CRCLEME
Cooperative Research Centre for
Landscape Evolution & Mineral Exploration

OPEN FILE
REPORT
SERIES

CSIRO
EXPLORATION
AND MINING

AMIRA
Australian Mineral Industries Research Association Limited ACN 004 448 266

STRATEGIES AND METHODS FOR THE INTERPRETATION OF GEOCHEMICAL DATA

E.C. Grunsky

CRC LEME OPEN FILE REPORT 59

December 1998

(CSIRO Division of Exploration Geoscience, Unnumbered 1991.
Second impression 1998.

CRC LEME is an unincorporated joint venture between The Australian National University, University of Canberra, Australian Geological Survey Organisation and CSIRO Exploration and Mining, established and supported under the Australian Government's Cooperative Research Centres Program.

AMIRA
Australian Mineral Industries Research Association Limited ACN 004 448 266

CRCLEME
Cooperative Research Centre for
Landscape Evolution & Mineral Exploration

CSIRO
EXPLORATION
AND MINING

STRATEGIES AND METHODS FOR THE INTERPRETATION OF GEOCHEMICAL DATA

E. C. Grunsky

CRC LEME OPEN FILE REPORT 59

December 1998

(CSIRO Division of Exploration Geoscience, Unnumbered 1991.
Second impression, 1998

CRC LEME is an unincorporated joint venture between The Australian National University,
University of Canberra, Australian Geological Survey Organisation and CSIRO Exploration and Mining
Headquarters: CRC LEME c/o CSIRO Exploration and Mining, Private Bag, PO Wembley, Western Australia, 6014

RESEARCH ARISING FROM CSIRO/AMIRA REGOLITH GEOCHEMISTRY PROJECTS 1987-1993

In 1987, CSIRO commenced a series of multi-client research projects in regolith geology and geochemistry which were sponsored by companies in the Australian mining industry, through the Australian Mineral Industries Research Association Limited (AMIRA). The initial research program, "Exploration for concealed gold deposits, Yilgarn Block, Western Australia" (1987-1993) had the aim of developing improved geological, geochemical and geophysical methods for mineral exploration that would facilitate the location of blind, buried or deeply weathered gold deposits. The program included the following projects:

P240: Laterite geochemistry for detecting concealed mineral deposits (1987-1991). Leader: Dr R.E. Smith.
Its scope was development of methods for sampling and interpretation of multi-element laterite geochemistry data and application of multi-element techniques to gold and polymetallic mineral exploration in weathered terrain. The project emphasised viewing laterite geochemical dispersion patterns in their regolith-landform context at local and district scales. It was supported by 30 companies.

P241: Gold and associated elements in the regolith - dispersion processes and implications for exploration (1987-1991). Leader: Dr C.R.M. Butt.
The project investigated the distribution of ore and indicator elements in the regolith. It included studies of the mineralogical and geochemical characteristics of weathered ore deposits and wall rocks, and the chemical controls on element dispersion and concentration during regolith evolution. This was to increase the effectiveness of geochemical exploration in weathered terrain through improved understanding of weathering processes. It was supported by 26 companies.

These projects represented "an opportunity for the mineral industry to participate in a multi-disciplinary program of geoscience research aimed at developing new geological, geochemical and geophysical methods for exploration in deeply weathered Archaean terrains". This initiative recognised the unique opportunities, created by exploration and open-cut mining, to conduct detailed studies of the weathered zone, with particular emphasis on the near-surface expression of gold mineralisation. The skills of existing and specially recruited research staff from the Floreat Park and North Ryde laboratories (of the then Divisions of Minerals and Geochemistry, and Mineral Physics and Mineralogy, subsequently Exploration Geoscience and later Exploration and Mining) were integrated to form a task force with expertise in geology, mineralogy, geochemistry and geophysics. Several staff participated in more than one project. Following completion of the original projects, two continuation projects were developed.

P240A: Geochemical exploration in complex lateritic environments of the Yilgarn Craton, Western Australia (1991-1993). Leaders: Drs R.E. Smith and R.R. Anand.
The approach of viewing geochemical dispersion within a well-controlled and well-understood regolith-landform and bedrock framework at detailed and district scales continued. In this extension, focus was particularly on areas of transported cover and on more complex lateritic environments typified by the Kalgoorlie regional study. This was supported by 17 companies.

P241A: Gold and associated elements in the regolith - dispersion processes and implications for exploration. Leader: Dr. C.R.M. Butt.
The significance of gold mobilisation under present-day conditions, particularly the important relationship with pedogenic carbonate, was investigated further. In addition, attention was focussed on the recognition of primary lithologies from their weathered equivalents. This project was supported by 14 companies.

Although the confidentiality periods of the research reports have expired, the last in December 1994, they have not been made public until now. Publishing the reports through the CRC LEME Report Series is seen as an appropriate means of doing this. By making available the results of the research and the authors' interpretations, it is hoped that the reports will provide source data for future research and be useful for teaching. CRC LEME acknowledges the Australian Mineral Industries Research Association and CSIRO Division of Exploration and Mining for authorisation to publish these reports. It is intended that publication of the reports will be a substantial additional factor in transferring technology to aid the Australian Mineral Industry.

This report (CRC LEME Open File Report 59) is a Second impression (second printing) of CSIRO, Division of Exploration Geoscience Restricted Report Unnumbered, first issued in 1991, which formed part of the CSIRO/AMIRA Project P240.

Copies of this publication can be obtained from:
The Publication Officer, c/- CRC LEME, CSIRO Exploration and Mining, PMB, Wembley, WA 6014, Australia. Information on other publications in this series may be obtained from the above or from http://leme.anu.edu.au/

Cataloguing-in-Publication:
Grunsky, E.C.
Strategies and methods for the interpretation of geochemical data. Discussion paper applied to laterite geochemistry
ISBN 0 642 28258
1.Geochemistry - methodology 2. Laterite - Western Australia
I. Title
CRC LEME Open File Report 59.
ISSN 1329-4768

Strategies and Methods for the Interpretation of Geochemical Data

Discussion Paper Applied to Laterite Geochemistry

Eric Grunsky

Present address: Geological Survey Branch, Ministry of Energy Mines and Petroleum Resources, 553 Superior St, Victoria, BC V8V 1X4, Canada

CSIRO/AMIRA Laterite Geochemistry Project P240

CSIRO, Division of Exploration Geoscience, Wembley, WA 6014, Australia

27 August, 1991

This discussion paper has been produced, for sponsors of the CSIRO/AMIRA Laterite Geochemistry Project P240, to facilitate an understanding of some of the important methods for interpretation of multi-element exploration geochemical data.

ABSTRACT

This paper summarizes a systematic approach to analyzing the geochemical data collected over lateritic terrains in the Yilgarn Block of Western Australia.

Critical steps in the sequence of data analysis are:

1) preliminary data analysis,
2) exploratory multivariate data analysis, and
3) specific multivariate data analysis and modelled multivariate analysis.

For the evaluation of geochemical data in laterites of the Yilgarn Block the following sequence of investigation is recommended:

1) Preliminary Data Analysis

- The use of histograms, box & whisker plots, Q-Q plots, scatter plot matrix, data ranking;
- Preparation of summary statistical tables;
- Maps of elements with each sample ranked into percentile ranges;
- Elimination of gross outliers;
- Investigate outliers for each element: analytical error or atypical abundance;
- Adjust data for censored values;
- Transformation of data based upon samples below the 95th-98th percentile;
- Scatterplot matrix for transformed data;
- Threshold selection after transformation.

2) Exploratory Multivariate Data Analysis

- Robust estimates to compute means and covariances to enhance the detection of outliers;
- Application of dimension reducing techniques such as principal components analysis;
- The use of methods to delineate structure in the data (cluster analysis, multidimensional scaling, non-linear mapping, and projection pursuit);
- The use of χ^2 plots applied to transformed data to isolate outliers based upon all of the elements of interest; maps of large Mahalanobis distances (>95th percentile) may identify anomalous areas.

3) Specific Multivariate Data Analysis and Modelled Multivariate Analysis

- Calculation of empirical indices specifically tailored to areas in which multi-element associations are understood.
- Multiple regression applied to areas where a linear model of the multi-element association can be computed with good results (i.e. high R^2 coefficients). Residuals can be examined for the potential of being associated with mineral deposits.
- The establishment of background and target groups that characterized the geochemical variation of the regional geochemistry and the mineral deposits.
- Analysis of variance and canonical variate analysis to test the statistical uniqueness of the groups.
- The use of all possible subsets to compare reference groups with each other and determine which group of elements enhance the group separations.
- The application of allocation/typicality procedures to test unknown samples from a regional exploration programme. Each samples is assigned a probability of belonging to one of the reference groups. Maps of typicality or posterior probability can be made to indicate group membership.

This approach of systematically analyzing the laterite geochemistry forms the basis of an effective exploration programme strategy in geochemistry.

TABLE OF CONTENTS

1.0 STATISTICAL DESCRIPTION OF DATA

1.1 Introduction

The objective of these notes is to review strategies and methodologies that can be used for the evaluation and interpretation of geochemical data. The general outline of these notes is subdivided by two main strategies.

Exploratory Data Analysis of Geochemical Data (Sections 3 and 4) is concerned with analysing geochemical data for the purpose of recognizing and enhancing trends or structures in the data that are not immediately obvious. These trends/structures/patterns provide insight into the geochemical/geological processes that have occurred. There are a great many methods and procedures that can be carried out on geochemical data which enable the investigator to discover patterns in the data.

Modelled Investigations of Geochemical Data (Section 5) are based upon the knowledge that certain geochemical patterns reflect particular geological processes (i.e. ore deposits) as recognized through exploratory data analysis. Through orientation studies, geochemical characteristics can be obtained for specific geological environments which through interpretation can be used as models. This information can lead to the definition of background and target groups, which can then be used as the models with which unknown samples can be compared using a variety of statistical methods. *Background groups* are composed of data that represent regional background variation. *Target groups* are composed of data that characterize the geochemistry around selected mineral deposits.

Both strategies can be important for a successful exploration programme and they are commonly carried out in parallel. However, the interpretation of the results of these strategies must be based upon a thorough understanding of the geology of the region(s) being investigated.

The use of graphical procedures is perhaps one of the most important aspects of geochemical data interpretation. Many of the numerical procedures are difficult to interpret from the examination of the numerical results alone. However, from the results of the numerical methods, if the interrelationships between the samples and elements can be shown graphically, the nature of the relationships can be easily observed.

The ultimate goal of the investigation of geochemical data is the detection of a spatially-continuous zone which has elevated abundances of strategic elements and which may host mineral deposits. A zone of elevated abundances is usually referred to as an anomaly. However, the definition of an anomaly is one of the most contested definitions in current exploration geochemistry concepts. An important assumption is that the background and target populations are adequately represented. In regional sampling programmes, this may not always be the case. Sampling strategies are crucial and the design of sampling strategies should be carefully considered prior to collecting samples (Garrett, 1983, Chapter 4). Data interpretation can be enhanced significantly if the design of the sampling strategy ensures an adequate representation of geochemical features.

Gold, massive sulphide deposits, and other rare metal deposits often are characterized by a halo of elevated abundances of other elements that are normally present in very low abundances in the country rocks. These elements are called *pathfinder elements* as their aerial extent is larger than the extent of the target element and thus provide a diagnostic key to the presence of the target being sought. This has led to the use of multi-element geochemistry being an important aspect of exploration geochemistry programmes. Multi-element geochemical distributions can yield useful results for mineral exploration purposes. Many elements have associations (correlations) with other elements which indicate that the elements are not independent from each other. Thus, inferences between the relationships of elements provide patterns that can be recognized with geological processes. These relationships necessitate the use of multivariate methods of analysis and statistics which assist in the recognition of mineral deposits.

These notes do not cover statistical theory, rather they emphasize methods that can be applied to multi-element geochemical data. The notes assume that the reader has a basic knowledge of statistics. Davis (1986, Chapter 2) gives an outline of basic statistics and discusses probability.

1.2 Data Sets Used In This Study

The laterite geochemical data used in these notes will provide contrasting examples between regional background geochemical data (background groups) and geochemical data associated with precious metal and massive sulphide base metal deposits (target groups). The study will focus on methods of exploring these datasets and contrasting the geochemistry of the datasets in order to examine the differences between background and target groups.

Dataset 1: Murchison Greenstone Belt: Regional Background Data

This dataset, part of the CSIRO/AGE Yilgarn geochemical database, represents geochemical data collected over the granite-greenstone terrain of the Murchison area in the Yilgarn Block of Western Australia. The area is underlain by supracrustal rocks, "greenstones", and surrounding felsic intrusive stocks and batholiths with enclaves of gneissic material, all of Archaean age. The supracrustal areas are comprised of predominantly mafic volcanic, felsic volcanic, and sedimentary sequences that have been deformed, metamorphosed, and intruded by

later post-kinematic plutons. Late stage faulting has occurred throughout the preserved greenstone belts. Many Au deposits have a proximity to these large scale fault systems.

As part of the AGE programme, 1983 to 1986, the region was sampled at 3-km triangular spacing intervals with follow-up sampling at closer spacings of 1 km and 300 m. A variety of surface exposed lateritic materials was sampled and classified as to the type of sample. Most samples belong to one of two groups consisting of either lateritic residuum, which dominates, or subordinate lag of ferruginous pebbles or cobbles. The most common form of the laterites is nodular or pisolitic. Only the lateritic materials are used in the examples presented here. Reports on the laterite and other Fe-rich materials that were sampled over the area are covered by Grunsky *et al.* (1988, 1989).

Gold mineralization is widespread throughout the area. Most deposits are viewed as epigenetic and many shears and mineralized vein systems are associated with carbonate, soda, and potash metasomatism.

Dataset 2: Yilgarn Block Granite/Gneiss Terrain Background Group from the Albany-Fraser region (Grunsky, 1991).
The region was sampled at 3-km triangular spacing intervals with follow-up sampling at closer spacings of 1 km and 300 m as part of the AGE programme. As in Dataset 1, a variety of surface-exposed lateritic materials were sampled and classified as to the type of sample. Almost all of the samples are lateritic, the most common form being nodular or pisolitic. The area is composed of large tracts of compositionally-banded gneiss and foliated and massive granitoid rocks ranging in composition from tonalite to granite. Enclaves of greenstone material occur throughout the area and are of middle to upper amphibolite facies grade metamorhpism.

Dataset 3: Mt. Gibson Au Deposit Target Group
The Mt. Gibson Au deposit occurs within a sequence of altered and sheared felsic and mafic volcanics of amphibolite grade metamorphism. The sequence occurs as a synformal enclave within gneissic Archaean granitoids. The volcanics are primarily tholeiitic basalts overlain by volcaniclastic sediments, felsic volcanics, and some banded iron formation. It occurs within the Murchison greenstone terrain of the regional background group of Dataset 1. The Au mineralization is situated in auriferous veins within a predominantly shear controlled vein network. It is typical of many Archaean lode Au deposits (Groves, 1988; Colvine *et al.*, 1988). Orientation sampling was carried out along grid lines and at specific sites of interest. The sampling density is approximately 18 samples per square kilometre. Several sample types were collected over the area (Smith *et al.*, in prep.), only lateritic duricrust samples are used in this paper. The geochemistry and geomorphology of the area is discussed in Anand *et al.* (1989a).

Dataset 4: Golden Grove Cu-Zn-Au deposit (Murchison belt)
The Golden Grove deposit is most similar to a massive sulphide type of environment. The geology and exploration geochemistry is discussed in Smith and Perdrix (1983) and Smith *et al.* (1984). Copper and Zn ore deposits occur at both the Gossan Hill and Scuddles locations within a sequence of stratiform metavolcanics with Au being enriched within the overlying laterite. The ore deposits occur within a sequence of intercalated acid pyroclastic, volcaniclastic, and associated felsic flows. This target group is also located within the Murchison greenstone terrain of Dataset 1. The mineralization is stratiform, hosted by felsic lapilli tuff. The deposits contain Cu, Pb, Au, Zn, and Ag. The samples that comprise this dataset consist of 100 samples of individual loose lateritized gossan-nodules (which will be a new category, LT164, in the Atlas (Anand *et al.*, 1989b)) collected from Gossan Hill. An initial study of the area indicated that associated pathfinders associated with Cu, Au, and Ag are Bi, Sn, Mo, In, Sb, and As. The samples represent a specialized Golden Grove geochemical signature which will be compared with the Mt. Gibson and Lawlers samples.

Dataset 5: Lawlers Au Deposit Target Group
The laterite samples collected for this dataset cover a range of bedrock lithologies. This Au deposit is situated within a sequence of mafic/ultramafic volcanic rocks. The mineralization is not well understood. Only the lateritic samples were used in this dataset. Sampling of the laterite was carried out over a small area directly over the mineralization.

1.3 Special Problems

There are some fundamental problems that commonly occur in geochemical data.

1) Most elements have a "censored" distribution, meaning that values at less than the detection limit can only be reported as being less than that limit.

2) The data do not occur as normally-distributed abundances.

3) The data have missing values. That is, not every sample has been analysed for the same number of elements.

4) Not every element has been analysed by the same method or the limits of detection of the method have changed over time.

These problems create difficulties when applying mathematical or statistical procedures to the data. Statistical procedures have been devised to deal with all except the last problem. To overcome the problems of censored distributions, procedures have been developed by applying transformations to estimate replacement values for the purposes of statistical calculations by the CSIRO Division of Mathematics and Statistics. Non-normally distributed data can be transformed using procedures discussed above. When the data have missing values, several procedures can be applied to estimate replacement values. Most procedures use a multiple regression procedure which estimates the replacement value based on a regression with samples that have complete analyses. The use of replacement values has obvious limitations and precautions must be kept in mind.

1.4 Frequency Distributions

1.4.1 SAMPLE POPULATIONS AND DISTRIBUTIONS

The chemical analyses of samples collected in a sampling programme generally show a range of abundances for each of the elements that are analysed. If the values for a given element of the sample population are ordered, then the "distribution" of the sample population can be observed. The distribution is commonly expressed in the form of frequency of value versus data value and can be expressed graphically by the histogram. Other graphical methods of distributions are box and whisker plots and quantile-quantile plots. Figures 1a,b graphically illustrate two types of frequency distributions. Figure 1a describes a symmetric distribution while Figure 1b displays a non-symmetric distribution. Each distribution has distinctively different properties. The frequency of various elements can have frequency characteristics that closely match those shown in Figures 1a or 1b.

The analysis of one element of a population of samples is known as **univariate** analysis. When several elements are being analyzed, the population of samples is **multivariate.** Figures 1a,b display characteristics associated with univariate populations. Multivariate populations are impossible to display beyond three dimensions. The statistics of multivariate populations are more complicated and less well understood. Figure 1c shows a graphical example of bivariate distribution which shows the frequency relationship between two variables.

1.4.2 ESTIMATES OF THE MEAN/MEDIAN

Estimates of the central values of a distribution are important for data analysis and statistical techniques that make use of correlation or covariance estimates. Techniques such as principal components analysis, regression procedures, and χ^2 plots are based upon the estimate of the mean. If the data are not normally distributed nor skewed, then the arithmetic mean can be significantly different from the median value. Figure 1b illustrates this difference. There are three measures of central tendency, the mode, median, and the mean. When a population is normally distributed, then the three measures are equal in value.

The **mean** is the maximum frequency point in a normal distribution (see Figure 1a). In a normal distribution for a random variable x and n samples in the population, the **mean** is defined as:

$$\bar{x} = \{\sum_{i=1}^{n} x_i\}/n$$

The **mode** is defined as the maximum frequency point of any distribution. In a non-normal distribution, there can be many modes, hence the term **polymodal** distribution. In a normal distribution, the mode and the mean are the same. The **median** is the midpoint in the ordered set of frequencies of any distribution. In a normal distribution, the median and the mean are the same. These three terms are illustrated in Figures 1a and 1b.

In statistical analysis, the use of the mean is the most commonly-used measure of central tendency. If the mean does not represent the maximum frequency of the distribution as illustrated in Figure 1a, then the results of a statistical analysis can be misleading. In the case of Figure 1b, it can be seen that the mean value is greater than the mode or the median value. Most geochemical distributions have characteristics similar to that of Figure 1b. In order to treat geochemical data statistically, there are a number of procedures that can be applied to transform the distribution to more normal characteristics as in Figure 1a. These techniques will be discussed below.

1.4.3 HISTOGRAMS

The histogram is probably the best known graphical means of displaying a distribution since it reflects the shape similar to theoretical frequency distributions. The example presented here represents As samples collected over the Murchison area greenstone terrain.

The histogram is constructed in the following way:

i) The interval of sample values is divided into a fixed number of classes. Figure 2 shows a histogram where the interval has been divided into 40 equal segments. The choice of interval is an arbitrary and trial and error procedure. If the number is intervals is too small, then the finer details of the distribution are smoothed over.

If the number of intervals are too many, then the distribution appears discontinuous. This is illustrated in Figures 2, 3, and 4 where the number of intervals are 40, 20, and 80 respectively. The shape of the histogram is sensitive to the data interval used to construct the frequency bars. If data intervals are poorly chosen then interpretation of the histogram can be misleading.

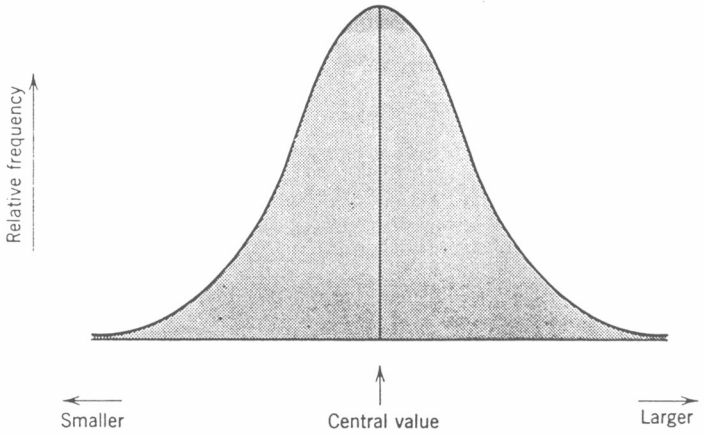


Figure 1a. Normal frequency distribution

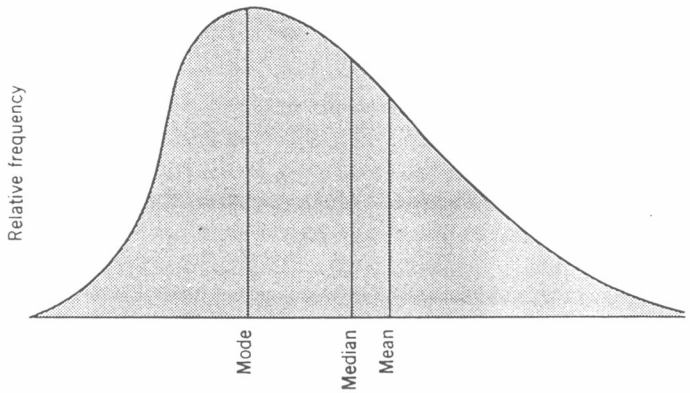


Figure 1b. Skewed frequency distribution

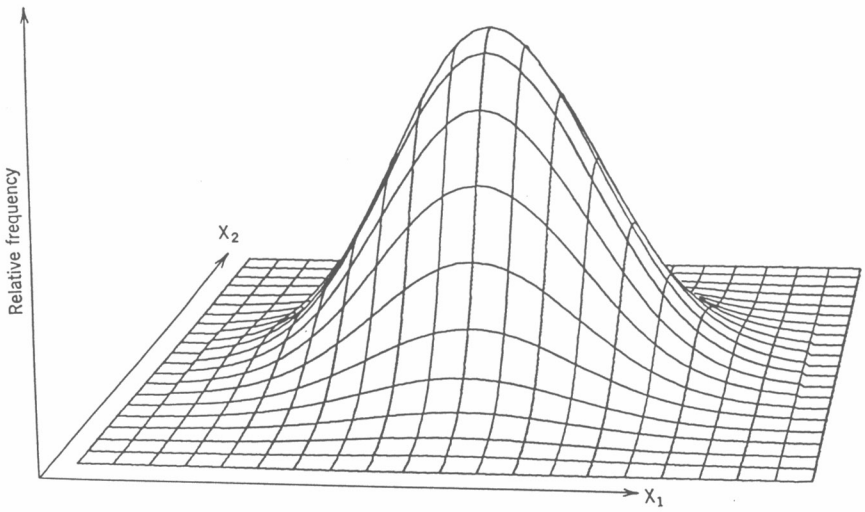


Figure 1c. Joint probability distribution

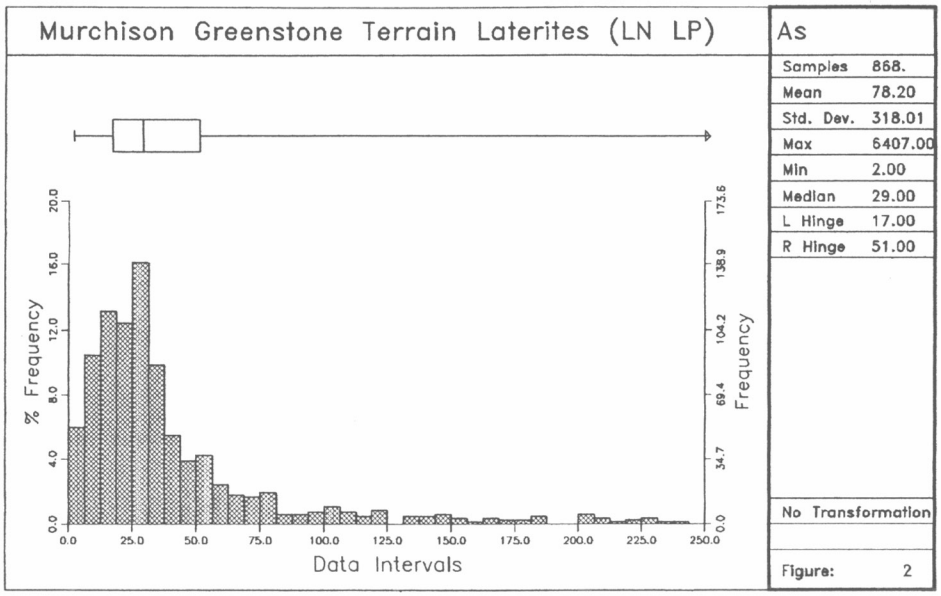


Figure 2. Histogram & Box-Whisker Plot: Arsenic in Laterites, Murchison Greenstone Terrain. Number of frequency intervals = 40.

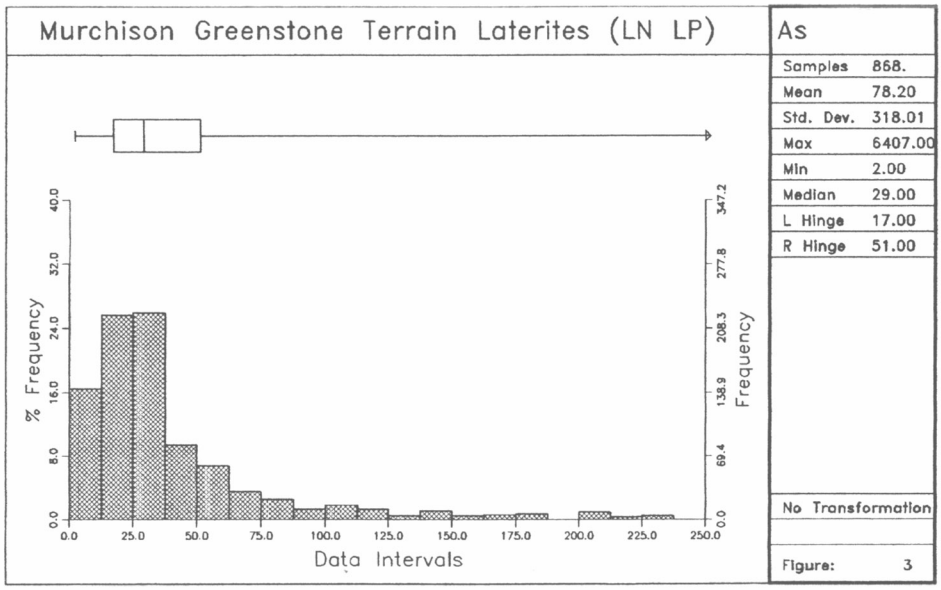


Figure 3. Histogram & Box-Whisker Plot: Arsenic in Laterites, Murchison Greenstone Terrain. Number of frequency intervals = 20.

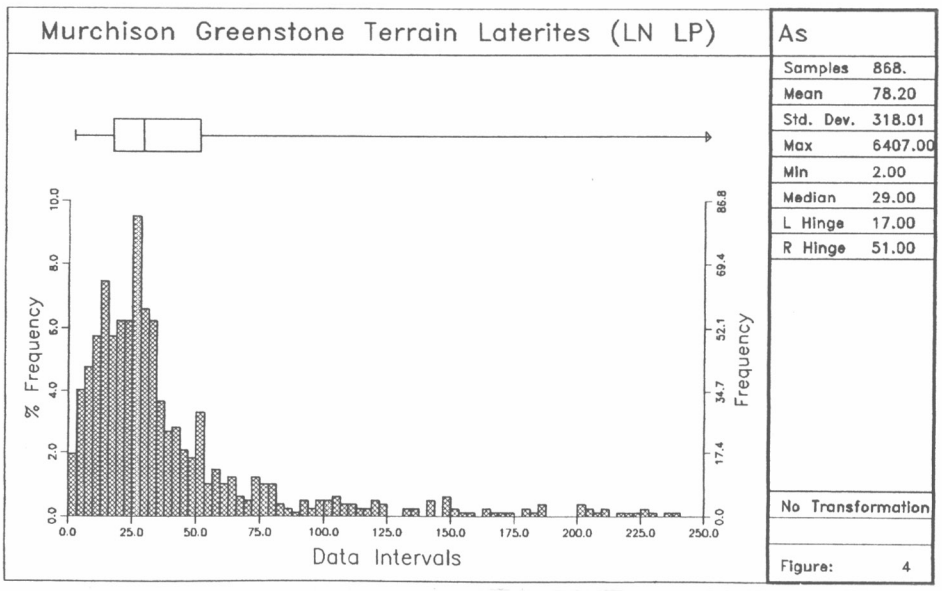


Figure 4. Histogram & Box-Whisker Plot: Arsenic in Laterites, Murchison Greenstone Terrain. Number of frequency intervals = 80.

ii) A cut-off value should be implemented in order to avoid distorting the histogram if extreme outliers are present. The graphical presentation of the distribution can be enhanced by choosing a suitable interval for the range of the data; however, this will hide the extreme values. Thus, histograms should be drawn for unscaled and scaled data. Figure 5a shows the same data as Figure 2; however, in this case the data are unscaled and it can be seen that the nature of the data distribution is masked by the extreme values that distort the scale of the histogram.

iii) Summary statistics alongside of the histogram can assist in viewing the range and spread of the data. In Figures 2-5, the number of samples, mean, standard deviation, maximum value, minimum value, median, left hinge, and right hinge are shown. These values provide a numerical basis for interpreting the distribution.

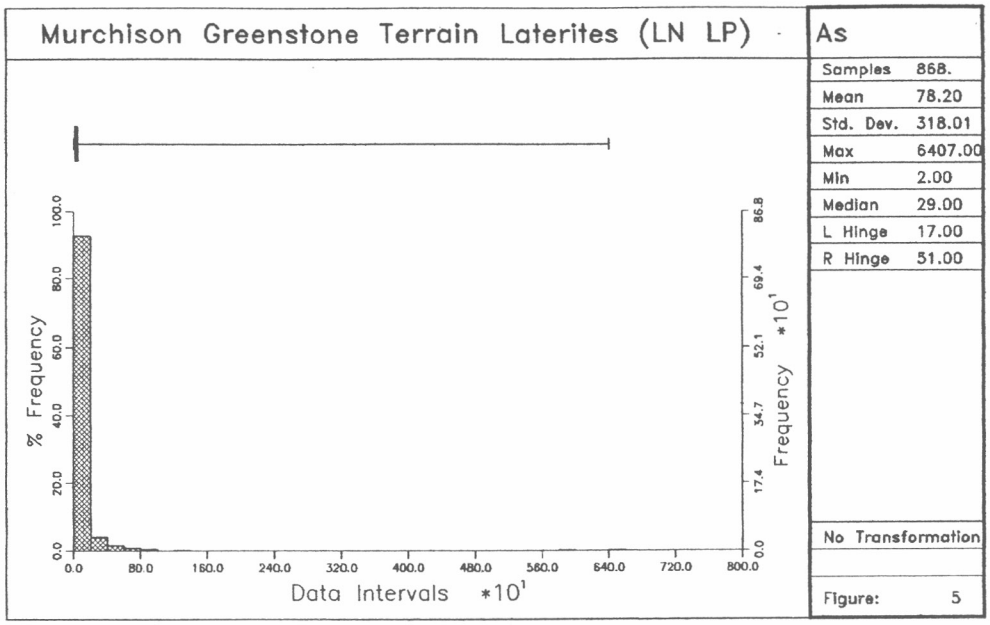


Figure 5a. Histogram & Box-Whisker Plot: Arsenic in Laterites,Murchison Greenstone Terrain. Number of frequency intervals = 40. The data is unscaled, all outliers are included.

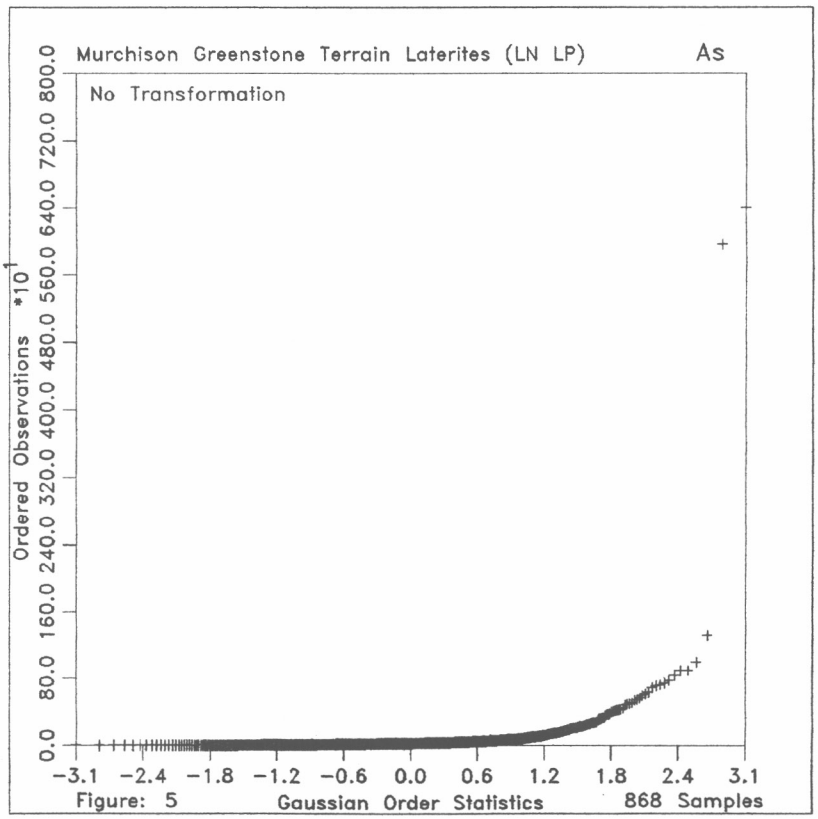


Figure 5b. Q-Q Plot: Arsenic in Laterites, Murchison Greenstone Terrain. All outliers are included.

1.4.4 BOX AND WHISKER PLOTS

The box and whisker plot was developed to express order statistics in a graphical form (Tukey, 1977). In the analysis of a univariate distribution, it is often convenient to present the box and whisker plot along with the histogram as is done in Figures 2 to 5. The box and whisker plot provides a graphical display of the median (50th percentile), left and right hinges (25th and 75th percentile) as well as the maximum and minimum values over the range of the data. The left and right hinges and median are presented as a "box" that displays the range over which 50% of the data is spread. The "whiskers" are the lines that extend to the left and right of the box. Several variations exist on the graphical presentation of box and whisker plots. Some plots display the values that are less than the 25th and greater than the 75th percentiles as points along the whisker. The extreme ends (maximum and minimum values) of the data are marked by vertical bars at the end of the whiskers. Alternatively, the whiskers can extend to the "fences", which are defined as 1.5*midrange of the data. Samples that plot beyond 3*midrange are plotted as special symbols. Samples that plot beyond 3*midrange can also be shown by separate symbols indicating their uniqueness. These samples have not been plotted on the box and whisker plots shown in Figures 2-9.

The location of the median line within the box gives an indication of how symmetric the distribution is within the range of the left to right hinges (midrange). The length of the whiskers on each side of the box also provides an estimate of the symmetry of the distribution. Figure 2 shows the asymmetric character of the As distribution for the Yilgarn greenstone terrain laterites. In this case, the median value occurs towards the left hinge line and the extreme values plot beyond the range of the axis.

The main advantage of the box and whisker plot is that its shape is independent of the interval used to present the histogram. Thus, providing the scale of presentation is reasonable, the box and whisker plot provides a fast visual estimate of the frequency distribution.

1.4.5 QUANTILE-QUANTILE (Q-Q) PLOTS

Quantile-Quantile (Q-Q) plots are a graphical means of comparing a frequency distribution with respect to an expected frequency distribution which is usually the normal distribution. Q-Q plots are generated by calculating quantile values for the normal frequency distribution (value of the normal frequency distribution over the range of probability, 0.0 to 1.0) and then plotting them against the observed ordered data. If a frequency distribution is normally distributed, when the quantile values are plotted against the ordered values of the sample population the plot will be a straight line. If the frequency distribution of the sample population is skewed or the population is polymodal, then the Q-Q plot will be curved or discontinuous.

Figure 5b shows a Q-Q plot for an asymmetric (non-normal) distribution. The data is the same data used to generate Figures 2-4. The histogram and box and whisker plot of Figure 4 show that the distribution is asymmetric with many samples occurring to the right of the median value. This asymmetry is also expressed in the Q-Q plot where the distribution of the data is highly distorted by a few outliers.

The main advantage of the Q-Q plot is that each individual sample is plotted and thus the detailed characteristics of groups of samples that are of interest can be observed. Another means of analysing frequency distributions is through the use of probability plots. Sinclair (1976) describes the use and applications of probability plots in the analysis of geochemical data. Probability plots are essentially the same as Quantile-Quantile plots.

1.5 Transformation of Data

In many distributions, the skewed nature of the data can be overcome by applying a suitable transformation that shifts the values of the distribution such that it becomes normally distributed. It has been common in the geological literature to apply logarithmic transformations to data as a way to correct positive skewness (long tail to the right of the median). However, the application of a logarithmic transform may not always be the best transform to apply and in some cases may even distort the distribution parameters even more so than the untransformed data (Link and Koch, 1975).

Transformations that can be applied are:

- Linear scaling, $y=kx$ or $y=x/k$
- Standardization, $y=(x_i - x_{mean})/s$
- Logarithmic, $y=\log_{10}(x)$, $y=\ln(a+x)$
- Exponential, $y=e^x$
- Box-Cox Generalized Power Transform, $y=(x^\lambda - 1)/\lambda$, $y=\ln(x)$ for $\lambda=0$

The linear scaling and standardization transformations do not change the shape of the distribution; however, the degree of dispersion can change. The logarithmic, exponential, and Box-Cox generalized power transforms modify both the shape and the dispersion characteristics of the distributions and are the transformations most commonly used. Howarth and Earle (1979) provide a computer program for estimating parameters for the generalized Box-Cox power transform based on the optimization of skewness and kurtosis and the optimization of the maximum likelihood criterion of Box and Cox (1964). Lindqvist (1976) has also published a computer program (SELLO) for transforming skewed distributions based on minimizing skewness.

Figure 6a shows a histogram of the log transformed data of Figure 2 after the extreme outliers have been eliminated. Figure 6a shows that the median line is more centrally located within the box and whisker plot with a value close to 3.0 (ln(20 ppm)=3.0). Also the whiskers are of more equal length indicating a more symmetric distribution. Figure 6b shows the same data on a Q-Q plot. The data clearly approximate a straight line and thus the transformation can be seen to result in a dataset that is closer to a normal distribution.

After applying the appropriate transformation, Q-Q plots can also be used for detecting multiple populations within data. Examination of Figure 6b indicates that there are changes in the slope of the data near the log values of 1.8, 3.5, and 5.6 (these translate to 6, 33, and 270 ppm respectively) for As. Such a change of slope indicates that there is a significant change in the frequency of observations at these abundances. The

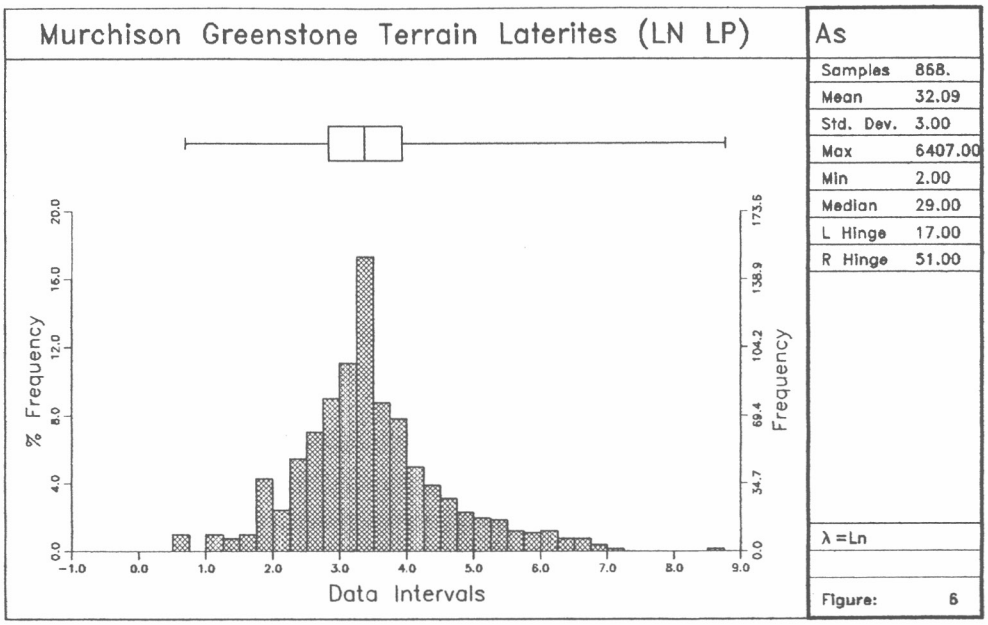


Figure 6a. Histogram & Box-Whisker Plot: Arsenic in Laterites [ln Transform], Murchison Greenstone Terrain. Number of frequency intervals = 40.

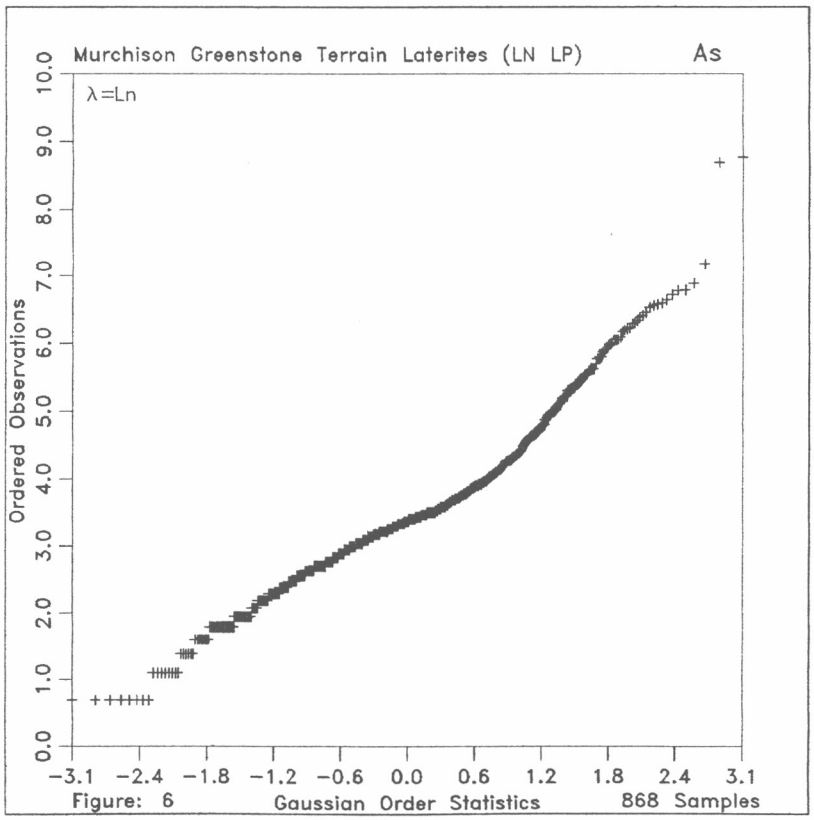


Figure 6b. Q-Q Plot: Arsenic in Laterites [ln Transform], Murchison Greenstone Terrain.

abrupt changes in frequencies can represent multiple populations or missing values within the frequency distribution. The two peaks in the range of 15 and 30 ppm in Figure 4 may reflect different As values within felsic and mafic volcanic rocks, whereas the extreme values above 1000 ppm (ln(1000)=6.9) in Figures 5a,b may be truly anomalous and possibly related to mineralization.

It is usual practice to apply a transformation procedure to data that have a non-normal distribution. The distribution should be examined for outliers before and after a transformation has been fitted to the data. Once the outliers are eliminated, the data should be subjected to an additional transformation fit to ensure that the outliers did not influence the parameters for the transformation. Campbell (1986a) has written computer programs that account for atypical abundances in the estimation of transformations and robust estimates of means and variances. If applying the procedures of Howarth and Earle (1979), the transformation parameters can be fitted after removing data above the 95-99th percentile ranking.

In exploratory data analysis, transformations are useful in assessing whether outliers are the result of an incorrect frequency distribution or are in fact truly atypical values. The use of transformations is more important in statistical testing between populations based upon established reference groups.

Table 1 displays the values of λ used to transform data from a sample population of laterites collected over the Archaean greenstone terrain of the Murchison area (Dataset 1). These values were obtained by procedures outlined by Howarth and Earle (1979) from samples below the 95th percentile.

Table 1

Transformation coefficients used for the Murchison Greenstone Belt Geochemical Data. Box-Cox Generalized Power Transform

$y=(x^{\lambda} - 1)/\lambda$ for $\lambda=0$ $y=\ln$

Fe_2O_3	0.6	Co	0.4
Ag	ln	As	ln
Mn	ln	Sb	ln
Cr	ln	Mo	ln
V	0.4	Sn	ln
Cu	ln	Ga	0.5
Pb	0.4	W	ln
Zn	ln	Nb	0.3
Ni	ln	Au	ln

1.6 Truncated and Censored Data

One of the biggest problems facing the analysis of geochemical data occurs when the data has an abundance less than the detection limit of the method of analysis. This results in large numbers of analyses that have a common value which results in a bias of the distribution. This effect is called censoring where the data are known to have a value, but is assigned some fixed value. Distributions in which censored data are ignored are known as truncated distributions.

Censored distributions are common in geochemical data. All geochemical analyses have a lower limit of detection (lld) imposed by the analytical method. In regional background samples, many of the chalcophile elements or other pathfinder elements have low abundance levels that are less than the lld. Figure 7 shows a theoretical normal distribution with the censored portion shaded. The distribution displays a break at a specific point which represents the lower limit of detection in geochemical analysis. A common practice is to substitute the censored value by one half or one third the detection limit. If the number of samples that fall below the lld are large, then this estimate will produce poor results.

The problem of censored data becomes more apparent when means and covariances are required. Clearly, using the substituted value biases the computation of the moments of the distribution. However, if the nature of the distribution can be assumed to be normal, then the replacement value of the censored data and parameters of the distribution (mean, variance) can be estimated based on the portion of the distribution that is not censored. The estimate of the distribution parameters is obtained using the EM algorithm (Dempster *et al.,* 1977) and is discussed by Campbell (1986b) and Chung (1985, 1988). Chung (1985) and Campbell (1986b) have published computer programs for the statistical treatment of geochemical data with observations below the detection limit. These procedures work by calculating the characteristics of the data above the lld. From these characteristics, an estimate can be made as to how the data is distributed below the lld. The assumption of normality is essential for the EM algorithm to work. Campbell (1986b) invokes an algorithm to transform the data to normality using Box-Cox transformations as described in the previous section.

Gold analyses from the regional background data of the Murchison greenstone terrain (Dataset 1) were subjected to analysis using the CENSOR procedure of Campbell (1986b). The procedure fits an appropriate transformation and substitute value for the distribution. The transformation is necessary since censoring procedures assume that the underlying censored or truncated distribution is normal. The CENSOR program estimated a replacement value of 0.51 ppb for values less than 1 ppb based on a logarithmic transformation of the data. The replacement value of 0.51 ppb ensures that the statistical parameters (mean, standard deviation, etc.) are for a complete distribution. Figure 8 shows the data prior to transformation with a replacement value of 0.51 ppb. Figure 9 shows the same data after a logarithmic transformation and using a replacement value of .51 ppb. The histogram still appears skewed; however, the mean and standard deviation calculated from this replacement value represent values that would be obtained if the data were not censored.

1.7 Summary Tables / Order Statistics

Ranking (ordering) the data is an effective means of observing the distribution of a sample population. Generally, abundances that are above a predetermined threshold or samples that rank above the 98th percentile may be considered anomalous. Although tabulated measures of the 1, 5, 10, 25, 50, 75, 90, 95, 99th percentiles are useful for obtaining actual numbers, graphical displays are the most effective means at describing the distributions. Distributions can be graphically displayed in a variety of ways. Figures 2-6 show three graphical ways in which a sample distribution can be expressed, the histogram, box and whisker plot, and the Q-Q plot.

Summary statistics in the form of a table provide a convenient means by which the distributional characteristics can be viewed from a strictly numerical basis. Summary statistics should include:

- Percentile Rankings: 1, 2, 5, 10, 25, 50, 75, 90, 95, 98, 99th percentiles.
- Maximum Value
- Minimum Value
- Arithmetic Mean
- Median (50th percentile)
- Mode
- Standard deviation
- Left Hinge (25th percentile)
- Right Hinge (75th percentile)

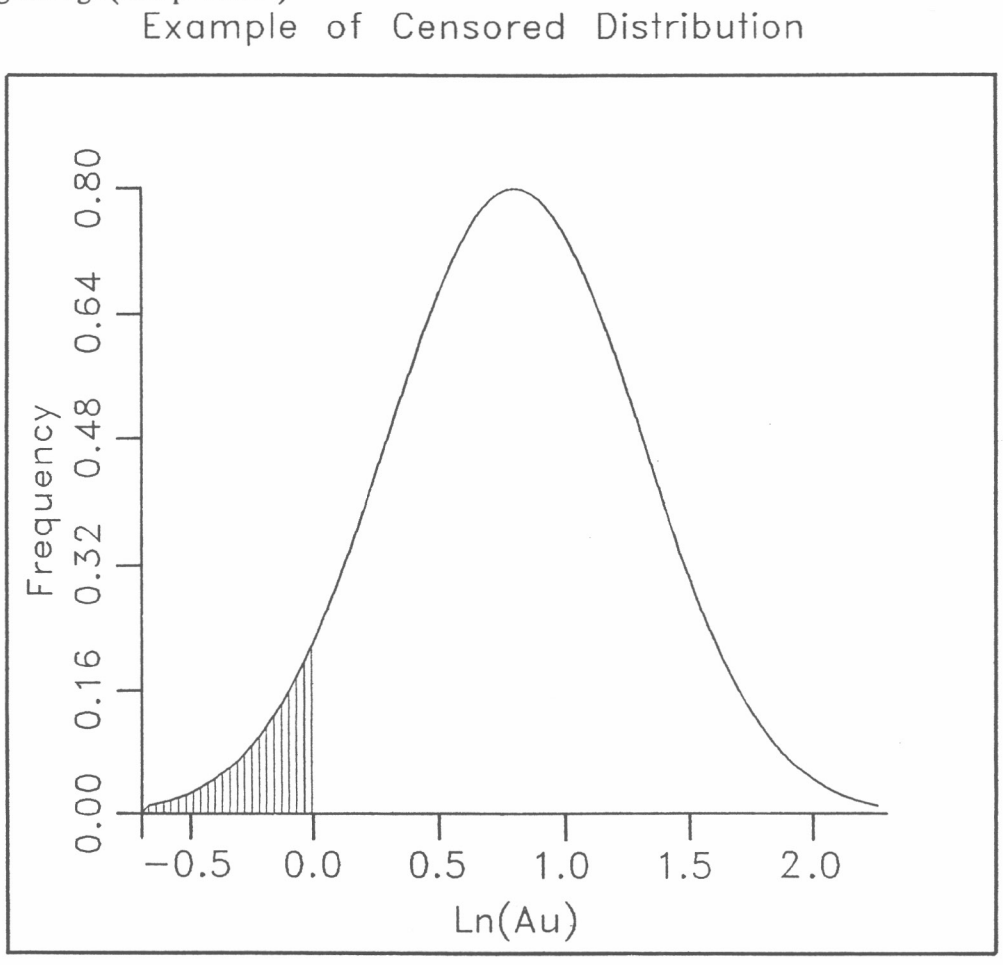
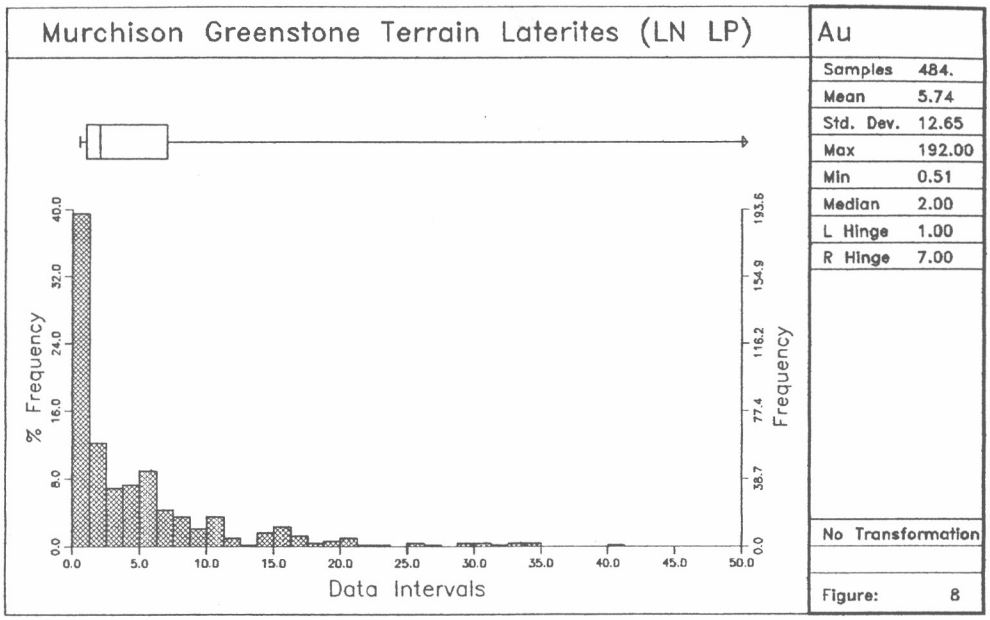


Figure 7. Normal Frequency Distribution: Censored data in Shaded area.

These numerical values are most useful when comparing distributions, particularly for the same element over several types of sample media. The percentile values can provide a quick assessment about the shape of the distribution. When adjacent percentile values are similar or equal then the distribution is skewed. The mode is the value of maximum frequency. It is often difficult to determine computationally since the mode is sensitive to the interval for binning the data required to create histograms. Also, there can be more than one mode (i.e. polymodal) in a distribution and thus a computational mode will not have any real meaning.



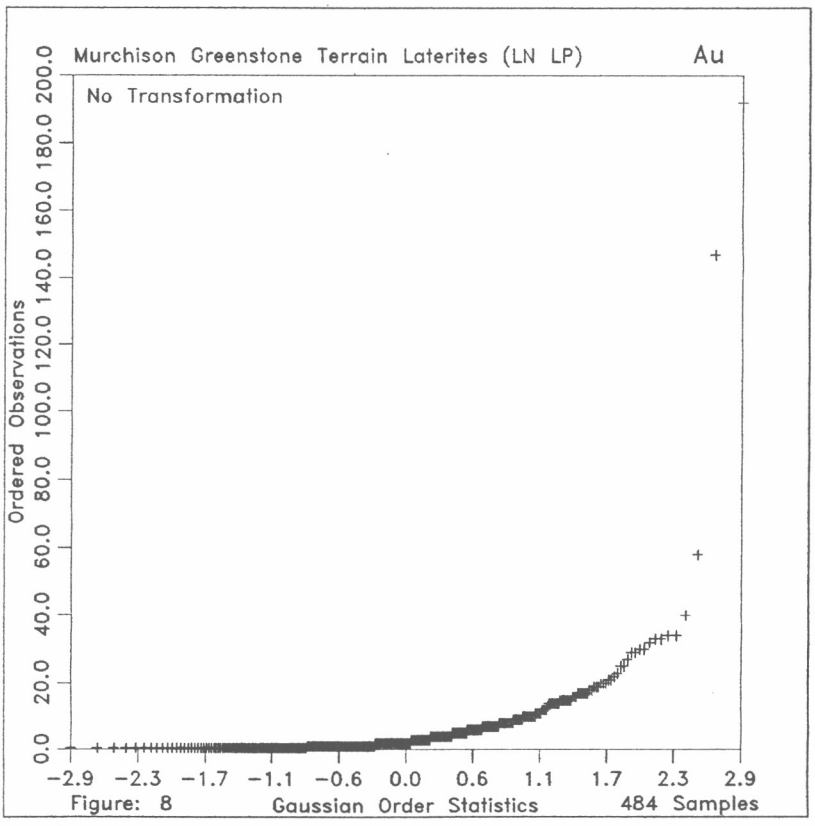
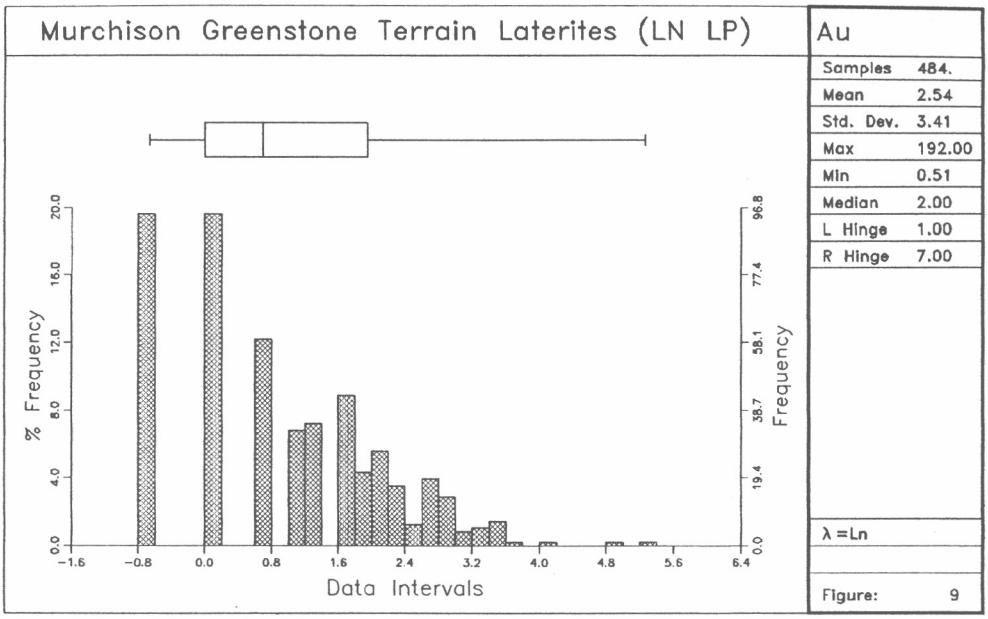


Figure 8. Histogram, Box-Whisker and Q-Q Plots of a Censored Distribution: Gold in Laterites Murchison Greenstone Terrain. Number of frequency intervals = 40. Censored values are those less than 1 ppb and form a peak at the left side of the histogram. A replacement value of 0.51 ppb for samples with values less than the detection limit was derived from analysis of the non-censored portion of the frequency distribution. Note that the censored values are masked by the untransformed histogram.

Tables 2, 3, and 4 are summary tables for three groups of data, a laterite sampling survey over the Murchison greenstone belt, 100 lateritized gossan-nodules collected over the Golden Grove deposit, and a group of lateritic duricrust collected over the Mt. Gibson Au deposit. Each Table displays selected percentile rankings, minimum value, maximum value, median, mode, mean, and standard deviation for selected groups of elements. Comparison between the tables of the upper percentiles, median, and mean values can assist in assessing the geochemical differences between the sample groups.



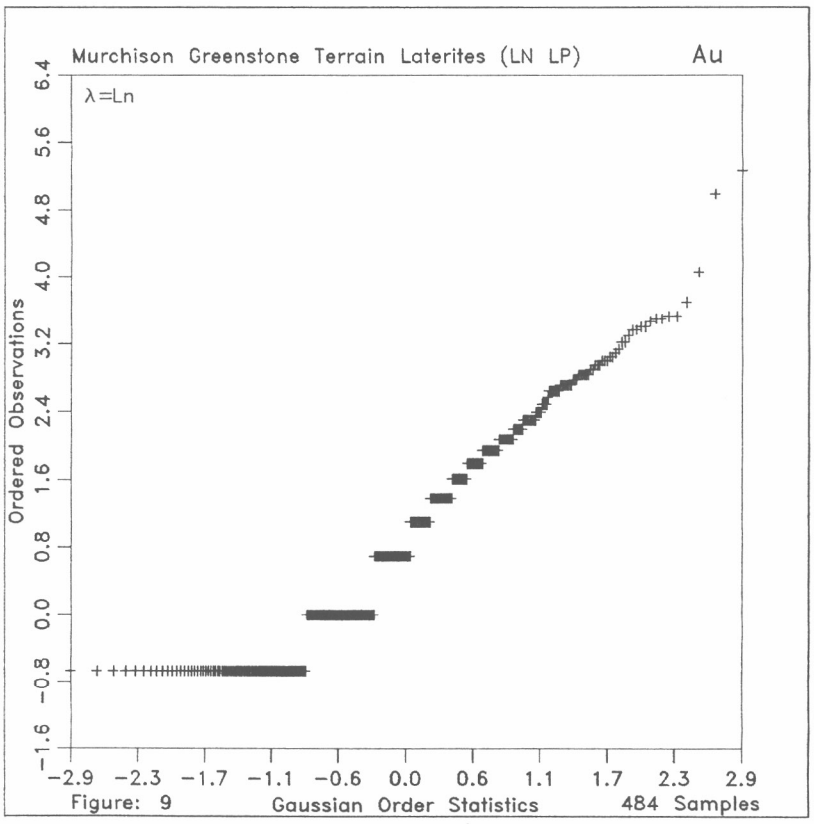


Figure 9. Histogram, Box-Whisker, and Q-Q Plots of a transformed Censored Distribution: Gold in Laterites Murchison Greenstone Terrain. Number of frequency intervals = 40. Censored values are those less than 0 (ln(1 ppb)) and form a peak at the left side of the histogram. A replacement of -0.67 (ln(.51ppb)) is used for samples that are less than the detection limit. The censored values stand out as a separate frequency bar in the transformed population.

Table 2
Summary statistics for Murchison Greenstone Laterites
Sample types: LN LP

No. of Samples in Group: 868

Element		Lab	Method	L.L.D.	#Samples	Percentiles 1%	5%	10%	25%	50%	75%	90%	95%	99%
Fe	Wt%	Amdel	AAS-HF	0.1	868	6.70	10.30	13.80	23.30	34.70	45.20	54.61	59.76	74.77
Ag	ppm	Amdel	OES	0.1	868	0.03	0.03	0.03	0.03	0.03	0.20	0.60	0.80	2.00
Mn	ppm	Amdel	AAS	5.0	868	11.00	21.00	29.00	51.00	110.00	210.00	450.00	806.00	3625.00
Cr	ppm	Amdel	XRF	5.0	868	95.00	172.00	225.00	347.00	707.00	1780.00	4481.00	8453.00	>9999.00
V	ppm	Amdel	XRF	10.0	868	30.00	112.00	178.00	355.00	669.00	1060.00	1473.00	1725.00	2691.00
Cu	ppm	Amdel	AA-HF	2.0	867	4.00	6.00	9.00	20.00	48.00	100.00	232.00	636.00	1025.00
Pb	ppm	Amdel	XRF	4.0	868	0.67	3.00	7.00	17.00	27.00	42.00	64.00	80.00	154.00
Zn	ppm	Amdel	AA-HF	2.0	868	7.00	10.00	12.00	18.00	26.00	38.00	58.00	80.00	139.00
Ni	ppm	Amdel	AA-HF	5.0	868	1.67	1.67	9.00	20.00	40.00	78.00	172.00	292.00	690.00
Co	ppm	Amdel	AA-HF	5.0	868	1.67	1.67	1.67	5.00	8.00	16.00	34.00	45.00	74.00
As	ppm	Amdel	XRF	2.0	868	3.00	6.00	9.00	17.00	29.00	52.00	141.00	258.00	740.00
Sb	ppm	Amdel	XRF	2.0	868	0.67	0.67	0.67	0.67	2.00	4.00	7.00	9.00	35.00
Mo	ppm	Amdel	XRF	2.0	868	0.67	0.67	0.67	0.67	3.00	5.00	9.00	15.00	36.00
Sn	ppm	Amdel	XRF	1.0	866	0.33	0.33	0.33	0.33	2.00	5.00	7.00	8.00	13.00
Ga	ppm	Amdel	OES	1.0	868	3.00	6.00	10.00	15.00	20.00	30.00	30.00	40.00	60.00
W	ppm	Amdel	XRF	10.0	868	3.33	3.33	3.33	3.33	3.33	3.33	16.00	27.00	114.00
Nb	ppm	Amdel	XRF	3.0	863	1.33	1.33	1.33	6.00	10.00	16.00	23.00	27.00	44.00
Au	ppb	Analb	334	1.0	484	0.33	0.33	0.33	1.00	2.67	7.00	14.00	19.00	34.00

Table 2
Summary statistics for Murchison Greenstone Laterites
Sample types: LN LP

Element		Lab	Method	L.L.D.	#Samples	Minimum	Maximum	Median	Mode	Mean	Std. Dev.
Fe	Wt%	Amdel	AAS-HF	0.1	868	2.90	83.78	34.60	33.32	34.61	15.38
Ag	ppm	Amdel	OES	0.1	868	0.03	10.00	0.03	0.04	0.20	0.49
Mn	ppm	Amdel	AAS	5.0	868	1.67	>9999.00	110.00	29.82	256.77	686.92
Cr	ppm	Amdel	XRF	5.0	868	33.00	>9999.00	702.00	243.50	1657.14	2329.25
V	ppm	Amdel	XRF	10.0	868	3.33	5228.00	669.00	304.63	776.40	571.88
Cu	ppm	Amdel	AA-HF	2.0	867	2.00	2211.00	48.00	5.17	115.36	216.79
Pb	ppm	Amdel	XRF	4.0	868	0.67	309.00	27.00	1.06	33.41	28.83
Zn	ppm	Amdel	AA-HF	2.0	868	4.00	592.00	26.00	21.86	33.31	33.00
Ni	ppm	Amdel	AA-HF	5.0	868	1.00	2000.00	40.00	2.45	80.02	158.44
Co	ppm	Amdel	AA-HF	5.0	868	1.67	170.00	8.00	1.88	13.76	17.65
As	ppm	Amdel	XRF	2.0	868	0.67	6407.00	29.00	14.82	78.19	318.01
Sb	ppm	Amdel	XRF	2.0	868	0.67	115.00	2.00	0.71	3.69	8.09
Mo	ppm	Amdel	XRF	2.0	868	0.67	780.00	3.00	0.74	5.15	27.11
Sn	ppm	Amdel	XRF	1.0	866	0.33	72.00	2.00	0.37	3.11	3.78
Ga	ppm	Amdel	OES	1.0	868	1.00	80.00	20.00	19.91	22.06	9.67
W	ppm	Amdel	XRF	10.0	868	3.33	5884.00	3.33	3.45	22.27	264.42
Nb	ppm	Amdel	XRF	3.0	863	1.33	125.00	10.00	1.46	12.11	9.47
Au	ppb	Analb	334	1.0	484	0.33	192.00	2.67	0.99	5.81	12.63

NOTE: Mode estimated by binning of data: # of bins= 100.
Bin width=(95%ile-minimum value)/100.0

Table 3
Summary statistics for Golden Grove Cu-Zn-Au Sulphide Deposit
Laterites
Sample types: LT164

No. of Samples in Group: 100

						Percentiles								
Element		Lab	Method	L.L.D.	#Samples	1%	5%	10%	25%	50%	75%	90%	95%	99%
SiO2	Wt%	Csiro	ICP-FS	0.5	100	4.84	5.64	8.04	11.00	15.66	20.47	27.75	34.63	57.81
Al2O3	Wt%	Csiro	ICP-FS	0.5	100	3.16	4.68	5.12	6.73	9.59	13.10	14.74	16.00	16.83
Fe2O3	Wt%	Csiro	ICP-FS	0.1	100	41.42	49.49	51.45	56.20	61.23	66.99	72.48	75.52	85.50
MgO	Wt%	Csiro	ICP-FS	0.01	100	0.00	0.00	0.00	0.00	0.01	0.02	0.02	0.02	0.04
CaO	Wt%	Csiro	ICP-FS	0.01	100	0.01	0.01	0.01	0.02	0.03	0.04	0.05	0.06	0.09
Na2O	Wt%	Csiro	XRF	0.01	100	0.00	0.00	0.00	0.00	0.01	0.05	0.11	0.14	0.18
K2O	Wt%	Csiro	XRF	0.01	100	0.02	0.04	0.05	0.08	0.28	0.82	1.66	2.07	2.70
TiO2	Wt%	Csiro	XRF	0.001	100	0.22	0.29	0.34	0.46	0.65	0.91	1.06	1.27	2.46
P2O5	Wt%	Csiro	XRF	0.002	100	0.01	0.01	0.02	0.02	0.02	0.03	0.04	0.04	0.05
Mn	ppm	Csiro	XRF	5.0	100	1.67	1.67	7.00	12.00	17.00	27.00	36.00	46.00	100.00
Cr	ppm	Csiro	XRF	10.0	100	66.00	83.00	102.00	128.00	167.00	196.00	231.00	259.00	293.00
V	ppm	Csiro	XRF	10.0	100	72.00	94.00	110.00	145.00	184.00	206.00	235.00	256.00	282.00
Cu	ppm	Csiro	XRF	10.0	100	12.00	23.00	26.00	38.00	81.00	111.00	145.00	159.00	212.00
Pb	ppm	Csiro	XRF	5.0	100	52.00	54.00	56.00	67.00	76.00	85.00	100.00	103.00	121.00
Zn	ppm	Csiro	XRF	2.0	100	0.67	3.00	4.00	7.00	10.00	14.00	25.00	40.00	151.00
Ni	ppm	Csiro	XRF	5.0	100	1.67	1.67	1.67	6.00	10.00	15.00	21.00	24.00	48.00
Co	ppm	Csiro	XRF	5.0	100	1.67	1.67	1.67	12.00	20.00	30.00	39.00	54.00	69.00
As	ppm	Csiro	XRF	2.0	100	239.00	281.00	295.00	376.00	451.00	557.00	636.00	683.00	732.00
Sb	ppm	Csiro	XRF	2.0	100	16.00	18.00	19.00	24.00	31.00	36.00	43.00	47.00	48.00
Bi	ppm	Csiro	XRF	1.0	100	35.00	49.00	53.00	60.00	74.00	94.00	106.00	143.00	168.00
In	ppm	Csiro	XRF	1.0	100	6.00	9.00	11.00	14.00	22.00	26.00	31.00	32.00	38.00
Mo	ppm	Csiro	XRF	1.0	100	0.33	0.33	1.00	1.00	2.00	4.00	5.00	6.00	14.00
Ag	ppm	Amdel	OES	0.1	100	0.03	0.03	0.10	0.20	0.30	0.40	0.70	0.80	1.50
Sn	ppm	Csiro	XRF	2.0	100	8.00	13.00	14.00	22.00	36.00	67.00	118.00	162.00	447.00
Ge	ppm	Amdel	OES	1.0	100	0.33	0.33	0.33	0.33	0.33	1.00	2.00	3.00	4.00
Ga	ppm	Csiro	XRF	3.0	100	27.00	34.00	36.00	44.00	52.00	61.00	69.00	73.00	132.00
W	ppm	Csiro	XRF	5.0	100	1.67	6.00	8.00	12.00	15.00	19.00	26.00	34.00	46.00
Ba	ppm	Csiro	XRF	10.0	100	3.33	12.00	14.00	28.00	87.00	226.00	356.00	437.00	2234.00
Zr	ppm	Csiro	XRF	3.0	100	161.00	252.00	315.00	408.00	543.00	704.00	865.00	1033.00	1130.00
Sr	ppm	Csiro	XRF	2.0	100	0.67	0.67	0.67	3.00	5.00	7.00	10.00	10.00	14.00
Nb	ppm	Csiro	XRF	2.0	100	0.67	0.67	0.67	0.67	4.00	7.00	11.00	14.00	29.00
Se	ppm	Csiro	XRF	2.0	100	4.00	5.00	6.00	7.00	9.00	11.00	14.00	17.00	20.00
Be	ppm	Csiro	ICP	0.01	96	0.00	0.00	0.00	0.00	0.03	0.18	0.31	0.32	0.60
Au	ppb	Becql	INAA	5.0	100	220.00	330.00	450.00	1430.00	3790.00	8480.00	10600.00	12900.00	18600.00
La	ppm	Becql	INAA	0.5	100	0.72	0.93	1.10	1.40	2.10	3.20	4.90	5.70	9.50
Rb	ppm	Becql	INAA	20.0	100	33.33	52.00	75.00	88.00	100.00	120.00	130.00	140.00	170.00
Ce	ppm	Becql	INAA	2.0	100	0.67	0.67	0.67	4.00	5.70	8.20	12.70	16.70	53.00
Sc	ppm	Becql	INAA	0.1	100	6.80	7.90	8.90	12.40	16.40	20.70	23.60	25.50	35.30
Y	ppm	Csiro	XRF	3.0	100	8.00	10.00	12.00	19.00	25.00	35.00	41.00	52.00	65.00
Yb	ppm	Becql	INAA	0.5	100	1.00	1.40	2.00	2.60	4.10	5.40	6.40	7.40	10.00
Th	ppm	Becql	INAA	0.5	100	6.40	13.00	19.00	23.00	30.00	39.00	50.40	65.90	84.60
Eu	ppm	Becql	INAA	0.5	100	0.17	0.17	0.17	0.17	0.17	0.67	0.78	0.91	1.00
Hf	ppm	Becql	INAA	1.0	100	4.40	6.90	8.60	12.00	15.00	21.00	26.00	29.00	34.00
Lu	ppm	Becql	INAA	0.2	100	0.07	0.22	0.32	0.52	0.75	1.00	1.20	1.40	1.80
Sm	ppm	Becql	INAA	0.2	100	0.55	0.66	0.79	1.00	1.30	1.50	1.60	1.80	3.20

Table 3
Summary statistics for Golden Grove Block Cu-Zn-Au Sulphide Deposit
Laterites
Sample types: LT164

No. of Samples in Group: 100

Element		Lab	Method	L.L.D.	#Samples	Minimum	Maximum	Median	Mode	Mean	Std. Dev.
SiO2	Wt%	Csiro	ICP-FS	0.5	100	4.39	57.81	15.49	10.29	16.93	9.02
Al2O3	Wt%	Csiro	ICP-FS	0.5	100	2.99	16.83	9.57	12.03	9.74	3.67
Fe2O3	Wt%	Csiro	ICP-FS	0.1	100	36.61	85.50	61.00	57.04	61.56	8.61
MgO	Wt%	Csiro	ICP-FS	0.01	100	0.00	0.04	0.01	0.00	0.01	0.01
CaO	Wt%	Csiro	ICP-FS	0.01	100	0.00	0.09	0.03	0.03	0.03	0.01
Na2O	Wt%	Csiro	XRF	0.01	100	0.00	0.18	0.01	0.00	0.03	0.05
K2O	Wt%	Csiro	XRF	0.01	100	0.02	2.70	0.28	0.05	0.57	0.66
TiO2	Wt%	Csiro	XRF	0.001	100	0.21	2.46	0.63	0.91	0.70	0.35
P2O5	Wt%	Csiro	XRF	0.002	100	0.01	0.05	0.02	0.02	0.02	0.01
Mn	ppm	Csiro	XRF	5.0	100	1.67	100.00	17.00	12.97	20.35	14.55
Cr	ppm	Csiro	XRF	10.0	100	37.00	293.00	165.00	126.91	164.50	50.78
V	ppm	Csiro	XRF	10.0	100	72.00	282.00	183.00	196.20	174.70	47.18
Cu	ppm	Csiro	XRF	10.0	100	10.00	212.00	81.00	89.71	79.36	45.23
Pb	ppm	Csiro	XRF	5.0	100	46.00	121.00	75.00	78.77	76.87	15.11
Zn	ppm	Csiro	XRF	2.0	100	0.67	151.00	10.00	7.16	13.99	19.33
Ni	ppm	Csiro	XRF	5.0	100	1.67	48.00	10.00	1.78	10.94	7.93
Co	ppm	Csiro	XRF	5.0	100	1.67	69.00	20.00	1.93	22.06	14.55
As	ppm	Csiro	XRF	2.0	100	235.00	732.00	449.00	385.08	458.53	123.23
Sb	ppm	Csiro	XRF	2.0	100	14.00	48.00	30.00	32.97	30.37	8.27
Bi	ppm	Csiro	XRF	1.0	100	30.00	168.00	73.00	70.12	78.97	26.17
In	ppm	Csiro	XRF	1.0	100	4.00	38.00	22.00	24.02	20.44	7.60
Mo	ppm	Csiro	XRF	1.0	100	0.33	14.00	2.00	2.01	2.75	2.03
Ag	ppm	Amdel	OES	0.1	100	0.03	1.50	0.30	0.20	0.33	0.24
Sn	ppm	Csiro	XRF	2.0	100	6.00	447.00	36.00	14.58	57.56	66.14
Ge	ppm	Amdel	OES	1.0	100	0.33	4.00	0.33	0.35	0.94	0.89
Ga	ppm	Csiro	XRF	3.0	100	27.00	132.00	51.00	43.79	53.02	14.69
W	ppm	Csiro	XRF	5.0	100	1.67	46.00	15.00	15.09	16.23	7.67
Ba	ppm	Csiro	XRF	10.0	100	3.33	2234.00	82.00	14.18	168.99	284.84
Zr	ppm	Csiro	XRF	3.0	100	117.00	1130.00	540.00	387.22	567.12	225.81
Sr	ppm	Csiro	XRF	2.0	100	0.67	14.00	5.00	6.97	5.50	3.07
Nb	ppm	Csiro	XRF	2.0	100	0.67	29.00	3.00	0.73	4.64	4.74
Se	ppm	Csiro	XRF	2.0	100	4.00	20.00	9.00	7.96	9.22	3.30
Be	ppm	Csiro	ICP	0.01	96	0.00	0.60	0.03	0.00	0.10	0.13
Au	ppb	Becql	INAA	5.0	100	180.00	18600.00	3740.00	498.00	4861.45	4135.75
La	ppm	Becql	INAA	0.5	100	0.61	9.50	2.10	1.60	2.59	1.68
Rb	ppm	Becql	INAA	20.0	100	25.00	170.00	100.00	109.52	101.86	25.70
Ce	ppm	Becql	INAA	2.0	100	0.67	53.00	5.70	0.75	7.11	6.76
Sc	ppm	Becql	INAA	0.1	100	5.30	35.30	16.40	13.89	16.56	5.66
Y	ppm	Csiro	XRF	3.0	100	8.00	65.00	25.00	20.10	26.69	12.14
Yb	ppm	Becql	INAA	0.5	100	0.50	10.00	4.00	2.40	4.12	1.85
Th	ppm	Becql	INAA	0.5	100	6.10	84.60	29.00	21.95	32.80	14.56
Eu	ppm	Becql	INAA	0.5	100	0.17	1.00	0.17	0.17	0.42	0.28
Hf	ppm	Becql	INAA	1.0	100	2.80	34.00	15.00	12.89	16.35	6.70
Lu	ppm	Becql	INAA	0.2	100	0.07	1.80	0.74	1.01	0.76	0.36
Sm	ppm	Becql	INAA	0.2	100	0.53	3.20	1.20	1.40	1.24	0.38

NOTE: Mode estimated by binning of data: # of bins= 100.
Bin width=(95%ile-minimum value)/100.0

Table 2 provides summary statistics for samples collected over the Murchison greenstone terrain using a regional grid spacing of 3 km for each sample. The abundances of the elements reflect variations over a diverse range of underlying lithologies and hence display large standard deviations. The abundances for the elements listed in Table 2 reflect regional background variation.

Table 3 displays the summary statistics for the Golden Grove massive sulphide deposit. In comparison with the regional background geochemical data of Table 2, this area shows increased Fe_2O_3, Ag, Cu, Pb, As, Sb, Sn, Ga, and Au.

Table 4 displays the summary statistics for lateritic duricrust from the Mt. Gibson Au mine. Using the median composition as a basis of comparison with the regional background samples of Table 2, V, Pb, Sb, Ga, and Au are more abundant in the Yilgarn Block Au mine samples. There is some suggested depletion of Cr and Ni, but this may reflect a lack of more mafic rocks associated with the Au deposit.

Comparison of the two target groups (Tables 3 and 4) indicates that they differ in their abundances of SiO_2, Al_2O_3, Fe_2O_3, TiO_2, Mn, Cr, V, Cu, Pb, As, Sb, Bi, Sn, Ge, W, Ba, Zr, Nb, Se, and Be and suggests that these two types of deposits are geochemically distinct.

1.8 Robust Estimation

The importance of an accurate estimate of distribution parameters (mean, variance) cannot be over-emphasized. The application of standard statistical procedures requires the assumption that the data are normally distributed. As illustrated in Figure 1b, if the distribution of an element is not normal, the arithmetic mean can be significantly different from the median or mode of the distribution. This type of distortion can have significant effects in the estimates of distribution parameters, particularly for the estimates of variance.

Table 4
Summary statistics for Mt. Gibson Gold Mine
Lateritic duricrust
Sample types: LT200

No. of Samples in Group: 121

						Percentiles								
Element		Lab	Method	L.L.D.	#Samples	1%	5%	10%	25%	50%	75%	90%	95%	99%
SiO2	Wt%	Analb	ICP	0.1	121	20.53	25.41	28.20	31.23	35.00	40.20	47.00	50.27	55.61
Al2O3	Wt%	Analb	ICP	0.1	121	12.59	15.68	17.91	19.70	21.91	24.21	26.41	27.94	31.00
Fe	Wt%	Analb	ICP	0.1	121	9.14	13.49	16.82	20.65	28.95	34.46	39.65	42.05	49.62
MgO	Wt%	Analb	ICP	0.003	121	0.03	0.05	0.05	0.08	0.13	0.25	0.54	0.71	2.57
CaO	Wt%	Analb	ICP	0.007	121	0.00	0.02	0.02	0.03	0.08	0.29	1.95	4.94	15.59
Na2O	Wt%	Analb	ICP	0.007	121	0.01	0.01	0.02	0.05	0.12	0.21	0.31	0.40	0.54
K2O	Wt%	Analb	ICP	0.06	121	0.02	0.02	0.02	0.06	0.09	0.14	0.24	0.34	0.47
TiO2	Wt%	Analb	ICP	0.003	121	0.50	0.66	0.81	0.92	1.18	1.39	1.70	2.13	2.50
Mn	ppm	Analb	ICP	15.0	121	22.00	28.00	31.00	46.00	73.00	117.00	171.00	215.00	268.00
Cr	ppm	Analb	ICP	20.0	121	184.00	267.00	321.00	415.00	548.00	718.00	945.00	1094.00	1460.00
V	ppm	Analb	ICP	5.0	121	273.00	337.00	384.00	549.00	785.00	955.00	1210.00	1349.00	2071.00
Cu	ppm	Amdel	AA-HF	2.0	121	8.00	24.00	25.00	32.00	48.00	60.00	94.00	150.00	240.00
Pb	ppm	Amdel	XRF	2.0	121	15.00	22.00	30.00	36.00	50.00	62.00	80.00	92.00	104.00
Zn	ppm	Amdel	AA-HF	2.0	121	0.67	5.00	6.00	10.00	15.00	22.00	30.00	36.00	48.00
Ni	ppm	Amdel	AA-HF	4.0	121	14.00	22.00	26.00	32.00	42.00	52.00	62.00	66.00	84.00
Co	ppm	Amdel	AA-HF	4.0	121	1.33	1.33	1.33	4.00	8.00	12.00	18.00	20.00	24.00
As	ppm	Amdel	XRF	2.0	121	6.00	10.00	13.00	17.00	27.00	40.00	54.00	90.00	132.00
Sb	ppm	Amdel	XRF	2.0	121	0.67	0.67	1.00	4.00	6.00	8.00	11.00	12.00	15.00
Bi	ppm	Amdel	XRF	2.0	121	0.67	0.67	0.67	3.00	5.00	7.00	11.00	14.00	18.00
Mo	ppm	Amdel	XRF	1.0	121	0.33	0.67	2.00	3.00	4.00	5.00	7.00	8.00	10.00
Ag	ppm	Amdel	AAS	0.1	121	0.03	0.03	0.03	0.03	0.40	1.00	1.50	2.00	2.50
Sn	ppm	Amdel	XRF	2.0	121	0.67	0.67	0.67	0.67	2.00	3.00	4.00	5.00	7.00
Ge	ppm	Amdel	XRF	2.0	121	0.67	0.67	0.67	0.67	1.33	1.33	2.00	3.00	4.00
Ga	ppm	Amdel	XRF	4.0	121	16.00	26.00	30.00	40.00	48.00	58.00	68.00	72.00	82.00
W	ppm	Amdel	XRF	4.0	121	1.33	1.33	1.33	1.67	6.00	10.00	12.00	14.00	19.00
Ba	ppm	Analb	ICP	5.0	121	7.00	18.00	21.00	58.00	157.00	356.00	912.00	1469.00	2970.00
Zr	ppm	Analb	ICP	5.0	121	62.00	83.00	93.00	111.00	130.00	154.00	239.00	275.00	314.00
Nb	ppm	Amdel	XRF	2.0	121	1.00	3.00	5.00	6.00	8.00	11.00	13.00	14.00	18.00
Se	ppm	Amdel	XRF	2.0	121	0.67	0.67	0.67	2.00	4.00	6.00	8.00	10.00	11.00
Be	ppm	Analb	ICP	1.0	121	0.33	0.33	0.33	0.33	1.00	1.00	2.00	2.00	2.00
Au	ppb	Analb	334	1.0	120	48.00	160.00	260.00	640.00	1400.00	2050.00	3020.00	4220.00	5900.00

Table 4
Summary statistics for Mt. Gibson Gold Mine
Lateritic Duricrust
Sample types: LT200

Element		Lab	Method	L.L.D.	#Samples	Minimum	Maximum	Median	Mode	Mean	Std. Dev.
SiO2	Wt%	Analb	ICP	0.1	63	11.80	56.70	28.76	23.99	29.32	9.45
Al2O3	Wt%	Analb	ICP	0.1	63	9.35	30.41	19.30	14.25	19.33	4.61
Fe	Wt%	Analb	ICP	0.1	63	19.66	67.56	35.57	29.34	38.75	10.95
MgO	Wt%	Analb	ICP	0.003	63	0.02	0.89	0.05	0.05	0.11	0.16
CaO	Wt%	Analb	ICP	0.007	63	0.02	1.14	0.04	0.03	0.11	0.22
Na2O	Wt%	Analb	ICP	0.007	63	0.01	0.39	0.02	0.01	0.07	0.10
K2O	Wt%	Analb	ICP	0.06	63	0.02	0.63	0.02	0.02	0.07	0.09
TiO2	Wt%	Analb	ICP	0.003	63	0.47	4.24	1.25	1.41	1.59	0.89
Mn	ppm	Analb	ICP	15.0	63	6.67	409.00	87.00	106.92	103.29	71.66
Cr	ppm	Analb	ICP	20.0	63	181.00	1842.00	702.00	618.46	822.29	326.41
V	ppm	Analb	ICP	5.0	63	383.00	3220.00	1037.00	835.88	1165.78	522.84
Cu	ppm	Amdel	AA-HF	2.0	63	13.00	300.00	40.00	28.01	49.37	39.77
Pb	ppm	Amdel	XRF	2.0	63	33.00	158.00	64.00	53.91	70.86	26.43
Zn	ppm	Amdel	AA-HF	2.0	63	2.00	56.00	14.00	11.90	16.11	10.40
Ni	ppm	Amdel	AA-HF	4.0	63	22.00	80.00	36.00	29.95	39.92	13.05
Co	ppm	Amdel	AA-HF	4.0	63	1.33	18.00	10.00	12.06	9.80	4.56
As	ppm	Amdel	XRF	2.0	63	3.00	115.00	35.00	41.60	38.41	24.76
Sb	ppm	Amdel	XRF	2.0	63	0.67	11.00	5.00	6.03	4.89	2.94
Bi	ppm	Amdel	XRF	2.0	63	0.67	52.00	5.00	4.06	7.90	8.47
Mo	ppm	Amdel	XRF	1.0	63	0.67	16.00	5.00	5.03	4.62	2.60
Ag	ppm	Amdel	AAS	0.1	63	0.03	2.10	0.50	0.04	0.71	0.59
Sn	ppm	Amdel	XRF	2.0	63	0.67	12.00	3.00	2.98	3.85	2.34
Ge	ppm	Amdel	XRF	2.0	63	0.67	9.00	0.67	0.69	1.69	1.56
Ga	ppm	Amdel	XRF	4.0	63	10.00	195.00	60.00	60.31	63.75	26.62
W	ppm	Amdel	XRF	4.0	63	1.33	46.00	8.00	5.98	10.77	8.33
Ba	ppm	Analb	ICP	5.0	63	6.00	215.00	27.00	18.49	50.95	48.69
Zr	ppm	Analb	ICP	5.0	63	72.00	356.00	151.00	119.43	161.48	52.25
Nb	ppm	Amdel	XRF	2.0	63	1.00	32.00	9.00	9.01	10.00	5.37
Se	ppm	Amdel	XRF	2.0	63	0.67	11.00	6.00	3.98	5.98	2.63
Be	ppm	Analb	ICP	1.0	63	0.33	3.00	0.33	0.34	0.91	0.72
Au	ppb	Analb	334	1.0	63	5.00	12757.00	1640.00	54.53	2435.49	2869.97

NOTE: Mode estimated by binning of data: # of bins = 100.
Bin width = (95%ile-minimum value)/100.0

Methods of robust estimation are primarily concerned with minimizing the influence of samples that are atypical. There are several methods by which this can be done. Rock (1987) provides a review of several methods of robust estimates and a number of computer subroutines for robust estimates of location (mean/median) and scale (variance). An example of robust estimation is given in Table 5. Most robust procedures are multivariate. Garrett (1989b, 1990) suggests that reasonable robust estimates of univariate or multivariate data can be obtained by selecting only those samples that are less than the 90th or 95th percentile. Naturally, some knowledge of the distribution of the data will assure that using such a procedure is realistic. By eliminating the upper percentile data, it is assumed that all of the outliers will be removed in the estimates of the means and covariances. Chork (1990) has demonstrated a more robust way of determining means and correlations/covariances with minimal effect from outliers. The procedure is known as robust estimation using the minimum volume ellipsoid method.

Robust estimates should be applied to transformed data. Robust estimates of distributions that are skewed, such as log normal distributions, may underestimate the mean and standard deviation if transformations are not applied first. This can be observed in Table 5 for As.

1.9 Spatial Presentation of Ranked Data

Since geochemical data are generally spatially distributed, it is useful to present geochemical analyses in map form. One of the simplest ways of identifying anomalous samples is to plot the ranked data onto the map with symbol sizes that are proportional to the ranking of the data. Thus, samples that are above the 95th percentile will have a symbol size larger than samples at the 50th percentile. This method of reporting geochemical data is discussed by Howarth (1983, Chapter 5).

Symbol maps have been used to outline the geochemical variation over 10 class-intervals for individual elements, principal component scores, and multivariate indices. The symbol sizes have been chosen such that only samples above the 95th percentile are easily recognized by sight. In the case of principal component scores, samples that exceed the 95th percentile and those that are less than the 5th percentile are given symbol sizes that are easily recognized. The symbol size is generated using a non-linear fourth order polynomial function as follows:

$$\text{Map Symbol Size} = \text{Minimum Symbol Size for Map} + \text{Constant Symbol Size} * (\text{Percentile}/100)^4.$$

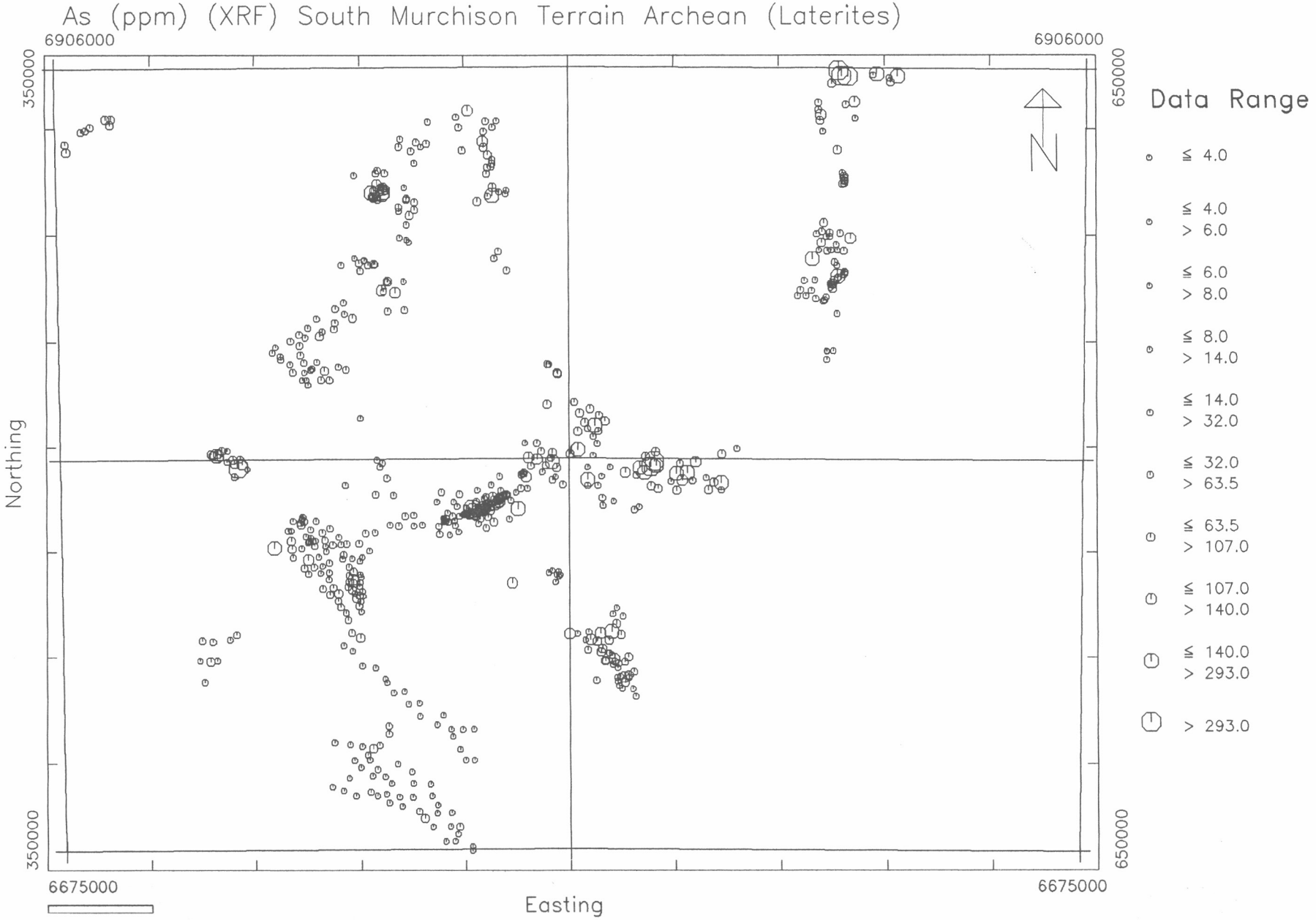


Figure 10. Map of Arsenic Values in Laterite from Archaean Greenstone Terrain: Samples greater than 140 ppm rank above the 95th percentile.

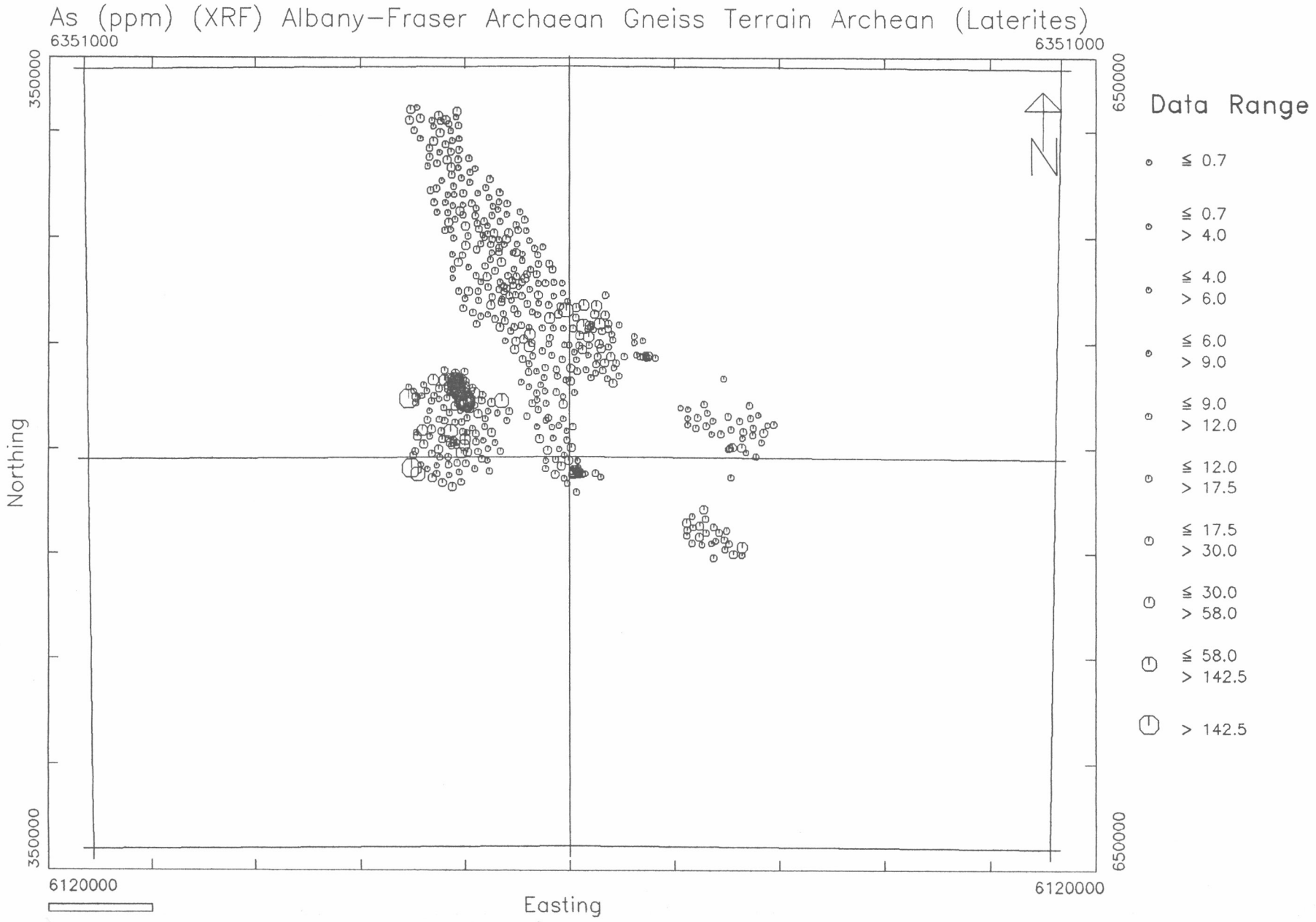


Figure 11. Map of Arsenic Values in Laterite from Archaean Gneiss Terrain: Samples greater than 58 ppm rank above the 95th percentile.

For a valid interpretation, the full range of values over an area should be represented in the population. A large symbol size does not necessarily imply that the sample is atypical (i.e. anomalous). Large symbol sizes merely indicate which samples exceed the 95th percentile rankings. Thus, it is important to know the range of values that are associated with the element. Comparison between Figures 10 and 11 illustrates this. Figure 10 shows a map of As for samples collected over a greenstone belt in the Yilgarn Block, while Figure 11 shows As values for samples collected over a granite/gneiss terrain also in the Yilgarn Block. Both maps show large symbols for As values above the 95th percentile. However, the 95th percentile ranking is 140 ppm for the greenstone belt area and 58 ppm for the gneiss terrain area. Clearly, the samples over the gneiss terrain area do not display significant As values relative to As values in the greenstone belt area.

The interpretation of these patterns requires knowledge about the underlying lithologies and geological processes that may have occurred. Generally, As abundance is considered to be independent of primary variations within igneous rocks and is usually associated with mineralized zones. Thus, the signature of As in the Yilgarn greenstone area is more likely to be indicative of zones of potential mineralization, particularly given the larger abundances relative to As observed in the granite/gneiss terrain.

Table 5
Robust Estimates of Murchison Greenstone Belt Loose Nodules
Observations: 484

	Initial Estimates (weights=1.0)		Final Robust Estimates		
	Mean	Std.Dev.	Mean	Std.Dev.	
Fe	37.1756	15.3438	35.5976	15.0435	Fe
Ag	0.2754	0.6392	0.1879	0.2317	Ag
Mn	236.4966	632.1769	127.0169	102.9410	Mn
Cr	1638.3368	2142.5525	1078.6084	967.4877	Cr
V	782.2824	619.8652	748.2899	571.8699	V
Cu	154.0868	262.1634	154.8221	261.6487	Cu
Pb	35.2789	29.3632	29.8245	17.2313	Pb
Zn	36.7066	40.2301	28.6707	17.9285	Zn
Ni	90.9277	154.5024	59.2149	48.0592	Ni
Co	11.6074	12.1092	8.7999	6.7564	Co
As	112.3237	418.5397	54.5992	59.9303	As
Sb	4.1687	8.4046	3.2118	3.2839	Sb
Mo	6.5475	35.7867	3.8618	3.4544	Mo
Sn	2.8939	3.0594	2.5815	2.5910	Sn
Ga	22.0909	9.7060	22.4347	8.7234	Ga
W	12.3430	69.3784	6.3215	6.7272	W
Nb	11.2700	9.4191	10.4555	8.3822	Nb
Au	5.8147	12.6286	4.8749	5.7870	Au

2.0 GEOCHEMICAL DEFINITIONS AND TERMINOLOGY

2.1 Definition of Threshold

There are several definitions of the term threshold. Garrett (1991) defines the threshold as the "outer" limit of local background variation. The term "outer" is used instead of "upper". This allows for the definition to include both "upper" and "lower" limits as it is common in some geochemical environments for depletion haloes to be as significant as enrichment haloes (see Robertson and Taylor, 1987). The threshold can be a very elusive value to obtain. The threshold of a geochemical distribution is considered to be the abundance level that is considered to exceed the accepted value for a background or regional population. Another way of expressing this concept is that the threshold is the minimum abundance level of an anomaly. Selection of anomalies depends upon choosing a suitable threshold. If the threshold is too low, then there will be too many areas to follow-up. If the threshold is too high, then some potentially-significant targets could be missed.

Samples from anomalous distributions usually overlap with samples from background distributions such that the threshold is more likely a range of values where the two distributions overlap. Rather than choose a specific threshold value, it may be better to assign a probability of the likelihood of an unknown sample belonging to each population. This modelled approach will be discussed in more detail later.

WARNING: One of the earliest definitions of a geochemical anomaly was by Hawkes and Webb (1962), where the threshold of a population is determined as the mean ± 2 standard deviations. This is based upon the assumption of normality. Later researchers (Rose *et al.*, 1979; Levinson, 1980; Garret, 1989a) have warned geochemists not to fall into the trap of defining thresholds as the mean ± 2 standard deviations. This is probably one of the least suitable ways in which anomalous values can be selected.

In univariate ranking, the choice of a suitable threshold is based upon information derived from orientation studies. Orientation studies may show that the threshold is at a level that is not very obvious (effects of lithology, etc.). An initial estimate of the threshold can be selected as the 95th percentile of the ranked samples. However, the distribution of the data should be examined first by the use of graphical methods, such as histograms, Q-Q plots, probability plots, or box and whisker plots. The choice of a threshold is largely a visually-based one. Table 2 indicates that the 95th percentile for As is 258 ppm. This suggests that values above this **threshold** value are anomalous. Figures 5a,b show that several values occur above the level of 258 ppm and could be considered suitable target sites for further evaluation in an exploration programme.

2.2 Definition of Geochemical Anomaly

There has been much debate about the definition of an anomaly and currently there is no clear consensus for the definition. Howarth and Sinding-Larsen (1983:208) discuss the concept of anomaly and suggest that anomalous concentrations are those values that exceed a given threshold. Workshops held by the Association of Exploration Geochemists (AEG) in 1983 and 1985 (Garrett, 1984; Aucott, 1987) failed to give any formal definition and concluded that an anomaly is a desired level of abundance in which the geologist has a particular interest and is different from the regional and/or background values and implies a spatially-continuous set of values. Anomalies imply spatial association and are small in area relative to the regional population that is being sampled.

A current working definition of anomaly used within the Laterite Geochemistry Section of CSIRO is a single or multiple group of samples composed of one or more elemental abundances that exceed the threshold expected for the regional or background population. The threshold values have been determined from orientation studies carried out over areas being sampled. In order to recognize a geochemical anomaly, the range of abundances must be different from the regional background abundances.

2.3 Outliers

Outliers can be defined as samples with abundance levels for one or more elements that are significantly different from samples with which they are initially grouped. These samples may be of significance from an exploration point of view and may be part of a mineralized zone (anomaly). Outliers can also be artifacts of erroneous analytical results or data entries. Outliers should always be carefully examined to be certain that the observed abundances are not the result of an error.

Samples are outliers when they do not fit within expected population parameters. A sample that is an outlier in one group may be indistinguishable from other samples within another group. In practice, outliers are assessed by examining the upper and lower percentile rankings of data, usually in the range of greater than the 95th or 98th percentile or less than the 5th or 2nd percentile. However, if the sample population if non-normal, bimodal, or polymodal, then it is necessary to make the appropriate adjustments to the data. In the case of multimodal populations, it will be necessary to split the populations into separate distinct populations through the analysis of Q-Q plots, probability plots, or by computer-based means using such methods as outlined by Campbell (1986a), Stanley and Sinclair (1987), or Bridges and McCammon (1980). Stanley has developed a

computer program, PROBPLOT, designed to be run on IBM-PC microcomputers and is available from the Association of Exploration Geochemists.

2.4 Indicator/Pathfinder Elements

Indicator elements are elements that are associated with mineralization. Pathfinder elements are indicator elements that occur in association with the target element being sought and usually have little economic significance (Joyce, 1984:9). An effective pathfinder element is usually more abundant and is distributed over a larger areal extent than the target element being sought. Pathfinder elements can be selected from multivariate data analysis techniques in conjunction with known geochemical/geological affinities of the elements.

The associations of indicator/pathfinder elements which occur within base metal and precious metals deposits are well documented in the geochemical literature. Many examples of multi-element haloes associated with Au deposits in Australian gossans and laterites have been cited in the literature (Smith *et al.,* 1984; Robertson and Taylor, 1987). In lode Au deposits in the Superior Province of Canada, pathfinder elements are commonly associated with Au (Colvine *et al.,* 1988). Pathfinder elements associated with lode Au deposits commonly include As, Sb, W, Mo, Bi, Ag, Li, Ba, Rb, Cr, Cu, Zn, and Pb. The presence of these elements can assist in choosing follow-up sites for further scrutiny in an exploration programme.

2.5 Target and Background Populations

Sample populations that represent the element(s) being sought are termed **"Target"** populations. These populations are derived from samples collected in orientation studies over known mineral deposits or areas of specific interest. Homogeneity of the target groups is important for the correct application of statistical procedures.

Sample populations that represent the regional geochemical background material are termed **"Background"** populations. Separation of the background population into similar subsets that represent homogeneous multivariate normal populations is important and forms the basis of the modelled approach of geochemical data analysis. This can be achieved using components analysis, spatial analysis, χ^2 plots, etc..

Typical background populations would be represented by regional geochemical sampling over greenstone or gneiss terrains. The background populations should be checked for atypical observations, outliers, anomalous values, or any trends that do not represent typical regional values. Some of these atypical values could represent mineralized zones and should be closely scrutinized. Target populations may represent samples collected over mineralized zones and associated alteration haloes. The geochemical characteristics of such a population would be distinctly different from the regional background geochemistry and thus a statistical distinction can be made between the two populations.

3.0 EXPLORATORY DATA ANALYSIS

Preliminary data analysis involves the examination of the frequency distributions, univariate statistics (parametric, non-parametric, and robust methods), bivariate plots, correlation coefficients, plus other numerical or graphical methods that assist in understanding the nature of the data. **It is at this stage that outliers should be eliminated. Failure to remove or compensate for outliers can have profound effects on any statistical interpretation of data.**

3.1 Univariate Methods of Exploratory Data Analysis

Univariate methods involve the analysis of a single element, usually within a single population of samples. The element may be the commodity being sought or it may be a pathfinder element. The primary goal is to split the population into groups of samples that represent geochemically-distinct domains from which samples can be isolated that may be related to mineralization.

Prior to any analysis, the removal of gross outliers is essential. These outliers may be related to areas of economic interest and should be followed up. If the outliers are not deleted from the sample population, then the results of most parametric statistical methods will be distorted.

One way of minimizing the effect of outliers is to use robust methods of parameter estimates. Robust methods down-weight the significance of outliers when statistical parameters are calculated. Rock (1987) provides a review and a computer program for analyzing univariate populations.

The use of exploratory data analysis (EDA) techniques can assist in recognizing patterns within geochemical data. The simplicity of approach and interpretation makes EDA a logical choice for the initial investigation of the data. EDA encompasses the use of order statistics, histograms, box and whisker plots, stem and leaf plots, and summary statistics. EDA methods have been routinely applied in the analysis of regional laterite geochemical data in Western Australia. Kürzl (1988) discusses the success of applying exploratory data analysis techniques in regional exploration programmes.

3.2 Interpretation of the Histogram, Box and Whisker, and Q-Q Plots

These three plots when used together can reveal a significant amount of information regarding the nature of the frequency distribution of a sample population. The histogram is useful as it estimates a graphical display of the shape of the distribution. Caution must be used in the construction of the histogram because data are binned into intervals which affect the display of the data. The box and whisker plots provide a quick visual summary of the order statistics for the sample population and the Q-Q plot provides a measure of how "normal" a distribution might be and the presence of individual outliers.

3.3 Scatterplot Matrix

A useful visual summary of geochemical data is the crosscorrelation plot. These plots are illustrated in Figures 12 and 13. The horizontal axis along the top provides a histogram of each element. The diagonal box provides a box-whisker summary of the data distribution and each non-diagonal element is a plot of each element against every other element. The main advantage of these plots is that they provide a **quick** summary of the distributions of the elements as well as their associations. Figure 12 shows the untransformed data plotted on a scatterplot matrix. Figure 13 shows the transformed data plotted on a scatterplot matrix.

3.4 Modelled Univariate Methods: Dissection of Mixtures of Populations and Separating Anomalous Samples

Bridges and McCammon (1980) developed a computer program for distinguishing anomalous populations from background populations based on the analysis of inflection points on cumulative frequency distribution curves. A combination of graphics and non-linear least-squares fitting routines are used to maximize the distinctions between the sub-populations.

Campbell (1986a) developed a procedure for unmixing populations based on a measure of typicality that each observation (sample) belongs to either population. The procedure requires that the component distributions are normal. If the data are not normally distributed, then the appropriate Box-Cox power transformation can be used to refine the resolution of samples into their respective component distributions.

Alternatively, Stanley and Sinclair (1987) have developed an approach whereby the multiple populations can be dissected by a background characterization approach. This procedure involves:

a) grouping data into subsets according to geological and geochemical characteristics and treating each subset independently,

b) choosing a dependent variable (i.e. the commodity element or suitable pathfinder element) for each subset,

c) selecting a threshold for each dependent variable using probability plots to unambiguously define a group of background samples characterized by concentrations lower than this threshold,

d) develop a multiple regression background model for the selected dependent variable(s) using only samples from the background population,
e) applying this background model to samples whose abundances are greater than the selected threshold to test the likelihood that these samples may in fact be part of the background population.

Implicit in this approach is the use of probability plots (or Q-Q plots) that have been suitably transformed to assist in choosing suitable thresholds for the elements that will be studied. Stanley (1987) has published a computer program, PROBPLOT, that assists in dissecting mixed univariate populations into a preselected number of populations based on inflection points. The program then chooses appropriate thresholds for each of the subpopulations

Gap statistic

A statistical procedure for estimating the threshold of distributions has been described by Miesch (1981). The method, known as the gap statistic, estimates the threshold value of a distribution from which all values above

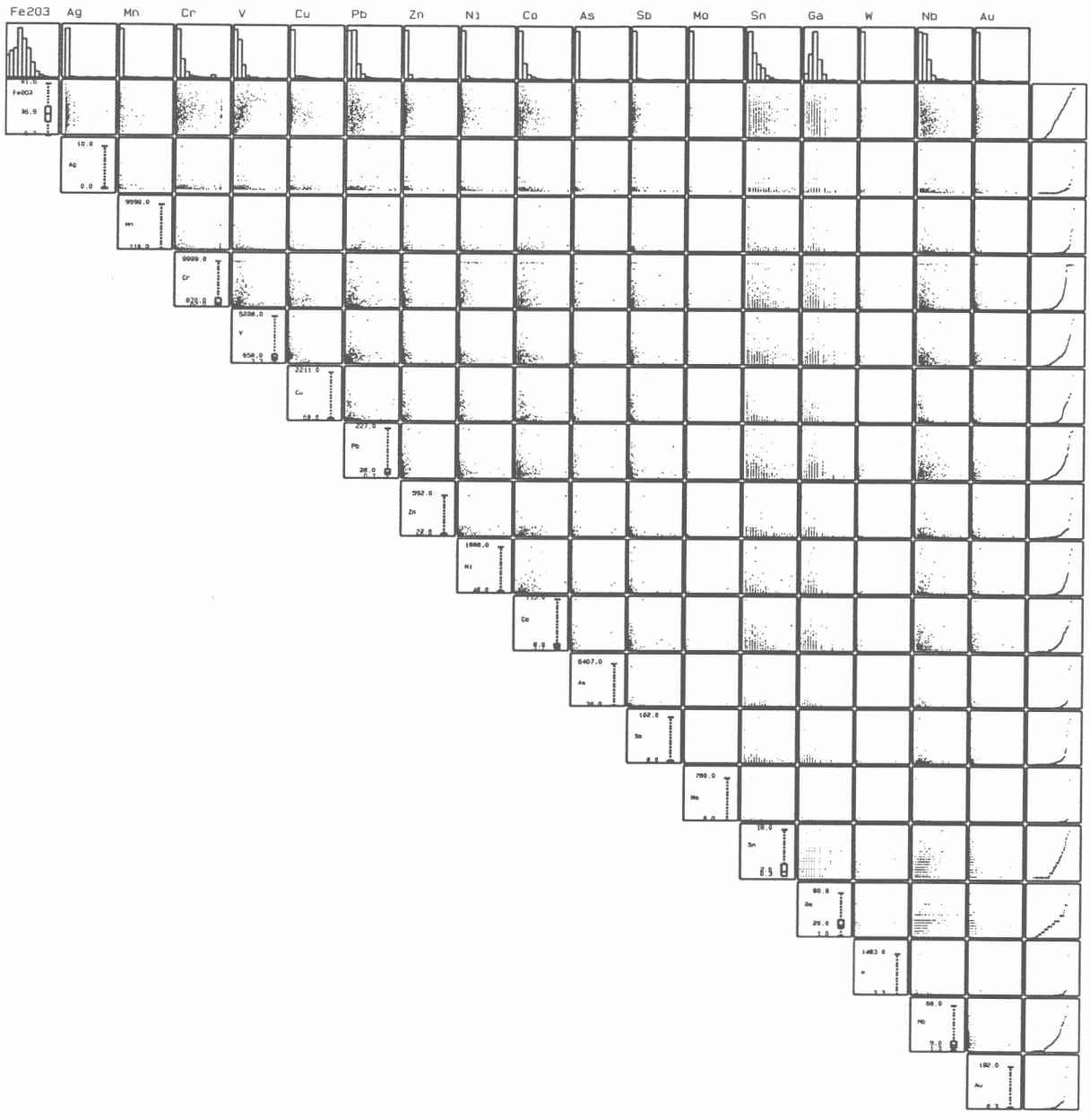


Figure 12. Scatterplot matrix of untransformed data from the Murchison Greenstone Terrain.

the threshold can be treated as outliers. The method also determines whether the threshold is realistic by estimating the probability that the outliers belong to the distribution below the threshold. A computer program has been written for this (Miesch, 1981).

The use of the gap statistic procedure is a useful adjunct to any programme of data evaluation. However, the program does not always provide a reliable estimate of distribution gaps and thus it should not be used without other methods to assess thresholds of distributions. Stanley and Sinclair (1989) compared the use of the gap statistic to the use of probability plots and found in several cases that the gap statistic was unreliable. Nonetheless, it can be used as an additional tool in the search for a threshold.

An example of the use of the gap statistic was applied to 868 As values from laterites over the Murchison greenstone terrain. Figures 4 and 9 show a histogram and Q-Q plot of the As distribution for the untransformed data. The application of the program (Miesch, 1981) resulted in a logarithmic transformation of the data from which the most significant gap in the resulting near-normal distribution occurs between 123-132 ppm. The gap statistic is 0.0117 which means that for approximately 1% of the time the gap may be due to random chance. Thus, the gap of 123-132 ppm can be considered to be meaningful. The gap statistic procedure also indicated that 91 values were found above the 132 ppm gap. These samples are considered to be separate from the main population of the background data and could be considered anomalous.

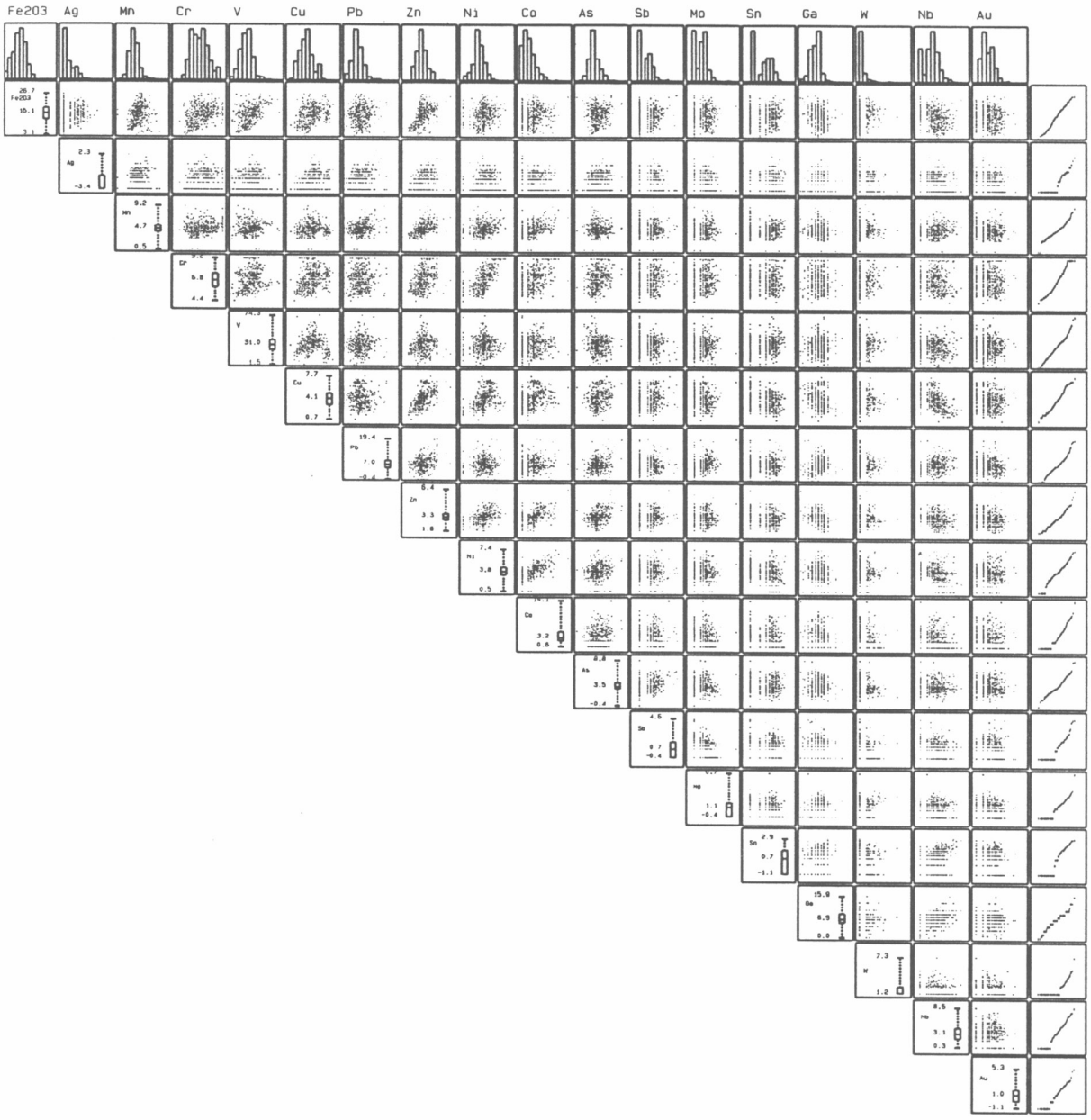


Figure 13. Scatterplot matrix of transformed data from the Murchison Greenstone Terrain.

4.0 MULTIVARIATE EXPLORATORY TECHNIQUES

The usefulness of multivariate data analysis methods applied to geochemical data has been well documented (Howarth and Sinding-Larsen, 1983, Chapter 6). Methods such as principal components analysis, cluster analysis, multidimensional scaling, and projection pursuit provide a numerical and graphical means through which the relationships of the elements and samples can be studied. These techniques are dimension-reducing, as they reduce the number of dimensions required to describe the variation of the data. An interpretation of the systematic relationships of 30 elements is almost impossible without applying some form of dimension-reducing technique. The outcome of these dimension-reducing techniques usually provides significantly fewer, "new variables", that describe variation and can be tied to specific geochemical/geological processes. The use of robust estimates for the correlation coefficient or covariance matrix assists significantly in assessing multi-element relationships. Otherwise, the presence of outliers can distort the resulting linear relationships that are obtained in methods such as principal components analysis. There are several good reviews that discuss the basics of multivariate data analysis techniques (e.g. Jöreskog *et al.*, 1976; Davis, 1986; Howarth and Sinding-Larsen, 1983). Mellinger (1987) provides a systematic approach to the application of multivariate methods in geological studies. Other methods include non-linear mapping (Sammon, 1969), and multidimensional scaling (Kruskal, 1964).

Multivariate techniques have been specifically applied to Archaean volcanic terrains from which a number of geological processes can be inferred, ranging from primary compositional variation to alteration and associated mineralization (Grunsky, 1986). Multivariate techniques also include empirical techniques such as the chalcophile and pegmatophile indices developed by Smith and Perdrix (1983).

Mineral deposits are usually associated with multi-element geochemical signatures that are atypical relative to regional background variation through depleted or increased abundances of a particular suite of elements. The diagnostic elements are dependent upon the type of mineral deposit. Experience in the exploration for Au deposits in lateritic environments has indicated the chalcophile elements, As, Sb, Bi, Mo, Ag, W, Sn, and Ta are commonly associated with Au deposits. The usefulness of choosing appropriate elements is outlined in Smith *et al.* (1984).

4.1 Robust Estimation of Mean and Covariance Matrices

In multivariate analysis, robust estimates of means, correlations, and covariances are determined from the Mahalanobis distances of samples (distance to multivariate mean). In most cases, a sample population will contain outliers and if these outliers are included in the calculations of the means and covariances, then the statistics will be distorted. Several procedures exist for reducing the influence of outliers. The benefit of applying robust estimates is that the application of statistical procedures will not be influenced by the presence of outliers.

One method of reducing the influence of outliers in the calculations of means, covariances, and correlations is through the use **W-Estimators**. This method reduces the influence of outliers in the distribution by using weights. Outliers are given lower weights in the procedure thereby reducing their effect in the estimates of the mean and variance. Campbell (1980) developed the use of W-estimators for the reduction of the influence of outliers in multivariate datasets. This procedure is used in the program WTHGRP (Campbell,1986a). The mean x_M and variance S_M is defined as:

$$\underset{s=1}{\bar{x}_M} = \{\sum_{s=1}^{n} w_S x_S\} / \{\sum^{n} w_S\}$$

$$S_M = \sum_{s=1}^{n} w_S^2 (x_S - \bar{x}_M)(x_S - \bar{x}_M)' / (\sum_{s=1}^{n} w_S^2 - 1)$$

where w_S is the weight attached to the sth sample.

The robust estimators give full weight to observations assumed to come from the main body of the data, but reduced weight or influence otherwise. The weight given to each observation in the calculation of the w-estimates is a function of the Mahalanobis distance. In practice, the influence of observations with unduly large Mahalanobis distances is reduced. Further details are given in Campbell (1980). A more readable description of the procedure can be found in Krzanowski, (1988:231-2).

Table 5 shows estimates of the mean and standard deviation for 484 laterite samples for which Au was determined from the Murchison Greenstone terrain (Dataset 1). The first set shows standard estimates and the second set shows the robust estimates. Comparison of the two sets clearly shows a difference between the means and standard deviations.

4.2 Principal Components Analysis

The objective of principal components analysis (PCA) is to reduce the number of variables necessary to describe the observed variation within a set of data. This is done by forming linear combinations of the variables (components) that describe the distribution of the data. Ideally, to the geologist, each component might be interpreted as describing a geological process, such as differentiation (partial melting, crystal fractionation, etc.), alteration/mineralization (carbonatization, silicification, alkali depletion, metal associations and enrichments, etc.), and weathering processes (bedrock-saprolite-laterite).

Components analysis methods are based on the extraction of linear combinations of samples and elements from some measure of association. Most commonly, the measures of association used are the inter-element correlations or covariances. Davis (1986, Chapter 6) gives a very readable account on the mathematics of principal components analysis.

Covariance relationships between the elements reflect the magnitudes of the elements and thus elements with large abundances tend to dominate the variance-covariance matrix. This has the effect of increasing the significance of elements with large magnitudes in the results of the principal components analysis.

The **correlation** matrix represents the inter-element correlations, it is actually the standardized equivalent of the variance-covariance matrix. When the correlation matrix is used, then all elements have equal representation and the linear combinations of the elements are based on their correlations only and not on the magnitudes of the elements. Other metrics of association can be used and they are discussed by Davis (1986), and Jöreskog *et al.* (1976).

Robust estimation of either covariance or correlation gives a better estimate of the means of the variables by down-weighting the influence of anomalous samples (Zhou, 1985, 1989).

A method of principal components analysis, known as simultaneous RQ-mode principal components analysis (Zhou *et al.,* 1983), has the advantage of presenting the component scores of the samples and the elements in the same scaled component space. Thus, scatter plots of component scores show the relationships of the samples with respect to each other and the elements with respect to each other.

The interpretation of the results of principal components is usually focused on placing a geological/geochemical interpretation on the linear combinations of elements that comprise the components. In regional geochemical studies within the Yilgarn Block, the first and second components usually reflect underlying lithologies. The less significant components can be related to processes of interest (i.e. mineralization) that are manifested by distinct linear combinations of chalcophile elements.

In multivariate ranking (χ^2 plots), the choice of a suitable "multivariate threshold" may be difficult. Initially, the application of methods such as principal components analysis can reveal relationships of the data that may be directly related to underlying lithologies. Princpial component analysis can be useful in choosing suitable pathfinder elements for further follow-up.

The following example illustrates the use of PCA as a tool for investigating the systematics of a regional geochemical dataset. The sampling was carried out using a 3-km triangular grid over the Murchison greenstone belt in the Yilgarn Block of Western Australia.

1) The data were analyzed by producing histograms, Q-Q plots, and summary statistics tables (see Tables 2-4). The data were scrutinized for outliers and atypical values. Atypical values were closely checked to verify that laboratory errors were not the cause.

2) The data were then transformed using the power transformation method outlined by Howarth and Earle (1979) (see Table 1). The transformations were derived from values that ranked less than the 95th percentile. This ensured minimal influence from outliers. The application of the transformations to each of the element ensures a valid estimate of means and covariances/correlations.

3) Robust estimates of means and covariance were computed for a selected group of elements considered important for exploration. Table 5 displays the robust estimates.

4) Principal components analyses were carried out as follows:

- PCA with untransformed data and non-robust estimates of means and correlations (Table 6),
- PCA with untransformed data and robust estimates of means and correlations (Table 8),
- PCA with transformed data and non-robust estimates of means and correlations (Table 7),
- PCA with transformed data and robust estimates of means and correlations (Table 9).

Comparison of the correlation coefficients between the Tables shows that the correlations change significantly depending upon the use of transformed/untransformed and the use of robust/non-robust estimates.

Outliers are more readily recognizable through robust estimates and data transformations. Figures 14a,15a,16a, and 17a show the variable and sample loadings plotted against the first two component axes.

Figure 14a is a plot of the variables and samples along the Component 1 (C1) and Componenet 2 (C2) axes for untransformed data using non-robust estimates of the means and correlations. The plot shows significant dispersion of samples along the positive sides of the C1 and C2 axes. These outliers merely represent the non-normal distribution of the data and do not necessarily reflect atypical values. The Figure indicates that

Table 6
Murchison Greenstone Laterites [Non-robust][Non-transformed]

Observations: 484

Var.	Mean	Std.Dev.	Coef.Var.
Fe	37.1756	15.3279	41.23
Ag	0.2754	0.6386	231.87
Mn	236.4966	631.5234	267.03
Cr	1638.3368	2140.3381	130.64
V	782.2824	619.2245	79.16
Cu	154.0868	261.8925	169.96
Pb	35.2789	29.3328	83.15
Zn	36.7066	40.1886	109.49
Ni	90.9277	154.3427	169.74
Co	11.6074	12.0967	104.22
As	112.3237	418.1071	372.23
Sb	4.1687	8.3959	201.40
Mo	6.5475	35.7497	546.00
Sn	2.8939	3.0562	105.61
Ga	22.0909	9.6960	43.89
W	12.3430	69.3067	561.51
Nb	11.2700	9.4094	83.49
Au	5.8147	12.6156	216.96

Correlation Matrix

	Fe	Ag	Mn	Cr	V	Cu	Pb	Zn	Ni
Fe	1.0000	0.0309	-0.0402	0.2330	0.4710	0.1963	0.0675	0.3115	0.1341
Ag	0.0309	1.0000	0.0302	0.0308	0.0542	0.0877	0.2651	0.1264	0.0039
Mn	-0.0402	0.0302	1.0000	0.0896	-0.0380	-0.0600	0.3491	0.1592	0.1394
Cr	0.2330	0.0308	0.0896	1.0000	0.0719	0.0088	-0.0032	0.1320	0.6993
V	0.4710	0.0542	-0.0380	0.0719	1.0000	-0.2010	0.0524	-0.0052	-0.0543
Cu	0.1963	0.0877	-0.0600	0.0088	-0.2010	1.0000	0.0150	0.2484	0.0866
Pb	0.0675	0.2651	0.3491	-0.0032	0.0524	0.0150	1.0000	0.2964	0.0226
Zn	0.3115	0.1264	0.1592	0.1320	-0.0052	0.2484	0.2964	1.0000	0.2567
Ni	0.1341	0.0039	0.1394	0.6993	-0.0543	0.0866	0.0226	0.2567	1.0000
Co	0.1858	0.0175	0.3443	0.4076	0.0435	0.1687	0.0561	0.5198	0.5585
As	0.1500	0.0577	-0.0175	0.0835	0.0107	0.1948	0.2905	0.2072	0.1383
Sb	0.1623	0.0195	-0.0007	0.0846	-0.0404	0.1329	-0.0371	0.1235	0.0372
Mo	0.0539	-0.0123	-0.0136	0.0055	-0.0377	0.1208	0.3267	0.1033	0.0235
Sn	0.0187	0.1211	-0.0247	0.0594	0.1710	-0.2408	0.0572	-0.0217	-0.0189
Ga	-0.2809	0.0547	-0.0348	-0.0978	0.0604	-0.1234	0.0592	-0.2612	-0.0928
W	0.0298	-0.0179	-0.0113	0.0889	0.0742	-0.0326	-0.0248	-0.0185	-0.0098
Nb	-0.1810	-0.0068	0.0507	-0.0001	-0.0205	-0.3008	0.1694	-0.1123	-0.0403
Au	-0.0448	-0.0334	0.0126	0.0127	-0.0157	-0.0592	-0.0559	0.0247	0.0581

	Co	As	Sb	Mo	Sn	Ga	W	Nb	Au
Fe	0.1858	0.1500	0.1623	0.0539	0.0187	-0.2809	0.0298	-0.1810	-0.0448
Ag	0.0175	0.0577	0.0195	-0.0123	0.1211	0.0547	-0.0179	-0.0068	-0.0334
Mn	0.3443	-0.0175	-0.0007	-0.0136	-0.0247	-0.0348	-0.0113	0.0507	0.0126
Cr	0.4076	0.0835	0.0846	0.0055	0.0594	-0.0978	0.0889	-0.0001	0.0127
V	0.0435	0.0107	-0.0404	-0.0377	0.1710	0.0604	0.0742	-0.0205	-0.0157
Cu	0.1687	0.1948	0.1329	0.1208	-0.2408	-0.1234	-0.0326	-0.3008	-0.0592
Pb	0.0561	0.2905	-0.0371	0.3267	0.0572	0.0592	-0.0248	0.1694	-0.0559
Zn	0.5198	0.2072	0.1235	0.1033	-0.0217	-0.2612	-0.0185	-0.1123	0.0247
Ni	0.5585	0.1383	0.0372	0.0235	-0.0189	-0.0928	-0.0098	-0.0403	0.0581
Co	1.0000	0.1721	0.0928	0.0035	-0.0752	-0.1651	0.0231	-0.1419	0.0177
As	0.1721	1.0000	0.1033	0.0238	0.0304	-0.1632	-0.0192	-0.0544	-0.0020
Sb	0.0928	0.1033	1.0000	-0.0307	0.0103	-0.1266	-0.0017	-0.0653	0.0368
Mo	0.0035	0.0238	-0.0307	1.0000	-0.0061	-0.0653	0.0340	-0.0259	-0.0230
Sn	-0.0752	0.0304	0.0103	-0.0061	1.0000	0.2288	0.0591	0.5806	0.0026
Ga	-0.1651	-0.1632	-0.1266	-0.0653	0.2288	1.0000	0.0113	0.4132	-0.0569
W	0.0231	-0.0192	-0.0017	0.0340	0.0591	0.0113	1.0000	0.0008	-0.0061
Nb	-0.1419	-0.0544	-0.0653	-0.0259	0.5806	0.4132	0.0008	1.0000	-0.0243
Au	0.0177	-0.0020	0.0368	-0.0230	0.0026	-0.0569	-0.0061	-0.0243	1.0000

	Eigenvalues	% Trace	S Trace
1	2.9413	16.3404	16.3404
2	2.0201	11.2226	27.5631
3	1.6434	9.1300	36.6931
4	1.5296	8.4975	45.1906
5	1.1642	6.4677	51.6584
6	1.0927	6.0706	57.7290
7	1.0438	5.7989	63.5279
8	0.9678	5.3767	68.9046
9	0.9475	5.2638	74.1684
10	0.8832	4.9065	79.0750
11	0.8138	4.5214	83.5963
12	0.7652	4.2514	87.8477
13	0.5919	3.2885	91.1362
14	0.5098	2.8321	93.9683
15	0.3149	1.7493	95.7176
16	0.2998	1.6656	97.3832
17	0.2518	1.3990	98.7822
18	0.2192	1.2178	100.0000

Principal Components R-Mode Loadings

	1	2	3	4	5	6	7	8
Fe	0.5063	-0.0064	0.0279	-0.7028	0.0694	-0.0094	-0.0159	-0.0935
Ag	0.1138	0.2323	0.3645	-0.0697	-0.1730	-0.1087	-0.4492	0.1995
Mn	0.2781	0.3111	0.1287	0.4032	0.4815	-0.3491	-0.0446	0.2270
Cr	0.5599	0.3098	-0.5212	0.0324	-0.0776	0.2604	-0.0424	-0.0780
V	0.0685	0.2515	-0.0682	-0.7690	0.3588	-0.0854	-0.1028	-0.0988
Cu	0.4016	-0.3943	0.2843	0.0712	-0.3681	0.2136	-0.1686	0.0755
Pb	0.2429	0.4442	0.6874	0.1194	0.1814	0.0228	0.0679	-0.0355
Zn	0.6739	0.0983	0.2641	0.0142	-0.0240	-0.1625	0.0413	0.0547
Ni	0.6463	0.2604	-0.4633	0.2524	-0.0882	0.1976	-0.0487	-0.1645
Co	0.7473	0.1737	-0.2108	0.2058	0.1039	-0.0931	-0.0903	0.0756
As	0.3858	0.0714	0.3175	-0.0827	-0.3562	-0.0395	0.1270	-0.1493
Sb	0.2405	-0.1016	-0.0066	-0.1465	-0.4951	-0.2995	0.1928	0.3424
Mo	0.1445	0.0544	0.4335	0.0531	0.1381	0.5714	0.4258	-0.1909
Sn	-0.2124	0.7005	-0.0400	-0.2430	-0.3114	-0.0479	0.1394	0.0056
Ga	-0.4624	0.4401	-0.0263	0.1281	-0.1084	0.1458	-0.2629	0.0346
W	0.0118	0.0859	-0.1292	-0.1726	0.1417	0.3868	0.2603	0.7875
Nb	-0.3722	0.7339	-0.0124	0.0833	-0.2351	-0.0079	0.1179	-0.0446
Au	0.0271	-0.0305	-0.1633	0.0826	-0.0035	-0.4154	0.6143	-0.0989

Relative Contributions: Variables

	1	2	3	4	5	6	7	8
Fe	25.6338	0.0040	0.0780	49.3984	0.4815	0.0089	0.0253	0.8741
Ag	1.2953	5.3947	13.2869	0.4858	2.9944	1.1824	20.1755	3.9798
Mn	7.7332	9.6791	1.6569	16.2577	23.1853	12.1876	0.1990	5.1546
Cr	31.3460	9.5992	27.1697	0.1050	0.6025	6.7825	0.1797	0.6090
V	0.4692	6.3236	0.4654	59.1360	12.8767	0.7298	1.0576	0.9755
Cu	16.1294	15.5466	8.0825	0.5064	13.5500	4.5638	2.8424	0.5697
Pb	5.8998	19.7326	47.2574	1.4248	3.2902	0.0520	0.4611	0.1258
Zn	45.4177	0.9656	6.9771	0.0202	0.0578	2.6413	0.1706	0.2997
Ni	41.7671	6.7807	21.4643	6.3724	0.7788	3.9045	0.2372	2.7059
Co	55.8494	3.0167	4.4432	4.2341	1.0804	0.8667	0.8147	0.5716
As	14.8808	0.5102	10.0820	0.6841	12.6895	0.1560	1.6138	2.2296
Sb	5.7828	1.0316	0.0043	2.1470	24.5143	8.9717	3.7168	11.7208
Mo	2.0885	0.2963	18.7925	0.2822	1.9071	32.6498	18.1281	3.6427
Sn	4.5122	49.0725	0.1601	5.9041	9.6978	0.2297	1.9424	0.0032
Ga	21.3785	19.3654	0.0689	1.6410	1.1756	2.1245	6.9102	0.1200
W	0.0139	0.7384	1.6685	2.9800	2.0069	14.9577	6.7771	62.0224
Nb	13.8570	53.8573	0.0154	0.6933	5.5293	0.0062	1.3894	0.1992
Au	0.0733	0.0930	2.6674	0.6831	0.0013	17.2562	37.7389	0.9776

Table 7
Murchison Greenstone Laterites [Non-robust][Transformed]

Observations: 484

Var.	Mean	Std.Dev.	Coef.Var.
Fe	14.8613	4.7093	31.69
Ag	-2.2775	1.3167	-57.82
Mn	4.7398	1.0561	22.28
Cr	6.8196	1.0499	15.39
V	30.9735	10.7038	34.56
Cu	4.1119	1.3349	32.47
Pb	7.2142	3.0260	41.94
Zn	3.3482	0.6538	19.53
Ni	3.8707	1.1394	29.44
Co	3.5139	2.3275	66.24
As	3.7627	1.1688	31.06
Sb	0.7469	1.0991	147.15
Mo	1.0347	1.1290	109.12
Sn	0.3869	1.2654	327.08
Ga	7.1584	2.1186	29.60
W	1.6712	0.8849	52.95
Nb	3.0518	1.7433	57.12
Au	0.9458	1.2708	134.37

Correlation Matrix

	Fe	Ag	Mn	Cr	V	Cu	Pb	Zn	Ni
Fe	1.0000	0.1233	0.2594	0.3908	0.4855	0.3683	0.0626	0.5127	0.2338
Ag	0.1233	1.0000	0.1157	0.2635	0.1793	0.4132	0.1583	0.3430	0.2528
Mn	0.2594	0.1157	1.0000	0.1529	0.1516	0.0629	0.2269	0.4659	0.1827
Cr	0.3908	0.2635	0.1529	1.0000	0.2592	0.3305	0.0035	0.3757	0.6403
V	0.4855	0.1793	0.1516	0.2592	1.0000	0.0182	0.0538	0.1292	0.0474
Cu	0.3683	0.4132	0.0629	0.3305	0.0182	1.0000	-0.0545	0.5700	0.4719
Pb	0.0626	0.1583	0.2269	0.0035	0.0538	-0.0545	1.0000	0.2203	0.0473
Zn	0.5127	0.3430	0.4659	0.3757	0.1292	0.5700	0.2203	1.0000	0.4255
Ni	0.2338	0.2528	0.1827	0.6403	0.0474	0.4719	0.0473	0.4255	1.0000
Co	0.2022	0.1762	0.3859	0.3898	0.0726	0.3995	0.0243	0.5040	0.6319
As	0.0958	0.2174	0.0139	0.2898	-0.0410	0.2256	0.1817	0.3383	0.2132
Sb	0.1033	0.1545	0.0473	0.1946	0.0102	0.1915	-0.0533	0.2016	0.1401
Mo	0.0206	-0.0633	-0.0253	0.0251	-0.0257	-0.1745	0.4053	-0.0457	0.0319
Sn	-0.0525	0.0939	0.0384	-0.0318	0.1948	-0.2745	0.0386	-0.0737	-0.2590
Ga	-0.3077	0.1355	-0.1619	-0.1241	0.0540	-0.3273	0.1638	-0.3535	-0.1898
W	0.2362	0.0343	0.0944	0.1530	0.1595	-0.0269	0.0558	0.0411	-0.0030
Nb	-0.2889	-0.0318	0.1098	-0.1718	-0.0680	-0.5229	0.2104	-0.2226	-0.3379
Au	-0.1059	-0.1676	-0.0207	-0.0666	-0.0194	-0.1734	-0.0766	-0.0938	-0.0795

	Co	As	Sb	Mo	Sn	Ga	W	Nb	Au
Fe	0.2022	0.0958	0.1033	0.0206	-0.0525	-0.3077	0.2362	-0.2889	-0.1059
Ag	0.1762	0.2174	0.1545	-0.0633	0.0939	0.1355	0.0343	-0.0318	-0.1676
Mn	0.3859	0.0139	0.0473	-0.0253	0.0384	-0.1619	0.0944	0.1098	-0.0207
Cr	0.3898	0.2898	0.1946	0.0251	-0.0318	-0.1241	0.1530	-0.1718	-0.0666
V	0.0726	-0.0410	0.0102	-0.0257	0.1948	0.0540	0.1595	-0.0680	-0.0194
Cu	0.3995	0.2256	0.1915	-0.1745	-0.2745	-0.3273	-0.0269	-0.5229	-0.1734
Pb	0.0243	0.1817	-0.0533	0.4053	0.0386	0.1638	0.0558	0.2104	-0.0766
Zn	0.5040	0.3383	0.2016	-0.0457	-0.0737	-0.3535	0.0411	-0.2226	-0.0938
Ni	0.6319	0.2132	0.1401	0.0319	-0.2590	-0.1898	-0.0030	-0.3379	-0.0795
Co	1.0000	0.2140	0.1331	-0.0607	-0.1367	-0.1814	0.0013	-0.1997	-0.0900
As	0.2140	1.0000	0.3655	-0.0826	0.0766	-0.1526	-0.1585	-0.0354	-0.0166
Sb	0.1331	0.3655	1.0000	-0.2254	0.0548	-0.0734	-0.0749	-0.0458	0.0011
Mo	-0.0607	-0.0826	-0.2254	1.0000	-0.0355	0.1442	0.2996	0.1010	-0.0354
Sn	-0.1367	0.0766	0.0548	-0.0355	1.0000	0.1683	0.0423	0.4837	0.1164
Ga	-0.1814	-0.1526	-0.0734	0.1442	0.1683	1.0000	0.0121	0.4133	-0.0499
W	0.0013	-0.1585	-0.0749	0.2996	0.0423	0.0121	1.0000	0.0213	-0.0077
Nb	-0.1997	-0.0354	-0.0458	0.1010	0.4837	0.4133	0.0213	1.0000	0.0847
Au	-0.0900	-0.0166	0.0011	-0.0354	0.1164	-0.0499	-0.0077	0.0847	1.0000

	Eigenvalues	% Trace	S Trace
1	4.1677	23.1537	23.1537
2	2.0710	11.5054	34.6591
3	1.6766	9.3143	43.9734
4	1.4573	8.0961	52.0696
5	1.2668	7.0377	59.1073
6	1.0954	6.0858	65.1930
7	1.0647	5.9151	71.1082
8	0.8355	4.6415	75.7496
9	0.8150	4.5278	80.2775
10	0.7156	3.9756	84.2531
11	0.5055	2.8082	87.0613
12	0.4896	2.7201	89.7814
13	0.4457	2.4759	92.2573
14	0.3886	2.1589	94.4161
15	0.3163	1.7574	96.1735
16	0.2509	1.3936	97.5671
17	0.2351	1.3060	98.8731
18	0.2028	1.1269	100.0000

Principal Components R-Mode Loadings

	1	2	3	4	5	6	7	8
Fe	0.6038	0.2051	-0.3276	0.4400	0.0287	0.1796	0.2031	-0.0369
Ag	0.4237	0.2918	0.2954	-0.1035	0.4779	0.1955	-0.1459	-0.1353
Mn	0.3823	0.3990	-0.0256	0.0688	-0.5653	0.1956	-0.3770	0.1210
Cr	0.6595	0.1863	0.0056	0.0492	0.1958	-0.4933	0.0199	0.0249
V	0.2424	0.4101	-0.1833	0.5893	0.2771	0.0240	-0.0656	-0.3123
Cu	0.7489	-0.2787	0.0114	-0.0756	0.2049	0.1991	0.0187	-0.0543
Pb	0.0749	0.6071	-0.0308	-0.4747	-0.1424	0.2670	0.2557	-0.2412
Zn	0.7927	0.1773	0.0611	-0.0142	-0.2469	0.2497	0.0354	-0.0185
Ni	0.7338	-0.0482	-0.0370	-0.2837	0.0433	-0.4320	-0.1801	-0.0703
Co	0.6765	0.0360	0.0536	-0.1853	-0.2258	-0.2367	-0.3829	0.0587
As	0.4002	0.0683	0.5614	-0.1538	-0.0404	-0.0312	0.4834	-0.0203
Sb	0.3080	-0.0532	0.5544	0.1448	0.0560	-0.0795	0.2971	0.3754
Mo	-0.1099	0.4402	-0.4823	-0.4542	-0.0001	-0.1510	0.3425	-0.0078
Sn	-0.2439	0.5095	0.4080	0.3967	-0.0145	-0.0622	0.0012	0.0388
Ga	-0.4320	0.4015	0.1581	-0.2283	0.4529	-0.1403	-0.2768	-0.0873
W	0.0644	0.3763	-0.5122	0.1594	0.1000	-0.2079	0.1679	0.4873
Nb	-0.4928	0.5788	0.3479	-0.0297	-0.1665	-0.0730	-0.1535	0.1526
Au	-0.1806	-0.0419	0.0931	0.2564	-0.4231	-0.4739	0.2041	-0.4710

Relative Contributions: Variables

	1	2	3	4	5	6	7	8
Fe	36.4526	4.2050	10.7298	19.3598	0.0824	3.2259	4.1254	0.1363
Ag	17.9511	8.5167	8.7242	1.0708	22.8401	3.8209	2.1276	1.8319
Mn	14.6155	15.9177	0.0654	0.4738	31.9555	3.8272	14.2151	1.4642
Cr	43.4953	3.4698	0.0031	0.2422	3.8336	24.3392	0.0395	0.0619
V	5.8774	16.8157	3.3614	34.7302	7.6797	0.0576	0.4308	9.7501
Cu	56.0815	7.7690	0.0129	0.5711	4.1974	3.9652	0.0348	0.2947
Pb	0.5608	36.8597	0.0951	22.5372	2.0280	7.1273	6.5394	5.8182
Zn	62.8415	3.1452	0.3737	0.0200	6.0936	6.2358	0.1250	0.0343
Ni	53.8474	0.2324	0.1368	8.0510	0.1872	18.6632	3.2452	0.4937
Co	45.7649	0.1293	0.2873	3.4326	5.1000	5.6048	14.6581	0.3440
As	16.0129	0.4666	31.5132	2.3649	0.1629	0.0971	23.3689	0.0411
Sb	9.4891	0.2827	30.7385	2.0980	0.3136	0.6319	8.8242	14.0940
Mo	1.2078	19.3785	23.2652	20.6283	0.0000	2.2813	11.7336	0.0061
Sn	5.9474	25.9596	16.6479	15.7355	0.0211	0.3872	0.0002	0.1506
Ga	18.6615	16.1201	2.4995	5.2134	20.5121	1.9674	7.6619	0.7630
W	0.4147	14.1574	26.2310	2.5393	0.9993	4.3235	2.8200	23.7501
Nb	24.2821	33.4966	12.1048	0.0881	2.7735	0.5332	2.3558	2.3274
Au	3.2630	0.1756	0.8677	6.5741	17.8982	22.4554	4.1666	22.1851

samples which plot along the positive side of the C1 axis are associated with laterites that are Co, Zn, Ni, Cr, and Fe enriched relative to samples that plot along the negative side of the C1 axis which are associated with Ga, Nb, and, Sn. The association of the elements along the C2 axis shows Sn is associated with more fractionated rocks containing Nb and Ga, while samples along the negative end of the C2 axis have a greater affinity with Cu and samples associated with Fe, Zn, Ni, Cr, and Co. An interpretation of the diagram suggests that samples which plot along the positive side of the C1 axis are associated with mafic volcanic rocks, while samples that plot along the negative side of the C1 axis are associated with felsic volcanic or fractionated plutonic rocks. Samples that plot close to the origin of the diagram indicate that they do not contribute much to the components which are plotted.

Figure 15a is a plot of component loadings of the variables and sample scores along the C1 and C2 axes for untransformed data using robust estimates of the mean and correlations. This Figure also shows significant dispersion along the C1 and C2 axes. In contrast with Figure 14a, the samples show a greater dispersion. This increased dispersion is due to the robust estimates of the means which tend to enhance the presence of the outliers for non-transformed data. The relationships between the variables are also different from Figure 14a. The results of this analysis suggest that Sb, As, and Ag have a more significant association with the mafic volcanic rocks represented by a relative enrichment in Cr, Co, Zn, Cu, and Ni. The association of W, Mn, V, Pb, and Mo along the C2 axis represents more fractionated rocks.

Figure 16a shows a plot of the variables and samples along the C1 and C2 axes for transformed data using non-robust estimates of the means and correlations. The Figure indicates that transforming the data significantly affects the dispersion of the samples in comparison with Figures 14a, and 15a. The association of the elements indicates that Zn, Ni, Cu, Co, Cr, Fe Ag, As, Sb, and possibly V are associated with lateritic materials derived from mafic volcanics and that Nb, Ga, Sn, and Mo are associated with materials derived from more fractionated rocks. Notice that Au plots very close to the origin of the diagram and has little significance in the first two components. Lead and W show a positive association with the elements associated with the more fractionated materials along the C2 axis, while Cu is associated more with the mafic-associated elements. The reduced dispersion of the diagram suggests that most of the samples represent a homogeneous population with very few outliers represented by the linear combination of elements in the first two components.

Figure 17a shows a plot of the variables and samples along the C1 and C2 axes for transformed data using robust estimates. The Figure is almost identical to Figure 16a and suggests that for this group of data, the effect of transforming the data brings the data distribution to normality and the resulting robust estimates of means and correlations are very similar to that of a non-robust estimate. The similarity of the first two principal components suggests that the elements associated with principal lithological variations are essentially normally distributed. This may not be the case for some of the other chalcophile elements.

One of the objectives of multivariate methods is to assess the associations of target elements. In this example, Au is chosen as the target element. Examination of the component loadings and relative contributions of Table 6 shows that Au makes significant contributions to the 7th component. Figure 14b shows a plot of the C1 versus C7 components. The associations of the samples and elements along the C1 axis show the mafic/felsic pattern as in Figure 14a, but the dispersion of samples along the positive side of the C7 axis indicates that a few samples are associated with Au, Mo, and W. These samples may represent sites that warrant further investigation in the form of more detailed sampling.

Table 7 shows that the seventh component is also the most significant component for Au association. Figure 15b shows a plot of the samples and element scores plotted onto the C1-C7 axes. Samples that plot along the positive side of the C7 axis show relative enrichment in Au and some association with Mo, As, and perhaps Co. This pattern outlines a problem with linear combinations based on untransformed data. It is difficult to judge which samples are truly atypical or anomalous since these apparent outliers may be part of the background population if the data were transformed. Exploration follow-up on these samples might be time consuming and costly.

Table 8 shows the results of the PCA for transformed data but without any robust estimate. Examination of the relative contributions indicates that Au contributes significantly to the C5, C6, and C8 components. Figures 16b,c, and d show plots of these components against the C1 component. Figure 16b shows that Au has an association with Mn and suggests that it is associated with some Mn-enriched lithologies. Figure 16c and Table 8 (Variable loadings) show that Au plots close to the C1 origin, but the association of Ni along the negative part of the C6 axis indicates some association with the Ni-enriched rocks that may be associated with mafic lithologies. Figure 16d shows the elements and sample scores plotted against the C1-C8 axes. The association of Au with Pb and V is suggested by the plot. Note that with the use of transformed data, the dispersion of the samples is not nearly as great as when using non-transformed data (cf. Figures 14, 15). In these plots, the few samples that plot away from the main cloud of points can be considered outliers and anomalous. These points should take first priority in a follow-up investigation.

Table 9 shows the results of the PCA applied to transformed data using robust estimates of the means and correlations. The analysis indicates that Au contributions to the fourth, fifth, sixth, and ninth components.

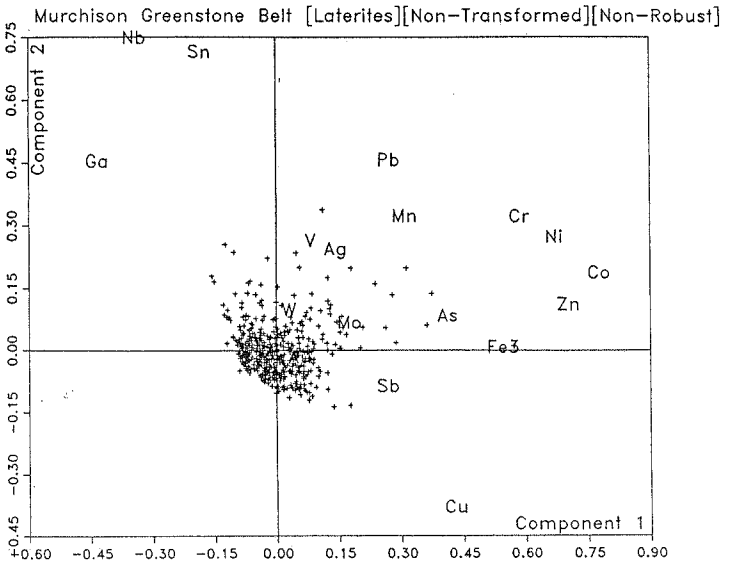


Figure 14a. Principal Component Plot: C1-C2: Murchison Greenstone Belt Laterites [Non-Transformed][Non-Robust]

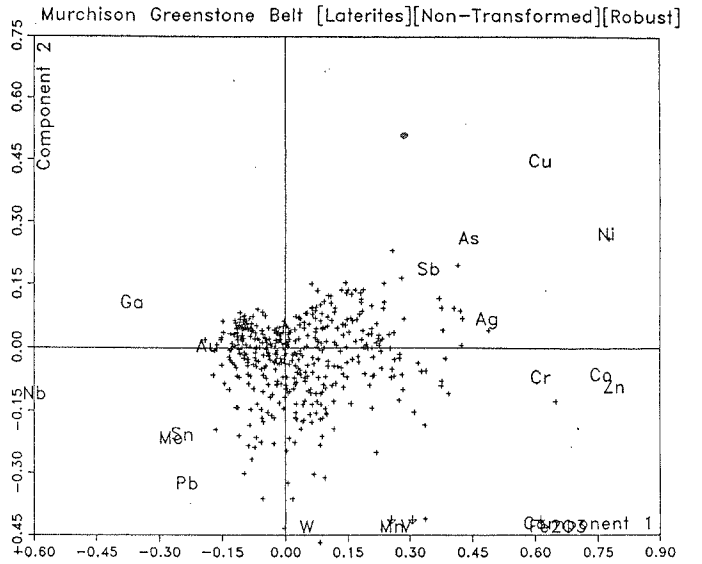


Figure 15a. Principal Component Plot: C1-C2: Murchison Greenstone Belt Laterites [Non-Transformed][Robust]

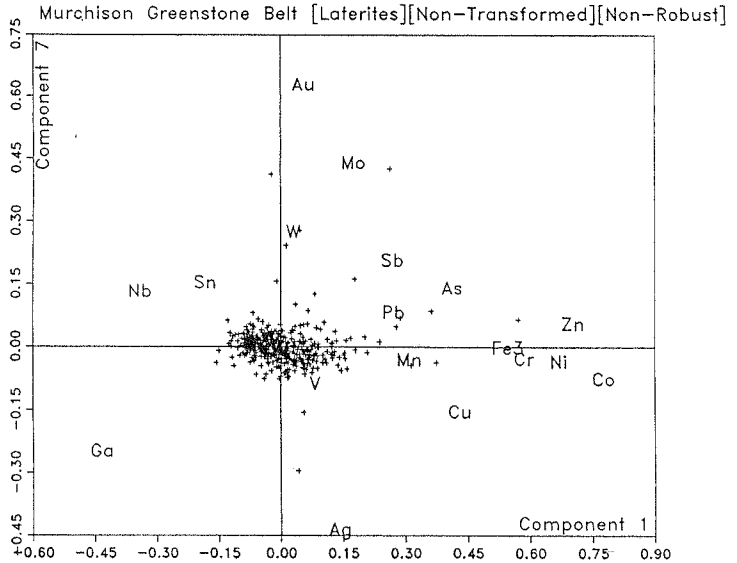


Figure 14b. Principal Component Plot: C1-C7: Murchison Greenstone Belt Laterites [Non-Transformed][Non-Robust]

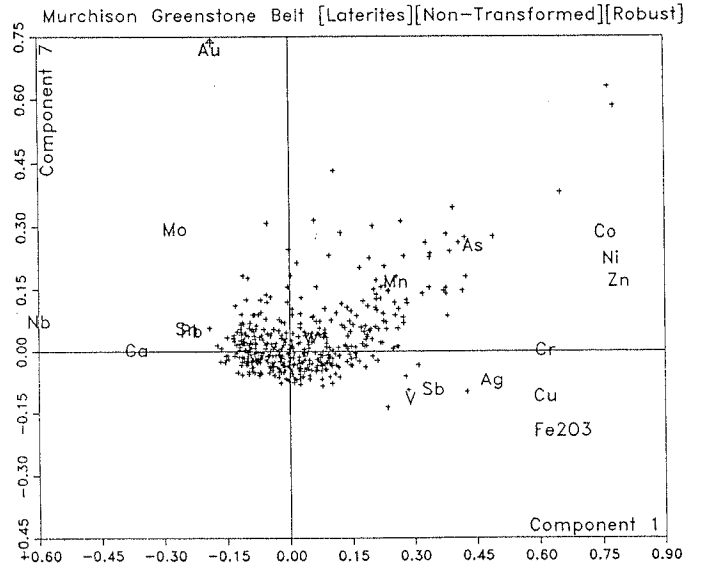


Figure 15b. Principal Component Plot: C1-C7: Murchison Greenstone Belt Laterites [Non-Transformed][Robust]

Table 8
Murchison Greenstone Belt Laterites [Robust][Non-transformed]
Robust Principal Components: Based on Previously Calculated
Robust Estimates of Means, Correlations, and Covariances

Observations: 484

Robust Means

Fe	35.60
Ag	0.19
Mn	127.02
Cr	1078.61
V	748.29
Cu	154.82
Pb	29.82
Zn	28.67
Ni	59.21
Co	8.80
As	54.60
Sb	3.21
Mo	3.86
Sn	2.58
Ga	22.43
W	6.32
Nb	10.46
Au	4.87

Robust Correlation Matrix

	Fe	Ag	Mn	Cr	V	Cu	Pb	Zn	Ni
Fe	1.0000	0.1095	0.3939	0.3751	0.5364	0.2051	0.0193	0.4806	0.1890
Ag	0.1095	1.0000	0.0725	0.2604	0.1016	0.3315	-0.0134	0.3150	0.3783
Mn	0.3939	0.0725	1.0000	0.0652	0.3741	-0.1205	0.1232	0.2632	0.0024
Cr	0.3751	0.2604	0.0652	1.0000	0.2063	0.2169	-0.0579	0.2868	0.5485
V	0.5364	0.1016	0.3741	0.2063	1.0000	-0.2203	0.0496	0.1809	-0.0169
Cu	0.2051	0.3315	-0.1205	0.2169	-0.2203	1.0000	-0.1543	0.3820	0.5071
Pb	0.0193	-0.0134	0.1232	-0.0579	0.0496	-0.1543	1.0000	-0.0950	-0.1179
Zn	0.4806	0.3150	0.2632	0.2868	0.1809	0.3820	-0.0950	1.0000	0.4715
Ni	0.1890	0.3783	0.0024	0.5485	-0.0169	0.5071	-0.1179	0.4715	1.0000
Co	0.3139	0.2892	0.3202	0.2377	0.2362	0.3613	-0.1193	0.5884	0.6250
As	0.0307	0.1584	-0.0936	0.2713	-0.0208	0.1806	-0.0836	0.3518	0.3163
Sb	0.0879	0.1029	-0.0130	0.2058	0.0553	0.1758	-0.1013	0.2156	0.1661
Mo	-0.0537	-0.1150	-0.0222	-0.0994	-0.0564	-0.1631	0.4280	-0.1729	-0.1184
Sn	-0.0760	-0.0330	0.0725	-0.0543	0.1231	-0.2878	-0.0004	-0.1251	-0.2712
Ga	-0.3801	0.1537	-0.1559	-0.1559	-0.0629	-0.0844	0.3369	-0.2990	-0.0881
W	0.2322	0.1155	0.1492	0.1391	0.1321	-0.0950	0.0746	0.0020	-0.0422
Nb	-0.3221	-0.1919	0.0700	-0.2827	-0.1833	-0.3407	0.3131	-0.3327	-0.3955
Au	-0.1565	-0.1568	0.0038	-0.1051	-0.0300	-0.1848	-0.1111	-0.1189	-0.1261

	Co	As	Sb	Mo	Sn	Ga	W	Nb	Au
Fe	0.3139	0.0307	0.0879	-0.0537	-0.0760	-0.3801	0.2322	-0.3221	-0.1565
Ag	0.2892	0.1584	0.1029	-0.1150	-0.0330	0.1537	0.1155	-0.1919	-0.1568
Mn	0.3202	-0.0936	-0.0130	-0.0222	0.0725	-0.1559	0.1492	0.0700	0.0038
Cr	0.2377	0.2713	0.2058	-0.0994	-0.0543	-0.1559	0.1391	-0.2827	-0.1051
V	0.2362	-0.0208	0.0553	-0.0564	0.1231	-0.0629	0.1321	-0.1833	-0.0300
Cu	0.3613	0.1806	0.1758	-0.1631	-0.2878	-0.0844	-0.0950	-0.3407	-0.1848
Pb	-0.1193	-0.0836	-0.1013	0.4280	-0.0004	0.3369	0.0746	0.3131	-0.1111
Zn	0.5884	0.3518	0.2156	-0.1729	-0.1251	-0.2990	0.0020	-0.3327	-0.1189
Ni	0.6250	0.3163	0.1661	-0.1184	-0.2712	-0.0881	-0.0422	-0.3955	-0.1261
Co	1.0000	0.2699	0.0831	-0.1306	-0.1527	-0.1596	-0.0356	-0.3637	-0.1270
As	0.2699	1.0000	0.2727	-0.1353	0.1096	-0.1215	-0.0875	-0.1404	-0.0875
Sb	0.0831	0.2727	1.0000	-0.2130	0.0373	-0.0897	-0.1060	-0.1185	-0.0333
Mo	-0.1306	-0.1353	-0.2130	1.0000	-0.1197	0.2343	0.2010	0.1294	0.0369
Sn	-0.1527	0.1096	0.0373	-0.1197	1.0000	0.1236	0.0147	0.4324	0.0713
Ga	-0.1596	-0.1215	-0.0897	0.2343	0.1236	1.0000	-0.0160	0.3519	-0.0486
W	-0.0356	-0.0875	-0.1060	0.2010	0.0147	-0.0160	1.0000	-0.0401	0.0159
Nb	-0.3637	-0.1404	-0.1185	0.1294	0.4324	0.3519	-0.0401	1.0000	0.0401
Au	-0.1270	-0.0875	-0.0333	0.0369	0.0713	-0.0486	0.0159	0.0401	1.0000

	Eigenvalues	% Trace	S Trace
1	4.1353	22.9739	22.9739
2	2.1216	11.7867	34.7606
3	1.7011	9.4505	44.2111
4	1.4992	8.3291	52.5402
5	1.1021	6.1226	58.6629
6	1.0106	5.6143	64.2771
7	0.9864	5.4801	69.7572
8	0.8583	4.7686	74.5258
9	0.7961	4.4229	78.9487
10	0.7114	3.9520	82.9007
11	0.6040	3.3555	86.2563
12	0.5418	3.0102	89.2665
13	0.4692	2.6067	91.8733
14	0.3952	2.1956	94.0689
15	0.3773	2.0959	96.1647
16	0.3038	1.6876	97.8524
17	0.2015	1.1193	98.9716
18	0.1851	1.0283	100.0000

Principal Component R-Mode Loadings

	1	2	3	4	5	6	7	8
Fe	0.5830	-0.5897	-0.0915	-0.1398	0.0477	0.1467	-0.2073	-0.0316
Ag	0.4563	0.0552	0.3965	0.2573	0.0815	-0.5216	-0.0842	0.0647
Mn	0.2240	-0.6850	-0.0623	0.0904	-0.3789	-0.0798	0.1544	-0.0554
Cr	0.5873	-0.0829	0.1158	0.1282	0.5097	0.0727	-0.0116	0.0757
V	0.2752	-0.7037	-0.1417	0.1162	0.0423	-0.0689	-0.1307	0.4147
Cu	0.5829	0.4333	0.2272	-0.1120	-0.1196	-0.0593	-0.1244	-0.1144
Pb	-0.2632	-0.3391	0.6362	0.1061	-0.0995	0.3776	0.0357	0.0907
Zn	0.7604	-0.1055	0.0078	0.0990	-0.2185	0.1071	0.1554	-0.1621
Ni	0.7480	0.2587	0.2925	0.0170	0.0330	-0.0552	0.2098	0.0400
Co	0.7291	-0.0769	0.1209	0.0436	-0.3693	-0.0983	0.2731	-0.0181
As	0.4148	0.2469	-0.0240	0.4649	0.1792	0.3498	0.2421	-0.2352
Sb	0.3159	0.1719	-0.1851	0.3983	0.2075	0.3467	-0.1076	0.3753
Mo	-0.3035	-0.2292	0.5661	-0.2910	0.1378	0.3180	0.2809	-0.0206
Sn	-0.2744	-0.2201	-0.2361	0.7125	0.1020	-0.1411	0.0413	-0.2292
Ga	-0.3949	0.0932	0.6073	0.3196	-0.0233	-0.2619	-0.0113	0.3148
W	0.0325	-0.4440	0.1908	-0.1928	0.5596	-0.2529	0.0191	-0.3887
Nb	-0.6294	-0.1238	0.1300	0.4679	-0.1660	0.0302	0.0585	-0.2283
Au	-0.2139	-0.0088	-0.3292	-0.1189	0.1811	-0.1826	0.7622	0.2673

Relative Contributions: Variables

	1	2	3	4	5	6	7	8
Fe	33.9846	34.7789	0.8380	1.9533	0.2279	2.1514	4.2985	0.0995
Ag	20.8215	0.3046	15.7222	6.6228	0.6637	27.2022	0.7084	0.4186
Mn	5.0160	46.9294	0.3877	0.8175	14.3566	0.6368	2.3851	0.3069
Cr	34.4887	0.6868	1.3405	1.6441	25.9781	0.5284	0.0134	0.5732
V	7.5719	49.5230	2.0078	1.3510	0.1792	0.4747	1.7085	17.1946
Cu	33.9764	18.7730	5.1618	1.2549	1.4316	0.3517	1.5472	1.3082
Pb	6.9296	11.4999	40.4803	1.1248	0.9897	14.2558	0.1273	0.8235
Zn	57.8262	1.1129	0.0061	0.9799	4.7754	1.1470	2.4146	2.6280
Ni	55.9431	6.6930	8.5564	0.0289	0.1091	0.3043	4.4037	0.1604
Co	53.1565	0.5909	1.4611	0.1900	13.6380	0.9663	7.4608	0.0327
As	17.2026	6.0975	0.0578	21.6139	3.2097	12.2391	5.8631	5.5300
Sb	9.9812	2.9554	3.4246	15.8649	4.3054	12.0171	1.1580	14.0879
Mo	9.2100	5.2539	32.0470	8.4709	1.8990	10.1102	7.8897	0.0426
Sn	7.5276	4.8427	5.5741	50.7695	1.0410	1.9912	0.1707	5.2550
Ga	15.5929	0.8678	36.8755	10.2113	0.0541	6.8592	0.0128	9.9110
W	0.1056	19.7101	3.6389	3.7165	31.3132	6.3964	0.0366	15.1052
Nb	39.6195	1.5339	1.6898	21.8965	2.7557	0.0914	0.3419	5.2116
Au	4.5764	0.0077	10.8389	1.4136	3.2796	3.3337	58.1014	7.1450

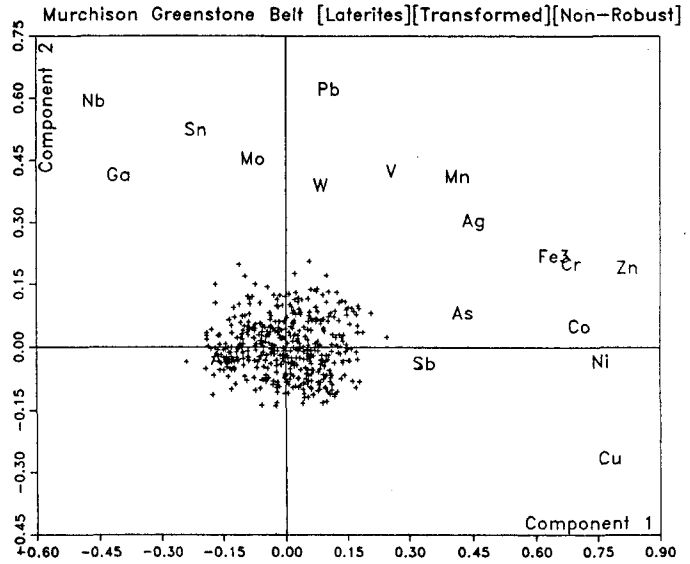


Figure 16a. Principal Component Plot: C1-C2: Murchison Greenstone Belt Laterites [Transformed][Non-Robust]

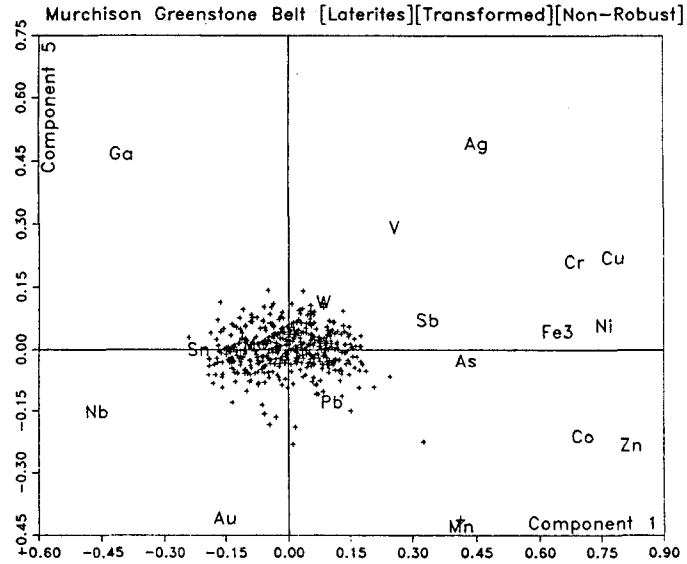


Figure 16b. Principal Component Plot: C1-C5: Murchison Greenstone Belt Laterites [Transformed][Non-Robust]

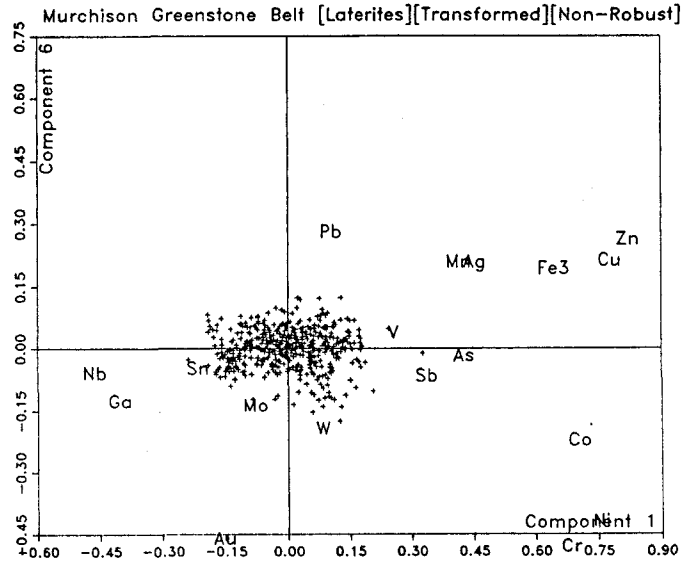


Figure 16c. Principal Component Plot: C1-C6: Murchison Greenstone Belt Laterites [Transformed][Non-Robust]

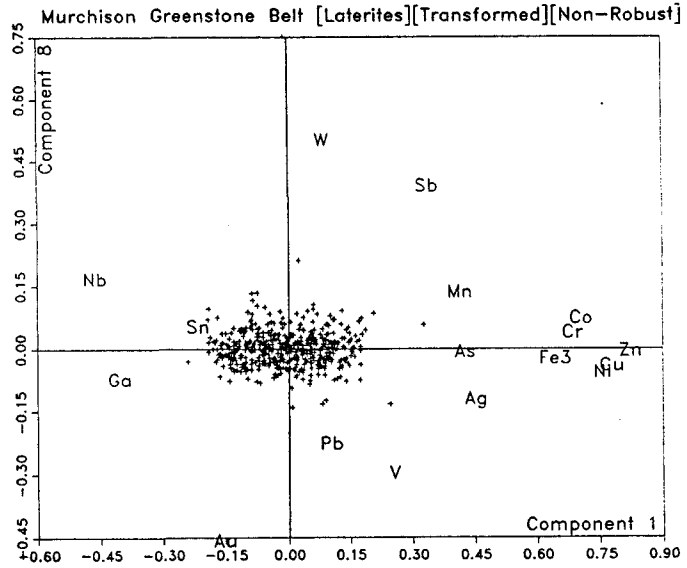


Figure 16d. Principal Component Plot: C1-C8: Murchison Greenstone Belt Laterites [Transformed][Non-Robust]

Table 9
Murchison Greenstone Belt Laterites [Robust][Transformed]
Robust Principal Components: Based on Previously Calculated
Robust Estimates of Means, Correlations, and Covariances

Observations: 484

Robust Means

Fe	14.78
Ag	-2.26
Mn	4.72
Cr	6.82
V	31.05
Cu	4.10
Pb	7.13
Zn	3.32
Ni	3.88
Co	3.47
As	3.76
Sb	0.73
Mo	1.03
Sn	0.37
Ga	7.18
W	1.65
Nb	3.05
Au	0.93

Robust Correlation Matrix

	Fe	Ag	Mn	Cr	V	Cu	Pb	Zn	Ni
Fe	1.0000	0.1405	0.3274	0.4002	0.5308	0.3595	0.0806	0.5363	0.2359
Ag	0.1405	1.0000	0.1571	0.2746	0.1708	0.4314	0.1688	0.3860	0.2745
Mn	0.3274	0.1571	1.0000	0.1799	0.2495	0.0824	0.1923	0.4955	0.1782
Cr	0.4002	0.2746	0.1799	1.0000	0.2546	0.3344	-0.0018	0.4154	0.6549
V	0.5308	0.1708	0.2495	0.2546	1.0000	0.0296	0.0924	0.1960	0.0739
Cu	0.3595	0.4314	0.0824	0.3344	0.0296	1.0000	-0.1140	0.5773	0.4580
Pb	0.0806	0.1688	0.1923	-0.0018	0.0924	-0.1140	1.0000	0.1444	0.0195
Zn	0.5363	0.3860	0.4955	0.4154	0.1960	0.5773	0.1444	1.0000	0.4363
Ni	0.2359	0.2745	0.1782	0.6549	0.0739	0.4580	0.0195	0.4363	1.0000
Co	0.2076	0.2271	0.3780	0.4069	0.1272	0.4108	-0.0262	0.4970	0.6484
As	0.0477	0.2235	-0.0171	0.3022	-0.0218	0.1882	0.1364	0.3211	0.1806
Sb	0.0716	0.1700	0.0436	0.1930	0.0325	0.1902	-0.0451	0.2016	0.1393
Mo	0.0279	-0.0518	-0.0221	0.0034	-0.0370	-0.2067	0.4201	-0.0730	0.0239
Sn	-0.0565	0.0976	0.0850	-0.0278	0.1873	-0.2690	0.0673	-0.0758	-0.2432
Ga	-0.2918	0.1440	-0.1539	-0.1438	-0.0139	-0.2914	0.2705	-0.3287	-0.1236
W	0.2451	0.0643	0.1184	0.1217	0.1612	-0.0284	0.0986	0.0477	0.0027
Nb	-0.2871	-0.0361	0.1415	-0.1807	-0.1027	-0.5198	0.2567	-0.2290	-0.3170
Au	-0.1329	-0.1692	-0.0617	-0.0676	-0.0072	-0.1968	-0.1027	-0.1338	-0.0906

	Co	As	Sb	Mo	Sn	Ga	W	Nb	Au
Fe	0.2076	0.0477	0.0716	0.0279	-0.0565	-0.2918	0.2451	-0.2871	-0.1329
Ag	0.2271	0.2235	0.1700	-0.0518	0.0976	0.1440	0.0643	-0.0361	-0.1692
Mn	0.3780	-0.0171	0.0436	-0.0221	0.0850	-0.1539	0.1184	0.1415	-0.0617
Cr	0.4069	0.3022	0.1930	0.0034	-0.0278	-0.1438	0.1217	-0.1807	-0.0676
V	0.1272	-0.0218	0.0325	-0.0370	0.1873	-0.0139	0.1612	-0.1027	-0.0072
Cu	0.4108	0.1882	0.1902	-0.2067	-0.2690	-0.2914	-0.0284	-0.5198	-0.1968
Pb	-0.0262	0.1364	-0.0451	0.4201	0.0673	0.2705	0.0986	0.2567	-0.1027
Zn	0.4970	0.3211	0.2016	-0.0730	-0.0758	-0.3287	0.0477	-0.2290	-0.1338
Ni	0.6484	0.1806	0.1393	0.0239	-0.2432	-0.1236	0.0027	-0.3170	-0.0906
Co	1.0000	0.2168	0.1297	-0.0814	-0.1129	-0.1470	-0.0361	-0.1932	-0.1126
As	0.2168	1.0000	0.3716	-0.1044	0.1002	-0.0866	-0.1474	0.0031	-0.0418
Sb	0.1297	0.3716	1.0000	-0.2325	0.0572	-0.0463	-0.1191	-0.0341	-0.0165
Mo	-0.0814	-0.1044	-0.2325	1.0000	-0.0480	0.1677	0.3016	0.1087	-0.0232
Sn	-0.1129	0.1002	0.0572	-0.0480	1.0000	0.1352	0.0316	0.4766	0.1380
Ga	-0.1470	-0.0866	-0.0463	0.1677	0.1352	1.0000	0.0070	0.3828	-0.0446
W	-0.0361	-0.1474	-0.1191	0.3016	0.0316	0.0070	1.0000	0.0119	-0.0079
Nb	-0.1932	0.0031	-0.0341	0.1087	0.4766	0.3828	0.0119	1.0000	0.0991
Au	-0.1126	-0.0418	-0.0165	-0.0232	0.1380	-0.0446	-0.0079	0.0991	1.0000

	Eigenvalues	% Trace	S Trace
1	4.2205	23.4470	23.4470
2	2.1646	12.0255	35.4725
3	1.7557	9.7538	45.2263
4	1.5070	8.3720	53.5982
5	1.1392	6.3289	59.9271
6	1.0838	6.0214	65.9485
7	1.0102	5.6120	71.5605
8	0.8406	4.6700	76.2304
9	0.7890	4.3836	80.6140
10	0.7135	3.9640	84.5780
11	0.5141	2.8561	87.4340
12	0.5025	2.7917	90.2257
13	0.4204	2.3356	92.5613
14	0.3913	2.1736	94.7350
15	0.3032	1.6843	96.4193
16	0.2328	1.2934	97.7126
17	0.2210	1.2280	98.9406
18	0.1907	1.0594	100.0000

Principal Component R-Mode Loadings

	1	2	3	4	5	6	7	8
Fe	0.6193	0.2287	-0.3921	-0.3385	-0.2245	0.0619	0.1490	-0.1611
Ag	0.4662	0.2810	0.2995	0.2018	-0.3940	-0.0127	-0.2843	0.4010
Mn	0.4149	0.4394	-0.0670	-0.2805	0.3121	-0.5278	-0.0214	0.0343
Cr	0.6824	0.1360	0.0541	0.0512	0.1817	0.4410	-0.1140	-0.1392
V	0.3235	0.3972	-0.2186	-0.4796	-0.2645	0.2046	-0.2190	-0.3737
Cu	0.7376	-0.3062	0.0057	0.1382	-0.2487	-0.0836	-0.0346	0.2775
Pb	0.0108	0.6679	0.0052	0.3838	-0.0539	-0.1379	0.3399	-0.1396
Zn	0.8102	0.1531	0.0370	-0.0747	0.0280	-0.2515	0.2067	0.1598
Ni	0.7206	-0.0507	0.0165	0.3353	0.3528	0.2021	-0.2380	-0.1035
Co	0.6834	0.0117	0.1092	0.1291	0.4177	-0.1420	-0.2529	-0.0767
As	0.3538	0.0510	0.5994	0.0842	-0.0299	0.2065	0.4612	-0.0613
Sb	0.2973	-0.0727	0.5600	-0.1266	-0.1624	0.1876	0.2485	-0.0875
Mo	-0.1415	0.4596	-0.4098	0.4787	0.1449	0.1921	0.2860	0.0116
Sn	-0.2071	0.4828	0.3942	-0.4528	-0.0093	0.1063	-0.1165	0.1516
Ga	-0.3766	0.4042	0.2347	0.3986	-0.1771	0.1097	-0.4332	-0.1346
W	0.0680	0.3890	-0.4972	-0.0203	-0.0744	0.3020	0.0221	0.4590
Nb	-0.4680	0.5737	0.3762	-0.0749	0.2154	-0.1452	-0.0556	0.0578
Au	-0.2196	-0.0445	0.0574	-0.3421	0.4870	0.4053	0.0677	0.2972

Relative Contributions: Variables

	1	2	3	4	5	6	7	8	9
Fe	38.3576	5.2292	15.3762	11.4606	5.0391	0.3826	2.2205	2.5945	0.1232
Ag	21.7357	7.8943	8.9700	4.0740	15.5235	0.0160	8.0833	16.0785	4.1504
Mn	17.2147	19.3032	0.4491	7.8675	9.7406	27.8604	0.0456	0.1175	0.3293
Cr	46.5621	1.8507	0.2929	0.2618	3.3023	19.4476	1.2988	1.9381	3.1702
V	10.4620	15.7745	4.7778	23.0043	6.9961	4.1873	4.7957	13.9622	4.1739
Cu	54.4082	9.3768	0.0033	1.9100	6.1849	0.6988	0.1195	7.7025	1.8685
Pb	0.0116	44.6105	0.0027	14.7277	0.2903	1.9018	11.5536	1.9491	8.5029
Zn	65.6418	2.3451	0.1370	0.5587	0.0784	6.3261	4.2714	2.5540	1.0237
Ni	51.9279	0.2569	0.0273	11.2417	12.4456	4.0853	5.6655	1.0719	0.0711
Co	46.7052	0.0138	1.1933	1.6664	17.4458	2.0157	6.3967	0.5889	0.3916
As	12.5143	0.2602	35.9266	0.7083	0.0897	4.2631	21.2746	0.3761	0.1718
Sb	8.8389	0.5287	31.3593	1.6038	2.6367	3.5200	6.1750	0.7665	4.1597
Mo	2.0021	21.1264	16.7941	22.9138	2.0998	3.6922	8.1809	0.0135	0.0324
Sn	4.2887	23.3109	15.5424	20.5022	0.0087	1.1300	1.3567	2.2992	1.0276
Ga	14.1838	16.3384	5.5096	15.8918	3.1348	1.2024	18.7620	1.8123	1.1349
W	0.4622	15.1318	24.7224	0.0414	0.5528	9.1201	0.0490	21.0691	16.2524
Nb	21.9068	32.9089	14.1540	0.5609	4.6380	2.1079	0.3089	0.3344	3.0499
Au	4.8229	0.1983	0.3297	11.7008	23.7127	16.4270	0.4587	8.8307	29.2704

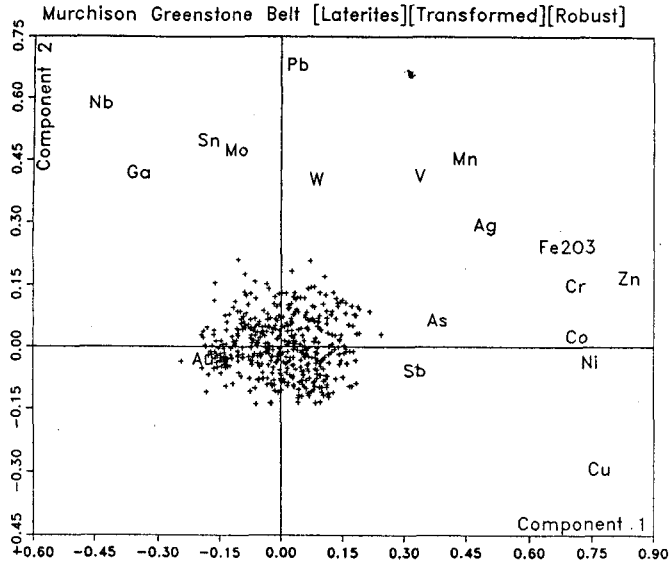


Figure 17a. Principal Component Plot: C1-C2: Murchison Greenstone Belt Laterites [Transformed][Robust]

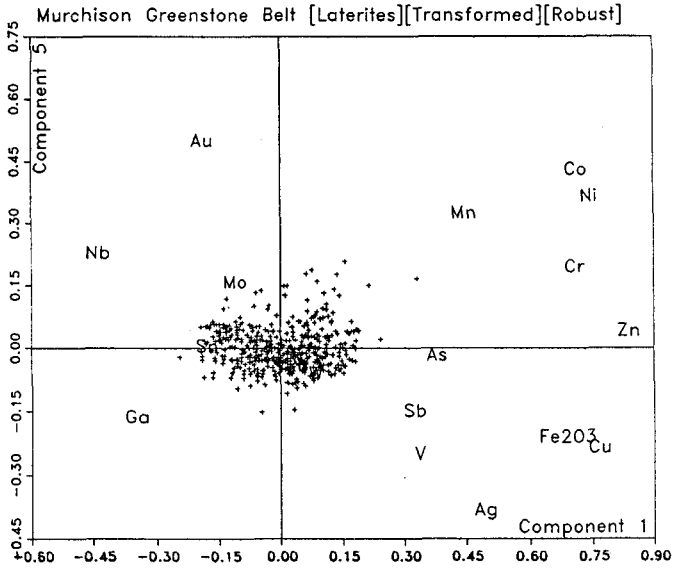


Figure 17b. Principal Component Plot: C1-C5: Murchison Greenstone Belt Laterites [Transformed][Robust]

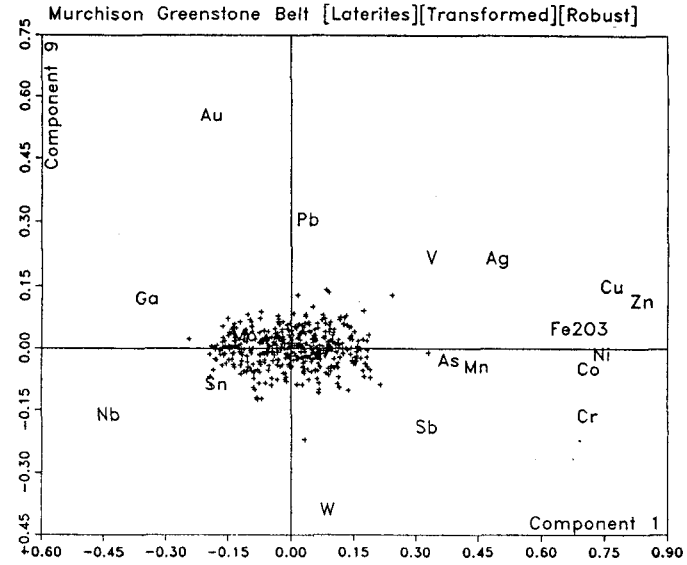


Figure 17c. Principal Component Plot: C1-C9: Murchison Greenstone Belt Laterites [Transformed][Robust]

Figures 17b and 17c show the scores of the samples and variables plotted onto the C1-C5 and C1-C9 axes respectively. Figure 17b shows that Au is relatively enriched along the C5 axis. The Figure suggests that Au enrichment may be associated with Mn, Co, and Ni enrichment, but that the association appears to be inversely related to relative enrichment in Fe, or Cu. Figure 17c shows that Au occurs associated only with Pb along the positive side of the C9 axis. This indicates that Pb is relatively enriched in samples that also have Au enrichment.

Experience within lateritic materials of the Yilgarn Block has found that elements such as Au can be significanct in as many as four components. The significance is due to the presence and absence of various chalcophile elements that may or may not be present with Au. Typically, Au may have a strong association with As in one component and an inverse association with As in another component. Part of this multiple component association is an artifact of the mathematics of PCA. Interpretation of these components must be placed within the context that within an area of elevated Au abundances, samples with elevated Au abundances may have As present or absent. In such a case, this relationship is expressed as two distinct components, but in fact represents the same locality.

Maps of the principal component scores of the samples can be useful in expressing linear relationships of the data that express meaningful geochemical processes. If a component expresses underlying lithologies, then a map of that component will clearly outline the major lithological variation of the area (Grunsky, 1986). Other components that outline other processes such as mineralization or alteration can also be clearly expressed on maps that display the component scores.

4.3 Cluster Analysis Methods

In an exploration programme, groups of samples defined by multi-element geochemistry may form distinct clusters that can represent background and target populations. Cluster analysis methods are useful as an exploratory tool for detecting groups of multi-element data that may not be readily observable in simple scatter plots or principal components analysis. The main objective of clustering algorithms is to represent natural groupings of multidimensional data in as few dimensions as possible. Clustering methods can be broadly divided into hierarchical and non-hierarchical methods. Davis (1986) provides a good introductory review of clustering methods.

Clustering procedures in geochemistry have been applied in many studies. Sinding-Larsen (1975) used clustering methods for the initial subdivision of a heterogeneous geochemical area. Jaquet *et al.* (1975) provide a detailed analysis of lake sediment geochemistry through the use of clustering procedures. Howarth and Sinding-Larsen (1983) provide a general discussion of clustering methods applied to geochemical exploration. Grunsky (1986) has shown how dynamic cluster analysis was used to detect different types of mineralization based on distinct geochemical differences between the mineral occurrences. More recently, the use of fuzzy clustering methods in geochemistry has been introduced (Bochang and Xuejing, 1985).

Hierarchical Methods

Hierarchical clustering is based on the linking variables (R-mode) or samples (Q-mode) from measures of similarity. The relationships between the variables or samples are usually expressed graphically on the dendrogram. Individual clusters can be discriminated by choosing an appropriate value of linkage at which the similarities between the objects is not realistic. Hierarchical clustering assumes a constant linkage of samples/variables which may not be a reasonable assumption for geological data.

The most common measure of similarity for clustering on the basis of variables (R-mode) is the correlation coefficient. For Q-mode analysis, the Euclidean distance can be used as a measure of proximity through which samples can be clustered.

Non-hierarchical Methods

Arbitrary Origin Methods are non-hierarchical and may offer some advantage over hierarchical methods since the clusters which are formed are based on multivariate similarities (proximities) rather than individual correlation coefficients. These methods start with an initial number of cluster centres that can be randomly or specifically chosen. Each sample is allocated to one of the groups based on the closest distance to group centres. The process is iterative and group centres change until a stable solution results. Methods such as K-means (McQueen, 1967; Everitt, 1974) or dynamic cluster analysis (Diday, 1973) are examples of this.

One method of non-hierarchical clustering known as dynamic cluster analysis has been applied to geochemical data. Like many non-hierarchical methods, it is commonly applied to component scores from methods such as principal analysis. The reasoning for clustering together component scores is based on the interpretation that linear combinations of the variables (component scores) may reflect geological processes that will be grouped together in a clustering procedure. Additionally, the component plots provide a reduced set of dimensions for viewing the multi-element associations of the data and thus provide additional visual assistance in examining grouped associations.

The dynamic cluster analysis procedure works by initially selecting random "seed" points, or group centroids, in the principal component space of the samples. The analysis proceeds through the following steps:
i) The Mahalanobis distance from each sample to each random seed point is computed.
ii) Each sample is then allocated to a group representing the shortest distance to one of the group centroids.
iii) From the samples that comprise each group, a new group centroid value is calculated which represents a "new" mean or group centroid.
iv) Mahalanobis distances for each sample are continued with corresponding allocation to new group centroids. Steps i) to iii) are repeated until the group centroids stabilize.
v) Each group is compared using Hotelling's T^2 and corresponding F-test for statistical uniqueness of the groups. If two groups are not statistically different at the 95% confidence level, then the groups are joined together and the statistics recomputed.
vi) Step v) is repeated until each group is statistically distinct.
vii) For each group, average compositions of the samples (from the original geochemical data) are computed so that compositions of the groups can be compared.

Dynamic cluster analysis was applied to the same transformed data as for the example describing the use of principal components analysis. Partial results of the analysis are shown in Table 10.

The first part of the Table shows that the data was subdivided into 22 statistically-distinct groups. Notice however, that Groups 14 to 22 consist of single values that represent outliers. Groups 1 to 13 represent groups of samples that are statistically distinct from each other based upon the principal component scores obtained from the robust principal components analysis procedure shown in Table 9.

The second part of the Table shows the mean composition of each group using the original *untransformed* data. This permits examination of the characteristics of the groups in units that are familiar to the geologist. Each mean value is accompanied by its coefficient of variation which is defined as (standard deviation/mean value)*100.0. Large coefficients indicate broad dispersion while small coefficients indicate limited dispersion. The main advantage of the ratio is that it is unitless and thus each element can be compared with every other element. Thus, relative variations of the elements can be observed with this coefficient.

Significant elements in groups can be done by scanning each element through each of the groups. Large and small mean values assist in determining the geochemical significance of the groups. As well, plotting the mean component scores onto the principal components axes, shown in Figures 18a,b,c,d, can assist in viewing multi-element associations with each group.

Groups 3 and 1 show relatively-large Fe values relative to the other groups. However, Group 3 has significantly more V, Ni, and Co, while Group 1 has more Mo, Sn, and W. This might suggest that these samples have distinct lithological associations. Group 5 shows significant Au accompanied by elevated Mo and lower Fe. Elevated W occurs in Group 1. Figures 18a, b, c, and d show projections of the mean component scores for each group. Figure 18a shows that Groups 3 and 1 are associated with the mafic rocks along the positive side of the C1 axis. Group 3, having a greater Fe abundance, plots closer to Fe on the plot.

Groups that may be of interest to the investigator are those with elevated values for chalcophile elements, pathfinder elements, and commodity elements. Associations with elevated Au occur in Groups 2, 5, 7, 8, 11, 12, and 13. Isolated Au samples occur as Groups 17, 18, 19, 20, and 21. These Groups can be investigated for their Au associations with other elements. Some of the associations may be lithologically controlled, while other associations may be with alteration assemblages. Further scatter plots, clustering or components analysis can be carried out on some of the larger groups to examine the multi-element associations. Figures 18a,b,c,d show that these groups plot closer to Au than other groups indicating their affinity to relative Au enrichment.

Other groups show elevated abundances for elements of interest. These include Group 13 for Ag; Groups 4, 6, 16, 17, 18, and 19 for Cu; Groups 9, 10, 13, 15, 16, 18 20, and 22 for Pb; Groups 9, 13, 16, 17, 18, and 22 for Zn; Groups 11, 16, 17, and 22 for Ni; Groups 14, 16, and 21 for Mo; and Groups 7, 10, 12, 13, 15, and 21 for Sn.

Some primary lithological associations are described by groups 1, 2, 3, 4, 5, 7, and 12. These groups show little deviation from the C1 axis. Mafic/ultramafic volcanic associations may be suggested by Groups with elevated Cr, Mn, Ni, and Co. Groups 1, 3, 4, 6, 9, 11, and 13 show elevated Cr values as well as individual samples that form groups 14, 15, 16, 18, 19, 21, and 22.

4.4 χ^2 (Chi-square) Plots: A multivariate extension of Q-Q Plots

Most anomaly recognition procedures are based upon determining the threshold that distinguishes background from anomalous values. However, the use of multivariate procedures can be useful in determining background from anomalous samples for a set of desired elements.

Garrett (1989b, 1990) and Chork (1990) describe the use of the covariance matrix as a tool for distinguishing background from anomalous sample populations. The covariance matrix contains information on the variability of the elements as well as their inter-relationships. The multi-element data define a hyper-ellipsoid in multidimensional space. The mean values of each element define the centroid of this hyper-ellipsoid

Table 10
Dynamic Cluster Analysis
Murchison Greenstone Belt Laterites
Loadings Determined from Robust PCA on Transformed Data
Observations: 484

Mean Principal Component Loadings of the Groups

Group	Size	1	2	3	4	5	6	7	8	9	10
1	68	0.0196	0.0431	-0.0716	0.0085	-0.0125	0.0246	0.0038	0.0339	-0.0218	-0.0081
2	50	-0.1304	-0.0200	0.0212	0.0349	0.0110	0.0155	-0.0026	-0.0167	0.0002	-0.0036
3	52	0.0433	-0.0225	-0.0368	-0.0154	-0.0060	0.0016	-0.0012	-0.0351	0.0205	-0.0166
4	32	0.1141	-0.0457	0.0176	0.0401	-0.0254	-0.0121	-0.0013	0.0020	-0.0083	0.0139
5	51	-0.0556	-0.0147	0.0010	0.0231	-0.0087	0.0250	-0.0074	0.0066	0.0269	0.0114
6	21	0.0745	-0.0761	0.0460	0.0536	-0.0434	0.0033	-0.0177	0.0182	0.0010	-0.0197
7	36	0.0047	0.0172	-0.0093	-0.0661	-0.0236	-0.0264	-0.0547	0.0037	0.0188	-0.0018
8	17	-0.0570	-0.0834	-0.0226	-0.0363	0.0308	-0.0535	-0.0182	0.0255	0.0027	0.0102
9	20	0.0990	0.0788	0.0097	0.0352	0.0426	-0.0767	0.0327	0.0170	0.0062	0.0194
10	43	-0.1134	0.0574	0.0108	-0.0101	0.0122	-0.0420	0.0106	-0.0186	-0.0021	-0.0071
11	40	0.0895	-0.0113	0.0424	-0.0220	0.0671	0.0274	-0.0127	-0.0217	-0.0333	0.0149
12	31	-0.0270	-0.0292	0.0452	-0.0620	-0.0157	0.0208	0.0468	0.0002	-0.0295	0.0027
13	14	0.0651	0.1036	0.0908	-0.0093	-0.0439	-0.0155	0.0592	0.0299	0.0594	0.0011
14	1	0.0667	-0.0040	-0.2182	0.0670	0.0546	0.0722	0.0385	0.0873	-0.0848	0.0038
15	1	0.0322	0.0791	0.0858	0.0997	-0.1457	0.0782	0.0695	0.0180	0.0620	0.0473
16	1	0.1379	0.0978	-0.0483	0.1596	0.0339	0.0102	0.2440	-0.0121	0.0516	0.1068
17	1	0.3319	-0.0498	0.0147	-0.0497	0.1669	-0.0638	0.1795	0.0574	-0.0103	-0.0677
18	1	0.2430	0.0304	0.0209	0.0488	0.0209	0.0097	0.1853	-0.0240	0.1292	0.1036
19	1	0.1663	-0.0975	0.0446	-0.0576	0.0345	0.0613	0.1738	0.0551	-0.0233	-0.1079
20	1	0.0084	0.0481	0.0312	0.0253	0.1488	-0.2335	0.0598	0.0161	0.0532	-0.1277
21	1	0.0321	0.1311	-0.1556	-0.0113	-0.0149	0.1128	-0.0605	0.0931	-0.2204	-0.0641
22	1	0.1567	0.1068	-0.0572	0.0691	0.2079	-0.0856	-0.0245	-0.0204	-0.0306	-0.0435

Table 10
Group Description
Mean Composition & Coefficient of Variation (C.V.)

Box-Cox Power Transformations

Transforms	λ=0.6		ln		ln		ln		λ=0.4		ln	
Group	Fe	C.V.	Ag	C.V.	Mn	C.V.	Cr	C.V.	V	C.V.	Cu	C.V.
1	17.67	20.14	-1.92	-69.29	4.88	16.82	7.08	11.70	34.18	25.16	4.29	26.31
2	8.39	28.66	-3.23	-14.96	3.68	16.03	5.88	7.86	24.23	17.38	2.66	24.91
3	18.22	18.21	-2.58	-41.04	4.91	14.43	7.15	14.13	38.36	18.95	4.33	18.93
4	16.94	14.64	-1.41	-74.06	4.56	13.56	7.40	6.96	27.05	41.67	5.98	12.71
5	12.44	26.16	-2.16	-60.66	3.74	25.52	6.54	10.30	29.18	33.26	3.96	28.26
6	13.94	13.19	-1.00	-77.76	4.02	8.30	7.05	11.04	23.73	43.75	6.11	16.37
7	16.73	21.39	-1.70	-74.35	5.33	13.92	6.44	10.63	42.56	27.19	4.32	18.82
8	11.19	54.35	-3.21	-13.46	5.11	15.74	5.96	14.98	21.44	51.74	3.92	17.18
9	14.79	36.15	-1.55	-90.21	6.40	13.53	7.09	15.53	29.17	37.01	4.89	16.03
10	13.42	35.31	-3.23	-15.42	5.13	16.59	5.78	11.20	27.78	43.39	2.48	28.39
11	15.67	22.15	-2.60	-48.31	5.22	16.51	8.36	10.65	32.45	21.27	4.18	23.70
12	15.13	25.48	-2.87	-28.64	4.49	16.50	6.78	13.31	29.10	30.13	3.64	28.38
13	16.69	23.11	0.13	705.36	5.38	16.90	7.05	11.64	36.28	16.39	4.63	24.28
14	15.80	0.00	-3.40	0.00	4.09	0.00	8.11	0.00	26.34	0.00	5.19	0.00
15	16.38	0.00	2.30	0.00	2.77	0.00	7.21	0.00	33.74	0.00	4.62	0.00
16	20.28	0.00	-3.40	0.00	3.87	0.00	7.47	0.00	20.00	0.00	6.88	0.00
17	25.39	0.00	-3.40	0.00	7.10	0.00	6.79	0.00	17.32	0.00	6.01	0.00
18	21.79	0.00	-1.20	0.00	5.01	0.00	8.06	0.00	40.06	0.00	6.61	0.00
19	26.33	0.00	-3.40	0.00	5.49	0.00	7.63	0.00	9.92	0.00	6.42	0.00
20	4.70	0.00	-3.40	0.00	9.21	0.00	5.04	0.00	12.28	0.00	3.50	0.00
21	14.55	0.00	-3.40	0.00	4.79	0.00	8.55	0.00	44.98	0.00	4.38	0.00
22	14.67	0.00	-3.40	0.00	8.59	0.00	9.21	0.00	21.31	0.00	3.40	0.00

Table 10
Group Description
Mean Composition & Coefficient of Variation (C.V.)

Box-Cox Power Transformations

Transforms	λ=0.4		ln		ln		λ=0.4		ln		ln	
Group	Pb	C.V.	Zn	C.V.	Ni	C.V.	Co	C.V.	As	C.V.	Sb	C.V.
1	7.84	28.58	3.35	14.75	3.97	16.46	3.19	56.05	3.25	28.18	0.34	315.55
2	7.28	24.09	2.48	15.07	3.32	22.31	2.13	51.36	3.54	12.05	0.60	143.45
3	6.90	26.45	3.48	14.02	4.15	22.21	3.68	57.94	3.40	22.60	0.80	122.19
4	6.49	20.10	3.94	11.45	4.84	10.96	5.57	31.36	4.48	28.21	1.13	86.69
5	6.99	27.25	3.01	13.76	3.57	20.53	2.36	55.85	3.51	22.50	0.34	266.98
6	5.89	22.75	3.57	6.10	4.75	9.40	4.30	28.98	4.33	18.11	1.66	30.67
7	5.60	46.86	3.42	10.76	3.44	27.71	3.56	48.98	2.81	30.19	0.42	233.16
8	4.37	59.04	3.09	9.58	3.43	14.25	2.45	59.74	2.74	33.74	-0.30	-144.81
9	11.18	22.03	4.52	10.20	4.46	21.68	6.54	27.55	4.37	25.96	0.42	231.46
10	9.09	29.65	2.95	11.21	2.68	50.03	2.06	69.35	3.30	20.43	0.39	214.50
11	4.89	61.65	3.64	12.84	5.34	17.95	6.27	41.08	4.76	23.24	1.26	87.71
12	5.32	45.62	3.12	9.87	2.97	34.74	1.86	79.31	4.63	24.25	1.87	64.28
13	13.32	16.84	4.28	15.51	3.51	26.28	2.56	70.12	5.59	18.00	1.60	53.42
14	7.37	0.00	3.40	0.00	4.44	0.00	6.56	0.00	1.95	0.00	-0.41	0.00
15	13.58	0.00	3.26	0.00	3.69	0.00	0.57	0.00	6.04	0.00	1.95	0.00
16	19.40	0.00	4.91	0.00	5.23	0.00	4.89	0.00	6.05	0.00	0.69	0.00
17	7.37	0.00	6.38	0.00	6.00	0.00	14.06	0.00	6.58	0.00	3.47	0.00
18	19.16	0.00	4.63	0.00	5.96	0.00	5.79	0.00	8.77	0.00	-0.41	0.00
19	7.37	0.00	3.74	0.00	4.54	0.00	4.47	0.00	6.57	0.00	4.53	0.00
20	18.27	0.00	3.89	0.00	2.94	0.00	9.45	0.00	2.71	0.00	1.39	0.00
21	4.68	0.00	3.26	0.00	3.91	0.00	5.79	0.00	1.61	0.00	1.61	0.00
22	14.12	0.00	4.17	0.00	6.66	0.00	11.18	0.00	3.09	0.00	-0.41	0.00

Table 10
Group Description
Mean Composition & Coefficient of Variation (C.V.)

Box-Cox Power Transformations

Transforms	λ=0.4		ln		λ=0.5		ln		λ=0.3		ln	
Group	Mo	C.V.	Sn	C.V.	Ga	C.V.	W	C.V.	Nb	C.V.	Au	C.V.
1	1.96	52.45	0.19	617.93	6.89	30.30	3.08	31.63	2.72	50.94	0.79	149.42
2	1.38	46.66	0.41	268.60	8.64	10.27	1.30	25.61	3.82	23.67	1.23	94.44
3	0.84	105.71	-0.60	-169.73	6.70	26.47	1.38	32.84	1.68	85.09	0.98	154.30
4	0.67	109.19	-0.18	-643.82	6.81	16.48	1.37	33.65	1.16	101.27	-0.31	-258.96
5	1.45	66.28	0.31	363.16	7.95	24.85	1.39	33.53	3.20	35.60	1.51	84.52
6	0.04	1722.48	-0.79	-88.59	7.80	13.75	1.35	34.85	1.54	76.72	0.20	363.18
7	0.19	437.91	1.25	74.18	7.57	35.16	1.54	43.40	3.24	51.78	1.06	99.64
8	-0.25	-173.57	0.05	1823.28	5.58	24.70	1.28	25.70	2.45	66.68	1.41	81.01
9	1.94	36.53	0.73	130.85	6.38	22.26	1.63	41.31	3.63	31.22	0.15	928.29
10	1.33	77.38	1.06	116.34	8.66	20.31	1.39	34.23	5.54	20.14	0.92	131.26
11	0.54	200.38	0.70	202.02	6.10	36.48	1.43	44.29	3.40	48.61	1.19	98.46
12	0.39	214.65	1.31	72.07	5.59	39.53	1.51	38.63	3.18	48.68	1.16	95.41
13	0.54	188.45	1.36	84.46	7.47	24.94	1.52	40.96	4.39	29.51	1.50	81.24
14	4.14	0.00	-1.10	0.00	2.90	0.00	5.08	0.00	0.30	0.00	0.69	0.00
15	2.83	0.00	1.39	0.00	8.00	0.00	1.20	0.00	2.07	0.00	0.00	0.00
16	6.66	0.00	1.10	0.00	2.90	0.00	1.20	0.00	1.72	0.00	0.00	0.00
17	1.10	0.00	-1.10	0.00	0.00	0.00	2.30	0.00	0.30	0.00	2.94	0.00
18	0.69	0.00	-1.10	0.00	1.46	0.00	1.20	0.00	0.30	0.00	1.95	0.00
19	-0.41	0.00	-1.10	0.00	2.90	0.00	2.64	0.00	0.30	0.00	3.53	0.00
20	-0.41	0.00	-1.10	0.00	5.75	0.00	1.20	0.00	3.11	0.00	1.95	0.00
21	3.00	0.00	1.95	0.00	8.00	0.00	7.30	0.00	3.69	0.00	-1.10	0.00
22	1.10	0.00	-1.10	0.00	6.94	0.00	2.64	0.00	3.51	0.00	1.10	0.00

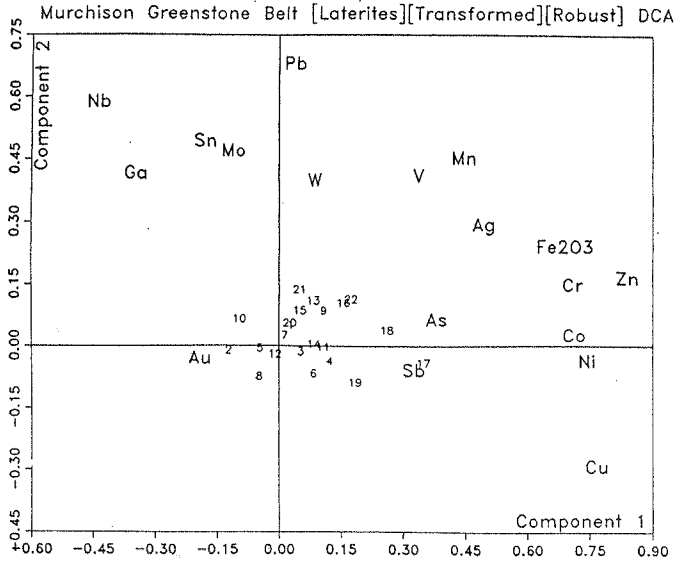


Figure 18a. Dynamic Cluster Analysi
Greenstone Belt Laterites [Transformed][I

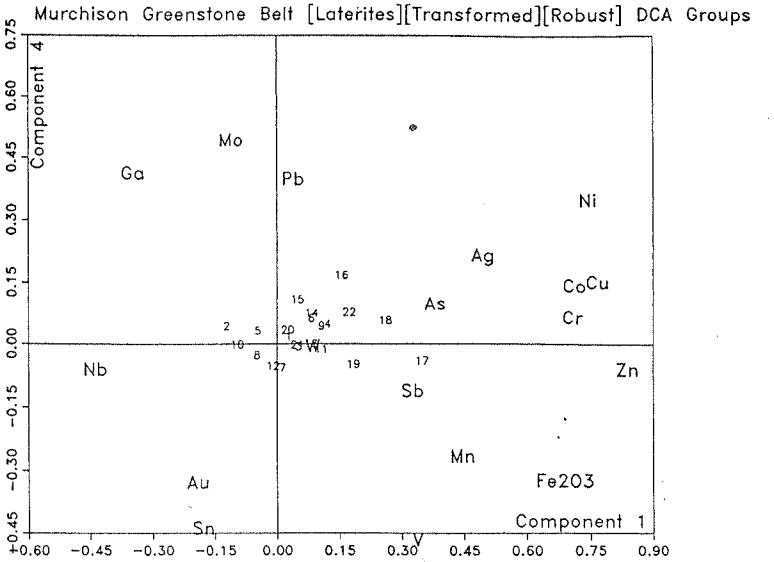


Figure 18b. Dynamic Cluster Analysis: C1-C3: Murchison Greenstone Belt Laterites [Transformed][Robust]

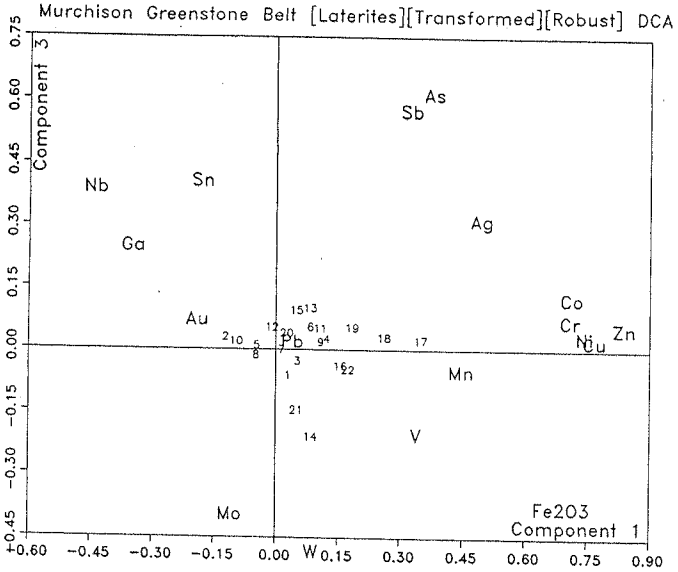


Figure 18c. Dynamic Cluster Analysis: C1-C4: Murchison Greenstone Belt Laterites [Transformed][Robust]

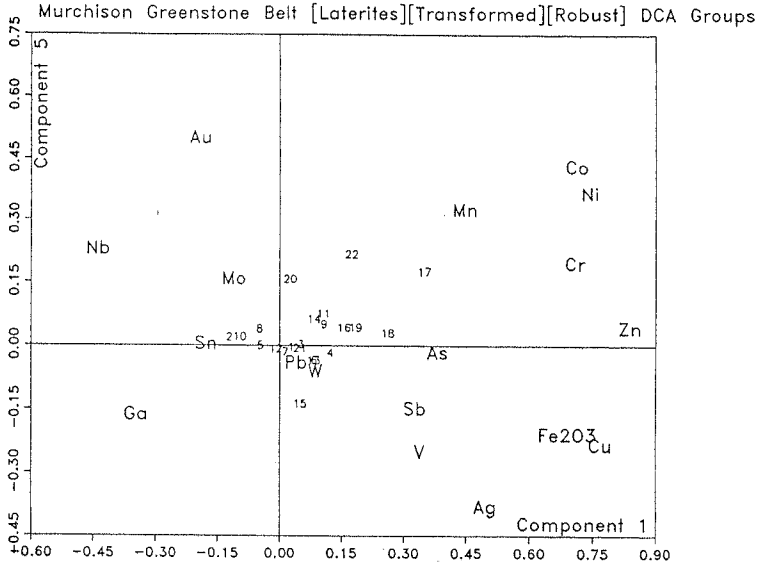


Figure 18d. Dynamic Cluster Analysis: C1-C5: Murchison Greenstone Belt Laterites [Transformed][Robust]

and the distance from each sample point to the centroid is known as the Mahalanobis distance. In a multivariate normal sample population, most samples lie within an expected radius of the centroid and are, by definition, the background group of samples. However, if outliers are included in the data, the shape of the hyper-ellipsoid that is defined by the covariance matrix changes. This resulting distortion affects the location of the centroid and thus affects the Mahalanobis distances for all of the samples.

The Mahalanobis distance is defined as:

For each sample x_i,

$$D_i^2 = (x_i - \bar{x})' S^{-1} (x_i - \bar{x})$$

where **S** is the covariance matrix and $\bar{x}$ is the mean vector for the data. The estimates of **S** and $\bar{x}$ can be obtained from robust procedures which enhance the presence of outliers.

Outliers can be distinguished from the main background population by determining the Mahalanobis distance of each sample to the group centroid. The distances can be compared to the "expected" distances of a multivariate normal population (cumulative probability with the number of degrees of freedom defined as the number of variables) by the use of χ^2 values. Garrett (1989b) uses robust estimates for determining the covariance matrix, S.

Table 11

Mahalanobis Distance vs. Theoretical χ^2 values
for Multielement Data [Box-Cox Power Tranformations]

Observations 866
Variables 8

Elements	Ag	As	Sb	Bi	Mo	Sn	W	Se
Transformations	ln	ln	ln	ln	ln	0.15	ln	0.66

Mahalanobis Distance vs. Chi-Square (χ^2)

Sample	Type	Mahalanobis Distance	χ^2 Value
G03121	LN	.74	.74
G04666	LN	1.03	1.00
G04664	LN	1.06	1.15
G04238	LP	1.28	1.26
G02605	LP	1.34	1.36
G03604	LP	1.44	1.44
G04082	LP	1.45	1.51
G04065	LP	1.47	1.58
G02598	LP	1.49	1.64
G03105	LN	1.54	1.69
G04236	LP	1.58	1.74
G02562	LP	1.58	1.79
G02459	LP	1.59	1.84
G03708	LP	1.59	1.88
G03341	LN	1.59	1.92
G02529	LP	1.61	1.96
.			
.			
.			
.			
.			

The following samples are considered outliers

G03352	LN	24.73	17.06
G03148	LN	26.67	17.18
G05693	LN	27.25	17.30
G05776	LN	27.98	17.42
G04043	LN	27.99	17.55
G03152	LN	28.00	17.69
G03832	LN	32.11	17.83
G02409	LP	32.70	17.98
G03939	LN	32.73	18.14
G04137	LP	33.01	18.31
G05780	LP	34.23	18.48
G06237	LN	35.42	18.67
G06085	LP	36.20	18.87
G06833	LN	37.93	19.08
G04034	LP	40.70	19.31
G04273	LP	41.66	19.56
G05097	LN	42.44	19.84
G03326	LN	44.48	20.14
G05798	LP	48.27	20.48
G06819	LN	54.00	20.87
G03962	LN	61.53	21.32
G03975	LN	75.93	21.85
G06084	LP	85.94	22.52
G02612	LN	91.43	23.40
G05795	LP	158.31	24.72
G05816	LN	195.14	27.51

A graphical procedure of plotting the Mahalanobis distances of the observed from the expected values allows for the detection of outliers. If the sample population is multivariate normal, then the χ^2 plot is a straight line. If the population contains outliers, then the observed Mahalanobis distances are greater than the expected χ^2 values and the plot becomes non-linear.

A χ^2 plot was prepared for the data collected over the Murchison greenstone belt. The data were transformed using λ values for Box-Cox transformations as shown in Table 1. Only chalcophile elements were selected as these might be the elements used in a regional geochemical exploration programme. Table 11 shows Mahalanobis distances and corresponding theoretical χ^2 values for a selected group of samples from the Murchison greenstone terrain (Dataset 1). The last 26 samples are listed at the bottom of the Table. Note that these samples have rather large Mahalanobis distances relative to the theoretical χ^2 values.

Figure 19a shows a χ^2 plot of the untransformed laterite data. The procedure used to calculate the Mahalanobis distances for this plot used robust estimates. The Figure shows a distorted curve with 18 samples that plot away from the main part of the data. Figure 19b shows the same data after they were transformed using the coefficients listed in Table 11. The line is less curved and there appears to be only 26 atypical samples. Figure 19c shows a χ^2 plot of the data after removing the 26 atypical samples. The Mahalanobis distances were recomputed with these outliers removed. The curve is much closer to a straight line and the data can be assumed to represent the background population data. Figure 20 shows a map of three of the 26 atypical samples plotted onto a map that covers the southern part of the Murchison greenstone area (see Grunsky *et al.,* 1988). The other eight samples plot on the adjoining map sheet north of the area (see Grunsky *et al.,* 1989). These samples can be considered anomalous for the elements used to compute the Mahalanobis distances and could be considered for further follow-up. From this example, it should be clear that the χ^2 plot is a useful tool for outlier recognition in multi-element datasets.

Since the χ^2 plot is a long tailed distribution, it is not symmetrical. Thus identifying outliers at the lower end of the χ^2 plot is difficult. Also, because the tail is very long, it is sometimes difficult to recognize outliers with apparently large Mahalanobis distances. An alternative to the χ^2 plot is the use of the cube root of the Mahalanobis distance plotted against the quantiles for a normal distribution. This plot has the advantage that it is symmetric, enabling the identification of outliers at both ends of the distribution.

4.5 The Use of Empirical Indices

The use of pathfinder elements has prompted the use of a number of numerical procedures through which selected elements can be elements can be used in an exploration programme by creating mineralization potential indices based upon the weighted sum scores of the pathfinder elements. The techniques used in this approach are described by Garret (1991), Garrett *et al.* (1980), Smith and Perdrix (1983), Smith *et al.* (1987), and Chaffee (1983).

Weighted Sum Index

Garrett *et al.* (1980:144) has suggested the use of a linear combination of a group of indicator elements that gives a weighted sum. This is defined as:

For m variables

$$V_i = \sum_{j=1}^{m} a_j z_{ij},$$

where

$$a_j = I_j \Big/ \left(\sum_{j=1}^{m} I_j^2\right)^{1/2},$$

$$z_{ij} = (x_{ij} - \bar{x}_j)/s_j$$

$\bar{x}_j$ is the robust mean estimate for variable j, and
s_j is the robust estimate of the standard deviation of variable j.
The quantity I_j is the importance of the jth variable and that

$$\sum_{j=1}^{m} I_j + 1.$$

In the multi-element survey, those elements which are considered pathfinders are given more weight than elements which may be more diagnostic of background. Background elements may be given weights of zero. Examples of the use of this index are given in Garrett *et al.* (1980).

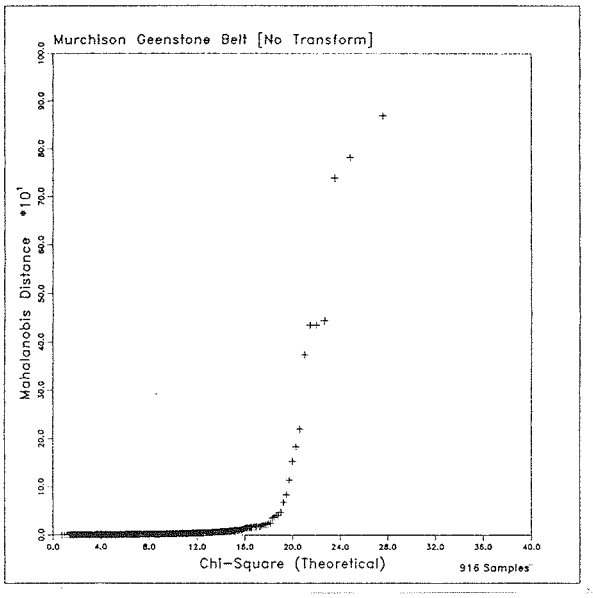


Figure 19a. χ^2 Plot: Murchison Greenstone Belt Laterites: 866 Samples [Non-Transformed]

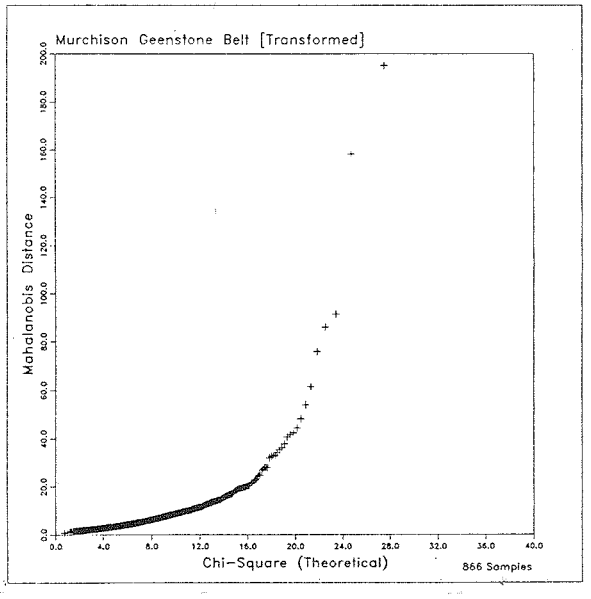


Figure 19b. χ^2 Plot: Murchison Greenstone Belt Laterites: 866 Samples [Transformed]

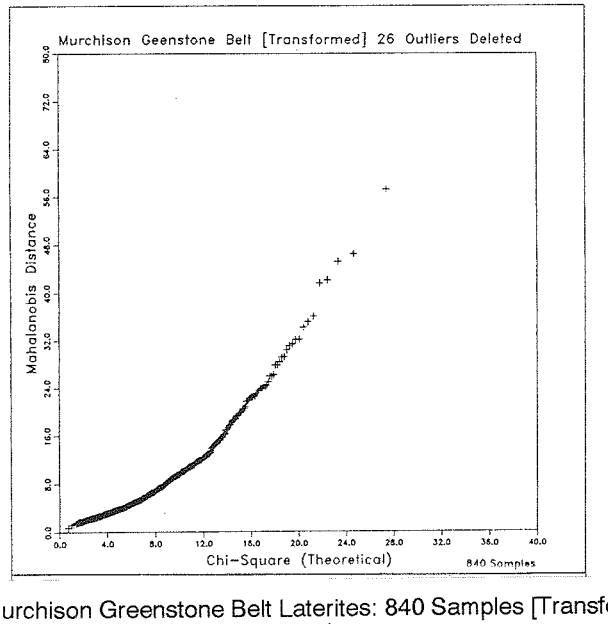


Figure 19c. χ^2 Plot: Murchison Greenstone Belt Laterites: 840 Samples [Transformed] 26 samples removed.

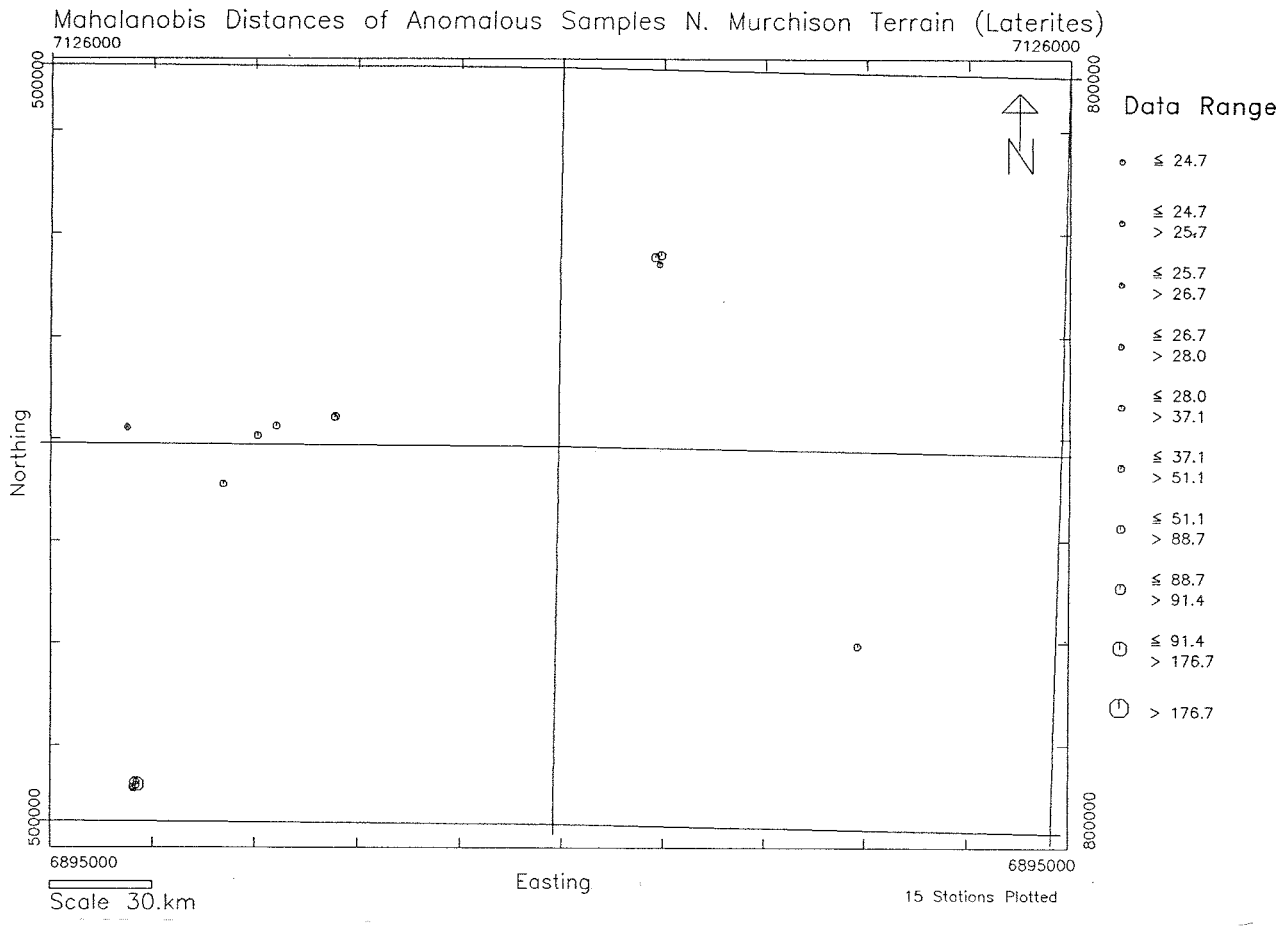


Figure 20a. Map of anomalous samples in North Murchison Greenstone area representing samples with the largest Mahalanobis distances.

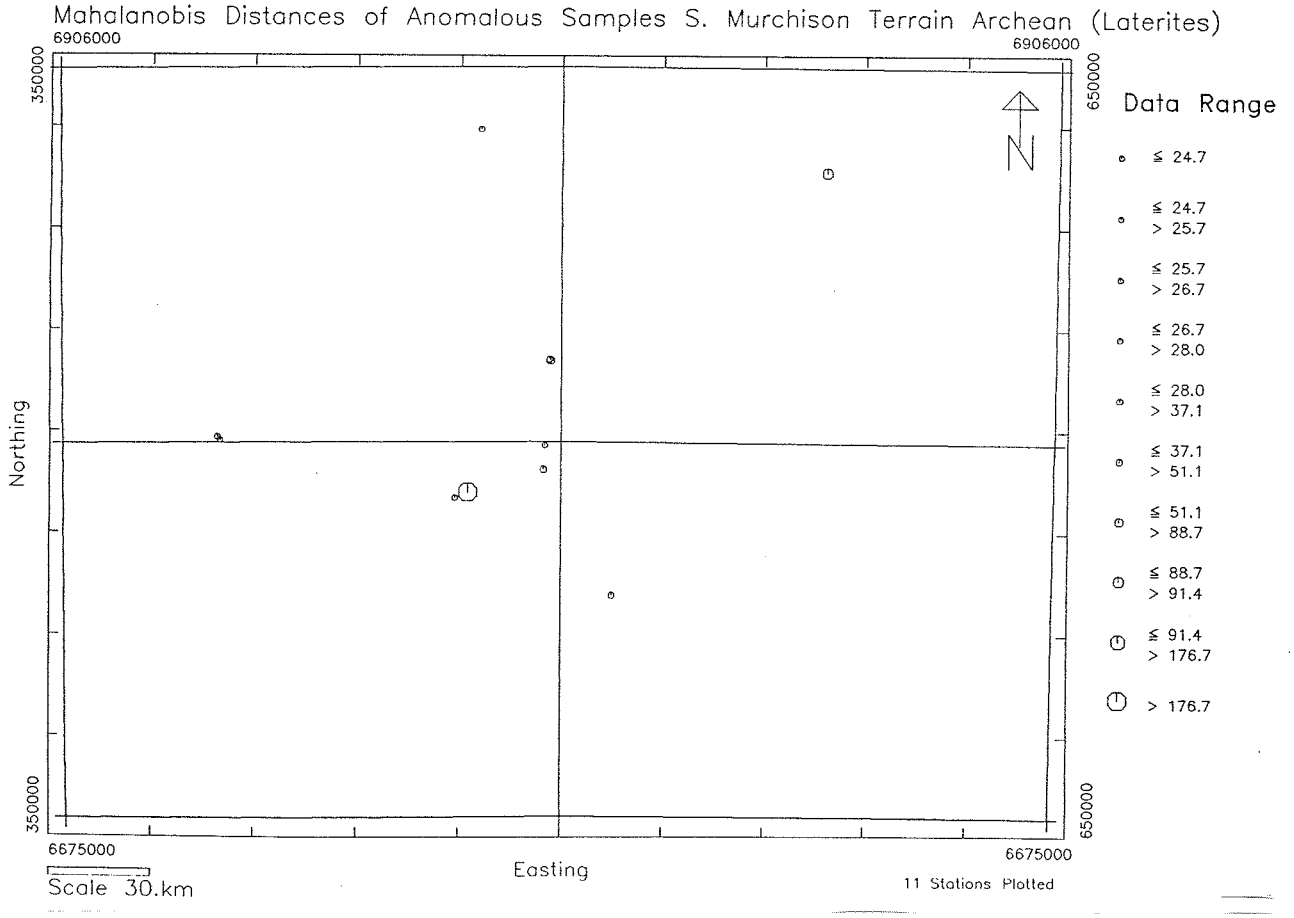


Figure 20b. Map of anomalous samples in South Murchison Greenstone area representing samples with the largest Mahalanobis distances.

Scoresum

Chaffee (1983) has developed a method of scoring samples for anomaly potential. Each element is evaluated such that the range of abundances are subdivided into 4 groups, or thresholds, with corresponding scores that represent background (0), weakly anomalous (1), moderately anomalous (2), and strongly anomalous (3). These ranges are derived from orientation studies over areas where the range of abundances and underlying geochemical distributions are reasonably-well understood. Each sample is then assessed with respect to each element. Samples with the highest scores are considered to be most anomalous and are targeted for further follow-up.

CHI-6*X, NUMCHI, PEG-4 Indices

Smith *et al.* (1989) have noted that there are broad linear trends, termed "chalcophile corridors", throughout the Yilgarn Block in Western Australia. The presence of elevated values of the pathfinder elements within these "corridors", namely As, Sb, Bi, Mo, Ag, Sn, and W form the basis of the empirical indices CHI-6*X, NUMCHI, and PEG-4 as developed by Smith *et al.* (1987) and Smith and Perdrix (1983). These trends show elevated abundances of these pathfinder elements in lateritic materials associated with greenstone belts, shear zones, base metal and precious metal deposits. These indices are based on simple equations as follows:

CHI-6*X= As + 3.56xSb + 10xBi + 3xMo + 30xAg + 30xSn + 10xW + 3.5xSe
PEG-4 = .09xAs + 1.33xSb + Sn + .14xGa + .4xW + .6xNb + Ta.

The coefficients provide weighting to the elements, such that samples with elevated chalcophile abundances have high CHI-6*X or PEG-4 indices. These coefficients were derived for lateritic materials only. The coefficients must be altered for other materials. The CHI-6*X index is suited more to isolating samples with elements associated with precious metal deposits, while the PEG-4 index is suited for isolating samples with elements associated with pegmatophile environments, such as Sn deposits within granitoid terrains.

The NUMCHI index is a score of the number of elements that exceed the threshold for each element. Thus for a given sample, if nine elements exceed their respective thresholds, then the NUMCHI index for that sample will have a value of 9. Threshold values are chosen from visual inspection of summary tables, order statistics, Q-Q plots, etc.

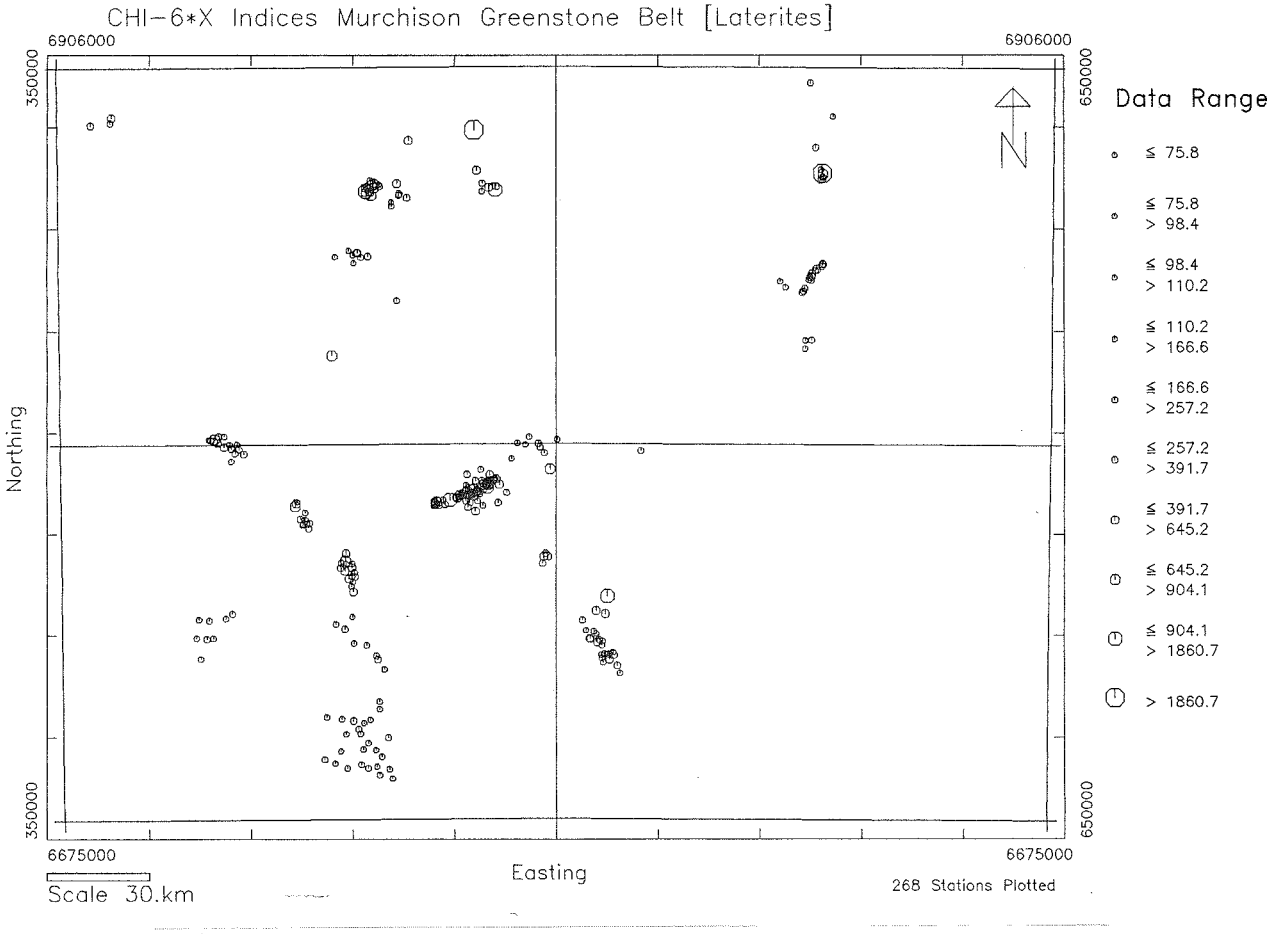


Figure 21a. Map of CHI-6*X scores for Murchison Greenstone area

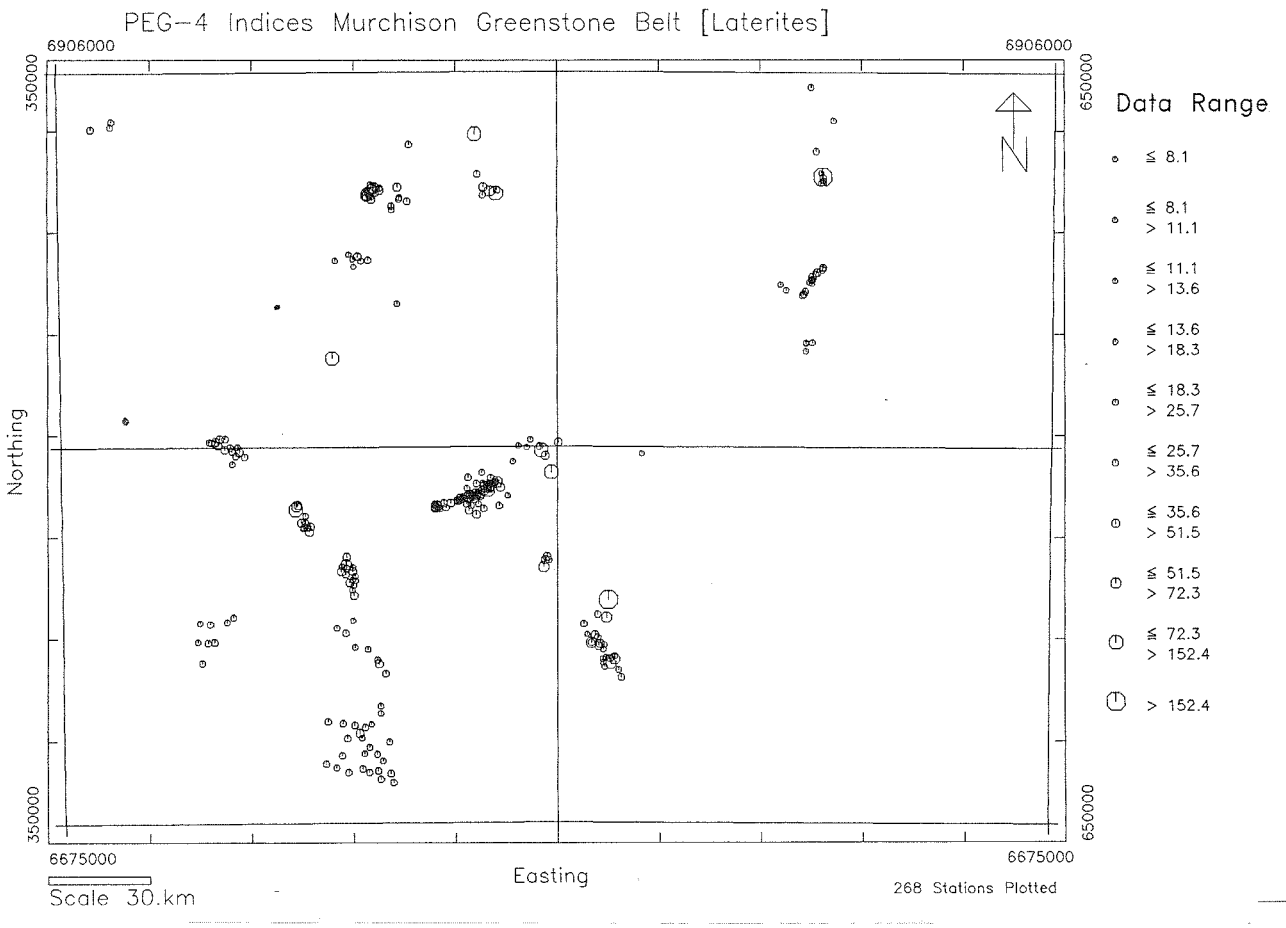


Figure 21b. Map of PEG-4 scores for Murchison Greenstone area

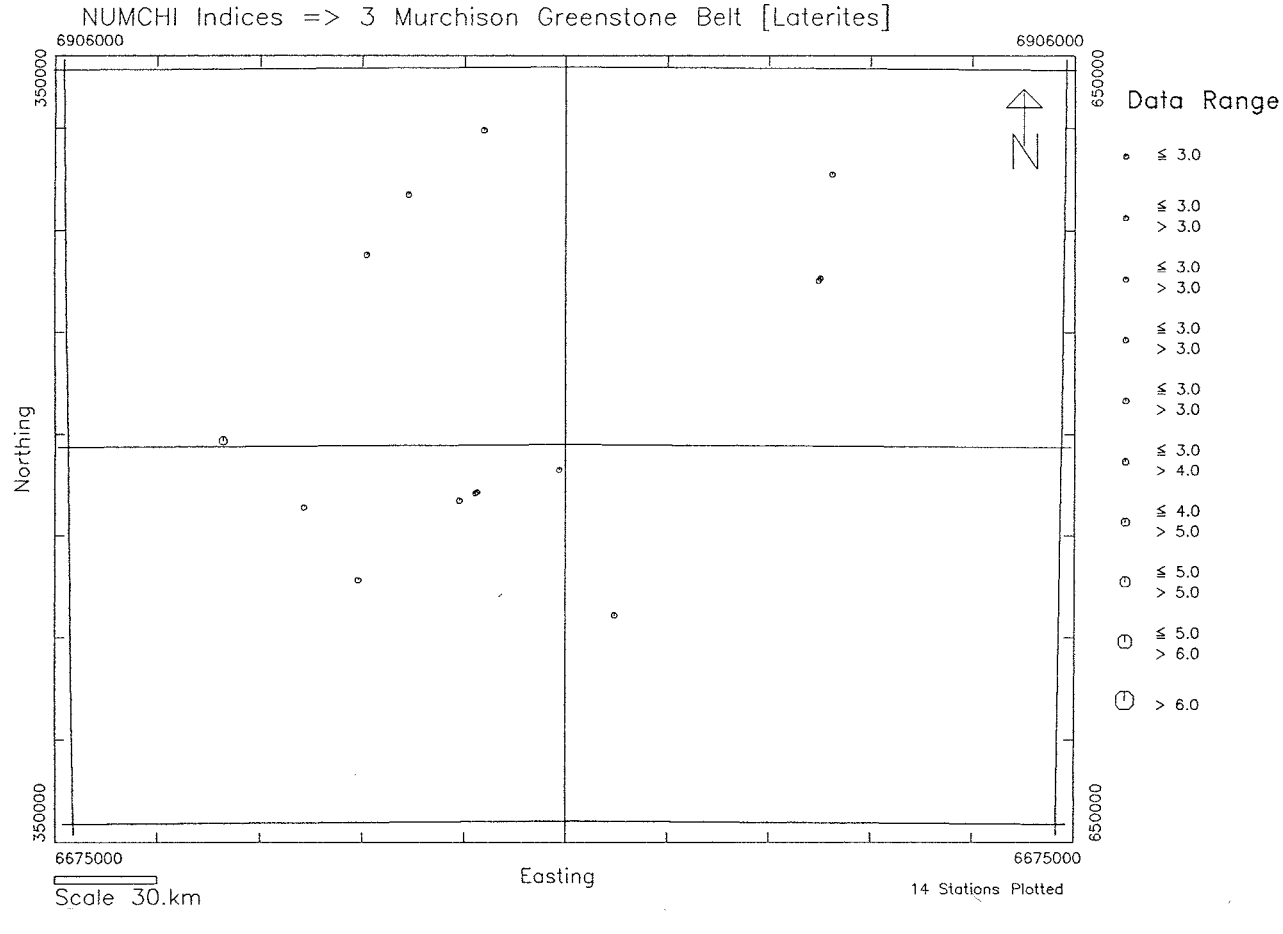


Figure 21c. Map of NUMCHI scores for Murchison Greenstone area

The indices were calculated for the Murchison greenstone data of Dataset 1. The scores for the samples were computed using the equations listed above. The larger scores are the scores of interest. Usually, investigation of CHI-6*X or PEG-4 scores that exceed the 95th or 98th percentile are considered worthy of follow-up. The samples with NUMCHI indices greater than 3 or 4 could be considered worthy of additional investigation.

Figures 21a,b,c are maps showing the CHI-6*X, PEG-4, and NUMCHI indices over part of the area covered by the Murchison greenstone terrain of Dataset 1. Figure 21a shows that several large scores occur throughout the area. These sample sites have elevated abundances of the elements listed in the formula quoted above. These sites would be considered suitable follow-up targets in an exploration programme. Figure 21c shows sample sites with NUMCHI indices of 3 or more elements that exceed thresholds obtained from the 95th percentile level of the regional geochemical data (see Table 2). A comparison of the sample sites with high CHI-6*X, PEG-4, and NUMCHI indices with the Mahalanobis distances of Figure 20 indicates that two of the three prominent sites in Figure 20b match high CHI-6*X, PEG-4 and NUMCHI indices.

Methods such as the CHI-6*X, PEG-4, NUMCHI, and χ^2 plots are useful because they employ the use of pathfinder elements. Methods such as the NUMCHI can assist in choosing a suitable multivariate threshold since it represents an integer number that reflects the number of anomalous elements that are present based on thresholds taken from univariate populations. Significant anomalies from CHI-6*X, PEG-4 indices, and the χ^2 plots can be chosen in conjunction with the NUMCHI index or by taking a conservative threshold such as the 95th percentile ranking of these indices.

5.0 MODELLED APPROACH TO ANOMALY RECOGNITION - THE USE OF STATISTICAL ANALYSIS

5.1 Multivariate Data Analysis: Grouped Data- Target vs. Background

Once structures and systematic processes can be recognized within the data, then additional numerically-based and statistically-based methods can be applied to further enhance the data interpretation. Methods such as regression analysis, canonical variate analysis (discriminant functions), and classification procedures can be applied to test target from background populations. This approach is part of a modelled approach in determining the differences between samples that may be associated with specific geological processes.

Reference groups are clearly-defined groups of samples that represent background and target groups. Unknown samples and populations can be statistically tested against these groups and a measure of similarity can be determined, thus assisting in finding samples that share the characteristics of the "target" populations. This approach has been clearly established and defined by classical statistical methods and has been successfully applied by Smith *et al.* (1984). Methods such as canonical variate analysis are well suited to be applied in areas where appropriate orientation studies have been carried out and the target and background geochemical populations have been well established.

5.2 Analysis of Variance

Selection of suitable pathfinder elements is based upon an understanding of elements that contribute to processes that underly both background and target populations. Incorporation of unimportant elements introduces uncertainty and confusion. Robust estimates of correlations can assist in determining elements that contribute to the multivariate geochemical signature. When groups of samples are being considered, methods such as Analysis of Variance (ANOVA) or Multivariate Analysis of Variance (MANOVA) can be used to determine the significance of an element. Davis (1986, Chapter 2) gives an account of ANOVA, and Cooley & Lohnes (1971) give an account of MANOVA. There are many different methods for analysis of variance which are dependent upon the relationships of the populations being tested. Rock (1988, Topics 11, 12) describes the various types of ANOVA's than can be carried out.

Analysis of variance is based upon testing the differences between populations of data. When only one variable (element) is involved then the procedure is analysis of variance (ANOVA). An evaluation of differences between batches of As analyses from the same set of samples sent to different laboratories can be tested using ANOVA. This is known as a one-way analysis of variance. Multivariate analysis of variance involves the comparisons of groups of samples that are composed of several variables (elements).

The application of MANOVA is particularly important in the use of regression and canonical variate analysis procedures. Before background and target groups can be established, the statistical uniqueness of each group must be tested. If the groups are not sufficiently distinct, then the application of statistical procedures, such as canonical variate analysis, allocation, or typicality procedures cannot be applied effectively. As an example, the laterite samples collected over the Mt. Gibson Au mine of Dataset 3 will be tested against the regional background samples of the Murchison greenstone terrain of Dataset 1.

In order to carry out the MANOVA procedure, the two datasets must be examined for outliers and then a common set of transformation parameters be calculated and applied to both sets of data. Each dataset was trimmed of samples that exceeded the 95th percentile and then values of λ were computed using the procedure outlined by Howarth and Earle (1979). The transformation parameters for both datasets were examined and where they were different, the values of λ were averaged and used for both datasets when the data were transformed.

Table 12 displays the results of the MANOVA. The Table shows the means and standard deviations for the two groups and for the total sample population. The first test of equality of dispersions is a measure of how similar the covariances of the groups are. The F-ratio of 15.088 exceeds the F-value of 1.18 for 171 and 390283 degrees of freedom at the 95% confidence level. From this, it can be concluded that the dispersions between the groups are not equal. In the case of unequal dispersions of covariances, Wilk's Λ must be cautiously interpeted. The Table also displays the mean square values between the two groups and within the groups and the corresponding F-ratio for each variable. For the corresponding degrees of freedom of 1 and 688, the F-value is 3.85. Only Mo and W show no significant difference between the two populations. Elements that appear to be most distinct between the two datasets are Au, Ga, Zn, Pb, and Ag. The Wilk's lambda (Λ) coefficient is a measure of how distinct the groups are. As lambda approaches zero, the groups are increasingly different. The results of the MANOVA on these two datasets indicates that they are sufficiently distinct from each other and can be used in further statistical procedures. Despite the inequality of covariances, the differences between the populations are significant and this can be further demonstrated in the application of Canonical Variate Analysis.

Table 12

MANOVA - MULTIVARIATE ANALYSIS OF VARIANCE

Background / Target Group [18 elements]
Murchison Greenstone Belt/ Mt. Gibson Au Deposit

Analysis for 18 variables and 2 groups

Group 1 NG= 507 Dispersion determinant= .48864819E+17
Group 2 NG= 183 Dispersion determinant= .69646212E+11
Pooled-samples Dispersion determinant= .68301370E+17

Group Means

	Fe	Ag	Mn	Cr	V	Cu	Pb	Zn	Ni	Co	As	Sb	Mo	Sn	Ga	W	Nb	Au
Group 1	37.12	-2.29	8.25	18.92	49.86	5.90	8.99	5.94	14.25	2.84	5.23	1.39	1.49	.52	2.00	1.99	6.39	1.35
Group 2	31.94	-1.16	7.02	16.05	57.16	5.18	12.94	4.12	10.85	2.52	4.39	2.52	1.59	.80	53.60	2.07	5.44	19.86
Total Sample	35.75	-1.99	7.92	18.15	51.79	5.71	10.04	5.46	13.34	2.7	5.01	1.69	1.52	.59	30.38	2.01	6.14	6.26

Standard Deviations

	Fe	Ag	Mn	Cr	V	Cu	Pb	Zn	Ni	Co	As	Sb	Mo	Sn	Ga	W	Nb	Au
Group 1	15.39	1.32	2.98	6.37	20.27	2.58	4.38	1.96	10.18	1.66	2.23	2.22	1.81	1.36	9.65	1.33	4.86	1.76
Group 2	11.09	1.46	1.61	2.30	14.14	1.06	3.11	1.57	2.26	1.34	1.09	1.46	.91	.90	20.69	1.13	2.50	6.53
Pooled samples	14.38	1.36	2.68	5.59	18.84	2.28	4.08	1.86	8.81	1.58	1.99	2.05	1.62	1.26	13.48	1.28	4.37	3.68

For test of h1 (equality of dispersions), M= 2680.320 and F= 15.088
For F ,N.D.F.1=171 N.D.F.2= 390283
Expected F(95%,171,390283)= 1.18

Univariate F-ratios,with NDF1= 1 and NDF2= 688

Variable	Among Mean square	Within Mean square	F-ratioquare
Fe	3597.16	206.85	17.39
Ag	172.97	1.85	93.47
Mn	202.95	7.20	28.18
Cr	1106.02	31.22	35.42
V	7177.42	355.11	20.21
Cu	69.64	5.21	13.38
Pb	2095.81	16.64	125.97
Zn	442.92	3.47	127.61
Ni	1553.05	77.65	20.00
Co	13.18	2.49	5.28
As	95.91	3.95	24.26
Sb	169.70	4.20	40.36
Mo	1.21	2.63	.46
Sn	10.61	1.58	6.70
Ga	134264.11	181.75	738.75
W	.67	1.64	.41
Nb	122.13	19.06	6.41
Au	46065.13	13.55	3399.61

Wilks Lambda= .1349

F-ratio for h2, Overall discrimination= 239.09
Ndf1= 18 and NDF2= 671
Expected F(95%,18,671)= 1.62

5.3 Regression Methods

Regression methods are usually applied when it is desired to model a predictor variable, that is an element of interest with elements that may or may not have an association with that element. Regression procedures have been used in exploration geochemistry for many applications. A review of regression applications can be found in Howarth and Sinding-Larsen (1983:255) and Rock (1988:252). Davis (1986, Chapter 4) gives a concise account of regression procedures. Regression is commonly employed to systematically characterize the background multi-element response through a linear model. Thus, samples that do not fit within the model, that is samples with high residuals, may be considered atypical and possibly anomalous.

The definition of the formula for linear regression is of the form:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p ,$$

where y is the dependent variable, $\beta_0,...,\beta_p$ are coefficients, and x_1 to x_p are the independent predictor variables. In matrix form the linear regression model can be expressed as:

$$Y = X\beta + E$$

where X is a matrix of independent variables, β is a matrix of coefficients, and E is a matrix of residuals. Variations in regression analysis include stepwise analysis which tests the significance of each of the variables in the regression calculations and iteratively deletes variables that do not add significantly to the regression model. Standard methods of regression are calculated using ordinary least squares (OLS) which means minimizing the difference between the observed value of Y from the predicted value of Y.

Methods such as stepwise backward elimination or forward stepwise regression methods are regression methods that delete or add variables to a regression model based on the overall improvement in the sum of squares explained by the regression versus the total sum of squares. These methods may be more desirable to use as they select the best set of elements that provide a consistent linear model.

Regression procedures make some assumptions that may make it difficult to justifiably apply them in geochemical data analysis. The first assumption is that the dependent variable has no associated errors. Clearly, this is not the case in geochemical data analysis. Secondly, regression models assume that the predictor variables are not strongly correlated. If they are, then procedures, such as ridge regression techniques should be employed. van Gaans and Vriend (1990) have published a ridge regression program that is also based on robust estimators.

Regression procedures also fail when the relationship between the variables is not linear. The most important aspect of applying regression procedures is that the results be studied carefully and that a number of regression techniques be applied so as to be certain that the regression is meaningful. In many regression-based studies, it is not always clear which are the dependent variables and which are the independent predictor variables.

Stanley and Sinclair (1987) used regression procedures to characterize anomalous samples from background samples using a reduced major axis regression technique. This technique differs from standard regression methods by regressing Y on X and X on Y. The rationale for using this method was based on the assumption that the predictor variables have errors that are not precisely known and thus OLS methods cannot be applied to the data.

The example of regression shown here was applied to the regional laterite geochemical data applied over the Murchison greenstone belt. In this particular case, Zn was chosen as the dependent variable and Fe, Mn, Cr, V, Cu, Pb, Ni, Co, and Ga as the independent predictor variables. A regression of this type exemplifies two of the problems encountered with analysis of these types of variables. Firstly, the independent variables have errors associated with them and secondly, there may be strong correlations between the variables. Correlation coefficients are shown in Table 9. Table 13a displays the results of the analysis. The program computes regression coefficients for both the ordinary least squares and the ridge regression methods. The first part of the Table provides mean values, standard deviations, squared multiple correlation coefficients R^2, the determinant of the correlation matrix (see Table 9), the eigenvalues, and the condition number.

The multiple R^2 coefficients are a measure of correlation between each predictor (independent) variable and all of the other predictor variables. It provides an assessment of how redundant a variable might be. The determinant, eigenvalues, and condition number indicate if high intercorrelations between the independent variables will interfere with the regression. If the condition number exceeds 100, then strong multicollinearity will exist. In this example, the condition number is 12.56. This suggests that the correlations between the variables are not high and will not distort the regression.

The Table also shows for each variable, the coefficients, standard error, normalized coefficients, contribution to the mean square error, and the contribution to the overall goodness of fit R^2. Multiple regression also includes a measure of how good the linear model is with the data. By computing the sum of squares of the total variation, the sum of squares calculated from the regression equation, and the sum of squares calculated from the residuals, an F-test can be applied to test if the regression equation is statistically significant. In this

Table 13a
Multiple Regression
Murchison Greenstone Belt
Box-Cox Transformations

NO. OF SAMPLES 484
NO. OF VARIABLES 10

	Zn	Fe	Mn	Cr	V	Cu	Pb	Ni	Co	Ga
Mean	3.348	14.86	4.740	6.820	30.97	4.112	7.214	3.671	3.514	7.158
St.Dev.	0.655	4.714	1.057	1.051	10.71	1.336	3.029	1.141	2.330	2.121
R^2	.468	.268	.493	.316	.368	.111	.624	.498	.233	

R^2 are the squared multiple correlations among the predictors

Determinant of the Predictor Correlation Matrix: .08240

Eigenvalues 2.99 1.39 1.21 1.03 .79 .59 .44 .33 .24

Condition Number: 12.56 [Measure of collinearity]

LEAST SQUARES SOLUTION

REGRESSION COEFFICIENTS
INTERCEPT 1.0370E+00 1.7870E-01

	COEFFICIENT	STANDARD ERROR	NORM. COEFF.	CONTRIBUTION TO MEAN SQUARE ERROR	CONTR. TO R^2
Fe	3.3894E-02	5.3517E-03	.2441	.0015	.1252
Mn	1.5609E-01	2.0341E-02	.2522	.0011	.1175
Cr	5.1923E-02	2.4576E-02	.0834	.0016	.0313
V	-4.3638E-03	2.0763E-03	-.0714	.0012	-.0092
Cu	1.7503E-01	1.7315E-02	.3574	.0012	.2037
Pb	4.0327E-02	6.4414E-03	.1866	.0009	.0411
Ni	-2.7625E-02	2.6293E-02	-.0481	.0021	-.0205
Co	5.4198E-02	1.1138E-02	.1929	.0016	.0972
Ga	-3.4285E-02	9.9007E-03	-.1111	.0010	.0393

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	APP. F-TEST
TOTAL VARIATION	206.8848	483		
REGRESSION	129.4235	9	14.3804	87.9962
RESIDUAL	77.4613	474	.1634	1.0000

STANDARD ERROR OF THE Y ESTIMATE = 4.0425E-01
GOODNESS OF FIT (R^2) = .6256
CORRELATION COEFFICIENT R = .7909

Table 13b
Multiple Regression
Murchison Greenstone Belt
Box-Cox Transformations

NO. OF SAMPLES 484
NO. OF VARIABLES 6

	Zn	Fe	Mn	Cu	Pb	Co
Mean	3.348	14.86	4.740	4.112	7.214	3.514
St.Dev.	0.655	4.714	1.057	1.336	3.029	2.330
R^2		.193	.1327	.306	.059	.248

R^2 are the squared multiple correlations among the predictors

Determinant of the Predictor Correlation Matrix: .571

Eigenvalues 1.83 1.20 .802 .704 .459

Condition Number: 4.00 [Measure of collinearity]

LEAST SQUARES SOLUTION

REGRESSION COEFFICIENTS

INTERCEPT 5.9017E-01 1.1251E-01

REGRESSION COEFFICIENTS

	COEFFICIENT	STANDARD ERROR	NORM. COEFF.	CONTRIBUTION TO MEAN SQUARE ERROR	CONTR. TO R^2
Fe	3.3335E-02	4.5212E-03	.2401	.0011	.1231
Mn	2.0059E-01	1.9433E-02	.3240	.0010	.1510
Cu	2.0558E-01	1.7197E-02	.4198	.0012	.2393
Pb	3.2340E-02	6.5153E-03	.1497	.0009	.0330
Co	6.0247E-02	1.9360E-02	.1050	.0011	.0447

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	APP. F-TEST
TOTAL VARIATION	206.8848	483		
REGRESSION	122.2634	05	24.4527	138.1256
RESIDUAL	84.6214	478	.1770	1.0000

STANDARD ERROR OF THE Y ESTIMATE = 4.2075E-01
GOODNESS OF FIT (R SQUARED) = .5910
CORRELATION COEFFICIENT R = .7687

example, the F-ratio is 87.9. At the 95% confidence level, for 484 samples and nine variables (degrees of freedom) the corresponding F-value is approximately 1.8. This indicates that it is unlikely that the regression coefficients are not significant. The F-ratio of 87.9 also says that the variation (sum of mean squares = 14.3804) of the regression is 87.9 times the variation (sum of squares = 0.1634) of the residuals or values not explained by the regression. The goodness of fit, R^2, is 0.6256 indicating that 62.5% of the variation of the data is accounted for by the regression equation.

The example presented here is for nine independent variables on Zn. However, not all of these variables may be important in the model. The low coefficients for V (-0.004), Ni (-0.002), and Ga (-.003) as well as their corresponding low R^2 contributions might suggest that these variables may not contribute significantly to the overall regression model.

To illustrate the point about the lack of significance of some of the variables, the multiple regression was repeated on Zn using a group of five elements, Fe, Mn, Cu, Pb, and Co. The results of the regression are shown in Table 13b. An examination of the analysis of variance shows that for these five elements the regression has a total sum of squares of 122. Compared to a sum of squares of 129 for nine variables, this suggests that these five variables are the most significant in terms of a regression model. Similarly, the F-ratio of 138 is higher than in Table 13a, also indicating that the regression model is 138 times more likely to be significant relative to a random chance. The R^2 coefficient is only slightly less than the R^2 coefficient for the nine variables. Thus, the linear regression model is more likely to be appropriate for the five element group of variables.

5.4 All Possible Subsets

The selection of suitable pathfinder elements can be determined from procedures, such as ALL-POSSIBLE-SUBSETS, which allow the selection of elements that give maximum separation between target and background groups while selecting only those elements which are most useful in the separation.

The all possible subsets procedure (McCabe, 1975) has been developed so that it determines the Wilk's likelihood ratio criterion:

$$\Lambda = |W| / |W+B|$$

where W is the within-groups sums of squares and products. B is the between-groups sums of squares and products. The likelihood ratio is computed for all possible subsets of variables. From the resulting values of Λ for each group of possible subsets, a group of variables can be chosen that indicates maximum group separation with a minimum number of variables.

For this example, five populations were selected. Background Group 1 is comprised of the Murchison greenstone belt area of Dataset 1 and Background Group 2 is a suite of samples collected over a the Archaean Albany-Fraser granitoid gneiss terrain that also occurs within the Yilgarn Block. Three target populations were also included for comparison. Target Group 3 is comprised of laterite samples collected over the Mt. Gibson Au deposit represented by Dataset 3, Target Group 4 is the suite of samples collected from the Golden Grove lateritic blanket, Dataset 4, and Target Group 5 is comprised of a suite of lateritic samples collected over the Lawlers area Au deposit that is also situated in the Yilgarn Block.

Table 14 displays a partial output from an all possible subsets program developed by Campbell (1986a). The procedure was applied to four subsets of elements in order to determine the minimum number of elements required to maximize the discrimination between the groups. The subsets were chosen using geochemical/geological considerations. The first subset used all 18 elements. This subset naturally gives maximum discrimination. Table 14a shows that for one element, Au gives maximum discrimination between the groups. As elements are added, the negative log ratio of Wilk's likelihood ratio increases to a maximum of 5.30. Table 14b shows a partial list of the all possible subsets using six chalcophile elements, three background elements, and Au. Group separation increases to a log ratio value of 4.47. Table 14c shows the results applied to six chalcophile elements only, and Table 14d shows the results applied to four commodity elements. Tables 14c,d both show a log ratio considerably less than the combinations of elements shown in Tables 13a,b.

Although Table 14a gives maximum discrimination between the groups, a subset of elements composed of fewer elements may be just as effective at discriminating the groups. One way of examining the differences between the subset choices is to plot the cumulative log ratio score versus the number of elements. This is shown in Figure 22. The Figure consists of four curves representing the four subsets that were analyzed. The curve that contains all 18 elements shows a steady increase in the log ratio starting with Au and ending with Ag. Notice that the curve rises steeply until Fe is included. At this point, the curve flattens out and suggests that the additional elements Cu, Co, Mn, Ni, Sb, Zn, W, Mo, and Ag do little to increase the separation between the groups. The elements, Au, Cr, As, Ga, Sn, Nb, Pb, and V appear to be useful discriminators for the five groups.

The curve for the log ratios of group separation for the six chalcophile elements, three background elements, and Au follows a similar path to the full 18 element in Figure 22. The differences between the two groups shows up with the use of Cu, Sb, and Zn as discriminators. The full 18 element set suggests that V, Fe, and Cu might be better at discrimination than Cu, Sb, or Zn. Nonetheless, the curves are nearly identical and

suggest that Au, Cr, As, Ga, Sn, Nb, and Pb maximize the differences between the groups. The log ratio curves of the six chalcophile elements and the four commodity elements display steep slopes, but do not reach the same level of group separation as the other two curves.

The analysis indicates that the subset of six chalcophile elements, three background elements, and Au provides good group separation without losing much information. Thus, these elements would be good to use in an exploration programme where unknown samples could be compared with the reference groups in order to detect samples associated with target group characteristics.

Table 14

All Possible Subsets [Box-Cox Transformed Data]

Group 1 456 samples Murchison Greenstone terrain
Group 2 451 samples Albany-Fraser Granite/Gneiss terrain
Group 3 173 samples Mt.Gibson Au deposit [felsic]
Group 4 100 samples Golden Grove Au Cu sulphide deposit
Group 5 48 samples Lawlers Au deposit [mafic]

Table 14a

18 Elements [Transformed]
Fe Ag Mn Cr V Cu Pb Zn Ni Co As Sb Mo Sn Ga W Nb Au

|W| / |W+B|	Log Ratio	Element(s)
0.174947	1.7432709	Au
0.858654E-01	2.4549742	Au Cr
0.497756E-01	3.0002301	Cr Au As
.		
.		
.		
0.498526E-02	5.3012705	Fe Mn Cr V Cu Zn Ni Co As Sb Mo Sn Ga Nb Au Pb W Ag

Table 14b
6 Chalcophile Elements + 3 Background Elements + Au [Transformed]

Cr Cu Pb Zn As Sb Sn Ga Nb Au

|W| / |W+B|	Log Ratio	Element(s)
0.174947	1.7432709	Au
0.858654E-01	2.4549742	Au Cr
0.497756E-01	3.0002301	Cr Au As
.		
.		
.		
0.132957E-01	4.3203163	Cr As Sn Ga Nb Au Pb Cu
0.121458E-01	4.4107680	Cr Cu As Sn Ga Nb Au Pb Sb
0.114211E-01	4.4722939	Cr Cu As Sn Ga Nb Au Pb Sb Zn

Table 14c
6 Chalcophile Elements [Transformed]

Cu Pb Zn As Sb Ga

|W| / |W+B|	Log Ratio	Element(s)
0.267145	1.3199652	Ga
0.139957	1.9664201	Ga As
0.944823E-01	2.3593428	As Ga Zn
0.720437E-01	2.6304827	As Ga Zn Pb
0.588621E-01	2.8325582	As Ga Zn Pb Sb
0.497424E-01	3.0008972	As Ga Zn Pb Sb Cu

Table 14d
4 Commodity Elements [Transformed]

Cu Pb Zn Au

|W| / |W+B|	Log Ratio	Element(s)
0.174947	1.7432709	Au
0.119741	2.1224275	Au Cu
0.101373	2.2889445	Cu Au Pb
0.883679E-01	2.4262469	Cu Pb Au Zn

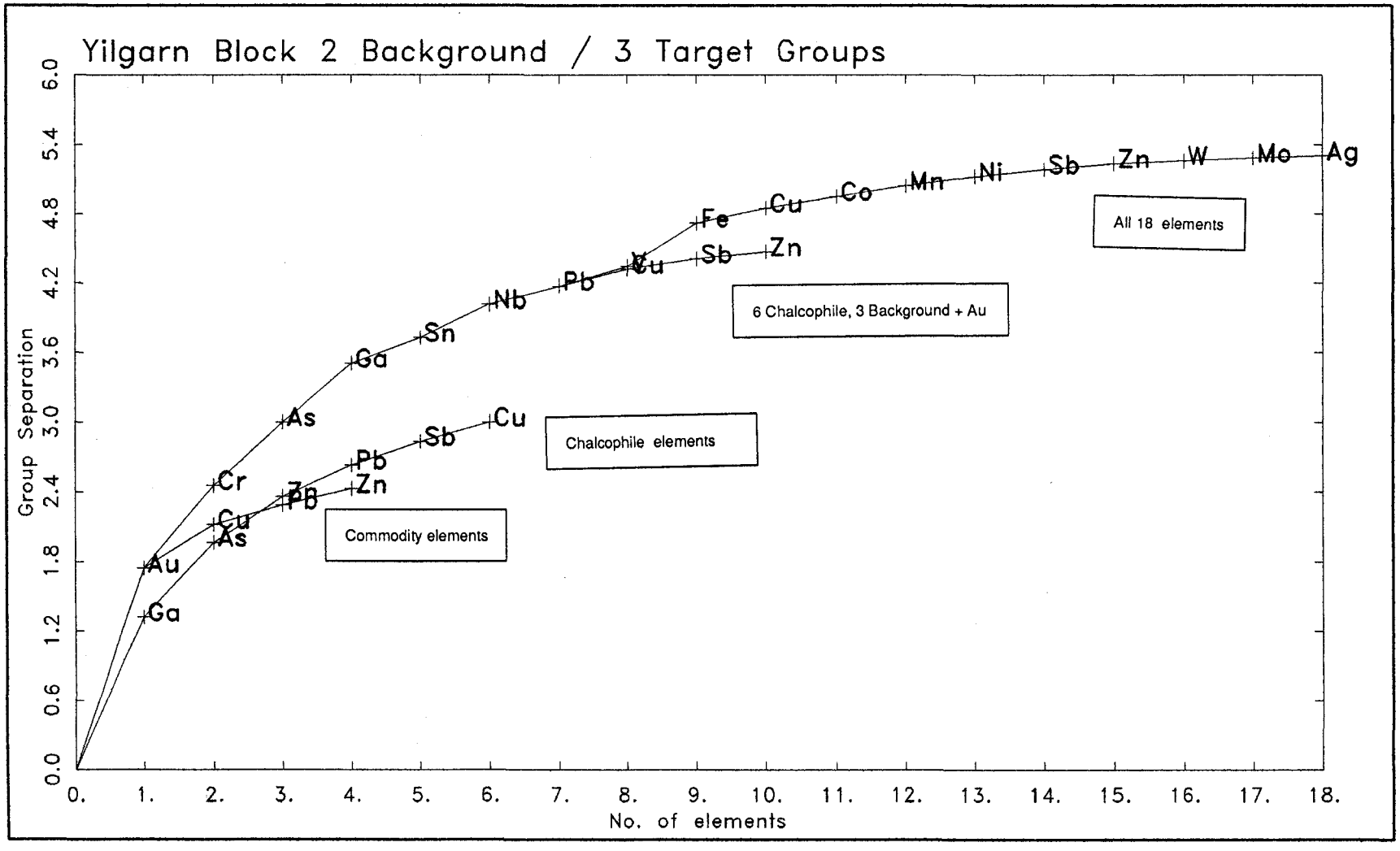


Figure 22. Plot of Log ratio separation vs. Number of elements for ALL POSSIBLE SUBSETS computations.

5.5 Canonical Variate Analysis

Statistical procedures, including canonical variate analysis (CVA) (multiple discriminant functions analysis) have been used for anomaly recognition by Smith *et al.* (1984). The most common application being the statistical comparison between background groups (regional lithological variation) and target groups (groups representing mineralization) from which reference populations are created with distinct geochemical attributes. CVA and the associated linear discriminant functions test the statistical uniqueness of the reference groups and indicate whether the distinctions between the groups are significant. A description of the method can be found in LeMaitre (1982), Cooley and Lohnes (1971), and Srivastava and Carter (1983). Caution must be applied using the CVA procedure. Samples that do not share any characteristics with the reference populations will still be classified into one of the discriminant reference groups. This implies that the samples that comprise the reference groups must be carefully chosen. In the analysis of geochemical data, the preceding methods of outlier detection and elimination and appropriate transformations help to ensure that atypical samples will not be included in the reference groups.

The method is based on the following procedure. Given a data matrix X, composed of N samples and m variables, three m by m matrices are computed that are composed of the following relationships:

T matrix: sum of squares and products of all samples and variables, (total covariance of all the data) defined as:

$$t_{ij} = \sum_{k=1}^{N} (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j),$$

where

$\bar{x}_i$ is the total sample mean of the ith major element oxide,

$\bar{x}_j$ is the total sample mean of the jth major element oxide,

x_{ik} is the value of the ith major element oxide of the kth observation in the total sample,

x_{jk} is the value of the jth major element oxide of the kth observation in the total sample.

A matrix: sum of squares and products of samples among each reference group (covariance among the different reference groups), defined as:

$$a_{ij} = \sum_{k}^{g} (\bar{x}_{ik} - \bar{x}_i)(\bar{x}_{jk} - \bar{x}_j),$$

where g is the number of reference groups.

W matrix: sum of squares and products of samples within each reference group (covariance within each reference group).

$$w_{ij} = \sum_{k=1}^{g} \sum_{m=1}^{N_k} (x_{imk} - \bar{x}_{ik})(x_{jmk} - \bar{x}_{jk}),$$

where N_k is the number of samples in each reference group,

x_{imk} is the ith major element oxide of the mth observation in the kth reference group,

x_{jmk} is the jth major element oxide of the mth observation in the kth reference group,

$\bar{x}_{ik}$ is the mean value of the ith major element oxide in the kth reference group,

$\bar{x}_{jk}$ is the mean value of the jth major element oxide in the kth reference group.

Given two or more distinct reference groups, then

$$\mathbf{T = A + W}$$

The method of canonical variate analysis is based on maximizing the ratio of the variation among the groups (**A**) to the variation within the groups (**W**). The number of canonical roots that can be obtained is g-1, where g is the number of groups.

The test statistic Wilk's Λ is a measure of similarity of the three reference groups. This statistic measures the ratio of variation within the reference groups (**W**) to the total variation over all of the groups (**T**). The lower the number, the less likely that the groups share common characteristics. The test assumes equality of the reference groups covariance matrices. Frequently, this is not the case and the Wilk's Λ test must be interpreted with caution. Campbell (1984a) outlines the CVA procedure for unequal covariances. Despite the problem of unequal covariances, the resulting discriminant functions will still separate the populations if they are significantly unique. Plots of the discriminant scores onto the canonical variate axes will show if the dispersions of the populations overlap and provide a visual assessment as to group separation. The χ^2 values provide a measure of the significance of each discriminant function. Values of χ^2 that are less than an accepted level of significance (99%) indicate that the function may not be significant. The Λ value associated with each function indicates how different the reference groups are with respect to the discriminant function. The trace indicates how much discriminating information is obtained from the discriminant function.

The use of the F-test allows the comparison of the similarities of the three groups. By assuming a level of significance and determining the number of degree of freedom, a critical value of F can be determined. If the F-test score is less than the critical value, then it is likely, within the level of significance, that the groups are similar. If the F-test score is greater than the critical value, then as the F-test score increases, the more distinct the groups become. The discriminant function coefficients (DF) provide the coefficients for functions that separate the reference groups. The canonical variate (factor pattern) scores (CV) for each function give the coordinates in the discriminant function space for each variable and provide the relative relationships of the variables.

Canonical variate procedures developed by Campbell (1986a) ensure that the covariances of the groups are based upon robust estimates and thus significantly reduce the effects of outliers. In the analysis of geochemical data, at the stage of applying CVA, it is assumed that reference groups have been established which have no outliers and that the data have been transformed to normality. Thus, robust estimation is not critical at this stage.

In CVA, there are various measures and tests which are applied that indicate how significant the discriminant functions are. These are described as follows. The value of Wilk's Λ, as in the all possible subsets analysis, is a measure of how distinct the groups are based on the determinant of the sums of squares representing variation between the groups from variation within the groups. As $\Lambda \rightarrow 0$, the groups are increasingly distinct. The F-ratio of overall discrimination between the groups is a measure that provides a means of determining whether the distinction between the groups is meaningful at a given confidence level. By determining the F-value at a given confidence interval (95%) and given degrees of freedom the significance of the groups can be ascertained. The canonical roots can be assessed for their significance by determining how much each linear combination of variables contributes to the distinction between the groups. The canonical correlation coefficient, R, is a measure of how well each discriminant function separates the groups. As R increases, the ability of the function to separate the groups increases. Generally, for each successive function in an analysis the value of R decreases. The χ^2 values are a measure of significance of discrimination for each discriminant function. Similarly, each successive discriminant function will be less significant and have a lower χ^2 value. The value of Λ for each discriminant function is also a measure of how well each function separates the groups. Each function is represented by a linear combination of elements given by the discriminant function coefficients. For each function, the factor pattern coefficients provide a measure of how well a variable correlates with that function. They are also similar to the loadings of principal components analysis as they provide the relative relationships of the variables in the discriminant function space. The measures of communalities indicate how much of the variable is accounted for by the full set of linear discriminant functions. Variables with low communalities are not important in describing group differences. The group centroids are useful when plotting the discriminant function scores for the samples in the discriminant function space.

The examples provided here are from two background reference group datasets and three target reference group datasets as discussed in the all possible subsets section previously.

Testing the Background Populations

CVA was applied to the two background datasets (Murchison and Albany-Fraser areas) to test their statistical uniqueness based on the combination of ten elements which was selected from the previous selection of all possible subsets (Table 14). The results of the CVA are shown in Table 15. Because there are only two groups, there is only one discriminant function. The various tests of the discriminant function indicate that it is significant. The canonical coefficient R of 0.880 suggests that the function is reasonably good at separating the two groups. The χ^2 test shows a large value of 1341 which exceeds the χ^2 value of 18.3 for ten degrees of freedom. The value of Wilk's Λ is 0.2254 and indicates a good degree of separation between the groups. The communalities of the variables show that Cr, Cu, Zn, As, and Ga are important in distinguishing the two groups and that Pb, Sn, and Au make virtually no contribution in distinguishing between the two groups.

Testing the Target Populations

Table 16 shows the results of applying CVA to the 3 taget populations. Wilk's Λ of 0.021 indicates that the groups are very distinctive from each other. The F-value of 101.73 exceeds the F-test value of 1.43 for 36 and 618 degrees of freedom. The canonical R of 0.956 for the first function indicates that the function discriminates between the groups very well. Similarly, the second function with R, 0.87 also separates the groups very well. Wilk's Λ for the first function is very low relative to the value of Λ of the second function and indicates that the first function is by far the best discriminator between the groups. The factor pattern of the first function indicates that Cu, Pb, As, Sb, Sn, Ga, and Au are associated with each other and negatively associated with Cr, Zn, and Nb. The communalities of the variables show that Cr, Pb, As, Sb, Sn are important to the discrimination

Table 15
Multiple Group Discriminant Analysis
2 Background Groups [Transformed]
Murchison Greenstone Belt/Albany Fraser Granitoids

Wilks Λ=.2254
F-ratio for overall discrimination = 192.49
ndf1= 16 and ndf2= 896
Expected F(95%,16,896)= 1.65

Chi-square tests with successive roots removed

Roots removed	Canonical R	Eigenvalue	χ^2	n.d.f.	Λ	% Trace
0	0.880	3.437	1341.	10	0.2254	100.00

	Function Coefficients	Factor Pattern for Disciminant Function	Communalities for Discriminant Function
Cr	0.0354	0.785	0.6154
Cu	0.0982	0.654	0.4271
Pb	-0.0743	-0.093	0.0087
Zn	0.1084	0.675	0.4559
As	0.1414	0.642	0.4123
Sb	0.0417	0.513	0.2632
Sn	-0.0032	0.014	0.0002
Ga	0.2370	0.765	0.5857
Nb	-0.0731	-0.362	0.1313
Au	0.0918	0.250	0.0625

Centroids for groups in 1 dimensional discriminant space
Group 1 0.8748
Group 2 -0.8845

Table 16
Multiple Group Discriminant Analysis
3 Target Groups [Transformed]
Mt. Gibson/Golden Grove/Lawlers

Wilks Λ=.0208
F-ratio for overall discrimination = 101.73
ndf1= 36 and ndf2= 618
Expected F(95%,36,618)= 1.43

Chi-square tests with successive roots removed

Roots removed	Canonical R	Eigenvalue	χ^2	n.d.f.	Λ	% Trace
0	0.956	10.666	1213.	20	0.0208	77.41
1	0.870	3.112	443.	9	0.2432	22.59

	Function Coefficients		Factor pattern for disciminant function		Communalities for Discriminant Function
	1	2	1	2	
Cr	0.0301	0.0342	0.576	0.545	0.6291
Cu	-0.0362	0.1915	-0.188	0.258	0.1023
Pb	-0.0307	-0.3172	-0.493	-0.682	0.7077
Zn	0.0206	0.0487	0.295	0.409	0.2548
As	-0.1008	0.4598	-0.753	0.519	0.8365
Sb	-0.2138	-0.0737	-0.768	0.213	0.6349
Sn	-0.3006	0.0377	-0.894	0.051	0.8011
Ga	-0.0053	0.0650	-0.030	-0.121	0.0155
Nb	0.1138	-0.0441	0.497	-0.100	0.2569
Au	-0.0014	-0.0307	-0.460	-0.332	0.3215

Centroids for groups in 2 dimensional discriminant space

	1	2
Group 1	0.6000	-0.5895
Group 2	-1.4160	0.0868
Group 3	0.7877	1.9436

between the groups, while Cu, Zn, Ga, Nb, and Au are less important variables in the discrimination. Relative to the background groups, Cr appears to be important in distinguishing all of the groups and Pb, As, Sb, and Sn are more important in distinguishing between the target groups.

The increase in the significance of Pb, As, Sb, and Sn is to be expected and supports the use of chalcophile elements as useful pathfinder elements for the recognition of various mineral deposits.

Testing the Background and Target Populations

The background and target groups were tested together using CVA. In this example, the results of the CVA are shown for the full set of 18 elements, ten elements [background + chalcophile], six chalcohpile elements, and the four commodity elements. This example will illustrate the importance of choosing the correct elements for the discrimination of populations.

18 Element CVA

Table 17 lists the results of the CVA applied to the full set of 18 elements for both the target and background groups. The value of Wilk's Λ of 0.0051 indicates that the distinctions between the groups are very significant.

Table 17
Multiple Group Discriminant Analysis
3 Target Groups/ 2 Background Groups [Transformed]
18 elements
Target: Mt. Gibson/Golden Grove/Lawlers
Background: Murchison Greenstone/Albany-Fraser

Wilks Λ=.0051
F-ratio for overall discrimination = 96.13
ndf1=140 and ndf2= 4744
Expected F(95%,140,4744)= 1.21

Chi-square tests with successive roots removed

Roots removed	Canonical R	Eigenvalue	χ^2	n.d.f.	Λ	% Trace
0	0.966	14.039	6426.	72	0.0051	74.90
1	0.852	2.638	132.	51	0.0760	14.07
2	0.798	1.755	1562.	32	0.2767	9.36
3	0.487	0.312	330.	15	0.7624	1.66

	Discriminant Function Coefficients				Factor pattern for disciminant functions				Communalites for 4 Discriminant Functions	
	1	2	3	4	1	2	3	4		
Fe	0.0549	-0.1093	-0.1200	-0.1427	0.586	0.204	-0.432	0.068	0.5765	Fe
Ag	0.0041	-0.0055	0.0872	0.0758	0.457	0.257	0.115	0.134	0.3057	Ag
Mn	-0.0088	0.1755	0.0129	-0.3950	-0.315	0.523	-0.126	-0.176	0.4196	Mn
Cr	-0.0011	0.0567	-0.0210	0.0442	0.092	0.741	-0.380	-0.067	0.7068	Cr
V	-0.0431	0.0685	0.0825	0.0500	-0.025	0.620	0.077	0.107	0.4019	V
Cu	0.0055	0.1010	-0.0892	0.6081	0.373	0.411	-0.329	0.376	0.5577	Cu
Pb	0.0087	-0.1273	0.1287	0.2634	0.408	-0.218	0.306	0.368	0.4438	Pb
Zn	-0.0150	0.1151	-0.0525	0.147	-0.067	0.487	-0.464	0.180	0.4890	Zn
Ni	-0.0553	0.0341	0.1941	-0.6537	-0.210	0.620	-0.120	-0.249	0.5045	Ni
Co	0.0452	-0.0738	-0.0738	-0.159	0.404	0.173	-0.346	-0.269	0.3852	Co
As	0.0394	-0.0017	-0.3766	-0.1225	0.674	0.045	-0.495	-0.011	0.7018	As
Sb	0.0818	-0.0440	0.0335	0.0224	0.788	0.076	-0.144	0.016	0.6472	Sb
Mo	-0.0037	0.0415	0.0418	0.218	-0.003	0.151	0.064	0.232	0.0805	Mo
Sn	0.0847	-0.1843	-0.0488	0.071	0.573	-0.339	-0.115	0.024	0.4569	Sn
Ga	0.0850	0.1610	-0.0249	0.016	0.800	0.408	0.138	0.034	0.8264	Ga
W	0.0063	-0.0003	-0.0977	-0.162	0.417	-0.014	-0.180	-0.184	0.2405	W
Nb	-0.0738	-0.0093	0.0205	-0.1174	-0.436	-0.065	0.199	-0.175	0.2644	Nb
Au	0.0364	0.0063	0.0828	-0.0504	0.888	0.011	0.365	-0.081	0.9280	Au

Centroids for groups in 4 dimensional discriminant space

	1	2	3	4
Group 1	-0.2401	0.5150	-0.6791	0.3583
Group 2	-0.7740	-0.7098	0.3416	-0.2190
Group 3	0.9251	0.9490	1.5632	0.1534
Group 4	2.5829	-1.5888	-0.6251	0.0843
Group 5	0.8387	1.6669	-1.0891	-2.0749

Table 18
Multiple Group Discriminant Analysis
3 Target Groups/ 2 Background Groups [Transformed]
10 Elements [Chalcophile + Background + Au]
Target: Mt. Gibson/Golden Grove/Lawlers
Background: Murchison Greenstone/Albany-Fraser

Wilks Λ=.0116
F-ratio for overall discrimination = 135.81
ndf1= 76 and ndf2= 4605
Expected F(95%,76,4605)= 1.28

Chi-square tests with successive roots removed

Roots removed	Canonical R	Eigenvalue	χ^2	n.d.f.	Λ	% Trace
0	0.958	11.117	5438.	40	0.0116	77.10
1	0.828	2.183	2396.	27	0.1402	15.14
2	0.706	0.996	984.	16	0.4463	6.91
3	0.330	0.123	141.	7	0.8908	0.85

	Discriminant Function Coefficients				Factor pattern for disciminant functions				Communalites for 4 Discriminant Functions
	1	2	3	4	1	2	3	4	
Cr	-0.0019	-0.0627	-0.0419	0.0489	0.139	-0.837	0.032	0.097	0.7299
Cu	0.0207	-0.1183	0.0896	-0.6084	0.391	-0.486	0.228	-0.498	0.6889
Pb	0.0116	0.1741	-0.0388	-0.3951	0.411	0.361	-0.101	-0.544	0.606 1
Zn	0.0024	-0.1529	-0.0064	0.0235	-0.051	-0.646	0.230	-0.230	0.5259
As	0.0417	-0.0932	0.5812	0.2987	0.636	-0.191	0.604	0.175	0.8370
Sb	0.1006	0.1040	0.1109	0.0456	0.781	-0.068	0.257	0.066	0.6853
Sn	0.0856	0.1943	0.3205	-0.0428	0.526	0.314	0.392	0.080	0.5346
Ga	0.0963	-0.1573	-0.0596	-0.0787	0.852	-0.260	-0.189	-0.060	0.8326
Nb	-0.0888	-0.0420	-0.1462	0.0805	-0.426	0.107	-0.263	0.185	0.2965
Au	0.0437	0.0307	-0.0822	0.0668	0.915	0.208	-0.216	0.116	0.9401

Centroids for groups in 4 dimensional discriminant space

	1	2	3	4
Group 1	-0.2350	-0.7599	0.3944	-0.2284
Group 2	-0.8198	0.7263	-0.1405	0.1402
Group 3	1.2206	-0.0934	-1.4469	-0.1690
Group 4	2.2460	1.2723	1.3041	0.0214
Group 5	0.8567	-1.9197	0.0718	1.4166

The value of F is 96.13 and far exceeds the F-statistic of 1.21 for 140 and 4744 degrees of freedom. Because there are five groups, there are four discriminant functions. The first three functions all show high canonical correlations and indicate that they are good at discriminating between the groups. The fourth function is less significant at group discrimination as indicated by the low canonical R and the high Wilk's Λ for that function. The communalities of the variables indicate that Cr, As, Sb, Ga, and Au are the most significant variables in the discrimination.

The relationships of the groups can be described graphically by plotting the discriminant function score of each sample for each function into the discriminant function space. The scores of the samples for the CV1 vs. CV2 and CV1 vs. CV3 are shown in Figures 23a,b. Each sample is labelled with the group with which it is associated. Figure 23a shows that the Yilgarn background groups form two distinct clouds of points with some degree of overlap. This is not unexpected as many samples in the greenstone terrain can also relate to granitic or felsic gneisses. The three target groups plot away from the two background populations and indicate their distinctiveness. The two Au deposits merge in Figure 23a; however, they are quite distinct in the projection of the samples onto the CV1-CV3 plane of Figure 23b. The samples associated with the massive sulphide deposit are quite distinct from any of the other groups. The first two discriminant functions clearly show the group separation and indicate that the populations of samples are sufficiently unique.

A graphical presentation of the factor patterns can assist in visualizing the associations of the elements with the different groups. Figure 23c shows the factor pattern of the variables for the first two discriminant functions and Figure 23d shows the pattern for the first and third discriminant functions. Both Figures describe similar relationships of the variables with respect to the groups of samples. The locations of the variables indicate the relative associations of the samples with the element. These relative associations can be observed in Figure 23a, where the relative position of Group 1 indicates that it is relatively enriched in Mn, Ni, V, Zn, and Cr while Group 2 is relatively enriched in Nb. The target groups show relative enrichment of the chalcophile elements, Au, and Fe.

10 Element CVA

The results of the CVA for the three background elements + six chalcophile elements + Au are shown in Table 18. The F-ratio and Wilk's Λ both indicate that this suite of elements is almost as good at separating the groups as the full set of elements shown by the analysis in Table 17. This had already been determined by the all possible subsets calculations of Table 14 and Figure 22. The first three discriminant functions provide for 99% of the total discrimination of the groups. The communalities of the variables show that Au, As, Ga, Cr, Cu, Pb, and Sb are the most important variables in the discrimination of the groups. Figures 24a,b show the sample scores plotted onto the CV1-CV2 and CV1-CV3 axes respectively.

In both of the Figures, the five groups are clearly distinct. The samples of the massive sulphide deposit of Group 4 stand out as being very different from those of the other groups. The mafic association of the Au deposit of Group 5 is shown by the proximity of Group 5 and Group 1 samples. Group 3 samples which are from the Au deposit associated with both mafic and felsic volcanics show a proximity to the more felsic portion of the Group 1 cloud.

Figures 24c,d show the relationships of the variables plotted onto the CV1-CV2 and CV1-CV3 axes. The chalcophile elements display affinity for the three target groups in Figure 24c. Zinc and Cr show affinity with the Murchison greenstone samples of Group 1 and Nb shows association with the Albany-Fraser granite/gneiss samples of Group 2. Figure 24d displays similar features, although As appears to be more distinctly associated with the Golden Grove dataset.

6 Element CVA

Table 19 shows the results of the CVA applied to the six element subset of samples. The value of Wilk's Λ, and the F-value indicate that the groups are significantly different. However, in this analysis, only the first discriminant function is really significant. The other functions have significantly-lower canonical correlation coefficients and the corresponding Wilk's Λ for each successive function is large. The communalities of the variables are all high and indicate that all of the variables are required to distinguish between the groups. Figures 25a,b show the discriminant scores of the samples plotted onto the CV1-CV2 and CV1-CV3 axes respectively. In both Figures, the distinction between the Murchison greenstone samples (Group 1) and the Albany-Fraser granite/gneiss samples (Group 2) is present, but there is considerable overlap of the target groups among themselves and with the Group 1 greenstone samples. The target groups show a trend of separation along the CV2 axis, but there is considerable overlap among them. Group 5, in particular, shows dispersion throughout the entire range of the samples along the CV2 axis. Figures 25c,d show the relative relationships of the variables along the CV1-CV2 and CV1-CV3 axes. Figure 25c shows that Zn is associated with the more mafic samples in the Group 1 greenstones, Cu is associated with the massive sulphide samples of Group 4, As, Sb, and Ga are associated with the three target Groups 3, 4 and 5, and Pb is associated with the Au deposit

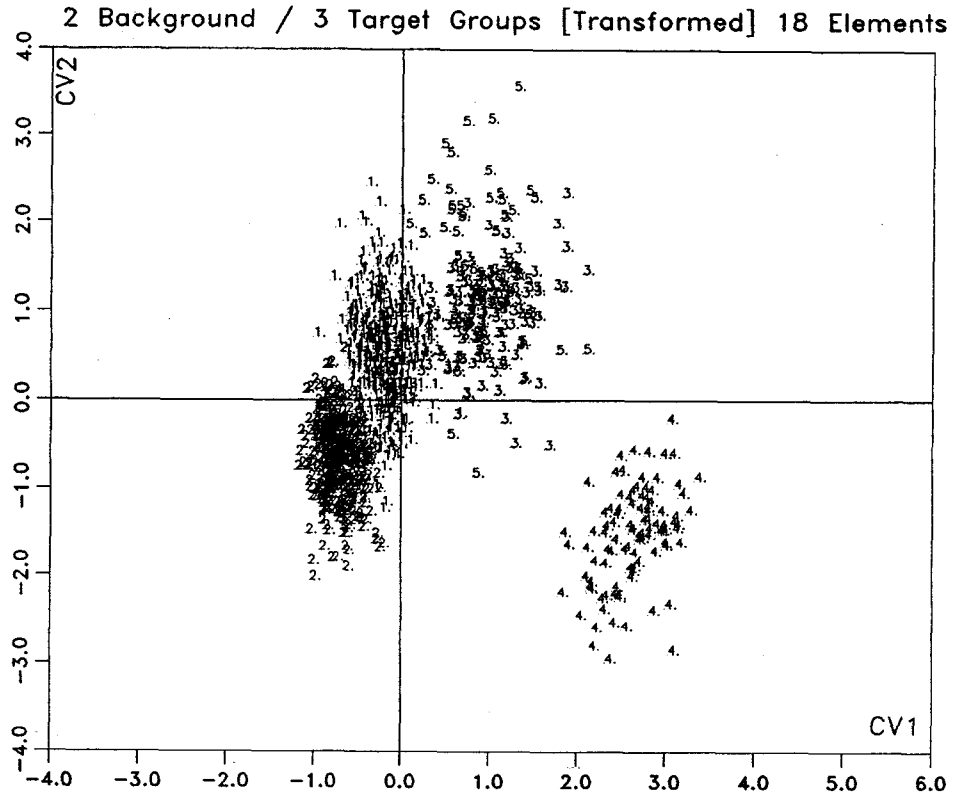


Figure 23a. Plot of discriminant function scores onto the first two canonical variate axes of the 2 background and 3 target groups for the full set of 18 elements. Each sample is labelled with the group from which it originates. Group 1 is the Murchison Greenstone belt dataset, Group 2 is the Albany-Fraser granite/gneiss terrain dataset, Group 3 is the data from the Mt. Gibson Au deposit in the Murchison greenstone belt, associated with felsic and mafic volcanics, Group 4 is the data associated with the Golden Grove massive sulphide deposit, and Group 5 is the data from with the Lawlers area Au deposit associated with mafic rocks.

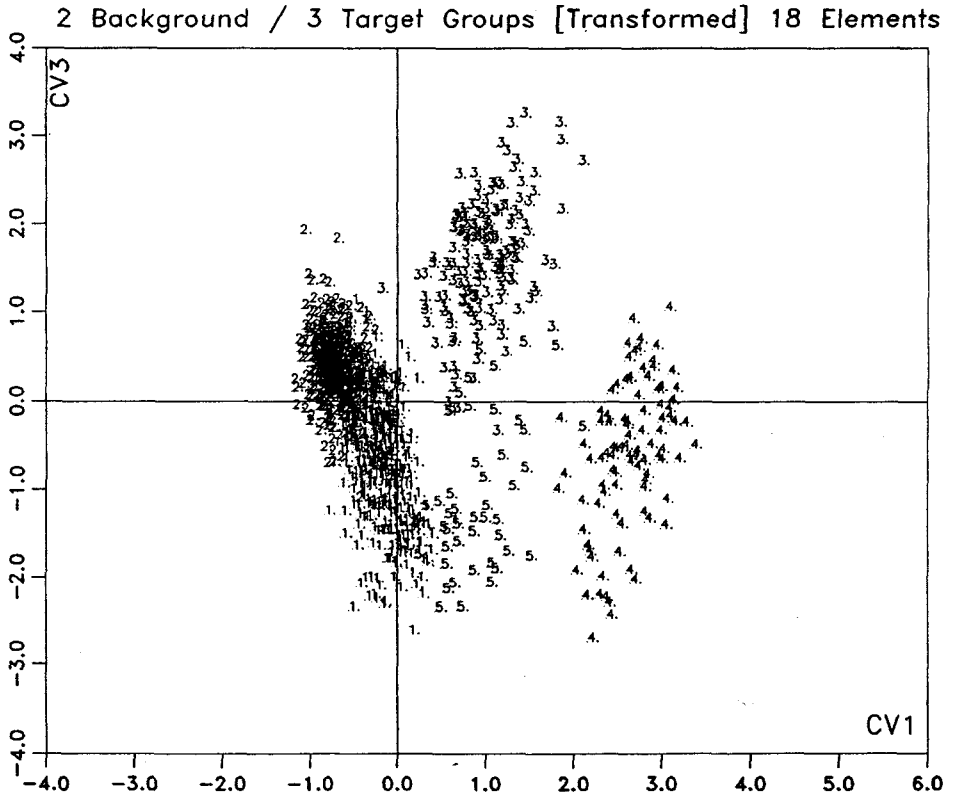


Figure 23b. Plot of discriminant function scores onto the first and third canonical variate axes of the 2 background and 3 target groups for the full set of 18 elements.

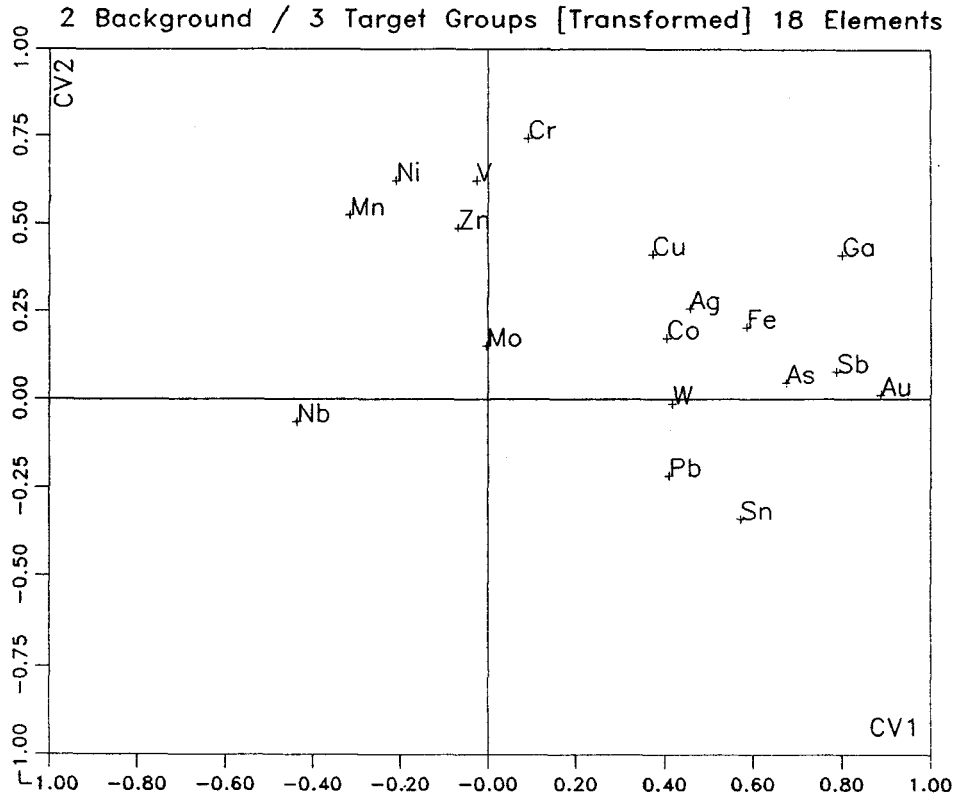


Figure 23c. Plot of the factor pattern scores projected on to the first and second canonical variate axes. The relative positions of the variables indicate the relative association of the samples to the variables in Figure 23a.

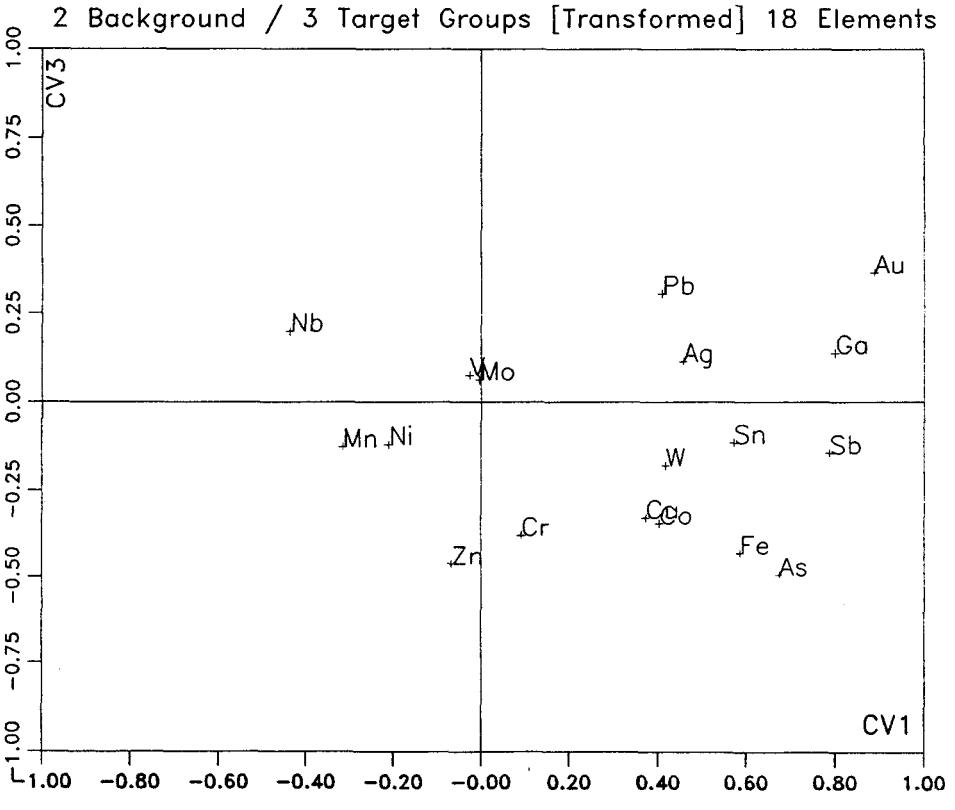


Figure 23d. Plot of the factor pattern scores projected on to the first and third canonical variate axes. The relative positions of the variables indicate the relative association of the samples to the variables in Figure 23b.

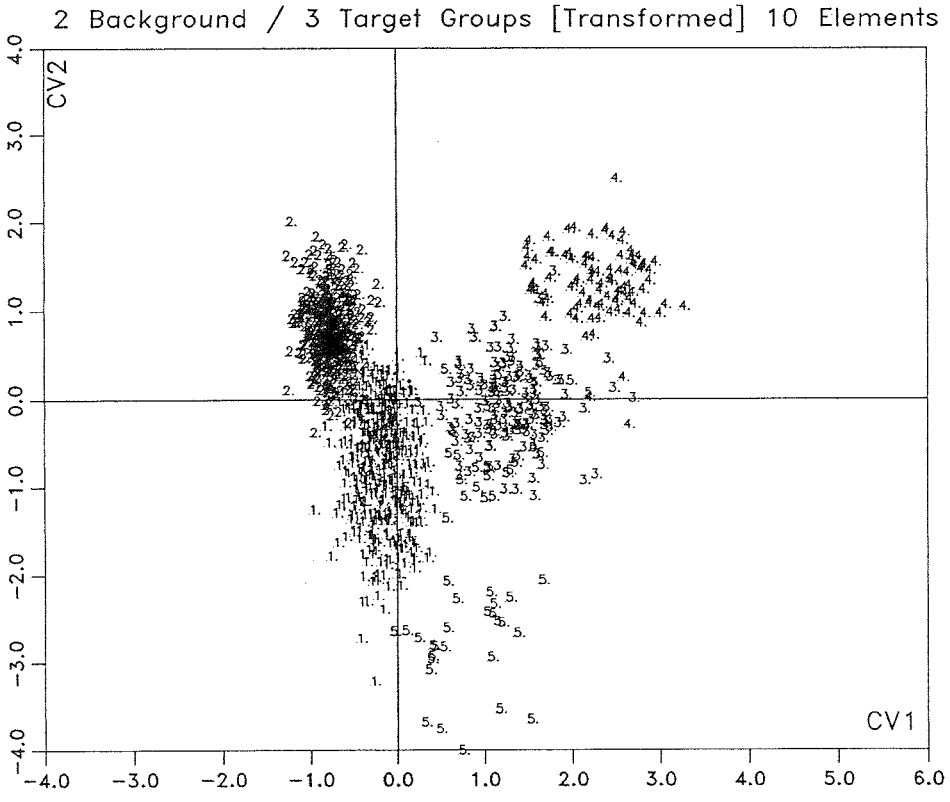


Figure 24a. Plot of discriminant function scores onto the first and second canonical variate axes of the 2 background and 3 target groups for the subset of 10 elements [3 Background + 6 Target + Au]. The separation between the 5 groups is as good as that shown in Figure 23.

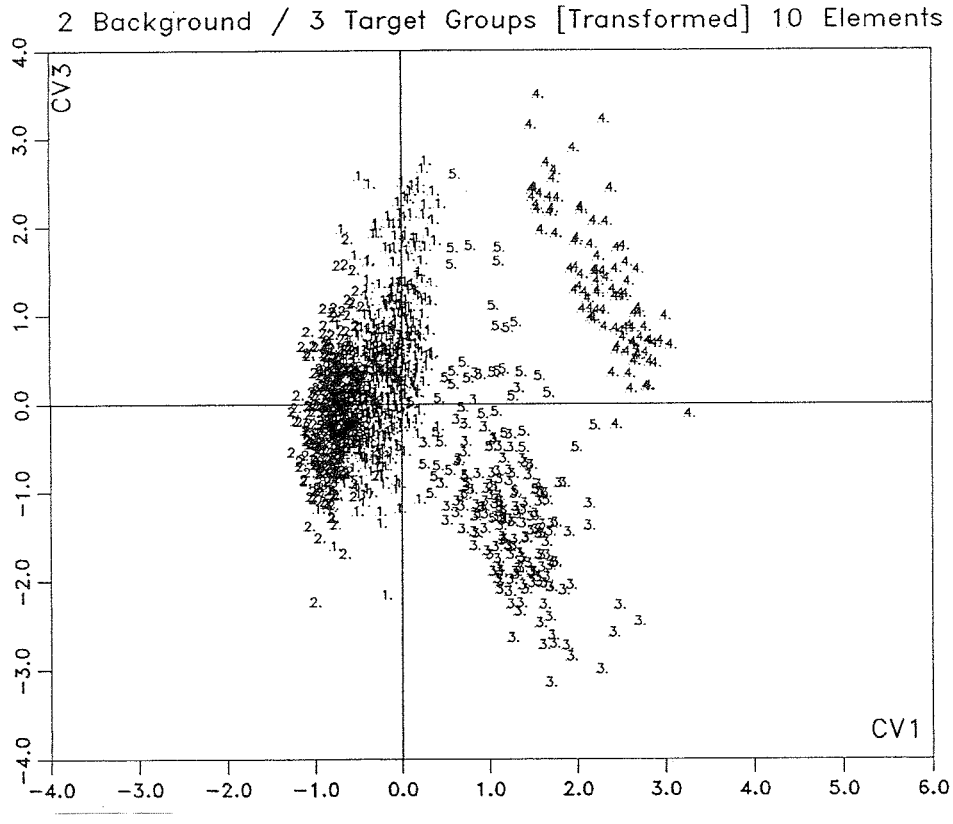


Figure 24b. Plot of discriminant function scores onto the first and third canonical variate axes of the 2 background and 3 target groups for the subset of 10 elements [3 Background + 6 Target + Au]. The separation between the 5 groups is as good as that shown in Figure 23.

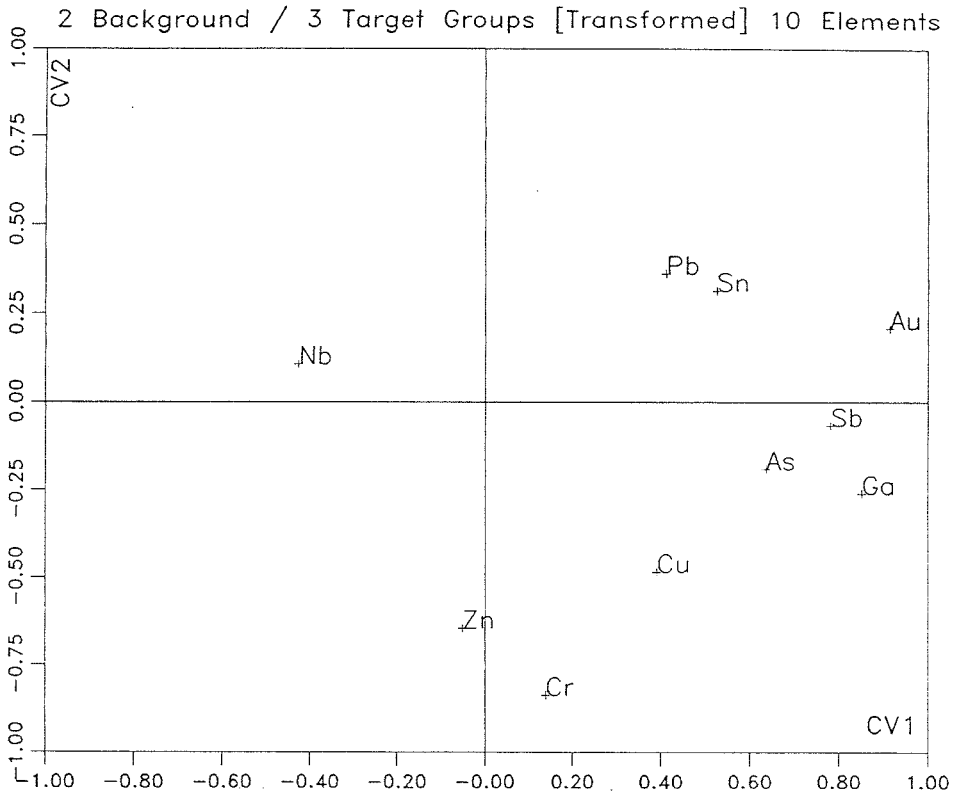


Figure 24c. Plot of the factor pattern scores projected on to the first and second canonical variate axes. The relative positions of the variables indicate the relative association of the samples to the variables in Figure 24a.

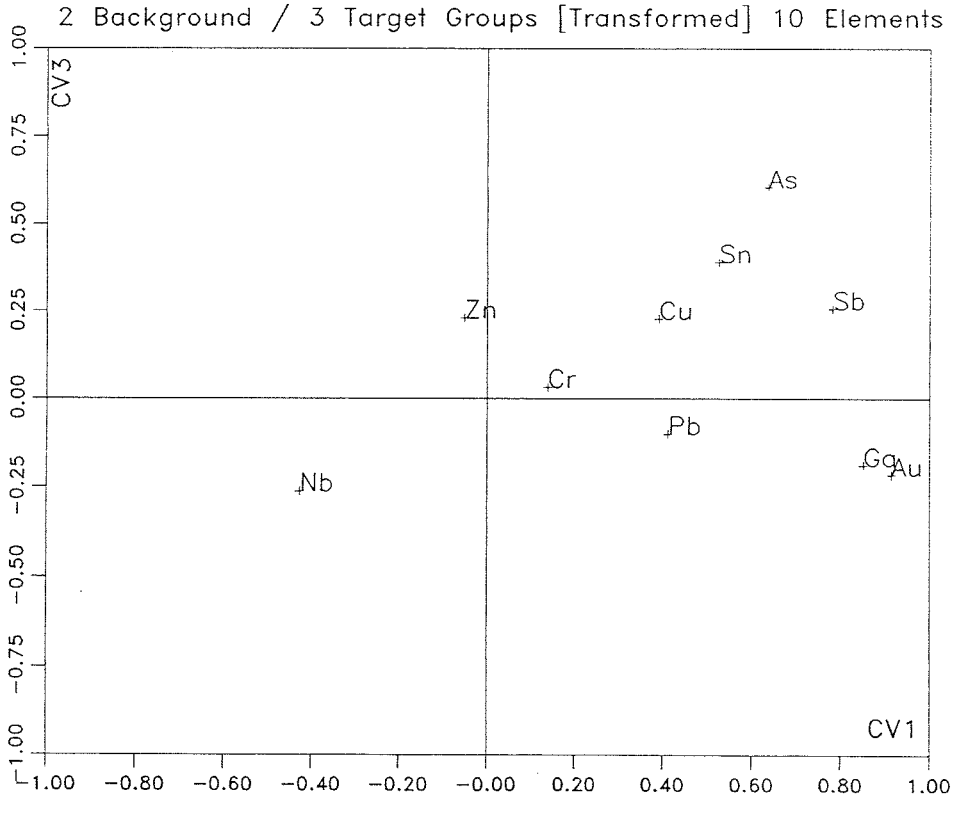


Figure 24d. Plot of the factor pattern scores projected on to the first and third canonical variate axes. The relative positions of the variables indicate the relative association of the samples to the variables in Figure 24b.

target Groups 3 and 5. Figure 25d indicates that As and Sb are associated with Groups 1 and 5, which are more mafic in character than the other three groups. Gallium appears to show an association with the Mt. Gibson samples.

The increased dispersion and overlap of samples from the CVA applied to the six element chalcophile suite of elements suggest that it is not nearly as good at distinguishing the groups as the ten element set. This demonstrates the importance of choosing a valid set of elements on which to distinguish background from target samples in an exploration programme.

4 Element CVA

The results of CVA applied to the chalcophile suite of four elements are shown in Table 20. The value of Wilk's Λ for the overall discrimination indicates that the group separation is good. However, only the first discriminant function contributes significantly to the group separation as indicated by the canonical R coefficients and the

Table 19
Multiple Group Discriminant Analysis
3 Target Groups/ 2 Background Groups [Transformed]
6 Chalcophile Elements
Target: Mt. Gibson/Golden Grove/Lawlers
Background: Murchison Greenstone/Albany-Fraser

Wilks Λ=.0502
F-ratio for overall discrimination = 131.14
ndf1= 44 and ndf2= 4250
Expected F(95%,44,4250)= 1.38

Chi-square tests with successive roots removed

Roots removed	Canonical R	Eigenvalue	χ^2	n.d.f.	Λ	% Trace
0	0.919	5.415	3655.	24	0.0502	78.33
1	0.694	0.929	1384.	15	0.3220	13.44
2	0.570	0.482	582.	8	0.6211	6.97
3	0.282	0.086	101.	3	0.9204	1.25

	Discriminant Function Coefficients				Factor pattern for disciminant functions				Communalites for 4 Discriminant Functions
	1	2	3	4	1	2	3	4	
Cu	0.1544	-0.0618	0.0985	0.6091	0.535	-0.407	-0.077	0.626	0.8506
Pb	-0.0226	0.3344	0.0274	0.4320	0.311	0.601	0.122	0.519	0.7422
Zn	-0.0133	-0.4500	0.0466	0.0179	0.136	-0.776	-0.099	0.381	0.7762
As	0.1612	-0.0950	-0.7730	-0.0855	0.721	-0.040	-0.663	0.016	0.9606
Sb	0.1664	0.3421	-0.1206	-0.1760	0.810	0.210	-0.256	-0.055	0.7687
Ga	0.1682	-0.0905	0.2266	-0.1222	0.902	0.089	0.341	-0.151	0.9610

Centroids for groups in 4 dimensional discriminant space

	1	2	3	4
Group 1	0.0908	-0.6876	-0.1467	0.2246
Group 2	-1.0091	0.4032	-0.0369	-0.1175
Group 3	0.9937	0.4054	1.2205	0.0082
Group 4	1.7844	1.3129	-1.1219	0.0627
Group 5	1.3192	-1.4527	-0.3215	-1.1901

Table 20
Multiple Group Discriminant Analysis
3 Target Groups/ 2 Background Groups [Transformed]
4 Commodity Elements
Target: Mt. Gibson/Golden Grove/Lawlers
Background: Murchison Greenstone/Albany-Fraser

Wilks Λ=.0893
F-ratio for overall discrimination = 160.41
ndf1= 28 and ndf2= 3727
Expected F(95%,28,3727)= 1.48

Chi-square tests with successive roots removed

Roots removed	Canonical R	Eigenvalue	χ^2	n.d.f.	Λ	% Trace
0	0.914	5.085	2953.	16	0.0893	86.90
1	0.629	0.655	745.	9	0.5436	11.19
2	0.314	0.109	129.	4	0.8996	1.86
3	0.047	0.002	3.	1	0.9978	0.04

	Discriminant Function Coefficients				Factor pattern for disciminant functions			
	1	2	3	4	1	2	3	4
Cu	-0.0306	0.3816	-0.5890	-0.6633	-0.262	0.846	-0.384	-0.259
Pb	-0.0804	-0.1679	-0.5213	0.3060	-0.500	-0.266	-0.676	0.470
Zn	0.0312	0.3566	0.3069	0.6766	0.215	0.843	-0.017	0.490
Au	-0.0915	0.0129	0.0769	0.0206	-0.990	0.047	0.119	-0.033

Centroids for groups in 4 dimensional discriminant space

	1	2	3	4
Group 1	0.4961	0.6352	-0.1926	0.0011
Group 2	0.5344	-0.7142	0.0854	-0.0061
Group 3	-1.4005	-0.0455	0.0715	0.0905
Group 4	-2.0981	-0.1692	-0.2917	-0.1060
Group 5	-0.3158	1.1932	1.3777	-0.0587

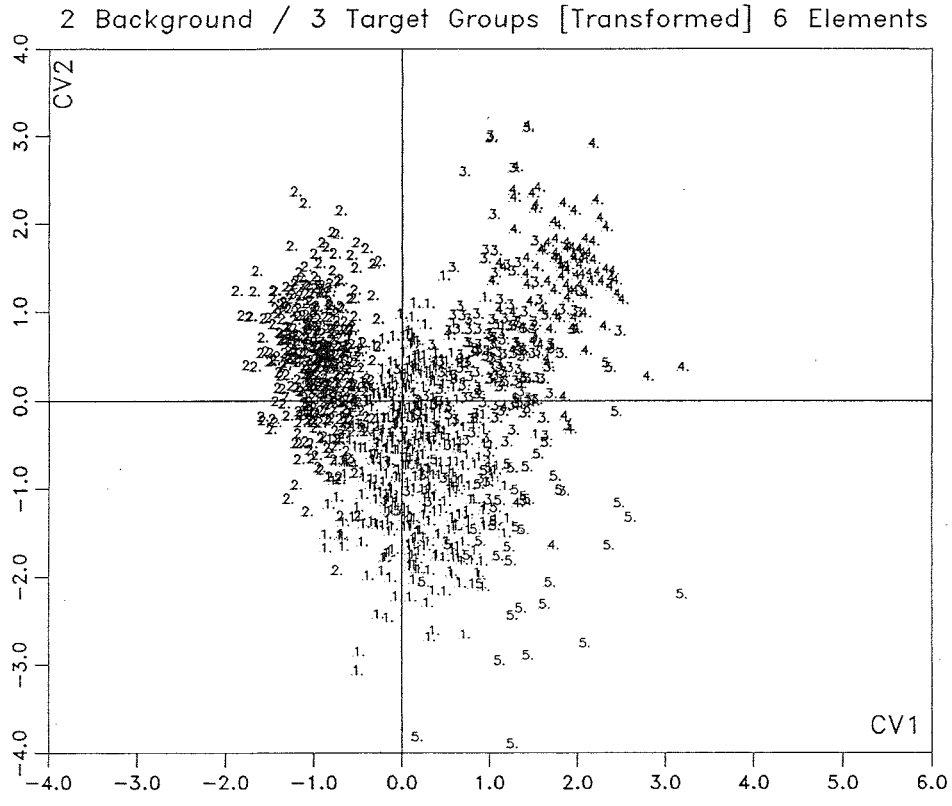


Figure 25a. Plot of discriminant function scores onto the first and second canonical variate axes of the 2 background and 3 target groups for the subset of 6 chalcophile elements [Cu Pb Zn As Sb Ga]. Note the good separation between background populations but poor separation of target groups from the Murchison greenstone population.

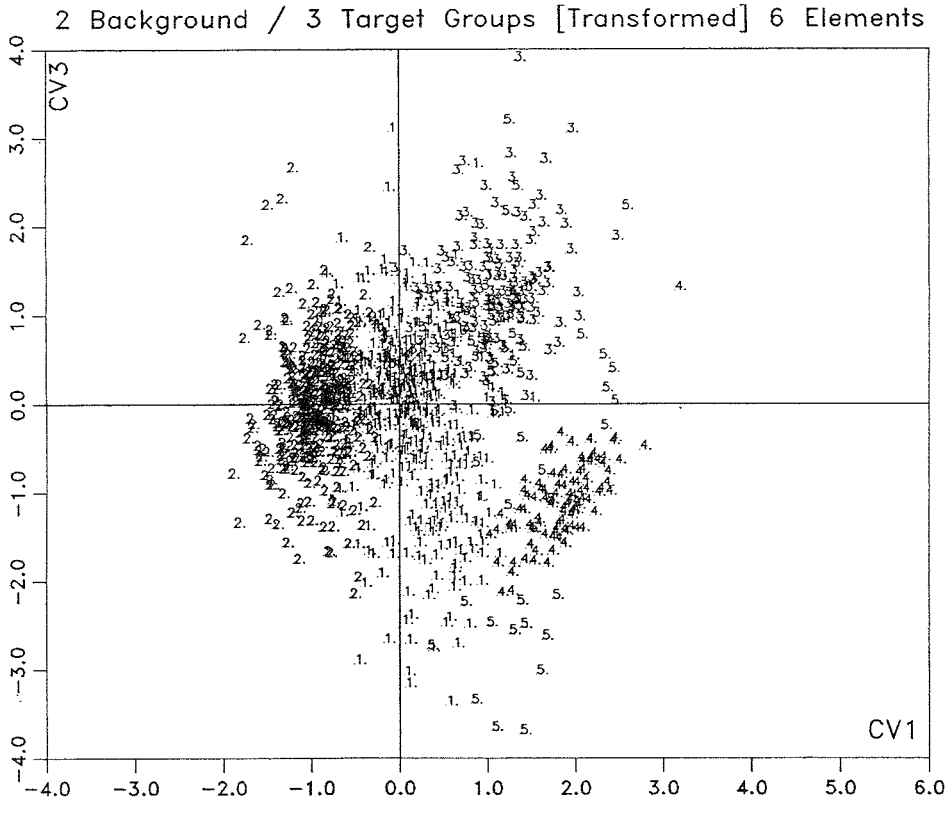


Figure 25b. Plot of discriminant function scores onto the first and third canonical variate axes of the 2 background and 3 target groups for the subset of 6 chalcophile elements [Cu Pb Zn As Sb Ga]. Note the good separation between background populations but poor separation of target groups from the Murchison greenstone population.

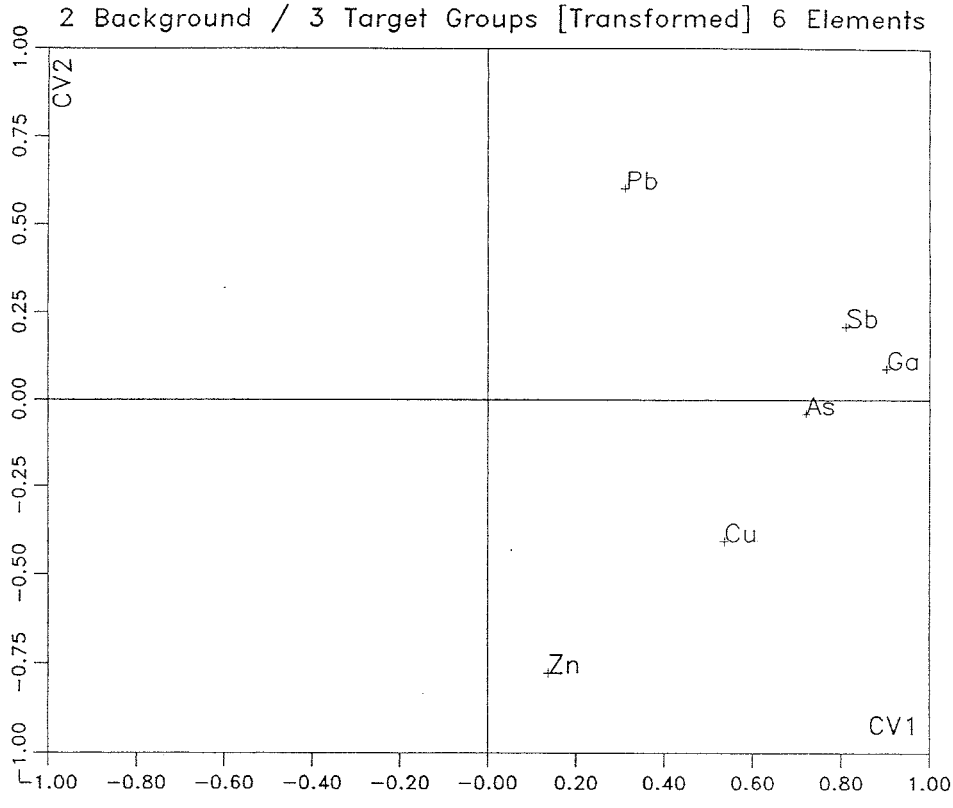


Figure 25c. Plot of the factor pattern scores projected on to the first and second canonical variate axes. The relative positions of the variables indicate the relative association of the samples to the variables in Figure 25a.

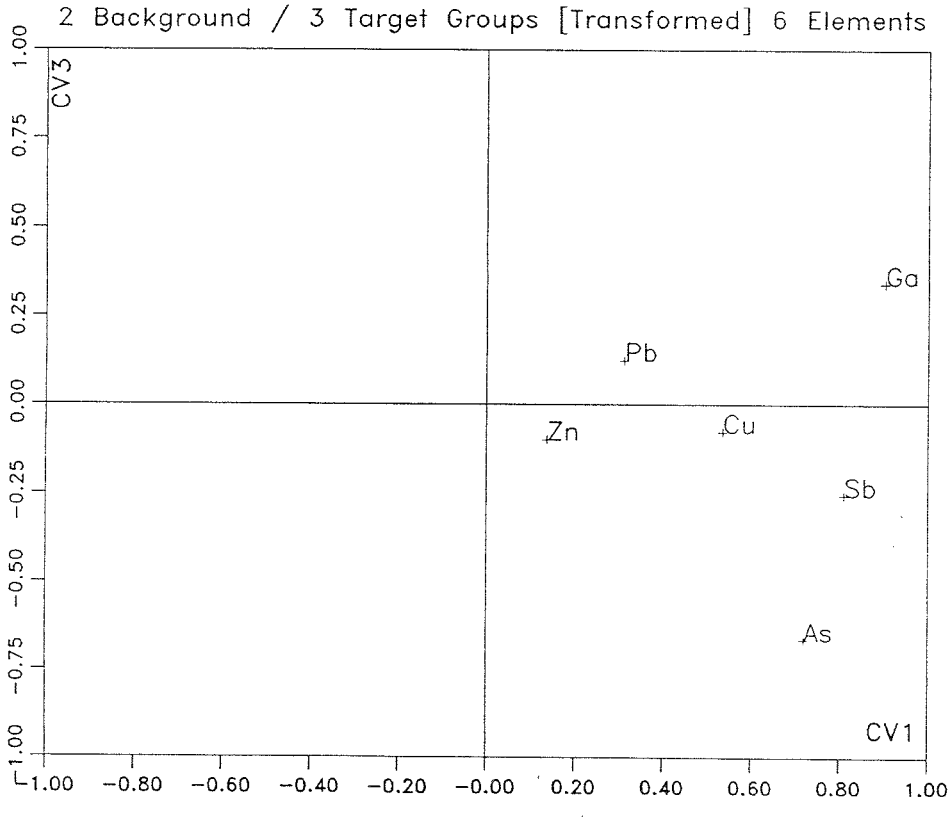


Figure 25d. Plot of the factor pattern scores projected on to the first and third canonical variate axes. The relative positions of the variables indicate the relative association of the samples to the variables in Figure 25b.

successive Wilk's Λ values for each function. Because the number of groups is equal to the number of variables, all of the communalities of the variables are set to 1.0, as they all must contribute equally to the group separation. Figures 26a,b show the discriminant scores of the samples plotted onto the CV1-CV2 and CV1-CV3 axes. The Figures show target group and background group separation along the first discriminant function (CV1) and Group 1 and Group 2 separation along the second discriminant function (CV2). CV3 does not appear to delineate any meaningful separation. From Figure 26a, the separation between the target and background groups can be seen, but Group 5 samples merge with the Murchison greenstone samples of Group 1 making the distinction of samples associated with this type of Au deposit difficult. Figure 26c shows that the background group, Group 1, is associated with Zn and Cu and the target Groups 3, 4, and 5 have relative Au and Pb enrichment. Group 2 is separated on the basis of being relatively depleted in these four elements.

The use of the four element suite of samples for discriminating background from target groups is even poorer than the six element suite and serves as a warning to avoid choosing a limited suite of elements that are inadequate for distinguishing between various geochemical populations.

5.6 Allocation / Typicality

When a statistical analysis is carried out on a particular group of samples that have been analysed for a specific element, it is assumed that the samples may belong to one or more groups of data which may represent one or more geological processes. This concept is crucial in any analysis of geochemical data.

The use of canonical variate analysis has provided a means of obtaining a statistical measure of the distinction of the reference groups. The analyses indicate that all of the magma clan and rock type reference groups are statistically distinct. A consequence of the establishment and testing of the reference groups is the ability to test unknown samples for possible membership into none, one, or more of the groups. Two types of procedures that can be used to do this are allocation and typicality.

Allocation procedures (also known as Classification Procedures) and measures of typicality make it possible to predict the probability of an unknown sample belonging to none, one, or more of the reference groups. The method is based on the assumption that the variables are the same between the reference groups and the unknown sample(s). A general reference to allocation procedures can be found in Cooley and Lohnes (1971), LeMaitre (1982), and Campbell (1984b). Allocation can also be used to test the group membership of samples used to create the reference groups in a canonical variate analysis. Typicality is described by Campbell (1984b).

Allocation methods and measures of typicality have not been used extensively in geological classification analysis. Exceptions to this have been for purposes of geochemical exploration as documented by Smith *et al.* (1984) and Garrett (1989a). These more recent contributions clearly explain the usefulness of the method.

Allocation procedures work on the basis of measuring the distance of a sample from each reference group centroid. The mathematics of allocation are straight forward and can be found in Cooley and Lohnes (1971) and LeMaitre (1976). By using the covariance estimates of the populations of samples used for the reference groups, the Mahalanobis distance can be computed between each unknown sample and the group centroids. LeMaitre (1982:171) provides a schematic concept of the allocation procedure. The covariance matrix contains the characteristics of the dispersion of the reference population. If the Mahalanobis distance of a sample, with respect to a particular reference group centroid, is within the dispersion matrix defined by the reference group covariances, then the probability of that sample belonging to that reference group is greater than zero. The closer a sample is to a group centroid, the higher the probability of membership; however, any sample with a probability >0 can be considered to be similar to the reference group. The dispersion between groups can overlap, which results in some samples having probabilities of belonging to more than one reference group. Similarly, if a sample has a Mahalanobis distance outside the dispersion matrix then it has a zero probability of belonging to any of the reference groups. A significant test is the determination of whether or not the groups have similar covariances. If they do not, then comparison of samples must be modified according the different covariance matrices.

Posterior Probability

Allocation of individual samples through the computation of posterior probabilities can be briefly described as: For the estimation of posterior probability of membership in the kth population is given by:

$$pr(k;\ \bar{x}_m) = p_k\, f(\ \bar{x}_m;P_k) \Big/ \sum_{j=1}^{g} p_j(f\ \bar{x}_m;P_j),$$

where $pr(k;\ \bar{x}_m)$ is the posterior probability of membership in the kth group with group vector (centroid) $\bar{x}_m$,
p_k is the prior probability that the unknown sample belongs to the kth group,
$f(\ \bar{x}_m;P_k)$ is the value of the probability density function for the group vector $\bar{x}_m$.

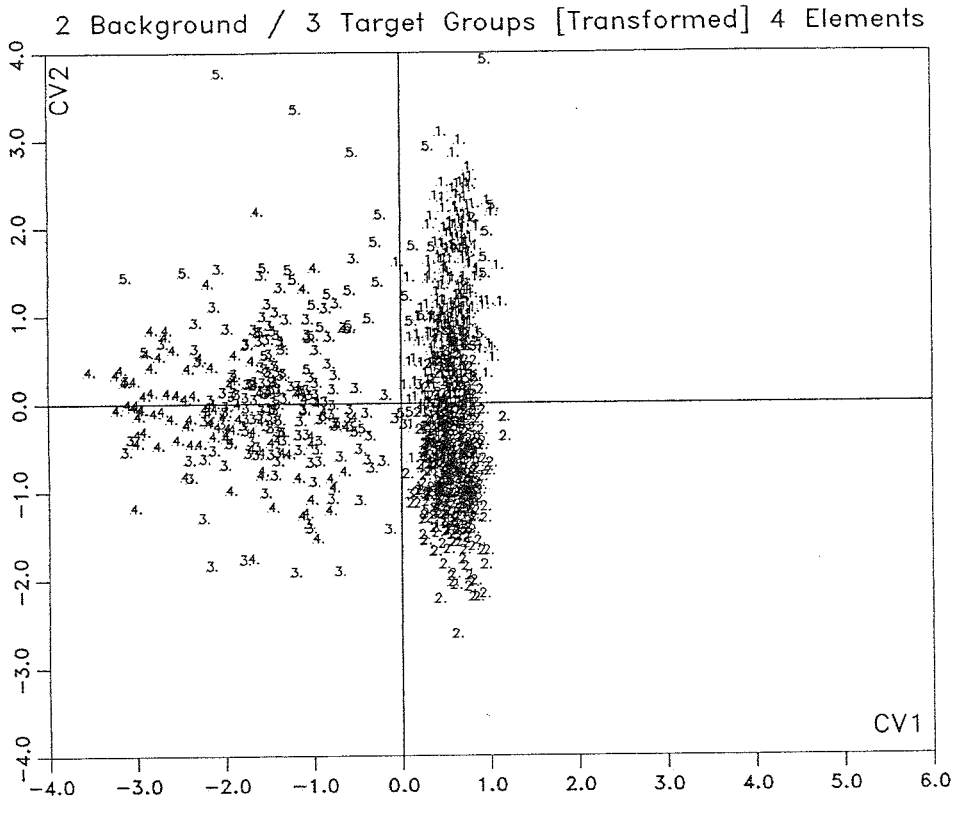


Figure 26a. Plot of discriminant function scores onto the first and second canonical variate axes of the 2 background and 3 target groups for the subset of 4 commodity elements [Au Cu Pb Zn]. Note the poor separation between the target groups and between the background groups.

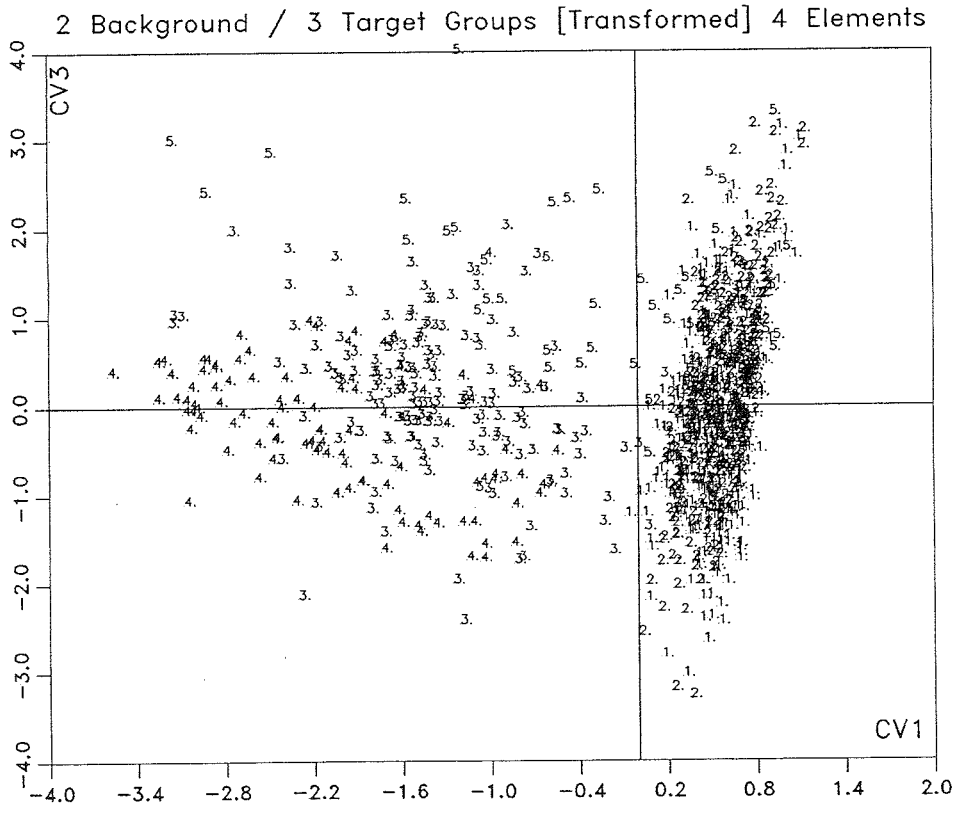


Figure 26b. Plot of discriminant function scores onto the first and third canonical variate axes of the 2 background and 3 target groups for the subset of 4 commodity elements [Au Cu Pb Zn]. Note the poor separation between the target groups and between the background groups.

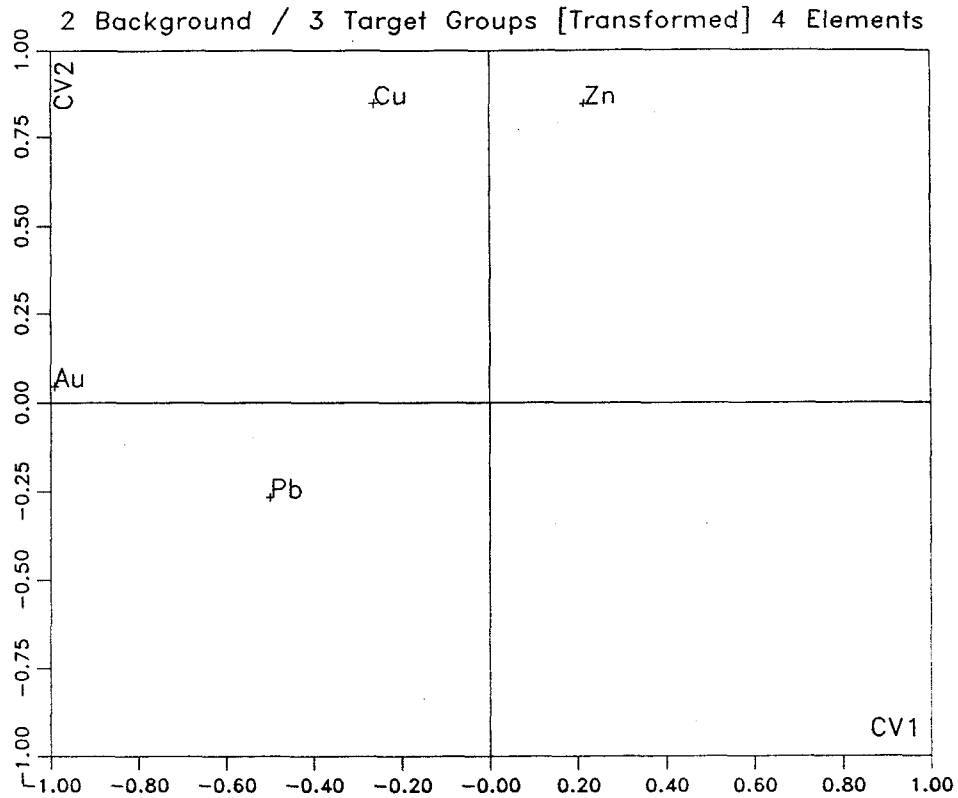


Figure 26c. Plot of the factor pattern scores projected on to the first and second canonical variate axes. The relative positions of the variables indicate the relative association of the samples to the variables in Figure 26a.

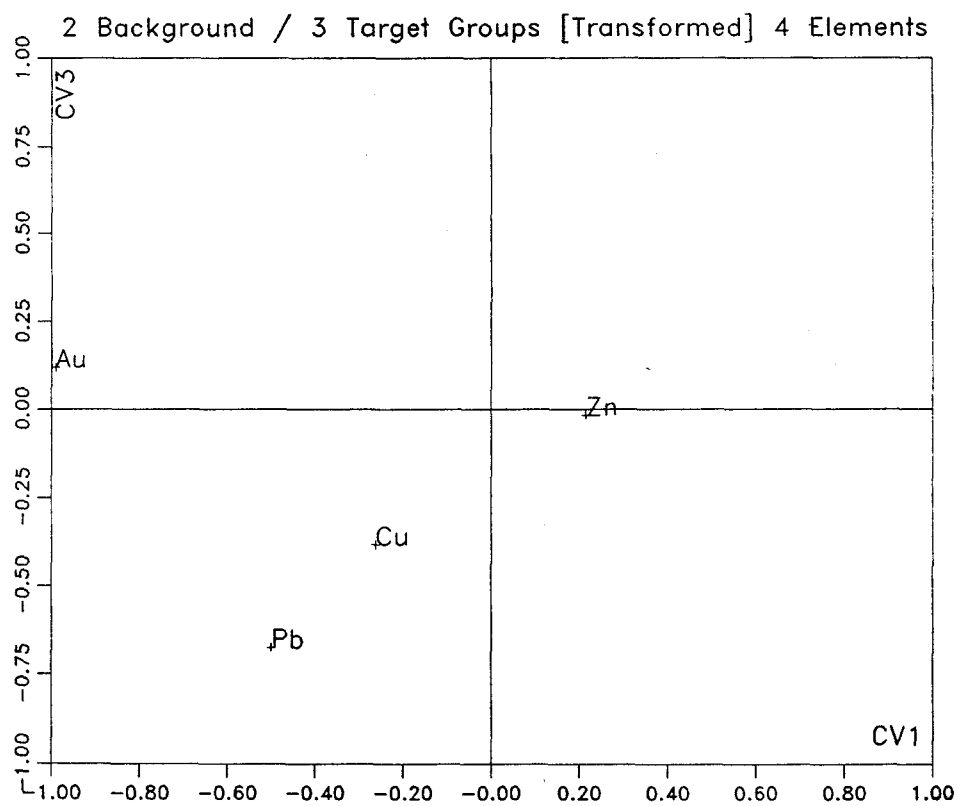


Figure 26d. Plot of the factor pattern scores projected on to the first and third canonical variate axes. The relative positions of the variables indicate the relative association of the samples to the variables in Figure 26b.

The posterior probability is the probability of belonging to the kth group divided by the sum of the probabilities of belonging to all g groups. An unknown sample is then assigned to the reference group with the highest posterior probability (smallest Mahalanobis distance).

Index of Typicality

The index of typicality is explained by Garrett (1990) (although it is referred to as allocation) and Campbell (1984b). Given a covariance (dispersion) matrix **S**, **k** groups, with samples composed of **p** variables, then $\mathbf{D}^2$, the generalized Mahalanobis distance is calculated such that

$$\mathbf{D}_k^2 = (\mathbf{x}_i - \mathbf{x}_{ik})' \, \mathbf{S}_k^{-1} \, (\mathbf{x}_i - \mathbf{x}_{ik}),$$

where k is the kth group, and i is the ith sample.

For the index of typicality, the unknown sample is provisionally allocated to the jth of the g groups such that

$$\mathbf{D}_j^2 + \ln |S_j| = \min[\mathbf{D}_k^2 + \ln |S_k|]$$

where k=1,...,j,...g reference groups, $|S_j|$ is the determinant of the covariance of the jth group.

If the covariances of the groups are equal then the logarithmic term is dropped.

The probability of group membership of each sample is predicted for each reference group using the statistic:

$$\mathbf{F} = \frac{(N - g - p + 1)\, n_k}{p\,(N - g)\,(n_k + 1)}\, \mathbf{D}^2_k$$

where

N= total number of samples over the groups being tested,

g= number of groups,

n_k= number of samples in group k,

p= number of variables.

This statistic is distributed as F with p (numerator) and N-g-p+1 (denominator) degrees of freedom. This is a "predictive" probability rather than an "estimative" probability that would normally be computed using the χ^2 distribution. Estimative procedures have been shown to underestimate group membership probabilities (Garrett, 1989a), thus the estimative approach is a better choice for classification.

An example of the use of allocation through posterior probabilities and the computation of the index of typicality is shown for a number of samples classified over the five reference groups listed in Table 21. The Table shows 18 samples that have been selected from a number datasets. The Group number associated with each sample represents the Group (Dataset) number mentioned in Section 1.2. Samples from Groups 6 and 7 are samples removed from Groups 1 and 2 respectively prior to establishing the references groups. The first part of the Table displays the composition of each sample and the second part of the Table shows the index of typicality and the posterior probability for each sample. The posterior probabilities force a fit of the sample to be allocated to at least one of the groups. The index of typicality does not force a fit. Comparison of samples 1 to 14 shows that generally the index of typicality and the posterior probability for each sample are as expected for the group with which it is associated. Samples that have probabilities in more than one group indicate that they overlap between the groups.

Samples 16, 17, and 18 have zero values for typicality indices. The application of the posterior probabilities appears to allocate these three samples to be most like Group 1. These samples were originally included in Group 1, but were eliminated in the reference group selection as their Mahalanobis distances exceeded the 95th percentile in the reference group selection. In this case, the application of posterior probability is reasonable. Sample 16 also shows a significant posterior probability of belonging to Group 5. The Mahalanobis distance of this sample to the Group 5 centroid may exceed the acceptable 95% confidence level for the F-statistic; however, it appears to have significant affinity to Group 5.

The application of posterior probabilities and indices of typicality have obvious advantages in an exploration programme where the geochemical characteristics of the commodity being sought are known. Samples that classify within the background populations can be recognized immediately. Samples from areas which have high typicality indices for any of the target groups are obvious areas for follow-up exploration. Maps of target group membership probability can be created which may outline areas of high mineralization potential. Such an approach was taken by Smith *et al.* (1984). Samples that have typicality indices of zero for any of the target or background groups should be investigated for their likelihood of representing some previously unrecognized target or background geochemical model. The use of posterior probabilities can assist in "forcing" a fit to one of the reference groups and to examine which group is the closest for an unknown sample.

Table 21
Typicalities 2 Background / 3 Target Groups [Transformed]
Target: Mt. Gibson/Golden Grove/Lawlers
Background: Murchison Greenstone/Albany-Fraser

Sample	Group	Cr	Cu	Pb	Zn	As	Sb	Sn	Ga	Nb	Au
1	1	869	100	36	22	25	4	4	25	15	3
2	1	552	32	74	14	21	<2	3	20	12	1
3	1	558	21	34	11	27	4	2	40	8	3
4	1	1105	73	5	28	45	9	<1	25	5	29
5	1	3469	100	19	26	26	6	<1	20	<4	17
6	6	634	54	38	20	20	<2	4	20	6	10
7	7	2028	90	14	30	23	<2	<1	15	5	4
8	7	364	32	78	12	30	7	3	30	9	11
9	2	98	4	25	10	10	<2	1	15	15	<1
10	2	137	24	36	10	13	<2	3	15	10	1
11	3	973	20	40	10	28	<2	5	38	15	5
12	3	625	28	46	15	38	6	2	62	12	34
13	4	196	60	74	5	385	34	35	63	5	1070
14	5	553	70	20	22	40	4	<2	44	10	840
15	5	560	62	18	11	360	5	2	54	8	1090
16	6	1701	46	18	26	691	175	<1	15	7	1
17	6	885	406	31	592	723	32	<1	1	<4	19
18	7	670	27	30	24	612	74	<1	10	6	<8

Typ: Index of Typicality
PstP: Posterior Probability

		Group 1		Group 2		Group 3		Group 4		Group 5	
Sample	Group	Typ	PstP	Typ	PstP	Typ	PstP	Typ	PstP	Typ	PstP
1	1	.9977	.9447	.0001	.0553	.0060	.0000	.0000	.0000	.0000	.0000
2	1	.7882	.2757	.2717	.7243	.0113	.0000	.0000	.0000	.0000	.0000
3	1	.8246	.9837	.0000	.0156	.5490	.0003	.0000	.0000	.3659	.0003
4	1	.3244	.9772	.0000	.0020	.0003	.0001	.0000	.0000	.5722	.0208
5	1	.7610	.9940	.0000	.0036	.0000	.0000	.0000	.0000	.6941	.0024
6	6	.8809	.8776	.0043	.1224	.3032	.0000	.0000	.0000	.0166	.0000
7	7	.9889	.9794	.0000	.0206	.0000	.0000	.0000	.0000	.4652	.0000
8	7	.4519	.7719	.0000	.2251	.3265	.0030	.0000	.0000	.0010	.0000
9	2	.5972	.0232	.9166	.9768	.0000	.0000	.0000	.0000	.0000	.0000
10	2	.7951	.0984	.9135	.9016	.0528	.0000	.0000	.0000	.0000	.0000
11	3	.9133	.9632	.0000	.0367	.0195	.0001	.0000	.0000	.0003	.0000
12	3	.0044	.4529	.0000	.0002	.6910	.3867	.0000	.0000	.6201	.1602
13	4	.0000	.0000	.0000	.0000	.0000	.0000	.8776	1.0000	.0000	.0000
14	5	.0000	.0016	.0000	.0000	.7036	.9766	.0000	.0000	.8750	.0218
15	5	.0000	.0000	.0000	.0000	.0000	.0679	.0000	.0000	.9927	.9320
16	6	.0000	.2486	.0000	.0010	.0000	.0000	.0000	.0002	.0000	.7502
17	6	.0000	.9610	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0390
18	7	.0000	.9399	.0000	.0483	.0000	.0000	.0000	.0000	.0000	.0118

6.0 CONCLUDING REMARKS - A SUGGESTED SEQUENCE OF DATA INTREPRETATION FOR ANOMALY RECOGNITION

Following the AEG workshop on Thresholds and Anomaly Interpretation, an appendix was assembled by Mellinger *et al.* (1984) which provided a suggested sequence of investigation for the anomaly interpretation based on the integration of geochemical, geographical, and geophysical data.

Three critical steps were defined: preliminary data analysis, descriptive multivariate analysis, and specific multivariate analysis. For the evaluation of geochemical data in laterites of the Yilgarn Block, the following sequence of investigation is recommended:

1) Preliminary Data Analysis

- Examine each element with histograms, box & whisker plots, Q-Q plots, scatter plot matrix, data ranking.
- Prepare summary statistical tables, plot elements onto maps.
- Trim the distribution of each element of gross outliers.
- Investigate outliers for each element; analytical error or atypical abundance?
- Adjust data for censored values if required.
- Transform each element with Box-Cox power transformations using samples below the 95th-98th percentile. This depends upon a prior visual examination of Q-Q plots or histograms.
- Create a scatter plot matrix for transformed data. Look for trends/associations.
- Techniques such as PROBPLOT (Stanley, 1987) or the Gap statistic (Miesch, 1981) can be used to dissect multiple populations.
- Set thresholds for elements after transformation.

2) Exploratory Multivariate Data Analysis

- The use of robust estimates to compute means and covariances to enhance the detection of outliers.
- Application of dimension reducing techniques, such as principal components analysis to show systematic linear relationships of the variables and the samples. The use of robust methods on transformed data assists in detecting outliers. Dimension-reducing techniques also indicate which elements are associated with commodity elements. Maps of the component scores can assist in outlining regional lithological variation and areas that are anomalous.
- The use of methods to delineate structure in the data. Methods such as cluster analysis, multidimensional scaling, non-linear mapping, and projection pursuit isolate groups of samples with similar characteristics. Atypical samples stand out as single outliers. Target groups can often be isolated using these methods. Maps of the locations of the groups can help to isolate mineralized areas.
- The use of χ^2 plots applied to transformed data to isolate outliers based on all of the elements of interest. The use of χ^2 plots assists in the elimination of outliers for the creation of background and target groups which can subsequently be used in canonical variate analysis and for allocation/typicality procedures. Maps of large Mahalanobis distances (>95th percentile) may identify anomalous areas.

3) Specific Multivariate Data Analysis and Modelled Multivariate Analysis

- The calculation of empirical indices can be used as methods which are specifically tailored to areas in which multi-element associations are well understood. The indices are based on a linear combination of pathfinder elements with coefficients that are selected for each area and the commodity being sought. Samples with high indices can be investigated for mineralization potential.
- The use of multiple regression can be applied to areas where a linear model of the multi-element association can be computed with good results (high R^2 coefficients). Residuals can be examined for the potential of association with mineral deposits.
- Once target and background groups have been established, the use of analysis of variance and canonical variate analysis test the statistical uniqueness of the groups. Groups that are statistically distinct can be used as reference groups against which unknown samples may be compared.
- The use of all possible subsets can be applied to compare the reference groups with each other and determine which group of elements enhances the group separations.
- The application of allocation/typicality procedures can test the samples used to make up the reference group populations. Additionally, unknown samples from a regional exploration programme can be used to assign the probability of belonging to one of the reference groups. Maps of tyicality or posterior probability can be made to indicate group membership.

7.0 REFERENCES

Anand, R.R., Smith, R.E., Innes, J., Churchward, H.M.. April 1989. *Exploration Geochemistry about the Mt. Gibson Gold Deposits, Western Australia.* CSIRO/AMIRA Laterite Geochemistry Project P240, Exploration Geoscience Restricted Report 20R, 93 pp.

Annand, R.R., Smith, R.E., Innes, J., Churchward, H.M., Perdrix, J.L. and Grunsky, E.C., August 1989. Laterite types and associated ferruginous materials, Yilgarn Block, WA. *Terminology, Classification and Atlas.* CSIRO Division of Exploration Geoscience, Restricted Report 60R (unpaginated).

Aucott, J.W., 1987. Workshop 5. Geochemical Anomaly Recognition, *J. Geochem. Explor.*, **29**:375-6.

Bochang, Y. and Xuejing, X., 1985. Fuzzy cluster analysis in geochemical exploration, *J. Geochem. Explor.*, **23**:281-92.

Box, G.E.P., and Cox, D.R., 1964. An Analysis of Transformations, *Jl. R. statist. Soc.*, Series B, **26**:211-52.

Bridges, N.J., and McCammon, R.B., 1980. Discrim. A computer program using an interactive approach to dissect a mixture of normal or lognormal distributions, *Computers Geosci.*, **6**:361-96.

Campbell, N.A., 1980, Robust procedures in multivariate analysis. I. Robust covariance estimation. *Appl. Statist.*, **29**:231-7.

Campbell, N.A., 1984a. Canonical Variate Analysis with Unequal Covariance Matrices: Generalizations to the Usual Solution, *Mathl Geol.*, **16**(2):109-24.

Campbell, N.A., 1984b. Some Aspects of Allocation and Discrimination. In: W.W. Howells and G.N. van Vark (eds.) *Multivariate Statistical Methods in Physical Anthropology*, D. Reidel, Dordrecht, pp. 177-92.

Campbell, N.A., 1986a. A General Introduction to a Suite of Multivariate Programs, *CSIRO Division of Mathematics and Statistics*, unpublished report.

Campbell, N.A., 1986b. Censored, grouped and truncated geochemical data, *Appl. Statist.*.

Chaffee, M. A., 1983. Scoresum - A Technique for Displaying and Evaluating Multi-Element Geochemical Information, With Examples of its use in Regional Mineral Assessment Programs, *J. Geochem. Explor.*, **19**:361-81.

Chork, C.Y., 1990. Unmasking multivariate anomalous observations in exploration geochemical data from sheeted-vein tin mineralization near Emmaville, N.S.W., *J. Geochem. Explor.*, **37**:205-23.

Chung, C.F., 1985. Statistical treatment of geochemical data with observations below the detection limit. In: *Current Research*, Part B, Geological Survey of Canada, Paper 85-1B, pp. 141-50.

Chung, C.F., 1988. Statistical analysis of truncated data in geosciences, in *Sci. de la Terr. Ser. Inf.*, Nancy, **27**:157-80.

Colvine, A.C. *et al.*, 1988. Archaean Lode Gold Deposits in Ontario; *Ontario Geological Survey Miscellaneous Paper 139*, 136 pp.

Cooley, W.W. and Lohnes, P.R., 1971. *Multivariate data analysis*, Wiley, New York, 364 pp.

Davis, J.C., 1986. *Statistics and Data Analysis in Geology*, John Wiley & Sons Inc., second edition, 646 pp.

Dempster, A.P., Laird, N.M., and Rubin, D.B., 1977, Maximum Likelihood from Incomplete Data via the EM Algorithm, *J. R. Stat. Soc.*, Series B, **39**:1-38.

Diday, E., 1973. The dynamic clusters method in non-hierarchical clustering, *Int. J. Computer Inf.*, **2**:61-88.

Everitt, B., 1980. *Cluster Analysis*, Heinemann, London, 2nd Edition, 122 pp.

Garrett R.G., 1983. Sampling Methodology, Chapter 4, Statistics and Data Analysis in Geochemical Prospecting, edited by R.J. Howarth, Vol. 2, in *Handbook of Exploration Geochemistry*, edited by G.J.S. Govett, Elsevier, 437 pp.

Garrett, R.G., 1984. Workshop 5. Thresholds and Anomaly Interpretation, *J. Geochem. Explor.*, **21**:137-42.

Garrett R.G., 1989a. A Cry from the Heart, *Explore, Newsletter of the Association of Exploration Geochemists*, **66**:18.

Garrett, R.G., 1989b. The chi-square plot. a tool for multivariate outlier detection, *J. Geochem. Explor.*, **32**:319-41.

Garrett, R.G., 1990. A Robust Multivariate Procedure with Applications to Geochemical Data. In. Proceedings of the Colloquium on "*Statistical Applications in the Earth Sciences*", Geological Survey of Canada Paper 89-9, pp. 309-18

Garrett, R.G., 1991. The Management, Analysis and Display of Exploration Geochemical Data, Paper #9, *Exploration Geochemistry Workshop*, Prospectors and Developers Association of Canada, Toronto, March, 1991.

Garrett, R.G., Kane, V.E., Zeigler, R.K., 1980. The Management and Analysis of Regional Geochemical Data, *J. Geochem. Explor.*, **13**:113-52.

Groves, D.I., 1988. Gold Mineralization in the Yilgarn Block, Western Australia. In: *Bicentennial Gold 88, Extended Abstracts, Oral Programme*, Geological Society of Australia Inc., Abstracts No. 22, p.13-23.

Grunsky, E.C., 1986. Recognition of Alteration in Volcanic Rocks Using Statistical Analysis of Lithogeochemical Data, *J. Geochem. Explor.*, **25**:157-83.

Grunsky, E.C., 1991: Laterite Geochemistry in the CSIRO-AGE Database for the Albany-Fraser Region. CSIRO/AMIRA Laterite Geochemistry Project P240, Exploration Geoscience Restricted Report 161., 2 Volumes, 145 pp., 1 5.25" diskette.

Grunsky, E.C., Innes, J. Smith, R.E., Perdrix, J.L., 1988. Report on laterite geochemistry in the CSIRO-AGE database for the southern Murchison region. CSIRO/AMIRA Laterite Geochemistry Project P240, Exploration Geoscience Restricted Report 2R, 92 pp., 1 5.25" diskette, 1 map.

Grunsky, E.C., Smith, R.E., and Perdrix, J.L., 1989. Report on laterite geochemistry in the CSIRO-AGE database for the northern Murchison region. CSIRO/AMIRA Laterite Geochemistry Project P240, Exploration Geoscience Restricted Report 68R, 148 pp., 1 5.25" diskette.

Hawkes, H.E. and Webb, J.S., 1962. *Geochemistry in Mineral Exploration*, First Edition, Harper and Row, New York, 415 pp.

Howarth, R.J., 1983. Mapping, Chapter 5, Statistics and Data Analysis in Geochemical Prospecting, edited by R.J. Howarth, Vol. 2, in *Handbook of Exploration Geochemistry*, edited by G.J.S. Govett, Elsevier, 437 pp.

Howarth, R.J. and Earle, S.A.M., 1979. Application of a Generalized Power Transformation to Geochemical Data, *Mathl Geol.*, **11**(1):45-62.

Howarth, R.J. and Sinding-Larsen, R., 1983. Multivariate Analysis, Chapter 6, Statistics and Data Analysis in Geochemical Prospecting, edited by R.J. Howarth, Vol. 2, in *Handbook of Exploration Geochemistry*, edited by G.J.S. Govett, Elsevier, 437 pp.

Jaquet, J.-M., Froidevaux, F., Bernet, J.-P., 1975, Comparison of Automatic Classification Methods Applied to Lake Geochemical Samples, *Mathl Geol.*, **7**(3):237-66.

Jöreskog, K.G., Klovan, J.E. and Reyment, R.A., 1976. *Geological Factor Analysis*. Elsevier Scientific Publishing Company, New York, 178 pp.

Joyce, A.S., 1984. *Geochemical Exploration*, The Australian Mineral Foundation, Glenside, South Australia, 184 pp.

Kruskal, J.B., 1964. Multidimensional scaling by optimising goodness of fit to non-metric hypothesis, *Psychometrika*, **29**:1-27.

Krzanowski, W.J., 1988. *Principles of Multivariate Analysis, A User's Perspective*, Clarendon Press, Oxford, 563 pp.

Kürzl, H., 1988. Exploratory data analysis: recent advances for the interpretation of geochemical data, *J. Geochem. Explor.*, **30**:309-22.

LeMaitre, R.W., 1976. A New Approach to the Classification of Igneous Rocks Using the Basalt-Andesite-Dacite-Rhyolite Suite as an Example, *Contributions to Mineralogy and Petrology*, **56**:191-203.

LeMaitre, R.W., 1982. *Numerical Petrology, Statistical Interpretation Geochem. Data*, Elsevier, 281 pp.

Levinson, A.A., 1980. *Introduction of Exploration Geochemistry*, Second Edition, Applied Publishing, Wilmette, 924 pp.

Lindqvist, L., 1976. SELLO, A Fortran IV program fo the transformation of skewed distributions to normality, *Computers Geosci.*, **1**:129-45.

Link, R.F., and Koch, G.S., 1975. Some consequences of applying lognormal theory to pseudolognormal distributions, *Mathl Geol.*, **7**:117-28.

McCabe, G.P., 1975. Computations for variable selection in discriminant analysis, *Technometrics*, **17**:103-9.

McQueen, J., 1967. Some methods for classification and analysis of multivariate observations, *5th Berkeley Symposium on Mathematics, Statistics, and Probability*, **1**:281-98.

Mellinger, M. 1987. Multivariate Data Analysis. Its Methods, *Chemometrics and Intelligent Laboratory Systems*, **2**:29-36.

Mellinger, M., Chork, S.C.Y., Dijkstra, S., Esbensen, K.H., Kürzl, H., Lindqvist, L., Saheurs, J.-P., Schermann, O., Siewers, U., and Westerberg, K., 1984. The Multivariate Chemical Space, and the Integration of the Chemical, Geographical, and Geophysical Spaces, *J. Geochem. Explor.*, **21**:143-8.

Miesch, A.T., 1981. Estimation of the Geochemical Threshold and it Statistical Significance, *J. of Explor. Geochemistry*, **16**:49-76.

Robertson, I.D.M. and Taylor, G.F. 1987. Depletion Haloes in Fresh Rocks Surrounding the Cobar Orebodies, N.S.W., Australia: Implications for Exploration and Ore Genesis, *J. Geochem. Explor.*, **27**:77-101.

Rock, N.M.S., 1987, Robust. An Interactive Fortran-77 Package for Exploratory Data Analysis using Parametric, Robust and Nonparametric Location and Scale Estimates, Data Transformations, Normality Tests, and Outlier Assessment, *Computers Geosci.*, **13**(5):463-94.

Rock, N.M.S., 1988. Numerical Geology, A Source Guide, Glossary and Selective Bibliography to Geological Uses of Computers and Statistics, *Lecture Notes in Earth Sciences*, Vol. 18, edited by Somdev Bhattacharji, Gerald M. Friedman, Horst J. Neugebauer and Adalf Seilacher, Springer-Verlag, 427 pp.

Rose, A.W., Hawkes, H.E., and Webb, J.S. 1979. *Geochemistry in Mineral Exploration*, Second Edition, Academic Press, London, 657 pp.

Sammon, J.W., 1969. A non-linear mapping for data structure analysis. *IEEE Transactions in Computing*, C18, pp. 401-409.

Sinclair, A.J., 1976. Application of Probability Plots in Mineral Exploration, *Association of Exploration Geochemists Special Volume 4*, 95 pp.

Sinding-Larsen, R., 1975. A computer method for dividing a regional geochemical survey area into homogeneous subareas prior to statistical interpretation. in I.L. Elliott and W.K. Fletcher (Editors), *Geochemical Exploration 1974*. Elsevier, Amsterdam, pp. 191-217.

Smith, R.E. and Perdrix, J.L. 1983. Pisolitic laterite geochemistry in the Golden Grove massive sulphide district, Western Australia, *J. Geochem. Explor.*, **18**:131-64.

Smith, R.E., Campbell, N.A., and Litchfield, R., 1984. Multivariate Statistical Techniques Applied to Pisolitic Laterite Geochemistry at Golden Grove, Western Australia, *J. Geochem. Explor.*, **22**:193-216.

Smith, R.E., Perdrix, J.L., Davis, J.M., 1987. Dispersion into Pisolitic Laterite from the Greenbushes Mineralized Sn-Ta Pegmatite System, Western Australia, *J. Geochem. Explor.*, **28**:251-65.

Smith, R.E., Birrell, R.D., and Brigden, J.F., 1989. The implications to exploration of chalcophile corridors in the Archaean Yilgarn Block, Western Australia, as revealed by laterite geochemistry, *J. Geochem. Explor.*, **32**:169-84.

Smith, R.E., *et al.*, in prep. *Reference Geochemical Datasets from the Mt. Gibson Orientation Study, Western Australia*. CSIRO/AMIRA Laterite Geochemistry Project P240, Exploration Geoscience Restricted Report 157R.

Srivastava, M.S., and Carter, E.M. 1983. *An Introduction to Applied Multivariate Statistics*, North Holland, New York, 394 pp.

Stanley, C.R. 1987. PROBPLOT, An Interactive Computer Program to Fit Mixture of Normal (or Log normal) Distribution with Maximum Likelihood Optimization Procedures, *Association of Exploration Geochemists Special Volume 14*, 1 diskette.

Stanley, C.R. and Sinclair, A.J., 1987. Anomaly recognition for multi-element geochemical data- A background characterization approach. *J. Geochem. Explor.*, **29**:333-53.

Stanley, C.R. and Sinclair, A.J., 1989. Comparison of probability plots and the gap statistic in the selection of thresholds for exploration geochemistry data. *J. Geochem. Explor.*, **32**:355-7.

Tukey, J.W., 1977. *Exploratory Data Analysis*, Addison-Wesley, Sydney, 506 pp.

van Gaans, P.F.M. and Vriend, S.P., 1990. Multiple Linear Regression With Correlations Among the Predictor Variables. Theory and Computer Algorithm RIDGE (Fortran 77), *Computers Geosci.*, **16**(7):933-52.

Zhou, D., 1985. Adjustment of geochemical background by robust multivariate methods, *J. Geochem. Explor.*, **24**:207-22.

Zhou, D., ROPCA, 1989. A Fortran Program for Robust Principal Components Analysis, *Computers Geosci.*, **15**(1):59-78.

Zhou, D., Chang, T. and Davis, J.C., 1983. Dual Extraction of R-Mode and Q-Mode Factor Solutions, *Mathl Geol.*, **15**(5):581-606.