NON-LINEAR MODELLING OF GEOLOGICAL CONTINUITY

John Henstridge
Data Analysis Australia

Abstract

A frequent problem in the understanding of a mineral deposit is the classification of samples into rock types. While there may be several types of data used in this classification, one of statistical interest is multi-element geochemical analysis. Provided there is a training dataset of samples where rock types have been manually classified and for which the geochemical information exists, it is possible to use discriminant analysis to classify subsequent rock samples into specific types. This is a logical extension of the first step in indicator kriging where samples are divided into two types – above and below a certain grade.

The continuity problem is that individual samples can be subject to significant random variation and thus a simple statistical classification is not smooth or does not reflect the geological continuity. One answer to this when there are just two rock types is that used in indicator kriging. Where there are more than two rock types the kriging approach is no longer satisfactory and a more probabilistic approach is necessary. This study investigates methods developed for regularly spaced data and used a combination of techniques developed by theoretical physicists and statisticians over the past twenty years. The example relates to an iron ore deposit.

Key Words: *geostatistics, non-linear estimation, classification, Metropolis algorithms.*

Introduction

One of the principal applications of geostatistics is the classification of blocks in a resource model into ore and waste. More generally there is a need to classify material into a number of different rock types. This is particularly the case where impurities must be considered as well as the ore grade itself. Examples are where iron ore mine plans must consider phosphorus levels or bauxite mines must consider clay content.
However the classification cannot be properly done on a point-by-point or block-by-block basis since that ignores geological continuity.
Continuity has two distinct though related roles in geostatistics. The first assumption of continuity is that despite the random variation in the data , the underlying structure

at sampling points close to each other is likely to be similar. This provides the basis of the estimation of the variogram and the application of this to point or block estimates through kriging. This essentially sees kriging as an estimation method. The second assumption is that important aspects of the underlying geological structure being studied is continuous and hence the resulting output of the kriging process should reflect this. This leads to kriging as a smoothing process.

Classical kriging assumes that the process being modelled is a continuous variable such as an ore grade. Furthermore the methods, being based on least squares and variances, are at best inefficient and at worst misleading when applied to data that is not normally distributed. However least squares methods are both computationally and theoretically the most tractable. While there is a need to consider models that do not depend upon normality, the penalty for such methods is usually greatly increased computation. Some such methods include disjunctive kriging that introduces some non-linearity (Matheron, 1976) or spatial generalised linear models that introduce the wider class of exponential family distributions (Diggle et al, 1998). The approach taken here relates to spatial models for categorical variables as occur with classification problems.

Discriminant analysis and classification

Discriminant analysis applied to geochemical data simultaneously considers a number of elements and models their joint distributions in each rock type. Typically with iron ore the measures are Fe, SiO_2, Al_2O_3, P, TiO_2 and loss on ignition (LOI). The simplest model assumes that these measures have a joint normal distribution in each rock type, perhaps after suitable transformation such as a power or logarithm. On the basis of these distributions it is possible to define a simple set of rules that can classify samples on the basis of their composition.

This is illustrated by Figure 1 where the contours of the probability density functions for several rock types are illustrated in two dimensions (two elements). The normal distribution assumption implies that the contours are elliptical. On this basis the whole range of possible data values can be divided into several regions such that sample can be allocated to a rock type on the basis of the region they fall into. The boundaries between the regions are generally segments of ellipses. If an assumption is made that the shapes of the distributions are the same for all rock types, the boundaries become straight lines.

While this might seem complex, it is quite simple to implement in practice. At the sample point the probability density function is calculated for the distributions corresponding to each of the rock types and the sample is assigned to the type which has the highest value of the density. The essence of this is the calculation of posterior probabilities $p_1, p_2 \ldots p_k$ that give the likelihood of the sample coming from each of the k rock types and then choosing the rock type that is most likely.

There are a number of other ways in which such a classification could be carried out - it is an area of great activity in statistics. Of particular interest is modern work on non-parametric classification that no longer needs to assume a particular distribution

but does generally require more data to be effective. McLachlan (1992) gives a somewhat mathematical summary.
In what follows we simply assume that there is a classification method that works via posterior probabilities as above.

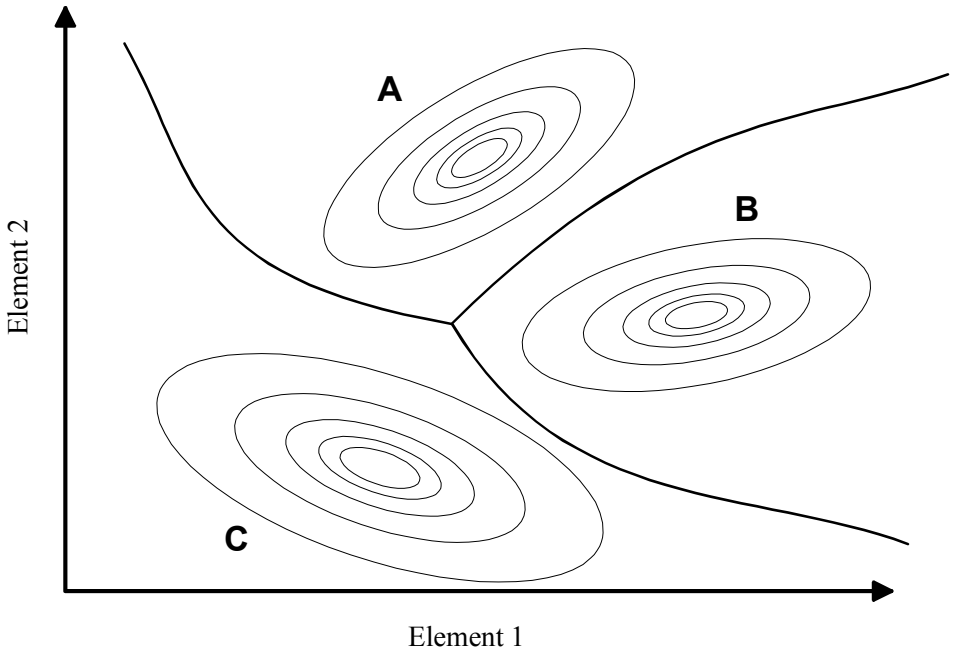

Figure 1. A typical application in two dimensions where three rock types A, B and C have been identified according to a two element assay. The regions correspond to the classification functions.

Continuity Models

The standard geostatistical model assumes that the quantity being measured is defined continuously in the resource volume. Such a model does not readily carry over to the classification problem. Instead it is easier to use a lattice model where just a grid of points in three dimensional space is defined. A good review of lattice models is found in Cressie (1991). This can be thought of as a block model except that we just think of the centroid of each block. As a further simplification we just consider a regular lattice where the samples might be from a regular grid of drillholes and they have a regular vertical spacing.

Continuity can now mean that each lattice point is likely to be similar to each of the adjacent points. The simplest model for adjacency is to take the points immediately to the north, south, east, west, above and below as neighbours. This is just one of the possibilities but with a rectangular grid it is effectively the minimum - it is illustrated in Figure 2. Of course via this nearest neighbour scheme every point in the lattice is generally related to every other point. However it assumes that the relationship works only by these nearest neighbours - a form of Markov property.

The database is structured so that every point has a list of its neighbours. Most points in the interior have six such neighbours while boundary points or those along known discontinuities will have fewer. The database structure is usually such that an arbitrary number of neighbours can be managed which means that virtually any model for nearest neighbours can be used.

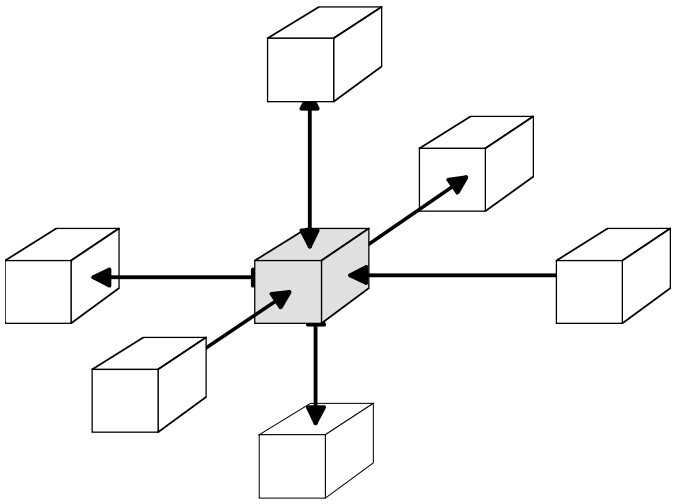

Figure 2. The nearest neighbours of a lattice point. The neighbours of the shaded point are the points immediately to the north, south, east, west, above and below.

The basis of the continuity model is a statement about the probability that a point is of a particular rock type given the knowledge of the rock types at its nearest neighbours. This probability model can take a number of forms but the simplest with that we have had success has the prior probability of a point being of rock type *i* proportional to:

$$\begin{aligned}\exp(&\alpha\,(\#\,N\text{–}S \text{ neighbours of type } i - \#\,N\text{–}S \text{ neighbours not of type } i) + \\ &\beta\,(\#\,E\text{-}W \text{ neighbours of type } i - \#\,N\text{–}S \text{ neighbours not of type } i) + \\ &\gamma\,(\#\,\text{vertical neighbours of type } i - \#\,\text{vertical neighbours not of type } i)).\end{aligned}$$

Here the parameters α, β and γ measure the continuity in the north-south, east-west and vertical dimensions respectively. The posterior probability is then the product of this prior probability and the classification probability discussed earlier. The point is classified to the rock type with the greatest posterior probability.

Notice that this probability model for continuity can be enhanced in many ways. We will give two examples:

- In general the parameters α, β and γ could depend upon the rock type being considered. A typical application might be to enforce certain stratigraphic rules that may say that certain vertical sequences of rock types are impossible. It would also be possible to introduce knowledge that some rock types tend to be more extensive than others are.

- The above model cannot allow for the main axis of variation to be rotated to the grid. This is a problem of the neighbourhood definition rather than the probability function itself. An appropriate neighbourhood model with more points as neighbours can readily correct this. However if each point is defined to have too many neighbours the calculations can get complex.

Algorithms

The fitting of the above models requires three steps, all of which require some care. The first and easiest is defining the basic point-by-point classification method. This is straightforward multivariate statistics. The second and third steps are the estimation of the parameters of the continuity model (the α, β and γ above) and the subsequent classification of the main body of the data. It is convenient to consider the final classification step first.

The classification of each point requires knowledge of how the points around it have been classified but to classify them properly the first point must be classified. There is no direct solution to this and an iterative solution must be used. However the iterative solution is not straightforward since there tend to be many locally optimal solutions that must be avoided. The solution we have adopted is a version of the Metropolis-Hastings or annealing algorithm, a method first applied to such problems by Geman and Geman (1984). Initially the classification is carried out by adding perturbations to each of the posterior probabilities so that the classification is carried out with a controlled level of error. The starting point is the simple classification that ignores the continuity model. The points are then classified sequentially using the perturbed probability. This is repeated while slowly decreasing the scale of the perturbations to zero.

Provided the perturbation process is scaled down slowly enough this method is sure to reach the global maximum likelihood solution even though it never evaluates the likelihood function. The impracticality of direct likelihood solutions to problems of this type has been a long-standing problem that first arose in statistical mechanics. The method also has many connections with the modern Markov Chain Monte Carlo (MCMC) approach that has revolutionised modern Bayesian statistics.

If the method has one criticism apart from the computation involved it is that it is never possible to know whether the annealing or scaling down of the perturbations is applied slowly enough. However it is possible to make the process faster by constantly changing the order in which the points are classified at each step.

The estimation of the continuity model parameters can form a second level of iteration. It is relatively straightforward for the example above and can be fitted by generalised linear model techniques (McCullagh and Nelder, 1989) in a package such as GLIM or S Plus.

An example

A small example of the application of this approach is presented here for an iron deposit that had been drilled on a 50 metre grid, with two metre composite samples taken vertically. A proportion was used as a training data set and manually classified into seven rock types ranging from ore through to the foot-wall underlying the ore body. Approximately 3500 samples were used in the example.

Figures 3 and 5 show the result of the simple classification using a six element multivariate discrimination, after transforming the variables so that a normal distribution was a reasonable model. It can be seen that there is significant irregularity, which is hardly surprising since the element distributions for the different rock types overlapped somewhat. Figures 4 and 6 show the results after an annealing process with just one step, estimating the continuity parameters from the initial classification.

The improvements that come from the continuity model are immediately obvious, giving a result that is physically much more reasonable, particularly for the vertical section. This could be improved even more with a more sophisticated continuity model, such as one that assumes that if a point has foot-wall above it then it must be foot-wall itself.

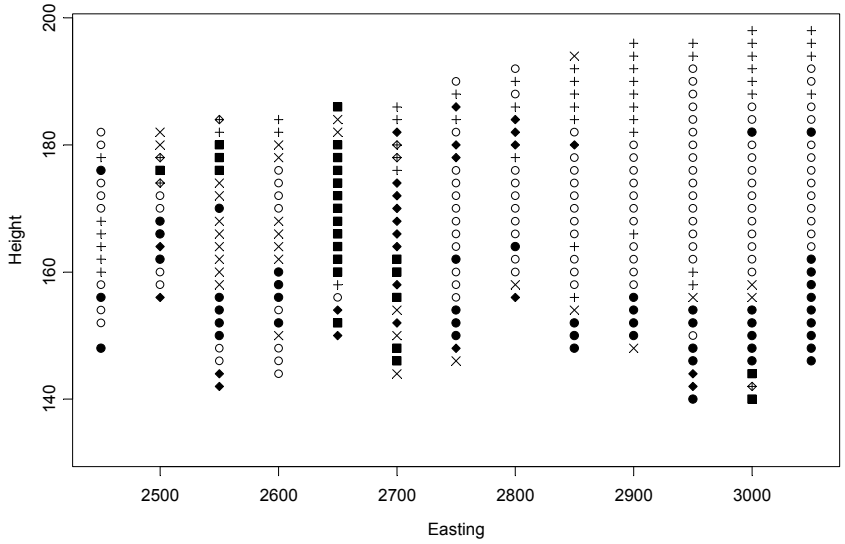

Figure 3. Vertical section of the deposit with the simple classification before applying the continuity model. (+ = weathered layer, ○ and ● = ore types, ■= foot-wall, ◆= clay)

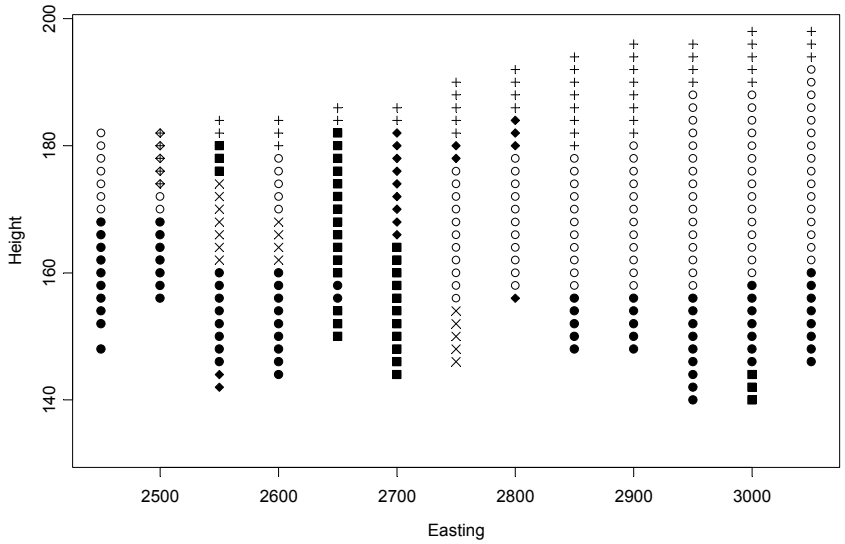

Figure 4. Vertical section of the deposit after applying the continuity model.

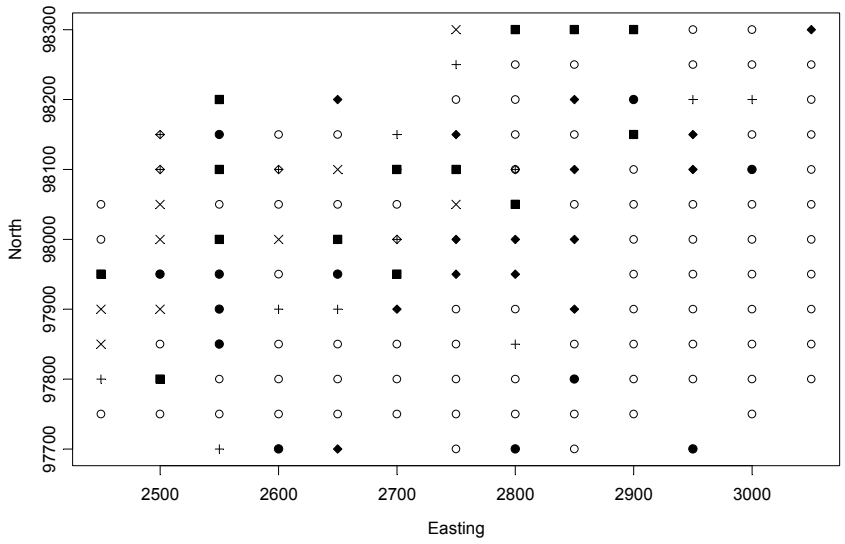

Figure 5. Horizontal section of the deposit with the simple classification before applying the continuity model.

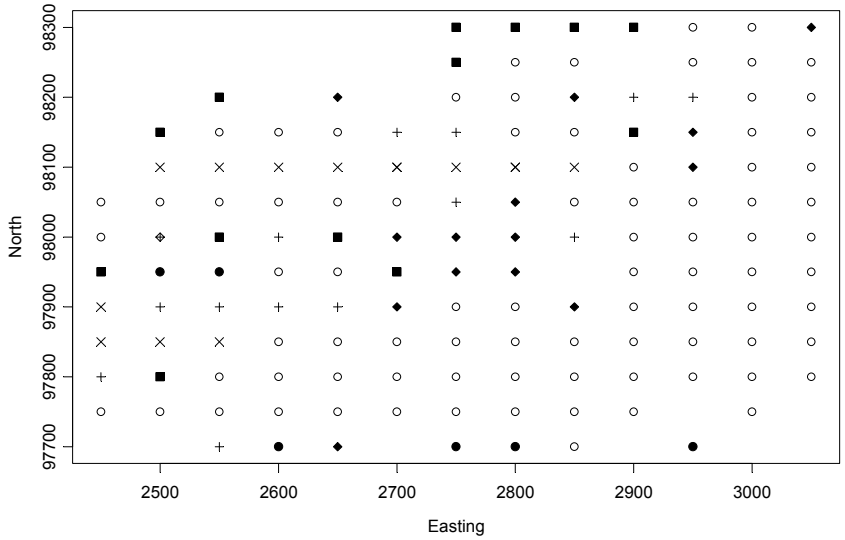

Figure 6. Horizontal section of the deposit with the simple classification after applying the continuity model.

Summary

The cost of these approaches is essentially greater computation for the annealing stage, something that made the methods impractical when first introduced in the early 1980s but which is now less of an issue. Their application is potentially quite wide since they are in some regards an extension to indicator kriging that divides the deposit into just two rock types. There is also a useful modularity to the model since it separates the continuity model from the model for what is happening at a specific point.

Acknowledgments

Cliff Robe River Mining provided the data for the initial development of these ideas. The diagrams were made with the assistance of Amanda Mellican of Data Analysis Australia.

References

Cressie, N.A.C., 1991. *Statistics for Spatial Data*, Wiley, New York.

Diggle, P.J., Tawn, J.A. and Moyeed, R.A., 1998. *Model-based geostatistics*. Applied Statistics, Vol. 47, pp299-350.

Geman, S. and Geman, D., 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, IEEE Transactions of Pattern Analysis and Machine Intelligence, Vol 6, pp 721-741.

Matheron, G., 1976. A simple substitute for conditional expectation: the disjunctive kriging. In *Advanced Geostatistics in the Mining Industry*, Reidal, Dordrecht.

McCullagh, P. and Nelder, J.A., 1989. *Generalised linear Models*, Chapman and Hall, London.

McLachlan, G.J., 1992. *Discriminant Analysis and Statistical Pattern Recognition*, Wiley, New York.