

## 1. Introduction

Mineral resource and ore reserve estimation requires a critical decision regarding the geological domains that will be used for the grade modeling, as well as the type of boundaries between these domains. The statistical characteristics of one domain can have a very different impact than another domain in the final mineral inventory. The most common geostatistical techniques, such as kriging and sequential simulation, are based on strong assumptions of stationarity of the estimation domains. In particular, they are based in a second order stationary hypothesis, that is, the mean and covariance remain constant along the domain and they do not depend on the location of the support points but only in the distance between them. Therefore, these domains should be chosen as statistically homogeneous zones, which are geologically significant and coherent, but still holds enough data to allow reliable inference of first and second order statistics.

Wherever possible, geology should be used to define appropriate estimation domains. Some deposits will have a very simple geometry and mineralisation pattern, making the choice of estimation domains straight forward from geological units. In other cases the mineralisation of interest is not entirely defined by a single geological unit or may transgress their boundaries. In these deposits, it is common to find an important structural or lithological influence on grade distribution and/or multiple events of mineralisation. In these cases, a combination of different geological attributes may reflect an estimation domain. When it is not possible to identify and isolate the geological controls of grade distribution, the estimation domains can be defined purely by grade boundaries; however, this method is considerably more dangerous in that under or overestimation of grade and tonnage can result.

Once certain geological or grade boundaries are selected to represent an estimation boundary, the choice should be validated statistically. Differences in the mean, variance and spatial variability between domains are common. A probability plot or a high coefficient of variation may indicate that the chosen domain still has mixed populations. Indicator variograms can be used to test for consistency in anisotropy and grade continuity of different grade ranges within a proposed domain.

Following the definition of the estimation domains, an analysis on how grades change across the boundaries between domains should be done, this aids in the validation of the proposed units and to determine the nature of their boundaries. Domain boundaries are often referred to as either 'hard' or 'soft'. Hard boundaries are found when an abrupt change in grade average or variability occurs at the contact between two domains, such as coal seams or sedimentary zinc deposits. In deposits where the disseminated mineralisation has a gradational nature, such as some porphyry Cu-Au deposits, and grades change transitionally across a boundary, the contact is referred to as a soft boundary (Figure 1.1). Other examples are nickel, chromium and platinum orthomagmatic deposits that often show an increasing grade

profiles towards the bottom of a layer due to magmatic segregation due to gravity, hypothermal deposit where successive injections of circulating fluids produce concentric fracturing and telescopic mineralisation.

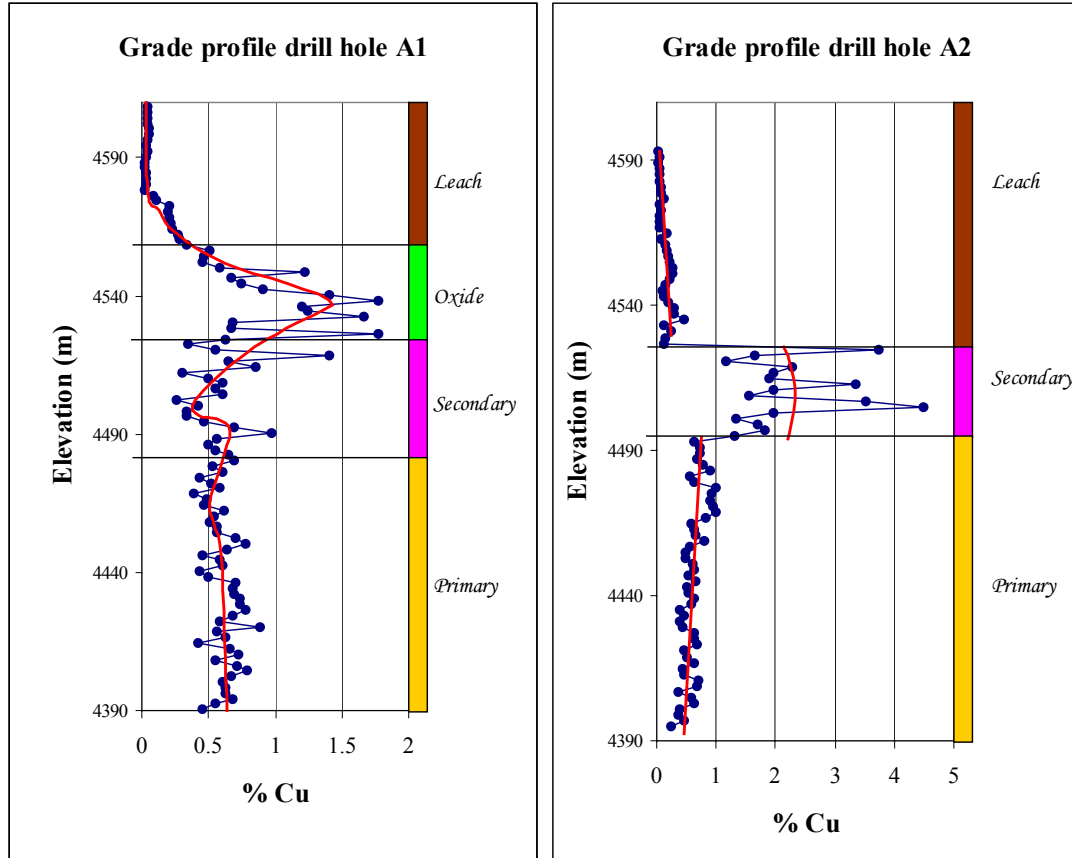


Figure 1.1: Grade profile of two drill holes in a porphyry copper deposit in north Chile, showing an example of soft boundaries (left) between mineralisation units as well as an example of hard boundaries (right). Notice how not only the mean (red line) change across the boundaries but also the variance does.

A good estimation of soft boundaries should insurance the correct reproduction of the correlation of the variable of interest across the boundary, reproduce non-stationary variations of the mean, variance and covariance in the zone of influence of each rock type into the other and reduce the uncertainty over all and specifically around the boundary. Boundaries are of special interest in the short mine planning and the above properties would benefit the design and operation stage in both underground and open pit deposits.

Although soft boundaries are found in several types of deposits due to the transitional nature of the geological mechanisms involve in the formation of a deposit which produce some degree of overlapping between geological grade

controls, conventional grade estimation usually treats the boundaries between geological units as hard boundaries. This is primarily due to the limitations of current estimation and simulation procedures.

The estimation of the hard boundary domain is straightforward, using only the samples within the domain, since no interpolation across domains is allowed. Soft boundaries allow grades from either side of a boundary to be used in the estimation of each domain, common practice is to share samples (from the conditioning data and/or from previously simulated nodes) within a given extent of influence of one domain over the other. The outside samples are treated equal to those within the domain, that is, the same mean, variance and covariance model from the samples within the domain are assumed. This generally has the effect of corrupting the representative statistics of the domain of interest. This corruption of the final grades, especially in the transition zones, often dissuades practitioners from handling these boundaries as soft boundaries and the result is that the boundaries are left as hard boundaries.

These work present two different methodologies for grade estimation in the presence of soft boundaries. The first is thought to be apply when the correlation of the variable of interest between two adjacent rock types remain constant within both units and is due to a underlying common factor. One example could be the supergene zone in a porphyry copper deposit (Figure 1.2); supergene enrichment of these systems began as the portions above the water table are oxidized and leach of copper by meteoric fluids, copper is transported in solution and precipitated below the ground water table by replacement of pre-existing iron sulfides. The mean and variance in the supergene zone is likely to be higher than the primary zone immediately below, but the spatial correlation structure will remain an underlying common factor.

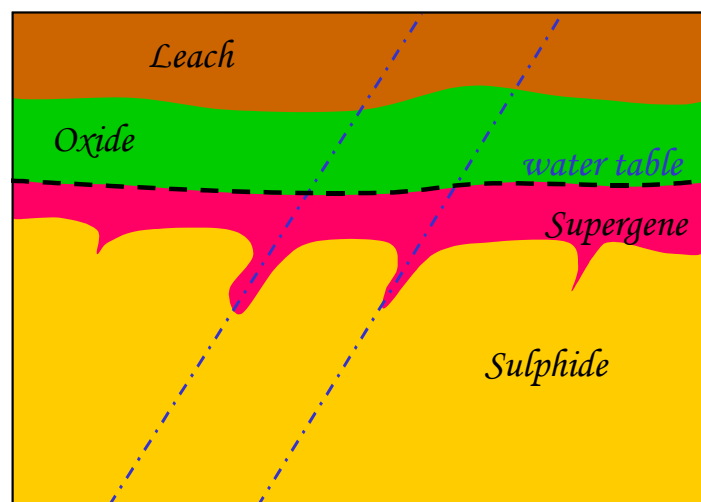


Figure1.2: Schematic cross-section illustrating porphyry copper mineralisation. The supergene zone develops below the water table as fluids percolated from the oxide zone and downward along faults.

In the case of a global soft boundary we propose to use a conventional linear model of coregionalization (LMC) to simulate grades using data from adjacent domains. Although the LMC is traditionally used to characterize the spatial variability of multiple properties or metal grades in one domain, we will show that it can be applied to model the spatial variability of one property across the boundary between multiple domains (Figure 1.3). A full model of coregionalization allows us to capture the spatial correlation of grades across the boundaries through a legitimate spatial model that can later be used to cokrig or cosimulate grades using data from adjacent domains. This approach guarantees the correct reproduction of representative statistics of each geological domain and improves the resource estimation by reduce the uncertainty in transitional zones around boundaries.

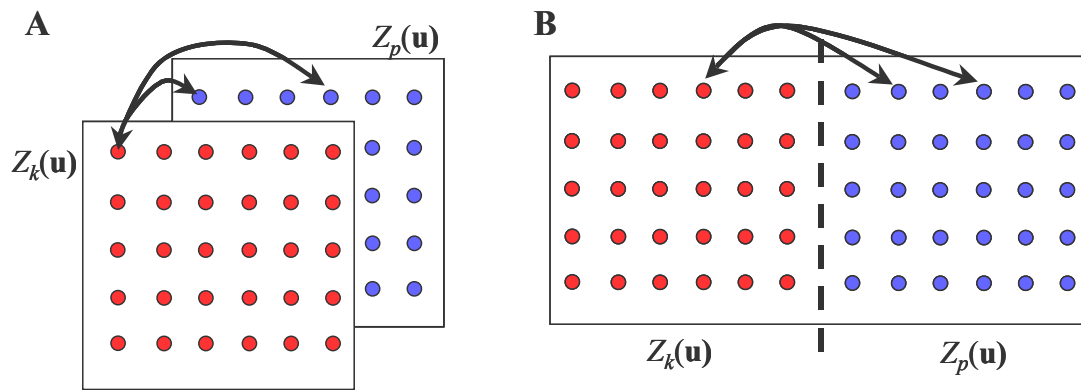


Figure 1.3: Conventional application of a local mode of coregionalization (A): model spatial variability of multiple properties in one domain. The proposed application (B): model the spatial variability of a single property across a boundary between two geological domains.

A global stationary soft boundary it is not very realistic, in general soft boundaries are characterized by a non-stationary behavior of the variable of interest in the proximities of the contact between units, that is, the mean, variance or covariance are no longer constant within a zone of influence of one rock type into the other and their values depends on the location relative to the boundary (Figure 1.4). Most common cases correspond to an increase or decrease in the mean or variance towards the boundary. Figure 1.5 show for an exhaustive training image the difference between a global stationary soft boundary and a local non-stationary one, both domains have a common underlying isotropic covariance model with a range of 100 meters. An example of the latest is the increased frequency of fractures towards a boundary between geological domains of structural nature (Figure 1.6). Faults or brittle zones are examples of this transition. The fractures may be mineralized so the average metal grade will increase closer to the boundary. Alternatively, fractures near the surface of the deposit may be leached by meteoric fluids, which may translate to a decrease in the average grade. The increase in the presence of factures will also lead to an increase or decrease in the variance closer to the boundary.

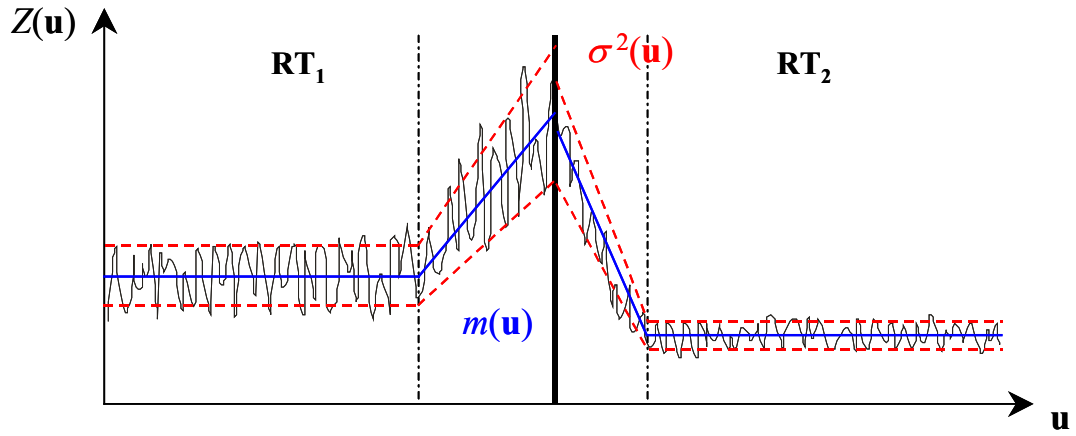


Figure 1.4: Local non-stationary soft boundary between rock type 1 and 2. The mean and variance are functions of the locations relative to the boundary within a zone of influence of one rock type into the other.

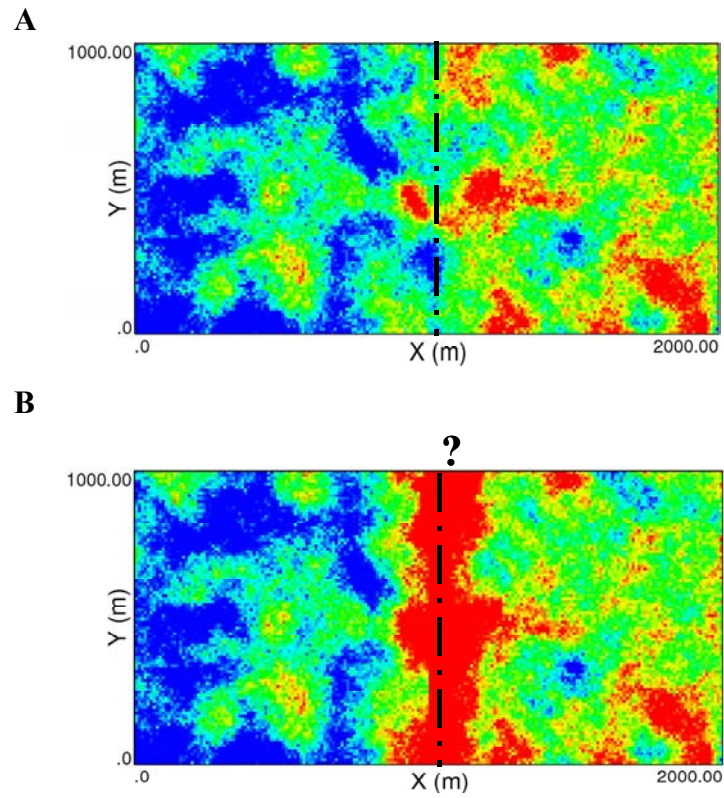


Figure 1.5: Difference between a global soft boundary (A) and a local non-stationary boundary (B). Both domains have a common underlying spatial isotropic structure of 100 meters range.

For local non-stationary soft boundaries we propose a new technique that accounts for stationary variables within domains and additional non-stationary components near boundaries for the mean, variance and covariance. The technique involves the following distinct phases: (i) identification of the rock type and boundary zones based on the geological modeling and timing of the different events, (ii) optimization of the non-stationary components of mean and variance in the boundary zone given the stationary statistical parameters of each domain, (iii) decomposition of data into stationary and non-stationary components of a linear model of coregionalization and optimization of the latest, and (iv) estimation of grades using a non-stationary form of simple cokriging. This technique provides an appealing alternative to capture grade distribution for deposits where complex contacts between different rock types exist.

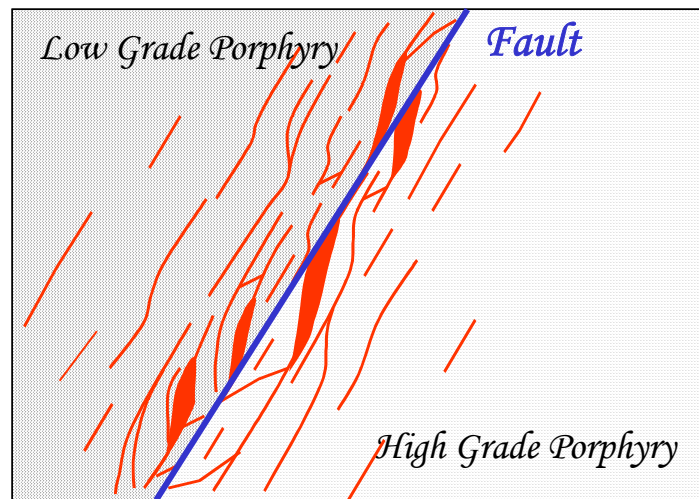


Figure 1.6: Schematic cross-section of a structural boundary between porphyries. The frequency of fractures increase toward the boundary defines by a fault. Besides the background-disseminated mineralisation, mineralisation of the fractures produces a local increase of the grade mean and variability at the boundary.

## Dissertation Outline

Chapter 2 discusses the application of a conventional linear model of coregionalization to estimate and simulate grades using data from multiple domains. Include the theoretical background as well as some small illustrative examples and a full application and comparison with current alternatives.

Chapter 3 corresponds to definitions, concepts and theoretical background of the proposed technique to estimate in the presence of local non-stationary soft boundaries.

Chapter 4 present the implementation of all the programs involves in the optimization of the statistical parameters; mean, variance and covariance, plus the non-stationary form of cokriging. The programming structure of these programs is illustrated with a simple 1-D example. A large 3-D example demonstrates the practical application of this technique in Chapter 5.

Finally, discussion on the results of the proposed methodologies, as well as future work and conclusions are presented in Chapter 6.

## 2. Global Linear Model of Coregionalization

As presented in the introduction, most geological mechanisms are transitional in nature; therefore some degree of overlap between geological units can be expected. This translates to correlation of the variable of interest between units, across the boundary.

In this chapter we will review how a linear model of coregionalization can be used to model grade distribution using data from different domains. We will illustrate the theoretical background with a few examples as well as some sensitivity analysis. This proposed methodology is applied to a synthetic deposit, and compared to the conventional approach of modeling using hard boundaries.

This methodology assumes that the variable is stationary in each domain. A **global** spatial relationship is used for the grades in different domains. Local non-stationary behavior at the boundary is discussed in the next chapter.

### Theoretical Background

The linear model of coregionalization (LMC) provides a method to model the cross covariance of two or more variables. The LMC model is legitimate, that is, that the variance of any possible linear combination of these variables is always positive. Given a set of  $K$  second order stationary random variables,  $\{Z_k(\mathbf{u}), k=1, \dots, K\}$ , the LMC provides a means to model the cross covariance functions,  $Cov_{Z_k Z_p}(\mathbf{h})$ ,  $k=1, \dots, K, p=1, \dots, P$ .

Usually  $Z_k(\mathbf{u})$  and  $Z_p(\mathbf{u})$  represent different properties measured at the same location, for example, gold and copper grades. We consider that each random variable  $Z_k(\mathbf{u})$ ,  $k=1, \dots, K$  corresponds to the outcomes of the variable of interest for each one of the  $K$  geological domains or rock types. The model could also be used for multiple grades within multiple rock types.

A linear model of coregionalization assumes that each variable  $Z_k(\mathbf{u})$  is a linear combination of  $n$  second order stationary independent random variables  $Y_i(\mathbf{u})$  with mean  $m_i$  and variance  $\sigma_i^2$ , with  $i=1, \dots, n$ . These  $n$  random variables are independent, that is, their cross covariances are zero:  $Cov_{Y_i Y_j}(\mathbf{h}) = 0, \forall \mathbf{h}$  and  $i \neq j$ . Each  $Z_k(\mathbf{u})$  variable is assumed to be a linear sum of the independent factors:

$$Z_k(\mathbf{u}) = \sum_{i=1}^n a_{ki} Y_i(\mathbf{u}) \quad k=1, \dots, K \quad (2.1)$$



The coefficients  $a_{ki}$  can be positive, negative or zero.

The mean of each  $Z_k(\mathbf{u})$  variable is:

$$\begin{aligned} m_k(\mathbf{u}) &= E\{Z_k(\mathbf{u})\} = E\left\{\sum_{i=1}^n a_{ki} Y_i(\mathbf{u})\right\} \\ &= \sum_{i=1}^n a_{ki} m_i \end{aligned}$$

The variance can be derived as follows,

$$\begin{aligned} \sigma_k^2(\mathbf{u}) &= E\left\{\left(\sum_{i=1}^n a_{ki}^2 Y_i(\mathbf{u})\right)^2\right\} - \left(\sum_{i=1}^n a_{ki} E\{Y_i(\mathbf{u})\}\right)^2 \\ &= \left(\sum_{i=1}^n a_{ki}^2 E\{Y_i(\mathbf{u})^2\} + \sum_{i=1}^n \sum_{j \neq i}^n a_{ki} a_{kj} E\{Y_i(\mathbf{u}) Y_j(\mathbf{u})\}\right) \\ &\quad - \left(\sum_{i=1}^n a_{ki}^2 (E\{Y_i(\mathbf{u})\})^2 + \sum_{i=1}^n \sum_{j \neq i}^n a_{ki} a_{kj} E\{Y_i(\mathbf{u})\} \cdot E\{Y_j(\mathbf{u})\}\right) \end{aligned}$$

But  $E\{Y_i(\mathbf{u}) \cdot Y_j(\mathbf{u})\} = E\{Y_i(\mathbf{u})\} \cdot E\{Y_j(\mathbf{u})\}$  since  $Y_i(\mathbf{u})$  is independent of  $Y_j(\mathbf{u})$ . Then,

$$\begin{aligned} \sigma_k^2(\mathbf{u}) &= \sum_{i=1}^n a_{ki}^2 E\{Y_i(\mathbf{u})^2\} + \sum_{i=1}^n \sum_{j \neq i}^n a_{ki} a_{kj} E\{Y_i(\mathbf{u})\} \cdot E\{Y_j(\mathbf{u})\} - \sum_{i=1}^n a_{ki}^2 (E\{Y_i(\mathbf{u})\})^2 \\ &\quad - \sum_{i=1}^n \sum_{j \neq i}^n a_{ki} a_{kj} E\{Y_i(\mathbf{u})\} \cdot E\{Y_j(\mathbf{u})\} \\ &= \sum_{i=1}^n a_{ki}^2 E\{Y_i(\mathbf{u})^2\} - \sum_{i=1}^n a_{ki}^2 (E\{Y_i(\mathbf{u})\})^2 \\ \sigma_k^2(\mathbf{u}) &= \sum_{i=1}^n a_{ki}^2 \sigma_i^2 \end{aligned}$$

The covariance of  $Z_k(\mathbf{u})$  at a vectorial distance  $\mathbf{h}$ , can also be calculated as an expression of the coefficients  $a_{ki}$  and the covariances of  $Y_i(\mathbf{u})$  for  $i=1, \dots, n$ :

$$Cov_{Z_k}(\mathbf{h}) = \sum_{i=1}^n a_{ki}^2 Cov_{Y_i}(\mathbf{h})$$

Similar to the derivation of the variance, the cross-covariance of  $Z_k(\mathbf{u})$  and  $Z_p(\mathbf{u})$ ,  $\forall k \neq p$ , with  $k, p = 1, \dots, K$  can also be derived as a linear combination of  $Y_i(\mathbf{u})$  covariances and coefficients  $a_{ki}$  for  $Z_k(\mathbf{u})$ , and  $a_{pj}$  for  $Z_p(\mathbf{u})$ ,  $i, j = 1, \dots, n$ ,

$$\begin{aligned} Cov_{Z_k Z_p}(\mathbf{h}) &= E \left\{ \sum_{i=1}^n a_{ki} Y_i(\mathbf{u}) \sum_{j=1}^n a_{pj} Y_j(\mathbf{u} + \mathbf{h}) \right\} - E \left\{ \sum_{i=1}^n a_{ki} Y_i(\mathbf{u}) \right\} \cdot E \left\{ \sum_{j=1}^n a_{pj} Y_j(\mathbf{u} + \mathbf{h}) \right\} \\ &= \left( \sum_{i=1}^n a_{ki} a_{pi} E \{ Y_i(\mathbf{u}) \cdot Y_i(\mathbf{u} + \mathbf{h}) \} + \sum_{i=1}^n \sum_{j \neq i}^n a_{ki} a_{pj} E \{ Y_i(\mathbf{u}) \cdot Y_j(\mathbf{u} + \mathbf{h}) \} \right) \\ &\quad - \left( \sum_{i=1}^n a_{ki} a_{pi} E \{ Y_i(\mathbf{u}) \} \cdot E \{ Y_i(\mathbf{u} + \mathbf{h}) \} + \sum_{i=1}^n \sum_{j \neq i}^n a_{ki} a_{pj} E \{ Y_i(\mathbf{u}) \} \cdot E \{ Y_j(\mathbf{u} + \mathbf{h}) \} \right) \end{aligned}$$

But since  $Y_i(\mathbf{u})$  is independent of  $Y_j(\mathbf{u})$  there is no cross spatial correlation between  $Y_i(\mathbf{u})$  and  $Y_j(\mathbf{u} + \mathbf{h})$ , that is,  $E \{ Y_i(\mathbf{u}) \cdot Y_j(\mathbf{u} + \mathbf{h}) \} = E \{ Y_i(\mathbf{u}) \} \cdot E \{ Y_j(\mathbf{u} + \mathbf{h}) \}$ ,  $\forall i \neq j$ . Then,

$$\begin{aligned} Cov_{Z_k Z_p}(\mathbf{h}) &= \sum_{i=1}^n a_{ki} a_{pi} E \{ Y_i(\mathbf{u}) \cdot Y_i(\mathbf{u} + \mathbf{h}) \} - \sum_{i=1}^n a_{ki} a_{pi} E \{ Y_i(\mathbf{u}) \} \cdot E \{ Y_i(\mathbf{u} + \mathbf{h}) \} \\ Cov_{Z_k Z_p}(\mathbf{h}) &= \sum_{i=1}^n a_{ki} a_{pi} Cov_{Y_i}(\mathbf{h}) \end{aligned}$$

If the random variables  $Y_i(\mathbf{u})$ ,  $i = 1, \dots, n$  are grouped by those who have the same direct covariance, that is,  $C_{Y_i}(\mathbf{h}) = C_{Y_j}(\mathbf{h})$ , while the cross terms remain zero due to orthogonal assumption for the random function  $Y(\mathbf{u})$ , Equation 2.1 can be rewritten as,

$$Z_k(\mathbf{u}) = \sum_{i=1}^{n_l} \sum_{l=0}^L a_i^l Y_i^l(\mathbf{u}) \quad k = 1, \dots, K$$

where  $L+1$  is the number of groups with different direct covariances and  $n_l$  is the number of random variables with the same covariance in group  $l$ .

The cross covariance of  $Z_k(\mathbf{u})$  and  $Z_p(\mathbf{u})$ , can also be rewritten as,

$$Cov_{Z_k Z_p}(\mathbf{h}) = \sum_{i=1}^{n_l} \sum_{l=0}^L a_{ki}^l a_{pi}^l C^l(\mathbf{h})$$

where the term,  $\sum_{i=1}^{n_l} a_{ki}^l a_{pi}^l = b_{kp}^l$  correspond to the sill contribution of the  $l^{th}$  covariance structure,  $C^l(\mathbf{h})$ .

In summary, a model will be a linear model of coregionalization if all direct and cross-covariances are derived as the linear combinations of  $l$  direct covariances  $C_{Y_i}(\mathbf{h})$ , that is,

$$Cov_{Z_k Z_p}(\mathbf{h}) = \sum_{l=0}^L b_{kp}^l C^l(\mathbf{h}) \quad \text{with } b_{kp}^l = b_{pk}^l, l = 1, \dots, L$$

The coefficient matrix  $\begin{bmatrix} b_{kp}^l \end{bmatrix}$  must be positive definite (Journel and Huijbregts, 1978). Since this matrix is symmetric, it will be positive definite, if all determinant and sub-determinants, as well as the diagonal terms are positive or zero.

For a two variable example and replacing the covariances  $Cov_{Z_k Z_p}(\mathbf{h})$  and  $Cov_{Y_i}(\mathbf{h})$  by the corresponding semi-variograms, the linear model of coregionalization takes the following form:

$$\begin{aligned} \gamma_{Z_k Z_k}(\mathbf{h}) &= \sum_{l=0}^L b_{kk}^l \cdot \Gamma^l(\mathbf{h}) \\ \gamma_{Z_k Z_p}(\mathbf{h}) &= \sum_{l=0}^L b_{kp}^l \cdot \Gamma^l(\mathbf{h}) \\ \gamma_{Z_p Z_p}(\mathbf{h}) &= \sum_{l=0}^L b_{pp}^l \cdot \Gamma^l(\mathbf{h}) \end{aligned}$$

where  $\Gamma^l(\mathbf{h})$ ,  $l=1, \dots, L$  are the nested structures made up of legitimate variogram models and the coefficients  $b$  follow the following constraints to ensure a positive definite model:

$$\begin{aligned} b_{kk}^l &> 0 \\ b_{pp}^l &> 0 \quad l = 0, \dots, L \\ b_{kk}^l \cdot b_{pp}^l &\geq b_{kp}^l \cdot b_{kp}^l \end{aligned}$$

A linear model of coregionalization, with  $Z_k(\mathbf{u})$  and  $Z_p(\mathbf{u})$  representing the distribution of the variable of interest in rock type  $k$  and rock type  $p$ , respectively, is a legitimate spatial correlation model that yield the correct statistics at unknown locations near the boundary where samples from both domains are used for the estimation or simulation. The calculated LMC spatial model can be used in cokriging or cosimulation to model locations near geological boundaries using samples from adjacent domains. This is a more consistent alternative to the estimation of domains with soft boundaries.

## Illustration of Theory

To illustrate how a linear model of coregionalization can be use to characterize the spatial variability of multiple rock types, consider a 2D example with two domains. The corresponding random variables  $Z_I(\mathbf{u})$  and  $Z_2(\mathbf{u})$  were constructed as a linear combination of three underlying non-standard normal random variables:

$$\begin{aligned} Z_1(\mathbf{u}) &= \sqrt{0.5} \cdot Y_1(\mathbf{u}) + \sqrt{0.5} \cdot Y_2(\mathbf{u}) + 0.0 \cdot Y_3(\mathbf{u}) \\ Z_2(\mathbf{u}) &= \sqrt{0.5} \cdot Y_1(\mathbf{u}) + 0.0 \cdot Y_2(\mathbf{u}) + \sqrt{0.5} \cdot Y_3(\mathbf{u}) \end{aligned}$$

where,

$$\begin{aligned} Y_1(\mathbf{u}) &\sim N(0.5, 0.5) \text{ with } \gamma_{Y_1}(\mathbf{h}) = 0.05 + 0.45 \cdot Sph_{\left(\begin{smallmatrix} h \text{ max}=200 \\ h \text{ min}=200 \end{smallmatrix}\right)}(\mathbf{h}) \\ Y_2(\mathbf{u}) &\sim N(2.0, 1.0) \text{ with } \gamma_{Y_2}(\mathbf{h}) = 0.1 + 0.9 \cdot Sph_{\left(\begin{smallmatrix} h \text{ max}=50 \\ h \text{ min}=300 \end{smallmatrix}\right)}(\mathbf{h}) \\ Y_3(\mathbf{u}) &\sim N(1.0, 0.5) \text{ with } \gamma_{Y_3}(\mathbf{h}) = 0.05 + 0.45 \cdot Exp_{\left(\begin{smallmatrix} h \text{ max}=400 \\ h \text{ min}=100 \end{smallmatrix}\right)}(\mathbf{h}) \end{aligned}$$

The random variables  $Y_i(\mathbf{u})$  were obtained by unconditional Gaussian simulation for a grid of 1000 by 2000 meters. The simulated values were transformed to a non-standard normal distribution with the corresponding mean and variance. 10 realizations were simulated.

The covariance of  $Z_I(\mathbf{u})$  and  $Z_2(\mathbf{u})$ , obtained from the simulated values, were checked against the analytically derived models:

$$\begin{aligned} Cov_{Z_1}(\mathbf{h}) &= 0.5 \cdot Cov_{Y_1}(\mathbf{h}) + 0.5 \cdot Cov_{Y_2}(\mathbf{h}) \\ &= 0.75 - 0.075 - 0.225 \cdot Sph_{\left(\begin{smallmatrix} h \text{ max}=200 \\ h \text{ min}=200 \end{smallmatrix}\right)}(\mathbf{h}) - 0.45 \cdot Sph_{\left(\begin{smallmatrix} h \text{ max}=50 \\ h \text{ min}=300 \end{smallmatrix}\right)}(\mathbf{h}) \\ \\ Cov_{Z_2}(\mathbf{h}) &= 0.5 \cdot Cov_{Y_1}(\mathbf{h}) + 0.5 \cdot Cov_{Y_3}(\mathbf{h}) \\ &= 0.5 - 0.05 - 0.225 \cdot Sph_{\left(\begin{smallmatrix} h \text{ max}=200 \\ h \text{ min}=200 \end{smallmatrix}\right)}(\mathbf{h}) - 0.225 \cdot Exp_{\left(\begin{smallmatrix} h \text{ max}=400 \\ h \text{ min}=100 \end{smallmatrix}\right)}(\mathbf{h}) \end{aligned}$$

The covariance over all realizations matches the analytical models, as it must.

The cross covariance between  $Z_I(\mathbf{u})$  and  $Z_2(\mathbf{u})$  was calculated and checked against its analytical model,

$$\begin{aligned}
Cov_{Z_1, Z_2}(\mathbf{h}) &= 0.5Cov_{Y_1}(\mathbf{h}) \\
&= 0.25 - 0.025 - 0.225 \cdot Sph_{\left( \begin{smallmatrix} h_{\max}=200 \\ h_{\min}=200 \end{smallmatrix} \right)}(\mathbf{h})
\end{aligned} \tag{2.2}$$

three different spatial arrangement of  $Z_1(\mathbf{u})$  and  $Z_2(\mathbf{u})$  were considered: (1) collocated (just as a check), (2) the two domains adjacent to each other (Figure 2.1A), and (3) the two domains merged (Figure 2.1B) using a categorical binary model obtained via a Boolean simulation program, `ellipsim`, that generate a 2D map of ellipsoids of variable size and anisotropies for a given target proportion (Deutsch and Journel, 1998).

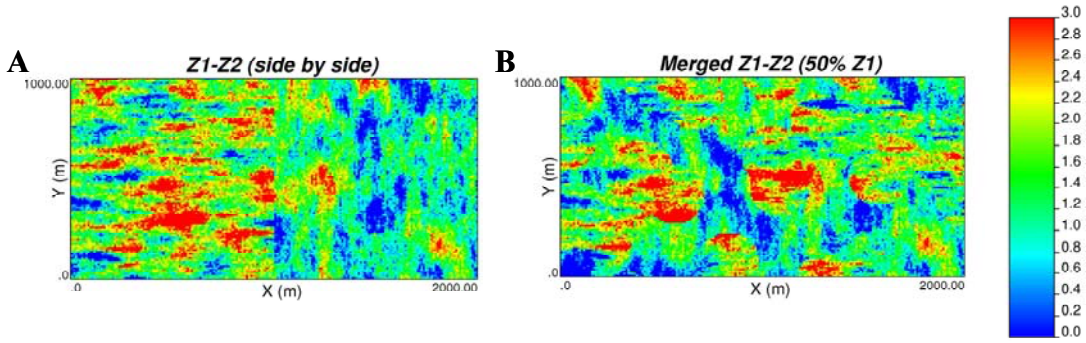


Figure 2.1: Example of two domains: (A) the domains are side by side,  $Z_1(\mathbf{u})$  correspond to the right-half and  $Z_2(\mathbf{u})$  to the left-half, (B) the two domains are merged using a categorical binary model.  $Z_1(\mathbf{u})$  shows anisotropy along the X-axis, and have on average lower values than  $Z_2(\mathbf{u})$ , which shows anisotropy along the Y-axis.

As a check of our derivations we compare the cross-covariance between  $Z_1(\mathbf{u})$  and  $Z_2(\mathbf{u})$  when both variables are collocated with the analytical derived model, as shown in Figure 2.2 the average variogram over all realizations is very close to the analytical model (Equation 2.2). The ergodic fluctuations associated with the different realization are very low.

In the case where the two domains are side by side, the covariances of simulated values correspond to the analytical model fairly well (Figure 2.3A), although configurations where the boundary is parallel to the major anisotropy of one of the domains ( $Z_2(\mathbf{u})$  in this case), showed a systematically lower covariance at shorter lag distances than the analytical model (Equation 2.2), and the dispersion of the ergodic fluctuations is greater at lag distances near zero. This latter observation means that inference of the nugget effect of the cross covariance is more uncertain in geometrical configurations similar to this one.

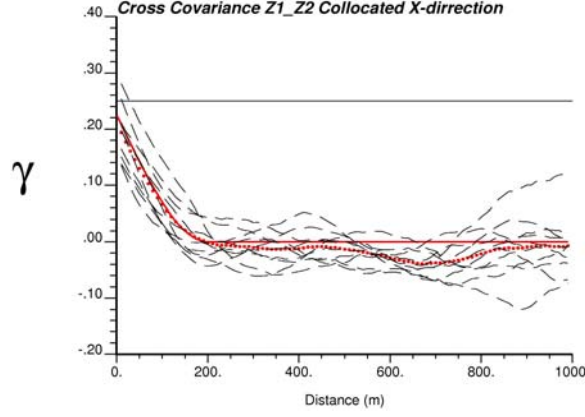


Figure 2.2: Cross-covariance reproduction of the simulated random variables  $Z_1(\mathbf{u})$  and  $Z_2(\mathbf{u})$ , assuming both variables are collocated. The dots are the average taken over all realizations; individual realizations are in dashed lines; and the thin red solid line corresponds to the analytical model. The analytical model is very close to the average over all realizations, which makes it difficult to differentiate the dots from the solid line.

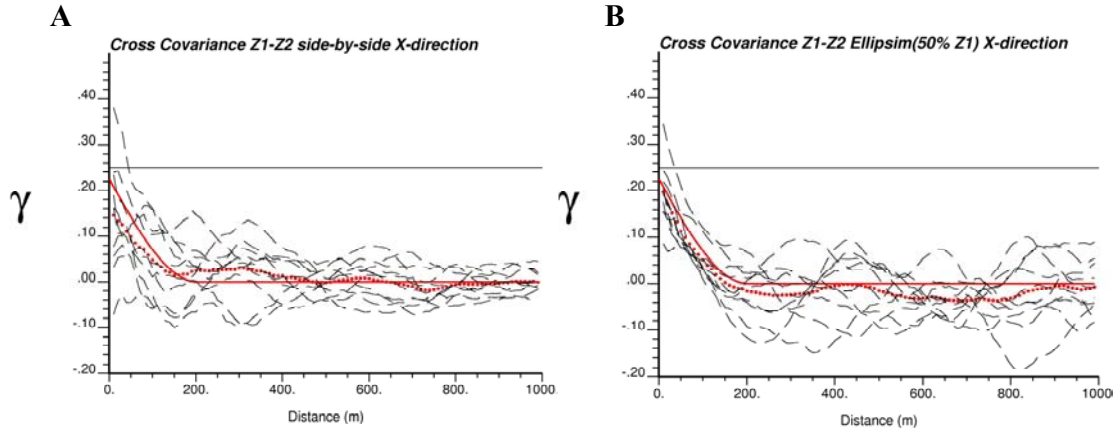


Figure 2.3: (A) Cross-covariance between  $Z_1(\mathbf{u})$  and  $Z_2(\mathbf{u})$  combined side by side. (B) Cross-covariance between  $Z_1(\mathbf{u})$  and  $Z_2(\mathbf{u})$  combined using `ellipsim` categorical model as a boundary model with a target proportion of  $Z_1$  of 50%. The dots are the average taken over all realizations; individual realizations are in dashed lines; and the thin solid line corresponds to the analytical model.

For the second scheme, using a circular shape with radius of 150 meters and three target proportions of 25, 50 and 75%, the cross covariance between the experimental points derived from the average over all realizations compares well with the analytical model (Figure 2.3B) (Equation 2.2). The fluctuations at short lag distances are small. This confirms our expectation that when more contact surfaces between domains are available and are more irregularly oriented; the determination of the nugget effect should have less uncertainty, compared to the case where a single contact surface exists between the two domains. A completely straight or planar boundary gives the least possible surface area to the boundary. This leads to the smallest possible transition zone between rock types and the fewest possible pairs for

variogram calculation. This was also confirmed by a poorer reproduction at shorter lag distances, with lower covariances than the analytical model, when the target proportion of the domain  $Z_I$  was lower than 10%. In addition, when the proportion of one domain decreases, the dispersion of the ergodic fluctuations increases (Figure 2.4).

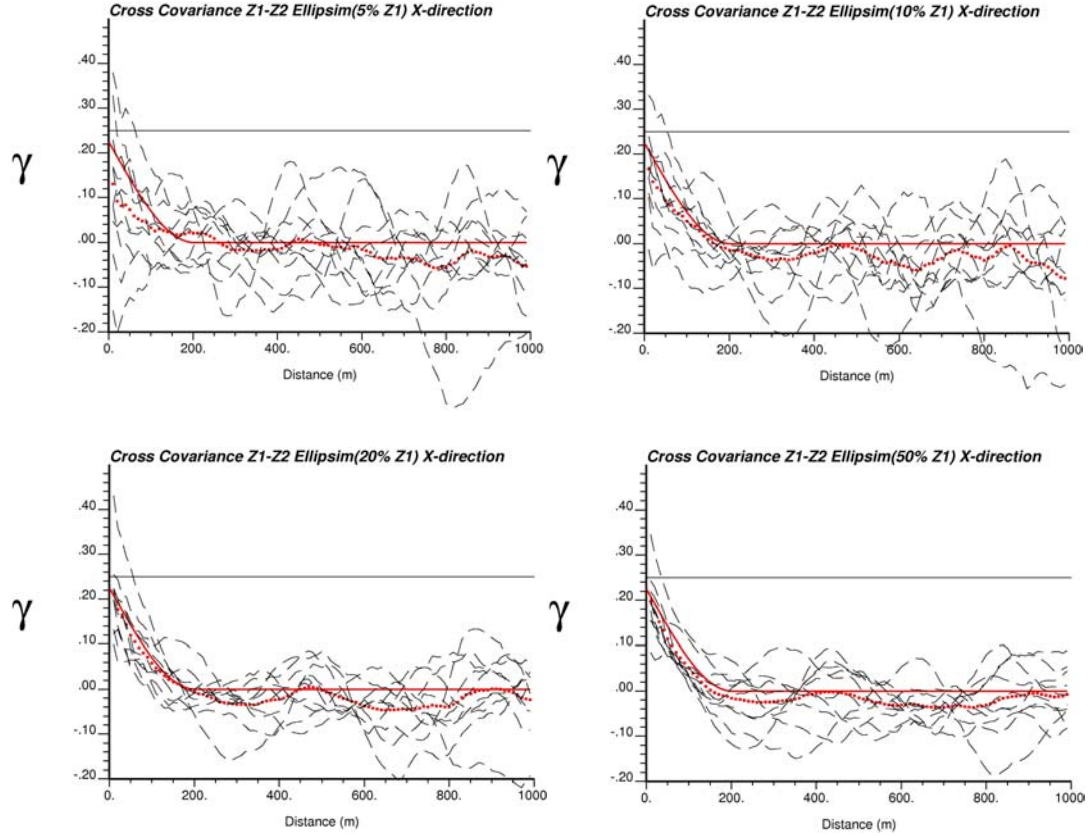


Figure 2.4: Cross-covariance between  $Z_I(\mathbf{u})$  and  $Z_2(\mathbf{u})$  combined using `ellipsim` categorical model as a boundary model, for target proportion of  $Z_I$  of 5, 10, 20 and 50%. Note that as the target proportion of one of the domains decreases the experimental derived from the average over all showed a systematically lower covariance at shorter lag distances than the dashed analytical model. The dots are the average taken over all realizations; individual realizations are in dashed lines; and the thin solid line corresponds to the analytical model.

To determine the influence of the size of the domain, six different radii were checked (25, 50, 100, 150, 200 and 300 meters) using a target proportion of 50%. For every radii, the average cross-covariance between  $Z_I(\mathbf{u})$  and  $Z_2(\mathbf{u})$  is very close to the analytical model and no significant differences were found (Figure 2.5).

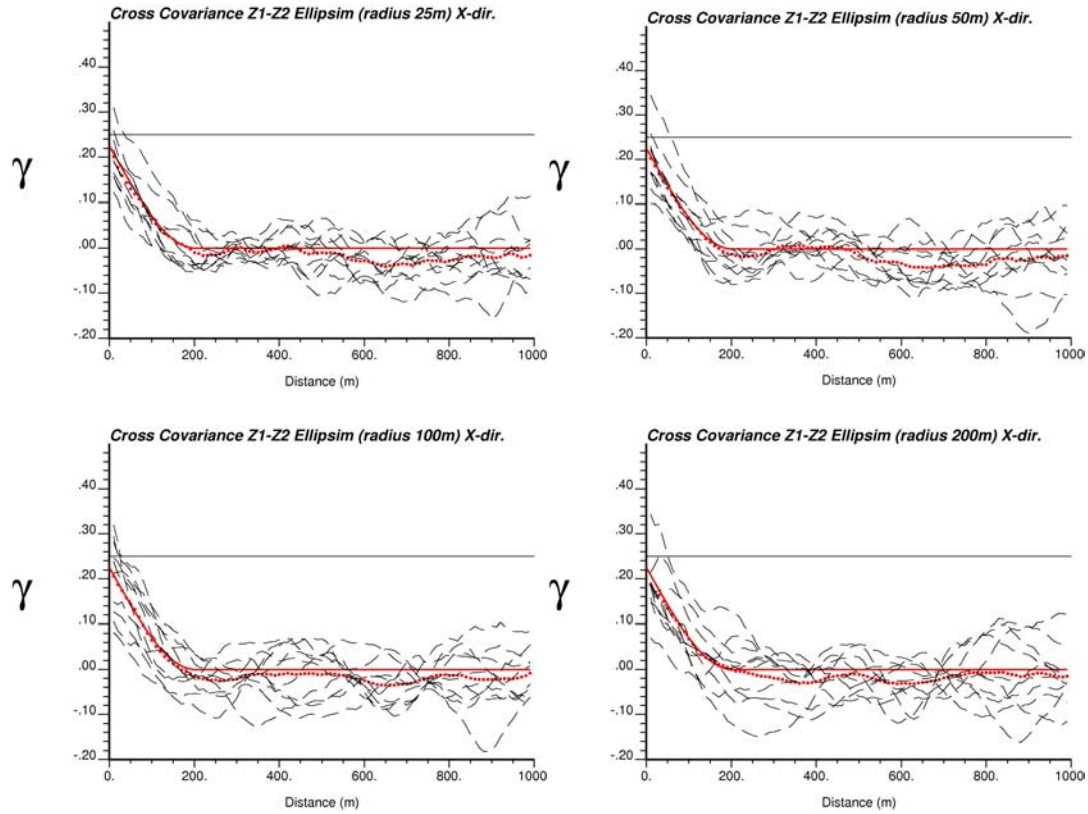


Figure 2.5: Cross-covariance between  $Z_1(\mathbf{u})$  and  $Z_2(\mathbf{u})$  combined using `ellipsim` categorical model as a boundary model, for different radii: 25, 50, 100 and 200 m. The dots are the average taken over all realizations; individual realizations are in dashed lines; and the thin solid line corresponds to the analytical model.

Using the same synthetic examples, the impact of different drill hole data spacing was examined. Overall, the reproduction of the cross covariance analytical model is as good as when all simulated values were used, although a wider range of fluctuation between realizations is observed. If the grid or data spacing is larger than the range of the cross-covariance between data from the two domains, the calculation of a cross-covariance will be meaningless. Also, the grid spacing will be critical if there are insufficient pairs to infer a model.

## 2D Example

For this example a real categorical geological model (Figure 2.6) was used to forecast the results for geometries of a real deposit. The grades, assumed to be percentage of copper, within the five rock types,  $Z_1(\mathbf{u})$  to  $Z_5(\mathbf{u})$  were constructed as a linear combination of four underlying non-standard normal random variables,



$$Y_1(\mathbf{u}) \sim N(0.01, 0.5) \quad \text{with } \gamma_{Y_1}(\mathbf{h}) = 0.05 + 0.45 \cdot Sph_{\left(\begin{smallmatrix} h \max=100 \\ h \min=100 \end{smallmatrix}\right)}(\mathbf{h})$$

$$Y_2(\mathbf{u}) \sim N(2.0, 1.0) \quad \text{with } \gamma_{Y_2}(\mathbf{h}) = 0.1 + 0.9 \cdot Sph_{\left(\begin{smallmatrix} h \max=400 \\ h \min=50 \end{smallmatrix}\right)}(\mathbf{h})$$

$$Y_3(\mathbf{u}) \sim N(0.2, 0.75) \quad \text{with } \gamma_{Y_3}(\mathbf{h}) = 0.05 + 0.70 \cdot Exp_{\left(\begin{smallmatrix} h \max=50 \\ h \min=300 \end{smallmatrix}\right)}(\mathbf{h})$$

$$Y_4(\mathbf{u}) \sim N(0.75, 1.5) \quad \text{with } \gamma_{Y_4}(\mathbf{h}) = 0.3 + 1.2 \cdot Exp_{\left(\begin{smallmatrix} h \max=400 \\ h \min=250 \end{smallmatrix}\right)}(\mathbf{h})$$

where  $Y_2(\mathbf{u})$  and  $Y_4(\mathbf{u})$  have a  $55^\circ$  anisotropy. The coefficients that multiplied the underlying variables in the summation that originates  $Z_1(\mathbf{u})$  to  $Z_5(\mathbf{u})$  are:

	$Z_1$	$Z_2$	$Z_3$	$Z_4$	$Z_5$
$Y_1$	$\sqrt{0.5}$	0.0	0.0	$\sqrt{0.2}$	$\sqrt{0.7}$
$Y_2$	0.0	$\sqrt{0.7}$	$\sqrt{0.2}$	0.0	0.0
$Y_3$	$\sqrt{0.5}$	$\sqrt{0.3}$	0.0	0.0	$\sqrt{0.25}$
$Y_4$	0.0	0.0	$\sqrt{0.8}$	$\sqrt{0.8}$	$\sqrt{0.05}$

The variables  $Z_1(\mathbf{u})$  up to  $Z_5(\mathbf{u})$  were merged together using the categorical rock type model (Figure 2.7)

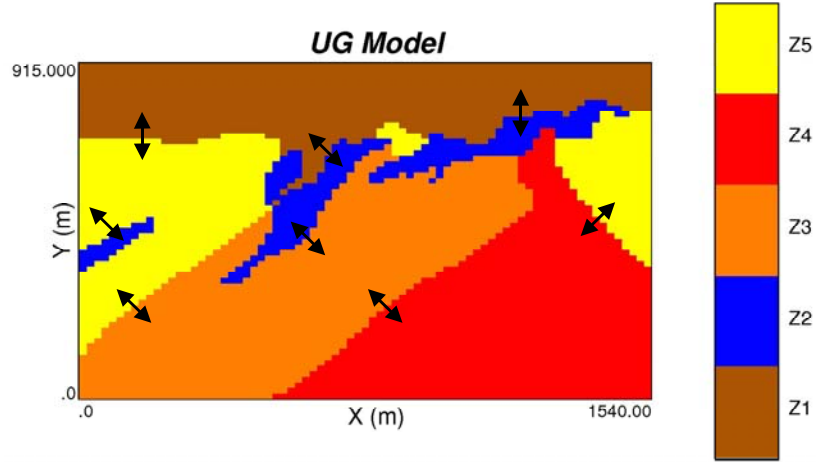


Figure 2.6: Categorical rock type model. The arrows indicate the directions in which the cross-covariance between domains was calculated.

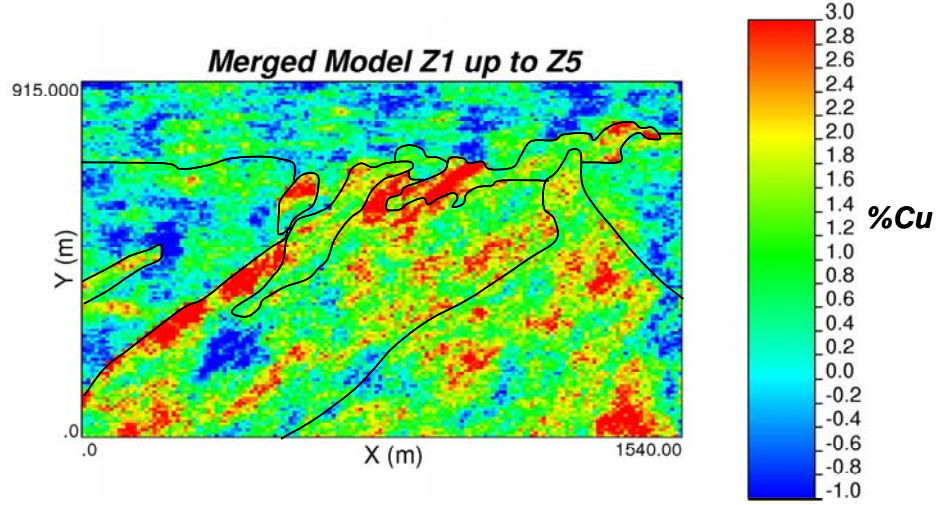


Figure 2.7: A realization of the merged grades  $Z_1(\mathbf{u})$  up to  $Z_5(\mathbf{u})$  using the categorical rock type model of Figure 2.6.

The cross covariance was calculated for each pair of  $Z_i(\mathbf{u})$  and  $Z_j(\mathbf{u})$ ,  $i \neq j$ , in the directions sketch in Figure 2.7 and compared with the corresponding analytical modes:

$$\begin{aligned}
 Cov_{Z_1, Z_2}(\mathbf{h}) &= 0.29 - 0.019 - 0.271 \cdot Exp_{\left( \begin{smallmatrix} h \text{ max}=50 \\ h \text{ min}=300 \end{smallmatrix} \right)}(\mathbf{h}) \\
 Cov_{Z_1, Z_5}(\mathbf{h}) &= 0.561 - 0.047 - 0.266 \cdot Sph_{\left( \begin{smallmatrix} h \text{ max}=100 \\ h \text{ min}=100 \end{smallmatrix} \right)}(\mathbf{h}) - 0.248 \cdot Exp_{\left( \begin{smallmatrix} h \text{ max}=50 \\ h \text{ min}=300 \end{smallmatrix} \right)}(\mathbf{h}) \\
 Cov_{Z_2, Z_3}(\mathbf{h}) &= 0.374 - 0.037 - 0.337 \cdot Sph_{\left( \begin{smallmatrix} h \text{ max}=400 \\ h \text{ min}=50 \end{smallmatrix} \right)}(\mathbf{h}) \\
 Cov_{Z_2, Z_5}(\mathbf{h}) &= 0.206 - 0.014 - 0.192 \cdot Exp_{\left( \begin{smallmatrix} h \text{ max}=50 \\ h \text{ min}=300 \end{smallmatrix} \right)}(\mathbf{h}) \\
 Cov_{Z_3, Z_4}(\mathbf{h}) &= 1.2 - 0.24 - 0.96 \cdot Exp_{\left( \begin{smallmatrix} h \text{ max}=400 \\ h \text{ min}=250 \end{smallmatrix} \right)}(\mathbf{h}) \\
 Cov_{Z_3, Z_5}(\mathbf{h}) &= 0.3 - 0.06 - 0.24 \cdot Exp_{\left( \begin{smallmatrix} h \text{ max}=400 \\ h \text{ min}=250 \end{smallmatrix} \right)}(\mathbf{h}) \\
 Cov_{Z_4, Z_5}(\mathbf{h}) &= 0.487 - 0.079 - 0.168 \cdot Sph_{\left( \begin{smallmatrix} h \text{ max}=100 \\ h \text{ min}=100 \end{smallmatrix} \right)}(\mathbf{h}) - 0.24 \cdot Exp_{\left( \begin{smallmatrix} h \text{ max}=400 \\ h \text{ min}=250 \end{smallmatrix} \right)}(\mathbf{h})
 \end{aligned}$$

The experimental cross covariance obtained from the average overall realizations compares very well with the analytical models (Figure 2.8), except for the pairs  $Z_3(\mathbf{u})$  and  $Z_4(\mathbf{u})$  and  $Z_4(\mathbf{u})$  and  $Z_5(\mathbf{u})$ , that have a side by side arrangement that shows lower covariances at shorter lag distances.

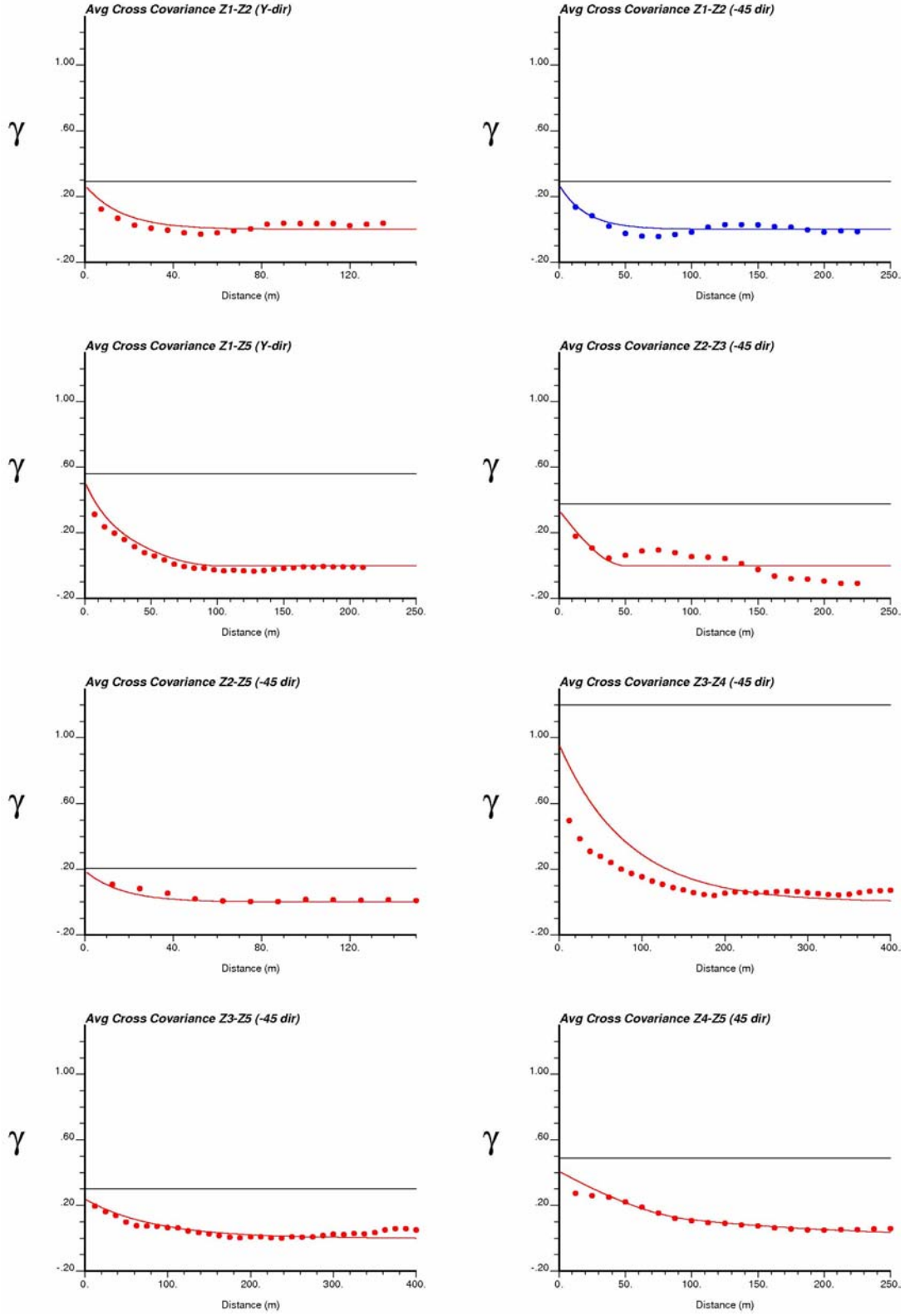


Figure 2.8: Average cross-covariance reproduction of the simulated pairs  $Z_i(\mathbf{u})$  and  $Z_j(\mathbf{u})$  for  $i \neq j$ , combined by the categorical rock type model. The experimental points correspond to the average over ten realizations.

## Application

A synthetic example was created in order to use a full LMC cosimulation and compare it with the results obtained from simulating two adjacent rock types independently. The LMC model was obtained by calculating the cross variograms between values of the different domains and the direct variograms within each rock type.

Using a similar methodology, the linear combination of three underlying standard normal random variables were used to populate a synthetic geological model; this will be considered as the ‘true’ image (Figure 2.9) for comparison.

$$\begin{aligned} Z_1(\mathbf{u}) &= \sqrt{0.5} \cdot Y_1(\mathbf{u}) + \sqrt{0.5} \cdot Y_2(\mathbf{u}) + 0.0 \cdot Y_3(\mathbf{u}) \\ Z_2(\mathbf{u}) &= \sqrt{0.5} \cdot Y_1(\mathbf{u}) + 0.0 \cdot Y_2(\mathbf{u}) + \sqrt{0.5} \cdot Y_3(\mathbf{u}) \end{aligned}$$

with,

$$\begin{aligned} Y_1(\mathbf{u}) &\sim N(0,1) \text{ with } \gamma_{Y_1}(\mathbf{h}) = 0.1 + 0.9 \cdot Sph_{\left(\begin{smallmatrix} h \text{ max}=200 \\ h \text{ min}=200 \end{smallmatrix}\right)}(\mathbf{h}) \\ Y_2(\mathbf{u}) &\sim N(0,1) \text{ with } \gamma_{Y_2}(\mathbf{h}) = 0.1 + 0.9 \cdot Sph_{\left(\begin{smallmatrix} h \text{ max}=50 \\ h \text{ min}=300 \end{smallmatrix}\right)}(\mathbf{h}) \\ Y_3(\mathbf{u}) &\sim N(0,1) \text{ with } \gamma_{Y_3}(\mathbf{h}) = 0.1 + 0.9 \cdot Exp_{\left(\begin{smallmatrix} h \text{ max}=400 \\ h \text{ min}=100 \end{smallmatrix}\right)}(\mathbf{h}) \end{aligned}$$

Later the random variables  $Z_1(\mathbf{u})$  and  $Z_2(\mathbf{u})$  were transform to lognormal distributions to resemble a more realistic distribution of grades assumed to be copper,%:

$$\begin{aligned} Z_1(\mathbf{u}) &\sim \text{lognormal} (m=1, \sigma=1) \\ Z_2(\mathbf{u}) &\sim \text{lognormal} (m=2, \sigma=2) \end{aligned}$$

The 2D reference image (2000 by 1000 meters, with a 10 meters grid spacing in both directions) was sampled at a spacing of 70 meters in the X-direction and 10 meters in the Y-direction yielding a total of 2800 samples.

## Variography

Variograms were calculated from the normal scores transform values from each rock type,  $\mathbf{RT}_1$  and  $\mathbf{RT}_2$ . Cross variograms can not be calculated if the variables are not collocated, which is the case here since we are trying to characterize the spatial variability across the boundary between  $\mathbf{RT}_1$  and  $\mathbf{RT}_2$ . An alternative (Wawruch *et. al.* 2003) is to (1) calculate the cross covariance between the variables, (2)

extrapolate the experimental points at lags near to zero to obtain the structured cross covariance ( $B_{Z_1-Z_2}$ ) (Figure 2.10), (3) determine the relative nugget effects for  $Z_1$  and  $Z_2$ , and (4) calculate the sill of the cross variogram between  $Z_1(\mathbf{u})$  and  $Z_2(\mathbf{u})$  as:

$$Cov_{Z_1-Z_2}(0) = \frac{B_{Z_1-Z_2}}{1 - \left[ \frac{1}{2} \left( \frac{Cov_{Z_1}^0}{\sigma_{Z_1}^2} + \frac{Cov_{Z_2}^0}{\sigma_{Z_2}^2} \right) \right]}$$

In this example, the relative nugget effects obtained from the direct variograms of each rock type were both 0.1, the structured cross covariance was chosen at 0.4, so the sill of the cross variogram is 0.44. With this value the experimental points from the cross covariance can be inverted to obtain the cross variogram between  $Z_1(\mathbf{u})$  and  $Z_2(\mathbf{u})$ .

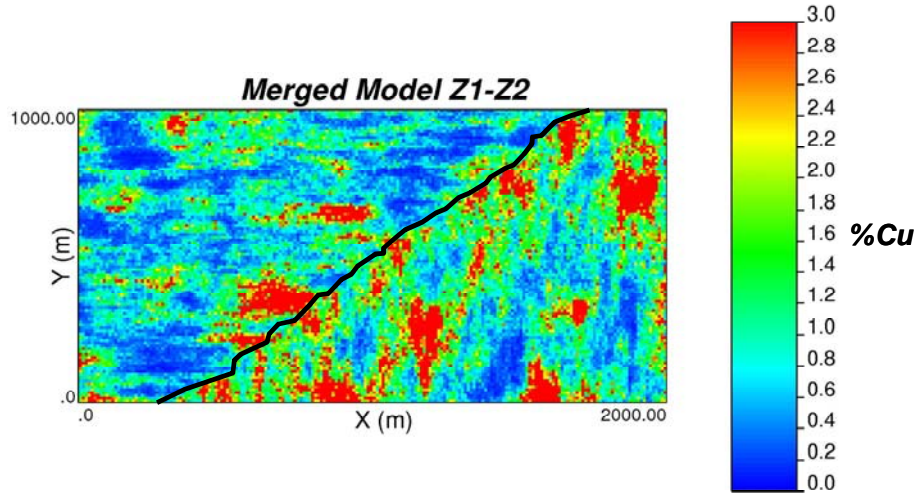


Figure 2.9: Reference or ‘true’ image.  $Z_1(\mathbf{u})$  is represented by  $\mathbf{RT}_1$ ,  $Z_2(\mathbf{u})$  by  $\mathbf{RT}_2$ .

As we discuss in the previous section, there will be some surface contacts that will present a higher uncertainty in the determination of the structured cross covariance. But in this case the nugget effect between the grade at each side of the boundary is not needed in any calculations because there is no collocated data nor do we estimate collocated grid blocks; most cokriging and cosimulation programs require the LMC to be defined with variogram models, which requires the nugget effect and the sill of the cross variogram.

The direct and cross variograms of  $Z_1(\mathbf{u})$  and  $Z_2(\mathbf{u})$  were model using a linear model of coregionalization (Table 2.1 and Figure 2.11) obtained by a semi-automatic variogram fitting program (Larrondo *et. al.*, 2003). Since independent simulations of

$Z_1(\mathbf{u})$  and  $Z_2(\mathbf{u})$  were also performed, the direct variograms of each variable were modeled independently (Table 2.2 and Figure 2.12).

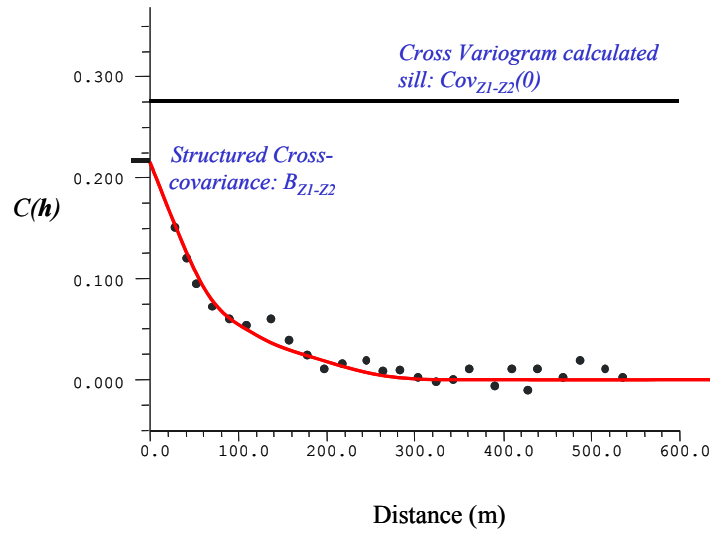


Figure 2.10: Sketch with the structured cross covariance and calculated sill of a cross variogram given an experimental cross covariance between two non-collocated variables.

Variable	Structure Type	Sill Contribution	Range Hmax	Range Hmin	Range Vert
Z1	Nugget Effect	0.100			
	Spherical	0.102	125.0	83.0	125.0
	Exponential	0.600	125.0	290.0	125.0
	Spherical	0.198	433.0	290.0	433.0
Z2	Nugget Effect	0.100			
	Spherical	0.524	125.0	83.0	125.0
	Exponential	0.005	125.0	290.0	125.0
	Spherical	0.371	433.0	290.0	433.0
Z1-Z2	Nugget Effect	0.100			
	Spherical	0.231	125.0	83.0	125.0
	Exponential	0.053	125.0	290.0	125.0
	Spherical	0.056	433.0	290.0	433.0

Table 2.1: LMC model for  $Z_1(\mathbf{u})$  and  $Z_2(\mathbf{u})$ .

Variable	Structure Type	Sill Contribution	Range Hmax	Range Hmin	Range Vert
Z1	Nugget Effect	0.10			
	Exponential	0.35	51.0	241.0	51.0
	Spherical	0.46	175.0	241.0	175.0
	Exponential	0.09	2010.0	2340.0	2010.0
Z2	Nugget Effect	0.10			
	Spherical	0.36	288.0	91.8	288.0
	Exponential	0.37	288.0	192.0	288.0
	Exponential	0.17	288.0	244.0	288.0

Table 2.2: Variogram models for  $Z_1(\mathbf{u})$  and  $Z_2(\mathbf{u})$ , obtained considering the two variables independently.

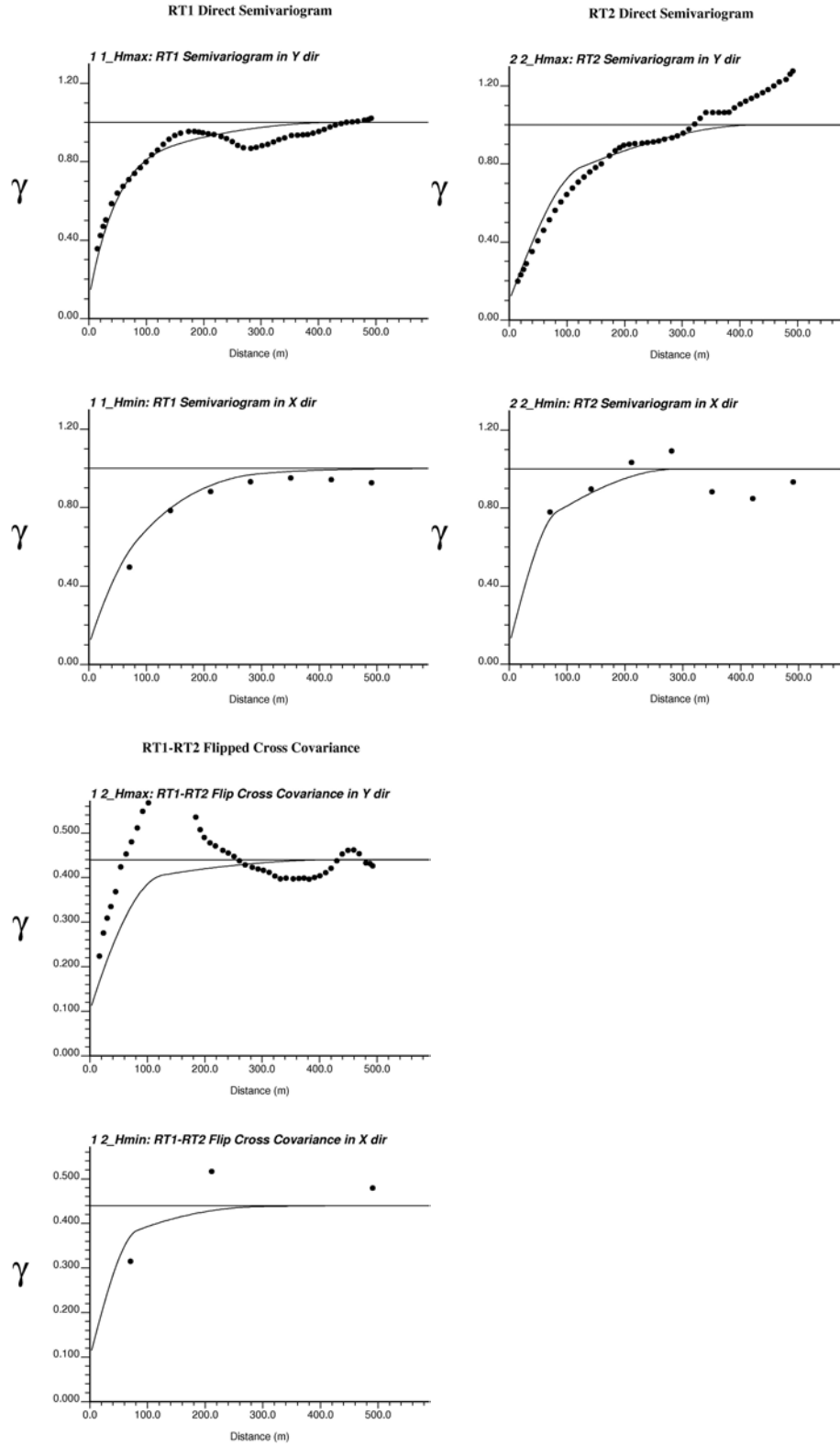


Figure 2.11: Direct and cross variograms and LMC model for  $Z_1(\mathbf{u})$  and  $Z_2(\mathbf{u})$ .

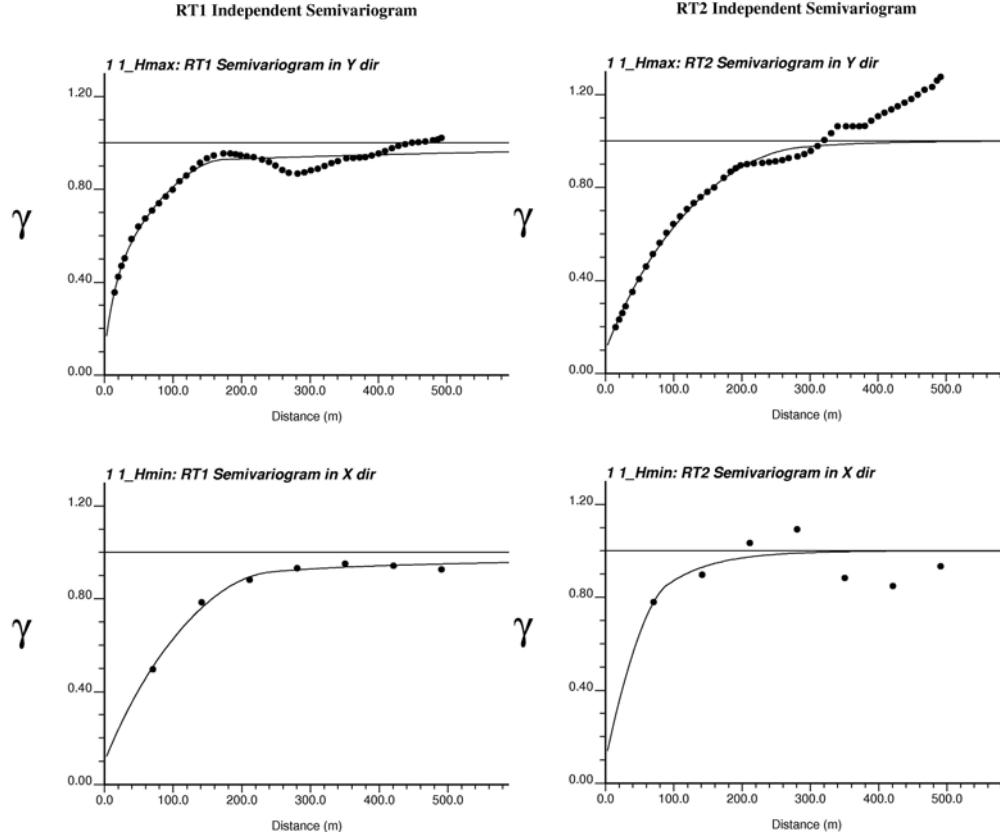


Figure 2.12: Variogram model for  $Z_1(\mathbf{u})$  and  $Z_2(\mathbf{u})$ , obtained independently.

### ***Simulation***

The cosimulation was performed using the full LMC cokriging option of the ultimate `sgsim` program (Deutsch and Zanon, 2002) and the parameters in Table 3; in this case each rock type was simulated using the samples of the other rock type, as a secondary variable. For the comparative case, sequential gaussian simulation (Table 4) was used to simulate each rock type independently as the contact between  $\mathbf{RT}_1$  and  $\mathbf{RT}_2$  was a hard boundary.

<b><i>Number of Nodes</i></b>	<i>24, maximum of 6 per octant</i>
<b><i>Search Radius</i></b>	<i>450 x 450 x 10 m</i>
<b><i>Number of Realizations</i></b>	<i>10</i>

Table 3: Parameters used for cosimulation of  $Z_1(\mathbf{u})$  and  $Z_2(\mathbf{u})$ .



<b><i>Number of Nodes</i></b>	<i>24, maximum of 6 per octant</i>
<b><i>Search Radius</i></b>	<i>300 x 300 x 10 m</i>
<b><i>Number of Realizations</i></b>	<i>10</i>

Table 4: Parameters used for independent simulations of  $Z_I(\mathbf{u})$  and  $Z_2(\mathbf{u})$ .

### ***Validation***

The reproduction of the direct variograms, both for the cosimulation or independent simulation schemes, was fairly good (Figure 2.13). Although the reproduction of the cross variogram was poor compared with the analytical model, the first 100 meters (total range) in the X-direction showed a similar amount of correlation (Figure 2.14). The case where the contact between  $\mathbf{RT}_1$  and  $\mathbf{RT}_2$  was assumed to be a hard boundary, resulted in almost no correlation for lags less than the range of the cross variogram, and is significantly lower than the correlation of the conditional data across the boundary. This correlation is a remnant correlation from data, not from modeling. While for a soft boundary assumption, the correlation of the average over all realizations is closer to the correlation shown by the ‘truth’ reference.

The validation of the parameters and variogram models used for independent simulation or cosimulation was done estimating a location where the value was removed from the dataset. This is done iteratively for all samples, but removing the entire drill hole. The estimation was done using the program `kt3d` in a cross validation mode for independent simulations, and a modified version of `cokb3d` program for cosimulation. Both programs return the estimated values for each sample location, the estimate variance and the error (true minus the estimate). With these results we can check the reproduction of the values, the distribution of the errors and whether the distributions of uncertainty are accurate and precise. Accuracy is achieved when the fraction of true values that fall in a probability interval of width  $p$  exceeds  $p$  for all  $p$  in  $[0,1]$  (Deutsch 2002). Precision is how closeness the fraction of true values is equal to  $p$  for all  $p$  in  $[0,1]$ . These properties are reflected in an accuracy plot (Figure 2.15) by the closeness of the points to the 45° line, if the points fall on the line, the distribution representing the model is accurate and precise, if the points fall above the line the distribution is accurate but not precise, if they fall below the model is neither accurate nor precise.

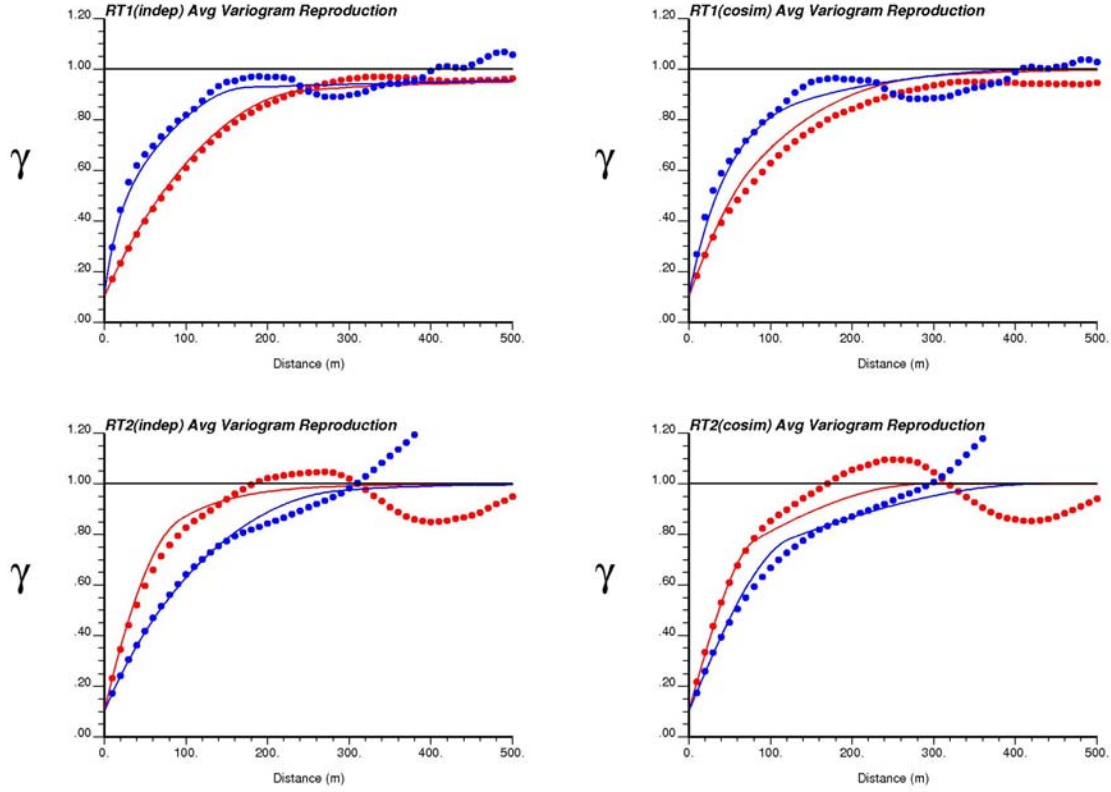


Figure 2.13: Direct variograms reproduction for  $Z_1(\mathbf{u})$  and  $Z_2(\mathbf{u})$ , cosimulated (right) and independently simulated (left). The dots represent the average of simulated values over ten realizations, the dash line correspond to the cross covariance calculated for the training image, and the solid line is the analytical model derived from the theoretical expression.

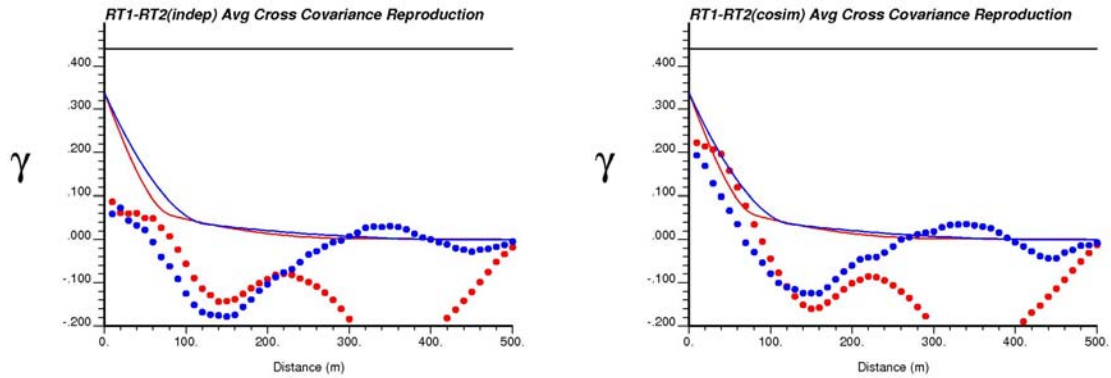


Figure 2.14: Cross covariance reproduction for  $Z_1(\mathbf{u})$  and  $Z_2(\mathbf{u})$ , cosimulated (right) and independently simulated (left). In a soft boundary scheme (right) the correlation between the simulated values is very close to the ‘truth’ reference. In the contrary, in a hard boundary assumption, the correlation at short lag distances is significantly lower. The dots represent the average of simulated values over ten realizations, the dash line correspond to the cross covariance calculated for the training image, and the solid line is the analytical model derived from the theoretical expression.

Cross validation of the model obtained by independently simulating  $Z_1(\mathbf{u})$  and  $Z_2(\mathbf{u})$  showed that the model is accurate and precise. The cosimulated model is also accurate, and equally precise for  $\mathbf{RT}_1$ , while for  $\mathbf{RT}_2$  is slightly less precise than the model obtained from independent simulations (Figure 2.15). This is not surprising since the fitted LMC model for this rock type did not fit the data as well as for  $\mathbf{RT}_1$ . This is a common disadvantage when using a linear model of coregionalization. Some variables are fit closer than others, which do not occur when the variables are modeled independently. The cosimulated model did, however, show less smoothing (Figure 2.16), which translates to less conditional bias in the estimation.

Although both model are similarly accurate and precise the overall uncertainty, defined as the average kriging variance of all samples, is significantly lower for the cosimulated model (0.3 for both  $Z_1(\mathbf{u})$  and  $Z_2(\mathbf{u})$ ) than for the independently simulated model (0.62 for  $Z_1(\mathbf{u})$  and 0.91 for  $Z_2(\mathbf{u})$ ).

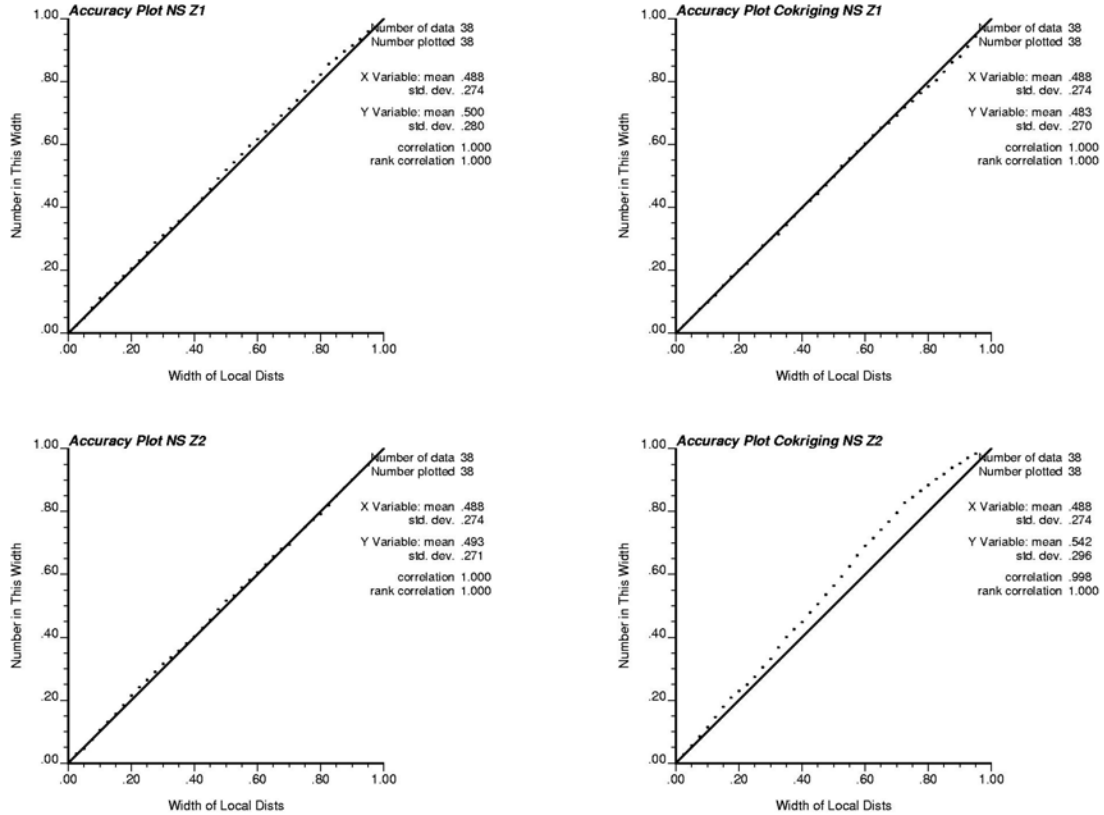


Figure 2.15: Accuracy plot for  $Z_1$  and  $Z_2$ , estimate independently (left) versus cosimulated (right). Cross validation show that the models from independent simulation or cosimulation of  $Z_1$  are accurate and precise. For  $Z_2$  the parameters used for cosimulation results a slightly less precise model than in the case of independent simulation.

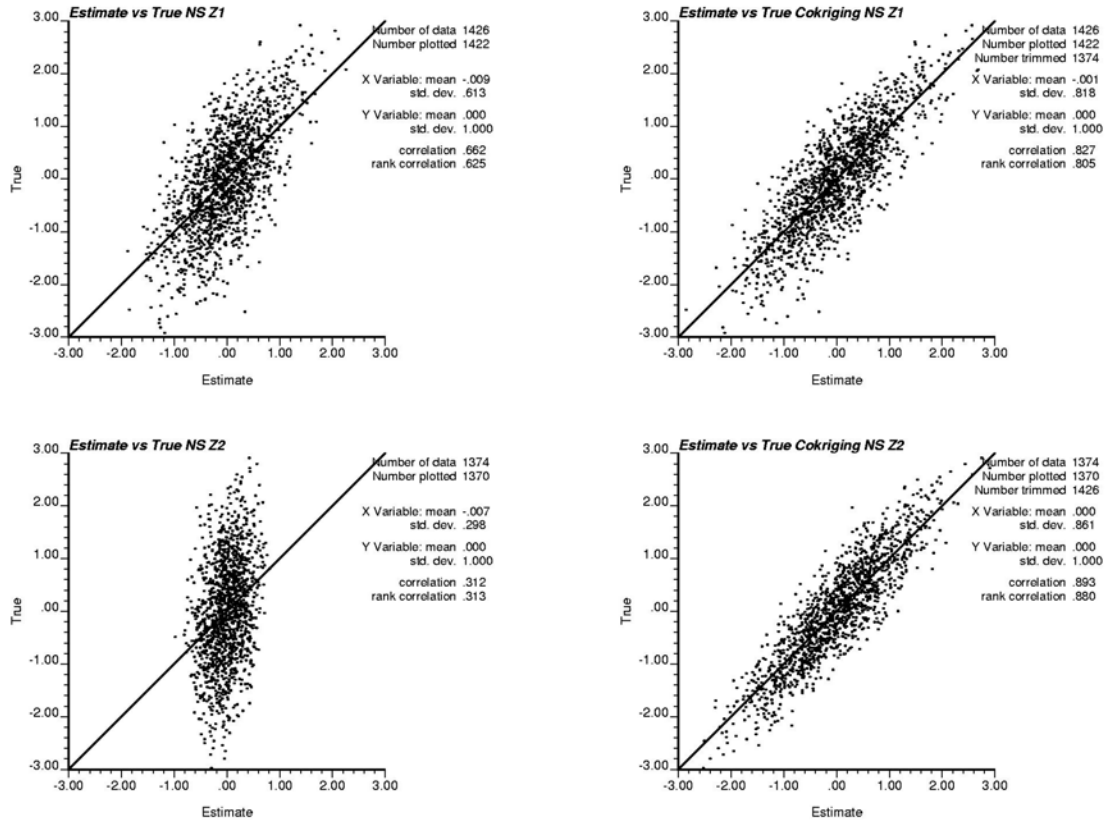


Figure 2.16: Cross validation of data values in  $RT_1$  and  $RT_2$ , estimate independently (left) versus cosimulated (right). The cokriging cross validation show far less conditional bias than the estimation of each rock type independently, especially for  $RT_2$ .

The distribution of errors (true-estimated) should be symmetric and centered at zero, as occurs for both schemes, but the standard deviation of the errors for cokriged estimates is significantly lower than independent kriged values, as shown in Figure 2.17.

The cumulative distribution of back transformed simulated values shows a very good reproduction of the data histograms, for both schemes. The target mean and variance are well reproduced for both cosimulation and independent simulations (Figure 2.18).

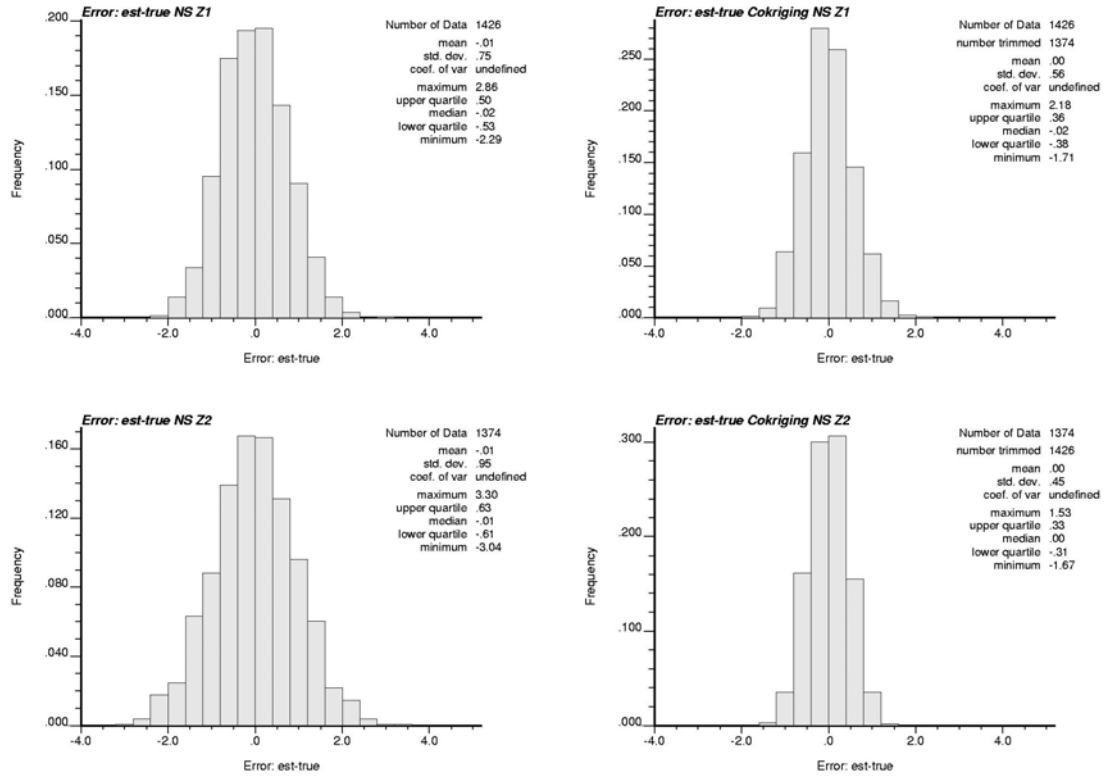


Figure 2.17: Error (true-estimated) distribution for  $Z_1$  and  $Z_2$ , estimate independently (left) versus cosimulated (right). The cosimulation scheme shows significantly lower standard deviation of the errors, in both rock types, than independent simulation.

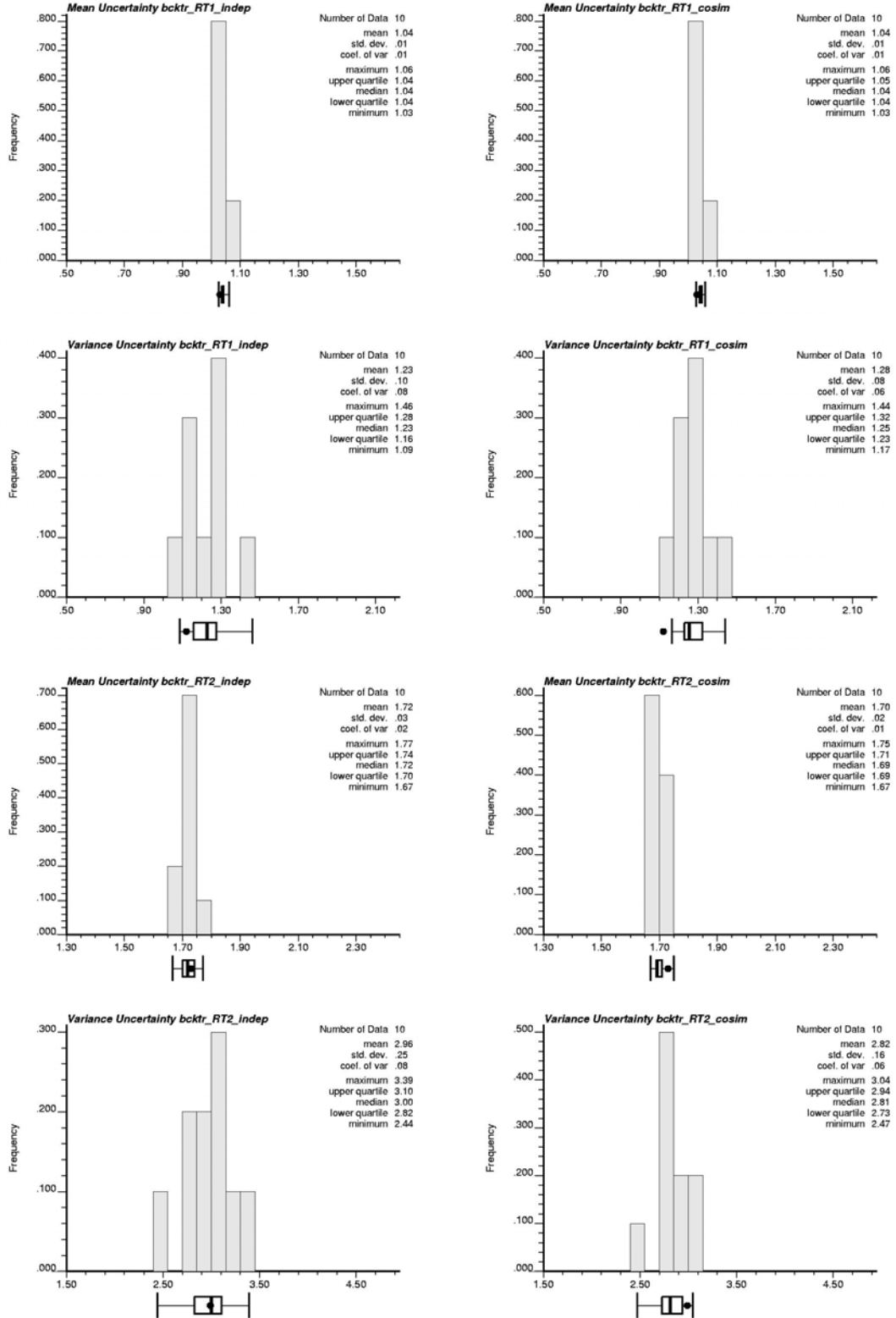


Figure 2.18: Global mean and variance reproduction for  $Z_1$  and  $Z_2$ ; cosimulation (right side) and independent simulation (left side). The dots represent the target values obtained from the conditioning data.

## Comparison at the boundary

In order to compare the performance of the two methods, we need to focus on the results near the boundary where we can expect to have greater differences.

One comparison was done using the expected value (E-type value) in original units at each location compared to the ‘true’ value in the reference map. The expected value is the average of the simulated realizations at each location. The block values obtained from cosimulation show systematically higher correlation coefficients with the true values. As expected, the difference between the two methods becomes smaller beyond the range of correlation of the cross variogram (Figure 2.19).

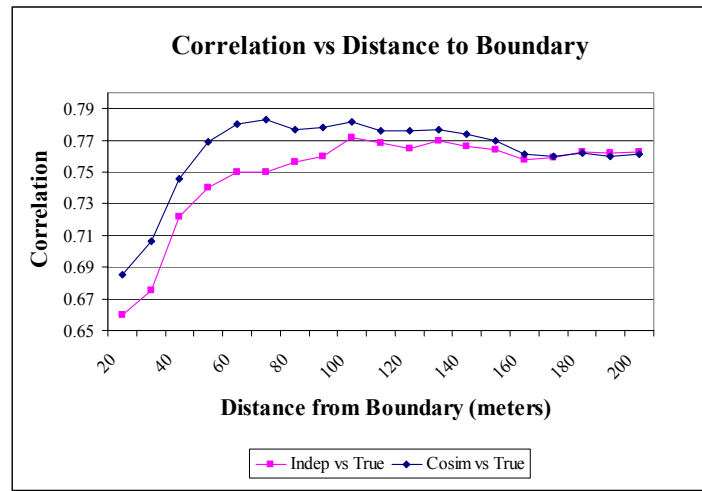


Figure 2.19: Correlation coefficient between E-type estimates of cosimulated and independently simulated models, and the “true” values considering blocks within a given distance from the boundary between  $Z_1$  and  $Z_2$ . The higher correlation coefficient with the true values shown by the blocks estimated by cosimulation indicate this model better represents the underlying correlation that exists between  $Z_1(\mathbf{u})$  and  $Z_2(\mathbf{u})$ .

The variance of the blocks from each realization within a given distance from the boundary was also compared. As expected, the average of the variance over all the realizations showed lower variance for the block values obtained using cosimulation (Figure 2.20). This variance is also closer to the average variance calculated from the same group of blocks in the ‘true’ reference map.

The methodology proposed in this section has the advantage of improved resource estimation by reducing the uncertainty in transitional zones near boundaries and reproducing the correlation of the conditioning data across a soft boundary. It also shows a decrease of smoothing in the estimates if kriging is the tool to obtain the resources.

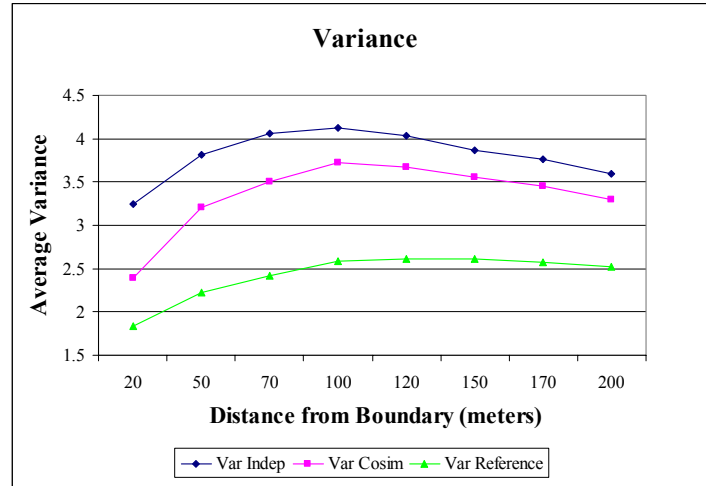


Figure 2.20: Average variance calculated from blocks within a given distance from the boundary between  $Z_1$  and  $Z_2$ . The average variance obtained from cosimulation is lower than the variance obtained from independent simulations. The difference between the two methods decreases as we consider blocks further away from the boundary, where the influence of the data from the adjacent rock type in the cosimulation decreases. The average variance from cosimulated blocks is closer to the variance of the same blocks in the ‘true’ reference map, than the average variance obtained from independent simulations.



### 3. Local Non-stationary Model of Coregionalization

The global LMC described in the previous chapter assumes that the grades are stationary within each domain and any cross correlation is also stationary and extends across the entire domain. This chapter will describe a new methodology to account for non-stationary factors near boundaries. We will show that non-stationary features in the vicinity of a boundary can be parameterized into a local model of coregionalization. With a legitimate spatial model, estimation of grades can be performed using a form of non-stationary cokriging.

#### Theoretical Background

For the case of a geological model with  $K$  rock types or domains, there are a maximum of  $K(K-1)/2$  boundary zones to be defined. These regions may be characterized by a non-stationary behavior of the variable of interest. The zones of influence of one region into an adjacent are not necessarily symmetric on each side of the boundary and there may be cases of truly hard boundaries where there are no special features at the boundary. When more than two rock types converge at a boundary, two or more rock types may influence the boundary zone in the adjacent domain. In this case, precedence or ordering rules should determine the dominant boundary zone. Only one non-stationary factor is dominant at each location.

The number of rock types to model  $K$ , the extent of the boundary zones and a set of precedence rules are required to partition a geological model into stationary and non-stationary regions. Each of these concepts will be described below.

The continuous random function  $Z(\mathbf{u})$  that represents the distribution of the property of interest, e.g. metal content, over the area of study, can be decomposed into  $K$  stationary random variables  $Z_k(\mathbf{u})$   $k=1, \dots, K$  and a maximum of  $K(K-1)/2$  non-stationary boundary variables  $Z_{kp}(\mathbf{u})$ , with  $k, p=1, \dots, K$  and  $Z_{kp}(\mathbf{u})=Z_{pk}(\mathbf{u})$ . Then at all locations,  $Z(\mathbf{u})$  can be explained by the sum of a stationary component from the collocated rock type and perhaps a single non-stationary boundary variable, that is,

$$Z(\mathbf{u}) = Z_k(\mathbf{u}) + Z_{kp}(\mathbf{u}) \quad \text{where } \mathbf{u} \in \text{rock type } k \quad (3.1)$$

and  $p$  is any of the surrounding rock types that share a boundary with rock type  $k$  ( $\mathbf{RT}_k$ ).

### Maximum Distance of Influence

The maximum distance of influence orthogonal to the boundary of rock type  $k$  into rock type  $p$  is denoted  $dmax_{kp}$ . A boundary zone is defined by two distances:  $dmax_{kp}$  and  $dmax_{pk}$ , since there is no requirement that the regions at each side of the boundary are symmetric, that is,  $dmax_{kp} \neq dmax_{pk}$ . Figure 3.1 shows a schematic illustration of this.

The modeler using all geological information available and his expertise should establish these distances. Creating an average value profile against distance to boundary could be helpful in identifying the zones of influence between domains. An automatic optimization algorithm is unlikely to work given that the stationary portion of the mean and variance, as well as the non-stationary factors that affect this statistics, are also unknown.

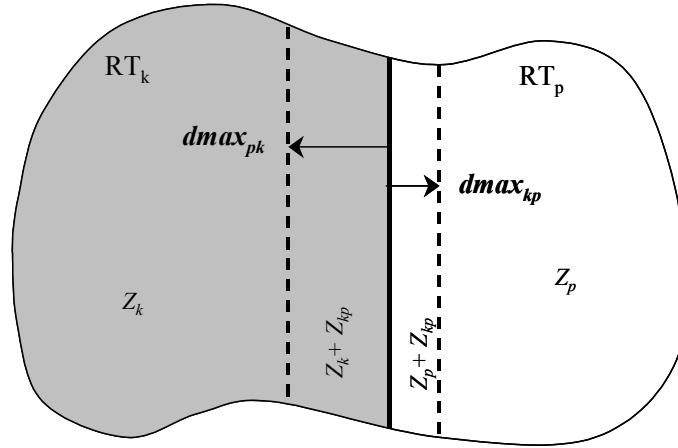


Figure 3.1: Maximum distances of influence for the boundary zone between rock type  $k$  ( $RT_k$ ) and rock type  $p$  ( $RT_p$ ).

All maximum distances can be arranged in a non-symmetric square  $K$  by  $K$  matrix:

$$\begin{bmatrix} -1 & dmax_{12} & \cdots & dmax_{1K} \\ dmax_{21} & \ddots & & \vdots \\ \vdots & & \ddots & dmax_{K-1 K} \\ dmax_{K1} & \cdots & dmax_{K K-1} & -1 \end{bmatrix}$$

The maximum distance of influence of a rock type to itself is meaningless and any default value could be assigned (-1 in the sketch above).

A rock type is assigned at all locations within the geological model. Where the distance to the boundary is lower than the maximum distance of influence of the adjacent rock, a boundary zone rock type is also assigned. We will denote the region corresponding to the zone of influence of rock type  $p$  into rock type  $k$  (Figure 3.2) as the boundary zone  $\mathbf{BZ}_{pk}$  and it is defined as all locations  $\mathbf{u}_i$  in rock type  $k$ , such that  $d_{pk}(\mathbf{u}_i) \leq dmax_{pk}$ , where  $d_{pk}(\mathbf{u})$  is the shortest distance to  $k$ - $p$  boundary. Equivalently, the region on the other side of the boundary between rock type  $k$  and  $p$ , is denoted as the boundary zone  $\mathbf{BZ}_{kp}$ .

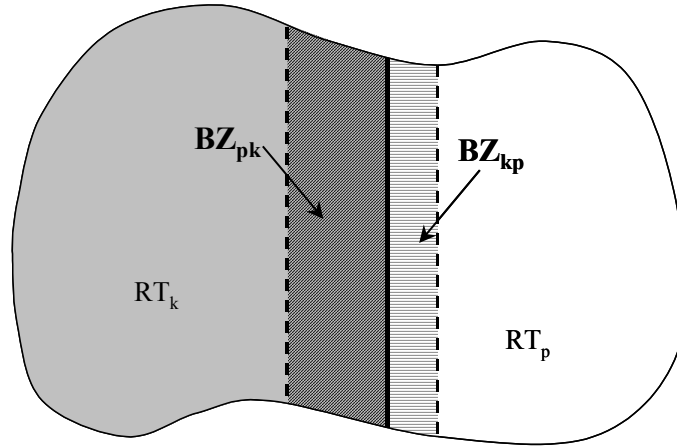


Figure 3.2: Boundary zone  $\mathbf{BZ}_{pk}$  and  $\mathbf{BZ}_{kp}$  corresponding to the area of influence of rock type  $p$  into rock type  $k$  and vice versa.

### ***Precedence Rules***

A set of precedence rules is needed to identify the predominant non-stationary factor at a boundary where more than two rock types converge. In Figure 3.3, a three rock type spatial arrangement is shown where precedence rules are used to establish the precedent rock types in the region A and B: in region A of  $\mathbf{RT}_k$ ,  $\mathbf{RT}_p$  is precedent over  $\mathbf{RT}_q$ , while in region B of  $\mathbf{RT}_p$ ,  $\mathbf{RT}_q$  is precedent over  $\mathbf{RT}_k$ .

Although the behavior of a property near such a boundary, could be explained by the overlapping of different geological controls, the task of identifying the individuals effects of each rock type and their interactions can be quite difficult. Geological properties are not usually additive and therefore the response of a combination of different rock types is complex.

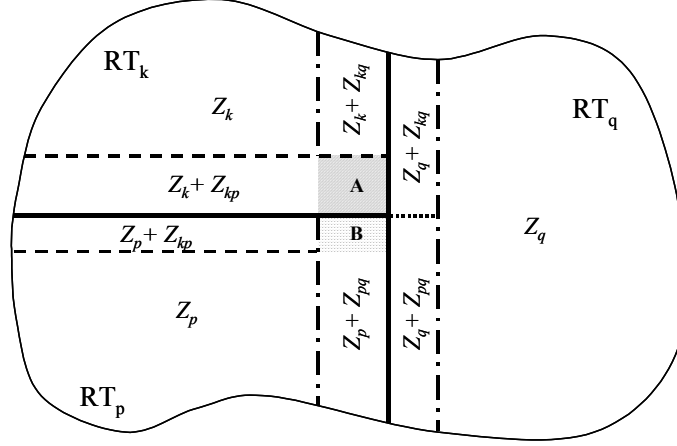


Figure 3.3: Three rock type example where a predefined precedence rule is used to determine the precedent rock type over two possibilities. In the region A of  $\mathbf{RT}_k$ ,  $\mathbf{RT}_p$  is precedent over  $\mathbf{RT}_q$ , while in region B of  $\mathbf{RT}_p$ ,  $\mathbf{RT}_q$  is precedent over  $\mathbf{RT}_k$ .

The precedence rules should be built by the modeler using all geological information available. The relative timing of intrusion, deposition or mineralisation events, geochemistry response of the protolith to an alteration or mineralisation process could be used to resolve timing and important variables. If the geological data does not provide sufficient information to establish a geological order of events, a neutral arrangement can be chosen. In this case, the precedent rock type  $p$  at a location  $\mathbf{u}$  will be the one to which the distance to the boundary is the minimum over all surrounding rock types, that is,  $\min \{d_{pk}(\mathbf{u})\}$  with  $p=1, \dots, K-1$  and  $\mathbf{u}$  belongs to  $\mathbf{BZ}_{pk}$ .

For the same example illustrated in Figure 3.3, the result of a neutral arrangement is shown in Figure 3.4.

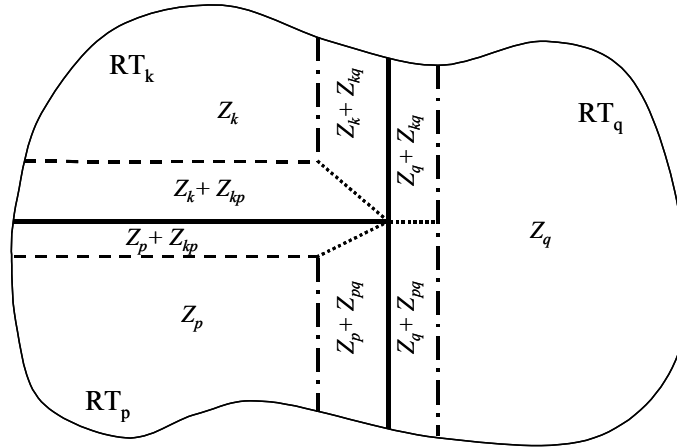


Figure 3.4: Three-rock type example where a neutral arrangement is chosen to define the shape of the zones of influence where two boundaries converge.

### Statistical Parameters

The continuous random function,  $Z(\mathbf{u})$  that represents the variable of interest over the area study, can be decomposed into stationary and non-stationary random variables by the rock type model, as shown in Equation 3.1. By definition, the non-stationary variable will take values only for locations within a distance to the boundary:  $d_{pk}(\mathbf{u}) \leq dmax_{pk}$ . A one dimensional example of how the random variable is decomposed into stationary and non-stationary components is presented in Figure 3.5.

#### Mean

The mean function of the continuous random function  $Z(\mathbf{u})$  for an specific rock type  $k$  will be the mean of the stationary variable  $Z_k(\mathbf{u})$  plus the mean of any corresponding non-stationary variable  $Z_{kp}(\mathbf{u})$ :

$$E\{Z(\mathbf{u}_i)\} = E\{Z_k(\mathbf{u}_i)\} + E\{Z_{kp}(\mathbf{u}_i)\} = m_k + m_{kp}(\mathbf{u}_i) \quad \text{where } \mathbf{u}_i \in \mathbf{RT}_k \quad (3.2)$$

where  $p$  is the adjacent rock type that shares a boundary with rock type  $k$ .

The stationary component of the mean ( $m_k$ ) is independent of location and is a constant value. The non-stationary component of the mean ( $m_{kp}$ ) is a function of the distance to the boundary,  $d_{pk}(\mathbf{u})$  and takes values different than zero for locations within the boundary zone  $\mathbf{BZ}_{pk}$ :

$$m_{kp}(d_{pk}(\mathbf{u}_i)) = \begin{cases} 0 & , \text{ if } d_{pk}(\mathbf{u}_i) \geq dmax_{pk} \\ f(d_{pk}(\mathbf{u}_i)) & , \text{ otherwise} \end{cases} \quad \text{where } \mathbf{u}_i \in \mathbf{RT}_k$$

The non-stationary mean could be, for example, approximated by a linear function as we have chosen for the implementation presented later (see Chapter 4). There may be cases where a different approximation could be better. The methodology presented here could easily be adapted.

Rewriting Equation 3.2, the mean of rock type  $k$  in the presence of  $P$  non-stationary boundaries is:

$$E\{Z(\mathbf{u}_i)\} = \begin{cases} m_k & , \text{ if } d_{pk}(\mathbf{u}_i) \geq dmax_{pk} \\ m_k + f(d_{pk}(\mathbf{u}_i)) & , \text{ otherwise} \end{cases} \quad \text{where } \mathbf{u}_i \in \mathbf{RT}_k \quad (3.3)$$

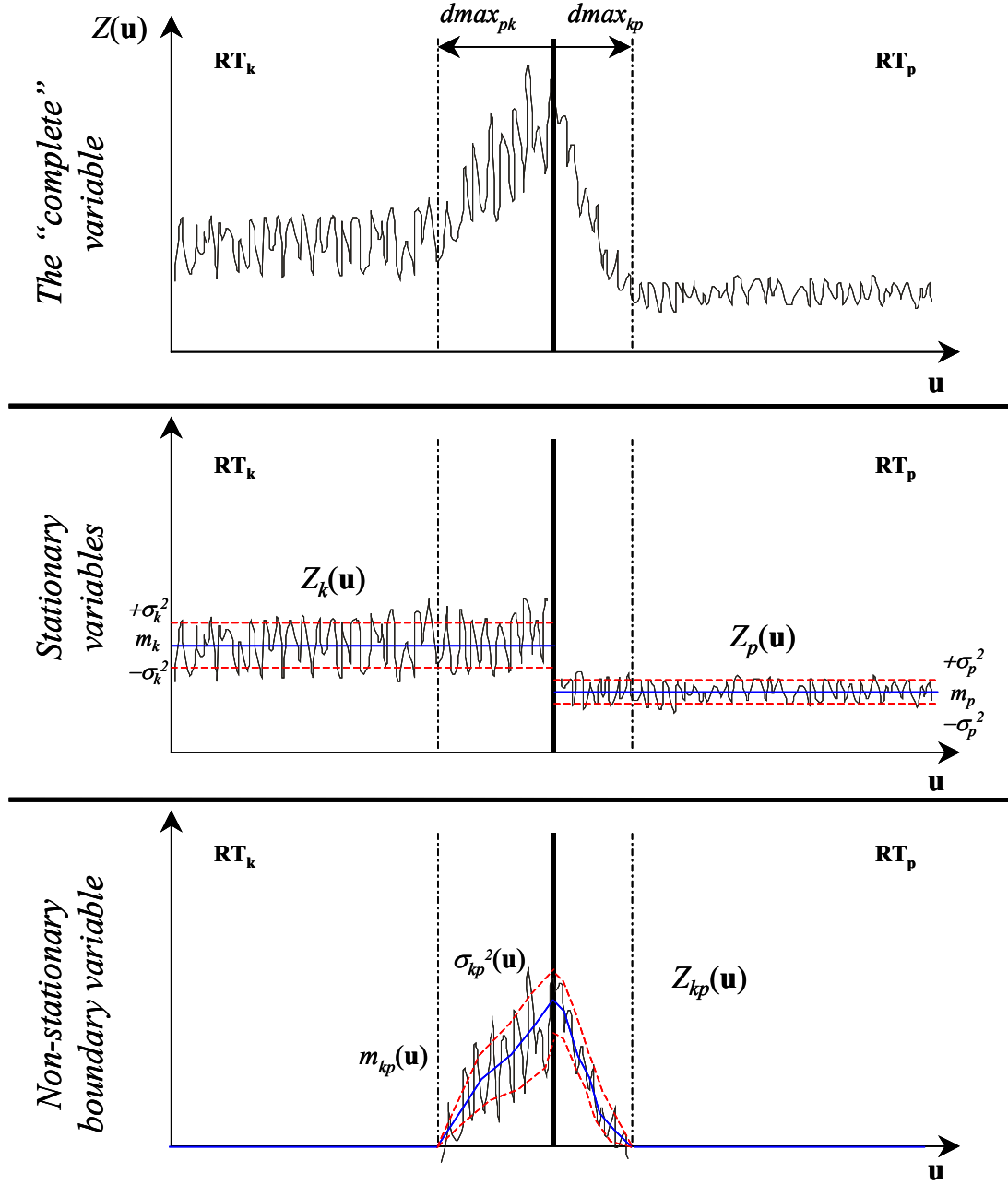


Figure 3.5: Decomposition of the random function  $Z(\mathbf{u})$  in two stationary variables  $Z_k(\mathbf{u})$  and  $Z_p(\mathbf{u})$ , with constant mean and variance, and a non-stationary boundary variable  $Z_{kp}(\mathbf{u})$ , with a mean and variance that are functions of the distance to the boundary.

### Variance

Similarly, the variance of  $Z(\mathbf{u})$  for rock type  $k$  will be the sum of the stationary variance due to  $Z_k(\mathbf{u})$  and the independent non-stationary variance due to  $Z_{kp}(\mathbf{u})$  for the  $p$  adjacent rock type to rock type  $k$ , that is,

$$E\left\{\left(Z(\mathbf{u}_i) - E\{Z(\mathbf{u}_i)\}\right)^2\right\} = \sigma_k^2 + \sigma_{kp}^2(\mathbf{u}_i) \quad \text{where } \mathbf{u}_i \in \mathbf{RT}_k \quad (3.4)$$

Since  $Z_k(\mathbf{u})$  and  $Z_{kp}(\mathbf{u})$  are independent random variables, the cross terms of the squared product are zero;

$$\begin{aligned} E\left\{\left(Z(\mathbf{u}_i) - E\{Z(\mathbf{u}_i)\}\right)^2\right\} &= E\left\{\left[\left(Z_k(\mathbf{u}_i) + Z_{kp}(\mathbf{u}_i)\right) - \left(m_k + m_{kp}(\mathbf{u}_i)\right)\right]^2\right\} \\ &= E\left\{\left(Z_k(\mathbf{u}_i) + Z_{kp}(\mathbf{u}_i)\right)^2 - 2\left(Z_k(\mathbf{u}_i) + Z_{kp}(\mathbf{u}_i)\right) \cdot \left(m_k + m_{kp}(\mathbf{u}_i)\right) + \left(m_k + m_{kp}(\mathbf{u}_i)\right)^2\right\} \\ &= E\left\{Z_k(\mathbf{u}_i)^2\right\} + 2E\left\{Z_k(\mathbf{u}_i) \cdot Z_{kp}(\mathbf{u}_i)\right\} + E\left\{Z_{kp}(\mathbf{u}_i)^2\right\} - m_k^2 - 2m_k \cdot m_{kp}(\mathbf{u}_i) - m_{kp}(\mathbf{u}_i)^2 \\ &= E\left\{Z_k(\mathbf{u}_i)^2\right\} + 2E\left\{Z_k(\mathbf{u}_i)\right\} \cdot E\left\{Z_{kp}(\mathbf{u}_i)\right\} + E\left\{Z_{kp}(\mathbf{u}_i)^2\right\} - m_k^2 - 2m_k \cdot m_{kp}(\mathbf{u}_i) - m_{kp}(\mathbf{u}_i)^2 \\ &= E\left\{Z_k(\mathbf{u}_i)^2\right\} - m_k^2 + E\left\{Z_{kp}(\mathbf{u}_i)^2\right\} - m_{kp}(\mathbf{u}_i)^2 \\ &= \sigma_k^2 + \sigma_{kp}^2(\mathbf{u}_i) \end{aligned}$$

As with the mean, the stationary component of the variance ( $\sigma_k^2$ ) is independent of location and is a constant value. The non-stationary component of the variance ( $\sigma_{kp}^2$ ) is a function of the distance to the boundary. Beyond the maximum distance of influence it is equal to zero. Within the boundary zone  $\mathbf{BZ}_{pk}$  it is defined as:

$$\sigma_{kp}^2(d_{pk}(\mathbf{u}_i)) = \begin{cases} 0 & , \text{ if } d_{pk}(\mathbf{u}_i) \geq dmax_{pk} \\ g(d_{pk}(\mathbf{u}_i)) & , \text{ otherwise} \end{cases} \quad \text{where } \mathbf{u}_i \in \mathbf{RT}_k$$

As will be shown in following examples and implementation,  $g(d_{pk}(\mathbf{u}_i))$  was chosen as a linear function.

A summarized expression for the variance of a random function  $Z(\mathbf{u})$  in a rock type  $k$  with  $P$  non-stationary boundaries is:

$$E\left\{\left(Z(\mathbf{u}_i) - E\{Z(\mathbf{u}_i)\}\right)^2\right\} = \begin{cases} \sigma_k^2 & , \text{ if } d_{pk}(\mathbf{u}_i) \geq dmax_{pk} \\ \sigma_k^2 + g(d_{pk}(\mathbf{u}_i)) & , \text{ otherwise} \end{cases} \quad \text{where } \mathbf{u}_i \in \mathbf{RT}_k$$

## Covariance

As with the mean and variance, the covariance structure between two rock types that share a local non-stationary boundary consists of: a stationary and a non-stationary component.

$$Cov_Z(\mathbf{u}_i, \mathbf{v}_i) = E\{(Z(\mathbf{u}_i) - m(\mathbf{u}_i)) \cdot (Z(\mathbf{v}_i) - m(\mathbf{v}_i))\} = Cov_Z^S(\mathbf{h}) + Cov_Z^{NS}(\mathbf{u}_i, \mathbf{v}_i)$$

where  $\mathbf{h} = \mathbf{u}_i - \mathbf{v}_i$ .

Since  $Z_k(\mathbf{u})$  and  $Z_{kp}(\mathbf{u})$  are independent random variables, the cross terms are zero, therefore the covariance of  $Z(\mathbf{u})$  is the sum of the stationary and non-stationary components. The combination of these components corresponds to a local linear model of coregionalization.

Depending whether  $\mathbf{u}_i$  and  $\mathbf{v}_i$ , only one of them or neither of them, fall at the boundary zone  $\mathbf{BZ}_{pk}$ , the covariance structure of  $Z(\mathbf{u})$  will be sum of the stationary and non-stationary component or only the stationary component,

$$Cov_Z(\mathbf{u}_i, \mathbf{v}_i) = \begin{cases} Cov_Z^S(\mathbf{h}) & , \text{if } \mathbf{u}_i \in \mathbf{RT}_k, \notin \mathbf{BZ}_{pk} \text{ and } \mathbf{v}_i \in \mathbf{RT}_k, \notin \mathbf{BZ}_{pk} \\ Cov_Z^S(\mathbf{h}) + Cov_Z^{NS}(\mathbf{u}_i, \mathbf{v}_i) & , \text{if } \mathbf{u}_i \in \mathbf{RT}_k, \notin \mathbf{BZ}_{pk} \text{ and } \mathbf{v}_i \in \mathbf{RT}_k, \in \mathbf{BZ}_{pk} \\ Cov_Z^S(\mathbf{h}) + Cov_Z^{NS}(\mathbf{u}_i, \mathbf{v}_i) & , \text{if } \mathbf{u}_i \in \mathbf{RT}_k, \in \mathbf{BZ}_{pk} \text{ and } \mathbf{v}_i \in \mathbf{RT}_k, \in \mathbf{BZ}_{pk} \\ Cov_Z^{NS}(\mathbf{u}_i, \mathbf{v}_i) & , \text{if } \mathbf{u}_i \in \mathbf{RT}_k, \in \mathbf{BZ}_{pk} \text{ and } \mathbf{v}_i \in \mathbf{RT}_p, \in \mathbf{BZ}_{kp} \\ 0 & , \text{if } \mathbf{u}_i \in \mathbf{RT}_k, \notin \mathbf{BZ}_{pk} \text{ and } \mathbf{v}_i \in \mathbf{RT}_p, \in \mathbf{BZ}_{kp} \\ 0 & , \text{if } \mathbf{u}_i \in \mathbf{RT}_k, \notin \mathbf{BZ}_{pk} \text{ and } \mathbf{v}_i \in \mathbf{RT}_p, \notin \mathbf{BZ}_{kp} \end{cases} \quad (3.5)$$

If the head and tail of the pair are in different rock types, the stationary component of the covariance structure is zero, since there is no correlation between the stationary variables  $Z_k(\mathbf{u})$  and  $Z_p(\mathbf{u})$  across the boundary.

The stationary component of the covariance can be calculated and model from data pairs within the same internal stationary portion of a rock type, that is  $\mathbf{u}_i$  and  $\mathbf{v}_i$  belong to rock type  $k$ , and do not belong to any  $\mathbf{BZ}_{pk}$ .

The inference of the covariance structure will be possible provided a reasonable number of pairs exist within the stationary and non-stationary regions.

To obtained the non-stationary component of the covariance model we will assume that the shape, anisotropies and relative nugget effect of the correlation for the non-stationary variable  $Z_{kp}(\mathbf{u})$   $k, p = 1, \dots, K$  are stationary and that they can be specified by the modeler. But due to the non-stationary nature of variable  $Z(\mathbf{u})$  at the boundary zone, this stationary spatial model shape has to by scaled at each point by a non-



stationary mean and variance. The relative standardized variogram model for the boundary zone,  $\mathbf{BZ}_{pk}$ , corresponding to the stationary shape is:

$$\hat{\gamma}_{kp}(\mathbf{u}_i, \mathbf{v}_i) = \frac{1}{2} \cdot E \left\{ \left[ \frac{Z(\mathbf{u}_i) - (m_{kp}(\mathbf{u}_i) + m_k)}{(\sigma_{kp}(\mathbf{u}_i) + \sigma_k)} - \frac{Z(\mathbf{v}_i) - (m_{kp}(\mathbf{v}_i) + m_k)}{(\sigma_{kp}(\mathbf{v}_i) + \sigma_k)} \right]^2 \right\}$$

Expanding the squared difference we obtained that:

$$2\hat{\gamma}_{kp}(\mathbf{u}_i, \mathbf{v}_i) = E \left\{ \begin{aligned} & \frac{Z(\mathbf{u}_i)^2 - 2Z(\mathbf{u}_i) \cdot m(\mathbf{u}_i) + m(\mathbf{u}_i)^2}{\sigma(\mathbf{u}_i)^2} \\ & - 2 \left( \frac{Z(\mathbf{u}_i) \cdot Z(\mathbf{v}_i) - Z(\mathbf{v}_i) \cdot m(\mathbf{u}_i) - Z(\mathbf{u}_i) \cdot m(\mathbf{v}_i) + m(\mathbf{u}_i) \cdot m(\mathbf{v}_i)}{\sigma(\mathbf{u}_i) \cdot \sigma(\mathbf{v}_i)} \right) \\ & + \frac{Z(\mathbf{v}_i)^2 - 2Z(\mathbf{v}_i) \cdot m(\mathbf{v}_i) + m(\mathbf{v}_i)^2}{\sigma(\mathbf{v}_i)^2} \end{aligned} \right\}$$

To simplify the notation the mean and variance of  $\mathbf{u}_i$  and  $\mathbf{v}_i$  has been replace by:

$$\begin{aligned} m(\mathbf{u}_i) &= m_{kp}(\mathbf{u}_i) + m_k \quad \text{if } \mathbf{u}_i \in \mathbf{RT}_k \quad \text{or} \quad m(\mathbf{u}_i) = m_{kp}(\mathbf{u}_i) + m_p \quad \text{if } \mathbf{u}_i \in \mathbf{RT}_p \\ \sigma(\mathbf{u}_i) &= \sigma_{kp}(\mathbf{u}_i) + \sigma_k \quad \text{if } \mathbf{u}_i \in \mathbf{RT}_k \quad \text{or} \quad \sigma(\mathbf{u}_i) = \sigma_{kp}(\mathbf{u}_i) + \sigma_p \quad \text{if } \mathbf{u}_i \in \mathbf{RT}_p \end{aligned}$$

and

$$\begin{aligned} m(\mathbf{v}_i) &= m_{kp}(\mathbf{v}_i) + m_k \quad \text{if } \mathbf{v}_i \in \mathbf{RT}_k \quad \text{or} \quad m(\mathbf{v}_i) = m_{kp}(\mathbf{v}_i) + m_p \quad \text{if } \mathbf{v}_i \in \mathbf{RT}_p \\ \sigma(\mathbf{v}_i) &= \sigma_{kp}(\mathbf{v}_i) + \sigma_k \quad \text{if } \mathbf{v}_i \in \mathbf{RT}_k \quad \text{or} \quad \sigma(\mathbf{v}_i) = \sigma_{kp}(\mathbf{v}_i) + \sigma_p \quad \text{if } \mathbf{v}_i \in \mathbf{RT}_p \end{aligned}$$

Since  $E\{Z(\mathbf{u}_i)^2\} = \sigma(\mathbf{u}_i)^2 + m(\mathbf{u}_i)^2$  and  $E\{Z(\mathbf{u}_i)\} = m(\mathbf{u}_i)$ , we can simplify the previous expression as follows,

$$\begin{aligned} 2\hat{\gamma}_{kp}(\mathbf{u}_i, \mathbf{v}_i) &= \frac{\sigma(\mathbf{u}_i)^2 + m(\mathbf{u}_i)^2 - 2 \cdot m(\mathbf{u}_i)^2 + m(\mathbf{u}_i)^2}{\sigma(\mathbf{u}_i)^2} \\ &\quad - 2 \left( \frac{E\{Z(\mathbf{u}_i) \cdot Z(\mathbf{v}_i)\} - m(\mathbf{v}_i) \cdot m(\mathbf{u}_i) - m(\mathbf{u}_i) \cdot m(\mathbf{v}_i) + m(\mathbf{u}_i) \cdot m(\mathbf{v}_i)}{\sigma(\mathbf{u}_i) \cdot \sigma(\mathbf{v}_i)} \right) \\ &\quad + \frac{\sigma(\mathbf{v}_i)^2 + m(\mathbf{v}_i)^2 - 2 \cdot m(\mathbf{v}_i)^2 + m(\mathbf{v}_i)^2}{\sigma(\mathbf{v}_i)^2} \\ 2\hat{\gamma}_{kp}(\mathbf{u}_i, \mathbf{v}_i) &= 2 - 2 \left( \frac{E\{Z(\mathbf{u}_i) \cdot Z(\mathbf{v}_i)\} - m(\mathbf{v}_i) \cdot m(\mathbf{u}_i)}{\sigma(\mathbf{u}_i) \cdot \sigma(\mathbf{v}_i)} \right) \\ \hat{\gamma}_{kp}(\mathbf{u}_i, \mathbf{v}_i) &= 1 - \frac{Cov_z^{NS}(\mathbf{u}_i, \mathbf{v}_i)}{\sigma(\mathbf{u}_i) \cdot \sigma(\mathbf{v}_i)} \end{aligned}$$

From this is clear that the non-stationary factor of the covariance that scaled the relative standardized variogram is the non-stationary standard deviation of the  $\mathbf{u}_i$  and  $\mathbf{v}_i$ ; this standard deviation is a function of the distance to the boundary. Reordering the terms, we can obtained an expression for the non-stationary covariance model component:

$$\begin{aligned} \text{Cov}_Z^{\text{NS}}(\mathbf{u}_i, \mathbf{v}_i) &= E\{Z(\mathbf{u}_i) \cdot Z(\mathbf{v}_i)\} - (m_{kp}(\mathbf{v}_i) + m_k) \cdot (m_{kp}(\mathbf{u}_i) + m_k) \\ &= (1 - \hat{\gamma}_{kp}(\mathbf{u}_i, \mathbf{v}_i)) \cdot (\sigma_{kp}(\mathbf{u}_i) + \sigma_k) \cdot (\sigma_{kp}(\mathbf{v}_i) + \sigma_k) \end{aligned} \quad (3.6)$$

Since the shape, anisotropies and nugget effect of the relative standardize variogram are inputs from the modeler, the only parameter that remains to establish the non-stationary covariance model, is the range.

### *Optimization of the Statistical Parameters*

The statistical parameters such as the mean, variance and covariance are crucial for estimation and simulation. We must fit the distribution of the random variable  $Z(\mathbf{u})$  at the boundary zone, knowing the stationary, precedence rules and maximum distances of influence.

Although we know that the non-stationary behavior is a function of the distance of the sample to the boundary, there are several possible analytical expressions that fit the distribution of the non-stationary random variable. For the purpose of showing the proposed methodology we will consider that the non-stationary component of mean and variance follow a linear function (Figure 3.6) of the distance to the boundary ( $d_{pk}$ ). In this scenario, the optimization of the parameter  $m_{kp}$  and  $\sigma_{kp}^2$  will be equivalent to optimizing estimates of the intercepts:  $a_{kp}$  and  $b_{kp}$ .

### *Mean Optimization*

The mean  $m_{kp}$  is optimized given that  $m_k$  is known from the experimental average of data within rock type  $k$ , outside any boundary zone. Using Equation 3.2, the objective function is written:

$$O_m = \sum_{k=1}^K \sum_{p=1}^P \sum_{i=1}^{N_{kp}} (z(\mathbf{u}_i) - (\hat{m}_k + m_{kp}(\mathbf{u}_i)))^2$$

where  $z(\mathbf{u}_i)$  is the outcome value at every location  $\mathbf{u}_i \in \mathbf{BZ}_{pk}$ ,  $\hat{m}_k$  is the experimental average (Equation 3.7) and  $m_{kp}(\mathbf{u}_i)$  is the non-stationary mean at location  $\mathbf{u}_i$  derived from Equation 3.8.  $N_{kp}$  is the total number of data in  $\mathbf{BZ}_{pk}$ . The experimental mean of rock type  $k$  is calculated as:

$$\hat{m}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} w(\mathbf{u}_i) \cdot z_k(\mathbf{u}_i) \quad \forall \mathbf{u}_i \in \mathbf{RT}_k, \notin \mathbf{BZ}_{kp}; \forall p \quad (3.7)$$

where  $N_k$  is all data within the internal stationary portion and  $w(\mathbf{u}_i)$  are the declustering weights associated to the location  $\mathbf{u}_i$ .

Although the mean of rock type  $k$  should include all samples in this geological unit, the samples in any boundary zone are excluded since its non-stationary component is yet to be determined.

In Equation 3.7  $N_k$  denotes the number of data within the internal stationary portion of rock type  $k$ , that is, outside any boundary zone with any adjacent rock type  $p$ .

$$m_{kp}(\mathbf{u}_i) = \begin{cases} \frac{(dmax_{kp} - d_{kp}(\mathbf{u}_i))}{dmax_{kp}} \cdot a_{kp} & \text{for } 0 \leq d_{kp}(\mathbf{u}_i) \leq dmax_{kp} \\ \frac{(dmax_{pk} - d_{pk}(\mathbf{u}_i))}{dmax_{pk}} \cdot a_{kp} & \text{for } 0 \leq d_{pk}(\mathbf{u}_i) \leq dmax_{pk} \\ 0 & \text{for } d_{kp}(\mathbf{u}_i) \geq dmax_{kp} \text{ and } d_{pk}(\mathbf{u}_i) \geq dmax_{pk} \end{cases} \quad (3.8)$$

From Equation 3.8 it is clear that the optimization of  $m_{kp}$ , amounts to the optimization of  $a_{kp}$ , given  $dmax_{kp}$  and  $dmax_{pk}$ , then the optimization of the mean can be achieved by iteratively modified  $a_{kp} \forall k, p$ , in a random fashion while accepting all changes in  $a_{kp}$  that reduce objective function. This is a simplified version of the simulated annealing formalism, where only favorable perturbations are accepted (Deutsch, 2002). For our implementation we will assume that  $a_{kp} = a_{pk}$ , but this hypothesis can easily be changed if a discontinuity at the boundary is clear.

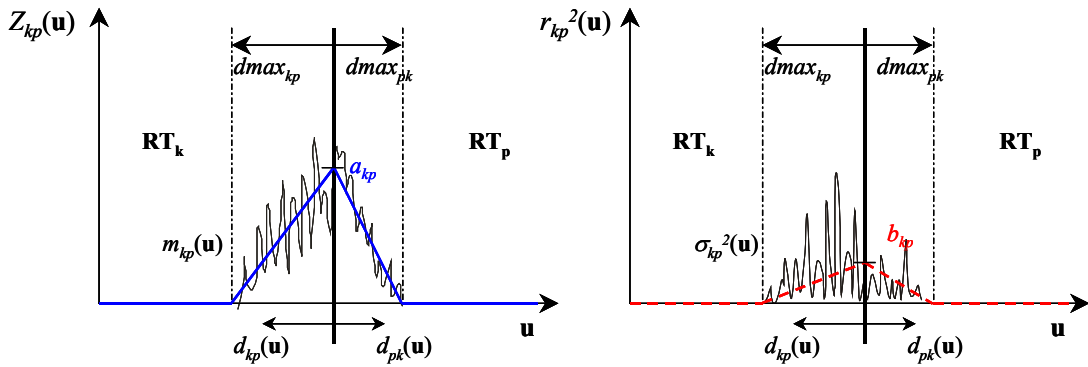


Figure 3.6: Mean and variance of the random variable  $Z_{kp}(\mathbf{u})$ , modeled by a linear function of the distance to the boundary.

### Variance Optimization

Since the variance is a statistic of second order that depends on the mean, once the optimum of  $m_{kp}$  is found, we can proceed to optimize  $\sigma_{kp}^2$  assuming the  $\sigma_k^2, k=1, \dots, K$  values are known from the experimental variance of data within the internal stationary portion of rock type  $k$ .

From Equation 3.4 the optimum  $\sigma_{kp}^2$ , will be the one that minimizes the following objective function:

$$O_{\sigma^2} = \sum_{k=1}^K \sum_{p=1}^P \sum_{i=1}^{N_{kp}} (r(\mathbf{u}_i)^2 - (\hat{\sigma}_k^2 + \sigma_{kp}^2(\mathbf{u}_i)))^2 \quad (3.9)$$

where  $r(\mathbf{u}_i)$  is the residual value at every location  $\mathbf{u}_i \in \mathbf{BZ}_{pk}$ , that is,  $r(\mathbf{u}_i) = z(\mathbf{u}_i) - (m_k + m_{kp}(\mathbf{u}_i))$ .  $\hat{\sigma}_k^2$  is the experimental variance (Equation 3.10), and  $\sigma_{kp}^2(\mathbf{u}_i)$  is the non-stationary variance at location  $\mathbf{u}_i$  derived from Equation 3.11.  $N_{kp}$  is total number of data in  $\mathbf{BZ}_{pk}$ . The experimental variance of rock type  $k$  is calculated as:

$$\hat{\sigma}_k^2 = \frac{1}{N_k} \sum_{i=1}^{N_k} w(\mathbf{u}_i) \cdot r_k(\mathbf{u}_i)^2 \quad \forall \mathbf{u}_i \in \mathbf{RT}_k, \notin \mathbf{BZ}_{kp}; \forall p \quad (3.10)$$

where  $N_k$  is all data within the internal stationary portion and  $w(\mathbf{u}_i)$  are the declustering weights associated to the location  $\mathbf{u}_i$ .

As with the mean, the optimization of  $\sigma_{kp}^2$ , is the same as the optimization of  $b_{kp}$ , given  $dmax_{kp}$  and  $dmax_{pk}$ . We assume that  $b_{kp} = b_{pk}$ .

$$\sigma_{kp}^2(\mathbf{u}_i) = \begin{cases} \frac{(dmax_{kp} - d_{kp}(\mathbf{u}_i))}{dmax_{kp}} \cdot b_{kp} & \text{for } 0 \leq d_{kp}(\mathbf{u}_i) \leq dmax_{kp} \\ \frac{(dmax_{pk} - d_{pk}(\mathbf{u}_i))}{dmax_{pk}} \cdot b_{kp} & \text{for } 0 \leq d_{pk}(\mathbf{u}_i) \leq dmax_{pk} \\ 0 & \text{for } d_{kp}(\mathbf{u}_i) \geq dmax_{kp} \text{ and } d_{pk}(\mathbf{u}_i) \geq dmax_{pk} \end{cases} \quad (3.11)$$

### Covariance Optimization

Find the covariance model of a boundary zone is equivalent to optimized the range of the relative standardize variogram that represent the stationary shape which is scaled by the standard deviations of the paired samples (Figure 3.7). This assumption provides some advantages over a full optimization algorithm to find the non-stationary covariance structure. Simplicity and fewer artifacts are the main advantages.

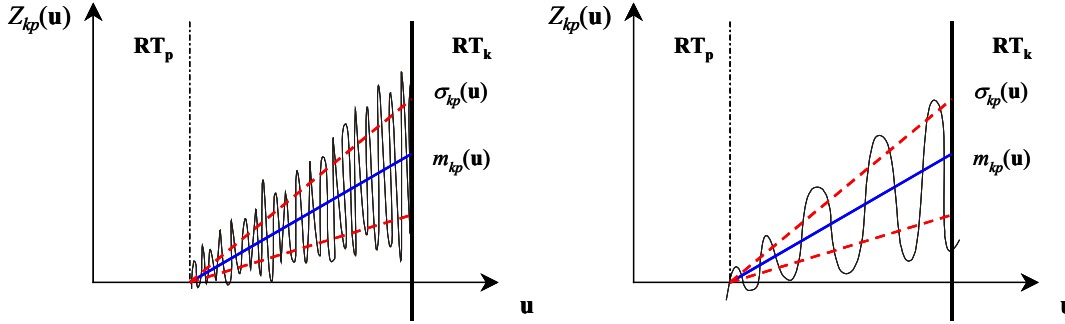


Figure 3.7: The non-stationary covariance of  $Z_{kp}(\mathbf{u})$  is defined by its non-stationary mean and variance and the shape of the correlation, which for the same mean and variance can be different as represented by the outcomes profile. The range of correlation is higher in the right.

To find the optimum parameter, in our case the total range, we need to minimize the following objective function:

$$O_{Cov} = \sum_{i=1}^N \left[ \hat{C}(z(\mathbf{u}_i), z(\mathbf{v}_i)) - C_{MOD}(z(\mathbf{u}_i), z(\mathbf{v}_i)) \right]^2$$

where  $\hat{C}$  denote the experimental covariance of the pair located at  $\mathbf{u}_i$  and  $\mathbf{v}_i$ , which is just the multiplication of the two residual values:

$$\hat{C}(z(\mathbf{u}_i), z(\mathbf{v}_i)) = r(\mathbf{u}_i) \cdot r(\mathbf{v}_i)$$

and  $C_{MOD}$  the modeled boundary covariance, which is the sum of the stationary and non-stationary component.

The squared difference between the experimental and modeled covariance is evaluated for randomly changed range that is iteratively modified until this difference is minimized. Since the anisotropies are fixed during the optimization process the ratios between the major and minor horizontal ranges and the major and vertical ranges are known. Consequently, we only need to find the total range.

More parameters could be optimized in the same fashion modifying the optimization algorithm, although we do not recommend a full optimization, since this could lead to data artifacts.

### *Estimation in presence of local non-stationary boundaries*

Once the mean, variance and covariance model have been determined both for the stationary and non-stationary regions of a geological model, we can proceed to estimation. We propose a non-stationary form of simple kriging. The basic linear regression estimator of kriging is:

$$z^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{\alpha=1}^n \lambda_{\alpha}(\mathbf{u}) \cdot [z(\mathbf{u}_{\alpha}) - m(\mathbf{u}_{\alpha})] \quad (3.12)$$

where  $z^*(\mathbf{u})$  is the estimate at unsampled location  $\mathbf{u}$ ,  $m(\mathbf{u})$  is the mean value at location  $\mathbf{u}$ ,  $\lambda_{\alpha}(\mathbf{u})$  is the weight assigned to datum  $z(\mathbf{u}_{\alpha})$ ,  $n$  are the closest data to the location  $\mathbf{u}$  being estimated, and  $m(\mathbf{u}_{\alpha})$  are the  $n$  mean values at the data locations. In this case, the expected values for the mean at the location being estimated and at each data location, will be calculated from Equation 3.3, that is, if any of the closest data or the location being estimated is within the zone of influence of a boundary, its mean will include a non-stationary component that will be calculated from a function of the distance to the boundary.

To find the optimal weights  $\lambda_{\alpha}(\mathbf{u})$ ,  $\alpha=1, \dots, n$  the following matrix system (the kriging equations) must be solved:

$$\sum_{\beta=1}^n \lambda_{\beta}(\mathbf{u}) \cdot \text{Cov}(\mathbf{u}_{\alpha}, \mathbf{u}_{\beta}) = \text{Cov}(\mathbf{u}, \mathbf{u}_{\alpha}) \quad \text{with } \alpha, \beta = 1, \dots, n$$

where  $\lambda_{\alpha}(\mathbf{u})$ ,  $\alpha=1, \dots, n$  are the simple kriging weights,  $\text{Cov}(\mathbf{u}_{\alpha}, \mathbf{u}_{\beta})$ ,  $\alpha, \beta=1, \dots, n$  correspond to the data-to-data covariances, and  $\text{Cov}(\mathbf{u}, \mathbf{u}_{\alpha})$ ,  $\alpha=1, \dots, n$  are the data-to-estimate covariances, which in matrix notation is equivalent to:

$$\begin{bmatrix} \text{Cov}(\mathbf{u}_1, \mathbf{u}_1) & \cdots & \text{Cov}(\mathbf{u}_1, \mathbf{u}_n) \\ \vdots & \ddots & \vdots \\ \text{Cov}(\mathbf{u}_n, \mathbf{u}_1) & \cdots & \text{Cov}(\mathbf{u}_n, \mathbf{u}_n) \end{bmatrix} \cdot \begin{bmatrix} \lambda_1(\mathbf{u}) \\ \vdots \\ \lambda_n(\mathbf{u}) \end{bmatrix} = \begin{bmatrix} \text{Cov}(\mathbf{u}, \mathbf{u}_1) \\ \vdots \\ \text{Cov}(\mathbf{u}, \mathbf{u}_n) \end{bmatrix} \quad (3.13)$$

The kriging weights are obtained by multiplying the inverse of the data-to-data covariance matrix by the vector of data-to-estimate covariance. This system have a solution and is unique, if the covariance model,  $\text{Cov}_{\mathbf{Z}}(\mathbf{h})$  is a legitimate model, that is, was fit with a positive definite function, and data are not collocated.

For a geological model with local non-stationary boundaries, the terms of the data covariance matrix and the vector of data-to-estimate covariances are obtained combining the stationary and non-stationary covariance model components as in Equation 3.5. The stationary covariance model is obtained from fitting the experimental variogram calculated with the data within the stationary regions within each rock type, while the non-stationary covariance component is calculated from Equation 3.6, using the correlation spatial model obtained from the optimization process describe in the previous section. The methodology and calculations are shown in more detail in an example in the next section.

Simple kriging estimator is unbiased and provides the minimum estimation variance estimate,

$$\sigma_E^2(\mathbf{u}) = Cov(0) - \sum_{\alpha=1}^n \lambda_{\alpha}(\mathbf{u}) \cdot Cov(\mathbf{u}, \mathbf{u}_{\alpha}) \quad (3.14)$$

where  $Cov(0)$  is the variance, which in our case has a stationary and a non-stationary component. Kriging also accounts for redundancy between the data and closeness of the data to where we are estimating. Spatial continuity is provided by the variogram model. The greatest disadvantage of kriging is that the estimates are smooth and the joint variability of the kriged estimates is incorrect.

The amount of missing variability is the kriging variance (Equation 3.14). Sequential Gaussian simulation overcomes this problem by adding back this missing variability, as it adds a random residual to the estimate, drawn from a normal distribution with zero mean and variance equals to the kriging variance. In this case described in this thesis, the estimation variance has also a non-stationary component that makes the implementation of sequential Gaussian simulation in the presence of local non-stationary boundaries delicate. This implementation is part of the proposed future work.

## 4. Implementation Details

### RT model and boundaries

The first step in order to perform any calculations and estimation in a geological model where local non-stationary boundaries are identified is the correct labeling of every single location with the corresponding rock type and closest precedent boundary zone and its distance to the boundary. To help with this task a small program coded in FORTRAN called `boundmod` was created.

The input data is a grid-type file of the geological model with the same format as the one used in GSLIB (Deutsch and Journel, 1998), the number of rock types available in the model, the matrix with all maximum distances of influence between rock types, and the set of precedence rules. The output file is a grid file as well, with 3 variables: the rock type (from input), the boundary zone identified by the rock type code of the precedent influencing rock type at that location or a default value if is beyond any maximum distance of influence (internal stationary regions), and the distance to the boundary (measured from the center of the node) with the precedent rock type.

In summary, the program loops over all nodes of the geological model grid, and checks within a search window if different rock types exist within the surrounding nodes. The cell may be assigned with a neighbor rock type if a boundary zone is defined for the two codes and the precedence rules are met. A more detailed structure of the program is shown in the flow chart of Figure 4.1.

The parameter file to run this program (Figure 4.2) has 4 major parts. The first one consist of four lines: the name of the file of the geological model, the variable number that contain the rock type codes at each node, the grid definition and the number of different rock types. In order for this and the following programs to work properly, the rock types codes must be correlative integers, starting at 1. The grid definition is given by the number of nodes, coordinate of the origin (locate at the center of the cell) and size of the cell for each axis: X, Y and Z.

The second part corresponds to the matrix of maximum distances of influence and it will have as many lines as different rock types in the model. Each line corresponds to the maximum distance of influence of that rock type over all others. For example in, the parameter file shown in Figure 4.2, the maximum distance of influence of rock type 2 into rock type 1, is 10 meters, i.e.,  $dmax_{21}=10$ . The maximum distance of influence of **RT**<sub>2</sub> with itself is denoted by, -1.0, a default number. If two rock types are not in contact with each other or a 'hard' boundary between them is assumed, the maximum distance of influence is set to zero, as in  $dmax_{23}=0.0$ .



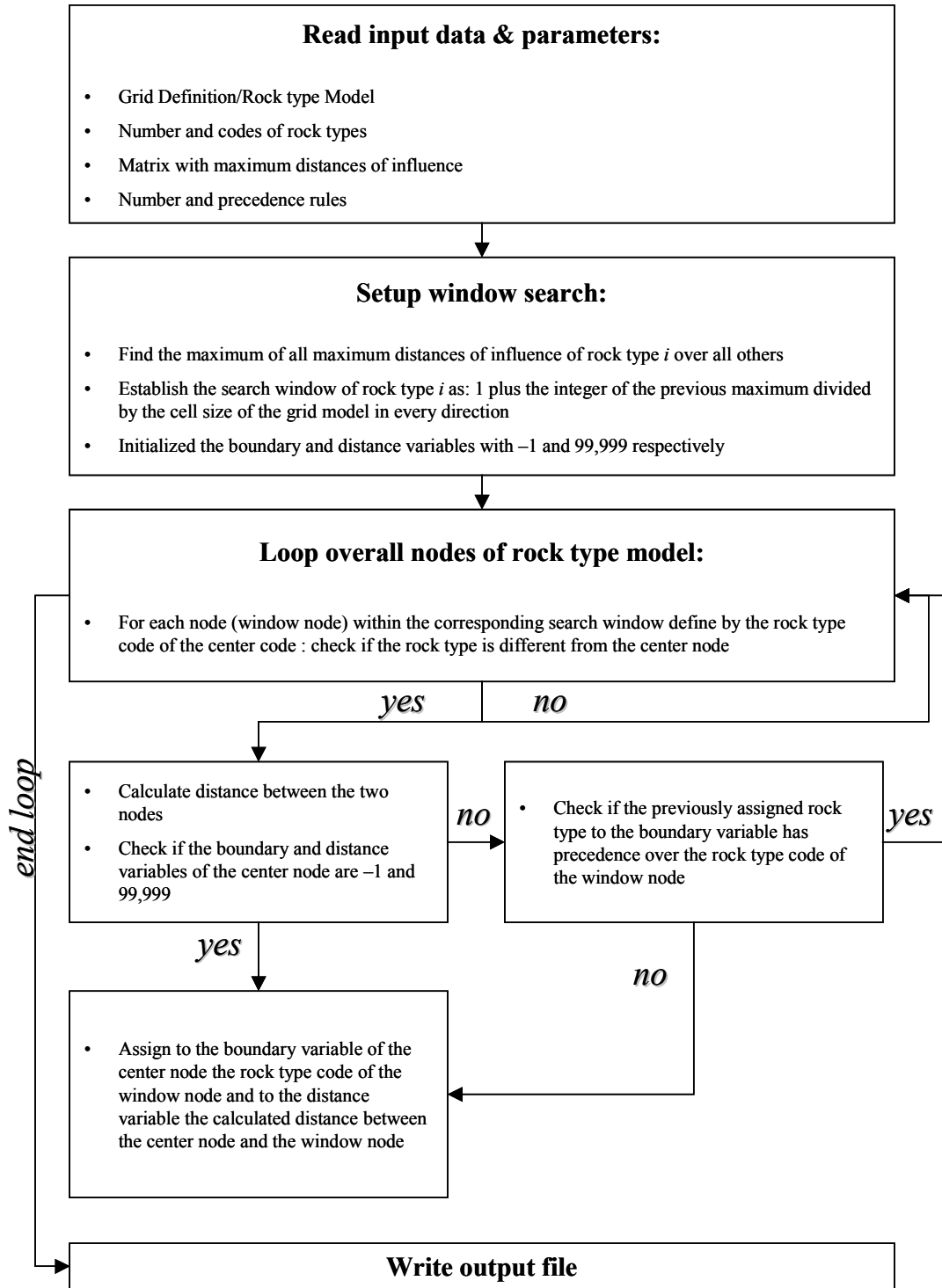


Figure 4.1: Flow chart of the program boundmod.

The third part of the parameter file corresponds to specification of the precedence rules: the number of them and the rules expressed as pairs of rock type codes. The order of the pairs is from oldest to youngest, or from less to more precedent, and is

given by: “rock type code – boundary code”. In the program, at each grid node, the array with all precedence rules is checked in order, for both pairs: the assigned boundary code in a previous iteration and the rock type code plus the possible boundary value obtained from the window node in the current iteration. If the pair is found within the list of pairs, the correlative number of the rule is assigned to the pair. By default at the beginning of this check both pairs are assigned with a correlative number of zero. The boundary code obtained from the window node overwrite a previous value of the boundary variable only if the rule number assigned to this pair is greater than the rule number of the previously assigned pair. In the parameter file shown in the example, the pair 2-4, rule number 4, is precedent over all previous pairs in the list: 3-1, 4-1 and 1-4. Leaving the number of precedence rules to zero is equivalent to choose a neutral arrangement where the precedent rock type is chosen as the one with the minimum distance to a boundary.

Finally the last line and part of the parameter file correspond to the name that will be given to the output file of the program.

```

Parameters for BOUNDMOD
*****

START OF PARAMETERS:
../data/rtnmodel.out      -input file with rock type model
1                          -column with RT value
100  0.5  1.0             -nx,xmn,xsiz
100  0.5  1.0             -ny,ymn,ysiz
1    0.0  1.0             -nz,zmn,zsiz
3                          -number of rock types (correlative starting at 1)
-1.0 50.0 20.0            -maximum distance of influence of RT 1 into RT j
10.0 -1.0 50.0            -maximum distance of influence of RT 2 into RT j
20.0 30.0 -1.0            -maximum distance of influence of RT 3 into RT j
2                          -number of precedence rules
1 3                        -pair RTi,RTj, with the oldest precedence
3 1                        -pair RTi,RTj, with the youngest precedence
output.out                -output file

```

Figure 4.2: Parameter file for “bound\_mod” program.

The following example illustrates the results of this program for a synthetic rock type model. The input rock type model has 4 rock types with a spatial configuration as shown in Figure 4.3, which reflects two sedimentary or volcanic strata cut by a reverse fault. The maximum distance of influence, are given in the parameter file of Figure 4.2, and reflect a broader influence across the fault than across the horizontal contact between the two stratums, with small differences between the upper and lower strata and the pendant and hanging walls of the fault.

## RT Model

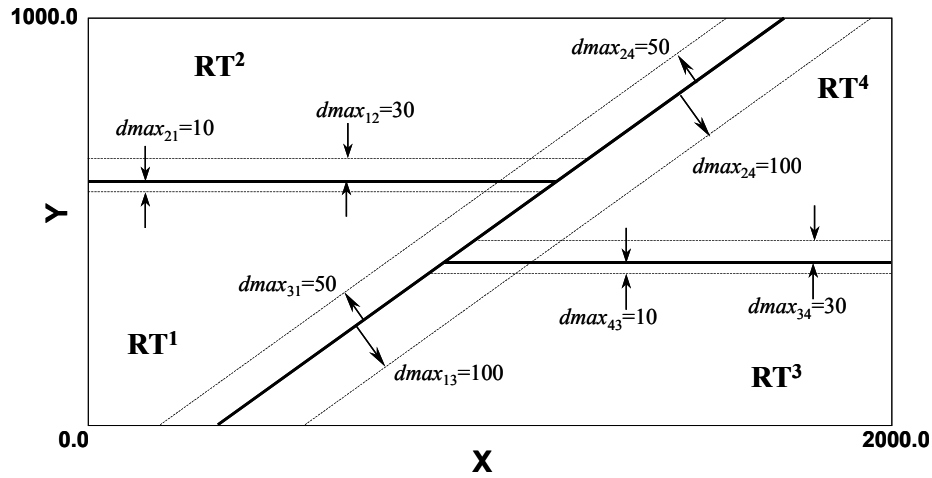


Figure 4.3: Example of a geological model with 4 rock types and maximum distances of influence define for each boundary. The influence of the fault has a larger extent compared to the influence of the stratigraphic contact.

For an exercise with 4 precedence rules (Figure 4.4), where the boundary zones generated by fault are precedent over the ones generate by the horizontal contact between layers, the program outputs for boundary zones and distance to boundary are shown in Figure 4.5. If a neutral arrangement is chosen, which is equivalent to set the number of precedence rules to zero, the boundary zone and distance to the boundary are assigned based on the closest boundary code and its distance to the node (Figure 4.6).

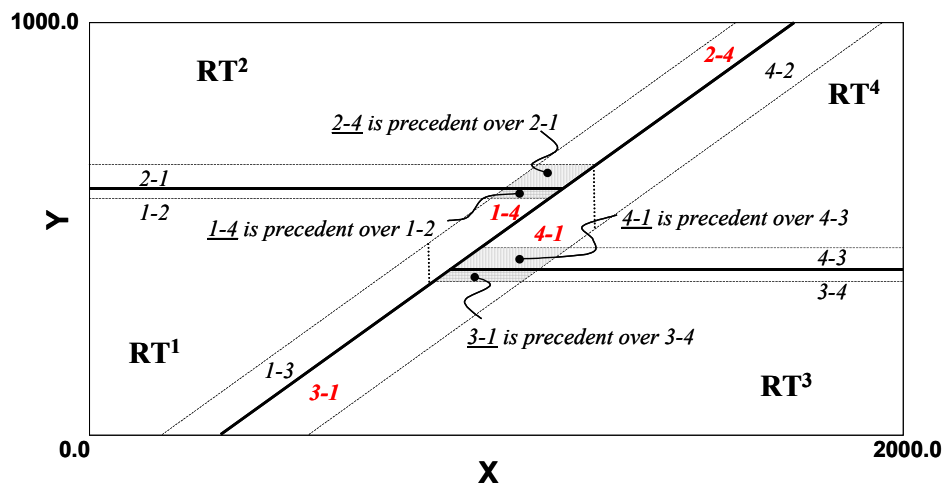


Figure 4.4: Set of four precedence rules: 3-1, 4-1, 1-4 and 2-4, for the example of Figure 4.3. With this precedence rules in the hashed regions the boundary zones define by the fault are predominant over the boundary zones defined by the horizontal contact between stratigraphic layers.

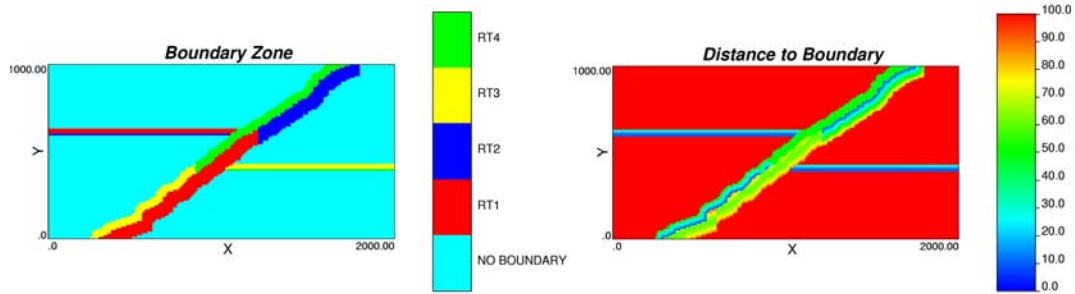


Figure 4.5: Output of the “bound\_mod” program for the example of Figure 4.3, considering a set of 4 precedence rules. The left map corresponds to the categorical variable boundary zone, a default code (NO BOUNDARY) is given to locations beyond the corresponding maximum distance of influence. The right map corresponds to the distances to the boundary.

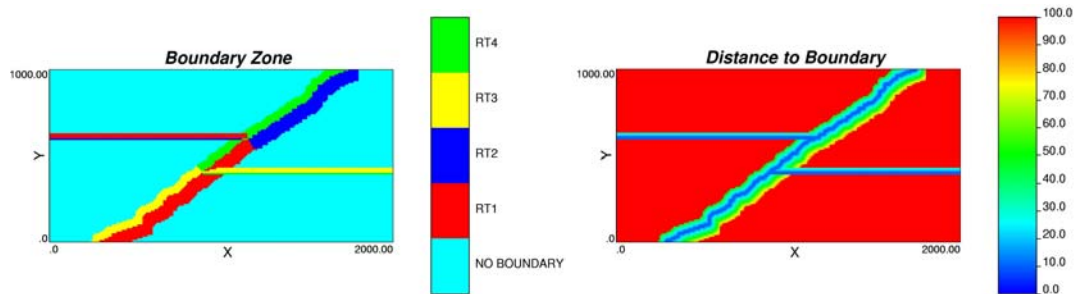


Figure 4.6: Output of the “bound\_mod” program for the example of Figure 4.3, considering a neutral arrangement. The left map corresponds to the categorical variable boundary zone. The right map corresponds to the distances to the boundary.

## Mean optimization

In order to estimate unsampled locations in a geological model, we need to calculate the mean, variance and covariance model from the available data, which we know from the previous chapter, involves the optimization of the stationary and non-stationary components. In this section we will review the implementation of the mean calculation and optimization in a FORTRAN code program called `opt_mean`.

Given a data set where every sample is tagged with the rock type, boundary zone and distance to the boundary, this program calculates the stationary component of the mean for each rock type code and the optimum intercept  $a_{kp}$  for each pair: rock type – boundary zone. The parameter  $a_{kp}$ , as shown in the previous chapter is the intercept at a distance from the boundary of zero. Since a linear fitting for the non-stationary mean has been chosen for the implementation of this methodology, the optimization of  $a_{kp}$  is equivalent to the optimization of the non-stationary mean (Equation 3.8). More complex mean surfaces can be considered by revising the code.

For this, the program first calculates the stationary means for each rock type, calculating the average over all samples that are within the internal stationary portion, that is, outside any of the boundary zones within the rock type. Then, a boundary zone and parameter  $a_{kp}$  is chosen randomly and perturbed within 20% of its original value. Using the modified value of  $a_{kp}$ , the objective function corresponding to the sum of the squared differences between the sample value and the mean for all samples is evaluated. The stationary mean of the rock type where the sample belongs plus the non-stationary component, calculated from the distance of the sample to the boundary and the parameter  $a_{kp}$  (Equation 3.8) are checked. If the objective function decreases the proposed change to  $a_{kp}$  is accepted; if not, the original value of the parameter  $a_{kp}$  from the previous iteration is restored. A flow chart with the program structure is shown in Figure 4.7.

The parameter file (Figure 4.8) consists of the name of the file with the samples tagged with rock type, boundary zone and distance to the boundary; the column number of the variable, rock type, boundary zone and distance to the boundary; trimming limits to exclude samples, either not assayed or from the upper or lower tails of the distribution. The number of rock types is also required, as well as, the matrix of maximum distances of influence as described for boundmod program. The program requires that the user enter the number of iterations. A name for the output file is also required.

### **Variance optimization**

The variance optimization is performed similarly to the mean, by a FORTRAN code program called `opt_var`. The program using the analytical expression for the non-stationary means for the different boundary zones and the corresponding stationary means, finds the stationary variances and the optimum intercept  $b_{kp}$  that defines the linear expression of the non-stationary variance for each boundary zone.

The inputs are the same as for the `opt_mean` program (Figure 4.9); a data file with the samples tagged with their corresponding rock type, boundary zone and distance to the boundary, the number of rock types and the corresponding matrix of maximum distances of influence, plus the output file from the optimization of the means with the stationary means and the intercepts for the non-stationary means.

The residual for each sample is the data value minus the mean. If the sample does not belong to a boundary zone, then the mean is just the stationary mean for the corresponding rock type. In the case where the sample belongs to a boundary zone, the mean will be the stationary mean plus the non-stationary component calculated from Equation 3.8 using the corresponding intercept obtained from the `opt_mean` program. After all residuals are calculated, the stationary variance is calculated for each rock type using all relevant samples. If there are no data for this calculation in one rock type a default value (-999) is assigned.

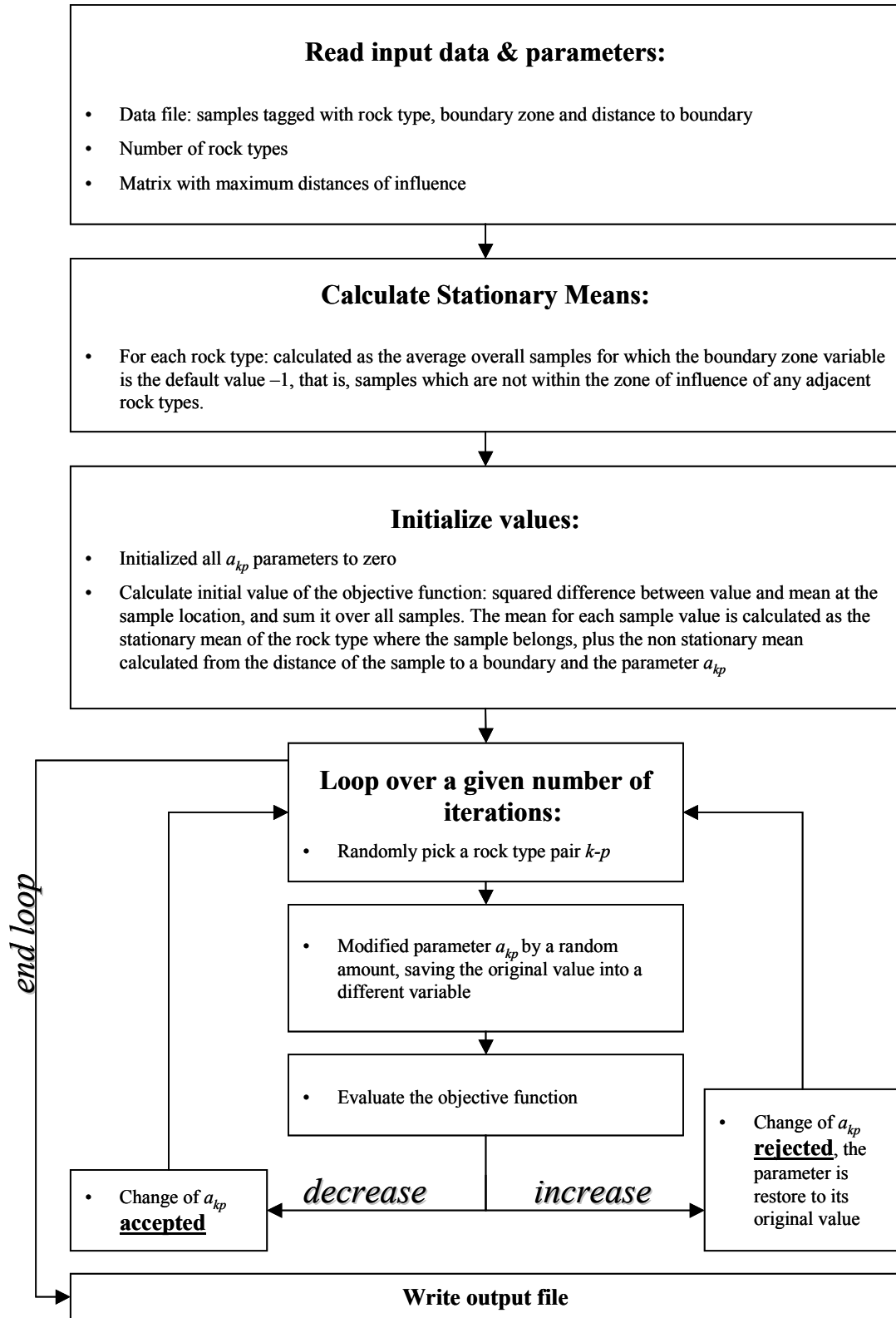


Figure 4.7: Flow chart of the program opt\_mean.

```

Parameters for OPT_MEAN
*****|

START OF PARAMETERS:
../data/data.dat          -file with data
1 2 3 4                  -columns for var,RT,Boundary RT,Distance
-1.0          1.0e21      -trimming limits
3                      -number of rock types
-1.0  50.0   20.0        -maximum distance of influence of RT 1 into RT j
10.0  -1.0   50.0        -maximum distance of influence of RT 2 into RT j
20.0  30.0   -1.0        -maximum distance of influence of RT 3 into RT j
100000                  -number of iterations
output.out              -output file

```

Figure 4.8: Parameter file for opt\_mean program.

The optimization procedure (Figure 4.10) has the same structure as the mean optimization program, that is, a boundary zone and parameter  $b_{kp}$  parameter is randomly perturbed within 20% of its previous value. The objective function (Equation 3.9) is then reevaluated, if its value decreases the proposed change is accepted; if not, the original value of the parameter  $b_{kp}$  is restored. The optimization of the non-stationary variance is done through the optimization of  $b_{kp}$  (Equation 3.11).

The parameter file (Figure 4.9) has the same entries described in the previous section for the opt\_mean program, plus the entry for the output file of this program with the stationary means and intercepts for each rock type and boundary zones.

```

Parameters for OPT_VAR
*****

START OF PARAMETERS:
../data/data.dat          -file with data
1 2 3 4                  -columns for var,RT,Boundary RT,Distance
-1.0          1.0e21      -trimming limits
3                      -number of rock types
../data/mean.dat          -file with means by RT and Boundary RT
1 2 3 4                  -columns for RT,Boundary RT,stationary mean, intercept
-1.0  50.0   20.0        -maximum distance of influence of RT 1 into RT j
10.0  -1.0   50.0        -maximum distance of influence of RT 2 into RT j
20.0  30.0   -1.0        -maximum distance of influence of RT 3 into RT j
100000                  -number of iterations
output.out              -output file

```

Figure 4.9: Parameter file for opt\_var program.

## Covariance optimization

The covariance optimization corresponds to find the optimum range of the pseudo stationary model that best fits the experimental spatial correlation of all pairs within a boundary zone. One minus the relative variogram model scaled by the standard deviation of the head and tail values corresponds to the non-stationary covariance. The FORTRAN code program that finds the optimum range is called `opt_cov`.

The inputs for this optimization are: (1) a data set tagged with the rock type, boundary zone and distance to the boundary, (2) the stationary components of mean and variance for each rock type plus the intercepts that define the non-stationary components of this statistics for each boundary zone, (3) the number of rock types and matrix of maximum distances of influences, (4) the direct variogram models of the stationary portions of each rock type, and (5) a proposed model (shape, relative nugget effect and anisotropies) for the cross variograms given by the user based on the residuals spatial correlation at the boundary. The anisotropy is specified and fixed through the ratio between the initial ranges input by the modeler. A generic parameter file is shown in Figure 4.11 as an example.

The program loops over all possible boundary zones. For each boundary zone, all pairs are found and the experimental covariance is calculated as the multiplication of the head and tail residual values. This experimental covariance is compared against the stationary plus non-stationary covariances derived from the direct variograms models and the proposed cross variogram model to be optimized. To determine the covariance value from a model given a distance, the subroutine of GSLIB `cova3` was used.

Since the modeler fixes the shape, nugget effect and anisotropies, the optimization of the non-stationary covariance model is done through the optimization of the range. The range is iteratively perturbed until the objective function corresponding to the sum over all pair squared differences between the experimental and modeled covariance of each pair, is minimized. The structure of the program is shown in the flow chart of Figure 4.12.

The `opt_cov` output file gives the optimum range for each cross variogram corresponding to the boundary zone defined by a pair of rock type  $k$  and boundary rock type  $p$ .



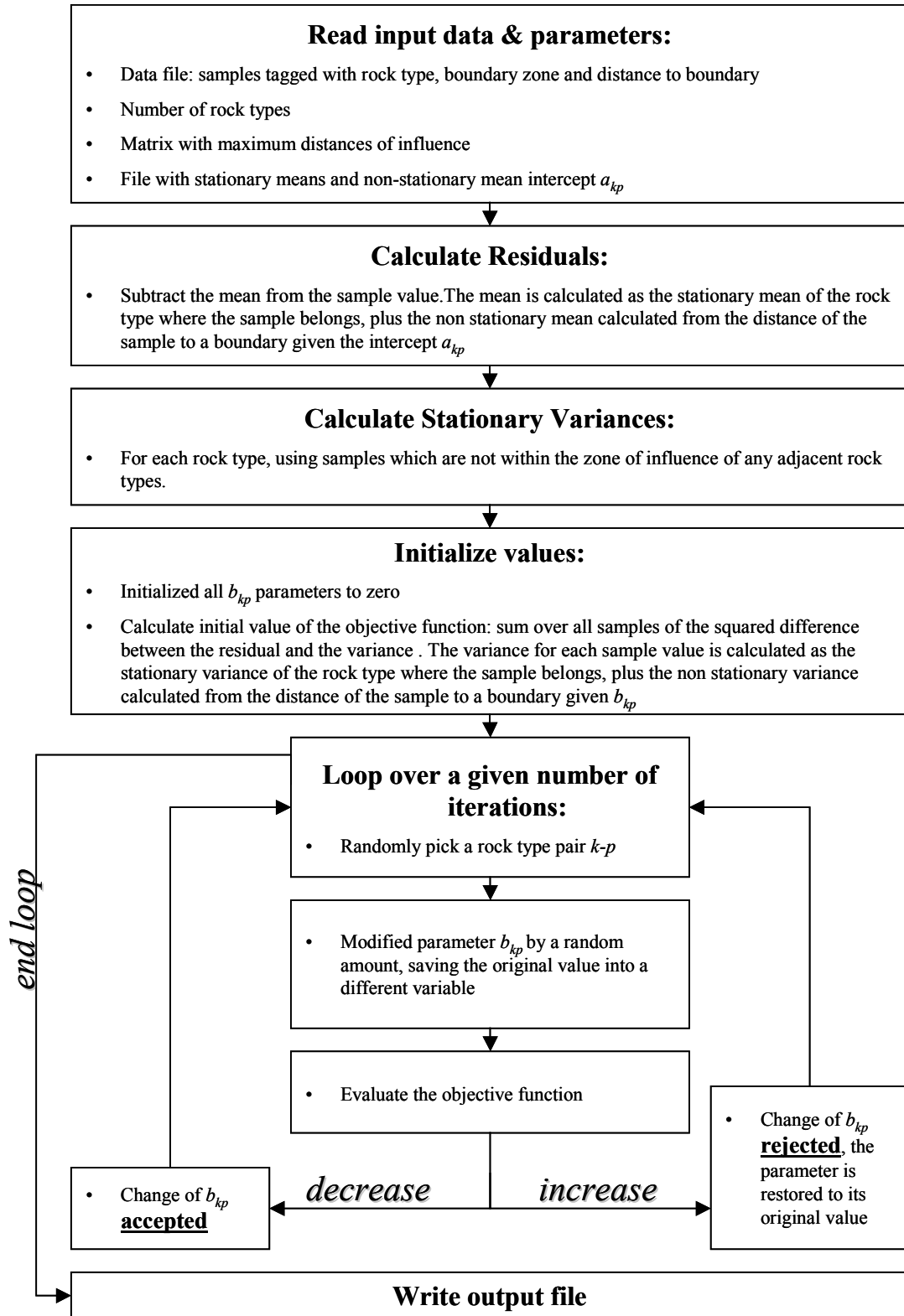


Figure 4.10: Flow chart of the program opt\_var.

```

Parameters for OPT_COV
*****

START OF PARAMETERS:
../data/data.dat          -file with data
1 2 3 4 5 6 7            -columns for X,Y,Z,var,RT,Boundary RT,Distance
-1.0                      -trimming limits
1.0e21                   -number of iterations
100000                   -output file
output.out               -number of rock types
3                         -file with means by RT and Boundary RT
../data/mean.dat          -columns for RT,Boundary RT,stationary mean,intercept
1 2 3 4                  -file with variances by RT and Boundary RT
../data/var.dat           -columns for RT,Boundary RT,stationary variance,intercept
1 2 3 4                  -maximum distance of influence of RT 1 into RT j
-1.0 50.0 20.0           -maximum distance of influence of RT 2 into RT j
10.0 -1.0 50.0           -maximum distance of influence of RT 3 into RT j
20.0 30.0 -1.0           -semivariogram for "RTi" and "RTj"
1 1                       - nst, nugget effect
1 0.1                     - it,cc,ang1,ang2,ang3
1 0.9 0.0 0.0 0.0        - a_hmax, a_hmin, a_vert
60.0 60.0 60.0           -semivariogram for "RTi" and "RTj"
1 2                       - nst, nugget effect
1 0.1                     - it,cc,ang1,ang2,ang3
1 0.4 0.0 0.0 0.0        - a_hmax, a_hmin, a_vert
60.0 60.0 60.0           -semivariogram for "RTi" and "RTj"
2 2                       - nst, nugget effect
1 0.1                     - it,cc,ang1,ang2,ang3
1 0.9 0.0 0.0 0.0        - a_hmax, a_hmin, a_vert
60.0 60.0 60.0

```

Figure 4.11: Parameter file for opt\_cov program.

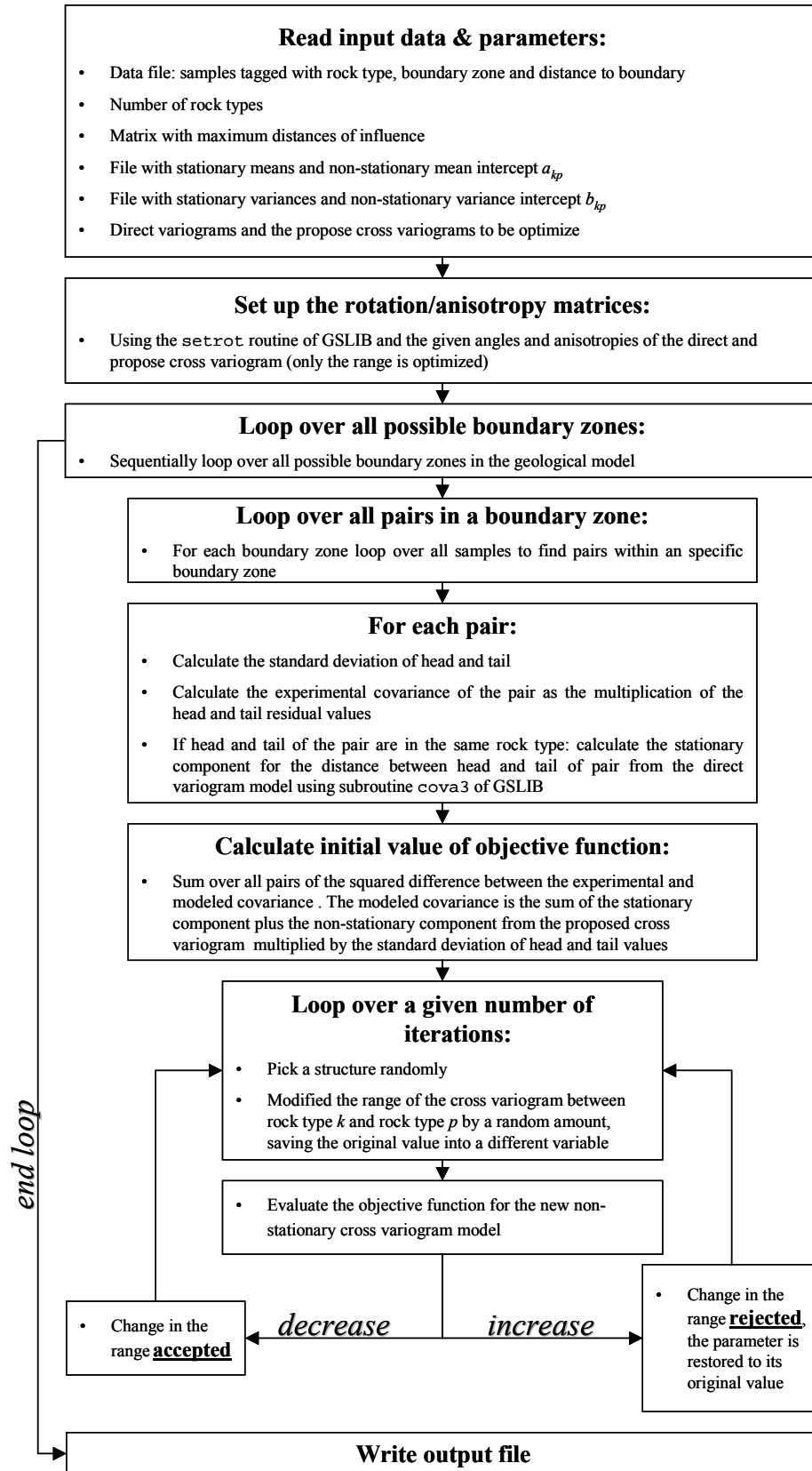


Figure 4.12: Flow chart of the program opt\_cov.

## Non-stationary kriging within boundaries

Finally, the program to perform a non-stationary version of simple kriging in the presence of soft boundaries is the FORTRAN coded program `kt3d_bound`. This program allows to estimate unsampled locations in a geological model where non-stationary boundary zones has been identified, by using a covariance model that takes into account both the stationary and non-stationary components of the conditioning data. It also considers that the mean and variance have non-stationary components at locations within the boundary zones.

The inputs to this program include a data file and rock type model tagged by the corresponding rock type, boundary zone and distance to the boundary of each location, the statistical optimized models for the mean, variance and covariance and the kriging parameters. The output is a GSLIB grid-type file with the estimates and kriging variances for all locations within the grid definition of the model.

The program (Figure 4.13) is a modified version of the GSLIB program `kt3d` (Deutsch and Journel, 1998) that performs simple kriging and/or cross validation using samples both from the geological domain of the location to estimate and from adjacent domains. The data-to-data covariance matrix and the data-to-estimate covariance vector (Equation 3.13) are fill accordingly to the spatial configuration of the samples and estimate in relation with their location outside or inside a boundary zone and their distances to the boundary (Equation 3.5). The calculated covariance between samples or with the estimate includes the non-stationary covariance component corresponding to the optimized model. A detailed scheme of how the covariance matrices and vectors are filled in this type of kriging is presented in Figure 4.14. Once the kriging system is solved, using the `ktso1` routine of GSLIB, the mean: stationary plus non-stationary component, of each sample and the location to be estimate are calculated in order to compute the solution as in Equation 3.12. To calculate the kriging variance (Equation 3.14), the block covariance is calculated for each location to be estimate, as the stationary covariance of the corresponding rock type plus the non-stationary covariance if the block is within a boundary zone.

The parameter file (Figure 4.15) was subdivided into four blocks to organize the input of data and parameters. The first block is for data and rock type model input, the second block is to define the grid to be estimate. In the third block the output files of the mean and variance optimization are specified, the direct variograms of each rock type and the cross variograms between rock types, also the matrix of maximum distances of influences can be entered following the same procedure explained for `boundmod` program. The fourth block is for the kriging parameters; the program can be run in an estimating or cross validation mode. The ellipsoid search parameters, the minimum, maximum of samples and maximum number per octant are specified in this block too.

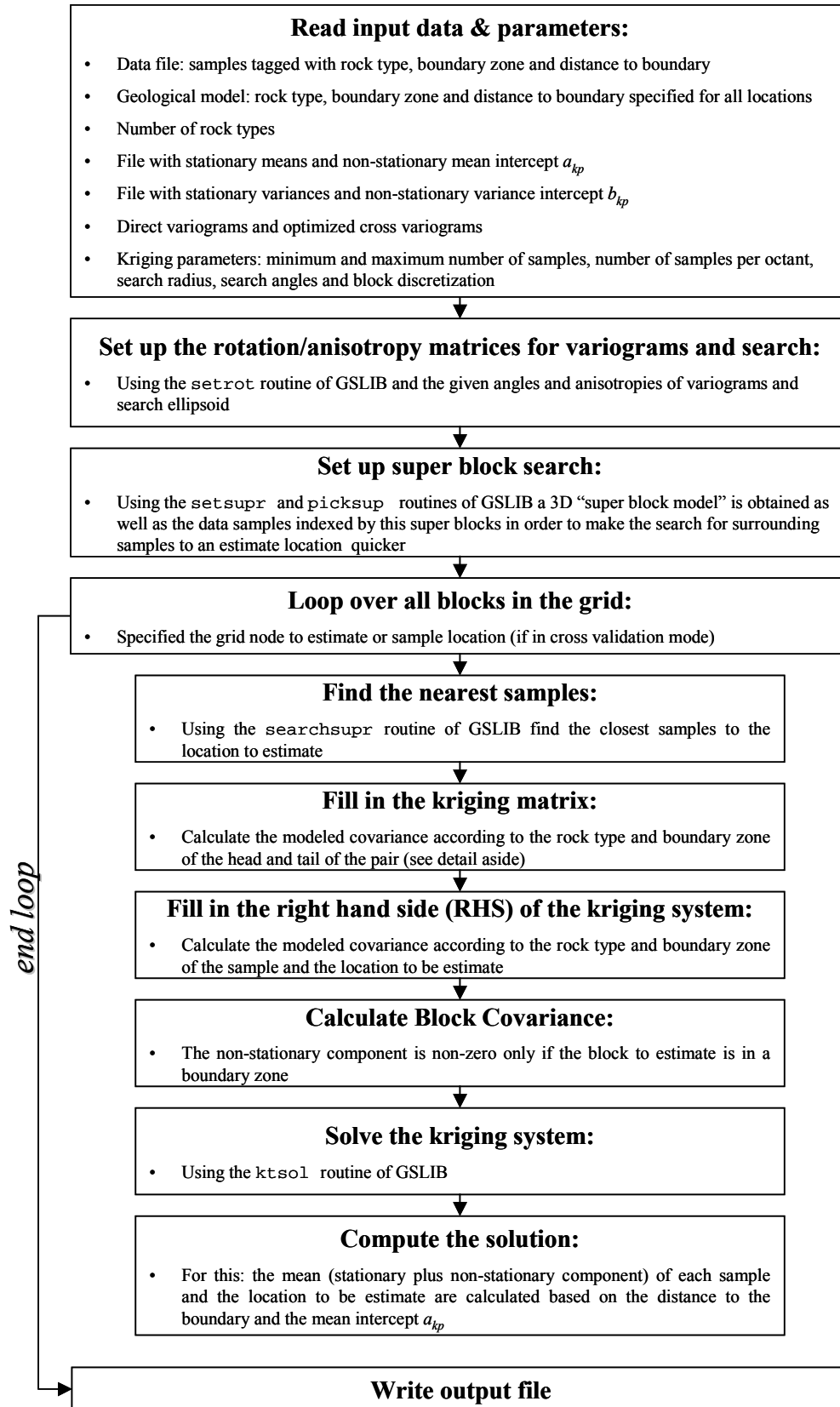


Figure 4.13: Flow chart of the program `kt3d_bound`.

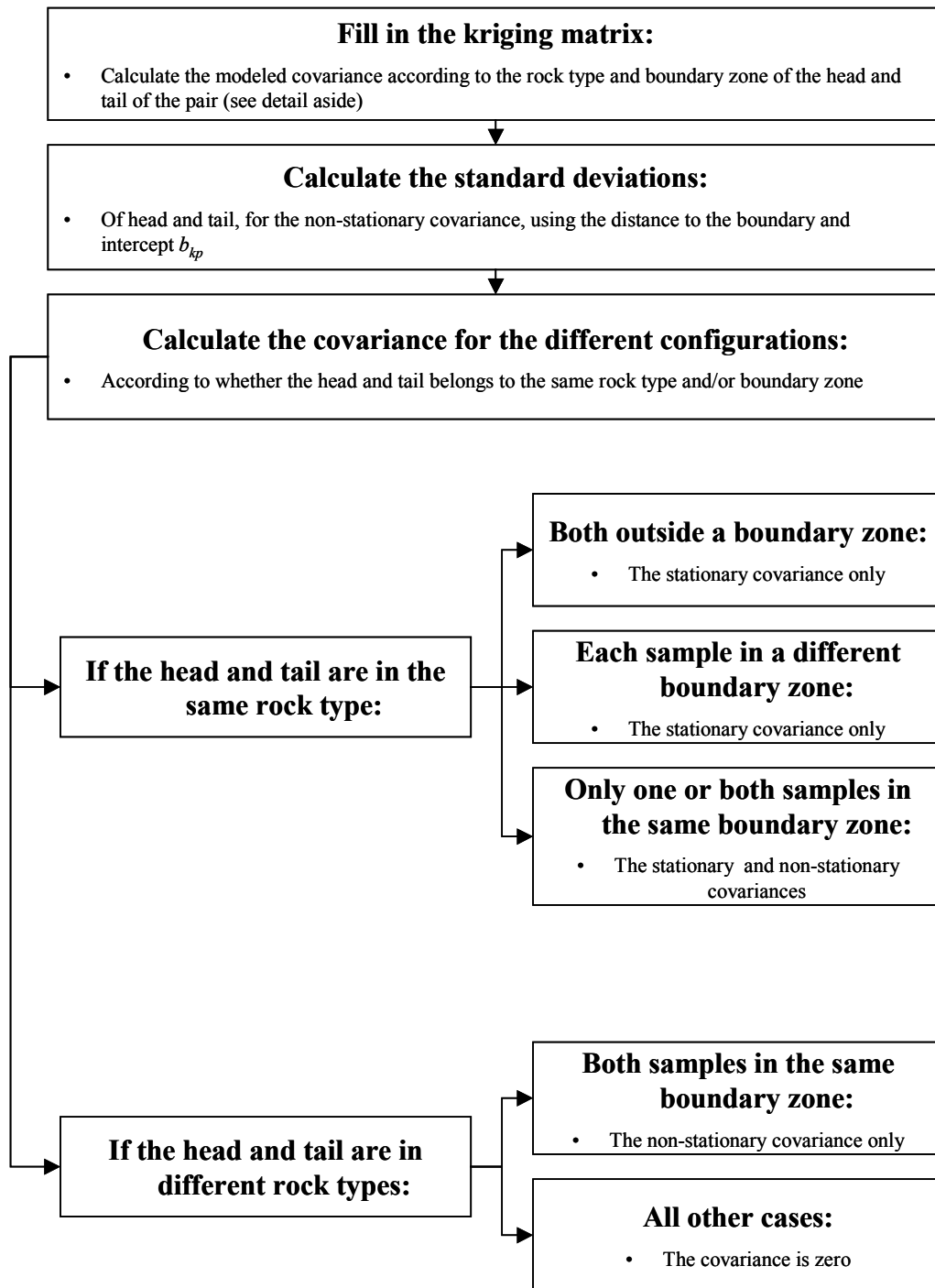


Figure 4.14: Flow chart for the filling of the kriging matrix in `kt3d_bound` program, the right hand side of the kriging system is fill in a similar way except that instead of the configurations of two samples is the configuration of the sample to location to be estimate, the one to determine how the covariance is calculated.

```

Parameters for KT3D_BOUND
*****

START OF PARAMETERS:

START OF DATA INPUT:
../data/data.dat          -file with data
0 1 2 3 4 5 6 7          -columns for DH,X,Y,Z,var,RT,Boundary RT,Distance
-1.0e21 1.0e21           -trimming limits
../data/rtmodel.out        -input file with rock type model
1 2 3                     -column with RT,Boundary RT,Distance
3                           -number of rock types

START OF GRID DEFINITION:
100 0.5 1.0               -nx,xmn,xsiz
100 0.5 1.0               -ny,ymn,ysiz
1 0.5 1.0                 -nz,zmn,zsiz

START OF STATISTICAL MODELS:
../data/mean.dat          -file with means by RT and Boundary RT
1 2 3 4                   -columns for RT,Boundary RT,stationary mean,intercept
../data/var.dat           -file with variances by RT and Boundary RT
1 2 3 4                   -columns for RT,Boundary RT,stationary variance,intercept
-1.0 50.0 20.0            -maximum distance of influence of RT 1 into RT j
10.0 -1.0 50.0            -maximum distance of influence of RT 2 into RT j
20.0 30.0 -1.0            -maximum distance of influence of RT 3 into RT j
1 1                       -semivariogram for "RTi" and "RTj"
1 0.1                     - nst, nugget effect
1 0.9 0.0 0.0 0.0         - it,cc,ang1,ang2,ang3
60.0 60.0 60.0           - a_hmax, a_hmin, a_vert
1 2                       -semivariogram for "RTi" and "RTj"
1 0.1                     - nst, nugget effect
1 0.4 0.0 0.0 0.0         - it,cc,ang1,ang2,ang3
60.0 60.0 60.0           - a_hmax, a_hmin, a_vert
2 2                       -semivariogram for "RTi" and "RTj"
1 0.1                     - nst, nugget effect
1 0.9 0.0 0.0 0.0         - it,cc,ang1,ang2,ang3
60.0 60.0 60.0           - a_hmax, a_hmin, a_vert

START OF KRIGING PARAMETERS:
0                           -option: 0=grid, 1=cross, 2=jackknife
xvk.dat                    -file with jackknife data
0 1 2 3 4 5 6 7           -columns for DH,X,Y,Z,var,RT,Boundary RT,Distance
3                           -debugging level: 0,1,2,3
kt3d.dbg                   -file for debugging output
kt3d.out                   -file for kriged output
4 8                         -min, max data for kriging
0                           -max per octant (0-> not used)
20.0 20.0 20.0             -maximum search radii
0.0 0.0 0.0               -angles for search ellipsoid

```

Figure 4.15: Parameter file for kt3d\_bound program.

## 1-D Example

To illustrate the concepts from Chapter 3 and the programs describe above will use a small 1-D synthetic example. Three independent unconditional simulations, with different variogram models where used to build the variable  $Z(\mathbf{u})$ , that will represent the metal grade across a boundary between two rock types. The variogram models were:

$$\text{SGS1} \sim \gamma(\mathbf{h}) = 1.0 \cdot Sph_{(h=15)}(\mathbf{h})$$

$$\text{SGS2} \sim \gamma(\mathbf{h}) = 1.0 \cdot Sph_{(h=30)}(\mathbf{h})$$

$$\text{SGS3} \sim \gamma(\mathbf{h}) = 1.0 \cdot Sph_{(h=10)}(\mathbf{h})$$

The first two consisted of 100 grid points spaced at 1 meter; the third consists of 40 grid points at the same spacing. The small size of the simulated area made the target statistics ( $m=0$  and  $\sigma^2=1$ ) difficult to reproduce exactly; one hundred realizations were generated and one realization was chosen with a mean and variance close to the target.

**SGS1** and **SGS2** were transformed to a non-standard normal distribution to reflect different average grade and variability across a boundary.

$$\begin{aligned}\mathbf{SGS1} &\sim N(2.0, 1.0) \\ \mathbf{SGS2} &\sim N(0.5, 0.25)\end{aligned}$$

The third simulation was transform to a non-standard normal distribution but with mean and variance as functions of distance to the boundary. The mean was assume to follow a linear function as the one describe in Equation 3.8, with a symmetric maximum distance of influence of 20 meters and an intercept  $a_{12}=2.0$  while the variance follows a linear function as well, with the same maximum distance of influence and an intercept  $b_{12}=1.0$  (Figure 4.16), that is,

$$\begin{aligned}m(\mathbf{u}_i) &= \frac{(20 - d(\mathbf{u}_i))}{20} \cdot 2 \quad \text{where } 0.5 \leq \mathbf{u}_i \leq 19.5 \\ \sigma^2(\mathbf{u}_i) &= \frac{(20 - d(\mathbf{u}_i))}{20} \cdot 1 \quad \text{where } 0.5 \leq \mathbf{u}_i \leq 19.5\end{aligned}$$

The transformed boundary values were calculated as follows:

$$Z(\mathbf{u}_i) = \sigma^2(\mathbf{u}_i) \cdot Y(\mathbf{u}_i) + m(\mathbf{u}_i)$$

where  $Y(\mathbf{u}_i)$  is the original simulated value in normal scores. This third variable will be the non-stationary component of the final random variable.

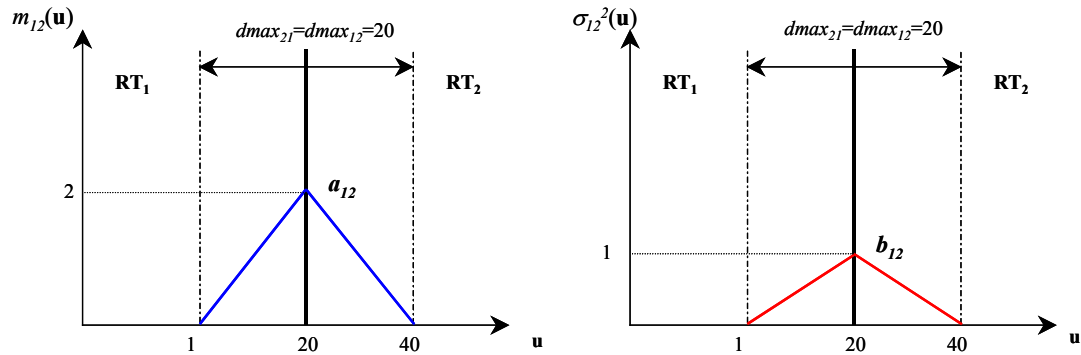


Figure 4.16: Linear functions for the mean and variance used to transform the normal score values of **SGS3**.



The final random variable was obtained by joining the two first simulations to obtain a 1-D array of 200 points and adding the third simulation to the values from locations 80.5 up to 199.5 (Figure 4.17 and 4.18).

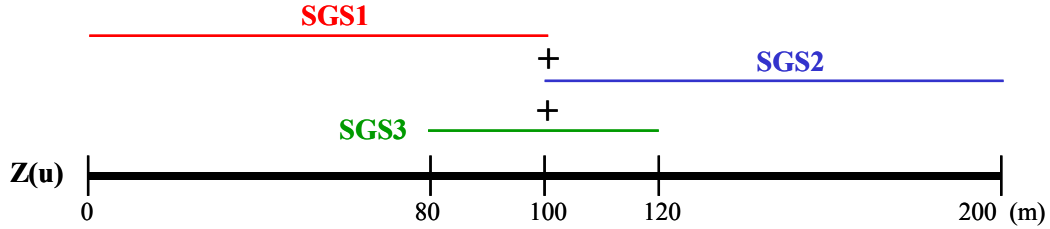


Figure 4.17: Scheme of how the three unconditional simulations are added to obtain  $Z(u)$ .

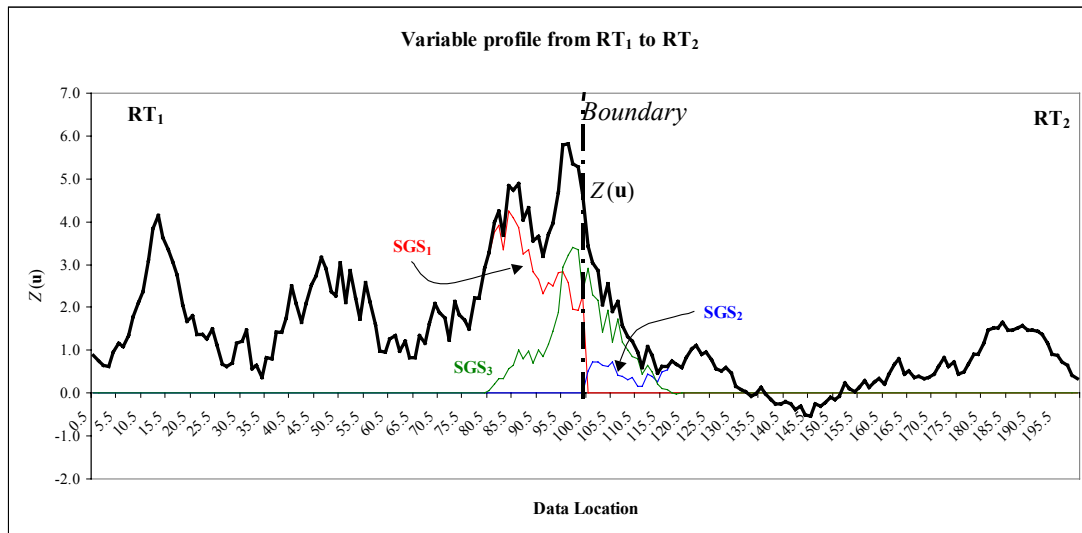


Figure 4.18: Dataset for 1-D example. Random variable profile of metal content against location along the X-coordinate. The dashed lines correspond to the underlying variables used to generate this synthetic example.

The first step in this methodology is to infer all stationary and non-stationary statistical parameters for each rock type and boundary zone. To find the stationary and non-stationary component of the mean the program `opt_mean`, was used.

The stationary mean of rock type 1, calculated with the values from location 0.5 to 79.5, is 1.75, which is lower than the mean of the full **SGS1** (2.0), because the eighty first values are relatively lower than the 20 values closest to the boundary that are not considered in the calculation of the stationary mean due to ergodic fluctuations in the simulation. The stationary mean of rock type 2 is 0.51.

For this example, after 100,000 iterations, the optimum intercept value  $a_{2I}$  is 3.22, therefore the mean of  $Z(\mathbf{u})$  is,

$$m(\mathbf{u}_i) = \begin{cases} 1.75 & \text{where } 0.5 \leq \mathbf{u}_i \leq 79.5 \\ 1.75 + \frac{(20 - d(\mathbf{u}_i))}{20} \cdot 3.22 & \text{where } 80.5 \leq \mathbf{u}_i \leq 99.5 \\ 0.51 + \frac{(20 - d(\mathbf{u}_i))}{20} \cdot 3.22 & \text{where } 100.5 \leq \mathbf{u}_i \leq 199.5 \\ 0.51 & \text{where } 120.5 \leq \mathbf{u}_i \leq 199.5 \end{cases} \quad (4.1)$$

The intercept is slightly higher than the original value of 2.0, used to create this synthetic data, but this is due again to the influence of systematically higher values in  $\mathbf{RT}_1$  near the boundary. The discontinuity in the optimum mean profile across the two rock types, after the boundary (Figure 4.19) is a consequence of the difference between the stationary means of  $\mathbf{RT}_1$  and  $\mathbf{RT}_2$ .

The output file of `opt_mean` is shown in Figure 4.20, this file gives the stationary means of rock types 1 and 2 and the intercept of the linear function that represent the non-stationary mean at the boundary zone.

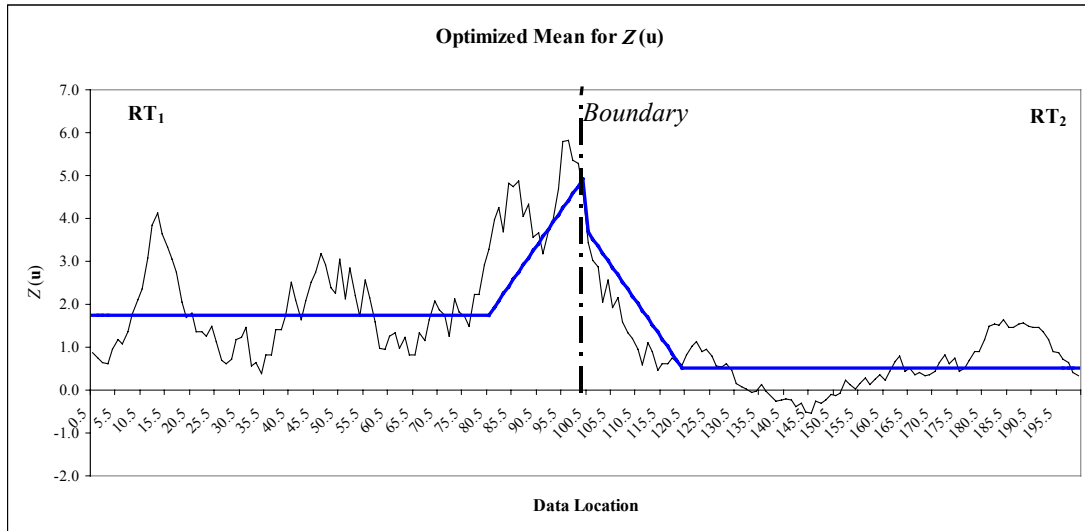


Figure 4.19: Optimized mean obtained for dataset of 1-D example.

To find the stationary and non-stationary component of the variance the program `opt_var` was used, this program use the output file of `opt_mean` as an input. The objective function in this case is the squared difference between the stationary plus the non-stationary variance and the residuals squared. The residuals are obtained using the already optimized expression for the mean (Equation 4.1). The stationary component of variance for rock type 1 is 0.71 and 0.33 for rock type 2. The optimum

intercept value of  $b_{2I}$ , in this example is attained at 0.50, thus the total variance of  $Z(\mathbf{u})$  (Figure 4.21) is,

$$\sigma^2(\mathbf{u}_i) = \begin{cases} 0.71 & \text{where } 0.5 \leq \mathbf{u}_i \leq 79.5 \\ 0.71 + \frac{(20 - d(\mathbf{u}_i))}{20} \cdot 0.50 & \text{where } 80.5 \leq \mathbf{u}_i \leq 99.5 \\ 0.33 + \frac{(20 - d(\mathbf{u}_i))}{20} \cdot 0.50 & \text{where } 100.5 \leq \mathbf{u}_i \leq 199.5 \\ 0.33 & \text{where } 120.5 \leq \mathbf{u}_i \leq 199.5 \end{cases}$$

The optimized intercept is slightly different to the one used to created this synthetic dataset, due to statistical fluctuations from the mean and variance near the boundary zone. The output file of `opt_var` is shown in Figure 4.22.

```
Optimized Means
4
RT
Boundary_RT
Stationary_Mean
NonStationary_Intercept
1 2 1.7491 3.2245
2 1 0.51318 3.2245
```

Figure 4.20: Output file of `opt_mean` program for the 1D example.

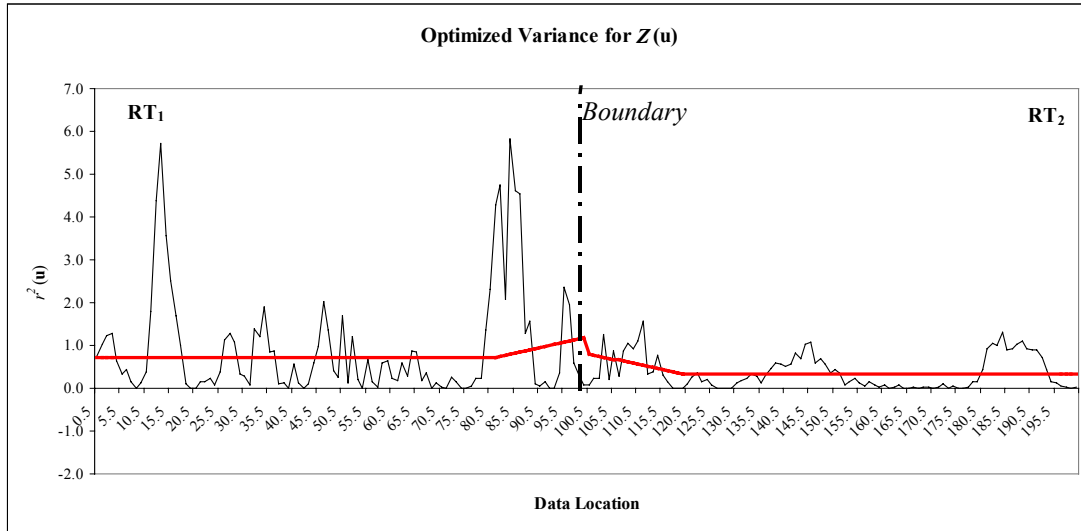


Figure 4.21: Optimized variance obtained for dataset of 1-D example.

```

Optimized Variances
4
RT
Boundary_RT
Stationary_Variance
NonStationary_Intercept
1 2 0.70569 0.49755
2 1 0.33372 0.49755

```

Figure 4.22: Output file of `opt_var` program for the 1D example.

For the estimation we also need the covariance models for both for the stationary and the non-stationary regions in the model. For the stationary regions of  $\mathbf{RT}_1$  and  $\mathbf{RT}_2$ , we will assume that the variogram models are the ones used to generate the underlying unconditional simulations. The FORTRAN code called `opt_cov` is used to find the optimum range of the non-stationary covariance structure that adjust better to the experimental covariance (Figure 4.23) calculated for each pair in which both points are within the 40 meters width zone of influence of the boundary. As explained previously, the modeler needs to input the shape and nugget effect of the relative variogram model for the non-stationary zone, while the range is optimized. In this case a spherical isotropic model with a nugget effect of 0.0 is adopted. The optimum range obtained is 6.37 (Figure 4.24), acceptably similar to the range used to build this synthetic dataset given the natural variations we have seen for mean and variance.

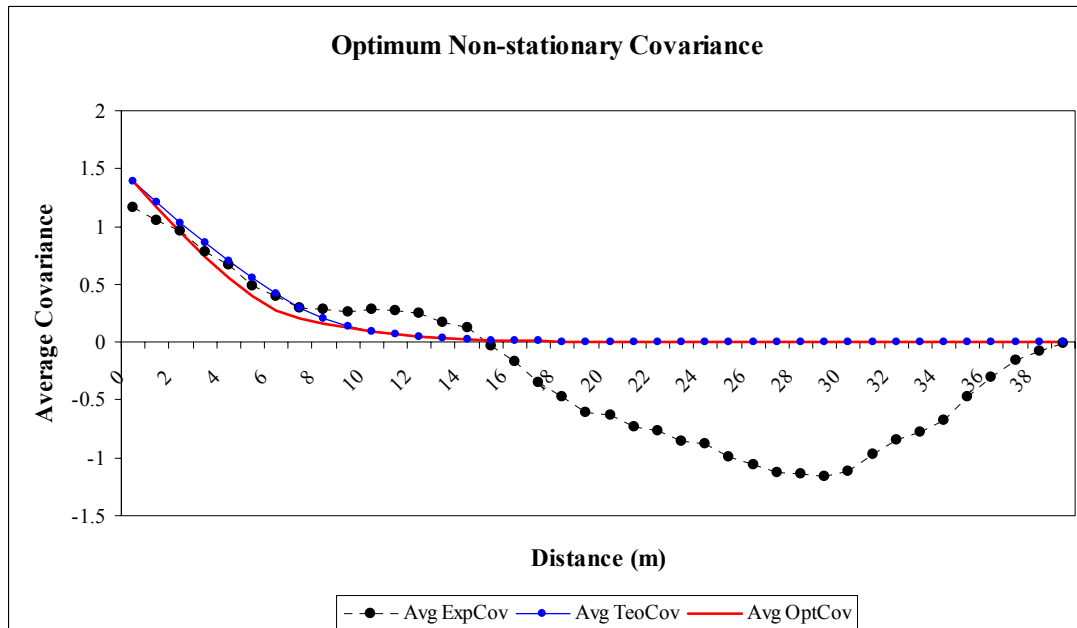


Figure 4.23: Experimental covariance from pairs within the boundary zone (black dash line), optimum non-stationary covariance obtained from `opt_cov` (red solid line) and original covariance of the non-stationary component, **SGS3**, used to build the synthetic dataset  $Z(u)$  (blue solid line).

```

Optimized Ranges
3
RT
Boundary_RT
Optimum_Range
1 2 6.369907

```

Figure 4.24: Output file of opt\_cov program for the 1D example.

The optimization convergence in this example of the mean and variance occur before the 250 iterations, while for the covariance is before 50 iterations as shown in Figure 4.25.

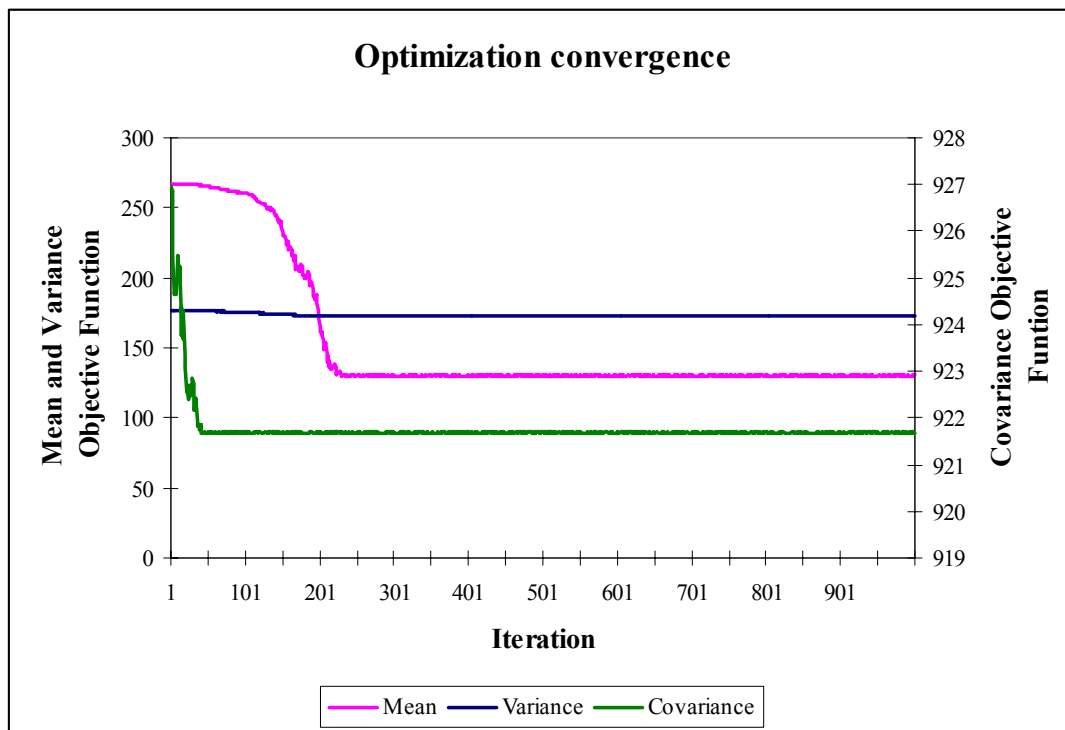


Figure 4.25: Optimization convergence of mean, variance and covariance parameters.

The estimation in the presence of soft boundaries is done using a FORTRAN program called kt3d\_bound, this program uses the mean, variance and covariance previously obtained from each optimization program. First we will show how estimation is performed with 8 surrounding data (Figure 4.26) at a single location, then we will review the results of estimation considering a conditioning dataset of one out four grid nodes from the reference.

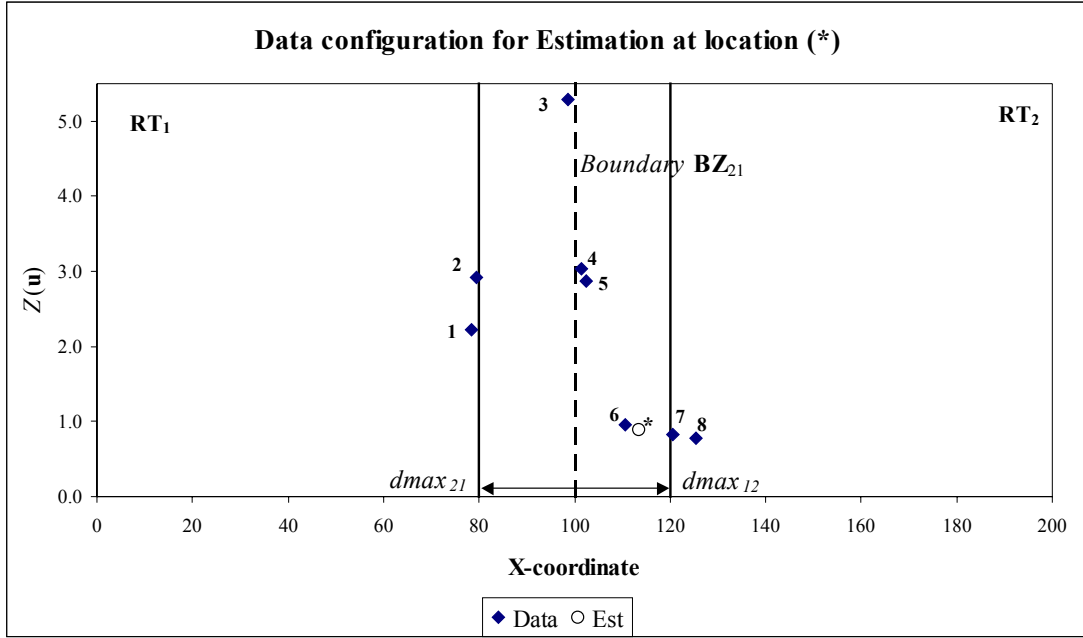


Figure 4.26: Data configuration for the estimation of an unknown location with the 1-D example.

The data covariance matrix in terms of the stationary and non-stationary component, for the example configuration is:

$$\begin{bmatrix}
 Cov(0)_{Z_1}^S & & & & & & & & \\
 Cov_{Z_1}^S & Cov(0)_{Z_1}^S & & & & & & & \\
 Cov_{Z_1+Cov_{Z_{12}}}^S & Cov_{Z_1+Cov_{Z_{12}}}^S & Cov(0)_{Z_1+Cov_{Z_{12}}}^{NS} & & & & & & \\
 0.0 & 0.0 & Cov_{Z_{12}}^{NS} & Cov(0)_{Z_{12}}^S + Cov(0)_{Z_{12}}^{NS} & & & & & \\
 0.0 & 0.0 & Cov_{Z_{12}}^{NS} & Cov_{Z_{12}}^S + Cov_{Z_{12}}^{NS} & Cov(0)_{Z_{12}}^S + Cov(0)_{Z_{12}}^{NS} & & & & \\
 0.0 & 0.0 & Cov_{Z_{12}}^{NS} & Cov_{Z_{12}}^S + Cov_{Z_{12}}^{NS} & Cov_{Z_{12}}^S + Cov_{Z_{12}}^{NS} & Cov(0)_{Z_{12}}^S + Cov(0)_{Z_{12}}^{NS} & & & \\
 0.0 & 0.0 & 0.0 & Cov_{Z_{12}}^S + Cov_{Z_{12}}^{NS} & Cov_{Z_{12}}^S + Cov_{Z_{12}}^{NS} & Cov_{Z_{12}}^S + Cov_{Z_{12}}^{NS} & Cov(0)_{Z_{12}}^S & & \\
 0.0 & 0.0 & 0.0 & Cov_{Z_{12}}^S + Cov_{Z_{12}}^{NS} & Cov_{Z_{12}}^S + Cov_{Z_{12}}^{NS} & Cov_{Z_{12}}^S + Cov_{Z_{12}}^{NS} & Cov_{Z_{12}}^S & Cov(0)_{Z_{12}}^S &
 \end{bmatrix}$$

Similarly, the data-to-estimate covariance vector in this case is:

$$\begin{bmatrix}
 0.0 \\
 0.0 \\
 Cov_{Z_{12}}^{NS} \\
 Cov_{Z_{12}}^S + Cov_{Z_{12}}^{NS} \\
 Cov_{Z_{12}}^S + Cov_{Z_{12}}^{NS} \\
 Cov_{Z_{12}}^S + Cov_{Z_{12}}^{NS} \\
 Cov_{Z_{12}}^S + Cov_{Z_{12}}^{NS} \\
 Cov_{Z_{12}}^S + Cov_{Z_{12}}^{NS}
 \end{bmatrix}$$

Completing the data covariance matrix and the data-to-estimate covariance vector, calculating the stationary and non-stationary component as shown before, the resultant kriging system is:

$$\begin{bmatrix} 1.0 & 0.9 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.9 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 2.149 & 0.382 & 0.227 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.382 & 1.027 & 0.842 & 0.141 & 0.044 & 0.014 \\ 0.0 & 0.0 & 0.227 & 0.842 & 1.003 & 0.152 & 0.052 & 0.019 \\ 0.0 & 0.0 & 0.0 & 0.141 & 0.152 & 0.805 & 0.130 & 0.078 \\ 0.0 & 0.0 & 0.0 & 0.044 & 0.052 & 0.130 & 0.25 & 0.188 \\ 0.0 & 0.0 & 0.0 & 0.014 & 0.019 & 0.078 & 0.188 & 0.25 \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \lambda_7 \\ \lambda_8 \end{bmatrix} = \begin{bmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 0.108 \\ 0.119 \\ 0.422 \\ 0.164 \\ 0.108 \end{bmatrix}$$

Given that the data covariance matrix is invertible, and calculating the mean at each data location and the mean at the estimate location, as the stationary plus the non-stationary mean, the estimate value is 1.15. The “true” value at this point was 0.89.

The reproduction of the reference values using a conditioning dataset of one out of four grid nodes is fairly good as shown in Figure 4.27 y 4.28. The only problem arises at the edges of the boundary zone, where unusual kriging weights occur leading to discrepancies between the estimate and reference. These weights are originated because the covariance of the estimate to data is higher than the covariance of the data to itself due to a non-stationary component in the first one, but not for the data-to-data covariance; in this case the estimate is inside the boundary zone, while the sample is outside. For example, the kriging system for the block at 80.5 meters is,

$$\begin{bmatrix} 1.0 & 0.748 & 0.610 & 0.276 & 0.276 & 0.056 & 0.056 & 0.0 \\ 0.748 & 1.793 & 0.276 & 0.765 & 0.056 & 0.276 & 0.0 & 0.056 \\ 0.610 & 0.276 & 1.0 & 0.056 & 0.610 & 0.0 & 0.276 & 0.0 \\ 0.276 & 0.765 & 0.056 & 1.892 & 0.0 & 0.783 & 0.0 & 0.276 \\ 0.276 & 0.056 & 0.610 & 0.0 & 1.0 & 0.0 & 0.610 & 0.0 \\ 0.056 & 0.276 & 0.0 & 0.783 & 0.0 & 1.992 & 0.0 & 0.802 \\ 0.056 & 0.0 & 0.276 & 0.0 & 0.610 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.056 & 0.0 & 0.276 & 0.0 & 0.802 & 0.0 & 2.091 \end{bmatrix} \cdot \begin{bmatrix} \lambda_1=1.929 \\ \lambda_2=-0.171 \\ \lambda_3=-0.575 \\ \lambda_4=-0.020 \\ \lambda_5=-0.017 \\ \lambda_6=0.028 \\ \lambda_7=0.086 \\ \lambda_8=-0.003 \end{bmatrix} = \begin{bmatrix} 1.447 \\ 0.967 \\ 0.566 \\ 0.351 \\ 0.208 \\ 0.097 \\ 0.026 \\ 0.0 \end{bmatrix}$$

In this case the kriging weight for the closest data sample is unusually higher, although the mismatch between the estimate and the reference it is not large. There are other examples for which the differences are more dramatic (see application below).

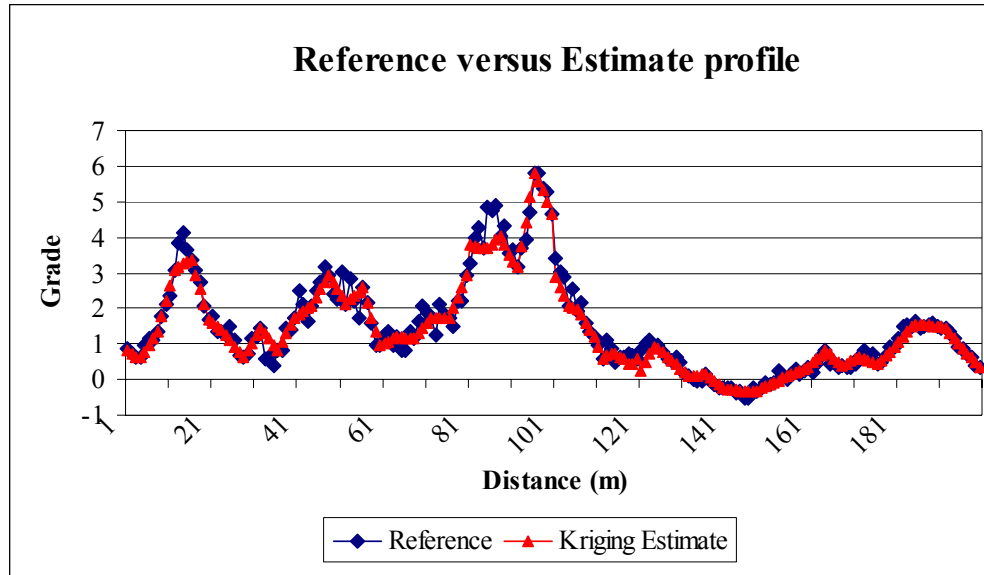


Figure 4.27: Grade reproduction profile along the X-coordinate. Reference values versus kriging estimates.

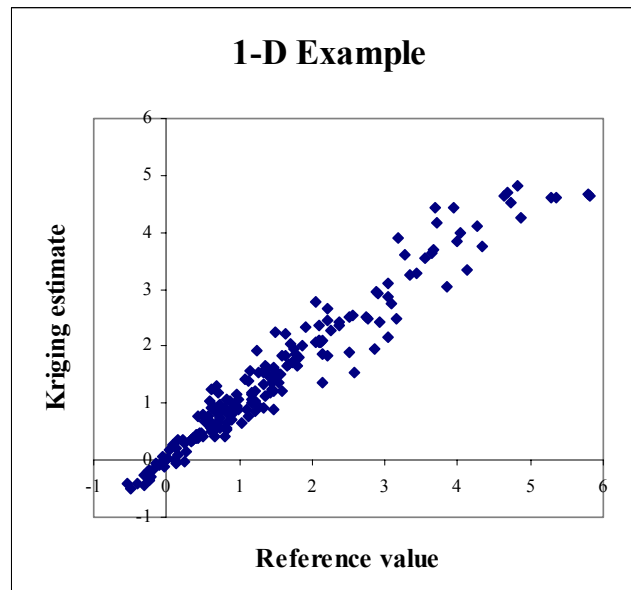


Figure 4.28: Scatter plot reference values versus kriging estimates, for the 1-D example.

For a larger grid, 2000 meters instead of 200, the reproduction of the reference improves (Figure 4.29) as the ergodic fluctuations have less influence in the underlying unconditional simulations use to build the reference and more samples are available to find the stationary and non-stationary components of mean, variance and covariance.



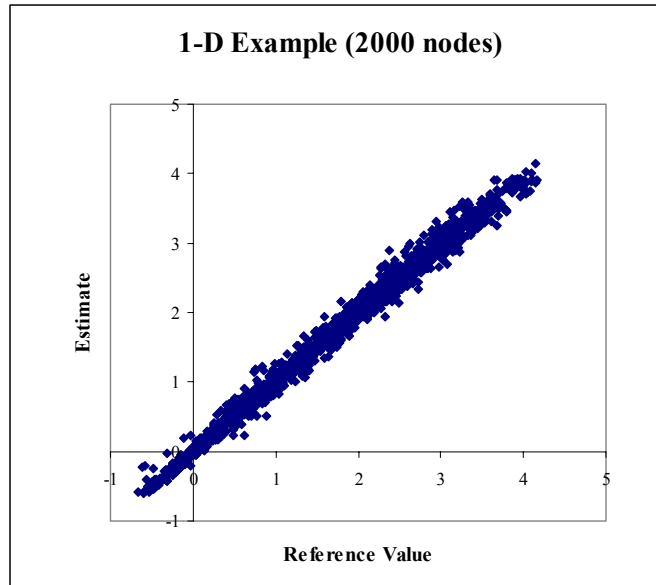


Figure 4.29: Scatter plot reference values versus kriging estimates, for a mimic of the 1-D example but considering 2000 nodes instead of 200.

## 5. Application

In this Chapter we will review the application and the steps involve in the non-stationary kriging in the presence of soft boundaries for a real deposit. The 3-D example was build using the geological model of a porphyry copper deposit from North Chile (Figure 5.1), but grades were simulated. For each rock type an unconditional simulations was generated and transformed to a non-standard normal distribution.

$$\begin{aligned}
 Y_2(\mathbf{u}) &\sim N(3.0, 1.2) & \text{with } \gamma_{Y_2}(\mathbf{h}) &= 0.2 + 0.8 \cdot Sph \begin{pmatrix} h_{\max}=200 \\ h_{\min}=200 \\ vert=100 \end{pmatrix}(\mathbf{h}) \\
 Y_3(\mathbf{u}) &\sim N(1.0, 1.35) & \text{with } \gamma_{Y_3}(\mathbf{h}) &= 0.1 + 0.9 \cdot Sph \begin{pmatrix} h_{\max}=400 \\ h_{\min}=400 \\ vert=100 \end{pmatrix}(\mathbf{h}) \\
 Y_4(\mathbf{u}) &\sim N(1.5, 0.5) & \text{with } \gamma_{Y_4}(\mathbf{h}) &= 0.1 + 0.9 \cdot Sph \begin{pmatrix} h_{\max}=500 \\ h_{\min}=500 \\ vert=100 \end{pmatrix}(\mathbf{h}) \\
 Y_5(\mathbf{u}) &\sim N(0.25, 0.05) & \text{with } \gamma_{Y_5}(\mathbf{h}) &= 0.2 + 0.8 \cdot Sph \begin{pmatrix} h_{\max}=250 \\ h_{\min}=250 \\ vert=100 \end{pmatrix}(\mathbf{h})
 \end{aligned}$$

All variograms are rotated in 90° azimuth and dip of 35° following the principal anisotropy in the deposit. Rock type 1 correspond to a leach zone, rock type 2 to a zone of secondary enrichment, rock type 3 and 4 to units of primary mineralisation corresponding to two intrusive events, and rock type 5 to a peripheral primary mineralisation. The values of rock type 1 were set to a default value (-9) since this unit has no economical interest and it will not be estimate.

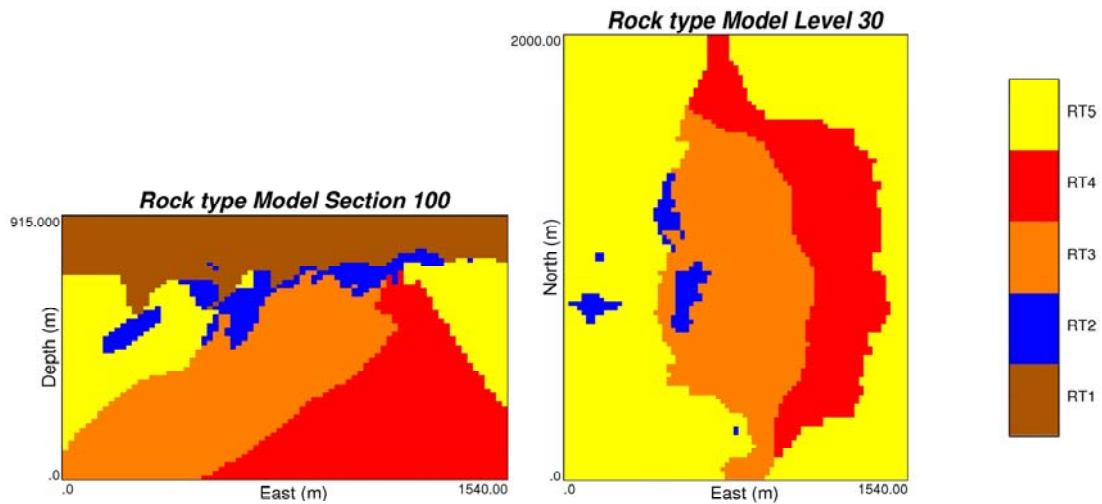


Figure 5.1: Categorical rock type model of a porphyry copper deposit in North Chile.

To perform non-stationary kriging its required a rock type model with the boundary zone and distance to boundary assigned to each block. The `boundmod` program requires the matrix of maximum distance of influence and a set of precedence rules as input parameters. Considering the geology of the deposit the matrix of maximum distance of influence (in meters) between rock types was chosen as follows:

$$\begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 30 & 30 & 30 \\ 0 & 0 & -1 & 120 & 60 \\ 0 & 0 & 90 & -1 & 90 \\ 0 & 0 & 30 & 30 & -1 \end{bmatrix}$$

The boundary zones define by the contact between the primary mineralisation units was assume to be more extended than the ones define by the secondary mineralisation, since they correspond to a wide fault zone. Also the maximum distance of influence of zero of rock type 3 to 5 into rock type 2 reflects that the secondary mineralisation grades into the primary mineralisation but not the opposite.

The set of precedence rules also reflects the timing of the mineralisation; the influence of rock type 2 is the youngest:

$$\left. \begin{array}{l} 4 - 5 \\ 3 - 4 \\ 4 - 3 \end{array} \right\} \text{primary mineralization}$$

$$\left. \begin{array}{l} 5 - 2 \\ 4 - 2 \\ 3 - 2 \end{array} \right\} \text{secondary mineralization}$$

For the same section and level shown in Figure 5.1 the boundary zones and distance to boundary are shown in Figure 5.2.

An additional unconditional simulation was generated to build the non-stationary zones around the boundaries:

$$Y_6(\mathbf{u}) \sim N(0,1) \quad \text{with } \gamma_{Y_6}(\mathbf{h}) = 0.1 + 0.9 \cdot Sph_{\left( \begin{array}{l} h \text{ max}=400 \\ h \text{ min}=400 \\ \text{vert}=50 \end{array} \right)}(\mathbf{h})$$

The variogram of this variable is also rotated in 90° azimuth and dip of 35°.

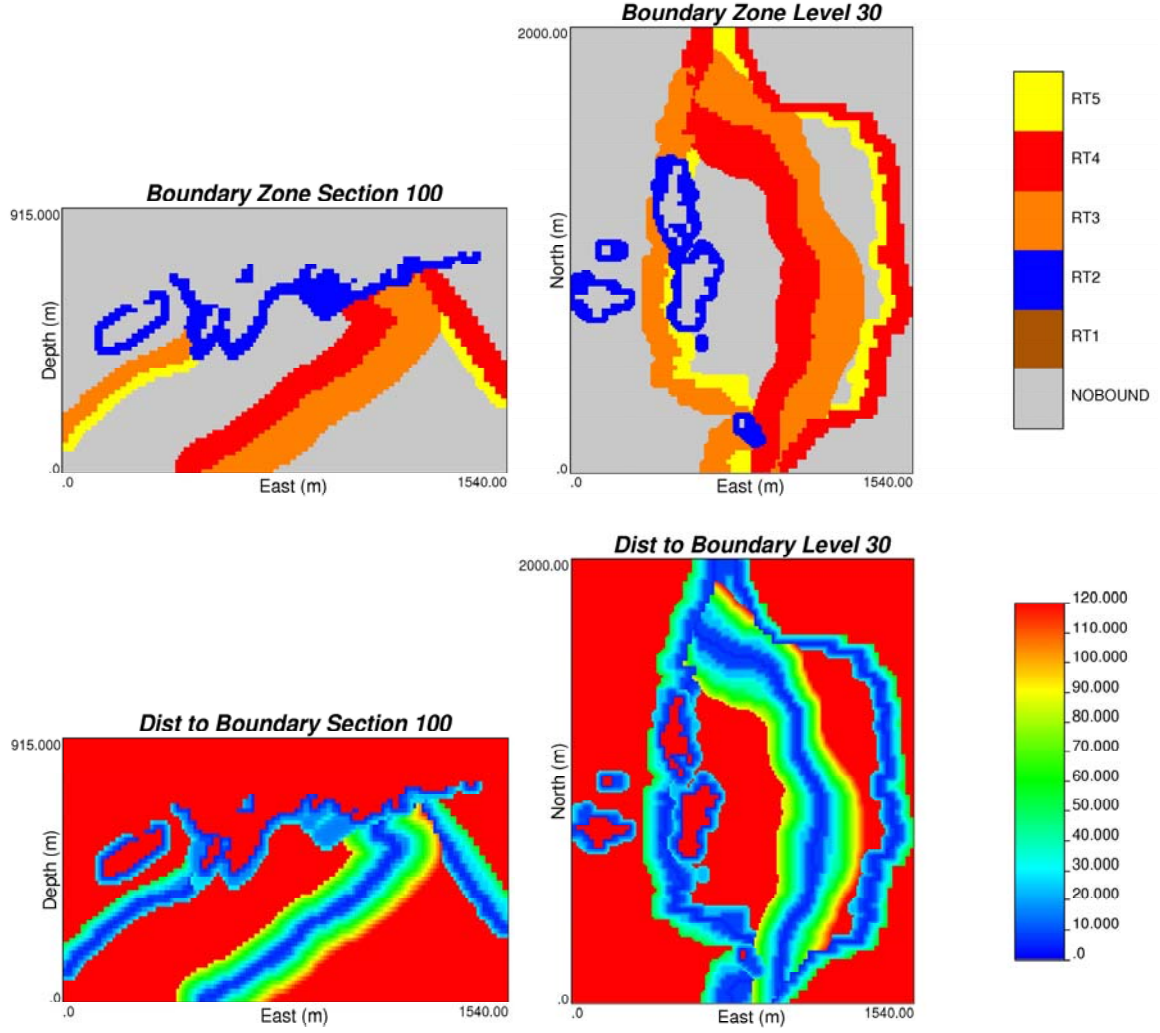


Figure 5.2: Boundary zone (up) and distance to boundary (bottom) maps for the same section and level of the rock type model of Figure 5.1. The default code (NO BOUND) is given to locations beyond the corresponding maximum distance of influence. Distances beyond the maximum distance of influence are set to the default value 9999, in red.

$Y_6(\mathbf{u})$  was transformed to a non-stationary variable (Figure 5.3) using the following expression:

$$Y_6^*(\mathbf{u}_i) = \left[ \frac{(dmax_{kp} - d(\mathbf{u}_i))}{dmax_{kp}} \cdot b_{kp} \right] \cdot Y_6(\mathbf{u}_i) + \left[ \frac{(dmax_{kp} - d(\mathbf{u}_i))}{dmax_{kp}} \cdot a_{kp} \right]$$

where the distance to the boundary as well as the boundary zone were obtained from the rock type model detailed before. For each boundary zone a mean and variance intercept were choose trying to reproduce the real trends in the deposit:

Boundary Zone (RT-BRT)	Intercept $a_{kp}$	Intercept $b_{kp}$
3-2	5.0	2.0
4-2	5.0	2.0
5-2	5.0	2.0
3-4 (and 4-3)	3.0	1.0
3-5 (and 5-3)	0.5	0.5
4-5 (and 5-4)	1.0	1.0

Outside the boundary zone this variable was set to zero, that way, it can be add to the grid file with the simulated values  $Y_2$  to  $Y_5$  to obtain the reference (Figure 5.4).

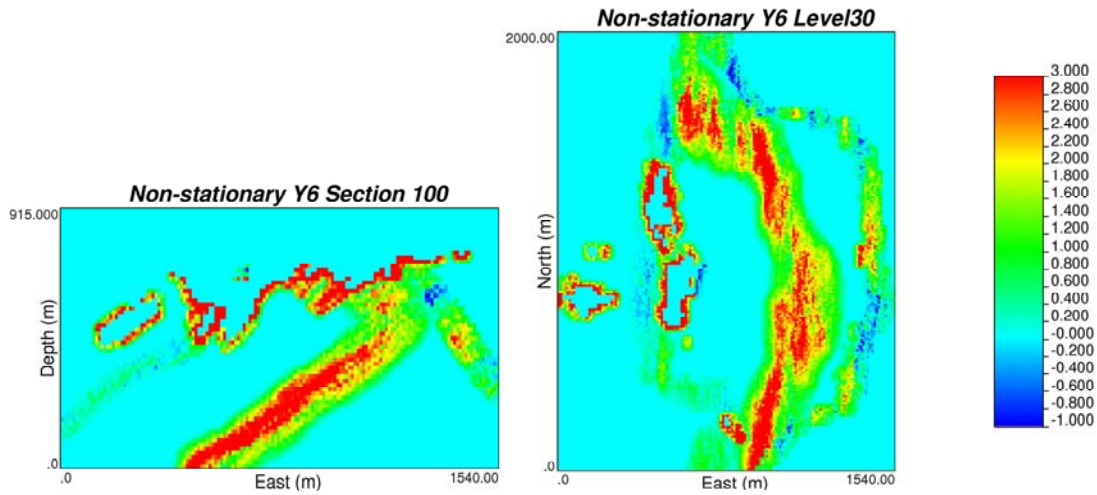


Figure 5.3: Transformed variable  $Y_6$ , blocks outside a boundary zone were set to zero.

Since variables  $Y_2$  to  $Y_5$  were transformed to non-standard normal distributions some negative values occurred. All negative grades (4.3%) were set to zero. Finally the reference grid file was sample in a 100x100x1 grid to obtain the conditioning data to do kriging.

Previous to kriging we need to find the optimum mean, variance and covariance range for the non-stationary boundaries for the conditioning data. The `opt_mean` output (Figure 5.5) shows that the stationary means of each rock type and the  $a$  intercepts are well reproduce for each boundary zone compared with the reference. The  $b$  intercepts corresponding to the non-stationary variance are acceptably well reproduced for all boundaries; Figure 5.6 shows the output of `opt_var` program. As already seen in the 1-D example the differences with target parameters increase for the non-stationary variance intercepts and then for the covariance range, since at

each optimization step the parameter became more sensitive to statistical fluctuations in the previously optimized parameters.

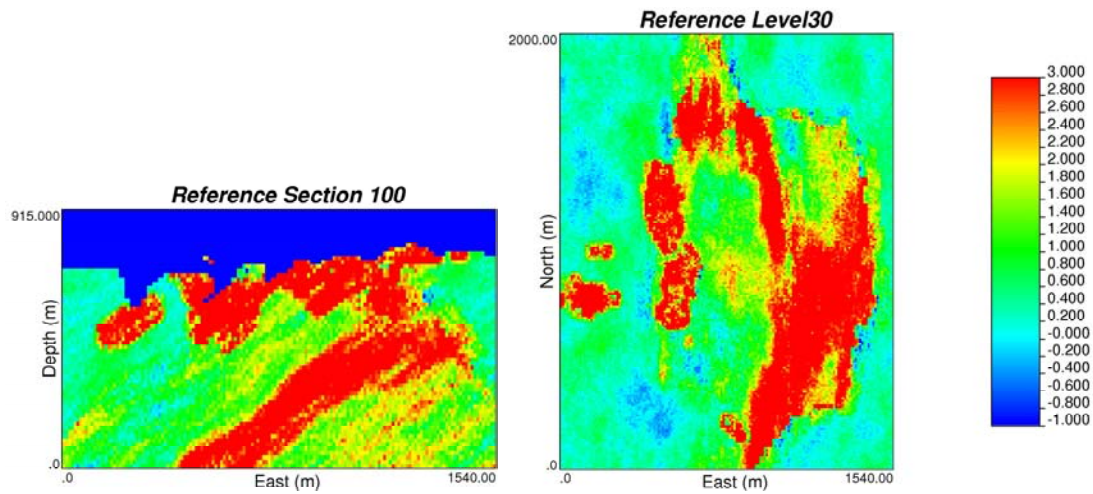


Figure 5.4: Section and Level maps of the reference distribution. Values from rock type 1 were assign a default value of -9 since this unit is of no economic interest therefore it will not be estimate

```
Optimized Means
4
RT
Boundary_RT
Stationary_Mean
NonStationary_Intercept
1 2 -9.0000 0.0000
1 3 -9.0000 0.0000
1 4 -9.0000 0.0000
1 5 -9.0000 0.0000
2 1 3.1180 0.0000
2 3 3.1180 5.0883
2 4 3.1180 5.7429
2 5 3.1180 4.9228
3 1 0.91914 0.0000
3 2 0.91914 5.0883
3 4 0.91914 2.9431
3 5 0.91914 0.60057
4 1 1.6637 0.0000
4 2 1.6637 5.7429
4 3 1.6637 2.9431
4 5 1.6637 1.1153
5 1 0.27290 0.0000
5 2 0.27290 4.9228
5 3 0.27290 0.60057
5 4 0.27290 1.1153
```

Figure 5.5: Output file of opt\_mean program.

```

Optimized Variances
4
RT
Boundary_RT
Stationary_Variance
NonStationary_Intercept
1 2 0.0000 0.0000
1 3 0.0000 0.0000
1 4 0.0000 0.0000
1 5 0.0000 0.0000
2 1 0.99811 0.0000
2 3 0.99811 3.0120
2 4 0.99811 2.3062
2 5 0.99811 2.5425
3 1 0.27125 0.0000
3 2 0.27125 3.0120
3 4 0.27125 0.87515
3 5 0.27125 0.18306
4 1 0.52029 0.0000
4 2 0.52029 2.3062
4 3 0.52029 0.87515
4 5 0.52029 0.56958
5 1 0.43380E-01 0.0000
5 2 0.43380E-01 2.5425
5 3 0.43380E-01 0.18306
5 4 0.43380E-01 0.56958

```

Figure 5.6: Output file of opt\_var program.

In order to find the optimum range of the non-stationary component for each boundary zone, we need: the variogram models of the stationary portions of each rock type and the shape and nugget effect of the relative standardize variogram. The relative standardize variogram correspond to the pseudo stationary shape of the non-stationary portion which is scale by the non-stationary standard deviation at each location.

The samples that belong to the stationary portion of each rock type were selected to calculate and modeled the variograms in the three principal directions. The stationary variance of each rock type was set as the sill. The models are close to the original variograms use in the unconditional simulations except for the minimum horizontal range of rock type 4 that turn to high (Table 5.1).

Rock type	Structure Type	Sill Contribution	Range Hmax	Range Hmin	Range Vert
2	Nugget Effect	0.2503			
	Spherical	0.7477	250.0	250.0	119.0
3	Nugget Effect	0.0300			
	Spherical	0.2400	531.0	373.0	200.0
4	Nugget Effect	0.0476			
	Spherical	0.4724	498.0	627.0	152.0
5	Nugget Effect	0.0092			
	Spherical	0.0308	267.0	248.0	136.0

Table 5.1: Variogram models for the stationary regions of rock type 2 to 5.

The shape, anisotropies and nugget effect of the relative standardize variogram of the non-stationary boundaries were chosen as the same of the variogram use to generate  $Y_6$ , the variable use to build the non-stationary component of this example. Normally the modeler will have to pick this values based on the geology of the deposit, if that would be the case here, the parameters would not be so different than the chosen ones, the non-stationary zone are strongly control by the fault system that coincides with the mayor anisotropy, most of the spatial correlation is explain by spherical models, only the nugget effect is more difficult to assess.

The optimization of the covariance ranges turns reasonable results (Figure 5.7) for all boundary zones except for the boundary defined by the rock type 2 and boundary rock type 4, which seems a little too high, and probably influence by the minimum horizontal range of rock type 4.

```
Optimized Ranges
3
RT
Boundary_RT
Optimum_Range
2      3      326.957855
2      4      660.023438
2      5      465.359070
3      4      425.019592
3      5      335.151306
4      5      576.773804
```

Figure 5.7: Output file of opt\_cov program.

In this example the optimization convergence of the mean and variance occurs around the 3000 iterations accordingly with a larger number of data as in the 1-D example of the previous chapter (Figure 5.8). The covariance range convergence occurs around the 30 iterations for all boundary zones. Considering the number of pairs involve in each calculation is surprisingly similar to the convergence time of the same parameter in the 1-D example.

The kriging parameters use for the estimation was: a minimum and maximum of 4 samples, no octants and an isotopic search radius of 650 meters.

In the first run of kt3d\_bound a significant number of estimates yield unexpected high or low values due to unusual kriging weights; this blocks are normally concentrate in the edges of a boundary zone. They represent a non-physical covariance model that need to be review as part of the future work. To make the kriging matrix stable for the purpose of this application two modifications to the kriging system were done. The first was for the case found in the 1-D examples, that is, when the covariance of the estimate to data is higher than the covariance of the data to itself. This occurs because the estimate is inside a boundary zone and therefore has a non-stationary component that is add to the stationary covariance model, while the data is outside. To fix this the diagonal term corresponding to the



covariance of the sample to itself was set to 1.1 the corresponding element in the estimate-to-data vector of the kriging system. But in this application it was also found locations where the kriging matrices did not have previous problem that yield unusual kriging weights. In this case and when the estimation variance turns to be negative all the diagonal terms of the data-to-data matrix were replace by the maximum value of all elements in the kriging system to make the matrices stable. The origin of these non-physical results in the covariance model adopted in this work to represent a non-stationary soft boundary needs to be review in the future in much more detail, especially for the implementation of this technique in simulation where the correct estimate and estimation variance are essential for the correct reproduction of the conditioning data and its variability.

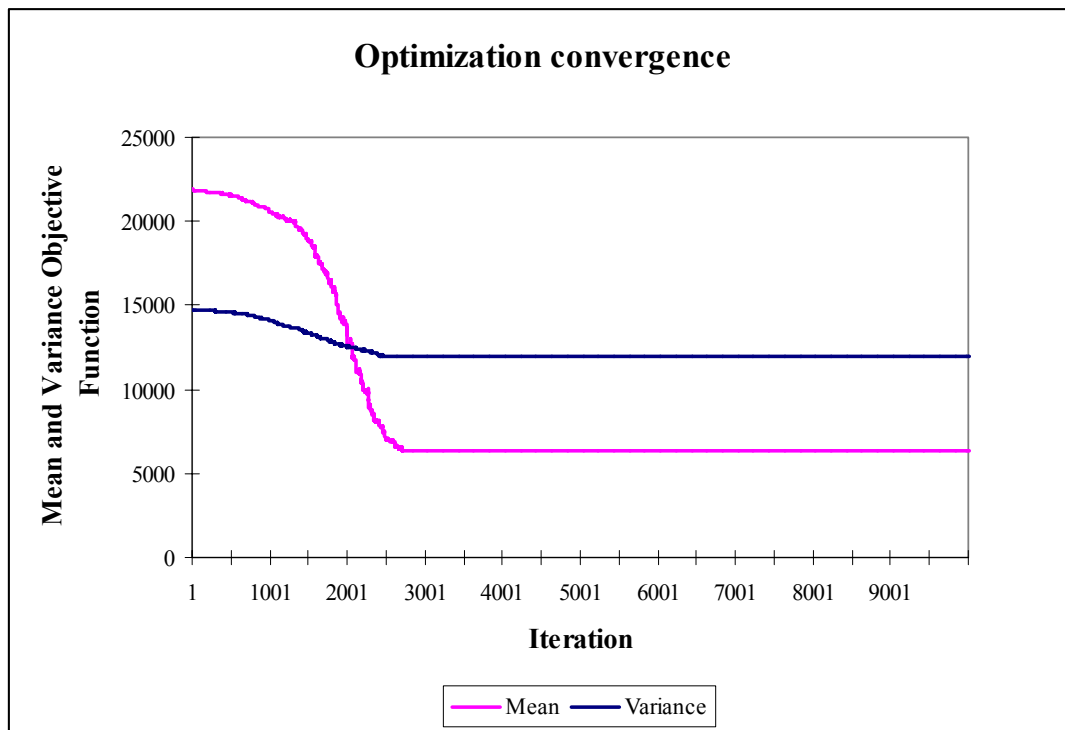


Figure 5.8: Optimization convergence of mean and variance in terms of the objective function value.

The correlation between the estimates and the reference or ‘true’ value is around 0.8 for each boundary zone (Figure 5.10). The stationary portions of rock type 2 and 5 show more conditional bias than rock types 3 and 4; the correlation between the estimate and reference ranges from 0.52 up to 0.8 (Figure 5.11).

Validation of the model and parameters was done running `kt3d_bound` in the cross validation mode removing the whole drill hole to which the sample to be estimate belongs. In this mode the correlation between the estimate and the true value is 0.93. The results show that the resultant model is accurate and precise

(Figure 5.12), the distribution of the error (estimate minus true) is symmetric and centered in zero and have a relatively small standard deviation.

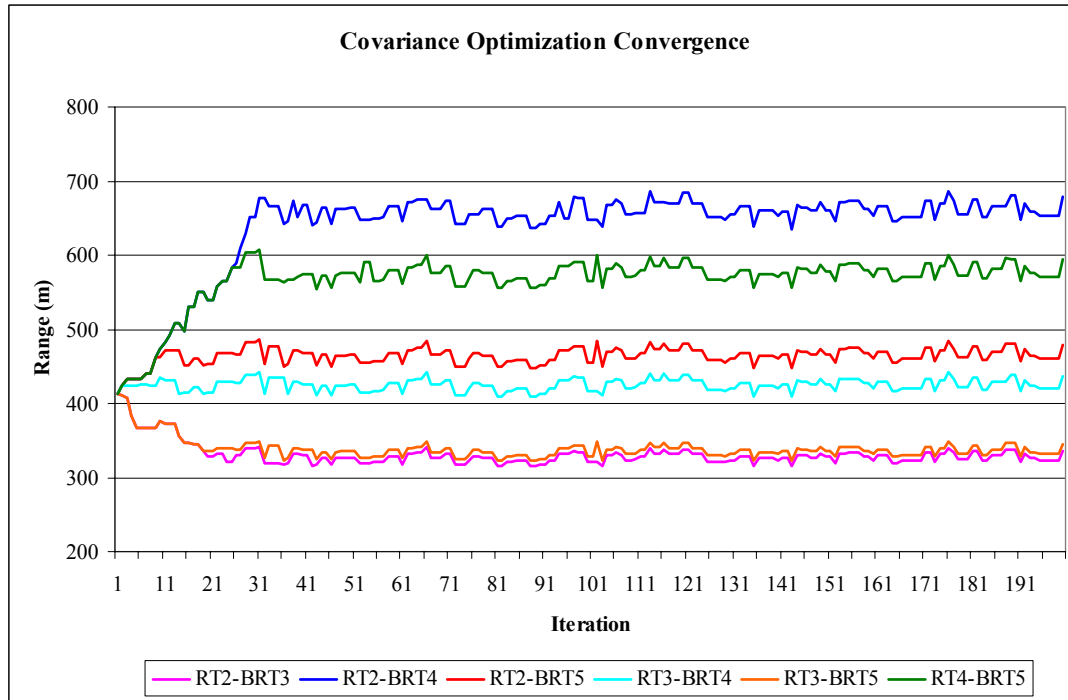


Figure 5.9: Covariance optimization convergence in terms of the range for the 6 boundary zones.

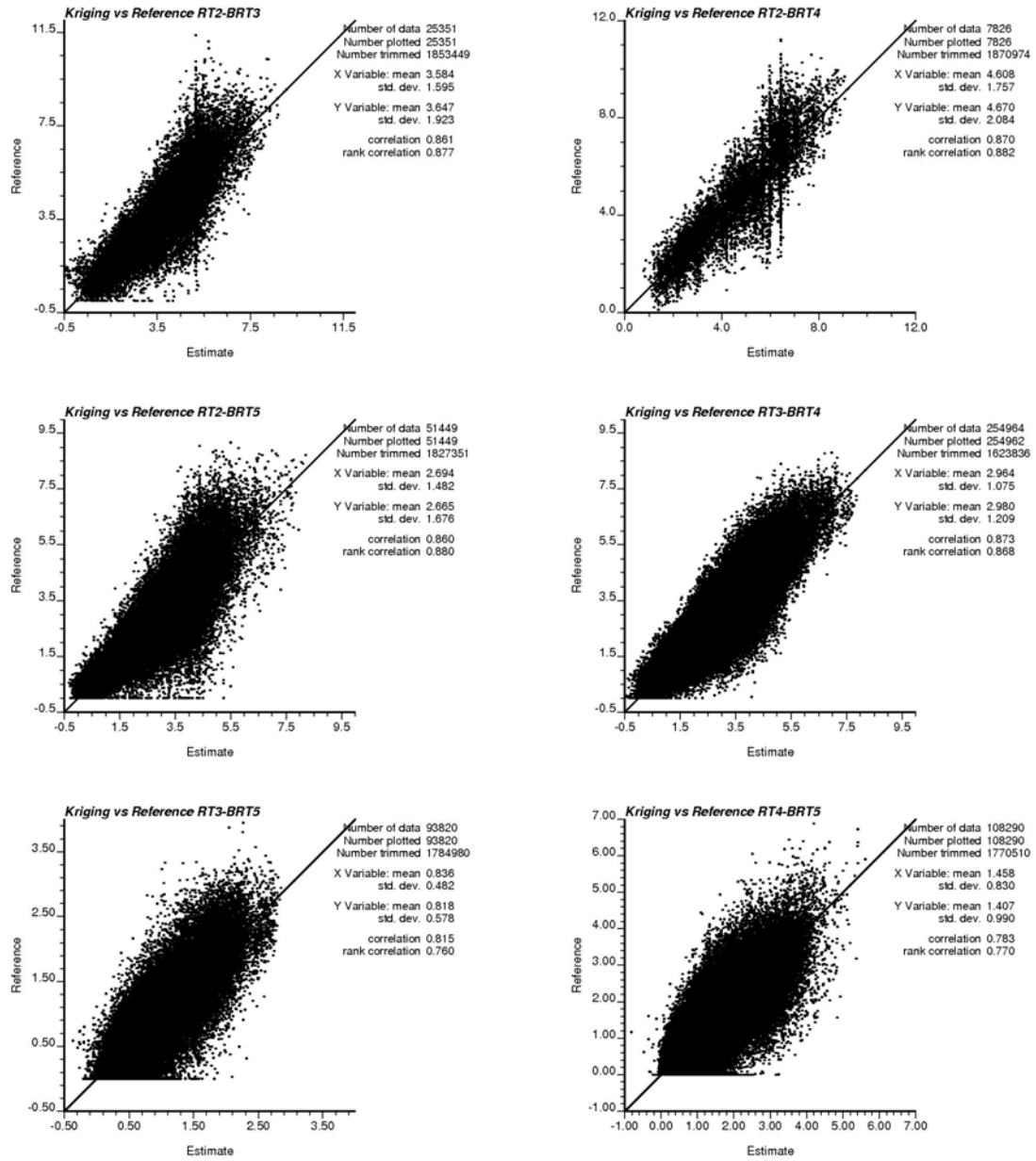


Figure 5.10: Scatter plot of the reference versus the estimate for each boundary zone.

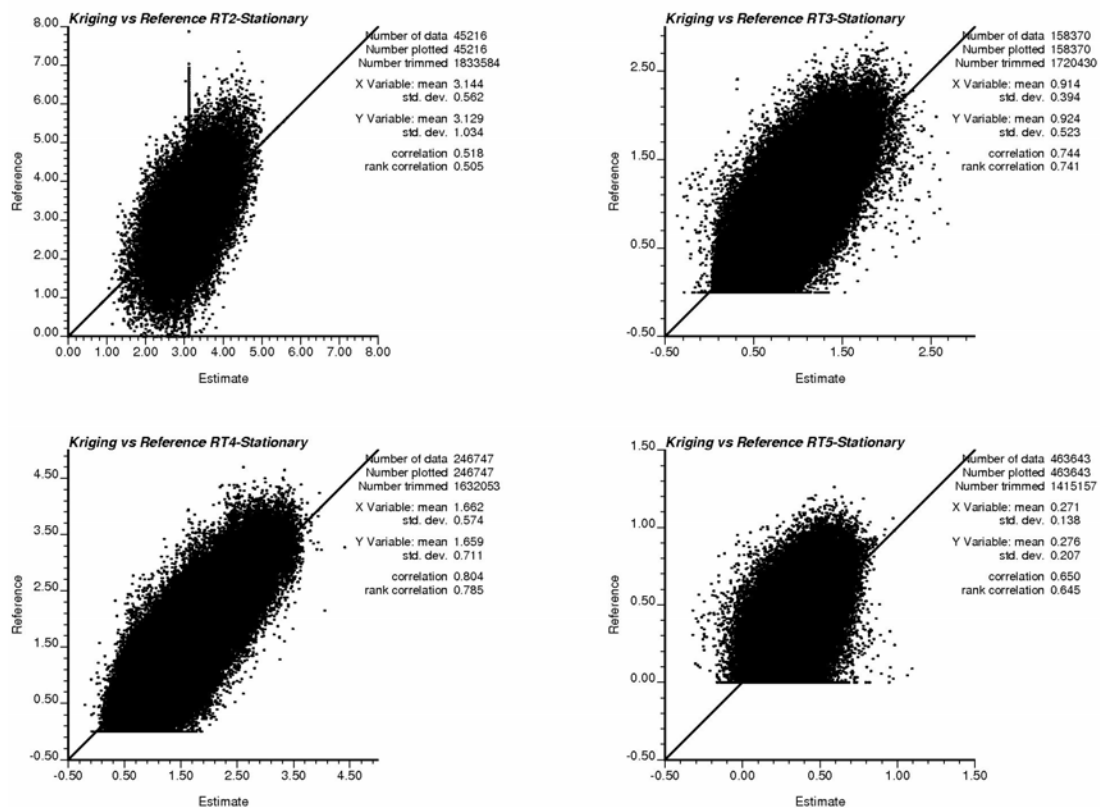


Figure 5.11: Scatter plot of the reference versus the estimate within the stationary portions of each rock type.

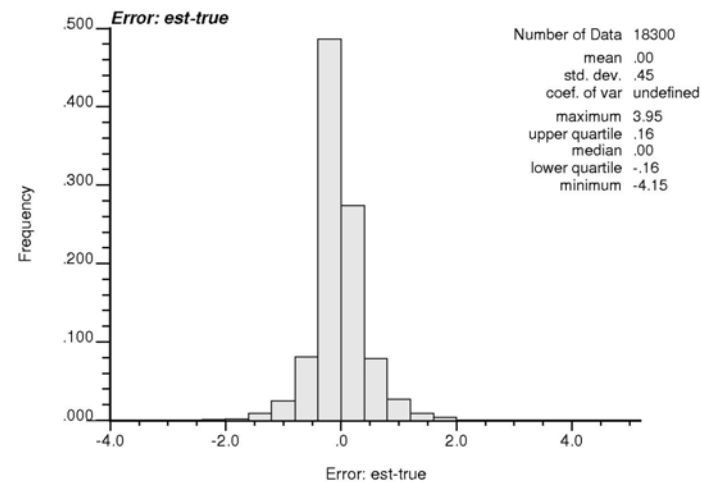
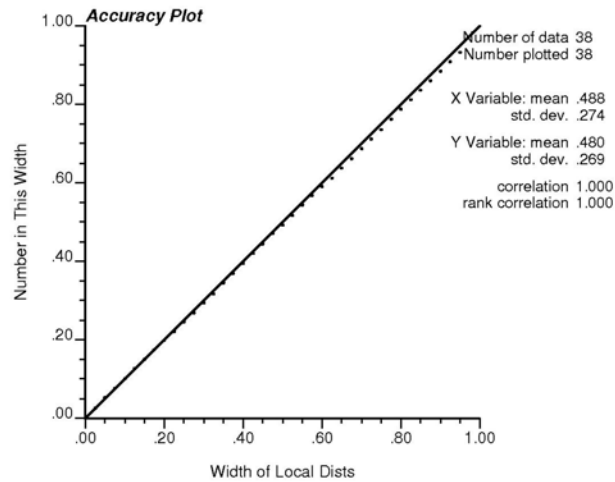
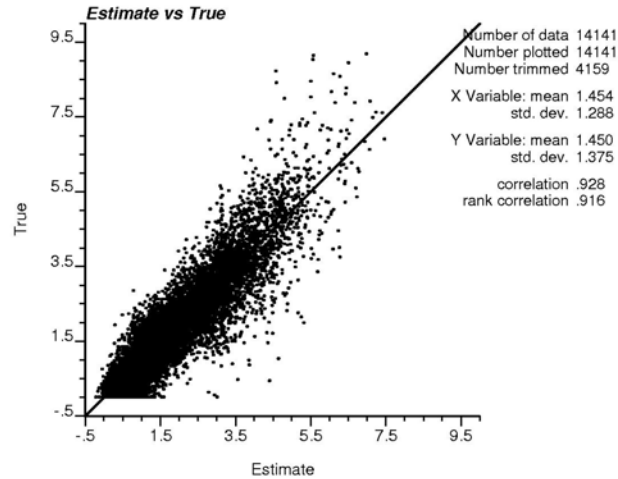


Figure 5.13: Cross validation results. Scatter plot of the estimate versus the true, the correlation is 0.92. The accuracy plot shows that the model is accurate and precise for the chosen parameters. The error (true-estimated) distribution is symmetric and centered in zero and with a relatively low standard deviation.