

IDENTIFYING GEOCHEMICAL ANOMALIES

K.G. McQueen

CRC LEME

Department of Earth and Marine Sciences

Australian National University, ACT 0200

Ken.McQueen@canberra.edu.au

Geochemical anomalies are geochemical features different from what is considered normal. They can be the result of:

1. unusual or uncommon processes concentrating particular elements (e.g. an ore-forming process, weathering and element dispersion from an unusual element concentration such as an orebody);
2. element accumulation or concentration by common processes acting over long periods (e.g. scavenging and concentration of certain elements by ironstones, ferruginous regolith or manganese oxides);
3. artificial contamination of sites or samples;
4. analytical noise or error (e.g. poor precision of the analytical method, particularly for element concentrations close to the detection limit).

Traditionally, geochemical anomalies have been identified by setting threshold values, which mark the upper and lower limits of normal variation for a particular population of data. Values within the threshold values are referred to as **background** values and those above or below as **anomalies**. In mineral exploration interest is generally in positive anomalies, on the assumption that ore deposits and their weathering have increased element abundances above normal crustal levels. However, negative anomalies can also be important, for example where they reflect depletion in some elements during host rock alteration accompanying ore formation.

Statistical methods have been widely applied to interpret geochemical data sets and define anomalies. Such methods need to be used cautiously because of the particular characteristics of geochemical data. Geochemical data sets seldom represent a single population or distribution, the data are typically spatially dependent and at each sample site a range of different processes have influenced the element abundances measured. The data are also imprecise due to unavoidable variability in sampling methods and media and the level of analytical precision. As a result no single universally applicable statistical test has been developed for identifying anomalies. Statistical investigation should use a range of techniques to explore the nature of geochemical data before selecting anomalous values (e.g. Reimann et al., 2005).

Univariate statistical methods for investigating geochemical data

Univariate statistical methods (i.e. involving observations with only one variable) can be used to organise and extract information from a data set of values for a single element (e.g. gold analyses for a group of samples). A first step is to examine the frequency distribution (spread of values) of the data set using frequency histograms, frequency plots or cumulative-frequency plots. This can help identify the type of distribution of the data, presence of multiple populations and outliers in the distribution. Box and whisker plots are another convenient way of examining the frequency distribution of a data set and for comparing the frequency distributions of multiple data sets. This type of plot shows the median (middle value or 50th percentile), a box with upper and lower hinges (or limits) defined by the 75th and

25th percentile values respectively (i.e. upper and lower quartile values), an inner fence defined as 1.5 times the length of the box (the interquartile range) towards maximum and minimum values and whiskers extending to the maximum and minimum values (Tukey, 1977). The central box will contain 50% of the data. Values below the whiskers are considered outliers and values more than 3 times the interquartile range (length of the box) from the box hinges are referred to as far outliers (Figure 1).

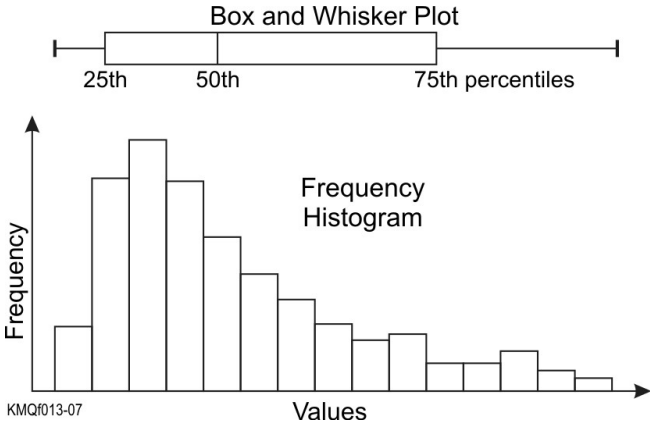


Figure 1. Example of a box and whisker plot with corresponding frequency distribution.

In the past, a simple way of statistically defining an anomaly in a single population of normally distributed data has been to consider values outside 2 standard deviations from the mean (statistical average) as anomalous (Hawkes and Webb, 1962). In other words the threshold or limit of normal variation was set at 2 standard deviations from the mean (Figure 2a) and the anomalous values taken as the top 2.5% of the population (positive anomalies) and the bottom 2.5% of the population (negative anomalies). This is somewhat arbitrary, and rarely do geochemical data fit a normal distribution pattern (they are typically positively skewed with a long tail towards higher values, Figure 2b). Commonly there is more than one population of data present in a geochemical data set (Figure 3). For example, samples collected of different media or of the same media derived from chemically different host rocks will contain multiple populations, each with their own threshold value. Further, anomalies of interest are defined by outliers that are not part of the background population.

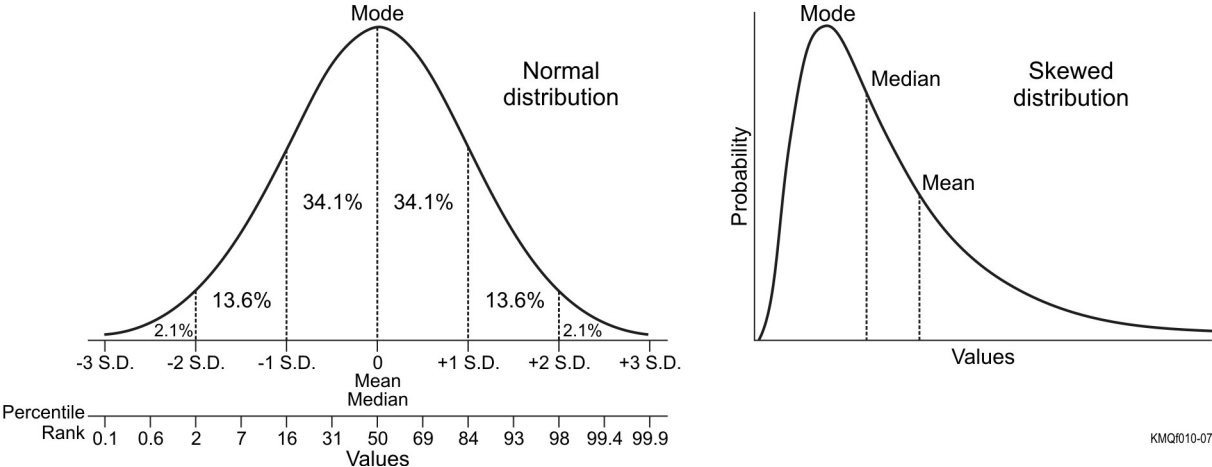


Figure 2. Examples of frequency plots for (a) normal (at left) and (b) skewed distributions.

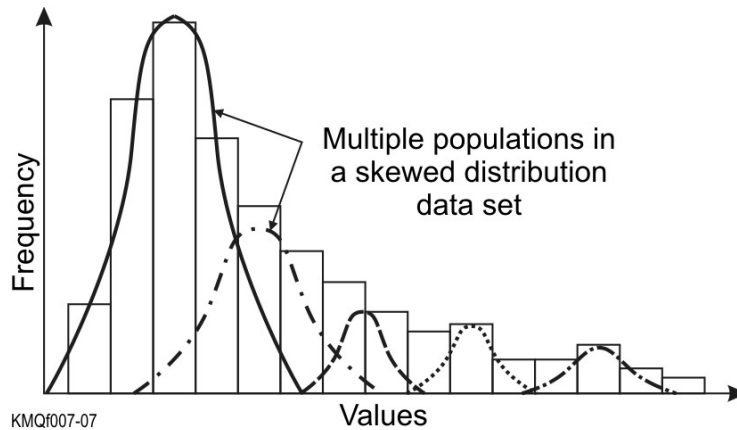


Figure 3. Example of frequency histogram and frequency curves for a data set with multiple populations.

Another common approach has been to transform geochemical data to a normal distribution pattern and then apply normal parametric tests. It has been argued that many natural data, including trace element abundances, approximate a lognormal (\log_{10}) distribution, so a simple log transformation has typically been applied. Again this is only valid for a single population of data. Also it has been shown that most geochemical data sets lie between normal and lognormal distributions (Reimann and Filzmoser, 2000).

An alternate method for handling skewed data is to set the threshold at 2 Median Absolute Deviations (MAD) from the median (middle or 50th percentile value). The median will lie away from the mean and the skewness of the data (typically to a lower value than the mean) and extreme values (outliers) will have less influence. The median absolute deviation is defined as the median value of the absolute deviations from the median of all the data (Tukey, 1977). The MAD approach is best applied when the data contain less than 10% outliers.

Background and anomalous values are commonly established empirically. During geochemical exploration programs, orientation surveys or case studies that compare typical background materials and sites with materials from areas of known mineralisation can be used to establish thresholds. This approach assumes that all the natural variability is covered in the orientation survey. It may miss very subtle anomalies or anomalies and element association patterns associated with a different or unknown style of mineralisation. A common deficiency with case studies of geochemical patterns around known mineralisation is insufficient coverage of the surrounding background.

Multivariate statistical analysis of geochemical data

Geochemical data sets are inherently multivariate (i.e. they generally have more than one variable for each sample, such as a large number of contained and associated elements). Geochemical anomalies are commonly expressed in more than one element. This is because the source or process that has generated the anomaly commonly has an association of elements. Different ore deposit types typically have specific element associations of target and pathfinder elements. A target element is the commodity being sought and a pathfinder element is one that accompanies this element, but may be more widely dispersed or easier to detect. Element associations can be used to advantage by taking a multi-element approach to anomaly detection. Multi-element analysis can also identify other non ore-related associations, such as those generated by normal regolith processes or the result of anthropogenic contamination.

A range of multivariate statistical methods can be used to assess the relationships within multi-element data sets. These methods commonly include:

1. “scatter plots” (bivariate plots comparing pairs of elements);
2. correlation matrices (using linear regression to test the correlation between pairs of elements);
3. cluster analysis (hierarchical grouping of elements in a data set with differing degrees of correlation of their abundance);
4. principal component factor analysis (useful for grouping elements into associations);
5. discriminant analysis (a method of optimising the distinction between two or more populations of samples).

The main difficulty in assessing multi-element data is multi-dimensional visualisation. A number of techniques can be used to help, including the use of multi-element “spider diagrams”, which plot values for a range of elements (connected by lines) in each sample. Typically these plots involve normalising the data to a reference sample. They have been widely used for example in comparing REE (rare earth element) data from different samples (see Rollinson, 1993). Linked scatter plots, in which particular samples can be identified across a number of bivariate element plots, are another convenient way for visually identifying samples with unusual multi-element characteristics. Triangular diagrams and computer generated rotatable 3D plots can be used to visually examine data sets for three elements at a time. Two-dimensional dendograms produced from cluster analysis are a simple way of assessing multi-element associations. A number of software programs are available for univariate and multivariate statistical analysis of geochemical data and visualisation of the results (e.g. ioGAS).

Where element association are well known for particular geological materials including ore deposits, suites of these of elements can be statistically combined to detect anomalies. Simple methods that have been used include addition or multiplication of different element concentration or weighted values (e.g. Beus and Grigorian, 1975; Smith and Perdix, 1983). The combined anomalies can be more robust or indicative of a particular type of source than single element anomalies. For example, anomalies of associated platinum group elements can be used to discriminate between nickel anomalies generated from komatiite-hosted nickel sulphide deposits and anomalies related to weathering of nickel-bearing ultramafic rocks. A number of empirical chalcophile (CHI) indices for ranking gossans and detecting anomalies and regional chalcophile corridors using samples of ferruginous lateritic residuum have been developed (e.g. Smith and Perdrix, 1983; Smith et al., 1989). These use various combinations of chalcophile and related pathfinder elements and simply weight these in an additive index. For example, $CHI-3 = As+3Sb+10Bi+10Cd+10In+3Mo+30Ag+30Sn$ was found useful for locating anomalies over massive Cu-Zn sulphide deposits at Gossan Hill and Scuddles in Western Australia.

Exploratory geochemical data analysis

In all studies of geochemical data it is valuable to make some initial assessment of the nature of the distribution of values, presence of outliers, and element correlations. This is referred to as exploratory data analysis (EDA) and commonly uses frequency plots, correlations matrices, bivariate scatter plots and in some cases cluster analysis or multivariate analysis to examine the data. Normal-probability plots (i.e. cumulative frequency plotted on a probability scale) are particularly useful for quick and simple first pass assessment of single data sets (see Box 1). EDA can indicate very obvious anomalies, the presence of multiple populations of

data and likely element associations. Multiple populations may be indicated by distinct groupings in the frequency distribution of a data set and in some cases these can be highlighted by careful assessment of transformed values (e.g. log transformed data vs probability is commonly used). The population of highest values may represent anomalies, but there may also be anomalies present in the upper values of other individual populations. EDA can also indicate deficiencies in the quality of the geochemical data. Using EDA natural breaks in data distributions can be highlighted and examined. Threshold values can be varied to see what affect this has on the allocation of anomalous values. The outliers defined by EDA can be investigated in more detail in terms of their other attributes.

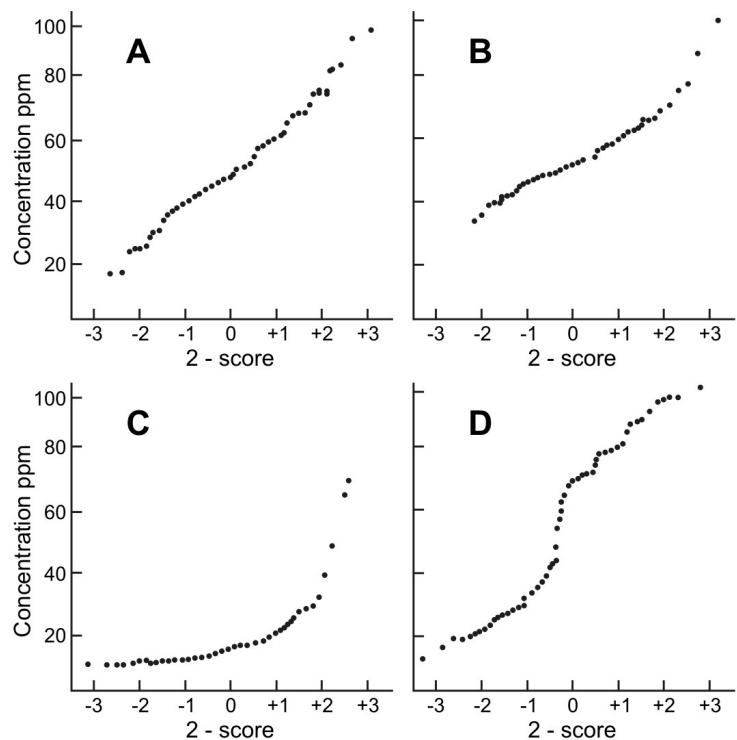
Box 1

Normal Probability Plots

The normal-probability plot adjusts the distribution of values in a data set such that if they have a normal distribution the raw measured values vs probability plot will be a near straight line. It is a simple test to check whether a data set is normally distributed or if there are multiple populations present. For the probability axis the raw data are transformed to standard normal values, that is where the mean is 0 and the standard deviation is 1.

These values or z-scores are calculated as $Z_i = (X_i - \mu) / \delta$ where

X_i = a value in the data set
 μ = the mean of the data set
 δ = the standard deviation.



Some examples of normal probability plots: A is a near normal distribution; B has longer tails than would be expected for normally distributed data; C is lognormally distributed data; D is two normally distributed populations with overlap.

KMQ052-07

In the exploratory analysis of multivariate data, a useful approach is to take into account not only the distance of observations from the centre or centroid of the multi-dimensional data distribution, but also the shape of the data distribution (Filzmoser et al., 2005). The shape and size of multivariate data can be quantified by the covariance matrix using a measure known as the Mahalanobis distance (see Box 2). The region for a particular Mahalanobis distance around the mean will form an ellipse when there are only two variables and an ellipsoid when there are more than two variables. The data can then be observed in terms of their Mahalanobis distance from the centroid of the distribution and outliers defined. Samples that are not anomalous in terms of univariate distributions may be revealed as anomalous associations when multivariate relationships are taken into account. Identified outliers can be investigated further using other methods and parameters. For large data sets this is a very useful approach to reducing the number of samples requiring detailed investigation in terms of anomaly forming processes.

Box 2

Detecting Multivariate Outliers using Mahalanobis Distance

Mahalanobis distances (MD) are based on the means variances and covariances of all the variables (e.g., elements) in a multivariate data set (Mahalanobis 1936).

For a ρ -dimensional multivariate sample

$$\chi_1 \dots \chi_n$$

$$MD_i = ((\chi_i - t)^T C^{-1} (\chi_i - t))^{1/2} \text{ for } i = 1 \dots n \text{ (number of variables).}$$

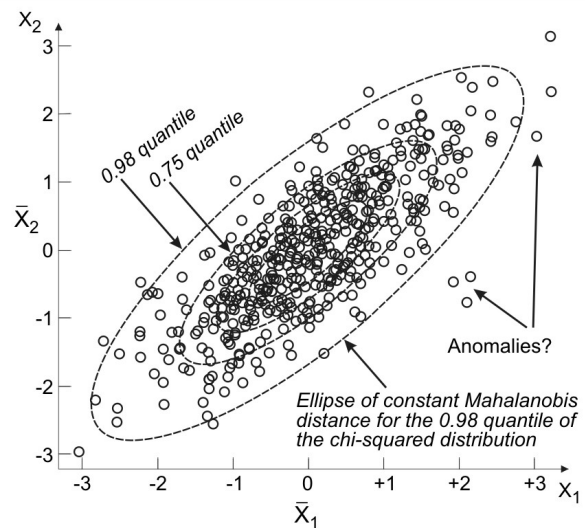
where:

t is the estimated multivariate location (usually the multivariate arithmetic mean i.e., the centroid)

C is the sample covariance matrix.

For multivariate normally distributed data the values of Mahalanobis distance squared (MD_i^2) are approximately chi-square distributed with ρ degrees of freedom (χ_i^2).

By setting the (squared) Mahalanobis Distance (MD^2) equal to a certain constant (i.e., to a certain quantile of χ_i^2) it is possible to define ellipsoids having the same Mahalanobis distance from the centroid. Multivariate outliers can then be identified as values having a large (squared) Mahalanobis distance from the centroid (e.g., outside the ellipsoid for the 98% quantile, $\chi_{\rho,0.98^2}$).



KMQI055-08

Space, time and source aspects of geochemical anomalies

Groups of geochemical values that are statistically defined should always be examined in terms of their spatial relationships (3D geographic and depth positions) and linked to their regolith and geological context. This can often clear up uncertainties about multiple populations and appropriate thresholds for defining anomalies. Ideally geochemical data should also be linked to their mineralogical controls. Spatial variance in anomalies and patterns of element associations are the basic method of vectoring to the source of an anomaly. A number of methods are widely used for visualising the 2-D spatial distribution of element values and derived data including: contouring, pixel plots and bubble plots. Contour maps are easy to visualise but may not be the best method of presentation as commonly the geochemical data do not possess the characteristics required for proper contouring. The data may lack linear dependence, even over short distances, and the contouring may also misrepresent the precision of the sampling and analytical methods. Pixel plots, colour coded for different values, or bubble plots, with bubble diameters representative of element concentrations, may be better options.

Element dispersion in the fourth dimension (i.e. time) is an important aspect of anomaly formation. Environmental conditions may have changed significantly through time at a particular site affecting the type and degree of element dispersion. If the relative expression of a geochemical anomaly is markedly different under different weathering regimes, setting the appropriate anomaly threshold and recognising the appropriate multi-element association will depend on knowing when the anomaly formed and under what conditions. Protracted

weathering or weathering through a range of contrasting chemical regimes commonly results in strong chemical leaching and marked depletion of most elements, so that any geochemical anomalies are very subtle. A multi-element approach may improve detection of such anomalies. Analytical or sampling techniques that improve anomaly to background contrast or reduce background variation (noise) may also be required to detect more subtle anomalies.

A geochemical anomaly may relate to an anomalous source for an element or suite of elements but lie within the level of background variation in terms of element abundance. These types of anomalies are difficult to detect and many have probably not yet been found. Methods for determining element sources include isotopic analysis, combined geochemical and mineralogical analysis to target particular host minerals and multi-element analysis to detect associations of elements related to a particular source. Isotopic analysis of low-level lead in ironstones/gossans, rock chips and soils has been used as a method to detect anomalies related to ore deposit types with particular lead isotopic ratios (Gulson, 1986). Isotopic analysis of groundwater, particularly using S, Sr and Pb isotopes, is a promising technique for locating ore deposits and other regolith element concentrations that have interacted with groundwater (e.g. Andrew et al., 1998; de Caritat et al., 2005).

References

- Andrew, A.S., Carr, G.R., Giblin, A.M. and Whitford, D.J., 1998. Isotope hydrogeochemistry in exploration for buried and blind mineralisation. *Geological Society of Australia Special Publication* 20, 222-225.
- Beus, A.A. and Grigorian, S.V., 1975. *Geochemical exploration methods for mineral deposits*. Applied Publishing Ltd. Wilmette, 287 pp.
- de Caritat, P., Kirste, D., Carr, G. and McCulloch, M., 2005. Groundwater in the Broken Hill region, Australia: recognising interaction with bedrock and mineralisation using S, Sr and Pb isotopes. *Applied Geochemistry* 20, 767-787.
- Hawkes, H.E. and Webb, J.S., 1962. *Geochemistry in Mineral Exploration*. Harper and Row, New York.
- Filzmoser, P., Garrett, R.G. and Reimann, C., 2005. Multivariate outlier detection in exploration geochemistry. *Computers and Geosciences* 31, 579-587.
- Gulson, B.L., 1986. *Lead isotopes in mineral exploration*. Elsevier, Amsterdam, 246 pp.
- Mahalanobis, P.C., 1936. On the generalised distance in statistics. *Proceedings of the National Institute of Science of India* 12, 49-55
- Reimann, C. and Filzmoser, P., 2000. Normal and lognormal data distribution in geochemistry: death of a myth. Consequences for the statistical treatment of geochemical and environmental data. *Environmental Geology* 39, 1001-1014.
- Reimann, C., Filzmoser, P. and Garrett, R.G., 2005. Background and threshold: critical comparison of methods of determination. *Science of the Total Environment* 346, 1-16.
- Rollinson, H., 1993. *Using Geochemical Data: Evaluation, Presentation, Interpretation*. Longman, Essex, 352 pp.
- Smith, R.E. and Perdrix, J.L., 1983. Pisolitic laterite geochemistry in the Golden Grove massive sulphide district, Western Australia. *Journal of Geochemical Exploration* 18, 131-164.
- 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. *Journal of Geochemical Exploration* 32, 169-184.
- Tukey, J.W., 1977. *Exploratory Data Analysis*. Addison-Wesley, Reading, 688 pp.