

Chapter 13 ---

Analysis of geochemical data sampled on a regional scale

K. Conradsen, A. A. Nielsen and K. Windfeld

When working with the analysis of geochemical data on a regional scale one will often encounter calibration problems arising from the chemical analysis of samples that might be collected over several years in geographical units. A procedure for removal of such geographical unit patterns is described. Stream sediment samples are collected at 33 992 sites and analyzed for the content of 26 elements. Semivariograms, the geostatistical analogue of spatial autocorrelation functions, are estimated, modeled and applied in interpolating to a regular grid. The interpolation method used is kriging, in which the mean square prediction error is minimized taking the spatial correlation into account. Heavy mineral samples are collected at 3094 sites and analyzed for the content of several economically interesting minerals. Based on the nonparametric method, CART (classification and regression trees) probability maps for the occurrence of gold and cassiterite are constructed using the kriged stream sediment data as predictors. This method is very useful for geologists when studying geological structures on a regional scale.

13.1 Geostatistics

The basis of geostatistics is the idea of considering the observed values of a geochemical, a geophysical or another natural variable at a given set of positions as a realization of a stochastic process in space. For each position \mathbf{x} in a domain \mathcal{D} there exists a measurable quality $z(\mathbf{x})$, a so-called *regionalized variable*. \mathcal{D} is typically a subset of \mathcal{R}^2 or of \mathcal{R}^3 . z is considered a particular realization of a *random variable* $Z(\mathbf{x})$. The set of random variables $\{Z(\mathbf{x}) | \mathbf{x} \in \mathcal{D}\}$ constitutes a random function. $Z(\mathbf{x})$ has mean value $E\{Z(\mathbf{x})\} = \mu(\mathbf{x})$ and covariance $\text{cov}\{Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h})\} = C(\mathbf{x}, \mathbf{h})$. If the covariance is translation invariant over \mathcal{D} , i.e. $C(\mathbf{x}, \mathbf{h}) = C(\mathbf{h})$, Z is said to be second-order stationary. Often $Z(\mathbf{x})$ is assumed to follow a normal or a lognormal distribution. This statistical view on natural phenomena was introduced by Georges Matheron in 1962–63 and is described in great detail in David (1977) and in Journel and Huijbregts (1978). An introductory textbook is Clark (1979). David (1988) looks back on ten years of application of geostatistics.

The classical application of geostatistics has been the calculation of ore reserves. Here it is applied to describe the spatial distribution of geochemical

elements over large areas (in the order of tens of kilometers by tens of kilometers). The spatial autocorrelation structure is described by means of semivariograms and if such autocorrelation is present it is utilized in interpolation performed by kriging, a best linear unbiased estimator (BLUE).

Point measurements of geochemical, geophysical or other natural variables or measurements taken over areas or volumes, also known as *supports*, are in principle continuous phenomena in space. If 'dense' sampling is performed the continuous nature of the variable in question will be reflected in the covariation of neighboring samples. If taken further apart from each other there will be little or no covariation between samples. Whether samples are 'dense' depends on the variable in question and sample sizes. Also, the autocorrelation revealed will depend on the scale at which one is operating. Different autocorrelation structures can be present simultaneously at different scales (mineralizations at the size of a few meters vs. regional variations at the size of tens of kilometers); this is referred to as *nested structures*.

The above remarks on autocorrelation applies to cross-correlations also if more than one variable is studied at a time.

13.1.1 The semivariogram

Consider two scalar values $z(\mathbf{x})$ and $z(\mathbf{x} + \mathbf{h})$ measured at two points in space \mathbf{x} and $\mathbf{x} + \mathbf{h}$ separated by \mathbf{h} . z is considered a particular realization of a random variable Z . The variability is described by the autocovariance function

$$C(\mathbf{x}, \mathbf{h}) = E\{(Z(\mathbf{x}) - \mu)(Z(\mathbf{x} + \mathbf{h}) - \mu)\}.$$

The *variogram* is defined as

$$2\gamma(\mathbf{x}, \mathbf{h}) = E\{[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})]^2\}.$$

In general the variogram will depend on the location in space \mathbf{x} and on the displacement vector \mathbf{h} . Note that the variogram represents a more general concept than that of the covariance function since the increment process $Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})$ may have well-defined properties which the basic process $Z(\mathbf{x})$ does not possess. The *intrinsic hypothesis* in geostatistics is that the variogram is independent of the location in space and that it depends on the displacement vector only:

$$2\gamma(\mathbf{x}, \mathbf{h}) = 2\gamma(\mathbf{h}).$$

Second-order stationarity of $Z(\mathbf{x})$ implies the intrinsic hypothesis. $\gamma(\mathbf{h})$ is called the *semivariogram*. The autocovariance function and the semivariogram are related by

$$\gamma(\mathbf{h}) = C(\mathbf{o}) - C(\mathbf{h}).$$

Note that $C(\mathbf{o}) = \sigma^2$.

An estimator for the semivariogram is the mean of the squared differences between any two measurements $z(\mathbf{x}_i)$ and $z(\mathbf{x}_i + \mathbf{h})$:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})]^2,$$

where $N(\mathbf{h})$ is the number of point pairs separated by \mathbf{h} . $\hat{\gamma}$ is called the *experimental semivariogram*. Pooling in both magnitude and argument of \mathbf{h} is often performed. Pooling in the magnitude of \mathbf{h} – i.e. the distance between samples – is done to obtain a sufficiently high $N(\mathbf{h})$ to ensure a small estimation variance ($N(\mathbf{h})$ is proportional to the estimation variance). Pooling in the argument of \mathbf{h} – i.e. the direction – is done to check for anisotropy.

There is an extensive literature on the problems one encounters when estimating experimental semivariograms on real world data, cf. e.g. Cressie (1985).

In order to be able to define characteristic quantities for the semivariogram (and in order to apply the semivariogram in kriging, see below) a model is often assumed. An often used semivariogram model is the spherical model with nugget effect. A reason for this is the easy interpretability of the parameters. Assuming isotropy and setting $|\mathbf{h}| = h$ the form of this model is

$$\gamma^*(h) = \begin{cases} 0 & \text{if } h = 0 \\ C_0 + C_1 \left[\frac{3h}{2R} - \frac{1}{2} \left(\frac{h}{R} \right)^3 \right] & \text{if } 0 < h < R \\ C_0 + C_1 & \text{if } h \geq R, \end{cases}$$

where C_0 is the *nugget effect* and R is the *range of influence*. $C_0/(C_0 + C_1)$ is the relative nugget effect and $C_0 + C_1$ is the *sill* ($= \sigma^2$). The nugget effect is a discontinuity in the autocorrelation function at $h = 0$ due to both measurement errors and to micro-variability, the structure of which is not available at the scale of study. This variability thus turns up as noise. The range of influence is the distance at which covariation between measurements stops; measurements taken further apart are uncorrelated. The spherical semivariogram model with nugget effect can easily be extended to e.g. a double spherical model with nugget effect to allow for nested structures:

$$\gamma^*(h) = \begin{cases} 0 & \text{if } h = 0 \\ C_0 + C_1 \left[\frac{3h}{2R_1} - \frac{1}{2} \left(\frac{h}{R_1} \right)^3 \right] + C_2 \left[\frac{3h}{2R_2} - \frac{1}{2} \left(\frac{h}{R_2} \right)^3 \right] & \text{if } 0 < h < R_1 \\ C_0 + C_1 + C_2 \left[\frac{3h}{2R_2} - \frac{1}{2} \left(\frac{h}{R_2} \right)^3 \right] & \text{if } R_1 < h < R_2 \\ C_0 + C_1 + C_2 & \text{if } h \geq R_2, \end{cases}$$

where C_0 is the nugget effect and R_2 is the range of influence. $C_0/(C_0 + C_1 + C_2)$ is the relative nugget effect and $C_0 + C_1 + C_2$ is the sill. Other models for the semivariogram such as linear, bilinear and exponential models are often used also.

The parameters in the above semivariogram models γ^* can be estimated from the experimental semivariograms $\hat{\gamma}$ by means of iterative, nonlinear least squares methods. Different weights of the estimated values in the experimental semivariogram $\hat{\gamma}$ may be considered. A weighting with the

number of point pairs included in the estimation for each lag distance seems natural. Also, if one is interested in a good model for small lags a weighting with the inverse lag distance applies.

Another important concept in geostatistics is *regularization*, i.e. the averaging of a random function over a domain \mathcal{D} . Let $\mathbf{x} \in \mathcal{D}$. The regularized value $Z_{\mathcal{D}}$ of Z is

$$Z_{\mathcal{D}} = \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} Z(\mathbf{x}) \, d\mathbf{x},$$

where $|\mathcal{D}|$ is the area (or volume) of \mathcal{D} . Similarly for the moment functions, e.g.

$$\bar{\gamma}(\mathcal{A}, \mathcal{B}) = \frac{1}{|\mathcal{A}| |\mathcal{B}|} \int_{\mathcal{A}} \int_{\mathcal{B}} \gamma(\mathbf{x} - \mathbf{y}) \, d\mathbf{y} \, d\mathbf{x}.$$

Thus $\bar{\gamma}(\mathcal{A}, \mathcal{B})$ is the regularized semivariogram when one end of the displacement vector $\mathbf{h} = \mathbf{x} - \mathbf{y}$ varies in \mathcal{A} and the other end of \mathbf{h} varies in \mathcal{B} . This integral can be either solved analytically for certain semivariogram models and supports of simple geometry or solved numerically.

What is said above about autocovariance functions and variograms can easily be extended to covariance functions and cross-variograms if more variables are studied simultaneously.

13.1.2 Kriging

Suppose that the random variable $Z(\mathbf{x})$ is sampled on a number of supports (could be points) $\mathcal{D}_1, \dots, \mathcal{D}_n$ giving the following scalar measurements $z(\mathbf{x}_1), \dots, z(\mathbf{x}_n)$. We now want to estimate $Z_{\mathcal{D}}$ on a support \mathcal{D} where Z is not sampled (or Z is sampled on a part of \mathcal{D} only). We are looking for a linear, unbiased estimator:

$$\hat{Z}_{\mathcal{D}} = \sum_{i=1}^n w_i Z_{\mathcal{D}_i}$$

$$E\{\hat{Z}_{\mathcal{D}} - Z_{\mathcal{D}}\} = 0.$$

The unbiasedness condition of the estimator gives

$$\sum_{i=1}^n w_i = 1.$$

The estimation variance (or the mean squared error) is

$$\sigma_E^2 = E\{(\hat{Z}_{\mathcal{D}} - Z_{\mathcal{D}})^2\}.$$

The *kriging estimate* is defined by the values of the weights w_i that minimize the estimation variance σ_E^2 subject to the constraint that the sum of the (kriging) weights is unity. This can be done by introducing a Lagrangian multiplier and setting each of the n partial derivatives

$\partial[\sigma_E^2 - 2\lambda(\sum_{i=1}^n w_i - 1)]/\partial w_i = 0$ leading to the $(n + 1) \times (n + 1)$ set of equations

$$\sum_{i=1}^n w_i \bar{\gamma}(\mathcal{D}_j, \mathcal{D}_i) + \lambda = \bar{\gamma}(\mathcal{D}_j, \mathcal{D}), \quad j = 1, \dots, n$$

$$\sum_{i=1}^n w_i = 1$$

with the *kriging variance* (or the minimum mean squared error)

$$\sigma_K^2 = \sum_{i=1}^n w_i \bar{\gamma}(\mathcal{D}_i, \mathcal{D}) + \lambda - \bar{\gamma}(\mathcal{D}, \mathcal{D}).$$

Of course this can be expressed in terms of the covariance function also

$$\sum_{i=1}^n w_i \bar{C}(\mathcal{D}_j, \mathcal{D}_i) - \lambda = \bar{C}(\mathcal{D}_j, \mathcal{D}), \quad j = 1, \dots, n$$

$$\sum_{i=1}^n w_i = 1.$$

with the kriging variance

$$\sigma_K^2 = - \sum_{i=1}^n w_i \bar{C}(\mathcal{D}_i, \mathcal{D}) + \lambda + \bar{C}(\mathcal{D}, \mathcal{D}).$$

Both kriging systems can be written in matrix form, here expressed by means of the covariance functions

$$\mathbf{C}\mathbf{w} = \mathbf{c}$$

with

$$\mathbf{C} = \begin{bmatrix} \bar{C}(\mathcal{D}_1, \mathcal{D}_1) & \dots & \bar{C}(\mathcal{D}_1, \mathcal{D}_n) & 1 \\ \vdots & & \vdots & \vdots \\ \bar{C}(\mathcal{D}_n, \mathcal{D}_1) & \dots & \bar{C}(\mathcal{D}_n, \mathcal{D}_n) & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix},$$

$$\mathbf{w}' = (w_1, \dots, w_n, -\lambda)$$

and

$$\mathbf{c}' = (\bar{C}(\mathcal{D}_1, \mathcal{D}), \dots, \bar{C}(\mathcal{D}_n, \mathcal{D}), 1).$$

The solution to the kriging system expressed in matrix form by means of the covariance function is

$$\mathbf{w} = \mathbf{C}^{-1} \mathbf{c}.$$

If the supports $\mathcal{D}, \mathcal{D}_1, \dots, \mathcal{D}_n$ can be considered points the kriging performed is referred to as *point kriging*, otherwise it is referred to as *block kriging* or *panel kriging*.

A few remarks on some very important properties of kriging:

- Kriging is an interpolation form that provides us with not only an estimate based on the covariance structure of the variable in question but also an estimation variance.
- The kriging system has a unique solution if and only if the covariance matrix $\{C_{ij}\}$, $i, j = 1, \dots, n$ is positive definite; this also ensures a nonnegative kriging variance.
- The kriging estimator is a best linear unbiased estimator (BLUE) and it is also exact, i.e. if the support to be estimated coincides with any of the supports of the data included in the estimation, kriging provides an estimator equal to the known measurement and a zero kriging variance.
- The kriging system and the kriging variance depend only on the covariance function (semivariogram) and on the spatial lay out of the sampled supports and not on the actual data values. If a covariance function is known (or assumed) this has important potential for minimizing the estimation variance in experimental design (i.e. in the planning phase of the spatial layout of the sampling scheme).

In all applications of kriging the problem of assuming stationarity arises. *Universal kriging* is a technique that allows for some forms of nonstationarity part of which is modeled as a trend in the mean value, that is described either as a linear combination of known functions or by means of local Taylor expansions. The type of spatial irregularity we are facing in the application below is not suitable for that type of solution. The nonstationarity is very nonlinear and all attempts at 'fitting' with 'regular' functions and models have failed. Another possible approach when analyzing multivariate observations is *co-kriging*. Here one could take the spatial covariation between different variables into account in the 'adjustment'. Again, the spatial irregularity in the case study below is not suited for that type of solution. Hence, the *ordinary kriging* method described above is applied in the case study. Universal kriging, co-kriging and other advanced types of kriging will not be elaborated on here; good references are Journel and Huijbregts (1978), Myers (1982) and Carr (1985).

13.1.3 Case study: central Spain

This section describes applications of geostatistical methods in the analysis of stream sediment geochemical data from a large area in central Spain. The work reported constitutes part of a project described in Conradsen *et al.* (1990).

(a) Geochemical data

Samples were collected at 33 992 sites in 16 mapsheets (a mapsheet covers approximately 30 km \times 20 km). All samples were analyzed for the contents of 27 geochemical elements by inductively coupled plasma emission spectrometry (ICP). The elements are

P, As, Sb, Sn, Pb, B, Zn, Cd, Hg, Cu, Ag, Ni, Co, Fe,
Mn, Cr, Mo, W, V, Nb, Y, Be, Ba, Al, Mg, Ti and Sc.

For proprietary reasons, all contents have been scaled by a constant factor. Each observation holds information on UTM coordinates (universal transverse Mercator projection), element contents and lithology code.

Sc is a multiplicative correction element added manually and not present in nature in these samples. For several elements there is a detection limit problem. Based on information from geologists the following procedure was decided upon:

- If the content z of an element is < 5 p.p.m. replace that value with 3.5 p.p.m. which is the approximate center of mass for a triangle with baseline $[0, 5]$ p.p.m.
- If the content z_i of an element at sample site i is ≥ 5 p.p.m. replace that value with $z_{i,corr} = z_i Sc_{mean} / Sc_i$.

The value of Sc_{mean} can be chosen as the overall mean or as a mean for individual mapsheets. As no visual difference was observed the simpler method applying the overall mean was preferred.

When inspecting sample site images of the individual elements a conspicuous mapsheet effect is noted, cf. plate. The effect is due to calibration problems in the chemical analysis of the samples. To facilitate a regional study the effect is removed by means of a method explained below.

(b) Semivariograms and kriging

To reveal the spatial structure of the individual elements and to prepare the interpolation by means of kriging experimental semivariograms were calculated. To allow for the mapsheet effect semivariograms were calculated before the removal of this effect but without allowing interaction between mapsheets, i.e. no point pair across any mapsheet border was included in the calculation. Spherical models with nugget effect were estimated by means of a weighted, iterative, nonlinear least-squares estimation procedure. Figure 13.1 shows semivariograms for the natural logarithm of the content of some elements (measured in p.p.m.).

Point kriging to a regular UTM grid with points in a 500 m \times 500 m grid was performed for individual elements using the nearest 20 neighbors resulting in 192 lines with 234 samples each. A speed-up feature used is the exploitation of the fact that all elements have the same neighbors, based on the sampling layout only to build a neighborhood table. The application of this speed-up feature and point (vs. block) kriging reduced the CPU time for kriging one element in the study area from approximately 26 hours to approximately 2 hours and 10 minutes on a DEC MicroVAX II. The plate shows kriged Cr and kriging variance for Cr.

In a small part of the study area (one mapsheet) a pilot study on co-kriging and kriging of principal components was carried out. This study showed very little visual difference between these more advanced forms of kriging and ordinary kriging.

(c) Mapsheet calibration

One of the major problems in the analysis of stream sediment data in the study area is the very distinct mapsheet pattern observed when displaying

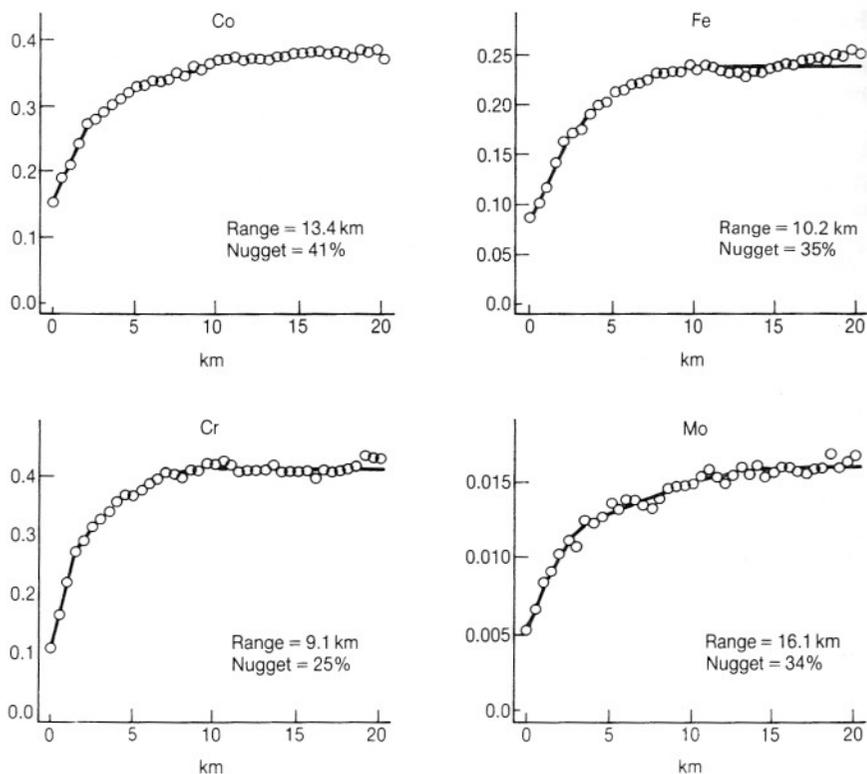


Figure 13.1 Experimental and modeled semivariograms for some elements in central Spain.

the samples as images. This obvious nonstationarity is due to problems in the chemical analysis of the samples and it violates the assumptions underlying many of the statistical techniques used in the analysis of spatial patterns. Therefore a method for removing the mapsheet pattern and thus calibrating the data is called for when the object is to analyze the area as a whole.

The general idea in the methods considered for mapsheet pattern removal is to transform observed values in the mapsheets using monotone transformations in order to preserve structure such that observations in neighboring mapsheets that are close to each other in distance have similar histograms.

Linear transforms are simple monotone transformations. It turns out that linear transformations are not sufficient for removing the mapsheet pattern. The phenomenon is more complex.

A method based on local histogram matching was found to have a satisfactory performance, removing the mapsheet pattern while preserving the spatial structure in the images. The method starts out in one mapsheet, transforming the neighboring mapsheets to match at the common borders. Afterwards the neighboring mapsheets of these transformed mapsheets are transformed. The scheme propagates throughout the whole area until all

mapsheets conform to one another in the sense that there are no discontinuities along the mapsheet limits in the image.

Consider for example the mapsheet layout in Figure 13.2.

The transformation scheme can start out in any mapsheet. For simplicity let us start in mapsheet 1 (MS1). Then the method works as follows:

Leave the observations in MS1 unchanged. Now, consider the observations in a narrow band on each side of the mapsheet limit shared by MS1 and MS2. To transform the observations in MS2 use the transformation that makes the histogram of the narrow band in MS2 equal to that of the narrow band in MS1.

In determining the transformation a number of empirical quantiles – for example, the 5, 10, 15, . . . per cent quantiles – are determined from each of the two narrow bands. The transformation takes the quantiles in MS2 to the corresponding quantiles in MS1. Between the quantiles we have used linear interpolation. Outside the range specified by the chosen quantiles a linear transformation from $[\min, \min]$ to the lowest quantile and from the highest quantile to $[\max, \max]$, respectively, is used; \min denotes the overall minimum, \max the overall maximum of the element in question. This transformation removes a possible discontinuity along the mapsheet limit between MS1 and MS2. Analogously, MS3 is transformed to remove discontinuities along the mapsheet limit shared by MS1 and MS3.

Now, in transforming MS4, two transformations are determined: one corresponding to MS2 and one corresponding to MS3. The actual transformation applied is a convex combination of these two. For each observation in MS4 the weight on each transformation is inversely proportional to the distance to the corresponding mapsheet limit. Thus, the observation z in MS4 with distance d_2 and d_3 to MS2 and MS3 respectively gets transformed

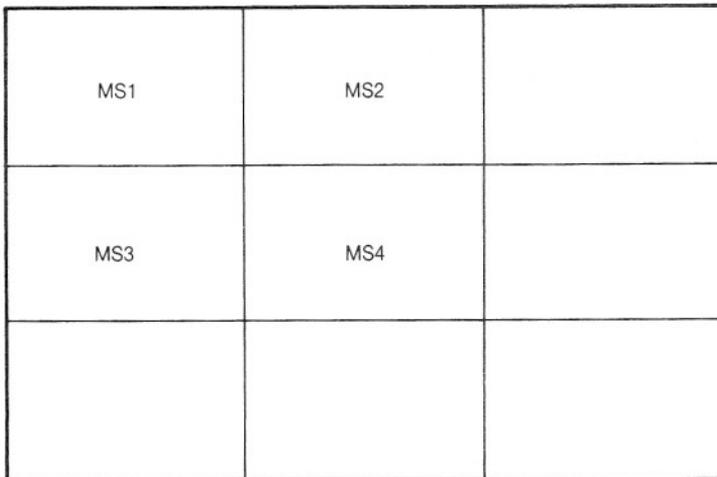


Figure 13.2 Example mapsheet layout.

with the transformation T :

$$T(z) = \frac{d_3}{d_2 + d_3} T_2(z) + \frac{d_2}{d_2 + d_3} T_3(z)$$

where T_2 and T_3 are the two transformations that correspond to MS2 and MS3. It is easy to generalize this idea to transforming a mapsheet to match more than two neighboring mapsheets by using convex combinations of transformations.

The transformation scheme propagates this way until all mapsheets have been transformed.

The mapsheet smoothing is carried out on the logarithm of the Sc-corrected variables. Observations below 5 p.p.m. have been simulated from a triangular distribution before Sc-correction in order to avoid problems with degenerate histograms in the determination of smoothing transformations.

Two examples of the transformation are shown in Figure 13.3. The result of the calibration can be seen in the plate. Note that the very distinct mapsheet pattern present before the calibration is removed. At the same time the overall spatial structure is preserved.

13.2 Prediction of heavy minerals

CART (Classification And Regression Trees) is a new and often powerful alternative to classical parametric methods in classification and regression. The methodology was developed in the 1970s and early 1980s and it is described in Breiman *et al.* (1984). In this section we will briefly describe what CART does, focusing only on the classification part.

In classification, one has measurements on an object and then one uses some sort of *decision rule* to decide to which class the object belongs. In our case we have measurements of concentrations of some geochemical variables

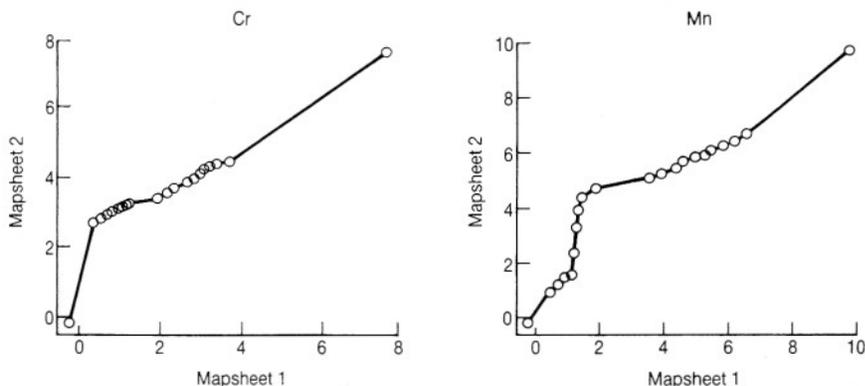


Figure 13.3 Examples of mapsheet transformations.

and we want to predict whether the concentration of gold, for example, is detectable or not. To construct a decision rule we need some data for which we know to what classes the cases belong. This is called the *training set* or the *learning set*. The decision rule that CART produces is in the form of a *binary decision tree*. An example is given in Figure 13.4.

CART grows trees by operating on the training set of the data. A *node* in the tree corresponds to a subset of the data. To begin with, we have a node containing all the cases in the learning set (training set).

In the tree-growing process, we need a strategy for determining the *splits* in the tree. Suppose we only have two classes: 0 and 1. A split of a node is good if it does a good job of separating class 0 and class 1 in the two descending nodes. One rule is to maximize decrease in node *impurity* for each split. For this we need a measure $i(t)$ of node impurity for a node t . The *Gini index* has been found useful. It has the form

$$\begin{aligned}
 i(t) &= \sum_{j \neq k} p(k|t)p(j|t) \\
 &= \left(\sum_{j \neq k} p(k|t) \right)^2 - \sum_k p^2(k|t) \\
 &= 1 - \sum_k p^2(k|t).
 \end{aligned}$$

This function has the desired properties of symmetry, concavity and non-negativeness required, and it is computationally attractive. This is not insignificant.

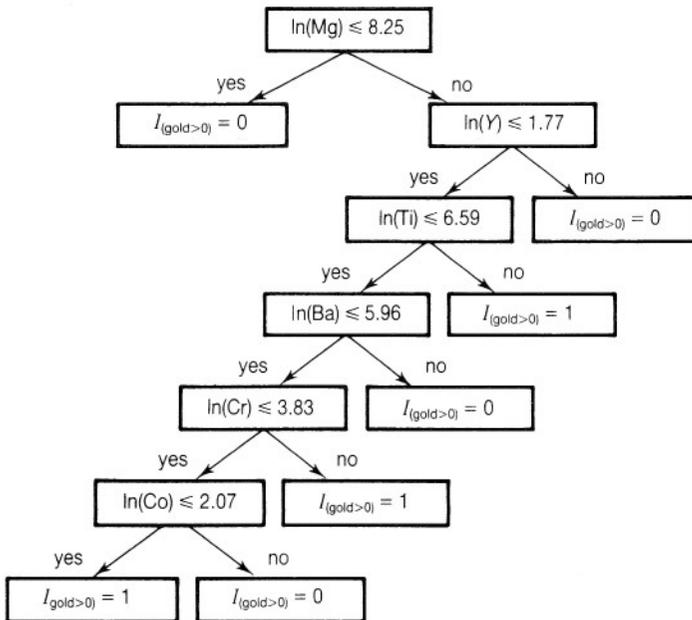


Figure 13.4 CART tree for gold.

nificant since the function is evaluated many times in the process of growing the tree.

CART examines all the coordinate splits of the form $x < c$, where $c \in [\min(x), \max(x)]$ and x is a measured variable. By this we mean: all cases satisfying $x < c$ go left and the rest go right.

For each x there will exist a value of c that yields the best split in terms of separating the classes. The split chosen by CART is the best of the best individual variable splits, resulting in the best overall split on a single variable. This procedure is repeated in a recursive manner for the descendant nodes, until a very large tree is constructed with pure terminal nodes or very few cases in these. This means that if we run all cases down the tree, use the *plurality rule* for predicting the class (i.e. the rule that assigns to a terminal node the class for which the proportion of learning set cases is largest), all cases will be predicted correctly, i.e. the resubstitution estimated misclassification rate is 0! Of course this is too optimistic. We have modeled the 'noise' as well as the structure in the training data. Hence, we need a more 'honest' estimate of the true error rate. This can be accomplished by setting aside some cases from the learning set in a *test* set not used in the actual tree-growing process. This test set is used to estimate the true error rate by running it down the tree and recording the proportion of misclassified cases. The concept of setting aside learning set cases for model selection and error estimation is generally denoted *cross-validation*, a common technique in the context of non-parametric computer-intensive models.

The large initial tree grown is too large so we need to prune it in a reasonable manner. A *cost-complexity* measure is introduced of the form

$$\text{total cost} = \text{misclassification cost} + \alpha \times \text{complexity of tree.}$$

For each value of α there exists a unique subtree of the initial tree, minimizing the total cost. By increasing α starting at 0 we obtain a sequence of smaller and smaller subtrees. The final tree selected in the sequence is the one that minimizes the error rate as estimated by use of the test set.

CART has features for handling missing data, linear combinations splits, variable importance ranking, class priors and misclassifications costs as well; for further details cf. Breiman *et al.* (1984).

13.2.1 Case study: Central Spain

The CART methodology has been used to construct tree-structured predictors for heavy minerals of economic interest in central Spain. The motivation for constructing models for prediction of heavy mineral data using the stream sediment data is that we want to estimate structural maps of heavy minerals in the area. Only the structural map for cassiterite is described below.

Conventional methods for interpolation such as kriging are not feasible because of the extreme skewness of the distribution of heavy minerals concentration at the sample points and also because of sparsity of heavy minerals samples. Essentially an *indicator* variable for the occurrence of a heavy mineral is recorded. The stream sediment data are sampled more densely so that if we can build a model for predicting heavy mineral occurrences based

on stream sediment variables, we can get an interpolated image using the kriged images of the stream sediments as predictors.

Stream sediment samples were collected at 33 992 sites in the study area. At some of these sites heavy mineral samples were collected also. Two different laboratories were used for the chemical analysis of heavy minerals. For calibration reasons only the samples analyzed by one laboratory were used in constructing the models. There are 3094 samples from this laboratory, i.e. there are 3094 observations of stream sediment and heavy mineral data at the same locations.

The CART methodology was chosen because it is a new and powerful technique used with success in other research areas. It handles different types of data in an elegant and unified way and interactions between predictors are automatically taken into account. Furthermore results have shown a good agreement between the structural images obtained using the CART trees and geological information from the area.

(a) CART trees for cassiterite

The distribution of cassiterite concentration is extremely skewed so an indicator $I_{(cass > 0.05)}$ for cassiterite concentration greater than 0.05 g/10l rather than the original concentrations was used in the analysis. The model should then predict whether the cassiterite concentration is detectable or not based on geochemical variables concentrations. Using the CART methodology, the tree shown in Figure 13.5 was selected.

In the construction process of the tree, 1/3 of the observations (994 cases, the test set) were used in the cross-validation of the model and for choosing

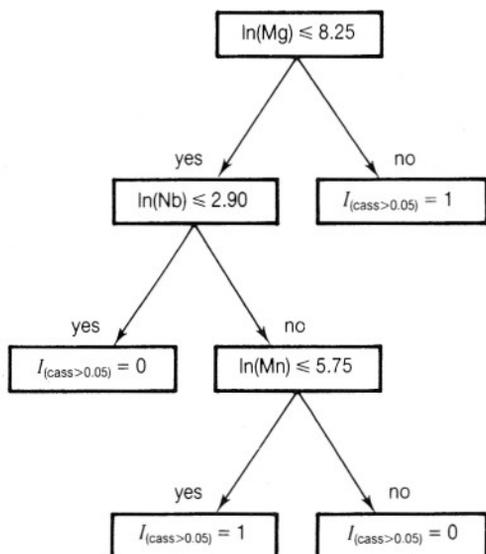


Figure 13.5 CART tree for cassiterite.

a right-sized tree. Running these cases down the tree yielded the classification matrix shown in Table 13.1.

As a standard feature CART makes a variable importance ranking based on a measure of predictive value of the individual predictors (here the stream sediment variates). The variable importance list for predicting cassiterite is given in Table 13.2.

Now, having constructed the tree we are able to run new cases down the tree and predict cassiterite occurrence. Also associated with each terminal node in the tree is an estimate of the probability of cassiterite occurrence, given that a case ends up at that particular node. These estimates can be obtained from running the test set down the tree.

To construct a structural image of the probability of cassiterite occurrence we run each pixel down the tree using the kriged images of stream sediment variables. The image for cassiterite is shown in the plate. Also shown in the plate is a structural image of the probability of gold occurrence. Different gray tones in these images correspond to different terminal nodes in the classification trees.

Table 13.1 Test set classification matrix for cassiterite.

	Predicted class		Total
	0	1	
True 0	768 (79%)	204 (21%)	972 (100%)
Class 1	6 (27%)	16 (73%)	22 (100%)
Total	774 (78%)	220 (22%)	994 (100%)

Table 13.2 CART variable importance ranking for predicting cassiterite.

Variable	Relative importance
Ti	100
Mg	52
Nb	46
Al	36
Zn	35
Cu	35
Ni	33
Cr	31
V	31
Ba	28
Co	28
Y	26
Fe	26
As	25
Mn	25
V	24
Sb	21
Pb	21
Sn	12
Be	12

13.3 Conclusion

A method for handling nonstationarity problems due to calibration problems for spatial data has been presented. The method uses a propagating weighted transformation scheme. It is simple and works well in practice.

Also we have shown an application of a combination of classical geostatistical interpolation and CART, a modern tree-structured classification method. We have found an interesting connection between the tree-structure of the classifier and the geological structures in the images. This can be seen in the structural image produced by running the pixels of the interpolated images down the classification tree.

One should be careful in interpreting the classification trees. Spatial correlation can be responsible for the correlation between the heavy minerals response and the stream sediment sample values. This means that a classification tree cannot be extrapolated to another area outside the images analyzed.

Acknowledgements

The work reported is part of a larger project funded by the Commission of the European Communities under Contract No. MA1M-0015-DK(B) and Minas de Almadén y Arrayanes, S. A. Also, the MOBS programme under the Danish Technical Research Council contributed under grant number 5.26.09.07.

The fruitful cooperation with other members of the IMSOR Image Group, the Minas de Almadén y Arrayanes, S. A., Geology Department and employees of the Commission is acknowledged. Especially, we would like to thank Enrique Ortega, MAYASA, for his enthusiasm and helpful comments.

References

- Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (1984). *Classification And Regression Trees*, Wadsworth.
- Carr, J.R. (1985). Co-kriging – a computer program. *Computers and Geosciences*, **11** (2).
- Clark, I. (1979). *Practical Geostatistics*, Elsevier Applied Science, London.
- Conradsen, K., Ersbøll, B.K., Nielsen, A.A., Pedersen, J., Stern, M. and Windfeld, K. (1990). *Development and Testing of New Techniques for Mineral Exploration Based on Remote Sensing, Image Processing Methods and Multivariate Analysis*. IMSOR, Technical University of Denmark.
- Cressie, N. (1985). Fitting Variogram Models by Weighted Least Squares. *J. Int. Assoc. Math. Geol.*, **17** (5).
- David, M. (1977). *Geostatistical Ore Reserve Estimation*, Developments in Geomathematics 2, Elsevier, Amsterdam.
- David, M. (1988). *Handbook of Applied Advanced Geostatistical Ore Reserve Estimation*, Developments in Geomathematics 6, Elsevier, Amsterdam.
- Journel, A.G. and Huijbregts, Ch.J. (1978). *Mining Geostatistics*. Academic Press, London.
- Myers, D.E. (1982). Matrix Formulation of Co-kriging. *J. Int. Assoc. Math. Geol.*, **14** (3).

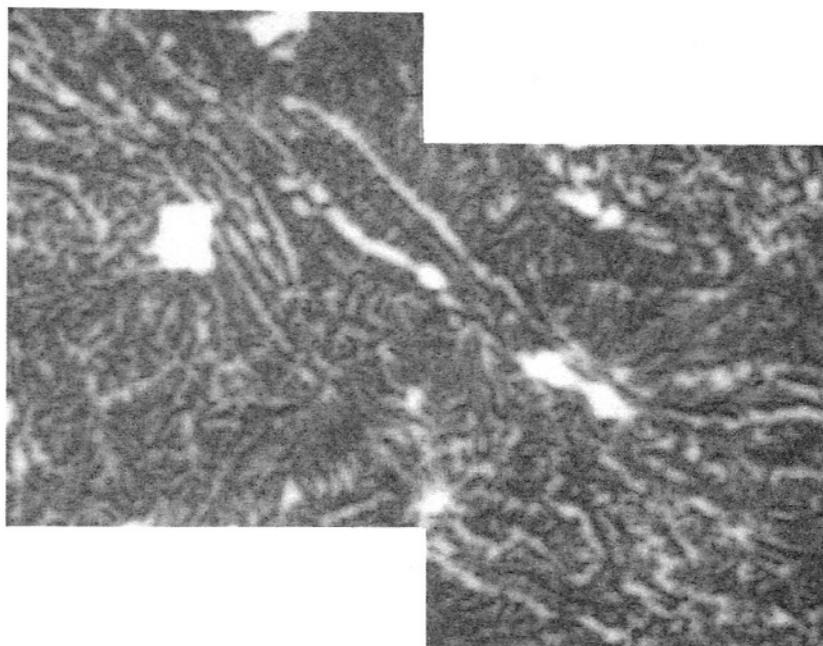


Figure 13.6 Cr before mapsheet pattern removal.

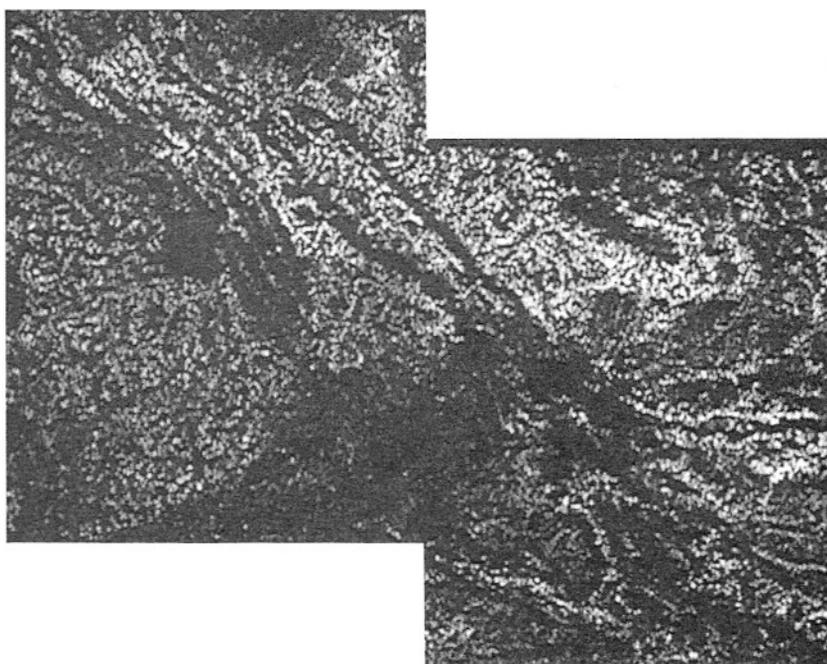


Figure 13.7 Cr after mapsheet pattern removal.

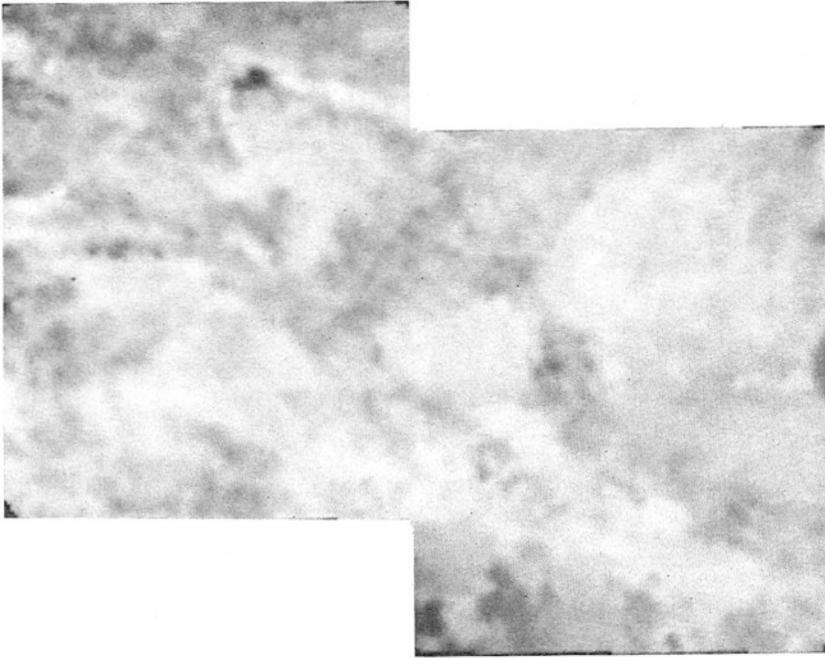


Figure 13.8 Kriged Cr.

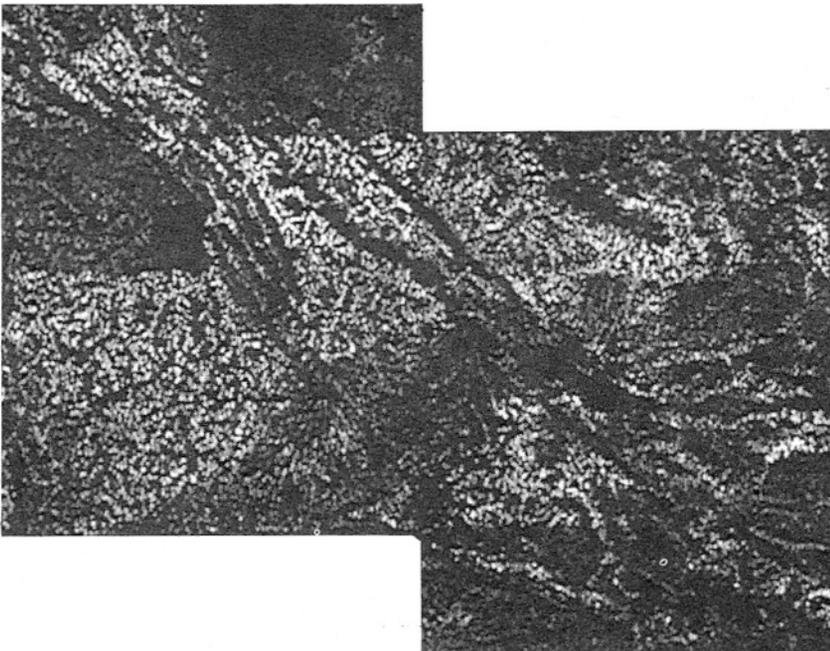


Figure 13.9 Kriging variance for Cr.

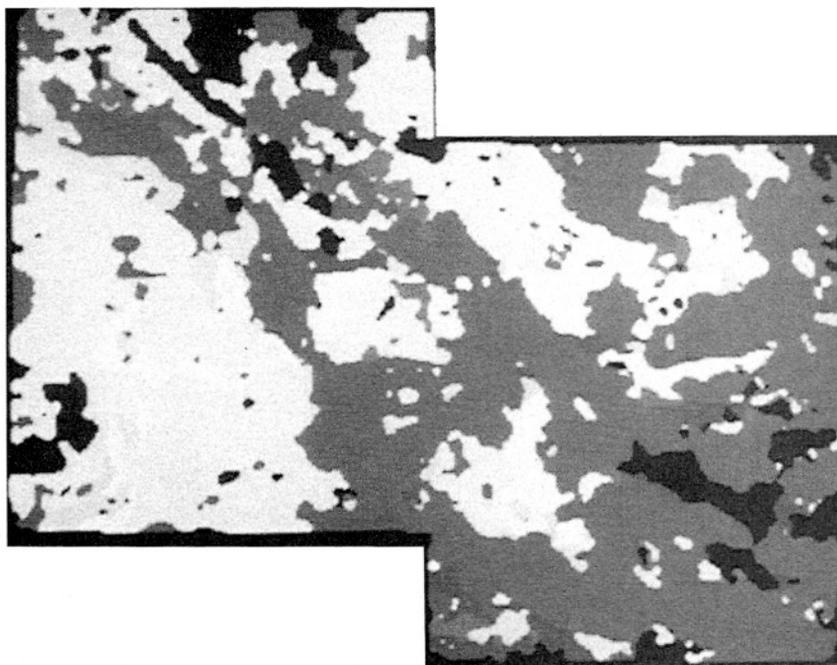


Figure 13.10 CART probability map for gold.

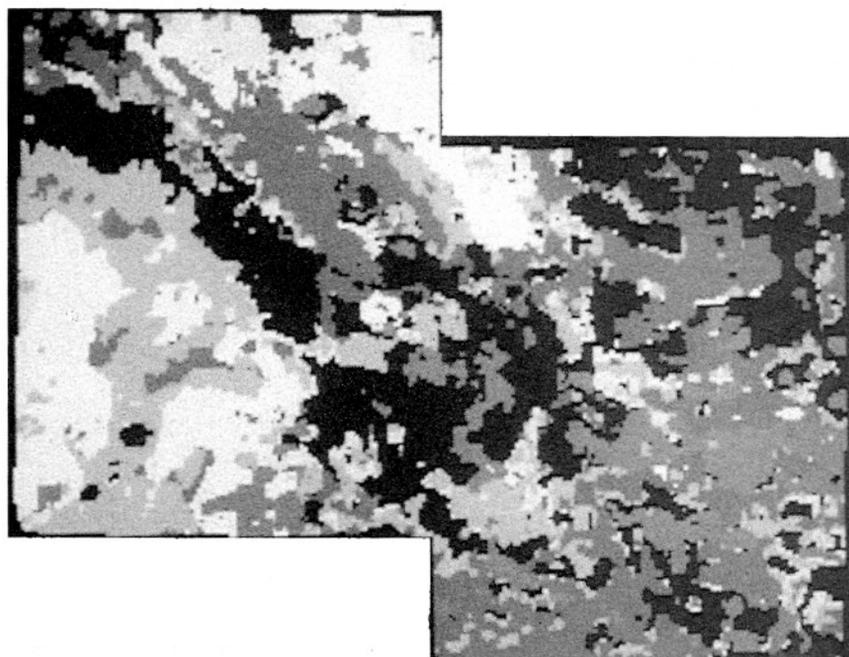


Figure 13.11 CART probability map for cassiterite.