *Bull. Geol. Soc. Amer.*, v. 81, p. 1273-1278.  
1970b: Effect of a single nonzero open covariance on the simple closure test; in *Geostatistics*, D.F. Merriam (ed.), Plenum Press, New York, p. 11-22.  
1971: *Ratio Correlation*; Univ. Chicago Press, Chicago, 99 p.
- Chayes, F. and Kruskal, W.  
1966: An approximate statistical test for correlations between proportions; *J. Geol.*, v. 74, p. 692-702.
- Chork, C.Y.  
1977: Seasonal, sampling and analytical variations in stream sediment surveys; *J. Geochem. Explor.*, v. 7, p. 31-47.
- Clark, I.  
1977: ROKE, a computer program for non-linear least-squares decomposition of mixtures of distributions; *Comput. Geosci.*, v. 3, p. 245-256.
- Clark, I. and Garnett, R.H.T.  
1974: Identification of multiple mineralization phases by statistical methods; *Inst. Min. Met. Trans.*, London, Sec. A, v. 83, p. 43-52.
- Clark, M.W.  
1976: Some methods for statistical analysis of multimodal distributions and their application to grain-size data; *Math. Geol.*, v. 8, p. 267-282.
- Cockburn, G.H.  
1977: *Atlas Géochimique des Sédiments du Ruisseau La Grande Rivière*; *Min. Rich. Nat. Quebec*, 500 p.
- Cohen, A.C.  
1959: Simplified estimators for the normal distribution when samples are singly censored or truncated; *Technometrics*, v. 1, p. 217-237.  
1961: Tables for maximum likelihood estimates: Singly truncated and singly censored samples; *Technometrics*, v. 3, p. 535-541.
- Connor, J.J., Tidball, R.R., Erdman, J.A., Ebens, R.J., and Felder, G.L.  
1974: Geochemical survey of the Western Coal Regions; *U.S. Geol. Surv., Open File Rep. 74-250*, 38 p.
- Connor, J.J. and Miesch, A.T.  
1964: Analysis of geochemical prospecting data from the Rocky Range, Beaver County, Utah; *U.S. Geol. Surv., Prof. Paper 475-D*, p. 79-83.
- Crisp, D.A.  
1974: Application of multivariate methods to geochemistry: The evaluation of a new technique; Unpub. M.Sc. thesis, London Univ., 113 p.
- Croissant, A.  
1977: La géostatistique comme outil dans la prospection géochimique; *Sci. Terre, Ser. Inform. No. 9*, p. 129-144.
- Culbert, R.  
1976: A multivariate approach to mineral exploration; *Can. Inst. Min. Met. Bull.*, v. 69, p. 39-52.
- Dahlberg, E.C.  
1969: Use of model for relating geochemical prospecting data to geologic attributes of a region, South Mountain, Pennsylvania; *Quart. J. Colo. Sch. Mines*, v. 64, p. 195-216.  
1970: Generalized Bayesian classification: K classes; Discussion; *Econ. Geol.*, v. 65, p. 220-222.  
1971: Algorithmic development of a geochemical exploration program; *Can. Inst. Min. Met., Spec. Vol. 11*, p. 577-580.
- Dall'Aglia, M.  
1971: Comparison between hydrogeochemical and stream-sediment methods in prospecting for mercury; *Can. Inst. Min. Met., Spec. Vol. 11*, p. 126-131.

- Geological Survey of Uganda  
1973: Geochemical Atlas of Uganda, Entebbe, 42 p.
- Glasby, G.P., Tooms, J.S., and Howarth, R.J.  
1974: Geochemistry of manganese concretions from the northwest Indian Ocean; *New Zealand J. Sci.*, v. 17, p. 387-407.
- Gordon, T. and Hutchison, W.W. (Ed.)  
1974: Computer use in projects of the Geological Survey of Canada; *Geol. Surv. Can.*, Paper 74-60, 108 p.
- Gordon, T. and Martin, G.  
1974: Computer-based data management in the Geological Survey of Canada — a preliminary appraisal; *Geol. Surv. Can.*, Paper 74-60, p. 7-14.
- Govett, G.J.S.  
1972: Interpretation of a rock geochemical exploration survey in Cyprus — statistical and graphical techniques; *J. Geochem. Explor.*, v. 1, p. 77-102.
- Govett, G.J.S., Goodfellow, W.D., Chapman, R.P., and Chork, C.Y.  
1975: Exploration geochemistry — distribution of elements and recognition of anomalies; *Math. Geol.*, v. 7, p. 415-446.
- Govett, G.J.S. and Pantazis, Th.M.  
1971: Distribution of Cu, Zn, Ni and Co in the Troodos Pillow Lava Series; *Inst. Min. Met. Trans.*, London, Sect. B, v. 80, p. 27-46.
- Gregory, P. and Tooms, J.S.  
1969: Geochemical prospecting for kimberlites; *Quart. J. Colo. Sch. Mines*, v. 64, p. 265-305.
- Guarascio, M., David, M., and Huijbregts, C.  
1976: Advanced Geostatistics in the Mining Industry; Ridel, Dordrecht, 461 p.
- Gustavsson, N. and Bjorklund, A.  
1976: Lithological classification of tills by discriminant analysis; *J. Geochem. Explor.*, v. 5, p. 393-395.
- Hawkes, H.E.  
1976: The downstream dilution of stream sediment anomalies; *J. Geochem. Explor.*, v. 6, p. 345-358.
- Hawkes, H.E. and Webb, J.S.  
1962: Geochemistry in mineral exploration; Harper and Row, New York, 415 p.
- Hesp, W.R.  
1971: Correlations between the tin content of granitic rocks and their chemical and mineralogical composition; *Can. Inst. Min. Met., Spec. Vol. 11*, p. 341-353.  
1973: Classification of igneous rocks by cluster analysis; *Acta Geol. Acad. Sci. Hung.*, v. 17, p. 339-362.
- Hesp, W.R. and Rigby, D.  
1973: Cluster analysis of rocks in the New England igneous complex, New South Wales, Australia; in *Geochemical Exploration 1972*, M.H. Jones (ed.), *Inst. Min. Met.*, London, p. 221-235.
- Hill, W.E.  
1975a: Analytical standards for the mineral industry; unpub. ms. (Presented at AIME Mtg., Broadmoor), 16 p.  
1975b: The use of analytical standards to control assaying projects; in *Geochemical Exploration 1974*, I.L. Elliott and W.K. Fletcher (ed.), Elsevier Publ. Co., Amsterdam, p. 651-657.
- Hodgson, W.A.  
1973: Optimum spacing for soil sample traverses; in *Application of Computer Methods in the Mineral Industry*, S. Afr. Inst. Min. Met., Johannesburg, p. 75-78.
- Hornbrook, E.H.W. and Garrett, R.G.  
1976: Regional geochemical lake sediment survey, east-central Saskatchewan; *Geol. Surv. Can.*, Paper 75-41, 20 p.
- Howarth, R.J.  
1967: Trend surface fitting to random data — an experimental test; *Amer. J. Sci.*, v. 265, p. 619-625.  
1971a: Fortran IV program for grey-level mapping of spatial data; *Math. Geol.*, v. 3, p. 95-121.  
1971b: Empirical discriminant classification of regional stream sediment geochemistry in Devon and east Cornwall; *Inst. Min. Met. Trans.*, London, Sect. B, v. 80, p. 142-149.  
1972: Empirical discriminant classification of regional stream sediment geochemistry in Devon and east Cornwall. Reply to discussion; *Inst. Min. Met. Trans.*, London, Sect. B, v. 81, p. 115-119.  
1973a: The pattern recognition problem in applied geochemistry; in *Geochemical Exploration 1972*, M.J. Jones (ed.), *Inst. Min. Met.*, London, p. 259-273.  
1973b: Preliminary assessment of a non-linear mapping algorithm in a geological context; *Math. Geol.*, v. 5, p. 39-57.  
1973c: Fortran IV programs for empirical discriminant classification of spatial data; *Geocom Bull.*, v. 6, p. 1-31.  
1974: The impact of pattern recognition methodology in geochemistry; *Proc. 2nd Intl. Joint Conf. Pattern Recognition*, Copenhagen, IEEE Pub. 74CH0885-4C, p. 411-412.  
1977: Automatic generation of randomized sample submittal schemes for laboratory analysis; *Comput. Geosci.*, v. 3, p. 327-334.
- Howarth, R.J. and Earle, S.A.M.  
1979: Application of a generalized power transform to geochemical data; *Math. Geol.*, v. 11, p. 45-48.
- Howarth, R.J. and Lowenstein, P.L.  
1971: Sampling variability of stream sediments in broad-scale regional geochemical reconnaissance; *Inst. Min. Met. Trans.*, London, Sect. B, v. 80, p. 363-372.  
1972: Sampling variability of stream sediments in broad-scale regional geochemical reconnaissance. Discussion; *Inst. Min. Met. Trans.*, London, Sect. B, v. 81, p. 122-123.  
1974: Data processing for the Provisional Geochemical Atlas of Northern Ireland; Applied Geochemistry Research Group, London, *Tech. Comm.* 61, 8 p.  
1976: Three-component colour maps from lineprinter output; *Inst. Min. Met. Trans.*, London, Sect. B, v. 85, p. 234-237.
- Howarth, R.J. and Thompson, M.  
1976: Duplicate analysis in geochemical practice. Part 2. Examination of proposed method and examples of its use; *Analyst*, v. 101, p. 699-709.

- Geological Survey of Uganda  
1973: Geochemical Atlas of Uganda, Entebbe, 42 p.
- Glasby, G.P., Tooms, J.S., and Howarth, R.J.  
1974: Geochemistry of manganese concretions from the northwest Indian Ocean; *New Zealand J. Sci.*, v. 17, p. 387-407.
- Gordon, T. and Hutchison, W.W. (Ed.)  
1974: Computer use in projects of the Geological Survey of Canada; *Geol. Surv. Can.*, Paper 74-60, 108 p.
- Gordon, T. and Martin, G.  
1974: Computer-based data management in the Geological Survey of Canada — a preliminary appraisal; *Geol. Surv. Can.*, Paper 74-60, p. 7-14.
- Govett, G.J.S.  
1972: Interpretation of a rock geochemical exploration survey in Cyprus — statistical and graphical techniques; *J. Geochem. Explor.*, v. 1, p. 77-102.
- Govett, G.J.S., Goodfellow, W.D., Chapman, R.P., and Chork, C.Y.  
1975: Exploration geochemistry — distribution of elements and recognition of anomalies; *Math. Geol.*, v. 7, p. 415-446.
- Govett, G.J.S. and Pantazis, Th.M.  
1971: Distribution of Cu, Zn, Ni and Co in the Troodos Pillow Lava Series; *Inst. Min. Met. Trans.*, London, Sect. B, v. 80, p. 27-46.
- Gregory, P. and Tooms, J.S.  
1969: Geochemical prospecting for kimberlites; *Quart. J. Colo. Sch. Mines*, v. 64, p. 265-305.
- Guarascio, M., David, M., and Huijbregts, C.  
1976: Advanced Geostatistics in the Mining Industry; *Ridel, Dordrecht*, 461 p.
- Gustavsson, N. and Bjorklund, A.  
1976: Lithological classification of tills by discriminant analysis; *J. Geochem. Explor.*, v. 5, p. 393-395.
- Hawkes, H.E.  
1976: The downstream dilution of stream sediment anomalies; *J. Geochem. Explor.*, v. 6, p. 345-358.
- Hawkes, H.E. and Webb, J.S.  
1962: Geochemistry in mineral exploration; *Harper and Row, New York*, 415 p.
- Hesp, W.R.  
1971: Correlations between the tin content of granitic rocks and their chemical and mineralogical composition; *Can. Inst. Min. Met.*, Spec. Vol. 11, p. 341-353.  
1973: Classification of igneous rocks by cluster analysis; *Acta Geol. Acad. Sci. Hung.*, v. 17, p. 339-362.
- Hesp, W.R. and Rigby, D.  
1973: Cluster analysis of rocks in the New England igneous complex, New South Wales, Australia; in *Geochemical Exploration 1972*, M.H. Jones (ed.), *Inst. Min. Met.*, London, p. 221-235.
- Hill, W.E.  
1975a: Analytical standards for the mineral industry; unpub. ms. (Presented at AIME Mtg., Broadmoor), 16 p.  
1975b: The use of analytical standards to control assaying projects; in *Geochemical Exploration 1974*, I.L. Elliott and W.K. Fletcher (ed.), Elsevier Publ. Co., Amsterdam, p. 651-657.
- Hodgson, W.A.  
1973: Optimum spacing for soil sample traverses; in *Application of Computer Methods in the Mineral Industry*, S. Afr. Inst. Min. Met., Johannesburg, p. 75-78.
- Hornbrook, E.H.W. and Garrett, R.G.  
1976: Regional geochemical lake sediment survey, east-central Saskatchewan; *Geol. Surv. Can.*, Paper 75-41, 20 p.
- Howarth, R.J.  
1967: Trend surface fitting to random data — an experimental test; *Amer. J. Sci.*, v. 265, p. 619-625.  
1971a: Fortran IV program for grey-level mapping of spatial data; *Math. Geol.*, v. 3, p. 95-121.  
1971b: Empirical discriminant classification of regional stream sediment geochemistry in Devon and east Cornwall; *Inst. Min. Met. Trans.*, London, Sect. B, v. 80, p. 142-149.  
1972: Empirical discriminant classification of regional stream sediment geochemistry in Devon and east Cornwall. Reply to discussion; *Inst. Min. Met. Trans.*, London, Sect. B, v. 81, p. 115-119.  
1973a: The pattern recognition problem in applied geochemistry; in *Geochemical Exploration 1972*, M.J. Jones (ed.), *Inst. Min. Met.*, London, p. 259-273.  
1973b: Preliminary assessment of a non-linear mapping algorithm in a geological context; *Math. Geol.*, v. 5, p. 39-57.  
1973c: Fortran IV programs for empirical discriminant classification of spatial data; *Geocom Bull.*, v. 6, p. 1-31.  
1974: The impact of pattern recognition methodology in geochemistry; *Proc. 2nd Intl. Joint Conf. Pattern Recognition, Copenhagen, IEEE Pub. 74CH0885-4C*, p. 411-412.  
1977: Automatic generation of randomized sample submittal schemes for laboratory analysis; *Comput. Geosci.*, v. 3, p. 327-334.
- Howarth, R.J. and Earle, S.A.M.  
1979: Application of a generalized power transform to geochemical data; *Math. Geol.*, v. 11, p. 45-48.
- Howarth, R.J. and Lowenstein, P.L.  
1971: Sampling variability of stream sediments in broad-scale regional geochemical reconnaissance; *Inst. Min. Met. Trans.*, London, Sect. B, v. 80, p. 363-372.  
1972: Sampling variability of stream sediments in broad-scale regional geochemical reconnaissance. Discussion; *Inst. Min. Met. Trans.*, London, Sect. B, v. 81, p. 122-123.  
1974: Data processing for the Provisional Geochemical Atlas of Northern Ireland; *Applied Geochemistry Research Group, London, Tech. Comm. 61*, 8 p.  
1976: Three-component colour maps from lineprinter output; *Inst. Min. Met. Trans.*, London, Sect. B, v. 85, p. 234-237.
- Howarth, R.J. and Thompson, M.  
1976: Duplicate analysis in geochemical practice. Part 2. Examination of proposed method and examples of its use; *Analyst*, v. 101, p. 699-709.

- Hubaux, A. and Smiriga-Snoeck, N.  
1964: On the limit of sensitivity and the analytical error; *Geochem. Cosmochim. Acta*, v. 28, p. 1199-1216.
- Huijbregts, C. and Matheron, G.  
1971: Universal kriging (an optimal method for estimating and contouring in trend surface analysis); *Can. Inst. Min. Met., Spec. Vol. 12*, p. 159-169.
- Hutchison, W.W. (Ed.)  
1975: Computer-based systems for geological field data; *Geol. Surv. Can., Paper 74-63*, 100 p.
- Ingamells, C.O., Engels, J.C., and Switzer, P.  
1972: Effect of laboratory sampling error in geochemistry and geochronology; *Proc. 24th Inter. Geol. Congr., Montreal, Sect. 10*, p. 405-415.
- James, C.H.  
1970: A rapid method for calculating the statistical precision of geochemical prospecting analysis; *Inst. Min. Met. Trans., London, Sect. B*, v. 79, p. 88-89.
- Jeffery, K.G. and Gill, E.M.  
1976: The design philosophy of the G-EXEC system; *Comput. Geosci.*, v. 2, p. 345-346.
- Jöreskog, K.G., Klován, J.E., and Reymont, R.A.  
1976: *Geological Factor Analysis*; Elsevier Publ. Co., Amsterdam, 178 p.
- Journel, A.  
1975: Geological reconnaissance to exploitation — a decade of applied geostatistics; *Can. Inst. Min. Met.*, June 1975, p. 75-84.
- Joyce, A.S. and Clema, J.M.  
1974: An application of statistics to the chemical recognition of nickel gossans in the Yilgam Block, Western Australia; *Proc. Austral. Inst. Min. Met.*, No. 252, p. 21-24.
- Kaiser, H.F.  
1959: Computer program for Varimax rotation in factor analysis; *Educ. Psychol. Meas.*, v. 19, p. 413-420.
- Kayser, R.B. and Parry, W.T.  
1971: A geochemical exploration experiment on the Texas Canyon Stock, Cochise Country, Arizona; *Can. Inst. Min. Met., Spec. Vol. 11*, p. 354-356.
- Kendall, M.G.  
1938: A new measure of rank correlation; *Biometrika*, v. 30, p. 81-93.
- Kork, J.O.  
1977: Examination of the Chayes-Kruskal procedure for testing correlations between proportions; *Math. Geol.*, v. 9, p. 543-562.
- Kremer, M., Lenci, M., and Lesage, M.T.  
1976: SIGMI: A user-oriented file-processing system; *Comput. Geosci.*, v. 1, p. 187-193.
- Larsson, J.O.  
1976: Organic stream sediments in regional geochemical prospecting, Precambrian Pajala district, Sweden; *J. Geochem. Explor.*, v. 6, p. 233-249.
- Lenthall, D.H.  
1972: The application of discriminatory and cluster analysis as an aid to the understanding of the acid phase of the Bushveld Complex; *Econ. Geol. Res. Unit, University of Witwatersrand, Inform. Circ. No. 72*, 70 p.
- Lenz, C.J.  
1972: Evaluation of geochemical data; in *Application of Computer Methods in the Mineral Industry*, M.D.G. Solomon and F.N. Lancaster (ed.), S. Afr. Inst. Min. Met., Johannesburg, p. 73-74.
- Lepeltier, C.  
1969: A simplified statistical treatment of geochemical data by graphical representation; *Econ. Geol.*, v. 64, p. 538-550.  
1977: Le rôle de l'ordinateur en exploration géochimique; *Sci. Terre, Ser. Inform. No. 9*, p. 15-38.
- Lindén, A.H. and Åkerblom, G.  
1976: Method of detecting small or indistinct radioactive sources by airborne gamma-ray spectrometry; in *Geology, Mining and Extractive Processing of Uranium*, M.J. Jones (ed.), Inst. Min. Met. London, p. 113-120.
- Ling, R.F.  
1974: Comparison of several algorithms for computing sample means and variances; *J. Amer. Stat. Ass.*, v. 69, p. 859-866.
- Link, R.F. and Koch, G.S.  
1975: Some consequences of applying lognormal theory to pseudolognormal distribution; *Math. Geol.*, v. 7, p. 117-128.
- Lockhart, A.W.  
1976: Geochemical prospecting of an Appalachian porphyry copper deposit at Woodstock, New Brunswick; *J. Geochem. Explor.*, v. 6, p. 13-33.
- Lowenstein, P.L. and Howarth, R.J.  
1973: Automated colour-mapping of three-component systems and its application to regional geochemical reconnaissance; in *Geochemical Exploration 1972*, M.J. Jones (ed.), Inst. Min. Met., London, p. 297-304.
- Mancey, S.J. and Howarth, R.J.  
1978: Factor score maps of regional geochemical data from England and Wales; *Applied Geochemistry Research Group, Imperial College, London*, 2 sheets.
- Mark, D.M. and Church, M.  
1977: On the misuse of regression in earth science; *Math. Geol.*, v. 9, p. 63-75.
- Matheron, G.  
1957: Théorie lognormale de l'échantillonnage systématique des gisements; *Ann. Mines*, v. 9, p. 566-584.  
1962: *Traité de géostatistique appliquée*; Editions Technip, Paris, Vol. 1 (1962), 334 p., Vol. 2 (1963), 172 p.  
1963: Principles of geostatistics; *Econ. Geol.*, v. 58, p. 1246-1266.
- May, J.P.  
1972: PROBPLT: A computer subroutine for plotting a cumulative frequency distribution on a probability scale; *Geol. Soc. Am. Bull.*, v. 83, p. 2867-2870.
- McCammon, R.B.  
1977: BINORM — A Fortran subroutine to calculate the percentiles of a standardized binormal distribution; *Comput. Geosci.*, v. 3, p. 335-339.

- Merriam, D.F. (Ed.)  
 1976: Capture, management and display of geological data; with special emphasis on energy and mineral resources; *Comput. Geosci.*, v. 2, p. 275-376.  
 1977: Computer software for the geosciences; *Comput. Geosci.*, v. 3, p. 385-545.
- Michie, U. McL.  
 1973: The determination of sampling and analytical errors in exploration geochemistry. Discussion; *Econ. Geol.*, v. 68, p. 281-282.
- Miesch, A.T.  
 1964: Effects of sampling and analytical error in geochemical prospecting; in *Computers in the Mineral Industries*, G.A. Parks (ed.), Stanford Univ., Stanford, p. 156-170.  
 1967a: Methods of computation for estimating geochemical abundance; *U.S. Geol. Surv., Prof. Paper 574-B*, 15 p.  
 1967b: Theory of error in geochemical data; *U.S. Geol. Surv., Prof. Paper 574-A*, 17 p.  
 1969: The constant sum problem in geology; in *Computer applications in the earth sciences*, D.F. Merriam (ed.), Plenum Press, New York, p. 161-176.  
 1971: The need for unbiased and independent replicate data in geochemical exploration; *Can. Inst. Min. Met., Spec. Vol. 11*, p. 582-584.  
 1976a: Geochemical survey of Missouri: Methods of sampling, laboratory analysis and statistical reduction of data; *U.S. Geol. Surv., Prof. Paper 954-A*, 39 p.  
 1976b: Sampling designs for geochemical surveys: Syllabus for a short course; *U.S. Geol. Surv., Open File Rep. 76-772*, 140 p.  
 1977: Logtransformations in geochemistry. Discussion; *Math. Geol.*, v. 9, p. 191-194.
- Montgomery, J.H., Cochrane, D.R., and Sinclair, A.J.  
 1975: Discovery and exploitation of Ashnola porphyry copper deposit, near Keremeos, B.C.: A geochemical case study; in *Geochemical Exploration 1974*, J.L. Elliott and W.K. Fletcher (ed.), Elsevier Publ. Co., Amsterdam, p. 85-100.
- Nackowski, M.P., Mardirosian, C.A., and Botbol, J.M.  
 1967: Trend surface analysis of trace chemical data, Park City district, Utah; *Econ. Geol.*, v. 62, p. 1072-1087.
- Nairis, B.  
 1971: Endogene dispersion aureoles around the Rudtjebacken sulphide ore in the Adak area, northern Sweden; *Can. Inst. Min. Met., Spec. Vol. 11*, p. 357-374.
- Nichol, I.  
 1971: Future trends of exploration geochemistry in Canada; *Can. Inst. Min. Met., Spec. Vol. 11*, p. 32-38.
- Nichol, I., Garrett, R.G., and Webb, J.S.  
 1966: Automatic data plotting and mathematical and statistical interpretation of geochemical data; *Geol. Surv. Can., Paper 66-54*, p. 195-210.  
 1969: The role of some statistical and mathematical methods in the interpretation of regional geochemical data; *Econ. Geol.*, v. 64, p. 204-220.
- Nichol, I., Thornton, J., Webb, J.S., Fletcher, W.K., Horsnail, R.J., Khaleelee, J., and Taylor, D.  
 1970: Regional geochemical reconnaissance of the Denbighshire area; *Rep. Inst. Geol. Sci., No. 70/8*, 40 p.
- Nichol, I. and Webb, J.S.  
 1967: The application of computerized mathematical and statistical procedures to the interpretation of geochemical data; *Proc. Geol. Soc. Lond., No. 1642*, p. 186-199.
- Nichols, C.E., Butz, T.R., Cagle, G.W., and Kane, V.E.  
 1977: Uranium geochemical survey in the Crystal City and Beeville Quadrangles, Texas; Union Carbide Corp., Oak Ridge Gaseous Diffusion Plant, Oak Ridge, Tennessee, Rept. K/UR-5, GJBX-19(77), 324 p.
- Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., and Bent, D.H.  
 1975: Statistical package for the Social Sciences; McGraw-Hill, New York, 675 p.
- Norcliffe, G.B.  
 1969: On the use and limitations of trend surface models; *Canad. Geog.*, v. 13, p. 338-348.
- Obial, R.C.  
 1970: Cluster analysis as an aid in the interpretation of multi-element geochemical data; *Inst. Min. Met. Trans., London, Sect. B*, v. 79, p. 175-180.
- Obial, R.C. and James, C.H.  
 1973: Use of cluster analysis in geochemical prospecting, with particular reference to southern Derbyshire, England; in *Geochemical Exploration 1972*, M.J. Jones (ed.), Inst. Min. Met. London, p. 237-257.
- Olea, R.A.  
 1975: Optimum mapping techniques using regionalized variable theory; *Kansas Geol. Surv., Series on spatial analysis, No. 2*, Univ. Kansas, Lawrence, 137 p.
- Pantazis, Th.M. and Govett, G.J.S.  
 1973: Interpretation of a detailed rock geochemistry survey around Mathiati mine, Cyprus; *J. Geochem. Explor.*, v. 2, p. 25-36.
- Parslow, G.R.  
 1974: Determination of background and threshold in exploration geochemistry; *J. Geochem. Explor.*, v. 3, p. 319-336.
- Pearson, K.  
 1899: Data for the problem of evolution in man. III. On the magnitude of certain coefficients of correlation in man, etc.; *Proc. Roy. Soc., London*, v. 66, p. 23-32.
- Pelet, R. and De Jekhowsky, B.  
 1972: Etude statistique sur ordinateur de la géochimie de certain elements dans la formations sédimentaires du Bassin de Paris; *Proc. 24th Inter. Geol. Congr., Montreal*, v. 16, p. 60-75.
- Plant, J.  
 1971: Orientation studies on stream sediment sampling for regional geochemical survey in Scotland; *Inst. Min. Met. Trans., London, Sect. B*, v. 80, p. 324-344.  
 1973: A random numbering system for geochemical samples; *Inst. Min. Met. Trans., London, Sect. B*, v. 82, p. 64-65.

- Plant, J., Jeffery, K., Gill, E., and Fage, C.  
1975: The systematic determination of accuracy and precision in geochemical exploration data; *J. Geochem. Explor.*, v. 4, p. 467-486.
- Plant, J. and Rhind, D.  
1974: Mapping minerals; *Geogr. Mag.*, Nov. 1974, p. 123-126.
- Rhind, D.  
1975: A skeletal overview of spatial interpolation techniques; *Computer Appl.*, Nottingham, v. 2, p. 293-309.
- Rhind, D., Shaw, M.A., and Howarth, R.J.  
1973: Experimental geochemical maps; *Cartog. J.*, Dec. 1973, p. 112-118.
- Rose, A.W.  
1972: Favorability for Cornwall-type magnetite deposits in Pennsylvania using geological, geochemical, and geophysical data in a discriminant function; *J. Geochem. Explor.*, v. 1, p. 181-194.
- Rose, A.W., Dahlberg, E.C., and Keith, M.L.  
1970: A multiple regression technique for adjusting background values in stream-sediment geochemistry; *Econ. Geol.*, v. 65, p. 156-165.
- Rose, A.W. and Keith, M.L.  
1976: Reconnaissance geochemical techniques for detecting uranium deposits in sandstones of northeastern Pennsylvania; *J. Geochem. Explor.*, v. 6, p. 119-137.
- Rose, A.W. and Suhr, N.H.  
1971: Major element content as a means of allowing for background variation in stream-sediment geochemical exploration; *Can. Inst. Min. Met., Spec. Vol. 11*, p. 587-593.
- Saager, R. and Esselaar, P.A.  
1969: Factor analysis of geochemical data from the Basal Reef, Orange Free State Goldfield, South Africa; *Econ. Geol.*, v. 64, p. 445-451.
- Saager, R. and Sinclair, A.J.  
1974: Factor analysis of stream sediment geochemical data from the Mount Nansen area, Yukon Territory, Canada; *Miner. Deposita*, v. 9, p. 243-252.
- Schweitzer, R.H.  
1973: Mapping urban America with automated cartography; U.S. Dept. of Commerce, Bureau of the Census, Washington, 21 p.
- Selvin, S.  
1976: A graphical estimate of the population mean from censored normal data; *Appl. Stat.*, v. 25, p. 8-11.
- Sharp, W.E. and Jones, T.L.  
1975: A topologically optimum prospecting plan for streams; in *Geochemical Exploration 1974*, I.L. Elliott and W.K. Fletcher (ed.), Elsevier Publ. Co., Amsterdam, p. 227-235.
- Shaw, B.R.  
1977: Evaluation of distortion of residuals in trend surface analysis by clustered data; *Math. Geol.*, v. 9, p. 507-517.
- Shiikawa, M., Wakasa, K., and Tono, N.  
1975: Geochemical exploration for kuroko deposits in north-east Honshu, Japan; in *Geochemical Exploration 1974*, I.L. Elliott and W.K. Fletcher (ed.), Elsevier Publ. Co., Amsterdam, p. 65-76.
- Siegel, S.  
1956: *Nonparametric statistics for the Behavioral Sciences*; McGraw-Hill, New York, 312 p.
- Sinclair, A.J.  
1974: Selection of threshold values in geochemical data using probability graphs; *J. Geochem. Explor.*, v. 3, p. 129-149.  
1975: Some considerations regarding grid orientation and sample spacing; in *Geochemical Exploration 1974*, I.L. Elliott and W.K. Fletcher (ed.), Elsevier, Amsterdam, p. 133-140.  
1976: Applications of Probability Graphs in Mineral Exploration; *Ass. Explor. Geochem., Spec. Vol. 4*, 95 p.
- Sinding-Larsen, R.  
1975: A computer method for dividing a regional geochemical survey area into homogeneous subareas prior to statistical interpretation; in *Geochemical Exploration 1974*, I.L. Elliott and W.K. Fletcher (ed.), Elsevier Publ. Co., Amsterdam, p. 191-217.  
1977: Comments on the statistical treatment of geochemical exploration data; *Sci. Terre, Ser. Inform.*, No. 9, p. 73-90.
- Singer, D.A.  
1972: ELIPGRID, a Fortran IV program for calculating the probability of success in locating elliptical targets with square, rectangular and hexagonal grids; *Geocom Programs No. 4*, Geosystems, London, 16 p.
- Slysz, W.D.  
1973: Software Compatibility in the Social Sciences; in *Proc. Educom. Fall Conf.*, Princeton, p. 235-243.
- Smeds, B.  
1976: Ink jet colour plotter for computer graphics; *Rep. Lund Inst. Techn., Dep. Elect. Measurements*, No. 1/1976.
- Smith, F.M.  
1971: Geochemical exploration over complex mountain glacial terrain in the Whitehorse Copperbelt, Yukon Territory; *Can. Inst. Min. Met., Spec. Vol. 11*, p. 265-275.
- Smith, E.C. and Webber, G.R.  
1973: Nature of mercury anomalies at the New Calumet Mines area, Quebec, Canada; in *Geochemical Exploration 1972*, M.J. Jones (ed.), Inst. Min. Met. London, p. 71-80.
- Spearman, C.  
1904: The proof and measurement of association between two things; *Amer. J. Psychol.*, v. 15, p. 72-101.
- Stevens, M.A.  
1974: EDF statistics for goodness of fit and some comparisons; *J. Amer. Stat. Ass.*, v. 69, p. 730-737.
- Summerhayes, C.P.  
1972: Geochemistry of continental margin sediments from northwest Africa; *Chem. Geol.*, v. 10, p. 137-156.
- Teil, H.  
1975: Correspondence factor analysis: an outline of its method; *Math. Geol.*, v. 7, p. 3-12.

- Thompson, M. and Howarth, R.J.  
1976: Duplicate analysis in geochemical practice. Part 1. Theoretical approach and estimation of analytical reproducibility; *Analyst*, v. 101,  
1978: A new approach to the estimation of analytical precision; *J. Geochem. Explor.*, v. 9, p. 23-30.
- Timperley, M.H. et al.  
1972: The improved detection of geochemical soil anomalies by multiple regression analysis of biogeochemical data; *Proc. Austral. Inst. Min. Met.*, v. 242, p. 23-36.
- Tourtelot, H.A. and Miesch, A.T.  
1975: Sampling designs in environmental geochemistry; *Geol. Soc. Amer., Spec. Paper 155*, p. 107-118.
- Trochimczyk, J. and Chayes, F.  
1977: Sampling variation of principal components; *Math. Geol.*, v. 9, p. 497-506.  
1978: Some properties of principal components; *Math. Geol.*, v. 10, p. 43-52.
- U.S. Geological Survey  
1969: Proceedings of the Symposium on map and chart digitizing; *U.S. Geol. Surv., Comput. Contr. No. 5*, 81 p.
- Van Trump, G. and Miesch, A.T.  
1977: The U.S. Geological Survey RASS-STATPAC system for management and statistical reduction of geochemical data; *Comput. Geosci.*, v. 3, p. 475-488.
- Walden, A.R.  
1972: Quantitative comparison of automatic contouring algorithms; *Kansas Oil Exploration (KOX) Project Rept.*, Kansas Geol. Surv., 137 p.
- Webb, J.S., Fortescue, J., Nichol, I., and Tooms, J.S.  
1964: Regional geochemical maps of the Namwala Concession area, Zambia; *Geol. Surv., Maps 1-1X*.
- Webb, J.S. and Howarth, R.J.  
Regional geochemical mapping; in *Proceedings of the Discussion Meeting on Geochemistry and Health*, The Royal Society, London. (in press)
- Webb, J.S., Nichol, I., Foster, R., Lowenstein, P.L., and Howarth, R.J.  
1973: Provisional Geochemical Atlas of Northern Ireland; Applied Geochemistry Research Group, Imperial College, London, 36 p.
- Webb, J.S., Thornton, I., Thompson, M., Howarth, R.J., and Lowenstein, P.L.  
1978: The Wolfson Geochemical Atlas of England and Wales; Oxford University Press, Oxford, 74 p.
- Wennervirta, H., Bolviken, B., and Nilsson, C.A.  
1971: Summary of research and development in Scandinavian countries; *Can. Inst. Min. Met., Spec. Vol. 11*, p. 11-14.
- Wennervirta, H. and Papunen, H.  
1974: Heavy metals as lithogeochemical indicators for ore deposits in the Iilinjärvi and Aijala fields, SW-Finland; *Bull. Geol. Surv. Finl.*, No. 269, 22 p.
- Whitehead, R.E.S. and Govett, G.J.S.  
1974: Exploration rock geochemistry detection of trace element haloes at Heath Steel Mines (N.B., Canada) by discriminant analysis; *J. Geochem. Explor.*, v. 3, p. 371-386.
- Wignall, T.K.  
1969: Generalized Bayesian classification: K classes; *Econ. Geol.*, v. 64, p. 571-574.  
1970: Generalized Bayesian classification: K classes; Reply to discussion; *Econ. Geol.*, v. 65, p. 221-222.
- Wolf, W.J.  
1974: Geochemical distribution of zinc, nickel and copper in volcanic rocks, Ben Nevis Township and parts of Clifford and Pontiac Townships, District of Timiskaming; Ontario Div. Mines, Prelim. Geochem. Maps P.915, P.916 and P.917.  
1975: Zinc abundance in early Precambrian volcanic rocks: its relationship to exploitable levels of zinc in sulphide deposits of volcanic-exhalative origin; in *Geochemical Exploration 1974*, I.L. Elliott and W.K. Fletcher (ed.), Elsevier Publ. Co., Amsterdam, p. 262-278.

## NOTE:

Enquiries for all reports with GJBX numbers should be addressed to:

Bendex Field Engineering Corp.,  
Technical Library,  
P.O. Box 1569,  
Grand Junction,  
Colorado, U.S.A.  
C081501