GEOSTATISTICS FOR FISH SURVEY DATA
Ecole d’été 3–7 Février 1992

M. ARMSTRONG, D. RENARD, J. RIVOIRARD, P. PETITGAS

C–148

Décembre 1992

CENTRE DE GÉOSTATISTIQUE
35, RUE SAINT–HONORÉ, 77305 FONTAINEBLEAU (France)

ECOLE DES MINES
DE PARIS
Acknowledgements

In recent years people involved in fisheries evaluation have shown great interest in spatial statistics techniques. Fruitful discussions arose during the two ICES/CIEM meetings on Spatial Statistics in Brest (4–6 April 90, CM1990/D34) and Reykjavik (5–9 Sept 91, CM1991/D40) at which the Centre de Géostatistique de Fontainebleau participated. Although some geostatistical points were presented there, briefly, it was felt that this was insufficient, and that a one week course would help understanding the basic geostatistical concepts. We are particularly grateful to ICES/CIEM for having publicized this course.

Especial thanks are due to the Commission of the European Communities, and its Programme FAR, for sponsoring this course. This has made it possible for 20 people to come from various places within the European community.

We would also like to thank people who have come from elsewhere.
# Table of contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>0. Introduction</td>
<td>2</td>
</tr>
<tr>
<td>0.1 The data</td>
<td>2</td>
</tr>
<tr>
<td>0.2 The Regionalized Variable</td>
<td>2</td>
</tr>
<tr>
<td>0.3 Random Functions</td>
<td>3</td>
</tr>
<tr>
<td>0.4 Stationary and intrinsic hypotheses</td>
<td>3</td>
</tr>
<tr>
<td>0.5 Other models</td>
<td>5</td>
</tr>
<tr>
<td>1. The Variogram</td>
<td>6</td>
</tr>
<tr>
<td>1.1 Properties</td>
<td>6</td>
</tr>
<tr>
<td>1.2 Models for Variograms</td>
<td>12</td>
</tr>
<tr>
<td>1.3 Computing experimental variograms</td>
<td>17</td>
</tr>
<tr>
<td>1.4 Fitting a Variogram Model</td>
<td>20</td>
</tr>
<tr>
<td>1.5 Troublesome Variograms</td>
<td>21</td>
</tr>
<tr>
<td>2. Dispersion variance and regularization</td>
<td>25</td>
</tr>
<tr>
<td>2.1 Variance of a point within a volume</td>
<td>25</td>
</tr>
<tr>
<td>2.2 Support of a Regionalized Variable</td>
<td>26</td>
</tr>
<tr>
<td>2.3 Dispersion Versus Block Size</td>
<td>26</td>
</tr>
<tr>
<td>2.4 Variance of v within V</td>
<td>29</td>
</tr>
<tr>
<td>2.5 The additivity relation</td>
<td>29</td>
</tr>
<tr>
<td>2.6 Regularization of variogram</td>
<td>30</td>
</tr>
<tr>
<td>2.7 Exercises</td>
<td>32</td>
</tr>
<tr>
<td>3. Estimation variance and global estimation</td>
<td>34</td>
</tr>
<tr>
<td>3.1 Estimation variance</td>
<td>34</td>
</tr>
<tr>
<td>3.2 Global estimation</td>
<td>35</td>
</tr>
<tr>
<td>3.3 Centered regular grid</td>
<td>37</td>
</tr>
<tr>
<td>3.4 Transitive theory</td>
<td>40</td>
</tr>
<tr>
<td>3.5 Exercises</td>
<td>44</td>
</tr>
<tr>
<td>4. Kriging</td>
<td>47</td>
</tr>
<tr>
<td>4.1 The purpose of kriging</td>
<td>47</td>
</tr>
<tr>
<td>4.2 Deriving the kriging equations</td>
<td>47</td>
</tr>
<tr>
<td>4.3 Proportional effect</td>
<td>54</td>
</tr>
<tr>
<td>4.4 Point and block kriging</td>
<td>54</td>
</tr>
<tr>
<td>4.5 Smoothing</td>
<td>54</td>
</tr>
<tr>
<td>4.6 Negative kriging weights</td>
<td>55</td>
</tr>
<tr>
<td>4.7 Screen effect</td>
<td>56</td>
</tr>
<tr>
<td>4.8 The neighbourhood</td>
<td>56</td>
</tr>
<tr>
<td>4.9 Exercise: kriging a segment</td>
<td>58</td>
</tr>
<tr>
<td>4.10 Exercise: kriging a block</td>
<td>58</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>5. Non stationarity</td>
<td>62</td>
</tr>
<tr>
<td>5.1 Generalities</td>
<td>62</td>
</tr>
<tr>
<td>5.2 &quot;Additive&quot; model</td>
<td>63</td>
</tr>
<tr>
<td>5.3 Multiplicative model</td>
<td>63</td>
</tr>
<tr>
<td>6. Universal Kriging</td>
<td>66</td>
</tr>
<tr>
<td>6.1 General Hypotheses</td>
<td>66</td>
</tr>
<tr>
<td>6.2 Estimation of the drift by Least Squares</td>
<td>66</td>
</tr>
<tr>
<td>6.3 Optimal linear estimation of the drift</td>
<td>67</td>
</tr>
<tr>
<td>6.4 Variogram of residuals</td>
<td>68</td>
</tr>
<tr>
<td>6.5 Universal Kriging System</td>
<td>68</td>
</tr>
<tr>
<td>6.6 Can Universal Kriging still be applied?</td>
<td>68</td>
</tr>
<tr>
<td>6.7 Exercise</td>
<td>69</td>
</tr>
<tr>
<td>7. The L.R.E.-k</td>
<td>70</td>
</tr>
<tr>
<td>7.1 Using Linear Combinations</td>
<td>70</td>
</tr>
<tr>
<td>7.2 Intrinsic Random Functions of order k</td>
<td>72</td>
</tr>
<tr>
<td>7.3 Generalized covariances</td>
<td>72</td>
</tr>
<tr>
<td>7.4 Kriging in the IRF-k model</td>
<td>74</td>
</tr>
<tr>
<td>7.5 Inference of the structure</td>
<td>74</td>
</tr>
<tr>
<td>7.6 External Drift</td>
<td>76</td>
</tr>
<tr>
<td>7.7 Exercise: Dual Kriging</td>
<td>76</td>
</tr>
<tr>
<td>8. Related variables</td>
<td>79</td>
</tr>
<tr>
<td>8.1 The structural tools</td>
<td>79</td>
</tr>
<tr>
<td>8.2 Coregionalization Models</td>
<td>80</td>
</tr>
<tr>
<td>8.3 Cokriging</td>
<td>81</td>
</tr>
<tr>
<td>8.4 Maitresse variable</td>
<td>81</td>
</tr>
<tr>
<td>8.5 Exercise</td>
<td>82</td>
</tr>
<tr>
<td>9. Other models</td>
<td>84</td>
</tr>
<tr>
<td>9.1 Gaussian model</td>
<td>84</td>
</tr>
<tr>
<td>9.2 Indicator residual model</td>
<td>86</td>
</tr>
<tr>
<td>9.3 Boolean scheme</td>
<td>87</td>
</tr>
</tbody>
</table>
A word of warning

The fact is that, in natural phenomena there is usually a spatial structure. And this structure can help to understand the phenomenon and to make estimations. The purposes of geostatistics are:

1) To help to find out the structure. We can ask whether it should be expressed in terms of the variogram, the trend, both, or something else?

2) To make explicit use of this structure in estimations.

The geostatistical techniques were first developed for mining evaluations. Since then, they have been enlarged and extended to other fields of application (oil evaluation, forestry, soil sciences...). Their use in fisheries is recent and is mainly limited to variograms and kriging interpolation. The concepts of geostatistical variances or the global evaluation techniques are usually not known, although their usefulness in fisheries is obvious. The way that trends, relations between variables, or some distributional features can be taken into account in geostatistics should also be of interest to fisheries specialists.

These course notes do not cover all topics in geostatistics, but they do give the basic concepts and techniques and also present some more advanced techniques or models that are potentially valuable for fisheries. These notes were initially drafted for the mining industry. What sometimes is called "a metal grade" in the notes, could just as well be a fish density, or any other concentration.

The course notes are complemented by fisheries case-studies, and by two other courses for theoretical references (Matheron 1971, Delfiner 1979).
0. INTRODUCTION

0.1 THE DATA

The first phase of any geostatistical study, before coming to the evaluation, is the structural analysis; that is, the study of the main features of the regionalization. The three main steps in this are firstly, the preliminary checking of the data and getting a feel for the problem, secondly the calculation of the experimental structure and thirdly fitting a mathematical model to the experimental structure. Here we will concentrate on the first step.

After the data has been collected and put into the computer, it has to be thoroughly checked to see that it is correct. This means that any numerical errors in the data have to be corrected, and some elementary statistics should be calculated. But more importantly, the geostatistician has to familiarise himself with the data and the problem to be solved. Most of the major errors in geostatistical studies occur at this point because the person doing the study did not understand the background to the problem correctly, or overlooked some of its essential features.

At the outset of the study it is important to find a biologist or an engineer who has been involved with the sampling program and find out

- what type of sampling procedure was used,
- whether the area is homogeneous,
- whether there has been preferential sampling of high valued areas.

If any of these factors is missed at the beginning of a study, the work may well have to be repeated when it is discovered.

Since one of the basic assumptions in geostatistics is that the data come from a homogeneous population, it is important to apply a few simple statistical tests to the data before starting to calculate the experimental variogram. The means, variances and correlations should be calculated. The histograms of the values should be examined carefully to check for outliers and to see if there is more than one mode. If there are several peaks the data should be rechecked to make sure that they do come from a homogeneous population. After this they should be plotted graphically to check for non-homogeneous regions or for abnormally high or low values. Sometimes these values can be picked up visually even though they do not show up on the histogram. While these tests are rather time-consuming, they are nevertheless vital. It is better to spend a little more time at the beginning than to have to recommence the study.

0.2 THE REGIONALIZED VARIABLE

Since our information about the phenomenon under study is fragmentary, we need a model to be able to draw any conclusions about parts that have not been sampled.

Since we need an estimate of the variance as well as the value itself, some of the well-known statistical methods such as trend surface analysis could be tested. The implicit assumption underlying these types of regression methods is that the surface under study can be represented,
at least locally, by a fairly simple deterministic function such as a polynomial, plus a random error component. Here “random” means that the error is uncorrelated from one place to another, and also uncorrelated with the function.

The problem is that most natural variables display a considerable amount of short scale variation in addition to the large scale trends that can reasonably be described by deterministic functions. If we insist on having uncorrelated errors then the function has to twist and turn a lot. This suggests that we might be better off to allow correlations between values different distances apart. This is the basic idea behind geostatistics.

The term “Regionalized Variable” was chosen by Matheron to emphasize the two apparently contradictory aspects of these types of variables:

i) a random aspect, which accounts for local irregularities, and

ii) a structured aspect, which reflects the large scale tendencies of the phenomenon.

The common statistical models including trend surfaces put all the randomness into the error term while all the structure is put into the deterministic term. This is not always realistic. A better way of representing the reality is to introduce randomness in terms of fluctuations around a fixed surface which we can call the “drift” to avoid any possible confusion with the term “trend”. Fluctuations are not necessarily "errors" but rather fully fledged features of the phenomenon, with a structure of their own. The first task in a geostatistical study is to identify these structures, hence the name "structural analysis". After this the geostatistician can go on to solve various types of estimation problems.

To avoid any confusion, we will make a clear distinction between:
- the Regionalized Variable, which is the real variable under study,
- and the model we use to represent this variable.

0.3 RANDOM FUNCTIONS

The observed value $z(x)$ at each data point $x$ is considered as the outcome of a random variable, $Z(x)$, whose mean at point $x$ is called the drift, $m(x)$. At points where no measurements have been made, the values $z(x)$ are well-defined even though they are unknown. They can also be thought of as being the outcomes (or realizations) of the corresponding random variables $Z(x)$. In mathematical terms, the family of all these random variables is called a random function. (Synonyms: stochastic process – in 1D – random field). A random function bears the same relation to one of its realizations as a random variable does to its outcome, except that the realization of a random function is a function whereas the outcome of a random variable is a number.

A random function is characterized by its spatial distribution, i.e. by the joint distributions of any set of variables $Z(x_1), Z(x_2), \ldots Z(x_k)$, for all $k$, and for all points $x_1, x_2, \ldots x_k$. Of course, it would be impossible to do anything with this model unless we are prepared to make some assumptions about the characteristics of these distributions. In particular since only one realization is available we have to make some assumptions about its stationarity. The same problem arises with random variables. What could be said about a random variable on the basis of one realization in which $z(x) = 34$ ?
0. 4 STATIONARY AND INTRINSIC HYPOTHESES

In statistics it is common to assume that the random variable is stationary, i.e. its distribution is invariant under translation. In the same way, a stationary random function is homogeneous, and its distribution is invariant under translation. This makes statistical inference possible. In its strictest sense stationarity requires all the moments to be invariant under translation but since this cannot be verified from the limited experimental data, we usually only require the first two moments (the mean and the covariance, which is sufficient for linear estimations) to be constant. This is called "weak" or second order stationarity. In other words we require:

(i) the expected value (or mean) of the function $Z(x)$ is constant for all points $x$. That is, $E(Z(x)) = m(x) = m$ which is independent of $x$.

(ii) the covariance function between any two points $x$ and $x + h$ is independent of the point $x$.

It depends only on the vector $h$. That is,

$E [Z(x) - m][Z(x+h) - m] = E[Z(x) Z(x+h)] - m^2 = C(h)$

In particular, when $h = 0$, the covariance comes back to the ordinary variance of $Z(x)$ which must also be constant.

In practice, it often happens that these assumptions are not satisfied. Clearly when there is a marked trend the mean value cannot be assumed to be constant. We shall see how to take account of trends later. For the moment we shall only consider cases where there is no trend. However, even in such cases, the variance and then the covariance need not exist. A particularly startling practical example of this was found by Krige (1978) for the grades of gold in South Africa. On both theoretical and practical grounds it is convenient to be able to weaken this hypothesis.

Under the "intrinsic hypothesis" we suppose that the increments of the function are weakly stationary: that is, the mean and variance of the increments $Z(x+h) - Z(x)$ exist and are independent of the point $x$.

$E[Z(x+h) - Z(x)] = 0$ intrinsic hypothesis

$\text{Var} [Z(x+h) - Z(x)] = 2\gamma(h)$ with zero mean

Note that the mean of the increments does not depend on $h$ either: it is zero, which means that there is no drift.

The function $\gamma(h)$ is called the semi-variogram (we say variogram for short).

We will see later that if a RF is stationary, there is an equivalence between the variogram $\gamma(h)$ and the covariance $C(h)$, and secondly that we can calculate the variance of any linear combination:

$\sum \lambda_i Z(x_i)$

in terms of either the variogram or the covariance. RF that are stationary always satisfy the intrinsic hypothesis but the converse is not necessarily true.

In contrast to the stationary case, when working with intrinsic variables the operations are defined only for increments. So we will see later that we can calculate the variance of linear combinations only if the sum of the weights is 0. The types of functions that can be variograms are much richer
than just those corresponding to covariances. So by using intrinsic variables instead of just stationary ones, we are obliged to work only with increments but by doing so the range of variogram models available is considerably enlarged.

The variogram is the basic tool for the structural interpretation of phenomena as well for linear estimation. It enables us to compute geostatistical variances, and to make global as well as local estimations.

0.5 OTHER MODELS

Drift and related variables

When a drift exists in the structure of the Regionalized Variable, the variogram may become difficult to use. We will see different models and another structural tool, the generalized covariance.

We will also see how to use a drift given by another variable, or how we can study different variables together.

Non linear models

The previous methods were linear and did not take account of the statistical distribution of the variable. To do this, different models of Random Functions are available. We will see some of these.

Transitive theory

Although modelling a Regionalized Variable as a Random Function proves to be very convenient, it is not the only possibility. For instance the relations between the geostatistical variances and variograms can be developed and proved outside this framework.

But besides that, there is a part of geostatistics, called the transitive theory, which is formulated outside the Random Function framework. It only provides linear and global estimation, but its validity does not require any hypothesis of stationarity. Moreover it does not suppose that the behaviour of the Regionalized Variable is independent of the geometry and the borders of its field, as is implicitly assumed in the intrinsic theory.
1. THE VARIOGRAM

1.1 PROPERTIES

The variogram measures the mean variability between two points \( x \) and \( x + h \) as a function of their distance \( h \). The variogram of an intrinsic random function is defined by the equation:

\[
\gamma(h) = 0.5 \ Var \ [Z(x+h) - Z(x)]
\]

As it has been assumed that the mean of \( Z(x+h) - Z(x) \) is zero, then \( \gamma(h) \) is just the mean square value of the difference. That is,

\[
\gamma(h) = 0.5 \ E \ [Z(x+h) - Z(x)]^2
\]

Here \( x \) and \( x + h \) refer to points in a \( n \)-dimensional space where \( n \) could be 1, 2 or 3. For example, when \( n = 2 \) (i.e. in the plane), \( x \) denotes the point \((x_1, x_2)\) and \( h \) is a vector \((h_1, h_2)\). As \( \gamma(-h) = \gamma(h) \), the variogram is a function of the modulus of the vector \( h \) and its orientation. For a fixed angle, the variogram indicates how different the values become as the distance increases. When the angle is changed, the variograms disclose directional features of the phenomenon such as its anisotropy.

The graph of \( \gamma(h) \) plotted against \( h \) presents the following features:

. It starts at 0 (for \( h = 0 \), \( Z(x+h) = Z(x) \))
. It generally increases with \( h \)
. It rises up to a certain level called the sill and then flattens out. Alternatively it could just go on rising.

The properties will now be treated in some detail.

Behaviour near the Origin

The behaviour for small values of \( h \) is related to the continuity and the spatial regularity of the variable. Four types of behaviour near the origin are shown on Figure 1.1.

(a) Parabolic. This indicates that the regionalized variable (Re. V.) is highly continuous. It is differentiable.

(b) Linear. In this case the Re V is continuous but not differentiable, and thus less regular than in (a).

(c) Discontinuous at the origin i.e. when the \( \gamma(h) \) does not tend to \( \gamma(0) = 0 \) as \( h \) tends to 0. This means that the variable is highly irregular at short distances. Many variables show this type of behaviour. This jump at the origin is called a nugget effect because it was first noticed in gold deposits in South Africa where it is associated with the presence of nuggets of gold in the ore. There the grade passes abruptly from zero outside the nugget to a high value inside
it. It is convenient to apply the term “nugget effect” to this sort of short range variability even when it is known to be due to some other factor e.g. the micro-structure, measurement error or errors in location.

(d) Flat. Pure randomness or white noise. The regionalized variables $Z(x+h)$ and $Z(x)$ are uncorrelated for all values of $h$, no matter how close they are. This is the limiting case of a total lack of structure.

Figure 1.1 The behaviour of the variogram near the origin

**Range and Zone of Influence**

The rate of increase of the variogram with $h$ indicates how quickly the influence of a sample drops off with distance. After the variogram has reached its limiting value (its sill) there is no longer any correlation between samples this far apart. This critical distance, called the range of the variogram, gives a more precise definition to the notion of “zone of influence”. The value of the variogram when there is no correlation between $Z(x+h)$ and $Z(x)$ corresponds to the stationary variance:

$$\gamma(h) = 0.5 \ Var \{ Z(x+h) - Z(x) \}$$

$$= 0.5 \ \{ Var \{ Z(x+h) \} + Var \{ Z(x) \} \}$$

$$= Var \{ Z(x) \}$$

Figure 1.2: Bounded and unbounded variograms

Not all variograms reach a sill. Some like the one shown in Figure 1.2, keep on increasing with $h$. This is one fundamental difference between the variogram and the covariance. The latter only exists for stationary variables and is bounded.
Nested structures

The range need not be the same in all directions. This merely reflects the anisotropy of the phenomenon. What is more, even for a given direction there can be several nested structures with different ranges, thus acting at different scales of distance.

![Nested structures](image1.png) ![Hole Effect](image2.png)

Figure 1.3

Hole effect

In between the origin and infinity, the behaviour of the variogram shows various features of the phenomenon, notably the presence of nested structures and in very rare cases, periodicities or a hole effect.

A strictly periodic phenomenon can only exist in 1D. After increasing, the theoretical variogram decreases and is strictly zero for the distance equal to the period and its multiples.

Sometimes the variogram presents a bump, then a hole (the covariance presents a hole, then a bump). The distance corresponding to the minimum of variability may represent a pseudo period.

However experimental variograms may present fluctuations, which are not statistically significant.

Anisotropies

When the variogram is calculated for all pairs of points in certain directions such as North-South or East-West, it sometimes shows different types of behaviour in some of them (i.e. anisotropy). If this does not occur, the variogram depends only on the magnitude of the distance between points \( h \) and is said to be isotropic.

Two different types of anisotropy will be distinguished: geometric anisotropy and zonal anisotropy.

(a) Geometric Anisotropy (also called “elliptic” anisotropy). In this case the anisotropy can be corrected by an affine change of coordinates. Typical situations are shown below.
Figure 1.4 Geometric or elliptic anisotropy

On the left in Figure 1.4 the variograms have the same sill in all directions even though their ranges are different while on the right they are both linear but have different slopes. We can draw a diagram of the variation of the range (in case a) and the slope (in case b) as a function of direction. If the curve is an ellipse (in 2-D), then the anisotropy is said to be geometric (or elliptic). In these cases a simple change of coordinates transforms the ellipse into a circle and eliminates the anisotropy.

Figure 1.5 The ellipses showing the major and minor axes in the case of geometric anisotropy.

This transformation is particularly simple when the major axes of the ellipse coincide with the coordinate axes as is shown on the left of Figure 1.5. Then if the equation of the variogram in direction 1 is \( \gamma_1(h) \), the overall variogram after correcting for the anisotropy is of the form:

\[
\gamma(h) = \gamma_1 \left( (x_1 - x_2)^2 + k^2(y_1 - y_2)^2 \right)
\]

when \( k \) is the anisotropy ratio, namely:

\[
k = \frac{\text{range 1}}{\text{range 2}} \quad \text{or} \quad k = \frac{\text{slope 2}}{\text{slope 1}}
\]

When calculating the variogram, it is important to use at least four directions if possible. If the variogram was only calculated in two perpendicular directions it would be possible to miss the anisotropy completely.
(b) Zonal (or stratified) Anisotropy

There are more complex types of anisotropy. For example, in 3-D the vertical direction often plays a special role because there is more variation between layers than within them. In such cases it is standard practice to split the variogram into two components, e.g. an isotropic one plus another which depends only on the vertical component:

$$\gamma_0 \left( \sqrt{h_1^2 + h_2^2 + h_3^2} \right) + \gamma_1 (h_3)$$

isotropic component vertical component

Presence of a Drift

Theory shows that for large distances the variogram must increase more slowly than a parabola. To be more specific

$$\frac{\gamma (h)}{h^2} \to 0 \text{ as } h \to \infty$$

However when computing $E[ Z(x + h) - Z(x)]^2$ in practice, we often find variograms which increase more rapidly than $h^2$. This indicates the presence of a drift $E[ Z(x + h) - Z(x)] \neq 0$. Figure 1.6.

$$E[ Z(x + h) - Z(x)]^2 = \text{Var} [ Z(x + h) - Z(x)] + (E[ Z(x + h) - Z(x)])^2$$

Figure 1.6 Variogram shape in presence of a drift.

With a linear drift in 1-D for instance, we have $E[ Z(x + h) - Z(x)] = ah$, which adds a parabolic term $0.5 a^2 h^2$ to the variogram. For small values of $h$ this is negligible but as $h$ increases it becomes the dominant term, which explains the rapid growth. Although this demonstration has been done for 1-D, the result is equally true for a linear drift in 2-D and 3-D.

The effect of any other drift is more difficult to evaluate because the value of $m(x + h) - m(x)$ depends on $x$ as well as $h$. The additional term then becomes the average of $0.5 (m(x + h) - m(x))^2$ over the whole region. It can produce quite curious effects on the variogram. For example a dome shaped drift (which is common with oil reservoirs) gives rise to a dome shaped variogram with its maximum at about half the diameter of the zone. It is intuitively clear that the greatest variation occurs when
one of the points in the pair is at the centre of the dome while the other is at the outside. For distances greater than this the difference between the values tends to be smaller.

Of course if the drift is known, it may be easy to take it off and to work on the residuals from the drift. But if it is unknown, we will see that things are not only difficult, but also misleading (see non stationary models: UK and IRF-k).

**Local stationarity**

Often the variogram is only used up to a certain distance. This limit could be the extent of homogeneity within the field (the variogram departs from its theoretical sill after a while) or the diameter of the neighbourhood used in kriging (i.e. estimation). Consequently, the phenomenon only has to be considered as stationary up to this distance. The problem is to decide whether we can find a series of moving neighbourhoods within which the values can be considered to be stationary and where there are enough data in these zones to give meaningful estimates. This assumption of local stationarity is really a compromise between the scale of homogeneity of the phenomenon and that of the sampling density. This can best be seen from an example.

![Diagrammatic representation of sulphur grades](imageFFECT)

*Diagrammatic representation of sulphur grades*

![Blow-up of central section](imageEFF)

*Blow-up of central section*

*Figure 1.7*
Consider the sulphur content of coal along a transect. (Figure 1.7). Over the total distance shown (9 km) there is a clear trend in the variable. However if we look at a blow-up of the central section of this, the fluctuations appear to cover up the trend. This means that at this scale the sulphur content could be considered as a locally stationary or, at least, intrinsic variable at this scale.

**Proportional Effect**

A variogram is said to have a proportional effect when its value (particularly its sill) increases with the local mean grade. This often occurs with lognormally or skewly distributed data. The variograms for different zones have the same shape but the sill of the variograms in rich zones is much higher than in poor ones. Figure 1.8. Often the sill turns out to be proportional to the square of the local mean. So the underlying variogram model can be found by dividing each of the local variograms by the square of the local mean and then averaging them before fitting a variogram model.

![Graph showing proportional effect](image)

**Figure 1.8 Proportional effect**

1.2 **MODELS FOR VARIOGRAMS**

Before the variogram can be used to estimate grades or tonnages, a mathematical model has to be fitted to it. The reason for this is that variograms have to satisfy certain conditions. Otherwise there is always a risk of finishing up with a negative variance which would be totally unacceptable.

**Variance of Admissible Linear Combinations**

We shall start by considering a stationary variable $Z(x)$ with a covariance function $C(h)$. Since we will eventually want to use a linear combination of the data as an estimator, we will need to be able to calculate the variance of linear combinations such as:
\[ Z' = \sum_{i} \lambda_i \, Z(x_i) \]

Here
\[ \text{Var} (Z') = \sum_{i} \sum_{j} \lambda_i \, \lambda_j \, C(x_i - x_j) \]

This must be non-negative whatever the points \( x_i \) and whatever the weights \( \lambda_i \). A function which satisfies this condition is said to be positive definite. Conversely, any positive definite function can be considered as the covariance of some stationary random function.

The situation is slightly different when the variable is intrinsic but not stationary. In this case the variance of an arbitrary linear combination need not exist; we can only be sure that this exists for linear combinations of increments. These combinations which are said to be "admissible", are characterized by:

\[ \sum \lambda_i = 0 \]

Clearly any linear combination of increments satisfies this condition since any single increment involves the weights +1 and -1. Conversely any linear combination satisfying this condition can be written as a linear combination of increments since, on choosing an arbitrary point as origin, we have:

\[ \sum \lambda_i \, Z(x_i) = \sum \lambda_i \, [Z(x_i) - Z(0)] \]

So its variance exists and is given by:
\[ \text{Var} \left[ \sum \lambda_i \, Z(x_i) \right] = \sum \sum \lambda_i \, \lambda_j \, \text{Cov}[Z(x_i) - Z(0), Z(x_j) - Z(0)] \]

To calculate the covariance of increments we use the identity:
\[ \text{Var}(Z(x) - Z(y)) = \text{Var}(Z(x) - Z(0)) + \text{Var}(Z(y) - Z(0)) - 2\text{Cov}(Z(x) - Z(0), Z(y) - Z(0)) \]

Since \( 2\gamma(h) = \text{Var}(Z(x + h) - Z(x)) \), we have
\[ 2\gamma(x - y) = 2\gamma(x) + 2\gamma(y) - 2\text{Cov}(Z(x) - Z(0), Z(y) - Z(0)) \]

Solving for \( \text{Cov}(.) \) and substituting this into the equation for \( \text{Var} \left( \sum \lambda_i \, Z(x_i) \right) \) gives:

\[ \text{Var} \left( \sum \lambda_i \, Z(x_i) \right) = \sum \sum \lambda_i \, \lambda_j \, (\gamma(x_i) + \gamma(x_j) - \gamma(x_i - x_j)) \]

\[ = \sum \lambda_i \, \sum \lambda_j \, \gamma(x_i) + \sum \lambda_i \, \sum \lambda_j \, \gamma(x_j) - \sum \sum \lambda_i \, \lambda_j \, (x_i - x_j) \]

The first two terms disappear since \( \sum \lambda_i = 0 \), leaving:

\[ \text{Var} \left( \sum \lambda_i \, Z(x_i) \right) = -\sum \sum \lambda_i \, \lambda_j \, \gamma(x_i - x_j) \]
Hence the very important result that the variance of any linear combination whose sum of weights is 0 exists and can be calculated by replacing the covariances in equation by \(-\gamma\).

As the variance must be non-negative, variogram models have to satisfy certain conditions. For any set of points \(x_1, x_2, \ldots, x_k\), any set of weights \(\lambda_1, \lambda_2, \ldots, \lambda_k\), such that \(\sum \lambda_i = 0\)

\[-\sum \sum \lambda_i \lambda_j \gamma (x_i - x_j) \geq 0\]

then \(-\gamma\) is said to be conditionally positive definite.

Note that this condition is weaker than the preceding condition for covariances which had to hold for all possible weights, whereas this one only has to hold for sets of weights whose sum is 0. Consequently the class of admissible models is richer than for covariances. It contains the bounded variograms associated with covariances, \(\gamma(h) = C(0) - C(h)\), and also unbounded ones which have no covariance counterpart.

**Admissible models**

We have seen that in order to ensure that the variance of any linear combination of a stationary regionalized variable is always non-negative, only certain functions can be used as covariance models. These must be positive definite functions. Similarly when working with intrinsic regionalized variables, only admissible linear combinations can be used; that is, those whose weights sum to 0. In that case the function used to model the variogram has to be conditionally negative definite.

It is difficult to test whether a certain function is or is not positive definite even in a 1-D space. The problem becomes even more complicated in higher dimensional spaces. A function which is positive definite in 2 or 3-D is positive definite in 1-D. But for example the piece-wise linear function shown in Figure 1.9 is admissible in a one dimensional space but not in two or three dimensions. As it is not easy to recognize functions that are positive definite or to test them for this property, it is best to choose variogram models from the range of suitable functions rather than trying to create them oneself. A list of the commonly used variogram models is given below.

![Figure 1.9: Piece-wise linear model that is admissible in 1D but not in 2D or higher dimensions.](image)

These can be combined additively with positive coefficients to obtain other admissible models (nested structures). That is, they can be added together but subtraction is not allowed. Nor can
they be combined over distance range. By this we mean that you cannot choose a model consisting of one spherical (say) up to a distance of 100m then another model from 100m onwards.

**Common Variogram Models**

The following variogram models are admissible. Those with a sill correspond to stationary regionalized variables; while the unbounded models are associated only with intrinsic variables. This list is not exhaustive.

1) **Nugget Effect**

\[ \gamma(h) = 0 \quad h = 0 \]

\[ = C \quad |h| > 0 \]

This model corresponds to a purely random phenomenon (white noise) with no correlation between values no matter how close they are.

2) **Spherical Model**:

\[ \gamma(h) = \begin{cases} 
  C \left[ \frac{3}{2} \frac{|h|}{a} - \frac{1}{2} \left( \frac{|h|^3}{a^3} \right) \right] & |h| < a \\
  C & |h| > a 
\end{cases} \]

The spherical model is probably the most commonly used model. It has a simple polynomial expression and its shape matches well with what is often observed: an almost linear growth up to a certain distance then a stabilization. The tangent at the origin intersects the sill at a point with an abscissa 2a/3. This can be useful when fitting models.

3) **Exponential Model**:

\[ \gamma(h) = C \left[ 1 - \exp \left( -\frac{|h|}{a} \right) \right] \]

For practical purposes, the range can be taken as 3a. So this is different to the spherical model. The tangent at the origin intersects the sill at a point with an abscissa a.

![Figure 1.10](image-url)
Figure 1.10 shows the spherical variogram model with a sill of 1.0 and a range of 1.0 on the left, and an exponential model with a sill of 1.0 and a scale parameter \( a \) of 0.33 (i.e. its practical range is 1.0). Although both models have the same sills and the practical range of the exponential equals the true range of the spherical, they are clearly different. The exponential rises more rapidly initially but only tends towards its sill rather than actually reaching it.

4) **Power Functions:**

\[
\gamma(h) = |h|^\alpha \quad \text{with} \quad 0 \leq \alpha < 2
\]

The linear model, \( \gamma(h) = |h| \), is a special case.

![Power Model](image)

**Figure 1.11**

5) **Gaussian Model:**

Figure 1.11 shows the power models for \( \alpha = 0.5, 1.0 \) and 1.5, and the gaussian model

\[
\gamma(h) = C[1 - \exp(-h^2/a^2)]
\]
The practical range is 1.73a. The gaussian model represents an extremely continuous phenomenon. Experience shows that numerical instabilities often occur when this is used without a nugget effect. Even a small nugget effect leads to a significant improvement.

6) Cubic Model:

This model has a parabolic behaviour at the origin and is generally similar to the gaussian model, except that it is not infinitely differentiable. Its equation is

\[
\gamma(h) = C \left[ 7r^2 - 8.75r^3 + 3.5r^5 - 0.75r^7 \right] \quad \text{if } r < 1
\]

\[
= C \quad \text{otherwise}
\]

where \( r = h/a \).

7) 1D Hole Effect Model:

Please note that this hole effect model is not authorized for spaces with dimension higher than 1D.

\[
\gamma(h) = C \left[ 1 - \exp(-|r|) \cos(2\pi r_2) \right]
\]

where \( r = h/a \) and \( r_2 = h/\lambda \). The value of \( \lambda \) controls the magnitude of the hole effect.

8) 2D Hole Effect Model:

\[
\gamma(h) = C \left[ 1 - \exp(-|r|) J_0(2\pi r_2) \right]
\]

where \( r = h/a \), \( r_2 = h/\lambda \) and \( J_0 \) is the Bessel function. The value of \( \lambda \) controls the magnitude of the hole effect.

1.3 COMPUTING EXPERIMENTAL VARIOGRAMS

The experimental variogram can be calculated using the following formula:

\[
\gamma^*(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i + h) - z(x_i)]^2
\]

where:

- \( x_i \) are the locations of the samples.
- \( z(x_i) \) are the data values.
- \( N(h) \) is the number of pairs of points \((x_i, x_i + h)\); that is the number of pairs of points separated by a distance \( h \), i.e. those which will actually be taken into account.

Data regularly located along lines

Variograms are then computed for the distance lag and its multiples. Averaging the variograms for two or more lines has to be weighted by the number of pairs.

Data irregularly located along lines

Distances are considered up to a given tolerance (distance classes).
Data in 2D

When the data are two dimensional, the variograms should be calculated in different directions to check for anisotropies (see exercise). If the data are arbitrarily spaced, the variograms are calculated for various angular classes as well as for different distance classes. This results in average values of the variogram within each of the given classes.

When data are densely located on lines (acoustic or seismic profiles, geological drillholes), the variogram should be computed along the lines first.

In three dimensions

Data are usually taken along one dimension: for instance along vertical drillholes. In this case a vertical variogram is computed along this vertical direction. Individual horizontal variograms are computed within horizontal slices and then averaged.

Preferential sampling

Structures are better known from a regular grid, or at least from an irregular grid providing it is not preferential. When different zones are sampled with a different grid, variograms should be calculated within each zone.

Example 1:

Use the formula given above to calculate the experimental variogram for the first three distance classes for the data given below. Samples are regularly spaced every 5m along a straight line.

| Grade: 8% 6% 4% 3% 6% 5% 7% 2% 8% 9% 5% 6% 3% |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|

Show that for the first variogram class, we have:

\[ \gamma'(5) = \frac{1}{2 \times 12} \left[ 2^2 + 2^2 + 1^2 + 3^2 + 1^2 + 2^2 + 5^2 + 6^2 + 1^2 + 4^2 + 1^2 + 3^2 \right] \]

\[ = 4.625 \]

Calculate \( \gamma'(10) \) and \( \gamma'(15) \), and plot the graph of \( \gamma'(h) \) against \( h \).

Example 2:

Suppose that one of the values (7%) was missing. The data are now:

| Grade: 8% 6% 4% 3% 6% 5% 2% 8% 9% 5% 6% 3% |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|

Recalculate the variogram and compare it to the previous experimental variogram.

Example 3:

Suppose that one very large value had occurred (17% instead of 7%). Recalculate the experimental variogram. This shows the effect of one large value (called an outlier) on the variogram.
Example 4: Calculating an experimental variogram in 2D

The data shown in Figure 1.12 were taken from a 2-D computer simulation of a RF with a linear variogram $\gamma(h) = |h|$. The data are arranged on a regular grid with a side of "a". You should calculate the experimental variograms in four directions (1 & 2 along the two axes and 3 & 4 at 45 degrees to them) for distances up to 3 lags.

![Data for calculating 2D variogram](image)

Figure 1.12 Data for calculating 2D variogram

Then plot all four experimental variograms. Note that along the diagonals one lag is 1.4142 times that along the axes. As it turns out these variograms are isotropic. Calculate the average experimental variogram for all directions and try to fit a model to it.
Figure 1.13: Experimental variograms for 2D example

<table>
<thead>
<tr>
<th>Mean Variogram</th>
</tr>
</thead>
<tbody>
<tr>
<td>h</td>
</tr>
<tr>
<td>N</td>
</tr>
</tbody>
</table>

A linear model with slope of 4.2/a and with no nugget effect gives a good fit to them. This example is, unfortunately, unrealistically good. In practice when there are as few data as this the variograms usually turn out to be much more erratic than these.

1.4 FITTING A VARIOGRAM MODEL

This is best seen from practical examples. By and large, experience has shown that the analytic form of the model does not matter very much as long as the major features of the phenomenon are respected. These are:

- the nugget effect
- the slope at the origin
- the range
. the sill

. the anisotropies

The behaviour at the origin (both the nugget effect and the slope) plays a crucial role in the fitting of the variogram because it has a tremendous influence on the results of the kriging and also on the numerical stability of the kriging system. The slope can be assessed from the first three or four variogram values; while the nugget effect can be estimated by extrapolation back to the origin.

The range can usually be assessed visually. If there is more than one range the intermediate ranges can be distinguished visually because the variogram has a change in its curvature at these points.

Modelling anisotropies requires more experience. It is often difficult to distinguish between a drift and an anisotropy.

In general a good fit can be obtained with the sum of two or three models. Using more than this increases the subsequent computing costs considerably and should be avoided. The fitting is done by trial and error, preferably with a graphic terminal.

1.5 TROUBLESOME VARIOGRAMS

The experimental variograms found in practice are often much more erratic than the examples presented in textbooks and journal articles. Since the causes of potential problems are extremely numerous and varied, it would be impossible to present them all. However it is worthwhile showing some of the more common ones, because this makes people more aware of the traps. A more complete list of these problems is given in Armstrong (1982).

Pseudo-periodic hiccups

The variogram calculated for a lag of 40m for ash data from the Bowen Basin illustrates this problem quite nicely. Figure 1.14.
At first one might wonder if the two rather strange peaks at n = 160m and h = 280–300m are of some sort of periodicity in the coal but this is not physically likely given the nature of the samples locations. Although this is not obvious from the plot of the distances between pairs of points shows that this was clearly the case. In this case the solution is quite simple. All we need do is to change the step length to some other than a submultiple of 160m. Calculating the variograms for 100m distance classes smoothes out the bumps.

Figure 1.15: Histogram of the number of pairs of points for each distance class.
Outliers

As was seen in the third exercise on calculating experimental variograms, the presence of even one outlier can lead to a highly erratic variogram. In a study on coal in the Bowen Basin in Australia, the variograms for three of the variables were very similar but the sulphur variograms for two seams were totally different. Figure 1.16. This was rather surprising. A closer examination of this data showed that it contained two extremely high sulphur values among a group of 207 values. (These can be seen quite clearly on a histogram). Moreover these outliers lie in a geographically distinct area that should be treated separately from the rest. When these two aberrant values were removed the variogram dropped back to about one fifth of its previous values and looked just like the sulphur variogram for the other seam. Removing additional points made no significant difference to the variogram.

In other cases (notably highly skew distributions such as gold or uranium grades) it is not quite so easy to find a good way of estimating the variogram. The high grade samples are not usually in a separate area. They are usually mixed in amongst lower grade material. More importantly the small percentage of high grades is often what makes the difference between opening the mine and not doing so.

![Variogram for top seam with 2 outliers included](image1)

![Variogram for lower seam](image2)

![Variogram for top seam with the outliers excluded](image3)

Figure 1.16: Sulphur variograms for top seam with and without the 2 outliers and for lower seam.

Artefacts

The next two examples have been included to point out that many of the problems with variograms are due to “operator error” rather than statistical problems with the data. Figure 1.17 shows a remarkable “saw-tooth” variogram. The variable under study was a highly variable quantity such as uranium grade, where there was a large percentage of values below the recording threshold which had all been recorded as zero. Since the data looked log-normal it seems advisable to take logs.
To avoid problems with the zeros these were arbitrarily set to 0.00001 and so became -5.0 after the log transformation. As all the other logs lay in the range from -3.0 to + 3.0 the -5.0's were then extreme values and their locations completely determined the shape of the variogram.

![Saw-tooth variogram](image.png)

Figure 1.17: Saw-tooth variogram

The second example is on the same lines. In this case the problem was caused by the presence of a number of zero values for the seam thickness. These could indicate that the seam has petered out in this region but as this seems unlikely the data were re-examined. This revealed that the seam thickness corresponded with missing values for both the top and bottom of the seam. As a student doing the study forgot to include a test for missing values in his program to calculate seam thickness, the subtraction of 0 from 0 resulted in a goodly number of zeros, which determined the form of the variogram.

These mistakes are rather silly. On reading them everyone naturally feels that he would not make such a mistake but experience shows that these types of errors are much commoner than most of us care to admit. What is more the only way to find out their cause and treat them is by playing around with the data.
2. DISPERSION VARIANCE AND REGULARIZATI

2.1 VARIANCE OF A POINT WITHIN A VOLUME

In our model, the variable under study is considered as a realization $z(x)$ of an intrinsic function $Z(x)$. If all the values within the volume (or area in 2D) $V$ were available it would be possible to find the mean over this volume

$$z(V) = \frac{1}{V} \int_V z(x) \, dx$$

and also the variance of the values within this volume

$$\sigma^2 (0|V) = \frac{1}{V} \int_V (z(x) - z(V))^2 \, dx$$

(Here $0|V$ denotes a point within $V$). If we let the realization vary, the variance of $z(x)$ can be obtained as the expected value of $\sigma^2 (0|V)$ over all possible realizations:

$$\sigma^2 (0|V) = E \{ \sigma^2 (0|V) \}$$

It can be shown that this variance is related to the variogram by the formula:

$$\sigma^2 (0|V) = \frac{1}{V^2} \int \int \gamma (x - y) \, dx \, dy$$

This integral is the average obtained by varying $x$ and $y$ independently throughout the space. It is therefore denoted by $\gamma(V, V)$. This gives

$$\sigma^2 (0|V) = \gamma (V, V)$$

The stationary case

If the RF is intrinsic with an unbounded variogram, e.g. $\gamma(h) = |h|$, this dispersion increases when we consider a larger and larger field $V$.

In the stationary case we get:

$$\sigma^2 (0|V) = C(0) - \bar{\gamma}(V, V)$$

If the field $V$ gets large compared to the range, $\bar{\gamma}(V, V) \to 0$, and the dispersion vanishes as the limit $C(0)$ which is the a priori variance in the model. Moreover we have

$$\bar{\gamma}(V, V) = \text{Var} \, Z(V) = \text{Var} [Z(V) - m]$$

and, if $V$ is large, its average $Z(V)$ equals the mean $m$. In other words, the parameter $\gamma$ is related to the average of $Z(x)$ on a large enough field.
2.2 SUPPORT OF A REGIONALIZED VARIABLE

In many cases a regionalized variable is defined as the average over a certain volume or a surface rather than at a point. The basic volume on which a regionalized variable is defined is called its support. If we change the support we obtain a new regionalized variable which is related to the preceding one but which has different structural characteristics. In mining evaluation for example the grade measured on 2" cores (i.e. with a core as support) has a higher variance than the same variable measured on larger diameter cores or on blocks (i.e. on a larger support). The problem is to know how one variable is related to the other. In other words, what can we say about the grade of blocks knowing the grade of cores?

The answer will be given in two stages. First we consider the dispersion of the values as a function of the support. Then we see how their variograms are related.

2.3 DISPERSION VERSUS BLOCK SIZE

Table 2.1 gives the grades of the 64 adjoining 1m x 1m blocks and also the average yields for the 16 2m x 2m blocks obtained by averaging 4 adjacent 1m x 1m blocks. (These values were actually taken from about 1000 values of millet studied by Sandjivy (1980)). As an exercise calculate the mean yield, the variance and the histogram of values for both sets of data.

As expected the means are the same (201) except for differences due to roundoff. But the variances are not. The variance for the larger support is smaller. If the values were statistically independent of each other the variance in the second case would be ¼ of the first one. However because of the correlations it is higher. The second thing that we see is that the shape of the histogram has changed. The second set of values is much less spread out and is less skew. The implications of this change are very important in mining. We often want to predict the percentage of blocks with a grade above a certain critical value (that is, the recoverable ore tonnage).

As we saw earlier when the polygonal method is used for reserve estimation, the grade of the sample inside the polygon is taken as the grade estimate for the whole polygon, which is the same as equating the histogram of core grades with that of blocks. This leads to serious errors in estimating the recoverable reserves because the histogram of block values is inevitably less spread out than that of the samples, as can be seen by comparing figures 2.1 and 2.2.

We shall now go on to see how to evaluate the variance of blocks given the variogram of core values. All that is needed is a small computer program (or the appropriate tables).
### 1 x 1m blocks

<table>
<thead>
<tr>
<th></th>
<th>735</th>
<th>325</th>
<th>45</th>
<th>140</th>
<th>125</th>
<th>175</th>
<th>167</th>
<th>485</th>
</tr>
</thead>
<tbody>
<tr>
<td>540</td>
<td>420</td>
<td>260</td>
<td>128</td>
<td>20</td>
<td>30</td>
<td>105</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>450</td>
<td>200</td>
<td>337</td>
<td>190</td>
<td>95</td>
<td>260</td>
<td>245</td>
<td>278</td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>250</td>
<td>380</td>
<td>405</td>
<td>250</td>
<td>80</td>
<td>515</td>
<td>605</td>
<td></td>
</tr>
<tr>
<td>124</td>
<td>120</td>
<td>430</td>
<td>175</td>
<td>230</td>
<td>120</td>
<td>460</td>
<td>260</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>135</td>
<td>240</td>
<td>35</td>
<td>190</td>
<td>135</td>
<td>160</td>
<td>170</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>95</td>
<td>20</td>
<td>35</td>
<td>32</td>
<td>95</td>
<td>20</td>
<td>450</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>35</td>
<td>100</td>
<td>59</td>
<td>2</td>
<td>45</td>
<td>58</td>
<td>90</td>
<td></td>
</tr>
</tbody>
</table>

### 2 x 2m blocks

<table>
<thead>
<tr>
<th></th>
<th>505</th>
<th>143</th>
<th>88</th>
<th>207</th>
</tr>
</thead>
<tbody>
<tr>
<td>270</td>
<td>328</td>
<td>171</td>
<td>411</td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>220</td>
<td>154</td>
<td>263</td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>54</td>
<td>44</td>
<td>155</td>
<td></td>
</tr>
</tbody>
</table>

*Table 2.1 Grades of the 64 1m x 1m blocks, and the 16 2m x 2m ones*
Figure 2.1: Histogram of the grades of the 64 1m x 1m blocks

Figure 2.2: Histogram of the grades of the 16 2m x 2m blocks
2.4 VARIANCE OF $v$ WITHIN $V$

We now consider the new variable defined as the spatial average of $Z(x)$ over a certain volume $v$:

$$Z(v) = \frac{1}{v} \int_V Z(x) \, dx$$

We want to find the dispersion of this new variable $Z(v)$ as it moves over a larger volume $V$. Typically $v$ could represent a core while $V$ could be a block; or $v$ could be a block and $V$ could be the whole deposit.

The variance of $v$ within $V$ is denoted by $\sigma^2(v|V)$ and is given by:

$$\sigma^2 (v|V) = E \left[ \frac{1}{V} \int_V (Z(v) - Z(V))^2 \, dv \right]$$

Expanding this gives:

$$\sigma^2 (v|V) = \frac{1}{V^2} \iint (x - y) \, dx \, dy - \frac{1}{V^2} \iint (x - y) \, dx \, dy$$

i.e.

$$\sigma^2 (v|V) = \overline{\gamma} (V|V) - \overline{\gamma} (v, v)$$

2.5 THE ADDITIVITY RELATION

Combining the different results we get

$$\sigma^2 (v|V) = \overline{\gamma} (V, V) - \overline{\gamma} (v, v)$$

$$= \sigma^2 (0|V) - \sigma^2 (0|v)$$

This formula can be generalized to any three volumes $v, V$ and $V'$ where: $v \subset V \subset V'$:

$$\sigma^2 (v|V') = \sigma^2 (v|V) + \sigma^2 (V|V')$$

For example,

$v =$ core section
$V =$ block
$V'$ = deposit
In this case the formula can be interpreted as “the variance of a core section within the deposit is equal to that of a core within a block plus the variance of a block within the deposit”. This can be checked experimentally for the millet values given earlier. Here $v$ corresponds to a 1m x 1m block while $V$ corresponds to a 2m x 2m block. From before we have:

$$\sigma^2 (v|V) = \sigma^2_{1x1} = 27,592$$

$$\sigma^2 (V|V') = \sigma^2_{2x2} = 16,641.$$  

A simple calculation gives 10,951 as the value of $\sigma^2 (v|V)$, which satisfies the additivity relation.

The variance of $Z(v)$ is less than the variance of $Z(x)$. We say that $Z(v)$ is the regularisation of $Z(x)$ on the support $v$.

The drop of variance when going from a point to a block $v$ is

$$\sigma^2 (0|V) = \overline{\gamma}(V,V)$$  

i.e. the "within-block" variability. This is very important when having to predict the future distribution of mining selective blocks, knowing the distribution of point sample values. We know that the mean over the deposit $V$ is the same, and this gives the way to compute the variance. However mean and variance are not enough to define a distribution. We use for that a change of support model (this is part of nonlinear geostatistics, very important for mining, but it will not be seen in this course).

### 2.6 Regularization of Variogram

It can be shown that the variogram of the regularized variable $Z(v)$ is:

$$\gamma_r(h) = \overline{\gamma}(v,v_h) - \overline{\gamma}(v,v)$$

where $v_h$ denotes the support $v$ moved through $h$ (translated by the vector $h$), and $\gamma(v,v_h)$ represents the average value of the variogram between an arbitrary point in $v_h$ and another in $v$.

When the distance $h$ is large compared to the size of $v$, the mean variogram value $\gamma(v,v)$ is approximatively equal to $\gamma(h)$. So we obtain the relation:

$$\gamma_r(h) = \gamma(h) - \overline{\gamma}(v,v)$$
Figure 2.3 Point and regularised variograms.

Figure 2.4 Experimental variograms for millet for three support sizes
Remarks

1) The relations between the different variances and the variogram hold also outside the RF framework.

2) In practice the variogram model may not be valid for the large distances. If so, the theoretical variances which are computed using the variogram for these distances will not be representative.

2.7 EXERCISES

Exercise 1: dispersion variances

Consider a spherical variogram with sill 2.0 and range 2500 m: \( \gamma(h) = 2.0 \text{ Sph} \ (2500 \text{ m}) \).

a) Calculate the dispersion variance of a point within
   - a 500 x 500 square;
   - a 1000 x 1000 square;
   - a 2000 x 2000 square;
   - a 25000 x 25000 square;
   - an infinite field.
   (Use the chart on p 93 in Matheron 1971).

b) Calculate the dispersion variance
   - of a 500 x 500 square within a 2000 x 2000 square;
   - of a 500 x 500 square within a 25000 x 25000 square;
   - of a 2000 x 2000 square within a 25000 x 25000 square;

c) Same questions with a supplementary nugget component equal to 1.0?

Exercise 2: nested structures

Consider the variogram

\[
\gamma(h) = 2.0 \text{ Nug} + 2.0 \text{ Sph} \ (500 \text{ m}) + 2.0 \text{ Sph} \ (2500 \text{ m}) + 2.0 \text{ Sph} \ (10000)
\]

which is the sum of a nugget component and 3 spherical structures with range 500 m, 10000 m. All the components correspond to the same sill 2.

a) What is the dispersion variance of a point within an infinite field? How is dispatched into the different components?

b) Calculate the dispersion variance of a point within a 2000 x 2000 square. How is dispatched into the different components?

(Use the chart on p 93 in Matheron 1971).

Exercise 3: anisotropy

Consider a spherical variogram with sill 2.0 and range 2500 m in EW and 1000 in anisotropy).
Calculate the dispersion variance of a point within
- a 500 x 2000 rectangle;
- a 2000 x 500 rectangle;
Compare the results.

(Use the chart on p 93 in Matheron 1971).

**Exercise 4: regularized variogram**

Consider the linear variogram |h| in a 1D space.
1) Dispersion variance on a segment $l$?
2) Regularized variogram for segments $l$?
3) What happens when adding a nugget effect?
GRAPH 7 - SPHERICAL SCHEME

Function \( \frac{1}{C} F \frac{\ell}{a} \cdot \frac{h}{a} \) = variance of a point within the rectangle \( \ell \times h \). (i.e. for \( C = 1 \))
3. ESTIMATION VARIANCE AND GLOBAL ESTIMATION

3.1 ESTIMATION VARIANCE

Suppose that we want to estimate the average value inside a region $V$:

$$Z(V) = \frac{1}{V} \int_{V} Z(x) \, dx$$

by using the known value on a support $v$ to estimate it:

$$Z(v) = \frac{1}{v} \int_{v} Z(x) \, dx$$

What error is made in doing this? First of all, if $Z(x)$ satisfies the stationary or the intrinsic hypotheses, the estimation is unbiased: the expectation of the estimation error $Z(V) - Z(v)$ is zero.

The variance of this error is what geostatistics calls the estimation variance. As we extend the value on $v$ to the region $V$, it is sometimes also called extension variance. It is usually denoted by $\sigma_k^2(v, V)$ or $\sigma_k^2$ for short. The estimation variance writes

$$\sigma_k^2(v, V) = \text{Var} [Z(V) - Z(v)]$$

$$= \frac{2}{vV} \int_{v} \int_{V} \gamma(x-y) \, dx \, dy$$

$$- \frac{1}{V^2} \int_{V} \int_{V} \gamma(x-x') \, dx \, dx' - \frac{1}{V^2} \int_{V} \int_{V} \gamma(y-y') \, dy \, dy'$$

or in short

$$\sigma_k^2(v, V) = 2 \bar{\gamma}(v, V) - \bar{\gamma}(v, v) - \bar{\gamma}(V, V) \quad [3.1]$$

where $\bar{\gamma}(v, V)$, $\bar{\gamma}(v, v)$ and $\bar{\gamma}(V, V)$ are the average variogram values when the end points of the vector $h$ sweep independently through $V$ and $v$ etc ...

\[\bar{\gamma}(v, V)\]  
\[\bar{\gamma}(V, V)\]  
\[\bar{\gamma}(v, v)\]
The factors influencing the estimation variance are:

- the geometry of \( V \) (through \( \gamma(V,V) \))
- the geometry of \( v \) (through \( \gamma(v,v) \))
- the location of \( v \) relative to \( V \) (through \( \gamma(v,V) \))
- the regularity of the variable (through \( \gamma \))

The estimation variance involves the variogram but not the actual sample values. Providing the variogram model is known, different sampling strategies can be tested.

This formula can be rewritten as:

\[
\sigma_E^2(v, V) = \left[ \mathcal{P}(v, V) - \mathcal{P}(V, V) \right] + \left[ \overline{\mathcal{P}}(v, V) - \overline{\mathcal{P}}(v, v) \right]
\]

This makes it clear that the variance decreases when

- the sample \( v \) is more representative of the region \( V \) to be estimated. In the limit when \( v = V \), \( \sigma_E^2(v, V) = 0 \).
- the variogram is more regular, i.e. the variable is more continuous.

People often tend to confuse the dispersion variance \( \sigma^2(v \mid V) \) with the extension variance \( \sigma_E^2(v, V) \). The dispersion variance has a physical meaning: it measures the dispersion of the samples of a given volume \( v \) within another volume \( V \). In contrast to this, the extension variance is an operational concept characterizing the error associated with a particular sampling pattern. Theoretically, the two types of variance are related in the following way: the dispersion variance is the average of the extension variance \( \sigma_E^2(v, V) \) when the sample \( v \) describes \( V \).

The formula [3.1] holds for any shape of \( v \) and \( V \). In particular \( v \) does not have to be included in \( V \). But also the support \( v \) may be the reunion of \( N \) sample points \( x_i \) in a given geometric configuration. It is also possible to give different weights to the sample points, as will be seen in kriging. This can be useful either for mapping a region or estimating the average over a region from irregularly spaced data.

### 3.2 GLOBAL ESTIMATION

Here we will consider the case where we estimate the region from the arithmetic mean of the data. The estimator of \( Z(V) \) is then

\[
Z(V)^* = \frac{1}{N} \sum Z(x_i)
\]

The corresponding estimation variance

\[
\sigma_E^2 = E \left[ Z(V) - \frac{1}{N} \sum Z(x_i) \right]^2
\]

can be written as

\[
= \frac{2}{N} \sum \overline{\mathcal{P}}(x_i, V) - \overline{\mathcal{P}}(V, V) - \frac{1}{N^2} \sum \gamma(x_i, x_j)
\]
where \( \bar{\gamma}(x_i, V) \) is the average of the variogram between \( x_i \) and the volume \( V \)

\[
\bar{\gamma}(x_i, V) = \frac{1}{V} \int_V \gamma(x_i, x) \, dx
\]

**Remark**

One must distinguish between:

- the estimation variance, which is the variance of the error \( Z(V) - \frac{1}{N} \sum Z(x_i) \)

  (and exists provided the variogram does since \( 1 - \frac{1}{N} \sum 1 = 0 \));

- and the variance of the estimator \( \frac{1}{N} \sum Z(x_i) \)

  (which only exists in the stationary case and equals \( \frac{1}{N^2} \sum \sum C(x_i, x_j) \)).

The estimation variance for \( Z(V) \) depends on \( V \) as well as the data points \( x_i \). It gives the precision of the estimation.

The variance of the estimator depends only on the data points \( x_i \). It corresponds to the dispersion variance of the sampling configuration within an infinite field. It indicates how variable we can expect the estimating value to be on different realizations. As

\[
Var \left( \frac{1}{N} \sum Z(x_i) \right) = Var \left[ m - \frac{1}{N} \sum Z(x_i) \right]
\]

it is also the estimation variance when estimating the mean parameter \( m \) by \( \frac{1}{N} \sum Z(x_i) \).

If the field is large compared to the range, then \( Z(V) \) equals the mean parameter \( m \) of the stationary model, and both variances are the same. In general one must be very careful because global estimation concerns the average \( Z(V) \) on the field \( V \), not the mean parameter \( m \).

**Random sampling**

In this case we have \( \gamma(x_i, V) = \gamma(V, V) = \gamma(x_i, x_i) \) \( i \neq j \) and \( \gamma(x_i, x_i) = 0 \). We find

\[
\sigma_e^2 = \frac{1}{N} \gamma(V, V) = \frac{1}{N} \sigma^2(0|V)
\]

as expected by classical statistics and could be found directly.

**Stratified sampling**

There might be confusion about this term since its current meaning is not the same in fisheries and geostatistics. In fisheries, strata correspond to a partition of the region into zones which have, in principle, different characteristics. Thus it can be a way to divide a nonstationary field into homogeneous strata.

In geostatistical jargon, the stratified sampling corresponds to the case where the region \( V \) is the union of \( N \) equal blocks \( v_j \) within each of which one randomly located sample is selected. There
is no prior knowledge about the blocks. The whole region is only supposed to be represented by its variogram.

Let $Z(v_i)$ be the true but unknown average over the $i$th block. The true average over the whole of $V$ is then:

$$Z(V) = \sum Z(v_i) / N$$

The error we make in taking the average of the $n$ sample values $Z_i$ as the estimator of $Z(V)$ is then:

$$\sum [Z(v_i) - Z_i] / N$$

that is, it is the average of the partial errors. As each sample is taken within its block independently from the others, the partial errors can be assumed to be uncorrelated, and the estimation variance is

$$\sigma_k^2 = \frac{1}{\bar N} \sigma^2(0|v) = \frac{1}{\bar N} \gamma(v,v)$$

Since $\sigma^2(0|V) - \sigma^2(0|v) = \sigma^2(v|V)$ is $>0$, stratified sampling is always better than pure random sampling.

**Regular grid with a random origin**

In this case we would have $\frac{1}{\bar N} \sum \gamma(x_i, V) = \gamma(V, V)$ and the estimation variance can be written as

$$\gamma(V, V) - \frac{1}{\bar N^2} \sum \sum \gamma(x_i, x_j)$$

**3.3 CENTERED REGULAR GRID**

**Direct composition of terms**

The region $V$ is the union of $N$ equal blocks $v_i$, and a sample $x_i$ is taken at the center of each block. If $N$ is large, the basic formula for the estimation variance becomes very unwieldy. In practice we generally use straightforward approximations for the basic formulae.

The error we make in taking the average of the sample values as the estimator of $Z(V)$ is:

$$\sum [Z(v_i) - Z(x_i)] / N$$

that is, it is the average of the partial errors, when estimating each block by its center. The approximation consists in considering that these partial errors are not correlated. We then find that the global estimation variance
\[ \sigma^2_E = E \left[ Z(V) - \frac{1}{N} \sum Z(x_i) \right]^2 \]

equals the extension variance of a sample to its block divided by the number of samples

\[ \sigma^2_E = \frac{1}{N} \sigma^2 (0, v) = \left[ 2 \gamma (0, v) - \gamma (0, 0) - \gamma (v, v) \right] / N \]

This formulae can be used in 1D: the region becomes a line, blocks are segments with one sample at their center.

In 2D, this method based on the direct composition of terms is valid only when blocks are roughly square. If the ratio of the length to the breadth is large the following approximation should be used. When deciding which approximation principle to use, the anisotropy in the variogram must be taken into account. The ratio of length to breadth should be calculated into terms of the variogram range rather the distance units.

**Composition by line and slice terms**

The method presented above is used when the samples are evenly spread in space. It is not suitable if the data are much more dense in one direction than the other, as for example happens in fisheries, seismic surveys, or sometimes in underground development sampling. See Figure 3.1.

---

Figure 3.1 : Samples along seismic profiles or galleries

When data are closely spaced along lines that are widely spaced, another approximation method must be used to calculate the estimation variances. It involves combining the errors made when extrapolating the sample values along the lines, and then extrapolating the line values out into the slices around them.

![Diagram of sample division](image)

\[ V = v_1 + v_2 \]

Figure 3.2: The area \( V \) has been sampled along two profiles \( d_1 \) and \( d_2 \), and has been divided into two slices (subareas) \( v_1 \) and \( v_2 \) each surrounding a profile.
At first we assume that the line sections $d_i$ have been analysed accurately, so we know $Z(d_i)$ exactly. To estimate $Z(V)$ we have to weight each $Z(d_i)$ by its volume $v_i$. If the distance between lines of samples is constant, $v_i$ is proportional to $d_i$. So we have

$$Z^*(V) = \frac{\sum_i v_i Z(d_i)}{\sum v_i} = \frac{\sum_i d_i Z(d_i)}{\sum d_i}$$

Now the true but unknown grade $Z(V)$ can be written as

$$Z(V) = \frac{\sum_i v_i Z(v_i)}{\sum v_i} = \frac{\sum_i d_i Z(v_i)}{\sum d_i}$$

So the estimation error made is the weighted average of the elementary estimation errors $Z(d_i) - Z(v_i)$:

$$Z^*(V) - Z(V) = \frac{\sum_i d_i [Z(d_i) - Z(v_i)]}{\sum d_i}$$

By the approximation principle the estimation variance is

$$\sigma_k^2 = Var [Z^*(V) - Z(V)] = \frac{\sum_i d_i^2 \sigma_k^2}{(\sum d_i)^2}$$

where $\sigma_k^2$ is the elementary extension variance of the central line section to its block of influence. Note that these are weighted by the squares of the lengths $d_i$.

More realistically, the line sections are themselves obtained by averaging the sample grades along the line section. To simplify the calculations, suppose that these samples are equally spaced $s$ apart, which is usually the case. So each estimate, $Z^*(d_i)$, is obtained as the average of samples $n_i = d_i / s$:

$$Z^*(d_i) = \sum_k Z(s_k) / n_i$$

The total estimation error can be split into two terms:

$$Z^*(V) - Z(V) = \frac{\sum_i d_i [Z(d_i) - Z(v_i)]}{\sum d_i} + \frac{\sum_i d_i [Z^*(d_i) - Z(d_i)]}{\sum d_i}$$

The left hand term corresponds to the extension of the line sections to the surrounding blocks while the right hand one corresponds to the extension of the samples to the line section. If we let $Z(s_k)$ and $Z^*(s_k)$ be the true and the estimated grades of the $k$th segment in the $i$th section, then the second term can be rewritten as

$$\frac{\sum_i d_i (1/n_i) \sum_k [Z^*(s_k) - Z(s_k)]}{\sum d_i}$$
The covariances are considered to be zero so that variances add up. We finally get:

\[
\sigma_k^2 = \frac{\sum d_i^2 \sigma_k^2}{(\sum d_i)^2} + \frac{\sum d_i^2 \sigma^2(O,s)}{(\sum d_i)^2} / n_i
\]

If \( N \) denotes the total number of samples \( N = \sum n_i \) this gives

\[
\sigma_k^2 = \frac{\sum d_i^2 \sigma_k^2}{(\sum d_i)^2} + \frac{\sigma^2(O,s)}{N}
\]

So the total estimation variance is the composition of a block term that accounts for the extension of the line sections to the block and a line term \( \frac{\sigma^2(O,s)}{N} \) that accounts for the error made when extending the samples to the line section.

In fish acoustic surveys, the data are dense along the profiles. But each datum already corresponds to the regularized value over an elementary segment (e.g. 1 min.). Thus their average on a profile gives already the value of the line, and there is no line term in the estimation variance.

Remark

The approximation principles amount to consider that the errors when estimating each block by its own sample have no correlation. But of course the block values themselves and/or the sample values can be (and generally are) correlated.

The principle can also be used, to a certain extent, to various configurations: partition into irregularly shaped zones, with one or more sample within each zone, providing we consider the errors when estimating each zone by its own samples only. If the different zones definitely do not have the same variogram, the variogram for each zone should be used.

3.4 TRANSITIVE THEORY

The transitive theory is a part of linear geostatistics but is not formulated in terms of Random Functions. It only provides linear and global estimates, but its validity does not require any stationarity hypothesis.

In the better known intrinsic theory, we consider the behaviour of the Regionalized Variable within its field. The limits of the field are supposed to be known and the behaviour of the Regionalized Variable is independent of the geometry and the borders of this field. Then we consider that we see a realization of a Random Function throughout the field–window.

In contrast to this, transitive theory does not require a distinction between the geometry of the field and the variations of the Regionalized Variable within the field. It just supposes that the values of the Regionalized Variable are known at the nodes of a regular grid. Outside the field, these values are zero.

Counting the number of non zero values gives an estimate of the area. Transitive theory gives an estimation variance for this surface, which quantifies the geometrical uncertainty.
When applying the intrinsic theory we have to fix the field, but the uncertainty on the geometry can be taken into account by adding a geometric term obtained from the transitive theory to the global estimation variance.

But adding the sample values gives directly an estimate of the total abundance, whatever the exact limits of the field. And the estimation variance can be computed by transitive methods. This is another way to compute the global estimation variance. This is particularly interesting when the behaviour of the variable is obviously influenced by the proximity of the borders of its field.

**Surface**

Let us imagine a mining deposit, or a fish school, which is sampled on a regular grid (Fig 3.3). We then define the indicator \( k(x) \) at each node of the grid to be equal to 1 if the point is within the school. At all other nodes the value is (or is supposed to be) 0. The exact limits of the school are unknown. But counting the positive \( k(x) \) values gives an estimate of its area.

To simplify the formula we will use 1D notation:

\[
\begin{align*}
x_0 & \quad \text{one node of the grid taken as its origin} \\
a & \quad \text{grid mesh} \\
x_0 + pa & \quad \text{the nodes of the grid } (-\infty < p < +\infty).
\end{align*}
\]

The estimator of the area can be written as:

\[
S^* (x_0) = a \sum_{-\infty}^{+\infty} k(x_0 + pa)
\]

Its exact value is

\[
S = \int k(x) \, dx
\]

The basic hypothesis in transitive theory is that the origin of the grid is random. So \( S^* (x_0) \) is an unbiased estimator of \( S \). The variance of the error \( S^* (x_0) - S \) can be written

\[
\sigma^2 = a \sum_{-\infty}^{+\infty} K(pa) - \int K(h) \, dh
\]

(which is also the variance of \( S^* (x_0) \) as there is nothing random in \( S \)). In this formula \( K(h) \) is the geometric covariogram of the area \( S \)

\[
K(h) = \int k(x) \, k(x + h) \, dx
\]

It is zero everywhere except where point \( x \) and \( x + h \) both belong to \( S \). So \( K(h) \) gives a measure of the intersection between \( S \) and \( S \) translated by \( -h \). The variance is all the larger as the area is irregularly shaped, and as the grid is large.
Figure 3.3: Data layout

N* of positive holes = 17
N* of horizontal sides = 12
N* of vertical sides = 10

In practice Matheron has developed an approximation formula for the geometric error variance relative to the mineralized area S:

\[
\frac{\sigma^2}{S^2} = \frac{1}{N^2} \left[ \frac{N_2}{6} + 0.061 \frac{(N_1)^2}{N_2} \right]
\]

where \(2N_1\) and \(2N_2\) are found by counting the number of elementary sides parallel to the axes of the sampling grid, where \(N_1\) is greater than or equal to \(N_2\), and where \(N\) is the total number of positive samples. From this we obtain for our figure

\[
\frac{\sigma^2}{S^2} = \frac{1}{17^2} \left[ \frac{5}{6} + 0.061 \frac{(6)^2}{5} \right] = 0.0044 = (6.6 \%)^2
\]

Use in intrinsic theory

In intrinsic theory we compute the estimation variance \(\sigma^2_k\) of the average value over a field which has to be fixed. However the exact field may not be known. In that case a geometric term has to be added to this variance. It is \(\frac{\sigma^2}{S^2} D^2(O|S)\). For details see the worked examples given in Journel and Huijbregts pp 428 – 438.

If we consider the global abundance \(Q = Z_s S\) instead of the average density \(Z_s\), the estimation variance \(\frac{\sigma^2}{Q^2}\) can be approximated by

\[
\frac{\sigma^2}{Z_s^2} + \frac{\sigma^2}{S^2} \left(1 + \frac{D^2(O|S)}{Z_s^3} \right)
\]
Global abundance

The same way that we obtained a direct estimate for a surface and its variance, can be used for
the global abundance. We then use the value of the fish density $f(x)$ at each node of the grid. This
value is zero outside the school which has to be estimated. Counting the positive $f(x)$ values gives
an estimate of the abundance

$$Q^\ast(x_0) = a \sum_{-a}^{+a} f(x_0 + pa)$$

while its exact value is

$$Q = \int f(x) \, dx$$

As the origin of the grid is random, $Q^\ast(x_0)$ is an unbiased estimator of $Q$. The variance of the
error $Q^\ast(x_0) - Q$ is also the variance of $Q^\ast(x_0)$ as there is nothing random in $Q$. It can be written

$$\sigma_0^2 = a \sum_{-a}^{+a} g(pa) - \int g(h) \, dh$$

In this formula $g(h)$ is the transitive covariogram of our Regionalized Variable

$$g(h) = \int f(x) f(x + h) \, dx$$

Like $K(h)$, this covariogram is zero for distances that exceed the size of the school. But there is
no distinction between the behaviour of the variable within its positive field, and the geometry of
this field.

In practice approximation formulae have been developed to express the variance according to the
behaviour of the covariogram at its origin. See Matheron 1965 or 1971 for more details.

There are also formulae in the case the grid is not regular but is random stratified (in its usual
geostatistical sense).

The above formula is general and is valid in 1, 2 or 3D. In the evaluation of mining veins or layers
it is current use to accumulated the grades along holes that are perpendicular to the extension.
Estimations are then carried out on thickness and accumulation. This procedure should be very
interesting for acoustic fish surveys, when transects are parallel. 2D fish acoustic densities should
then have been accumulated along transects, leading to a 1D estimation problem. See Petitgas 1992.
3. 5 EXERCISES

Exercise : a segment estimated by one or two points

Consider the variogram $\gamma(h) = h^4$ $(0 < \lambda < 2)$ in a 1D space.
1) Calculate the average variogram $\bar{\gamma}(l) = \bar{\gamma}(0, l)$ between a point which describes a segment l and one of its extremities.
2) Calculate the average variogram $F(l) = \gamma(l, l)$ over a segment l.
3) Deduce the estimation variance when estimating a segment l by its center.
4) Calculate the average variogram over two points distant of l.
5) Deduce the estimation variance when estimating a segment l by its two extremities.
6) Compare the estimation variances for the configurations 3) and 5). Which is the more precise?

Exercise : a line estimated from point sampling

As above, consider the variogram $\gamma(h) = h^4$ $(0 < \lambda < 2)$ in a 1D space.
Compute the estimation variance of the segment $L = n \Delta$ from n point samples, in the following three cases:
- centered regular grid;
- stratified random sampling;
- purely random grid.
(Use the results of the previous exercise: segment estimated by one point).

Exercise : a square estimated by one point

Consider a spherical variogram with sill 2.0 and range 250 m: $\gamma(h) = 2.0 \text{ Sph} (2500 \text{ m})$
and a 2000 x 2000 m square block $V$.
Compute the estimation variance:
- when estimating the block $V$ by its center $x$;
- when estimating the block $V$ by a random sample within the block.
(Take $\bar{\gamma}(V,V) = 1.13$ and $\gamma(x,V) = 0.88$)
What happens with a supplementary nugget component equal to 0.5?

Exercise : a field estimated from a square grid

Consider a spherical variogram with sill 2.0 and range 2500 m: $\gamma(h) = 2.0 \text{ Sph} (2500 \text{ m})$
and a 40000 x 10000 m field divided into 2000 x 2000 m square blocks.
Compute the estimation variance of the field in the two cases:
- regular grid centered in the blocks;
- stratified random grid.
(Use the results of the exercise: square estimated from one point)
What happens with a supplementary nugget component equal to 0.5?
Exercise: a block estimated from line

Consider a spherical variogram with sill 2.0 and range 2500 m: \( \gamma(h) = 2.0 \text{ Sph (2500 m)} \)
Compute the estimation variance of a block \( v(L, d) \) from its central line \( d \) in the following cases:

\[
\begin{array}{cc}
L & d \\
10000 & 2000 \\
2000 & 10000 \\
2000 & 6000 \\
6000 & 2000 \\
\end{array}
\]

(Use the chart p 95 in Matheron 1971)

What happens if a nugget component equal to 0.5 is added?

Exercise: a field estimated from lines

Consider a spherical variogram with sill 2.0 and range 2500 m: \( \gamma(h) = 2.0 \text{ Sph (2500 m)} \)
and a field made of two blocks \( v_1 \ 10000 \times 2000 \) and \( v_2 \ 6000 \times 2000 \).
Compute the estimation variance of the field knowing the central line \( d_1 \ 10000 \) and \( d_2 \ 6000 \) of the blocks.
(Use the results of the exercise: block estimated from line)

What happens with a supplementary nugget component equal to 0.5 ?

Exercise: block estimated from line, anisotropy

Consider a spherical variogram with sill 2.0 and range 2500 m in EW and 1000 m in NS (geometric anisotropy).

a) Compute the estimation variance of a block \( v(l, d) \) knowing its EW central line \( l \) in the following cases:

\[
\begin{array}{cc}
l & d \\
2000 & 2000 \\
2000 & 6000 \\
6000 & 2000 \\
\end{array}
\]

b) The same when knowing its NS central line \( d \).

(Use the chart p 95 in Matheron 1971)

Exercise: sampling effort on rectangle

Consider a rectangular field \( 10000 \times 6000 \text{ m} \). We would like to know which regular sampling configuration is the best:
- three \( 10000 \ \text{m} \) EW lines, or
- five \( 6000 \ \text{m} \) NS lines.
Both correspond to the same sampling effort of \( 30000 \ \text{m} \).

Compute the global estimation variance in the following cases:
GRAPH 9: SPHERICAL SCHEME

Extension variance of the median drift $h$ within its rectangle $L \times h$
- isotropic spherical variogram with sill 2.0 and range 2500 m
- spherical variogram with sill 2.0 and range 2500 m in EW and 1000 m in NS (geometric anisotropy)
- spherical variogram with sill 2.0 and range 1000 m in EW and 2500 m in NS (geometric anisotropy).

(Use the chart p 95 in Matheron 1971)

**Exercise: optimising sampling effort**

a) Imagine a field that extends a long way in the EW direction but the limits of which are unknown. It may be regularly sampled either by EW lines, or by NS lines. Which methods can be used to estimate the total abundance?

b) Suppose that the regionalization within the field can be represented by an intrinsic Random Function. For similar sampling efforts, which direction is better:
   - to estimate the surface;
   - to estimate the average over the field;
   - to estimate the total abundance?

(Do not do any computations, and consider the following cases: isotropy, geometric anisotropy with maximal range in EW or NS)
4. KRIGING

In many cases we need to give different weights to the samples. For instance if we want to make a global estimate from an irregular grid. Or if we want to make a map (local estimation). Kriging allows us to give each sample an appropriate weight.

There are different types of kriging depending on whether the variable is considered to be stationary, intrinsic or nonstationary. We can krig a point \( x \), a block \( V \), the whole field \( V \), the mean \( m \) of the variable (in the stationary case), the trend, the gradient, or, in fact, any linear function of the variable. Nonstationary geostatistics will be studied in the next chapter. Here we will only consider the stationary and intrinsic cases.

4.1 THE PURPOSE OF KRIGING

Sampling provides information about the variable at the data points. However this does not tell us what is happening in between them. We need an accurate way to estimate the values at intermediate points or the averages over blocks or over the region.

The accuracy of the estimates depends on several factors:

- the number of samples and the quality of the data at each point.
- the positions of the samples.
- the distance between the samples and the point or block to be estimated. If we are interested in a particular point or block it is natural to rely more heavily on samples close by, rather than on more distant ones. Similarly we expect the accuracy to be best in the vicinity of the samples and to deteriorate as we move further away. The dangers of extrapolating outside the zone sampled need not be spelled out here.
- the spatial continuity of the variable under consideration. It is more accurate to estimate the value of a fairly regular variable than one which fluctuates wildly.

"Kriging" is an estimation method which takes account of all these factors. It was named after Dr. D.G. Krige, a South African mining engineer, who had developed a kind of regression technique for estimating gold grades. G. Matheron then improved on this and the new method was called kriging. In essence, it is a way of finding the best (in the sense of least variance) linear unbiased estimator. That is, we choose the weighted average of the samples values which gives the minimum estimation variance.

4.2 DERIVING THE KRIGING EQUATIONS

The problem is as follows: we have \( N \) data values \( z(x_1), \ldots, z(x_N) \) at our disposal and we want to estimate linearly the variable \( Z(x) \). For example we might want to estimate the value of the variable at a particular point, \( Z(x_0) \), or its average over a certain region. To avoid having to write out all the cases separately we denote the quantity to be estimated by:

\[
Z_V = \frac{1}{V} \int_V Z(x) \, dx
\]
where the volume \( V \) would reduce to a single point in the case of point estimation. To estimate this, we consider a weighted average of the data:

\[
z_{V}^* = \sum_{i=1}^{N} \lambda_i \, z(x_i)
\]

The problem is to choose the weighting factors \( \lambda_i \) in the best way. This is where we make use of the geostatistical model. We consider the Random Function:

\[
Z_{V}^* = \sum_{i=1}^{N} \lambda_i \, Z(x_i)
\]

We choose the weights so that the estimator is:

1. unbiased: \( E[Z_{V}^* - Z_{V}] = 0 \)
2. minimum variance: \( Var[Z_{V}^* - Z_{V}] \) is minimum.

This variance will be called the kriging variance: it is an estimation variance (i.e. the variance of the estimation error) and should not be confused, in the stationary case, with the variance of the kriging estimator.

In the first instance we assume that \( Z(x) \) is stationary and that its mean, \( m \), is unknown. Kriging with an unknown mean is called ordinary kriging, which is abbreviated to OK. In fact we will see that OK only depends on the variogram, and corresponds to intrinsic kriging.

We will also see what happens to the kriging estimator if the mean, \( m \), is known. This is called simple kriging and is abbreviated to SK. After that we will see that OK is related to SK through the kriged estimate of the mean.

**Ordinary kriging**

If \( Z(x) \) is stationary with mean \( m \), then

\[
E[Z(x_i)] = m \quad \text{and so} \quad E[Z_{V}] = m
\]

and we have

\[
E[\sum \lambda_i \, Z(x_i) - Z_{V}] = \sum \lambda_i \, m - m = m \, [\sum \lambda_i - 1]
\]

Consequently, for the estimator to be unbiased whatever the unknown value of the mean, the weights must add up to 1

\[
\sum \lambda_i = 1
\]
The variance of the error \([ Z_V^* - Z_V ] \) can be expressed in terms of either the covariance or the variogram:

\[
= \sum \sum \lambda_i \lambda_j C(x_i, x_j) + \overline{C}(V, V) - 2 \sum \lambda_i \overline{C}(x_i, V) \\
= 2 \sum \lambda_i \overline{\gamma}(x_i, V) - \sum \sum \lambda_i \lambda_j \gamma(x_i, x_j) - \overline{\gamma}(V, V)
\]

where

\[\overline{\gamma}(x_i, V)\] is the average of the variogram between \(x_i\) and the volume \(V\), i.e.

\[\overline{\gamma}(x_i, V) = \frac{1}{V} \int_V \gamma(x_i - x) \, dx\]

\[\overline{\gamma}(V, V)\] is the average of the variogram between any two points \(x\) and \(x'\) sweeping independently throughout the volume \(V\):

\[\overline{\gamma}(V, V) = \frac{1}{V^2} \int \int \gamma(x - x') \, dx \, dx'\]

In order to minimize the estimation variance [4.1] under the constraint that the sum of the weights must be equal to 1, we introduce a Lagrange multiplier \(\mu\).

\[\phi = \text{Var} ( Z_V^* - Z_V ) - 2\mu \left[ \sum \lambda_i - 1 \right]\]

Since the sum of the weights must be 1.0, adding the term in \(\mu\) does not change the value of the expression. We then need to minimize this quantity. The partial derivatives with respect to all the \(\lambda_i\) and \(\mu\) are then set to zero. This leads to a set of \(N+1\) linear equations called the Kriging System. When written in terms of the variogram model it is:

\[
\sum_{j=1}^{N} \lambda_j \gamma(x_i, x_j) + \mu = \overline{\gamma}(x_i, V) \quad i = 1, 2, ..N \\
\sum_{i}^{N} \lambda_i = 1
\]

The minimum of the variance which is called the kriging variance, is given by:

\[\sigma_k^2 = \sum \lambda_i \overline{\gamma}(x_i, V) - \overline{\gamma}(V, V) + \mu\]
Clearly the equations could also have been obtained in terms of the covariance by minimising the first form of [4.1]. The kriging system is then:

\[
\begin{align*}
\sum_{j=1}^{N} \lambda_j \ C (x_i, x_j) - \mu &= \bar{C} (x_i, V) \quad i = 1, 2, \ldots, N \\
\sum_{i}^{N} \lambda_i &= 1
\end{align*}
\]

The corresponding kriging variance, is given by:

\[
\sigma_K^2 = \bar{C} (V, V) + \mu - \sum_{i}^{N} \lambda_i \ C (x_i, V)
\]

To solve the system numerically, it is convenient to write it in matrix form. We get

\[
A \ X = B
\]

\[
\begin{bmatrix}
\gamma_{11} & \gamma_{12} & \cdots & \gamma_{1N} & 1 \\
\gamma_{21} & \gamma_{22} & \cdots & \gamma_{2N} & 1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\gamma_{N1} & \gamma_{N2} & \cdots & \gamma_{NN} & 1 \\
1 & 1 & 1 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
\lambda_1 \\
\lambda_2 \\
\vdots \\
\lambda_N \\
\mu
\end{bmatrix}
= \begin{bmatrix}
\bar{C} (x_1, V) \\
\bar{C} (x_2, V) \\
\vdots \\
\bar{C} (x_N, V) \\
1
\end{bmatrix}
\]

If \( \gamma \) is an admissible model and if there are no multiple points, the matrix \( A \) is always non singular. The inverse matrix \( A^{-1} \) exists. So a solution exists and is unique:

\[
X = A^{-1} B
\]

The kriging variance can be written:

\[
\sigma_K^2 = X^T \ B - \bar{C} (V, V) \quad (X^T = X \text{ transposed})
\]

**Deriving the OK equations for intrinsic random functions**

In the preceding section the OK equations were derived for the case of a stationary RF. What happens if \( Z(x) \) is intrinsic but not stationary? In the definition of intrinsic variables we saw that underlying idea was to work only with increments rather than with the variable itself. In particular, two hypotheses were made:

\[
E \left[ Z(x + h) - Z(x) \right] = 0
\]
\[ \text{Var} [ Z(x + h) - Z(x) ] = 2\gamma(h) \]

where \( \gamma(h) \) depends on \( h \) but not on \( x \). So under this hypothesis the estimation error \( [ Z^\star - Z^\star ] \) is an increment provided that the sum of weights is 1.0, and consequently its expectation and variance exist and can be calculated. From this point on, the procedure is the same. The variance of the estimation error is calculated in terms of variograms; it is minimised and gives rise to the OK kriging system in terms of the variogram.

**Simple Kriging**

We are now going to see how the kriging system is changed if the mean \( m \) of the RF is known. Firstly we consider a RF \( Y(x) \) with zero mean. Clearly the initial RF is obtained from this as \( Z(x) = Y(x) + m \). Our estimator of \( Y(x) \) is going to be

\[ Y^\star = \sum_{i=1}^{N} \lambda^\prime_i Y(x_i) \]

As before this estimator must be unbiased and minimum variance. In order to be unbiased the estimation error must have an expected value of 0. That is,

\[ E [ Y^\star - Y ] = E [ \sum_{i=1}^{N} \lambda^\prime_i Y(x_i) - Y ] = 0 \]

As the mean of \( Y(x) \) is 0, this estimator is automatically unbiased. So there is no condition on the sum of the weights. The variance of the estimation error is

\[ \text{Var} [ Y^\star - Y ] = E [ \sum \lambda^\prime_i Y(x_i) - Y ]^2 \]

\[ = \sum_i \sum_j \lambda^\prime_i \lambda^\prime_j C(x_i,x_j) + \overline{C(V,V)} - 2 \sum_i \lambda^\prime_i \overline{C(x_i,V)} \]

As there is no condition on the sum of the weights, there is no need for a Lagrange multiplier. Consequently the kriging system is

\[ \sum_{j=1}^{N} \lambda^\prime_j C(x_i,x_j) = \overline{C(x_i,V)} \quad i = 1, 2, \ldots N \]

The corresponding kriging variance, is given by:

\[ \sigma_{SK}^2 = \overline{C(V,V)} - \sum_i \lambda^\prime_i \overline{C(x_i,V)} \]
Solving the kriging system gives us the kriging weights and hence the estimator of \( Y_v \). From this we can deduce the estimator of \( Z_v \) by replacing \( Y(x) \) by \( Z(x) - m \). This gives us

\[
Z_v^* = Y_v^* + m \\
= \sum \lambda_i' [ Z(x_i) - m ] + m \\
= \sum \lambda_i' Z(x_i) + m [ 1 - \sum \lambda_i' ]
\]

where \( [ 1 - \sum \lambda_i' ] \) is the weight given to the mean. The less information that is available in the kriging neighbourhood, the greater the importance of the mean. So this parameter gives an indication of how sparse the data are relative to the variogram model and also to what extent the hypothesis of stationarity will be relied upon.

**Kriging the mean**

In ordinary kriging the objective was to estimate the value at a point or the average value of a block. Here the objective is to estimate the value of the (unknown) mean \( m \), or equivalently the value of a very large block. If we use the index \( m \) to distinguish this estimator from the previous one then it can be written as

\[
m^* = \sum_{i=1}^{N} \lambda_{mi} Z(x_i)
\]

As before this estimator must be unbiased and minimum variance. In order to be unbiased the estimation error must have an expected value of 0. That is,

\[
E [ m^* - m ] = E [ \sum_{i=1}^{N} \lambda_{mi} Z(x_i) - m ] = 0
\]

Consequently as before since the mean of \( Z(x) \) is \( m \),

\[
\sum_i \lambda_{mi} = 1
\]

The variance of the estimation error is

\[
Var [ m^* - m ] = Var [ \sum_{i=1}^{N} \lambda_{mi} Z(x_i) - m ] \\
= \sum_j \sum_i \lambda_{mi} \lambda_{mj} C(x_i, x_j)
\]

As in ordinary kriging, this variance is minimised subject to the constraint on the weights using a Lagrange multiplier. The kriging equations are

52
The corresponding kriging variance can be calculated. It is not difficult to show that

\[ \sigma_K^2 = \text{Var} (m^*) = \mu_m \]

This gives us a meaning for the Lagrange multiplier in this case.

**The Additivity Theorem**

Starting out from the SK estimator

\[ Z_V^* = \sum \lambda'_i Z(x_i) + m \left[ 1 - \sum \lambda'_i \right] \]

we replace m by its kriged estimator m*

\[ Z_V^* = \sum \lambda'_i Z(x_i) + m^* \left[ 1 - \sum \lambda'_i \right] \]

To simplify let \( \lambda_M \) be the weight of the mean in simple kriging: \( 1 - \sum \lambda'_i \).

This gives the estimator

\[ Z_V^* = \sum \lambda'_i Z(x_i) + \lambda_M \sum \lambda_{mi} Z(x_i) \]

\[ = \sum Z(x_i) \left[ \lambda'_i + \lambda_M \lambda_{mi} \right] \]

This estimator is just OK one. In fact, it can be shown that the weights sum to 1 and are the solution of the OK system. So OK is just a SK in which the mean m is replaced by its kriged estimate.

Note that \( \lambda_M \) is the weight of the mean, whether this mean is known (SK) or estimated (OK). More information on the kriging weights and on the weight of the mean is given in Rivoirard (1984, 1987).

The different kriging variances also are related

\[ \sigma_K^2 = \sigma_{SK}^2 + \text{Var} (m^*) (\lambda_M)^2 \]

This shows that the kriging variance can be split into two parts: the first is the simple kriging variance when the mean is known; the second is the kriging variance of the mean multiplied by the square of its weight. The second term measures the loss of accuracy caused by having to use an estimated value for m.
When the kriging weight of the mean is low OK is close to SK. Little precision is lost in estimating the mean. Conversely when it is high, the OK variance is markedly higher then the SK variance. This has an influence on the choice of the kriging neighbourhood, as we will see.

4. 3 PROPORTIONAL EFFECT

Kriging weights and variance depend on the structure and on the geometrical configuration, not on the data values $Z(x_i)$ themselves. This makes it possible to see the improvement in the precision that would result from a supplementary sample. But on the other hand, the kriging variance is not conditional to the data values. That is it does not take into account the fact that variability is often all the higher than values are high.

However, multiplying the covariance or variogram by a constant factor does not change the kriging weights. The kriging variance is multiplied by this factor. In the case of a proportional effect, a factor of this type may be taken to depend on the kriged value, so that it gives a kind of conditional variance. (Otherwise this is only accessible using nonlinear geostatistics).

When the structure is only known up to a factor, the same property guarantees the weighting to be correct, even if variance is not.

4. 4 POINT AND BLOCK KRIGING

In the kriging system, the left side matrix corresponds to the covariances between the data, the right hand one to the covariance between the data and the point or block we wish to estimate. The weights depend linearly on this. As a consequence, kriging a block is exactly the average of the krigings of its points. Of course, this presupposes that the data neighbourhood is the same.

Kriging is an exact interpolator. At a data point $x_0$, the estimator $Z(x_0) = Z(x_0)$ minimizes the estimation variance (since this is zero). This means that the mapping will pass through the data.

In the case of a nugget component, the mapping presents a discontinuity at the sample points. When a point $x$ moves toward a data point $x_0$, its kriged value does not tend towards the kriged value of $Z(x_0)$ (which is $Z(x_0)$), but tends towards the kriged value of a micro–regularised $v$ around $x_0$. This limit also represents the kriged value at $x_0$ when the nugget component comes from a measurement error that we want to filter. In this case the mapping does not pass through the data. Of course, it is also possible to consider that only some of the data are subject to error.

4. 5 SMOOTHING

A kriged map does not look like the real field. Interpolation between data cannot reproduce all the details that exist in the real field, and so kriging is a smoothing operation. The more variable the real field, the less representative the data are, and consequently the smoother the kriged map looks, which may seem paradoxal.

If reproducing the real variability is more important than the precision, of the map the solution is to simulate the random function. Simulations look like a possible field.

It is possible (in the gaussian case) to force the simulation to go through the known values at the data points. We then speak of conditional simulation.
4.6 NEGATIVE KRIGING WEIGHTS

Suppose that samples have been taken at 4 points \( P_1, P_2, P_3 \), and \( P_4 \) that are regularly spaced one unit apart on a line. The value at their midpoint \( P_0 \) is to be estimated. The locations of the points are as shown in Figure 4.1.

\[ \begin{array}{c}
\bullet & P_1 & - & P_2 & - & P_0 & - & P_3 & - & P_4 \\
\end{array} \]

Figure 4.1: Location of the four samples \( P_1, P_2, P_3, \) and \( P_4 \) and the point \( P_0 \) to be estimated.

For comparison purposes, four variogram models were used to krig the point. These were

1. a power function model with exponent 1.5 (i.e. \( \gamma(h) = \beta h^{1.5} \)).

2. a Gaussian model with a distance parameter \( \alpha = 0.8 \) and a sill of 1.0 and no nugget effect.

3. a Gaussian model with a distance parameter \( \alpha = 0.8 \), a sill of 0.75 and a nugget effect of 0.25 (i.e. a total sill of 1.0) and

4. a spherical model with a range of 1.38 and no nugget effect.

The practical range of the gaussian is \( 0.8 \sqrt{3} = 1.38 \). The spherical given in (4) has the same range as the gaussian but is linear at the origin instead of quadratic. Table 6.1 gives the kriging weights corresponding to these variograms. By symmetry the weights for \( P_2 \) and \( P_3 \) are identical (\( \lambda/2 \)), as are those for \( P_1 \) and \( P_4 \) (\( \lambda/2 \)).

<table>
<thead>
<tr>
<th>Kriging weights</th>
<th>Model to power 1.5</th>
<th>Gaussian with no nugget effect</th>
<th>Gaussian with nugget effect</th>
<th>Spherical with no nugget effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_1 )</td>
<td>-0.047</td>
<td>-0.083</td>
<td>0.008</td>
<td>0.010</td>
</tr>
<tr>
<td>( \lambda_2 )</td>
<td>0.547</td>
<td>0.583</td>
<td>0.492</td>
<td>0.490</td>
</tr>
<tr>
<td>Kriging variance</td>
<td>0.201</td>
<td>0.227</td>
<td>0.563</td>
<td>0.590</td>
</tr>
</tbody>
</table>

Table 6.1: Kriging weights corresponding to the four variogram models.

There are negative weights for the outer points for the first two models (the power model with exponent 1.5 and the gaussian with no nugget effect) because these are highly structured models. In contrast to this the last two models (the gaussian with 25% nugget effect and the spherical) are much less structured and consequently do not have negative weights. When choosing a variogram model it is important to realise that a model that is quadratic at the origin (particularly with no nugget effect) corresponds to a more structured phenomenon than a linear model at the origin does, and consequently is more likely to give rise to negative weights and hence to negative kriged concentrations (which are not desirable in mining and in some other fields).
4.7 SCREEN EFFECT

Suppose that we had a line of samples each 2000m apart in order to estimate a 2000 x 2000m block centered on a sample. We can imagine increasing the number of samples from 1 (the central one), to 3, 5, 7 and so on by taking the most closest pair of samples. Once the variogram model is known, the kriging weights and the kriging variance can be calculated for each data configuration. Clearly each time more samples are added the kriging variance will decrease (or stay the same).

To illustrate how the screen effect works, three cases are considered: a spherical variogram with a range of 2500m and a sill of 2.0; a second spherical variogram with a range of 1000 m and a sill of 2.0 and thirdly a pure nugget effect of 2.0. This could be thought of as a spherical with a zero range. Figure 4.2 shows the OK weights for each point along the line as the height in a bar diagram. The first column corresponds to the long range spherical; the middle column, to the short range spherical and the right column to the pure nugget effect. The kriging variance associated with each configuration is given just below it.

The figures in the left column show that for the well structured spherical model (range 2500m) the kriging variance and the weights quickly tend to limit values. Increasing the number of points to more than 5 does not lead to any significant improvement in the kriging variance nor do the weights (and hence the estimated value) change much. So there seems to be little point in using more than the closest few data in this case.

In contrast to this when the variogram is poorly structured (pure nugget effect or a spherical with a short range) the kriging variance continues to drop as more samples are added and the weights for the outer points do not tend to zero quickly. So in this case a larger kriging neighbourhood is required. Please note that even points outside the range from the block to be estimated have nonzero OK weights. They are not necessarily zero.

4.8 THE NEIGHBOURHOOD

This point can be of great importance in practice. The kriging neighbourhood is the geometrical domain where are taken the sample data for kriging. If we want to estimate the average over the whole field (global estimation), then the neighbourhoods correspond to the field and contains all the samples. This neighbourhood is also used, and called unique neighbourhood, when making a map (local estimation) from not too many samples.

Sampling programs often produce hundreds or thousands of data values. If we want to make a map, then we should select the closest and the most influential samples. This can be achieved by using a moving neighbourhood (which is centered on the point to be estimated and moves with it) or by using a more sophisticated neighbourhood research (like the one in the Bluepack software). In special cases, data are dense along lines (seismic profiles, acoustic survey). The neighbourhood should be then chosen also in term of these lines. The neighbourhood research is even more crucial in nonstationary mapping.
Figure 4.2: The OK weights and the kriging variance for 5 different kriging configurations for three variogram models.
When there are enough data around the target point, often the closest ones screen out the others, and the neighbourhood can be limited to these closest data. For example only a few aureoles (i.e. rings) around the point or block to be estimated need be used.

Block + 1 Aureole

Block + 2 Aureoles

If it is not the case, the weight of the mean may be high. In SK there is generally no problem. Due to the drop off of the correlations with the distance, only close samples have a significant weight. So it is generally easy to determine the kriging neighbourhood for SK.

On the contrary OK includes an implicit estimation of the mean. As kriging the mean involves kriging a very large field, it can give weight to samples whatever their distance. As a consequence, if the weight of the mean is high, it is difficult to determine the ideal kriging neighbourhood, and this will rather have to be chosen. A larger neighbourhood theoretically gives a better precision, but a smaller one reduces the importance of the local stationarity hypothesis and reduces the computation time.

4. 9 EXERCISE: KRIGING A SEGMENT

Consider the variogram \( \gamma(h) = |h|^\lambda \) \((0 < \lambda < 2)\) in a 1D space. Suppose that samples have been taken at 4 points \( x = 0, l, 2l, 3l \). The average value over the segment \([l, 2l]\) is to be estimated. Calculate the kriging weights for each sample (Use the symmetries and some of the results from the exercise on the segment estimated by one or two points).

For which \( \lambda \) do the inner samples screen the outer ones? For which value of \( \lambda \) do the outer samples have a negative weight? What is the reason for this?

4. 10 EXERCISE: KRIGING A BLOCK

The shaded block (2000m x 2000m) is to be kriged using 5 samples on a regular 2000m grid. Suppose that the regionalized variable is stationary with an isotropic spherical variogram with a sill of 2.0 and a range of 2500m.

To make the calculations easier to do with a pocket calculator, the values of \( \gamma(V, V) \) and \( \gamma(V, x) \) are given. These can be calculated using a computer program or the appropriate tables.
\[ \bar{\gamma}(x_1, V) = 0.88 \]
\[ \bar{\gamma}(x_2, V) = 1.86 \]
\[ \bar{\gamma}(V, V) = 1.13 \]

**Solution**

The kriging system is:

\[
\begin{bmatrix}
\gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} & \gamma_{15} & 1 \\
\gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} & \gamma_{25} & 1 \\
\gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} & \gamma_{35} & 1 \\
\gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} & \gamma_{45} & 1 \\
\gamma_{51} & \gamma_{52} & \gamma_{53} & \gamma_{54} & \gamma_{55} & 1 \\
1 & 1 & 1 & 1 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
\lambda_1 \\
\lambda_2 \\
\lambda_3 \\
\lambda_4 \\
\lambda_5 \\
\mu
\end{bmatrix}
= 
\begin{bmatrix}
\gamma(V, x_1) \\
\gamma(V, x_2) \\
\gamma(V, x_3) \\
\gamma(V, x_4) \\
\gamma(V, x_5) \\
1
\end{bmatrix}
\]

When the variogram values are substituted, this becomes:

\[
\begin{bmatrix}
0 & 1.89 & 1.89 & 1.89 & 1.89 & 1 \\
1.89 & 0 & 2 & 2 & 2 & 1 \\
1.89 & 2 & 0 & 2 & 2 & 1 \\
1.89 & 2 & 2 & 0 & 2 & 1 \\
1.89 & 2 & 2 & 2 & 0 & 1 \\
1 & 1 & 1 & 1 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
\lambda_1 \\
\lambda_2 \\
\lambda_3 \\
\lambda_4 \\
\lambda_5 \\
\mu
\end{bmatrix}
= 
\begin{bmatrix}
0.89 \\
1.86 \\
1.86 \\
1.86 \\
1.86 \\
1
\end{bmatrix}
\]
This can easily be solved to give:

\[
\begin{align*}
\lambda_1 &= 0.60 \\
\lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 &= 0.10 \\
\mu &= 0.12
\end{align*}
\]

So the estimated average value over the square is:

\[
Z = 0.60 \, Z_1 + 0.10 \, (Z_2 + Z_3 + Z_4 + Z_5)
\]

The estimation variance is given by:

\[
\sigma_k^2 = \sum \lambda_i \, \bar{y} (V, x_i) + \mu - \bar{y} (V, V) = 0.26
\]

Symmetry in the equations

When kriging was first developed, computer facilities were poor. In particular inverting large matrices or simply solving large sets of linear equations was very time consuming. This led geostatisticians to look for ways of reducing the size of kriging systems. One way is by taking account of any symmetries in the system. For example, the previous exercise on ordinary kriging involved kriging a block using 5 data, four of which are set symmetrically outside the block. Moreover there is no anisotropy. These four weighting factors are clearly identical.

\[
\begin{align*}
\bullet \, Z_2 \\
\bullet \, Z_3 & \quad \bullet \, Z_1 & \quad \bullet \, Z_5 \\
\bullet \, Z_4
\end{align*}
\]

Figure 6.1: A symmetric data configuration

These samples can be regrouped and considered as a single unit S with a single weighting factor. Each of the individual samples will get one quarter of this. As before the first kriging weight is associated with the central point Z1. The second weight is associated with the group S.

The kriging system can be rewritten as:

\[
\begin{bmatrix}
\bar{y} (Z_1, Z_1) & \bar{y} (Z_1, S) & 1 \\
\bar{y} (Z_1, S) & \bar{y} (S, S) & 1 \\
1 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
\lambda_1 \\
\lambda_2 \\
\mu
\end{bmatrix}
= 
\begin{bmatrix}
\bar{y} (Z_1, V) \\
\bar{y} (S, V)
\end{bmatrix}
\]

60
Although it is not difficult to calculate the various variogram values, it is not self evident. So this will be presented in detail.

\[
\bar{\gamma} (Z_1, Z_1) = 0, \quad \bar{\gamma} (Z_1, V) = 0.88
\]

\[
\bar{\gamma} (Z_1, S) = \frac{1}{4} \left[ \gamma (Z_1, Z_2) + \gamma (Z_1, Z_3) + \gamma (Z_1, Z_4) + \gamma (Z_1, Z_5) \right]
\]

= 1.89

\[
\bar{\gamma} (S, S) = \frac{1}{16} \left[ \gamma (Z_2, Z_2) + \gamma (Z_2, Z_3) + \ldots + \gamma (Z_5, Z_5) \right]
\]

= \frac{1}{4} \left[ \gamma (0) + \gamma (100 \sqrt{2}) + \gamma (100 \sqrt{2}) + \gamma (200) \right]

= 1.50

\[
\bar{\gamma} (S, V) = \frac{1}{4} \left[ \bar{\gamma} (Z_2, V) + \ldots + \bar{\gamma} (Z_5, V) \right]
\]

= \bar{\gamma} (Z_2, V)

= 1.86

So now we have a 3 x 3 system instead of a 6 x 6 one:

\[
\begin{bmatrix}
0 & 1.89 & 1 \\
1.89 & 1.50 & 1 \\
1 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
\lambda_1 \\
\lambda_S \\
\mu
\end{bmatrix}
= \begin{bmatrix}
0.88 \\
1.86 \\
1.00
\end{bmatrix}
\]

Solving it gives \(\lambda_1 = 0.60\), \(\lambda_S = 0.40\) and \(\mu = 0.12\)

and hence:

\[Z^* = 0.60 Z_1 + 0.40 (Z_2 + Z_3 + Z_4 + Z_5) / 4.\]

Of course, \(\sigma_K^2\) is the same as before.
5. NON STATIONARITY

5.1 GENERALITIES

In linear geostatistics we need only calculate the expectation and the variance of linear combinations \( \sum \lambda_i Z(x_i) \). Obviously if we know the mean

\[
E[Z(x)] = m(x)
\]

and the covariance

\[
\text{Cov} [Z(x), Z(y)] = E [Z(x) - m(x)][Z(y) - m(y)]
\]

\[
= E[Z(x)Z(y)] - m(x)m(y)
\]

for any points \( x \) and \( y \), we can compute what we need

\[
E \left[ \sum \lambda_i Z(x_i) \right] = \sum \lambda_i E[Z(x_i)] = \sum \lambda_i m(x_i)
\]

\[
\text{Var} \left[ \sum \lambda_i Z(x_i) \right] = \sum \lambda_i \lambda_j \text{Cov} [Z(x_i), Z(x_j)]
\]

In the stationary case, the mean and the covariance are stationary, i.e. the mean \( m(x) = m \) does not depend on \( x \), and the covariance depends only on the difference \( x - y \)

\[
\text{Cov} [Z(x), Z(y)] = C(x - y)
\]

It is easy to imagine non stationary models, in which either \( m(x) \) is not stationary, or the covariance is not stationary. Generally we define the drift as the first moment of the variable of interest

\[
m(x) = E[Z(x)]
\]

Remark

The hypothesis of stationarity or non-stationarity not only depends on the phenomenon itself, but also on the scale at which it is studied. For instance a chain of mountains can be considered globally as stationary, whereas one individual peak is definitely not stationary. Conversely the sea floor depth, which increases from the coast line, can be globally considered as not stationary, whereas it can behave locally as a stationary variable.
5.2 "ADDITIVE" MODEL

The simplest is the additive model, in which the target variable $Z(x)$

$$Z(x) = m(x) + Y(x)$$

is considered as the sum of a drift $m(x) = E[Z(x)]$ and of an independent residual $Y(x)$ with mean zero and stationary covariance

$$\text{Cov} \ [Y(x), Y(y)] = E \ [Y(x)Y(y)] = C(x - y)$$

In this case the mean of $Z(x)$ is non stationary, but its covariance

$$\text{Cov} \ [Z(x), Z(y)] = E \ [Z(x) - m(x)][Z(y) - m(y)] = E \ [Y(x)Y(y)] = C(x - y)$$

is stationary.

In practice we must distinguish the cases where the drift $m(x)$ is known, from those where it is unknown. If the drift is known we just have to subtract this drift and are back at the stationary case with the residual $Y(x) = Z(x) - m(x)$.

The case where the drift is unknown (and has to be estimated) is more complex. It corresponds to the Universal Kriging approach. However as the covariance itself $\text{Cov} \ [Z(x), Z(y)]$ may be unknown, the more general IRF-k approach may be preferable, as we will see in the next chapters.

Note that the previous additive model cannot represent variables for which the variability changes with the drift, as will be the case in the following multiplicative model.

We must stress the fact that the drift is generally not something which is given a priori. In the case of a single realization of a Random Function, it is not clear which part of the variable corresponds to the drift, and which is the residual.

However it may happen that the phenomenon under study varies with time, but keeps the same drift. If we consider that different times correspond to different realizations of the same Random Function, the drift appears as a time average and is accessible through repeated surveys.

5.3 MULTIPLICATIVE MODEL

Let us consider the multiplicative model

$$Z(x) = m(x) \ Y(x)$$

in which the target variable $Z(x)$ is the product of a drift $m(x)$ and an independent variable $Y(x)$ with mean 1 and stationary covariance $C(h)$. 

63
It is possible to split the target variable $Z(x)$ into its mean $m(x)$ and the residual $Z(x) - m(x) = m(x) \{ Y(x) - 1 \}$. But the covariance of this residual, which is also the covariance of $Z(x)$, is not stationary

$$\text{Cov} [Z(x), Z(y)] = E [Z(x) - m(x)][Z(y) - m(y)] = m(x) \ m(y) \ C(x - y)$$

In particular the variance at a point $x$

$$m(x)^2 \ C(0)$$

is larger when the value of the drift is large at that point.

Here we will suppose that the drift $m(x)$ is known and different from zero. We then can work on the relative variable $Y(x) = Z(x)/m(x)$, which is stationary.

This makes it possible to compute estimation variances, even if the data points are more numerous in rich zones. The unknown quantity is

$$Z_N = \frac{1}{V} \int_V Z(x) \ dx = \frac{1}{V} \int_V m(x) \ Y(x) \ dx$$

and its estimate is

$$\sum \lambda_i \ Z(x_i) = \sum \lambda_i \ m(x_i) \ Y(x_i)$$

The estimation error

$$Z_N - \sum \lambda_i \ Z(x_i)$$

has a mean

$$\frac{1}{V} \int m(y) \ dy - \sum \lambda_i \ m(x_i) = m_Y - \sum \lambda_i \ m(x_i)$$

and a variance

$$\sum \sum \lambda_i \ m(x_i) \ \lambda_j \ m(x_j) \ C(x_i, x_j) + \frac{1}{V^2} \int m(x) \ m(y) \ C(x-y) \ dx \ dy - 2 \sum \lambda_i \ m(x_i) \ \frac{1}{V} \int m(y) \ C(x_i-y) \ dy$$

It is then possible to find the $\lambda_i \ m(x_i) / m_Y$, summing to 1 (no bias), which minimize the estimation variance.

The model can be written

$$Z(x) = m_Y \ l(x) \ Y(x)$$

where $l(x) = m(x)/m_Y$ is the relative drift. Such a model has been developed by P. Petitgas 1991 for sole eggs. But it was more complex in this case, as the global mean $m_Y$ was itself varying with time. He obtained, with more complete notation:

64
\[ Z_t(x) = m_Y(t) \ell(x) Y_t(x) \]

The value at point \( x \) and time \( t \) is conditioned:
1) by the global abundance at that time, and
2) by the value of the relative drift at that point.
Such a model is not new in biology. Petitgas's contribution was to establish it in terms of spatial structure, verifying the spatial independence between the different components.
6. UNIVERSAL KRIGING

Here we consider an additive model where the target variable $Z(x)$ is the sum of a drift and an independent residual. The drift is supposed to be a smooth continuous function that varies slowly and will be usually modelled by a low order polynomial.

The study of this unknown drift and of the corresponding residuals (obtained by subtracting the drift from the initial data) is precisely the topic of the Universal Kriging.

6.1 GENERAL HYPOTHESES

The target variable $Z(x)$ is the sum of a drift and an independent residual:

$$Z(x) = m(x) + Y(x)$$

The unknown drift $m(x) = E[Z(x)]$ is supposed to be a smooth continuous function that varies slowly. It will be written as a linear combination of a given number $k$ of known basic functions (denoted $f^l(x)$) such as monomials, with unknown coefficients $a_l$:

$$m(x) = \sum_{l=0}^{k} a_l f^l(x)$$

The covariance of the residuals is stationary and can be written:

$$\text{Cov}[Z(x), Z(x+h)] = \text{Cov}[Y(x), Y(x+h)] = C(h)$$

6.2 ESTIMATION OF THE DRIFT BY LEAST SQUARES

The unknown coefficients $a_l$ are determined by minimizing the cumulated distance between each datum (denoted $z_a = z(x_a)$) and the value of the drift at that point:

$$\text{Minimum} \sum_a \left( z(x_a) - \sum_{l=0}^{k} a_l f^l(x_a) \right)^2$$

which leads to the following linear system:

$$\sum_{l=0}^{k} a_l \sum_{a} f^l_a f^m_a = \sum_{a} f^m_a z_a \quad \text{for} \ m \leq k$$

This method does not require any assumptions about the structure of the variable and consequently does not provide, in general, an optimal linear estimator of the drift.
6.3 OPTIMAL LINEAR ESTIMATION OF THE DRIFT

We consider the optimal unbiased linear estimate of the drift:

\[ m^*(x) = \sum_a \lambda^a Z(x_a) \quad (\lambda^a \text{ unknown}) \]

\[ E[m^*(x_0)] = m(x_0) = > \sum_a \lambda^a f'_a = f'(x_0) \text{ for each } l \text{ (unbiased)} \]

\[ D^2(m - m^*) = E(m - m^*)^2 = \sum_a \sum_\beta \lambda^a \lambda^\beta C_{\alpha\beta} \text{ (optimal)} \]

The unknown coefficients \( a_l \) no longer appear.

This leads to the kriging system for the estimation of the drift introducing \( k+1 \) Lagrange multipliers:

\[
\begin{align*}
\sum_a \lambda^a C_{\alpha\beta} - \sum_l \mu_l f'_a &= 0 \\
\sum_a \lambda^a f'_a &= f'(x_0)
\end{align*}
\]

Under the usual positive definiteness condition on the covariance \( C(h) \), the previous system is regular and provides a unique solution as long as all data points are all distinct.

Note that, using a similar formalism, we could also expand the drift as follows:

\[ m(x) = \sum_{l=0}^k A_l f'_a \]

where each \( A_l \) is now considered as a random variable and use a similar kriging system to derive its optimal unbiased linear estimate.

We must keep in mind that the problem of estimating the drift has not been completely solved:

i) in order to perform Universal Kriging, we need to have done the structural analysis, i.e. to have obtained the covariance function \( C(h) \)

ii) to calculate the experimental covariance function, we have to subtract the drift \( m(x) \) from the raw phenomenon \( Z(x) \)

iii) the "true" drift (the one used in the model) is unattainable. We can only try to estimate it optimally

iv) the optimal estimator of the drift can only be obtained after the covariance function is already known, that is after the inference has been made.

This problem can be called the vicious circle of Universal Kriging.
6. 4 VARIOGRAM OF RESIDUALS

When we estimated the drift using kriging we needed to introduce the a priori structure of the (true) residuals. We would like to check this hypothesis, at least a posteriori.

The previous paragraph allow us to obtain the optimal linear estimation of the drift and therefore the optimal estimated residuals. The next step consists intuitively in calculating the variogram of these estimated residuals and comparing it to the initial hypothesis. Unfortunately, because we do not know the true drift or the true residuals, the variogram of the estimated residuals may be quite different from the variogram of the true residuals: not only it can be proved that it is biased

\[ \gamma_R'(h) \leq \gamma_R(h) \]

but even the shape of the two variograms may be significantly different.

The equality is obtained only when the estimated residuals are close enough to the true ones. As an illustration of this inequality, we can make the residuals as small as possible and therefore arbitrarily reduce their variogram. In fact, if we estimate the residuals by least squares, we can imagine using a polynomial with a degree equal to the number of data points and thus moving all the variability into the drift component and shrinking the residuals to zero.

6. 5 UNIVERSAL KRIGING SYSTEM

As in the stationary or the intrinsic cases, we now wish to produce the optimal linear unbiased estimation of the variable at a point \( x \).

This leads to the following kriging system:

\[
\begin{align*}
\sum_{\beta} \lambda^\beta C_{a\beta} + \sum_{l} \mu_{l}^a &= C_{ax} \\
\sum_{a} \lambda^a f^a_{l} &= f_x \\
\sigma^2 &= C_{xx} - \sum_{a} \lambda^a C_{ax} - \sum_{l} \mu_{l}^a
\end{align*}
\]

An interesting remark is that this system is very similar to the one for the optimal estimation of the drift: they have the same left-hand side. Therefore the remarks about the non singularity of the system still hold.

6. 6 CAN UNIVERSAL KRIGING STILL BE APPLIED?

Universal Kriging requires a dichotomy of the phenomenon splitting it into a drift component and residuals. For this model to be workable, we also need the residuals to be stationary (or at least intrinsic).

The technique may therefore still be applied nicely when the "true" drift or the "true" residuals are attainable.
i) "true" drift is attainable when the drift has a physical meaning (e.g. time average) or due to the construction of a random process.

ii) "true" residuals are attainable if the drift only operates in one direction: the orthogonal can be considered as "drift free" and one can believe that the variable measured in this direction coincides with the "true" residuals.

6.7 EXERCISE

A phenomenon $Z(x)$ defined on a segment has been regularly sampled and the following values obtained:

<table>
<thead>
<tr>
<th>$x$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z(x)$</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

This phenomenon is regarded as a realization of a random function on which we want statistical inference. Of course this case is rather academic since in practice it is difficult to make correct inference with so few data. The present data set has been designed to lead to results without prohibitive calculations.

i) Compute the raw variogram $\gamma_R(h)$ and observe that it indicates the presence of errors on the data themselves anyway).

ii) Assuming the drift is linear, evaluate it by least squares or alternatively by kriging. The variogram is linear). Compute the variogram of the residuals $\gamma_R(h)$ in b comment on them.

iii) In fact the given data originate from a coin tossing experiment: at each toss throw incremented by 1, and $Z$ by 0 or 2 according as heads or tails came out. So:

$$Z(n) = \sum_{i=1}^{n} X_i$$

where the $X_i$ are independent identically distributed Bernoulli

$$\text{Prob}(X_i = 0) = \text{Prob}(X_i = 1) = 1/2$$

and $E[X_i] = Var(X_i) = 1$

Derive:
- the "true" drift $m(x)$
- the "true" residuals $Y(x) = Z(x) - m(x)$
- the variogram of the "true" residuals
- the theoretical raw variogram $E[\gamma_R(h)]$

Compare these results with the ones above.

References on UK:

7. THE I.R.F.-K

The Universal Kriging technique required a splitting of the target variable into a drift component and a residual part. This dichotomy is responsible for the problems in the practical use of this technique.

Here, we prefer to study the target variable through a linear combination (increment) which filters out the drift. This supposes that this drift is additive (and not multiplicative). A new corresponding structural tool is introduced: the generalized covariance. We then derive the kriging system and also present some hints on the way to infer the parameters of this generalized covariance.

7.1 USING LINEAR COMBINATIONS

We start this paragraph by recalling the various hypotheses used in the first chapter.

In the (strict) stationary case, we assume that the random variable has a known constant mean. In the intrinsic case, only the mean of an increment is considered, and is set to zero, whereas the mean of the variable may not exist. If the mean exists, it is constant but unknown, and using increments is a way to eliminate this indeterminacy.

A similar technique will be used to get rid of a drift more complex than just a constant.

As for the Universal Kriging model, we usually assume that the drift can be considered as a polynomial of order k. The number of monomials $N(k)$ constituting a polynomial of order k obviously depends on the dimension of the study $n$:

$$N(k) = \frac{(k + n)!}{k! \ n!}$$

Finite Differences

Let us first illustrate this filtering technique with a regular sampling in 1-D.

\[
\begin{array}{cccccccc}
  z_1 & z_2 & z_3 & z_4 & \ldots & z_i & \ldots & z_n \\
\end{array}
\]

The variable $z_i' = z_{i+1} - z_i$ filters out any constant added to $z$. The variable $z'$ is usually called the first increment of the variable $z$.

The variable $z''$, built as the first increment of the variable $z'$, is also the second order increment of the variable $z$, then $z_i'' = z_{i+2} - 2z_{i+1} + z_i$ still filters out the constant but also any linear function of the coordinate.

Using this technique, we can build k-order increments:

$$\Delta^k z_i = \sum_{q=0}^{k} (-1)^q C^k_q z_{i+k-q}$$
We can demonstrate that this k-order increment filters out any polynomial of degree up to k−1. Therefore if we believe that the variable \( Z(x) \) is not stationary because of the presence of a polynomial drift, one might think of taking successive increments until stationarity is reached.

For example, if the target variable \( Z(x) \) is the sum of a polynomial drift of degree k and an independent stationary residual, then the drift is filtered out by increments of order k.

Exercise:

Let us consider the multiplicative model \( Z(x) = m(x)Y(x) \) where \( Y(x) \) is stationary with mean 1. This implies that the drift is \( E[Z(x)] = m(x) \). Assuming that \( m(x) \) is linear, show that the second order increments have a zero expectation but that their variance still depends on the drift. Conclude that the drift is not filtered out.

Divided Differences

The previous technique can be extended to the case where the 1-D sampling is no longer regular:

\[
\begin{align*}
z(x_1) & \quad z(x_2) & \quad z(x_3) & \quad \ldots & \quad z(x_i) & \quad \ldots & \quad z(x_n) \\
\hline
\end{align*}
\]

The first order increment is obtained by considering \( z'(x_i) = z(x_{i+1}) - z(x_i) \). Higher order increments are obtained by iterating the procedure.

Least Squares residuals

The most commonly used method of constructing k-increments in \( \mathbb{R}^n \) is take residuals from a least squares fit by a polynomial \( p(x) \) of degree up to k: \( r(x) = z(x) - p(x) \).

i) First each residual is a linear function of the data. As a matter of fact, we have already stated that the coefficients \( a_l \) of the polynomial result from the linear system of equations:

\[
\sum_{l=0}^{N(k)} a_l \sum_{a} f_a^l f_m^n = \sum_{a} f_a^m z_a \quad \text{for} \quad m \leq N(k)
\]

If we assume that the matrix \( A \) whose generic element \( A_{lm} = \sum_{a} f_a^l f_m^n \) is regular, we can derive each coefficient: \( a_l = \sum_{a} \left( \sum_{m=0}^{N(k)} A_{lm}^{-1} f_m^n \right) z_a \), and we can finally write the least squares residuals as a linear combination of the data:

\[
r(x) = z(x) - \sum_{l=0}^{N(k)} a_l f^l(x) = z_x - \sum_{a} \left( \sum_{l=0}^{N(k)} \sum_{m=0}^{N(k)} A_{lm}^{-1} f_m^n f^l(x) \right) z_a
\]
ii) Lastly, we must prove that this linear combination filters out each monomial up to degree k, and therefore any linear combination of them, such as \( p(x) \) for example, that is to demonstrate that the linear combination is zero when \( z(x) \) is replaced by \( p(x) \).

But transforming \( z_a \) into \( p(x_a) \) will obviously not modify the coefficients \( a_l \) of the fitted polynomial and therefore the residuals \( r(x) = p(x) - \sum_{l=0}^{N(k)} a_l f_a^l \) are effectively zero.

**Increment of order k**

We will define an *increment of order k* of a random function \( Z(x) \) as a linear combination \( Z(\lambda) = \sum_a \lambda^a Z(x_a) \) when the weights are chosen so that \( \sum_a \lambda^a P(x_a) = 0 \) for any polynomial \( P(x) \) of degree up to \( k \). We also say that this linear combination is authorized for the order \( k \).

### 7.2 INTRINSIC RANDOM FUNCTIONS OF ORDER K

**Definition:**

By analogy with the definition of the intrinsic case, we will say that a random function is intrinsic of order \( k \) (IRF-\( k \)) if its increments of order \( k \) are stationary, have a zero mean and have a finite variance. An intrinsic random function can therefore be considered as IRF-0.

**Properties:**

i) If the mean of the increments of order \( k \) is constant but not zero, increments of order \( k + 1 \) have a zero mean.

ii) If the intrinsic property is true for order \( k \), it remains true for order \( k' \geq k \). For example an intrinsic variable (IRF-0) is intrinsic at all others orders \( k \). Similarly a stationary variable is intrinsic at any order so that it can be convenient to consider such a variable as a IRF-(−1).

### 7.3 GENERALIZED COVARIANCES

Let \( Z(x) \) be a continuous IRF-\( k \). A continuous and symmetric function \( K(h) \) is a *generalized covariance* if, for any pair of increments of order \( k \), \( Z(\lambda) \) and \( Z(\mu) \):

\[
E[Z(\lambda)Z(\mu)] = \sum_a \sum_{\beta} \lambda^a \mu^\beta K(|x_a - x_\beta|)
\]

where \( |x_a - x_\beta| \) denotes the distance between the two points in \( \mathbb{R}^n \).

We can demonstrate that \( K(h) \) exists and is unique up to an even polynomial of degree \( \leq 2k \).

We still need to characterize a generalized covariance as, for any increment of order \( k \),

\[
Var[Z(\lambda)] = \sum_a \sum_{\beta} \lambda^a \mu^\beta K(|x_a - x_\beta|)
\]

must be non negative. We say that a generalized covariance
must be \( k \)-conditionally positive definite. By comparison, a covariance must be positive definite and a variogram must be \( 0 \)-conditionally negative definite.

Any covariance or variogram may be used as a generalized covariance and then the family of the generalized covariances is much larger than the one of the covariances or the variograms.

The generalized covariance is related to the order of the IRF by the limiting conditions:

- variogram
  \[
  \lim_{h \to \infty} \frac{\gamma(h)}{h^2} = 0
  \]

- generalized covariance
  \[
  \lim_{h \to \infty} \frac{K(h)}{h^{2k+2}} = 0
  \]

**Polynomial generalized covariances**

Among many possible functions, the polynomial generalized covariances are chosen:

\[
K(h) = \sum_{p=0}^{N(k)} b_p \ |h|^{2p+1} (-1)^{p+1} \text{ with all coefficients } b_p \geq 0.
\]

1) The polynomial generalized covariance is a linear combination of the parameters of the model \( b_p \), which facilitates their inference.

2) Even terms of a polynomial generalized covariance do not matter and can therefore be ignored.

3) Providing that this polynomial generalized covariance is completed by a nugget effect, \( m \) of the features of the variable are captured in this rather poor function. However, with \( k = 0 \), the polynomial generalized covariance is reduced to a combination of the nugget effect and a linear term.

4) The polynomial generalized covariance is isotropic but studies have shown that any possible anisotropy is taken into account by the drift. Note that this remark does not hold with \( k = 0 \).

5) If necessary, the polynomial generalized covariance can be nested with standard variograms and covariances. It will then lose the nice property of linearity against the parameters variogram is not linearly dependent on the range.

6) Another component can also be used: \( h^2 \log |h| \). When a 2-D variable is an IRF–1 wit structure, its kriging is equivalent to the thin–plate spline technique based on the minimization of its bending energy. The component \( h^2 \log |h| \) is thus called the 2-D spline. In 1-D, the thin–plate spline technique corresponds to a \( h^3 \) structure.
7.4 KRIGING IN THE IRF–K MODEL

We now establish the kriging system (linearity, non-bias and optimality conditions) in the scope of the IRF–k model:

\[
\begin{align*}
Z^* &= \sum_{\alpha} \lambda^\alpha Z_\alpha \\
E[Z_0^* - Z_0] &= 0 \\
\text{Var}[Z_0^* - Z_0] &= \text{minimum}
\end{align*}
\]

Before developing the non-bias and optimality conditions, we must first check that the linear combination \( \sum_{\alpha} \lambda^\alpha Z_\alpha - Z_0 \) is authorized for the order \( k \), which leads to the authorization conditions:

\[ \sum_{\alpha} \lambda^\alpha f_\alpha = f_0 \quad \text{for each } l \leq N(k) \]

It appears that the non-bias condition is automatically fulfilled.

The optimality condition which consists in minimizing an authorized linear combination of order \( k \) under the authorization conditions, introduces the generalized covariance \( K(h) \) and the \( N(k) \) Lagrange multipliers \( \mu_l \).

This yields to the kriging system and the kriging variance:

\[
\begin{align*}
\sum_{\alpha} \lambda^\alpha K_{\alpha\beta} + \sum_l \mu_l f_\alpha &= K_{00}^{a} \\
\sum_{\alpha} \lambda^\alpha f_\alpha &= f_0 \\
\sigma^2 &= K_{00} - \sum_{\alpha} \lambda^\alpha K_{\alpha0} - \sum_l \mu_l f_0
\end{align*}
\]

This kriging system is the same as in Universal Kriging except that \( K(h) \) is substituted for \( C(h) \). This is the benefit of the IRF–k model: we only need the generalized covariance rather than the ordinary covariance and therefore we have access to a much wider class of functions.

7.5 INFERENCE OF THE STRUCTURE

The inference of the structure for the IRF–k model, as it is carried out in the software package BLUEPACK 3-D, consists in finding:

i) the order \( k \) of the IRF

ii) the generalized covariance \( K(h) \)

Order of the IRF

To find the order \( k \), the idea is to delete each point in turn and to estimate it from the neighboring data values with several hypotheses for the order \( k \). The optimal order \( k \) corresponds to the one which minimizes the average error between the estimated value and the true data.
Obviously, as in the structure identification in the Universal Kriging model, we encounter a serious problem: when performing an estimation, we need to have already established the generalized covariance which precisely depends on the order of the IRF.

In fact, as a crude hypothesis, we temporarily consider that the generalized covariance is reduced to a pure nugget effect (which is authorized for any order of the IRF). The estimation will then coincide with the least squares fits of polynomials of different degrees.

Moreover, using the least squares technique also leads to a valuable by-product: for each tested point, the error between the true data and the value estimated using a polynomial of degree k constitutes an authorized linear combination of order k.

**Generalized covariance**

By definition of the generalized covariance, for any authorized linear combination \( Z(\lambda_m) \) is:

\[
\text{Var}[Z(\lambda_m)] = E[Z(\lambda_m)^2] = \sum_a \sum_\beta \lambda_m^a \lambda_m^\beta K_{a\beta}
\]

If we use a generalized covariance which is a linear combination of a finite set of elementary non-parametric structures \( K^p(h) \) with unknown coefficients \( b_p \), we are simply looking for the optimal (and authorized) set of \( b_p \) coefficients. We can write:

\[
E[Z(\lambda_m)^2] = \sum_p b_p \left( \sum_a \sum_\beta \lambda_m^a \lambda_m^\beta K_{a\beta} \right) = \sum_p K^p(\lambda_m)
\]

which is a regression equation of \( Z(\lambda_m)^2 \) on the \( K^p(\lambda_m) \). In order to determine the \( b_p \), we consider a large number of authorized linear combinations and minimize:

\[
Q(b) = \sum_m \omega_m^2 \left[ Z(\lambda_m)^2 - \sum_p b_p K^p(\lambda_m) \right]^2
\]

The weights \( \omega_m^2 \) are introduced to equalize the variances of \( Z(\lambda_m)^2 \).

Note that as the \( b_p \) must be positive, the minimization should be performed under constraints. Instead, as the number of possible models is usually small, we prefer trying all possible regressions, setting first one coefficient to zero, then two, and up to all of them but one ("pure" terms), and keeping only admissible results. Note that the regressions of the "pure" models always lead to admissible models as the corresponding coefficient:

\[
b_{p_0} = \frac{\sum_m (K^{p_0}(\lambda_m))^2}{\sum_m E[Z(\lambda_m)^2] K^{p_0}(\lambda_m)} \geq 0
\]

Among all the admissible results, we then have to select the best one. This is done by ranking the models according to the value of the ratio:
\[ r = \frac{\sum_{m} Z(\lambda_m)}{\sum_{m} \sum_{p} \hat{b}_p K_p(\lambda_m)} \]

which should be as close to 1 as possible.

### 7.6 EXTERNAL DRIFT

In terms of both the Universal Kriging and the IRF-k models, the drift was considered as a low order polynomial. This polynomial assumption for the drift may sometimes be improved if another variable is more likely to serve as an underlying shape factor.

For example in the petroleum industry, when \( Z(x) \) denotes the depth of a target horizon measured at a few wells, one may suspect that the marker pointed out by a seismic campaign could be used as this shape factor.

The only difference in the kriging system comes from the drift function as the standard monomials are substituted by the value of the shape factor: hence the external drift name.

The external drift is a first step in the multivariate approach of geostatistics although it must be noticed that:

- the shape factor variable is considered as the drift and its structural analysis needs not be performed.
- the main variable is honored whereas the shape factor variable only gives the underlying slow variations.
- the shape factor needs to be known at any point \( x_\alpha \) (where the main variable is measured) but also at each target point to be estimated.

### 7.7 EXERCISE: DUAL KRIGING

If we use genuine generalized covariance \( K(h) \), the kriging system is regular.

The kriging matrix is:

\[
\begin{pmatrix}
K_{\alpha\beta} & \lambda^\alpha_\beta \\
\lambda^\alpha_\beta & 0
\end{pmatrix}
\]

and its inverse matrix can be written:

\[
\begin{pmatrix}
b^{\alpha\beta} & \lambda^\alpha \\
\lambda^\alpha & \mu_m
\end{pmatrix}
\]

Let us write that their product is the identity matrix:

\[
\begin{align*}
\sum_{\gamma} b^{\alpha\gamma} K_{\gamma \beta} + \sum_{l} \lambda^\alpha_l f^l_\beta &= \delta^\alpha_\beta \\
\sum_{\gamma} \lambda^\alpha_{\gamma m} &= \delta^m_l \\
\sum_{\gamma} b^{\alpha\gamma} f^l_\gamma &= 0 \\
\sum_{\gamma} \lambda^\gamma K_{\alpha \gamma} + \sum_{m} \mu_m f^m_\alpha &= 0
\end{align*}
\]

76
The kriging weights are:

\[
\lambda^a = \sum_{\gamma} b^{a\gamma}K_{\gamma x} + \sum_{m} \lambda^{a,m}_{m^\gamma x}
\]

and the kriging estimate can be written:

\[
z^*(x) = \sum_{a} \lambda^a z_a = \sum_{a,\beta} b^{a\beta}z_{\beta x}K_{\beta x} + \sum_{a,m} \lambda^{a,m}_{m^\gamma x} d^{m}_{m^\gamma x}
\]

Introducing:

\[
b^\gamma = \sum_{a} b^{a\gamma}z_a
\]

\[
c^m = \sum_{a} \lambda^{a}_{m^\gamma x} z_a
\]

we can finally write the kriging estimate in its **dual form**:

\[
z^*(x) = \sum_{\gamma} b^\gamma K_{\gamma x} + \sum_{m} c^m_{m^\gamma x}
\]

The \( b^\gamma \) matrix has the following property:

\[
\sum_{\gamma} b^{\gamma\gamma} = \sum_{a,\gamma} b^{a\gamma}z_{\gamma a} = \sum_{a} z_{a} \left( \sum_{\gamma} b^{a\gamma} \right) = 0
\]

Suppose we have a 1-D irregular sampling, of a IRF-1, characterized by the 1-D thin plate spline generalized covariance: \( K(x) = |x|^3 \). Examine the shape of the kriging surface given by the dual formulation:

\[
z^*(x) = \sum_{\gamma} b^\gamma |x - x_{\gamma}|^3 + \sum_{m} c^{m}_{m^\gamma x}
\]

<table>
<thead>
<tr>
<th>( Z(x_1) )</th>
<th>( Z(x_2) )</th>
<th>( Z(x_3) )</th>
<th>( Z(x_4) )</th>
<th>( Z(x_5) )</th>
<th>( Z(x_6) )</th>
</tr>
</thead>
</table>

- Estimating below \( x_1 \) or above \( x_6 \), we can write:

\[
z^*(x) = x^3 \sum_{\gamma} b^\gamma + x^2 \sum_{\gamma} b^\gamma x_{\gamma} + x \sum_{\gamma} b^\gamma x^2_{\gamma} + \sum_{m} b^\gamma x^3_{\gamma} + \sum_{m} c^{m}_{m^\gamma x}
\]

but as \( \sum_{\gamma} b^{\gamma\gamma} = 0 \), then \( z^*(x) = \sum_{m} c^{m}_{m^\gamma x} \) is a polynomial or order 1.

- Within an interval \( [x_i, x_{i+1}] \), the kriging surface is a third order polynomial.
- At each conditioning point \( x_i \), because the kriging is an exact interpolator:

\[
z^*(x_i) = z(x_i)
\]

- The second order derivative of the kriging function is continuous at each data point:

\[
(z^*)''(x) = 6 \sum_{\gamma} b^\gamma |x - x_\gamma|
\]

Moreover, at the last point \( x_6 \) (and similarly at the first point \( x_1 \)), the second order derivative of the kriging function is zero:

\[
(z^*)''(x_6) = 6 \sum_{\gamma < 6} b^\gamma (x - x_\gamma) = 6x - 6 \sum_{\gamma} b^\gamma x_\gamma = 0
\]

References on IRF-k and external drift

8. RELATED VARIABLES

8.1 THE STRUCTURAL TOOLS

The extension of stationary and intrinsic RF to several variables is straightforward. Besides the usual covariance or variogram, there is now the cross covariance or the cross variogram.

The stationary cross covariance is

\[ C_{12}(h) = \text{Cov} [Z_1(x), Z_2(x + h)] \]
\[ = E[Z_1(x) - m_1] [Z_2(x + h) - m_2] \]

When divided by the standard deviations it is the correlation \( \rho_{12}(h) \) between \( Z_1(x) \) and \( Z_2(x + h) \). There is no spatial correlation when this is zero for all distances \( h \) (and not only for \( h = 0 \) which corresponds to the variables at the same point \( x \)).

Of course

\[ \rho_{12}(h) = \rho_{21}(-h) \]

but this may not be equal to

\[ \rho_{12}(-h) = \rho_{21}(h) \]

An example of this is when a spatial delay exists between the variables (e.g. if their maximal correlation is observed for a non zero distance).

The cross variogram is defined as

\[ \gamma_{12}(h) = \frac{1}{2} E[Z_1(x + h) - Z_1(x)] [Z_2(x + h) - Z_2(x)] \]

It can be negative. This happens when one variable increases while the other decreases (substitution of one metal by another one, or of one fish age class/species by another one).

When there is a stationary cross covariance we have

\[ \gamma_{12}(h) = C_{12}(0) - \frac{1}{2} [C_{12}(h) + C_{12}(-h)] \]

so no delay can appear on the cross variogram. If the cross covariance is symmetrical we have

\[ \gamma_{12}(h) = C_{12}(0) - C_{12}(h) \]

These tools make it possible to compute variances of linear combinations.
\[ \text{Var} \left( \sum_i \sum_\alpha \lambda_i^\alpha \; Z_i(x_\alpha) \right) = \sum_i \sum_j \sum_\alpha \sum_\beta \lambda_i^\alpha \lambda_j^\beta \; C_{ij}(x_\alpha - x_\beta) \]

or, under the conditions \( \sum_\alpha \lambda_i^\alpha = 0 \quad \forall \lambda \):

\[ \text{Var} \left( \sum_i \sum_\alpha \lambda_i^\alpha \; Z_i(x_\alpha) \right) = - \sum_i \sum_j \sum_\alpha \sum_\beta \lambda_i^\alpha \lambda_j^\beta \; \gamma_{ij} (x_\alpha - x_\beta) \]

### 8.2 COREGIONALIZATION MODELS

Fitting auto and cross variograms or covariances has to be done in a mathematically consistent way, if we want to avoid negative variances. However more specific models sometimes have to be used. For instance if we want to map the topography using topo values and gradient values, we need a model for the two variables topo and gradient. This model is entirely derived from the topo variable, which has to be differentiable. We can also introduce non stationary models of coregionalization. Here we only consider the linear model of coregionalization in the stationary case.

**Linear model**

In fact these models are already used in the univariate case. Let us consider a variable \( Z(x) \) with structure

\[ 9 \; \text{Sph}(20) + 16 \; \text{Sph}(100) \]

It amounts to saying that

\[ Z(x) = 3 \; Y_1(x) + 4 \; Y_2(x) \]

where \( Y_1(x) \) and \( Y_2(x) \) are spatially uncorrelated components, with variance 1 and spherical structures with ranges 20 and 100 respectively.

Let us now consider two variables with structures

- \( 25 \; \text{Sph}(20) + 100 \; \text{Sph}(100) \)
- \( 100 \; \text{Sph}(20) + 100 \; \text{Sph}(100) \)

and cross structure

\( -24 \; \text{Sph}(20) + 60 \; \text{Sph}(100) \)

This is an admissible linear coregionalization model as we can write

\[ Z_1 = 3 \; Y_1 + 4 \; Y'_1 + 6 \; Y_2 + 8 \; Y'_2 \]
\[ Z_2 = -8 \; Y_1 + 6 \; Y''_1 + 10 \; Y_2 \]
where the components $Y_1$, $Y'_1$, $Y''_1$, $Y_2$, $Y'_2$ are spatially uncorrelated, with variance 1, and spherical structures with ranges 20, 20, 20, 100, 100.

8.3 COKRIGING

Kriging estimates $Z_1$ linearly from data on the same variable $Z_1$. Similarly cokriging estimates $Z_1$ or $Z_2$ linearly from data on $Z_1$ and $Z_2$ (and other possible variables). The different variables do not need to be known at the same points (heterotopy).

Deriving the cokriging system can be done as an exercise. The estimation variance should be written and minimized. Conditions on the weights for any of the variables should be introduced when their mean is unknown.

The different variables can represent different types of measurements of the same original variable, which may differ in their nugget components (variance of error) and in their mean (bias coming from a systematic error).

It is also possible to cokrige one of the structural components (for instance $Y_1$) using data on $Z_1$ and $Z_2$. This is called factorial kriging analysis. However the decomposition of several variables into components is generally not unique, and their choice is conventional. In the previous example the decomposition

$$Z_1 = 4 \ Y_1 + 3 \ Y'_1 + 10 \ Y_2$$

$$Z_2 = -6 \ Y_1 + 8 \ Y''_1 + 6 \ Y_2 + 8 \ Y'_2$$

is also possible and results in different components.

8.4 MAITRESSE VARIABLE

When dealing with several variables, it is tempting to use regressions and residuals. However this does not generally correspond to a spatial decomposition, and we will see why. The case of a maitresse variable (in French a variable is feminine and maitresse means mistress in many connotations) is indeed particularly interesting.

Let us consider two variables $T(x)$ and $Q(x)$. It is always possible to regress $Q(x)$ knowing $T(x)$ at the same point $x$. This implies that the residual $R(x) = Q(x) - E[Q(x)|T(x)]$ is uncorrelated with $T(x)$. But in general this residual has no more reason to be uncorrelated with $T$ for other points, than the residual of $T(x)$ knowing $Q(x)$ has to be uncorrelated with $Q$ at other points. So splitting a variable into its regressed value and a residual does not correspond to a spatially interesting decomposition.

We say that $T(x)$ is a maitresse variable if, knowing $T(x)$ at point $x$, $Q(x)$ at the same point does not depend on $T$ at any other points. This implies that the residual $R(x) = Q(x) - E[Q(x)|T(x)]$ is uncorrelated, not only to $T(x)$, but also to $T(x + h)$. The variable $Q(x)$ can be split into its regressed value and the residual, and these two terms are spatially uncorrelated.

For cokriging only linear combinations are used, so that the strict regression should be replaced by the linear regression $Q_{(x)}^R = aT(x) + b$ with residual $R(x) = Q(x) - Q_{(x)}^R$. We have

$$Q(x) = aT(x) + b + R(x)$$

81
R(x) is spatially uncorrelated with T(x), and the cross variogram between these two is zero. Experimentally we usually compute the variograms of T(x) and Q(x) first. Then their cross variogram should be similar to the variogram of the maitresse variable T(x). The coregionalization model can thus be easily seen when T and Q are very differently structured.

Cokriging T and Q is simplified as it reduces to kriging the variables T and R separately

\[ T(x)^{ck} = T(x)^k \]

\[ Q(x)^{ck} = aT(x)^k + b + R(x)^k \]

This model has been used by Bordessoule & al. 1989 for a multilayer deposit evaluation. In that case the variable T was the 0–1 geometric indicator of mineralized layers, which regularizes into the ore proportion, whereas Q was the grade, which gives the metal quantity. The model says that the metal quantity at a place only depends on the ore quantity at this very same place. In average Q is proportional to T (with a ratio which is the mean grade of mineralized parts). The structure of Q is a combination of the structure of the geometry (which is more continuous) and the structure of the residual (which has a short structure because of erratic grades).

It is thought that such maitresse variables are likely to be conditioning geometrical or topological variables, as is also the case in the next example.

Chautru 1987 has studied the distribution of polymetallic nodules on the sea floor (North Pacific). Two main factors had been selected, to study the possible exploitation of the nodules: their abundance Q(x), and the local slope T(x). This topographic variable indicates whether it is possible to make the exploitation process pass at this point. But this variable also controls the nodules abundance and acts as a maitresse variable.

Chautru has made simulations of nodule fields. This made it possible to see how really are (or can be thought to be) the accessible and rich areas. The relations between the two variables were in fact non linear. Slope and residual could be considered as independent once normalized, and were then used to obtain 2D simulations of slope and nodule abundance.

**8.5 EXERCISE**

Consider two positive, independent, and stationary Random Functions T(x) and Z(x). Let Q(x) be the Random Function defined as their product Q(x) = Z(x) T(x).

1) Show that the expected value of Q(x) knowing T(x) at the same point x is:

\[ E \left[ Q(x) \mid T(x) \right] = m \cdot T(x) \]

where \( m = E[Z(x)] = E[Q(x)] / E[T(x)] \).

Note that the regression is linear.

2) Consider the residual of this regression

\[ R(x) = Q(x) - E \left[ Q(x) \mid T(x) \right] = [Z(x) - m] \cdot T(x) \]

It is stationary. Show that it is not correlated with T(x) at the same point x. Note however that it depends on T(x).
3) Show that \( R(x) \) is not correlated with \( T(x+h) \) at any other point \( x+h \) either. Deduce that \( Q(x) \) can be split into two spatially uncorrelated components

\[
Q(x) = m \ T(x) + R(x)
\]

4) This model can sometimes be used to fit two variables \( T(x) \) and \( Q(x) \) and to obtain their cokriging. Suppose that \( T(x) \) can take zero values. Show that \( Z(x) \) need not to be known at points where \( T(x) = 0 \).

5) What happens when \( T(x) \) is the 0-1 indicator of a stationary Random Set?

6) What happens if \( T(x) \) is not stationary?

7) What happens if \( T(x) \) is a deterministic function of \( x \)?
9. OTHER MODELS

When looking outside linear geostatistics, we find models that are used for nonlinear geostatistics and/or simulation. We have already mentioned the smoothing effect of kriging. In contrast to this performing a simulation reproduces the real variability. In mining or oil industries, simulated deposits are used to test some exploitation processes.

Nonlinear geostatistics is mainly used to estimate nonlinear functions of the variable $Z(x)$ under study. In certain cases such as pollution, we need to know whether the true value $Z(x)$ exceeds a specified threshold value $z$. For instance the fact that the estimated value for $Z(x)$ is 2 does not ensure that true unknown value $Z(x)$ does not exceed 3. We then have to estimate the 0–1 indicator variable $I [Z(x) ≥ z]$, which tells whether the true $Z(x)$ exceeds the cutoff $z$. Estimates of this indicator have the meaning of a probability (the probability that $Z(x)$ exceeds $z$ given the available information).

In mining estimation, we want to estimate $I [Z(v) ≥ z]$ for the future selection blocks $v$. So we need something more – a change of support model. This makes it possible to go from the known sample distribution to the unknown block distribution (same mean, variance diminished by $D^2(0 | v) = \bar{f}(v, v)$). These change of support models will not be discussed here.

Models for nonlinear geostatistics or simulation are richer and more powerful than just linear ones. But they have to be fitted to the same data. That means that using these models requires stronger assumptions, but also the data have to be looked at very carefully. Having to look more closely at data and coming to a better understanding of the results would be one of the immediate benefits to be obtained from using nonlinear methods for fisheries.

9.1 GAUSSIAN MODEL

This is the most commonly used model. A stationary Random Function is gaussian if its multivariate distributions are gaussian, which is much stronger assumption than just its marginal distribution. The gaussian model has very convenient properties. We know how to compute the conditional distribution of $Y(x)$ at point $x$ knowing the values $Y(x_1), \ldots, Y(x_n)$ at other points. The mean of this distribution, i.e. the conditional expectation, is linear and corresponds to the simple kriging (kriging with known mean).

Moreover the residual of this regression is independent of the data values (and consequently of all kriged values). This gives a way to build what is called conditional simulation, i.e. a simulation which honour the data values at the sample points. It is obtained by adding a map of independently simulated residuals to the kriged map.

Lognormal

In practice experimental distributions are rarely gaussian and so a transformation is used to transform them to a normal one. For instance if $Z(x)$ is lognormal, $\log Z(x)$ is (marginally) normal. Assuming that the pairs Log $Z(x)$, Log $Z(x + h)$ are gaussian, there is a relation between the stationary covariance function $C(h)$ of $Z(x)$ and the covariance $\sigma (h)$ of Log $Z(x)$

$$C(h) = m^2 [\exp \sigma(h) - 1]$$
where \( m = \text{E}(Z) \). With a large logarithmic variance \( \sigma \), the lognormal distribution is skew and has a few, relatively large values. This may make the direct inference of the raw structure difficult. Using the above formula gives a way to obtain this structure from the variogram of the gaussian transformed, which is usually better known.

As the structure of the normal variable is more continuous than for the raw variable, kriging it is easier. But be careful: taking the antilog of this kriged logarithm gives a biased estimator for the raw variable. See second exercise.

Exercise 1

Suppose \( Y(x) = \text{Log } Z(x) \) has a spherical structure with range 10 and sill (= logarithmic variance) 2. The mean \( m \) of \( Z(x) \) is 0.56. Compute the value of the structure of \( Z(x) \) for \( h = 0, 1, 2, \ldots, 10 \). Compare the two structures. What happens to the structure if the mean of \( Z(x) \) is 1.12 instead of 0.56?

Exercise 2

Let \( Y(x) = \text{Log } Z(x) \) be a normal Random Function with mean 0 and variance 1. Given the values \( Y(x_1), \ldots, Y(x_N) \), the value \( Y(x) \) at point \( x \) is normal with mean \( Y(x)^{CE} \) and variance \( \sigma^2 \) (the kriging variance).

The conditional expectation of \( Z(x) \) knowing the values \( Y(x_1), \ldots, Y(x_N) \) is the average value we can expect at this point. It can be proved to be

\[
Z(x)^{CE} = \exp \left( Y(x)^K + \frac{\sigma^2}{2} \right)
\]

Note that it is not merely the antilog of the kriged logarithm (which would give a biased estimator for the raw variable).

The conditional variance gives a measure of how badly known \( Z(x) \) is. It can be written as

\[
CV(x) = \left[ Z(x)^{CE} \right]^2 \left[ \exp \sigma^2 - 1 \right]
\]

We now consider 4 points \( x \) corresponding to the following values

<table>
<thead>
<tr>
<th>( Y(x)^{CE} )</th>
<th>( \sigma^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Compute the values of \( Z(x)^{CE} \) and \( CV(x) \) for these points. Suppose you can sample one of these points. Which would you choose? Which two points would you sample? Why?

85
Transformed gaussian

In practice distributions are not exactly lognormal. Fitting a lognormal distribution on the empirical distribution may cause problems. We then prefer to make an empirical transformation (called an anamorphosis) to get a normal distribution. This transformed gaussian model is a generalization of the lognormal model. There is a generalized formula for the covariance, but it is more complicated.

However such a transformation is not always possible because the normal distribution is continuous. It is not possible when there are many identical data values (for instance, many zeroes) or many values which are practically the same. Artificial forcing would result in bivariate distributions which would be far from gaussian. Other models (gamma and gamma transformed, negative binomial, ...) are available for these cases. The following model is not based on a given statistical distribution, as normal or gamma, but on geometrical properties.

9.2 INDICATOR RESIDUAL MODEL

Let us imagine a Random Function $Z(x)$ with 3 classes of values, say 1, 2 or "much". Class 1 corresponds to small values (they will be considered as equal to 1), class 2 to medium values (say 2), and class 3 corresponds to the larger values (the tail of the distribution). The set $A$ of points with values $>1$ may be considered as windows, through which we observe the values 2 and much. We consider two hypotheses:

First hypothesis: among these values $>1$ the much group form small patches which can be anywhere within these windows (on the borders as well as in the centers).

Second hypothesis: inside a patch, whatever its position, we can find any large values from the tail of the distribution.

Thus, providing the value of a point $x$ is $>1$, its value can be 2 or much, no matter whether a neighbouring point value $x+h$ is $>1$ or not. Moreover, if it is much, it can take any possible large value.

Because the much values can be very large, the inference of the structure of $Z(x)$ may be difficult. Moreover in such a case, the variogram of the variable $Z(x)$ poorly represents the structure of the phenomenon. The main structure is given by the geometry of the set $A$, i.e. the points with value $>1$. The second structure corresponds to the patches within this set, that is, it is conditioned by $A$.

Experimentally the relation between these geometries can be tested by computing

$$P(Z(x) > 2 \mid Z(x) > 1, Z(x + h) < 1)$$

As an exercise you can show that this probability is the ratio between the cross variogram of the indicators $I[Z(x) > 1]$ and $I[Z(x) > 2]$ and the variogram of $I[Z(x) > 1]$. Does this ratio increase with $h$? This means that large values are present less often on the borders of $A$. In this case "diffusion type" models should be used (Gaussian, gamma, negative binomial are diffusion models). Here this ratio does not depend on $h$.

The indicator $I[Z(x) > 1]$ plays the role of a maîtresse variable. The other indicator $I[Z(x) > 2]$ is structurally subordinated to the first one. It can be split into two terms

$$\frac{I[Z(x) > 2]}{P[Z(x) > 2]} = \frac{I[Z(x) > 1]}{P[Z(x) > 1]} + R(x)$$
where
\[ R(x) = \frac{P[Z(x) > 2]}{P[Z(x) > 2]} - \frac{P[Z(x) > 1]}{P[Z(x) > 1]} \]

is the residual of the regression of \( \frac{P[Z(x) > 2]}{P[Z(x) > 2]} \) knowing \( P[Z(x) > 1] \) and is spatially uncorrelated with \( P[Z(x) > 1] \). As an exercise, prove the formula of the regression between the two indicators.

The structure of the second indicator \( P[Z(x) > 2] \) appears as being made of the structure of the first indicator and of the structure of the residual.

This model can be extended to more classes, and the structure of \( Z(x) \) can be expressed with the structures of the first indicator and of the successive residuals. See Rivoirard 1989, 1990. This model has also been used by Petitgas 1991 to describe Norwegian herring acoustic densities.

9.3 BOOLEAN SCHEME

Attracted by their geometrical properties, we have diverged from Random Functions to Random Sets. There are also lots of other models which are used in Spatial Statistics: Random Point Processes, Random Partition models. Here we will present one basic Random Set model, which is build from Poisson points (Matheron 1967, Serra 1982).

Poisson points in 2D are independently and randomly distributed points, such that the numbers of points within each of any two regions are independent when there is no intersection between these regions. If \( a \) is the mean density of points, the number of points within an area \( A \) has a Poisson distribution with parameter \( aA \). For two non intersecting areas \( A \) and \( A' \) the distribution of the number of points is a Poisson with parameter \( a(A + A') \).

When the number of points over a region \( A \) is known, the different points are uniform and independent within \( A \). This gives a way of simulating this point process.

At each point we now put a "grain" (a circle for example). Two grains can overlap. The reunion of all grains is the Boolean scheme. In this model the porosity is
\[ q = \exp(-aK(0)) \]

and the non centered covariances of pores is
\[ \exp(-a(2K(0)-K(h))) \]

where \( K(h) \) is the geometric covariogram of a grain. Moreover grain size and shape can be random, in which case the covariogram has to be replaced by its expectation.

The Boolean scheme might help to describe the patches of high fish density within favourable zones.
REFERENCES

Some of these references are notes of the Centre de Géostatistique de Fontainebleau, which are publications of the Ecole des Mines de Paris and can be obtained by writing to the Librarian at the Centre.


Delfiner, P., Basic Introduction to Geostatistics.– Summer School, School, Fontainebleau, September 1979.– 129 p.


Rivoirard, J., Introduction to disjointive kriging and nonlinear geostatistics.— Décembre 1990.— 89 p. : bibliogr.

