Basic Linear Geostatistics

Springer-Verlag Berlin Heidelberg GmbH

Margaret Armstrong

Basic Linear Geostatistics

With 59 Figures

Springer

Dr. Margaret Armstrong

Ecole des Mines de Paris
Centre de Géostatistique
35 rue St.-Honoré
77305 Fontainebleau / France
E-mail: marg@cg.ensmp.fr

ISBN 978-3-540-61845-4

Library of Congress Cataloging-in-Publication Data
Armstrong, M., 1950- Basic linear geostatistics / Margaret Armstrong. p. cm.
Includes bibliographical references and index.
ISBN 978-3-540-61845-4 ISBN 978-3-642-58727-6 (eBook)
DOI 10.1007/978-3-642-58727-6
1. Mine valuation -- Statistical methods. I. Title
TN272.7.A76 1998 622--dc21 98-34424

springeronline.com

Originally published by Springer-Verlag Berlin Heidelberg New York in 1998

Typesetting: Camera-ready by author
SPIN:10995952 32/3111 - 5 4 3 2 1 - Printed on acid-free paper

Preface

The Centre de Géostatistique has had a long tradition of teaching geostatistics, starting with the summer schools in the early seventies. Over the past twenty years, the CFSG postgraduate course has trained more than 150 engineers and geologists to be mining geostatisticians. The aim of this 9 month postgraduate course which is partially funded by the French government via the CESMAT has always been to train professional geostatisticians – specialists who will work in industry after graduating. When the CFSG started, there were plenty of textbooks in both English and French. The legendary *Fascicule 5*, Matheron's course notes for the 1970 summer school, provided the theoretical benchmark; the classic books by André Journel and Charles Huigbregts, Michel David and Isobel Clark which were just hot off the presses, gave practical insights into how to apply the theory in the mining industry. These books are now a bit out of date. It seemed to be time for a new book on basic linear geostatistics – not that the theory has changed in the meantime, but ideas on how to apply it have evolved as more case studies were done.

This book gives my view on how to apply linear geostatistics (variograms and kriging), especially in a mining context. Getting geostatistics to work in practice requires some theoretical knowledge together with practical know–how. Not enough theory and people make silly mistakes like choosing functions that are not positive definite as variogram models, giving rise to negative variances. Not enough practical know–how and people cannot see how to interpret experimental variograms. I hope that this book will contain the right mixture of theory and practice to allow engineers and geologists, particularly those living in remote locations, to apply geostatistics to their own data.

The first chapter in the book illustrates two key concepts in geostatistics: the support effect and the information effect. Chapter 2 introduces the different stationarity hypotheses. The next three chapters concentrate on the variogram. Chapter 6 takes up the topic of support again, but this time from the point of view of regularization. The next three chapters are devoted to kriging, the name used for the different geostatistical estimation methods. The final chapter treats the topic of estimating the total reserves. Some reminders on basic mathematics and statistics are included in Appendix 1. Lastly given the increasing trend toward law suits, it seems important to propose guidelines on how to carry out case–studies in industry. The key to this is keeping a "logbook" for each case–study. This is described in Appendix 2.

These course notes have evolved over the years through teaching the CFSG students. Their comments and suggestions have been very helpful. Particular thanks are due to Malcolm Thurston and Mike Harley who kindly allowed part of their CFSG projects to be published in this book. Last but not least, I would like to thank Michel Schmitt, Chris Roth and Jane Bocquel for their invaluable help in preparing the manuscript.

Margaret Armstrong
Fontainebleau, May 1998

Table of Contents

1 Introduction

1.1 Summary

After outlining the types of problems in mining that geostatistics can be used to solve, an introductory exercise is presented. It illustrates the need for good estimators, particularly when selective mining is being used. The economic impact of the support and information effects on reserve calculations is stressed. Lastly some case studies comparing geostatistics with other estimation methods are reviewed.

1.2 Introduction

Over the past 30 years, geostatistics has proved its superiority as a method for estimating reserves in most types of mines (precious metals, iron ore, base metals etc.). Its application to the petroleum industry is more recent, but it has nevertheless demonstrated its usefulness, particularly for contour mapping and for modelling and simulating the internal heterogeneity of reservoirs. Its use has been extended to other fields such as environmental science, hydrogeology, agriculture and even fisheries, where the time component as well as the spatial variability is important.

The basic tool in geostatistics, the variogram, is used to quantify spatial correlations between observations. Once a mathematical function has been fitted to the experimental variogram, this model can be used to estimate values at unsampled points. This estimation procedure is called "kriging" after the South African engineer, Danie Krige, who with Herbert Sichel carried out the first developments in geostatistics in the Witwatersrand gold mines. After reading an early paper written by Krige, the French mathematician, Georges Matheron, saw its implications and went on to develop the theory in the sixties and seventies. Before going into detail about the variogram and the different types of kriging, the main uses of geostatistics in mining are outlined.

1.3 Applications of geostatistics in mining

1.3.1 Estimating the total reserves

The first important step in a feasibility study is to determine the in situ reserves. Geostatistics can help the mine planner get accurate estimates of the total tonnage in situ, the average grade and the quality from the available information, and thus help him decide whether further investment in the project is warranted.

1.3.2 Error estimates

No estimation method can give exactly the right value all the time since there is inevitably some error involved. So it is important to know how serious this error is. Decision makers need to know whether the estimated grade is accurate to $\pm 0.1\%$ or to $\pm 1\%$. As well as giving the estimated values, geostatistics provides a measure of the accuracy of the estimate in the form of the kriging variance. This is one of the advantages of geostatistics over traditional methods of assessing reserves.

1.3.3 Optimal sample (or drillhole) spacing

The estimation variance (calculated by geostatistics) depends on the variogram model chosen for the deposit and on the location of the samples, but not on their numerical values. So once the variogram has been selected for a particular deposit or region, the estimation variance can be found. This makes it possible to evaluate the estimation variance for a wide variety of possible sample patterns without actually doing the drilling, and hence to find the grid that just gives the required accuracy.

1.3.4 Estimating block reserves

Once a decision has been made to mine a deposit, estimates of the tonnage and the average grade are needed block by block. Here a block might represent the production for a shift, or for a month. In addition to estimating the ore tonnage and the average grade of mining blocks, geostatistics can provide estimates of quality variables. For coal these include ash content, sulphur content, F.S.I. and calorific value. For iron ore, they are the percentages of silica and phosporus, loss on ignition and sometimes manganese content.

1.3.5 Gridding and contour mapping

Although most mining companies usually want block estimates of their variables rather than contour maps, geostatistics can be used to estimate the values at the nodes of a regular grid. After this, a standard contouring package can be used to do the plotting. This has the advantage of being more accurate than other methods of evaluating grid node values. Over the past 25 years the petroleum industry has been turning more and more to kriging for this. More recently environmental scientists have also started using geostatistics.

1.3.6 Simulating a deposit to evaluate a proposed mine plan

Since kriging is designed to give the minimum variance linear estimates, the kriged values are smoother than any other unbiased linear estimators but they are also smoother than the real values. This means that if a numerical model of a deposit is being set up to test various proposed mine plans, the kriged values should not be fed into this because they would seriously under-estimate the inherent variability. In this case a conditional simulation of the deposit should be used. More information on when to use simulations rather than kriging is given in Chapter 9.

1.3.7 Estimating the recovery

In many mining operations engineers have to predict the recovery and the recovered grades when blocks of a specified size are selected for treatment (or mining) if their average grade is above an economic cutoff. When the sample grid is about the same size as the selection blocks their grades can be estimated individually with reasonable accuracy. But if the blocks are much smaller than the grid size as is usually the case at the feasibility stage, it can be misleading trying to get estimates of individual blocks. These are simply not accurate enough. The best that can be done is to predict the proportion of selection units that will be recovered, and their average grade. This leads in to nonlinear geostatistics.

Similar problems arise in soil rehabilitation work where scientists have to predict the total amount of material that is contaminated, i.e. contains unacceptably high levels of the pollutant.

In the subsequent chapters we go on to see what the variogram is and how kriging is used to estimate values and to obtain the estimation variance. As this text deals only with linear geostatistics, it does not go into more advanced topics such as conditional simulaton or recoverable reserve estimation using nonlinear methods.

1.4 The $64 question : does geostatistics work?

Having seen some of the possible applications of geostatistics in the mining industry, the $64 question is: "Does geostatistics work?" or "Does it work better than

the alternative reserve estimation methods?". First we present a simplified example showing the financial impact of poor block estimates. Then several comparative case studies on ore evaluation techniques are reviewed.

1.5 Introductory exercise

One of the most common uses of geostatistics (in the mining industry) is to estimate the average grades of mining blocks by kriging – for example, for day to day grade control. Many people wonder whether kriging really does give better results than other methods. This exercise has been designed to highlight the economic impact of estimation errors. First let us see the data. The grades of 64 blocks of size 1 x 1 were available in an area 8 x 8. Figure 1.1. shows 16 of these grades, each from the top left corner of a block of size 2 x 2. The values of the other 48 samples will be used later for comparison purposes.

735	45	125	167
450	337	95	245
124	430	230	460
75	20	32	20

Fig 1.1. Sixteen samples of size 1 x 1 to be used to estimate mining blocks of size 2 x 2

These 16 values will be used as the "samples" to estimate the values of mining blocks of size 2 x 2 (i.e. 4 times larger than the samples). The simplest way of estimating each of the 16 block grades is by equating the grade of the sample in each one to the block estimate. This is called the polygonal method. Figure 1.2.**a** gives these estimates.

Now it is your turn to design a way of estimating each of these 16 values. You may choose whatever method you like; for example by looking at the values and seeing how they vary, or by guessing, or by taking moving averages of neighbouring values. Write your estimates in the space provided on the right of Fig. 1.2.**b**.

735	45	125	167
450	337	95	245
124	430	230	460
75	20	32	20

a

b

Fig 1.2. (**a**) Polygonal estimates of block grades obtained by equating the block grade with the sample value inside it; (**b**) space for the reader's personal estimate

442	190	142	204
354	276	212	279
189	226	216	271
99	81	88	125

Fig 1.3. Kriged estimates of block grades

A third set of estimates was obtained by having a geostatistician krige the block values (Fig. 1.3.). Kriging is just a special sort of weighted moving average. You are not expected to understand how these numbers were obtained yet. For the present, they are just another possible set of block estimates.

1.5.1 Selective mining

In most mining operations, the high grade blocks are mined while the others are either left in place or are dumped as waste. Suppose that in this case the economic cutoff is 300. So mining a block with a grade of 301 leads to a profit of 1 unit, and conversely mining a block with a grade of 299 leads to a loss of 1 unit. For the time being we are going to ignore any geometric constraints due to the mining method, and we will assume that all pay blocks are mined.

The first step is to calculate the profits that would be predicted by each of the three estimation methods. We shall work through this together for the polygonal estimation and then you can repeat it for the other two. Firstly, shade in any blocks with a predicted grade above 300 (or equal to it). There are five of them. See Fig. 1.4.**a**.

So the expected profit is: 735+ 450 + 337 + 430 + 460 – 5 x 300 = 912

Now we calculate the actual profits that would be made when the blocks estimated to be above cut off are mined. The true grades of the 2 x 2 blocks are given in Fig. 1.4.**b**. For the polygonal method, five blocks (the shaded ones) are scheduled for exploitation. Their real grades are 505, 270, 328, 220 and 263 rather than the estimated 735, 450, 337, 430 and 460.

So the actual profit would be: 505 + 270 + 328 + 220 + 263 – 1500 = 86

Instead of earning a profit of 912 units, the mine makes only 86 units. The company could well end up in serious financial difficulties. Repeat these calculations for kriging. Show that only two blocks are scheduled for mining and that the actual profit is 175 compared to a predicted profit of 196. Lastly repeat the calculation for your own estimator and note the results.

735	45	125	167
450	337	95	245
124	430	230	460
75	20	32	20

a

505	143	88	207
270	328	171	411
102	220	154	263
101	54	44	155

b

Fig 1.4. (**a**) Polygonal estimates and (**b**) true block grades. Shaded blocks with a grade above 300 are scheduled for mining. Note the difference between true and estimated grades

1.5.2 Optimal recovery

Before comparing these results, we should calculate the optimal recovery (i.e. what we would recover had we known the true grades before mining). Clearly only three blocks would have been selected (505, 328 and 411) and the profit would have been 344. Compared to this, the profit predicted by the polygonal method (912) was quite

illusory. Kriging gave a much more realistic prediction of 196 compared to an actual profit of 175 (i.e. only 10% difference) but this is still suboptimal compared to 344.

Now it is interesting to see why kriging works better, on average, than other estimation methods. We will see that the problems met when estimating blocks are due to two effects: the information effect and the support effect.

1.5.3 Information effect

The information effect is due to the incomplete information available at the time when we must discriminate between waste and ore blocks. We have only estimates for the block grades instead of the real ones. To visualize this, we draw scatter diagrams of the true grade (Y axis) against the estimate (X axis) for different estimation methods. Ideally the estimated grade would be equal to the true one, so the points would fall on a 45° line passing through the origin. Unfortunately they do not. They form a cloud of points which has been represented here as an ellipse.

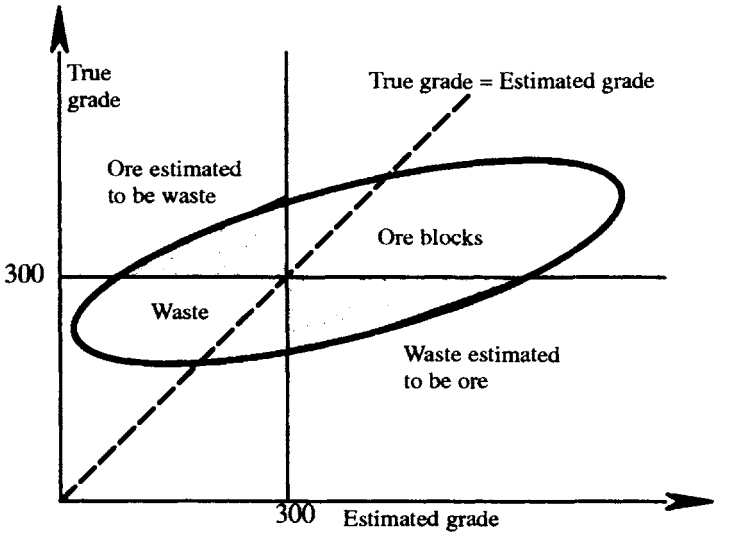


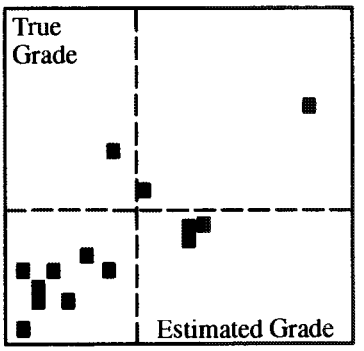
Fig 1.5. Crossplot of true grade versus estimated grade. The cloud of points has been represented as an ellipse. Blocks with an estimated grade above 300 are scheduled for mining whereas those blocks actually above 300 should be mined

When selecting blocks for mining, all the blocks whose estimated value is above the cutoff are considered to be ore. To show this graphically, a vertical line is drawn at X=300. The blocks to the right of this line are selected for mining. What we actually wanted was the blocks whose true grade is above 300. A horizontal line drawn at Y=300 represents this. The blocks above this line should have been mined. This divides the whole area into four zones:

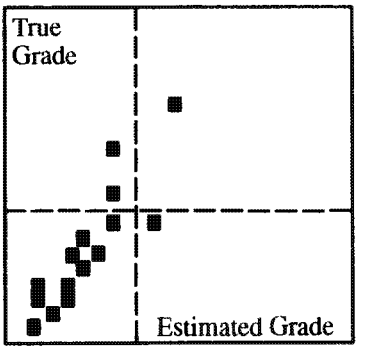
1. True grade > 300; estimated grade > 300. These ore blocks are correctly estimated as ore. They correspond to the upper right part of the diagram.

2. True grade < 300; estimated grade < 300. These waste blocks are correctly estimated as waste. They lie in the lower left part of the diagram.
3. True grade >300; estimated grade <300. These ore blocks have been considered to be waste; this estimation error can have costly consequences for the mine. These blocks lie in the upper left part of the diagram.
4. True grade <300; estimated grade >300. These waste blocks have been considered to be ore. This second type of estimation error does not cancel out the preceding one and can have expensive consequences for the mine. These blocks are in the lower right part of the diagram.

Going back to our example, Fig. 1.6. shows the crossplots corresponding to the polygonal method and to kriging. For kriging the slope of the regression is approximately 1.0 (i.e. at 45°) wheras it is less than 1.0 for the polygonal method. Now look at the "fatness" of the two clouds. Kriging effectively gives a "thinner" cloud. The reader can go back to the two scatter diagrams and see the misallocated blocks for each of the estimation methods (in the upper left and lower right quadrants). This confirms that kriging is better. In Chapter 8 we shall see that the criteria for judging an estimator include the slope of the regression line of the true value on the estimated one.



a Polygonal Estimates



b Kriged Estimates

Fig 1.6. Crossplot of true grade versus estimated grade; (**a**) for the polygonal estimator and (**b**) for kriging. Ideally points should lie along the diagonal (true grade = estimated grade)

1.5.4 Support effect

In geostatistics the term "support" refers to the size and volume of a sample or a block. Here the samples have a 1m x 1m support while blocks are 2m x 2m. In general, the support of samples is smaller than that of blocks. The true grades of the sixteen 2m x 2m blocks and of the sixty–four 1m x 1m blocks are shown in Figs. 1.4. and 1.7. Although the two means are the same, the variance of the samples is higher than that of the blocks.

735	325	45	140	125	175	167	485
540	420	260	128	20	30	105	70
450	200	337	190	95	260	245	279
180	250	380	405	250	80	515	605
124	120	430	175	230	120	460	260
40	135	240	35	130	135	160	170
75	95	20	35	32	95	20	450
200	35	100	53	2	45	58	90

Fig 1.7. The true grades of the sixty–four 1m x 1m blocks

Their histograms (Fig. 1.8.) show that the small blocks are more dispersed than the large ones. For the 300 cutoff more ore will be recovered if 1m x 1m blocks are mined rather than 2m x 2m ones. As the polygonal method equates the grades of the samples (i.e. a small support) with those of the blocks, it substitutes the histogram of the samples for the block histogram – even though they are quite different. This shows that a good estimator must take account of the difference between the supports of the samples and the blocks to be estimated; that is, of the support effect.

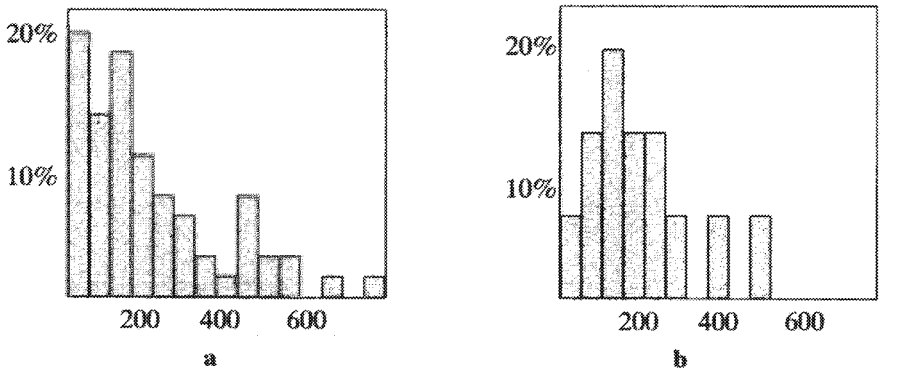


Fig 1.8. Histograms of the grades, (**a**) for small blocks and (**b**) for larger ones. Although the means remain the same (201) the variances are different and so are the shapes

So we have seen that the support effect and the information effect are two of the main causes of the incorrect prediction of reserves. We now know some of the properties that a good estimator should have. We can see that the way in which we combine data in the neighbourhood of the block to be estimated is important. The first part of this course will deal with the variogram; this is a statistical tool for assessing how similar values are, as a function of the distance between them. In the second part of the book the variogram is used to calculate the weights to be used when estimating blocks (for example); that is, when kriging blocks.

1.6 Does geostatistics work in the real world?

To answer this question we will have a look at some case-studies where the predictions made using geostatistics were compared with actual production figures. We will first look at two case-studies on coal, which is usually fairly easy to estimate, then at two others on gold, involving lognormal kriging and then review some more recent ones.

1.6.1 Early coal case studies

The studies on coal are Sabourin (1975) and Wood (1976). Both were favorably impressed by the results given by kriging. Sabourin estimated the sulphur content of blocks using channel samples. He then compared his estimates with the actual production figures. The average relative error between the estimates and the actual value was 9.8 %, which he judged to be "very satisfactory". It is important to note that this deposit had a marked trend in the sulphur values, which necessitated the use of universal kriging. As this is more complicated than ordinary kriging, the close agreement with the production figures is particularly impressive.

Wood did not run into the problem of a trend in the data in his study of South African coal. He was therefore able to use ordinary kriging. The aim of his study was to predict the seam width at a distance of 18m (one pillar) in advance of the workings in the No 2 seam at the Witbank Mine. Four estimation methods were considered:

1. the average of all measurements in the last 3 pillar advances,
2. the average of all measurements in the last pillar advance,
3. the closest single measurement, and
4. kriging.

The tests were carried out in two parts of the mine (one with 133 sets of data values, the other with 101 sets). The differences between the estimated values and the actual production figures were calculated for all four methods (Table 1.1.). This showed that the kriged estimates were consistently closer to the true values than the other methods considered.

Table 1.1. Mean square of the standardised estimation errors, taken from Wood (1976) Report No 2 South African Chamber of Mines

	Region A	Region B	Both regions
No of estimates	133	101	234
Mean of last 3 advances	7.50	2.69	5.42
Mean of last advance	7.38	1.84	4.99
Closest sample	5.79	3.17	4.66
Kriging	4.89	1.38	3.38

1.6.2 Gold case studies

We now go on to look at some comparative case-studies on gold deposits, which are clearly much more difficult to estimate because of the skewness of the distributions. Two interesting ones are Rendu (1979) and Krige and Magri (1982), who worked on South African deposits.

Rendu set out to test whether geostatistical predictions were verified in practice. He had about 5000 gold grades from one section of the Hartebeestfontein Mine. As the data were on a very close grid (25 ft), blocks of size 125 ft x 125 ft contained 25 samples and averaging these gives an accurate idea of the true grade of each block. Rendu then took the central one of the 25 grades as the "sample" and estimated the block grades by kriging using the "samples". By moving the center of the "sample" grid, he was able to carry out his procedure on 4808 blocks.

Since the data had a three parameter lognormal distribution, he used lognormal kriging as well as ordinary kriging and also two classical methods (polygons of influence and inverse distance weighting). To present his results he calculated the regression of the true grade against the estimated one (on a bi-logarithmic scale) for all the estimation methods considered. Ideally the regression line should be at 45 degrees. Lognormal kriging with a known mean came closer to this than any of the other methods. If the regression line is not at this angle, the estimates are said to be conditionally biased. When that occurs, the slope is usually less than 1.0 and so the grade of rich blocks is over-estimated, whereas that of poor ones is under-

estimated. This was certainly the case for the polygonal method. As we have seen, the economic consequences of this can be very serious when the estimates are being used to select blocks above a cutoff grade for mining, and for predicting the recoverable reserves.

The second comparative study by Krige and Magri was on the gold grades of a very variable reef in the Lorraine gold mine and on the lead grades in the Prieska copper-zinc mine. Their findings confirmed those by Rendu.

1.6.3 More recent case studies

Several more comparative studies were presented at the Ore Reserve Estimation Symposium held in Montreal in May 1986. Two particularly interesting ones were those by Raymond and Armstrong (1986) who worked on a porphyry copper deposit and by Blackwell and Johnston (1986) who studied a low grade copper molybdenum deposit. Raymond and Armstrong found a very close agreement between the grade of milled ore over a 17 month period. They used lognormal kriging (which is a special form of kriging designed for skew data with a lognormal distribution). In the other paper the authors described how they used both DDH data and blasthole data when kriging blocks. In their conclusion they cited three advantages of using geostatistical methods:

1. The mineral reserves results are easily duplicated by different mine personnel, as there is little need for subjective interpretation after the variogram models are selected.
2. Geostatistics effectively improves estimated grades, even when using grade data of differing size and reliability.
3. The improved mineral reserve permits better long and short range planning and allows the operator flexibility when dealing with downtime, breakdowns, wall slope instabilities, metal price changes.

Since these comparative studies confirm the superiority of kriging over other commonly used estimation methods for deposits ranging from coal to gold, it seems worthwhile looking more closely at the technique. We shall start by seeing how to use geostatistics to model these types of variables.

1.7 Exercises

Ex 1.1 The four tables below show the estimated grades obtained using three different methods and also the real grades found after mining.

Calculate the average for the 16 block grades for each method and for the true grades. Which of these estimators are unbiased?

Plot the scatter diagrams of the true grade (on the vertical axis) against the estimated grade. Look at their regression slopes (true against estimated) and see which is closer to 45° (i.e. a slope of 1.0).

Table 1.2. Grades estimated by three different methods (**a**, **b** and **c**) and the true grades (**d**)

30	30	20	45
20	50	50	35
40	25	35	25
15	40	20	20

a

10	30	10	55
5	20	50	35
40	15	40	30
5	35	10	10

b

15	35	20	35
15	20	45	25
40	10	30	30
15	35	20	10

c

20	25	15	40
15	25	45	30
35	15	35	25
10	35	20	10

d

2 Regionalized Variables

2.1 Summary

In this chapter the basic definitions in geostatistics including the concepts of *random function* and *regionalized variable* are presented. The underlying hypotheses (second order stationarity and the weaker intrinsic hypothesis) are introduced. The *variogram* and the *spatial covariance* are defined. The problem of how to decide whether to treat a variable as stationary, intrinsic or nonstationary is discussed. Some of the basic properties of the spatial covariance are introduced in this chapter as they are helpful in deciding on the degree of stationarity. The relationship between the variogram and the spatial covariance is derived but the rest of the variogram properties are left to the next chapter.

2.2 Modelling regionalized variables

Since the information available about the variable under study is fragmentary, we need a model to be able to draw any conclusions about points that have not been sampled. There are many ways of setting up models. Several will be discussed.

Genetic models. One of the most intuitively appealing ways of developing a mathematical model is by modelling the genesis of the phenomenon. As sedimentary processes are amongst the simplest to describe, attempts were made to model them in the early seventies by Jacod and Joathon (1970 a, b). Unfortunately, the geological factors controlling even simple sedimentary processes are extremely complicated and require many parameters to represent them. Not surprisingly it proved difficult to get meaningful estimates of these from limited sample data. These problems led researchers to give up this approach at the time. Recent work by Hu, Joseph & Dubrule (1994) in modelling oil reservoirs made up of prograding lobes has met with more success, but the idea of simulating the genesis of deposits

mathematically has been dropped. The geology of reservoirs and deposits is too complicated and not yet well enough known for this approach to work – at least at present.

Trend surfaces. By the late sixties, computers had become much more common which made it possible to carry out the calculations involved in statistical methods such as trend surfaces. So at the same time that Jacod and Joathon were working on reservoir genesis, two Americans applied trend surface analysis for predicting the properties of coal. 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 does not depend on the function. The difficulty with this approach can be seen from Table 2.1 which shows the equation fitted by Gomez and Hazen (1970) for the proportion of pyritic sulphur in a certain coal. The equation is very complicated and contains many terms like sines, cosines and exponentials. The problem is that most geological variables display a considerable amount of short scale variation in addition to the large scale trends that can reasonably be described by deterministic functions. Insisting on having uncorrelated errors means that the function has to twist and turn a lot, which explains the presence of all the exponential and trigonometric terms in Table 2.1. This suggests that it might be better to allow for correlations between values different distances apart. This is the basic idea behind geostatistics.

Geostatistics. The term *regionalized variable* was coined by Matheron (1963, 1965) to emphasize two apparently contradictory aspects of these types of variables: a random aspect, which accounts for local irregularities, and a structured aspect, which reflects large scale tendencies. 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. Unfortunately this is not realistic for geological phenomena. A better way of representing the reality is to introduce randomness in terms of fluctuations around a fixed surface which Matheron called the "drift" to avoid any confusion with the term "trend". Fluctuations are not "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". The geostatistician can go on to estimate or simulate the variables.

2.3 Random functions

The observed value at each data point x is considered as the outcome, $z(x)$, of a random variable, $Z(x)$. Its mean 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 variable $Z(x)$.

Table 2.1. Regression equation for pyritic sulphur in coal (From Gomez & Hazen, 1970) This method assumes uncorrelated errors, which forces the trend surface to twist and turn rapidly, hence the trigonometric and exponential terms

$$\text{Equation A}-1\ (\text{Pyritic sulfur, coarse coal}) = 14.0548 - 5.97910 \times AS$$
$$\div 1.35753 \times SU + 1.34232 \times SU \times (AS - SU) - 0.419448 \times AS$$
$$\times SU \times (AS - SU) + 4.95307 \times 10^{-3} [AS \times SU \times (AS - SU)]^2$$
$$- 2.69728 \times 10^{-4} [AS \times SU \times (AS - SU)]^3 + 5.88963 \times 10^{-11} \times e^{(AS \times SU)}$$
$$- 1.38995 \times 10^{-6} [e^{[AS \times (AS - SU)]}] + 0.0103637 [e^{[SU \times (AS - SU)]}]$$
$$- 16.70984 [AS \times SU/e^{(AS \times SU)}] + 5.67080 \times 10^{-3} [AS \times (AS - SU)$$
$$/e^{[AS \times (AS - SU)]}] - 1.96079 \times 10^{-5} [SU \times (AS - SU)\ /e^{[SU \times (AS - SU)]}]$$
$$\div 0.104688 \times \sin(AS \times SU)^3 \times \cos(AS \times SU)^2 - 9.53418 \times 10^{-3}$$
$$\times sin[AS \times (AS - SU)]^3 \times cos[AS \times (AS - SU)]^2 - 0.0848224$$
$$\times sin[SU \times (AS - SU)]^3 \times cos[SU \times (AS - SU)]^2 - 2.08676 \times sin(AS)$$
$$- 2.19124 \times sin(SU) + 1.49662 \times sin(AS - SU) - 0.107983 \times e^{AS}$$
$$- 0.0504733 \times e^{SU} + 0.312656 \times {}^{(AS - SU)} - 0.0491919 \times \sin(AS^3)$$
$$\times \cos(AS^2) + 0.0434771 \times sin(SU^3) \times \cos(SU^2) - 1.81229 \times 10^{-4}$$
$$\times sin(AS - SU)^3 \times cos(AS - SU)^2 - 2.01703 \times 10^{-4} \times e^{(AS + SU)} \times e^{(AS - SU)}$$

Variables used in equation.

AS	= Ash in coal, percent
SU	= Sulfur in coal, percent

In mathematical terms, the family of all these random variables is called a random function. (Synonyms: stochastic process, random field). A random function bears the same relation to one of its realizations as a random variable does to one of its outcomes, 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 finite dimensional distributions, i.e. by the joint distributions of any set of variables $Z(x_1)$, $Z(x_2)$, ... $Z(x_k)$, for all k, and for all points x_1, x_2, ... x_k. It would be impossible to do anything with this model unless we are prepared to make some assumptions about the characteristics of these distributions. The next section

presents the most commonly used hypotheses. Before going into these in detail, readers may be interested to know what types of variables can be modelled as random functions. Box No 1 lists some of the more common ones.

BOX No 1 : Variables that can be modelled by random functions.

- Metal grades, for precious metals, uranium, base metals, coal, diamonds, beach sands, industrial minerals,
- Quality parameters e.g. for iron ore, silica, alumina, loss on ignition and sometimes manganese; for gold, arsenic; for coal, calorific value, ash & sulphur content; for cement, iron content, magnesium oxide, moisture,
- Topographic variables such as seam thickness, overburden thickness, depth to a geological horizon, position of the sea floor,
- Rock type indicators e.g. for distinguishing between sandstone and shale in oil reservoirs, or between different facies in general,
- Porosity and permeability, for both oil reservoirs and aquifers, hydraulic head and transmissivity in hydrology,
- Geochemical trace element concentrations in soil samples and stream sediments,
- Pollutant concentrations in soil & water and in the atmosphere,
- For soil science, trace element concentrations (e.g. Cu & Co), nematode counts in soil,
- In fishery science, fish & egg counts, water temperature, salinity; density of shellfish per unit area,
- In hydrology, rainfall and runoff measurements,
- Tree density in tropical forests.

2.4 Stationary and intrinsic hypotheses

In statistics it is common to assume that the variable is stationary, i.e. its distribution is invariant under translation. In the same way, a stationary random function is homogeneous and self-repeating in space. For any increment h, the distribution of $Z(x_1)$, $Z(x_2)$, ... $Z(x_k)$ is the same as that of $Z(x_1 + h)$, $Z(x_2 + h)$, ... $Z(x_k + h)$ This makes statistical inference possible on a single realization. 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 require only the first two moments (the mean and the covariance) to be constant. This is called "weak" or second order stationarity. In other words, the expected value (or mean) of $Z(x)$ must be constant for all points x.

That is,

$$E(Z(x)) = m(x) = m \qquad [2.1]$$

Secondly the covariance function between any two points x and x+h depends on the vector h but not on the point x. That is,

$$E[Z(x)\, Z(x+h)] - m^2 = C(h) \qquad [2.2]$$

There is no need to make an assumption about the variance because it turns out to be equal to the covariance for a zero distance, C(0).

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. Another branch of geostatistics has been developed to handle "nonstationary" regionalized variables. It is outside the scope of this text. Interested readers could consult Matheron (1973) or Delfiner (1976).

For the moment we shall only consider cases where the mean is constant. However, even when this is true, the covariance need not exist. A particularly startling practical example of this was described by Krige (1978) for the gold grades in South Africa. On both theoretical and practical grounds it is convenient to be able to weaken this hypothesis. This is why Matheron (1963, 1965) developed the "intrinsic hypothesis". It assumes 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.

$$E[Z(x+h) - Z(x)] = 0 \qquad [2.3]$$

$$\mathrm{Var}[Z(x+h) - Z(x)] = 2\gamma(h) \qquad [2.4]$$

The function $\gamma(h)$ is called the semi-variogram (variogram for short). It is the basic tool for the structural interpretation of phenomena as well as for estimation.

Regionalized variables that are stationary always satisfy the intrinsic hypothesis but the converse is not necessarily true. Later in this chapter we will see that if a regionalized variable is stationary, there is an equivalence between its variogram $\gamma(h)$ and its covariance C(h).

Most estimators used in the earth sciences are linear combinations (i.e. weighted moving averages) of the data. This is true for the inverse distance method, and for kriging (as will be seen later) and even for the polygonal method where all the weights except one are zero. So it is important to be able to calculate the variance of linear combinations in terms of the variogram and/or the covariance. In contrast to the stationary case, when working with intrinsic variables the operations are defined only for increments. We will show later that the variance of linear combinations can be calculated only if the sum of the weights is 0. By using intrinsic regionalized variables instead of just stationary ones, we have to work with increments but the range of variogram models available is considerably enlarged.

2.5 How to decide whether a variable is stationary

Before going into more detail about the variogram, it is important to see how to decide whether a particular variable can be considered stationary or not. In practical situations the variogram is only used up to a certain distance. This limit could be the extent of a homogeneous zone within a deposit or the diameter of the neighbourhood used in kriging (i.e. estimation). Consequently, the phenomenon only has to be stationary up to this distance. The problem is to decide whether we can find a series of moving neighbourhoods within which the expected value and the variogram can be considered to be constant and where there are enough data to give meaningful estimates. This assumption of quasi-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.

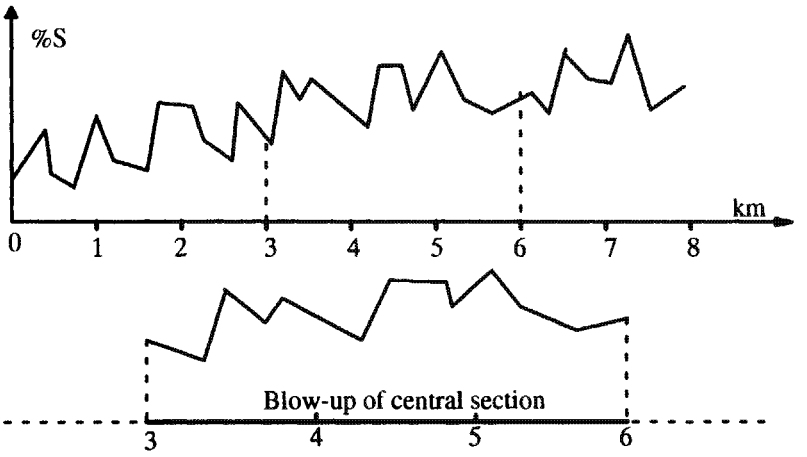


Fig 2.1. Diagrammatic representation of sulphur grades and a blow–up of the central section. Over the whole 8km length, the sulphur content is clearly not stationary because of the increase in the average. But over shorter sections it can be considered as being locally stationary because the fluctuations dominate the trend

Consider the sulphur content of coal along a transect (Fig 2.1.). Over the total distance shown (8 km) there is a clear increase from left to right. However looking at a blow-up of the central section, 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 whereas it is clearly nonstationary over longer distances. In practice the blocks of coal to be estimated are about 100m x 100m for underground mines and 200m x 60m in strip mining operations. Samples are generally on a 500m x 500m grid for wide spaced holes, down to 100m x 100m later on. With samples at these distances, there is no point in searching for data several kilometers away. There is plenty much closer.

2.6 Spatial covariance function

Before going into the properties of the variogram in detail, we present some of the basic properties of the spatial covariance and derive the relationship between it and the variogram for stationary random functions. (Note: no covariance exists for a random function that is intrinsic but not stationary). Three important properties are listed below. Proofs are given in Box No 2.

$$C(0) = \sigma^2 \qquad [2.5]$$

$$C(h) = C(-h) \qquad [2.6]$$

$$|C(h)| \leq C(0) \qquad [2.7]$$

Note that absolute values appear in some equations because the covariance can take negative values. Our next task is to establish the basic relation between the variogram and the corresponding covariance:

$$\boxed{\gamma(h) = C(0) - C(h)} \qquad [2.8]$$

Proof. The proof starts out from the definition of the variogram:

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

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

$$= 2C(0) - 2C(h) \qquad [2.9]$$

Hence the result. This shows that the corresponding covariance is obtained by "turning the variogram upside down". Figure 2.2. illustrates this idea.

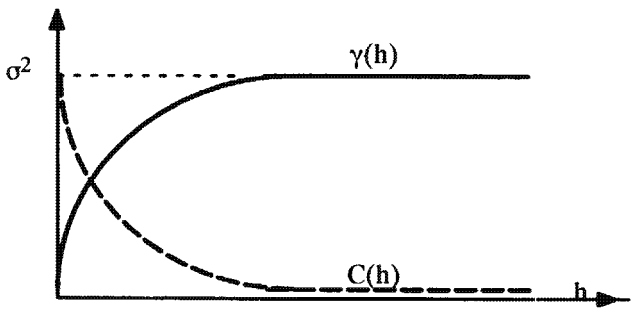


Fig 2.2. Whereas the variogram starts from zero and rises up to a limit, the spatial covariance starts out from the variance and decreases

BOX No 2 : Proofs of the properties of the covariance.

First property. For stationary variables, the mean m exists. The covariance is, by definition:

$$C(h) = E\,[(Z(x+h) - m)(Z(x) - m)\,] \qquad [2.10]$$

Substituting h = 0 gives

$$C(0) = E\,[(Z(x+0) - m)\,(Z(x) - m)] = \sigma^2 \qquad [2.11]$$

Second property. By definition

$$C(-h) = E\,[(Z(x-h) - m)(Z(x) - m)\,] \qquad [2.12]$$

Putting t = x – h gives

$$C(-h) = E\,[(Z(t) - m)(Z(t+h) - m)\,] = C(h) \qquad [2.13]$$

Third Property To prove the third property, we must prove that

$$C(h) \leq C(0) \text{ and } C(h) \geq -C(0) \qquad [2.14]$$

We start out from the relation

$$0 \leq E\,[\{Z(x+h) - Z(x)\}^2\,]$$

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

$$= 2C(0) - 2C(h) \qquad [2.15]$$

$$\text{Hence } C(h) \leq C(0) \qquad [2.16]$$

Similarly the other inequality can be obtained by starting out from the relation

$$0 \leq E\,[\{Z(x+h) + Z(x)\}^2\,]$$

$$\text{Hence } C(h) \geq -C(0) \qquad [2.17]$$

So we obtain

$$|C(h)| \leq C(0) \qquad [2.18]$$

This is called Schwartz's inequality.

Clearly this is possible only when the variogram is bounded above. It can be shown mathematically that variograms with an upper bound come from stationary regionalized variables. It would be more accurate to say that only stationary regionalized variables have bounded variograms. Consequently, unbounded variograms come from intrinsic regionalized variables or nonstationary ones.

Lastly it can be proved that for stationary and intrinsic variables

$$\lim_{h \to \infty} \frac{\gamma(h)}{h^2} = 0 \qquad [2.19]$$

This result is not easy to prove. Interested readers can consult Matheron (1972). This result means that if the variogram rises more rapidly than a quadratic for large h, the variable is nonstationary. Otherwise it can be considered to be stationary or intrinsic. This is helpful in deciding whether a variable is stationary or intrinsic or whether it has to be treated as nonstationary.

2.7 Exercises

Expressing variances in terms of covariances. One of the key steps in geostatistics is expressing the variance of a linear combination (a weighted average) in terms of the weights and the covariance function, and later the variogram. The first exercise develops the basic formula; in the second one, it is applied to a particular case.

Ex 2.1 Let Z(x) be a stationary random function. Its spatial covariance is denoted by C(h). Let Z* be the weighted average of the values at two points:

$$Z^* = \lambda_1 Z(x_1) + \lambda_2 Z(x_2) \qquad [2.20]$$

where λ_1 and λ_2 are two weighting factors and x_1 and x_2 are the two points. What is the expected value of Z*? Express its variance in terms of the weighting factors and its covariance C(h).

Now generalize this to an arbitrary linear combination Z**

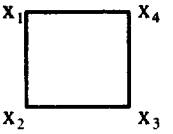
$$Z^{**} = \sum_i \lambda_i Z(x_i) \qquad [2.21]$$

What is the expected value of Z**? Show that its variance can be written in either of the following ways:

$$\mathrm{Var}(Z^{**}) = \sum_i \lambda_i^2 \mathrm{Var}(Z(x_i)) + 2 \sum \sum_{j>i} \lambda_i \lambda_j C(x_i - x_j)$$

$$= \sum_i \sum_j \lambda_i \lambda_j C(x_i - x_j) \qquad [2.22]$$

Ex 2.2 Let Z(x) be a stationary random function, and let Z* be a weighted average of the values at the four corners of a 100m x 100m square:



$$Z^* = 0.5\ Z(x_1) + 0.2\ Z(x_2) + 0.2\ Z(x_3) + 0.1\ Z(x_4) \qquad [2.23]$$

Evaluate the variance of Z* when the spatial covariance of Z(x) is an exponential:

$$C(h) = 2.5 \exp(-|h|/200) \qquad [2.24]$$

(In the next chapter we will see that this is an acceptable model for a covariance).

3 The Variogram

3.1 Summary

This chapter and the following one are devoted to the variogram. In this one, after defining the variogram, its theoretical properties are discussed (e.g. zone of influence, behaviour near the origin, anisotropies, presence of a drift, etc.). The common variogram models are presented. Images of variables having some of these variograms have been simulated to highlight the differences between the models. The formula for calculating the variance of a linear combination of regionalized variables in terms of the variogram is proved. The reason why only positive definite functions can be used as models for the variogam is stressed.

3.2 Definition of the variogram

In Chapter 2, the variogram of an intrinsic random function was defined as:

$$\gamma(h) = 0.5 \text{ Var } [Z(x+h) - Z(x)] \qquad [3.1]$$

For stationary and intrinsic variables, the mean of $Z(x+h) - Z(x)$ is zero, and so $\gamma(h)$ is just the mean square difference. Consequently,

$$\gamma(h) = 0.5 \text{ E } [Z(x+h) - Z(x)]^2 \qquad [3.2]$$

Here x and x+h refer to points in an 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. Consequently, the variogram is a function of the two components h_1 and h_2, or alternatively, 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 such as anisotropy. Figure 3.1. shows a typical variogram.

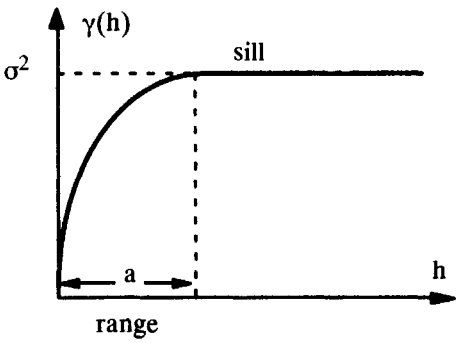


Fig 3.1.. A typical variogram which reaches a limit called its sill at a distance called the range

It presents the following features:
- it always starts at 0 (for h = 0, Z(x+h) = Z(x)). It could be discontinuous just after the origin.
- 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 of the variogram will now be treated in detail.

3.3 Range and zone of influence

The rate of increase of the variogram with distance 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 critical distance, called the range, gives a more precise definition to the notion of the "zone of influence". For stationary variables, γ(h) equals the variance for distances past the range. That is,

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

$$= 0.5 \ [\text{Var } (Z(x+h)) + \text{Var } (Z(x))] = \sigma^2 \qquad [3.3]$$

Not all variograms reach a sill. Some, like the one shown in Fig. 3.2.**b**, keep on increasing with distance. This is one fundamental difference between the variogram and the covariance. The latter only exists for stationary variables and is bounded.

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 more than one range. This occurs when there are several nested structures acting at different distance scales. Examples of anisotropy and nested structures will be given later.

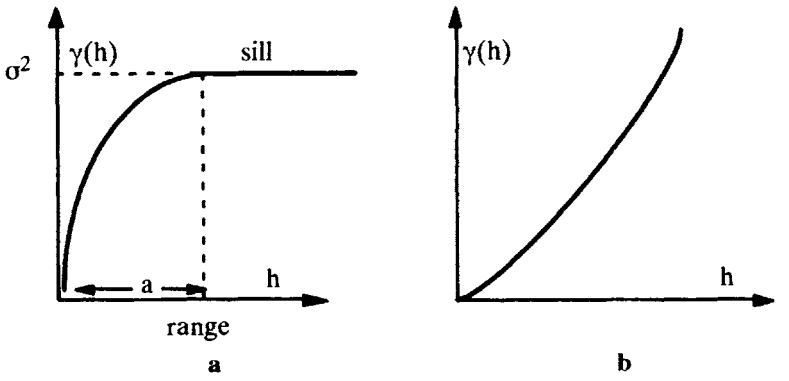


Fig 3.2. Bounded and unbounded variograms

3.4 Behaviour near the origin

We have just examined the behaviour of the variogram for large distances. But it is even more important to study its behaviour for small values of h because this is related to the continuity and the spatial regularity of the variable. Four types of behaviour near the origin are shown in Fig. 3.3.

1. **Quadratic.** This indicates that the regionalized variable is highly continuous. In fact it is differentiable. A quadratic shape can also be associated with the presence of a drift.

2. **Linear.** The regionalized variable is then continuous but not differentiable, and is thus less regular than above.

3. **Discontinuous at the origin** i.e. $\gamma(h)$ does not tend to zero as h tends to 0. This means that the variable is highly irregular at short distances.

4 **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. It is, incidentally, the model adopted in trend surface analysis.

The variograms of most geological variables, including metal grades, have this discontinuity at the origin. It 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. The grade passes abruptly from zero outside the nugget to a high value inside it. Gold is not the only substance that contains nuggets. Particles of pyrite randomly distributed in coal lead to erratic changes in its sulphur content. The term "nugget effect" is also applied to short range variability even when it is known to be due to some other factor such as micro-structure, measurement error or errors in location.

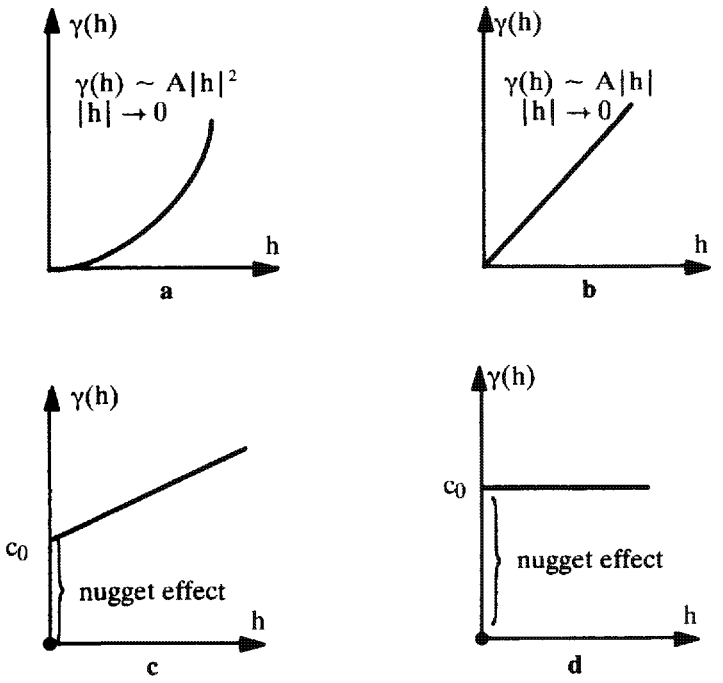


Fig 3.3. The behaviour of the variogram near the origin. The quadraticic shape (**a**) indicates a high degree of continuity in the variable; linear behaviour (**b**) shows moderate continuity. Discontinuities at the origin (**c**, **d**) indicate erratic short scale behaviour called a nugget effect

3.5 Anisotropies

When the variogram is calculated in different directions, it sometimes behaves differently in some of them (i.e. anisotropy). If this does not occur, the variogram depends only on the magnitude of the distance between the two points and is said to be isotropic. Two different types of anisotropy can be distinguished: geometric anisotropy and zonal anisotropy.

3.5.1 Geometric anisotropy

Figure 3.4. shows examples of geometric anisotropy. On the left, 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.

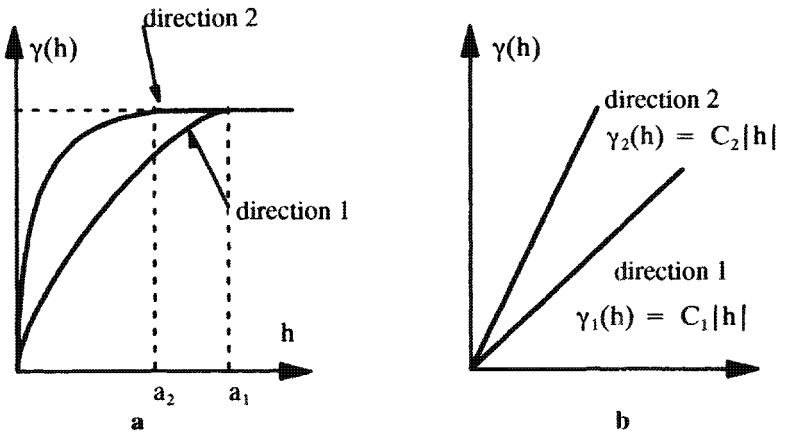


Fig 3.4. Elliptic or geometric anisotropy

We can draw a diagram showing the range or the slope as a function of direction. (Fig. 3.5.). 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.

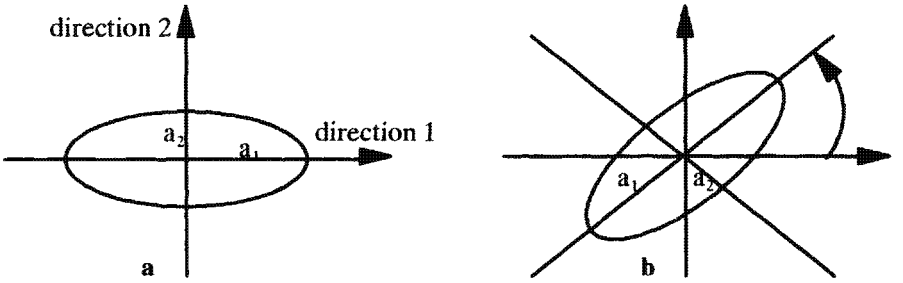


Fig 3.5. Ellipses showing the major and minor axes in the case of geometric anisotropy

This transformation is particularly simple when the major axis of the ellipse coincides with the coordinate axes as is shown in Fig. 3.5. (**a**). 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:

$$\gamma(h) = \gamma_1\left(\sqrt{h_1^2 + k^2 h_2^2}\right) \qquad [3.4]$$

where h_1 and h_2 are the two components of h and k is the anisotropy ratio, namely:

$$k = \frac{\text{range 1}}{\text{range 2}} \text{ or } k = \frac{\text{slope 1}}{\text{slope 2}} \qquad [3.5]$$

When calculating the variogram, it is important to use at least four directions. If the variogram was calculated in only two perpendicular directions it would be possible to miss the anisotropy completely. This would be the case if the major axis was at 45° to the directions in which the variogram was calculated, as in Fig. 3.5.**b**.

3.5.2 Zonal (or stratified) anisotropy

There are more complex types of anisotropy. For example, the vertical direction often plays a special role because there is more variation between strata than within them. In such cases the sills of the variograms are not the same in all directions. It is standard practice to split the variogram into two components, an isotropic component plus another one which acts only in the vertical direction:

Isotropic component — Vertical component:

$$\gamma_0\left(\sqrt{h_1^2 + h_2^2 + h_3^2}\right) \qquad \gamma_1(h_3) \qquad [3.6]$$

The overall variogram $\gamma(h)$ is

$$\gamma(h) = \gamma_0(h) + \gamma_1(h) \qquad [3.7]$$

3.6 Presence of a drift

As was mentioned at the end of Chapter 2, theory shows that for intrinsic and stationary variables, the variogram increases more slowly than a quadratic for large distances. To be more specific,

$$\frac{\gamma(h)}{h^2} \to 0 \text{ as } h \to \infty \qquad [3.8]$$

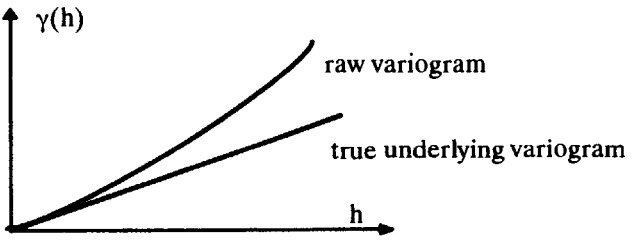


Fig 3.6. Variogram shape in presence of a drift

However, in practice it often occurs that variograms increase more rapidly than h^2. This indicates the presence of a drift (Fig. 3.6.). The experimental variogram

provides an estimate of $0.5E(Z(x+h) - Z(x))^2$, which is called the raw variogram, rather than the true (or underlying) variogram. These two coincide only if the increments have a zero mean. Otherwise

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

raw variogram = underlying variogram + (bias term)^2 [3.9]

If there is a drift, the empirical variogram overestimates the underlying variogram.

3.7 Nested structures

Nested structures can sometimes be seen when looking at experimental variograms. In Fig. 3.7. the longer range is apparent because the variogram flattens out at that distance. The shorter range can be distinguished by the characteristic change in the curvature. This change is obvious when the two ranges are quite different. If they are not, the change is more gradual and need not be obvious.

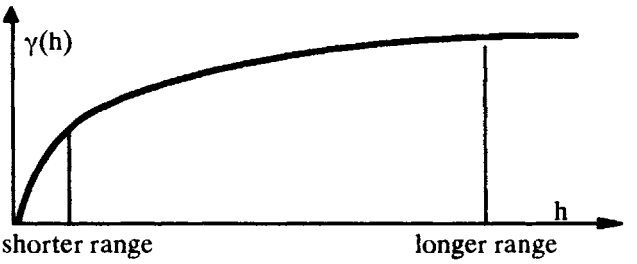


Fig 3.7. Nested structure composed of a short range structure and a longer one

Nested structures indicate the presence of processes operating at different scales. For example, there may be measurement error at the level of a sample, i.e. for h = 0. At the petrographic scale (i.e. h < 1 cm), there can be variability due to a transition from one mineralogical constituent to another. At the level of strata or mineralized lenses (i.e. for h < 100m) a third type of variability comes into play as the points pass from ore to waste or from ore facies to another.

3.8 Proportional effect

A variogram is said to have a proportional effect when its value (particularly its sill) is proportional to the square of the local mean grade. This often occurs with lognormally distributed data. The variograms for different zones have the same shape but the sill in rich zones is much higher than in poor ones. As the sill often

turns out to be proportional to the square of the local mean, 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.

3.9 Hole effects and periodicity

In between the origin and infinity, the behaviour of the variogram reflects different features such as the presence of nested structures or occasionally a hole effect or even periodicity. In some cases, the experimental variogram rises above its sill value then drops down. As this "hump" in the variogram corresponds to a hole in the covariance, the effect is called a hole effect. Sometimes this shape can be explained geologically. One such example is the variogram obtained by Serra (1968) from thin sections of iron ore from the Lorraine region in France. The calcite crystals tended to be separated by intervals roughly proportional to their size, because of the concentration process around randomly located seed crystals. This regular, almost periodic structure leads to the hole effect. But more generally bumps of this type are due to natural fluctuations in the variogram or to statistical fluctuations because too few pairs of points were used in calculating the experimental variogram.

Variograms, like covariances, can exhibit periodic behaviour. Whereas it is natural to find periodic phenomena when dealing with time series, this is much rarer and more difficult to explain with geological variables. Folded strata could exhibit periodicity. But it is more common to find that periodicity is an artefact due to human activity rather than Mother Nature. For example, grades obtained during the night shift may be statistically different to the day shift. It is important to check that the effect is real and not merely an artefact. One case where periodicity can occur, is when the ore lies in ridges and valleys, as sometimes happens in sedimentary gold deposits. The variograms calculated perpendicular to the ridges and valleys can show the periodicity but those parallel to the ridges do not. An example of this is presented in Chapter 5.

3.10 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.

3.10.1 Variance of admissible linear combinations

Since the common estimators are linear combinations of the data (i.e. weighted averages), we need to be able to calculate their variance. First we consider a stationary variable Z(x) with covariance C(h). Let the linear combination be Z*

$$Z^* = \sum_i \lambda_i Z(x_i) \qquad [3.10]$$

where λ_i are the weights and x_i are the sample locations. By definition its variance is

$$\text{Var}(Z^*) = E(Z^* - E(Z^*))^2 \qquad [3.11]$$

It is not difficult to show that if m is the mean of Z(x), $E(Z^*) = m \sum_i \lambda_i$. Hence

$$\text{Var}(Z^*) = E\left(\sum_i \lambda_i (Z(x_i) - m)\right)^2$$

$$= \lambda_1^2 C(x_1 - x_1) + \lambda_2^2 C(x_2 - x_2) + \ldots + \lambda_n^2 C(x_n - x_n) + 2\lambda_1\lambda_2 C(x_1 - x_2)$$
$$+ \ldots + 2\lambda_{n-1}\lambda_n C(x_{n-1} - x_n)$$

Consequently

$$\text{Var}(Z^*) = \sum_i \sum_j \lambda_i \lambda_j C(x_i - x_j) \qquad [3.12]$$

This must be non-negative whatever the points and whatever the weights. A function C(h) satisfying this condition is said to be positive definite.

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. Combinations are said to be "admissible" if the sum of the weights is zero.

$$\sum \lambda_i = 0 \qquad [3.13]$$

Clearly any linear combination of increments satisfies this condition since any single increment involves the weights +1 and −1. Conversely any combination satisfying this condition can be written as a linear combination of increments. Box 3 gives this proof and the formula for its variance in terms of the weights and the variogram model. As the covariance need not exist for intrinsic random functions, the formula must be in terms of the variogram.

$$\boxed{\text{Var}\left(\sum \lambda_i Z(x_i)\right) = -\sum\sum \lambda_i \lambda_j \gamma(x_i - x_j)} \qquad [3.14]$$

As this 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 $\sum \lambda_i = 0$, we require that

BOX No 3 : Calculating the variance of admissible linear combinations.

Firstly we want to show that any linear combination whose weights sum to 0, can be expressed as a combination of increments. By choosing an arbitrary point as origin, we have:

$$\sum \lambda_i \, Z(x_i) = \sum \lambda_i \left[Z(x_i) - Z(0)\right] \qquad [3.15]$$

So its variance exists and is given by:

$$\mathrm{Var}\left[\sum \lambda_i \, Z(x_i)\right] = \sum\sum \lambda_i \, \lambda_j \, \mathrm{Cov}\left[Z(x_i)-Z(0), Z(x_j)-Z(0)\right] \qquad [3.16]$$

To calculate the covariance of increments we use the identity:

$$\mathrm{Var}\left[Z(x_i) - Z(x_j)\right] = \mathrm{Var}\left[Z(x_i) - Z(0) + Z(0) - Z(x_j)\right]$$

$$= \mathrm{Var}\left[Z(x_i) - Z(0)\right] + \mathrm{Var}\left[Z(x_j) - Z(0)\right]$$

$$- 2\mathrm{Cov}\left[Z(x_i) - Z(0), Z(x_j) - Z(0)\right] \qquad [3.17]$$

Hence $2\mathrm{Cov}[Z(x_i) - Z(0), Z(x_j) - Z(0)] =$

$$= \mathrm{Var}\left[Z(x_i) - Z(0)\right] + \mathrm{Var}\left[Z(x_j) - Z(0)\right] - \mathrm{Var}\left[Z(x_i) - Z(x_j)\right]$$

$$= 2\gamma(x_i) + 2\gamma(x_j) - 2\gamma(x_i - x_j) \qquad [3.18]$$

Substituting this into formula [3.16] gives:

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

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

The first two terms disappear since $\sum \lambda_i = \sum \lambda_j = 0$, leaving:

$$\boxed{\mathrm{Var}\left(\sum \lambda_i \, Z(x_i)\right) = -\sum\sum \lambda_i \, \lambda_j \, \gamma(x_i - x_j)} \qquad [3.19]$$

Hence the very important result that the variance of any linear combination whose weights sum to 0, exists and can be calculated by replacing the covariances in Equation [3.12] by $-\gamma$.

$$-\sum\sum \lambda_i \lambda_j \gamma (x_i - x_j) \geq 0 \qquad [3.20]$$

Then $-\gamma$ is said to be conditionally positive definite. This condition is weaker than the preceding one for covariances which had to hold for all possible weights, because this one only has to hold for sets of weights whose sum is 0. Consequently the class of admissible variogram models is richer than for covariances. It contains the bounded variograms associated with covariances and also unbounded ones which have no covariance counterpart. So there is a trade–off between the two hypotheses. The intrinsic hypothesis allows us to use a wider range of variograms but the weights must sum to 0. The range of admissible variogram models is more restricted for the stationary hypothesis but any weighting factors may be used.

3.11 Admissible models

We have seen that in order to ensure that the variance of any linear combination never goes negative, only certain functions can be used as models for variograms and covariances. Covariances must be positive definite functions; variograms have to be conditionally negative definite.

As it is not easy to recognize functions that have this property or to test for it, it is best to choose variogram models from the range of suitable functions rather than try to create them oneself. A list of the common models is given in the next section. These can be added to obtain other admissible models because this is equivalent to adding independent random functions, but subtraction is not allowed. Nor can they be combined piecewise. By this we mean that you cannot choose one model up to a certain distance then a different one from there onwards as shown in Fig. 3.8.

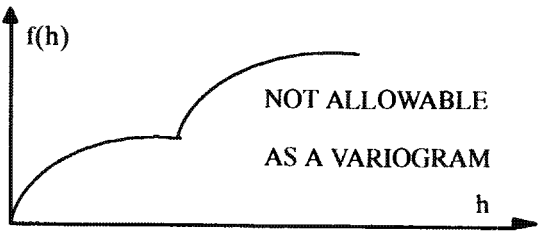


Fig 3.8. Example of a function that is NOT allowable as a variogram model

In order to work out whether a certain function is or is not positive definite, one has to calculate its Fourier transform. This is not always simple in 1-D but it becomes even more difficult in higher dimensional spaces. For more information on how to test for positive definiteness see Armstrong and Diamond (1984).

To make matters more complicated it is possible for functions to be positive definite in one space but not in higher dimensional spaces. For example the piece-wise linear function shown in Fig. 3.9. is an admissible variogram in 1–D but not in two or three dimensions. Exercise 3.11 shows how to construct it in 1–D while exercise 3.10 gives a counter example in 2D with a negative variance proving that it is not acceptable in 2–D or higher dimensional spaces.

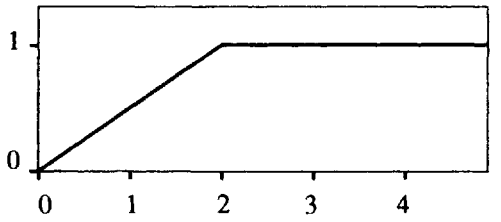


Fig 3.9. Piecewise linear model that is admissible in 1D but not in 2D or higher dimensions

Having been warned of the dangers of trying to invent their own variogram models, readers may be curious to know how the current variogram models came into existence. Basically they were developed by mathematically constructing a random function and calculating its variogram theoretically. The resulting model must, by construction, be positive definite or at least conditionally negative definite in the space in which it was built. Several exercises at the end of the chapter illustrate this procedure. For example, exercise 3.13 describes a method for building spherical balls enclosing a random number of Poisson points, leading to the so–called spherical variogram.

3.12 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.

3.12.1 Nugget effect

$$\gamma(h) = 0 \qquad h = 0 \tag{3.21}$$

$$= C \qquad |h| > 0$$

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

3.12.2 Spherical model

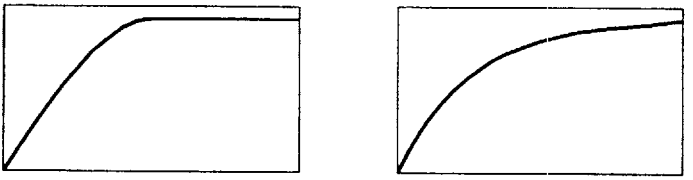
$$\gamma(h) = \begin{cases} C\left(\frac{3}{2}\frac{|h|}{a} - \frac{1}{2}\left(\frac{|h|^3}{a^3}\right)\right) & |h| < a \\ C & |h| \geq a \end{cases} \qquad [3.22]$$

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 $2a/\sqrt{3}$. This can be useful when fitting the parameters of the model.

3.12.3 Exponential model

$$\gamma(h) = C\,(1 - \exp(-|h|/a)) \qquad [3.23]$$

The practical range of this model is 3a, because that is the distance when it reaches 95% of its limit value. The tangent at the origin intersects the sill at a point with an abscissa a. As both the spherical and the exponential models are linear for small distances, it is helpful to compare them. Figure 3.10. shows two models with the same sills, with the practical range of the exponential equal to the true range of the spherical. The differences are quite obvious. The exponential rises more rapidly initially but only tends towards its sill rather than actually reaching it.



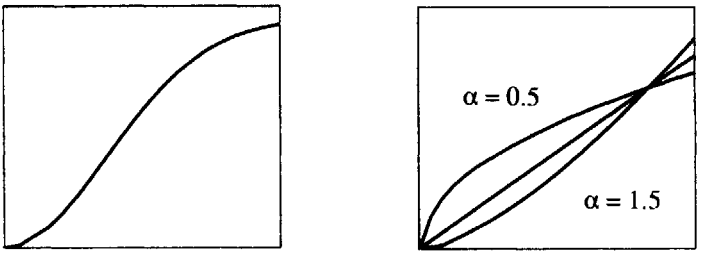
a Spherical Model **b** Exponential Model

Fig. 3.10. (**a**) The spherical variogram model with a sill of 1.0 and a range of 1.0 and (**b**) an exponential model with a sill of 1.0 and a scale parameter of 0.33 (i.e. its practical range is 1.0)

3.12.4 Power functions

$$\gamma(h) = C|h|^{\alpha} \qquad \text{with} \quad 0 < \alpha \leq 2 \qquad [3.24]$$

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



a Gaussian Model **b** Power Models

Fig. 3.11. (**a**) The gaussian variogram model and (**b**) three power function models with exponents $\alpha = 0.5$, 1.0 and 1.5

3.12.5 Gaussian model

$$\gamma(h) = C\left(1 - \exp\left(-\frac{|h|^2}{a^2}\right)\right) \qquad [3.25]$$

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.

3.12.6 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

$$\gamma(h) = C\left(7r^2 - 8.75r^3 + 3.5r^5 - 0.75r^7\right) \quad \text{if } r < 1 \qquad [3.26]$$

$$= C \quad \text{otherwise}$$

where $r = h/a$.

3.12.7 2D hole effect model

$$\gamma(h) = C\left(1 - \exp(-|r|)\, J_0(2\pi r_2)\right) \qquad [3.27]$$

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

3.12.8 Cardinal sine model

This model is one of the rare ones with a hole effect that is authorized in 3D. It corresponds to very continuous structures. Its equation is

$$\gamma(h) = C\left(1 - \frac{\sin r}{|r|}\right) \qquad [3.28]$$

where r = h/a. When calculating this on a pocket calculator, remember that the parameter "r" should be in radians not degrees. Working in degrees, [sin(r)]/r oscillates around the value of $\pi/180 = 0.0174532$.

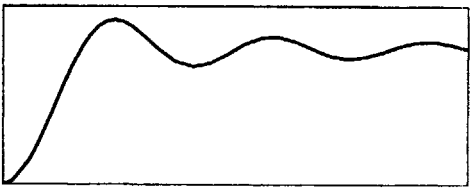


Fig. 3.12. The cardinal sine model

3.12.9 Prismato–magnetic model

$$\gamma(h) = C\left(1 - \frac{1}{(1 + r^2)^{1.5}}\right) \qquad [3.29]$$

where r = h/a.

3.12.10 Prismato–gravimetric model

$$\gamma(h) = C\left(1 - \frac{1}{(1 + r^2)^{0.5}}\right) \qquad [3.30]$$

where r = h/a. The last two models were developed to model different types of gravimetric and magnetic anomalies.

3.13 Simulated images obtained using different variograms

The variogram models given in the previous section range from very common ones like the exponential and the spherical to quite unusual ones like the hole effect

models. When most people look at the equations, they are unable to guess how regionalized variables having those variograms might behave. Are they very similar or do they have obviously different features? To highlight the differences between the variograms, we have generated one realization (one possible image) for each of four types of variograms: the exponential, the spherical, the gaussian and the cardinal sine. (Figs. 3.13. to 3.16.) A zero nugget effect was used throughout. These variograms all have sills so they correspond to stationary variables.

The images are 200 pixels x 200 pixels. The EW range (or the practical range) was set to 20 pixels whereas the NS one is half that. The darker and lighter patches in the figures are elongated with these dimensions on average. For those interested in the technical details the simulations were obtained using 400 turning bands and the resulting distributions are N(0,1). Simulation methods are beyond the scope of linear geostatistics. Details can be found in Journel and Huijbregts (1978) or Lantuéjoul (1994).

Comparing the four figures, it is clear that the exponential and the spherical variograms lead to "fuzzier" images than the gaussian and the cardinal sine. This obvious difference is due to the fact that the spherical and the exponential models are linear near the origin whereas the other two are parabolic. Variograms that are quadratic near the origin come from highly continuous variables. For a spherical and an exponential with the same sill and the same range, the exponential rises faster initially, which explains why the simulation obtained using the exponential model is slightly less structured than the corresponding spherical.

If a pure nugget effect had been used to generate a simulation, the resulting image would have been totally unstructured, with high and low valued pixels being spread around at random. This leads to very "spotty" pictures with a uniform grey colour and no distinct high or low patches.

It is important to keep the relation between the variogram model and its realizations in mind when fitting models to experimental variograms and later when kriging. Unless very closely spaced data are available, the geostatistician has to choose the shape of the variogram near the origin rather than fitting it to experimental values. So it is important to understand the implications of this choice in terms of the continuity of the variable or, on the contrary, its behaviour.

3.14 Exercises

Variogram Properties. Before fitting models to experimental variograms it is important to become more familiar with their properties.

Ex 3.1 Spherical model. Write down the equation for the spherical model with a range of 300m and a sill of 2. Plot its shape for distances up from h = –500m to h = +500m, remembering that $\gamma(-h) = \gamma(h)$. Note the mirror image around the y axis. The curve is continuous at the origin but what about its derivative? When we get to kriging, we will see that the kriged estimates "inherit" the discontinuities in the variogram function and its derivatives.

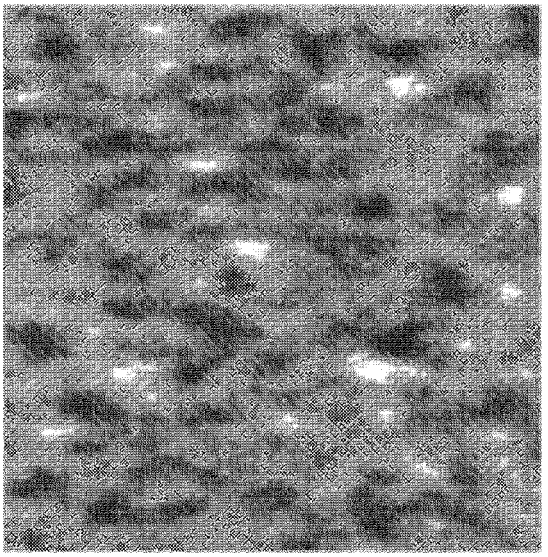


Fig 3.13. Simulation of a variable having an exponential variogram

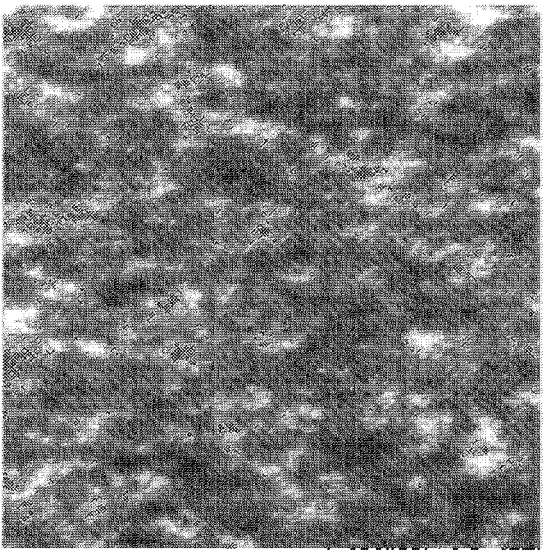


Fig 3.14. Simulation of a variable having a spherical variogram

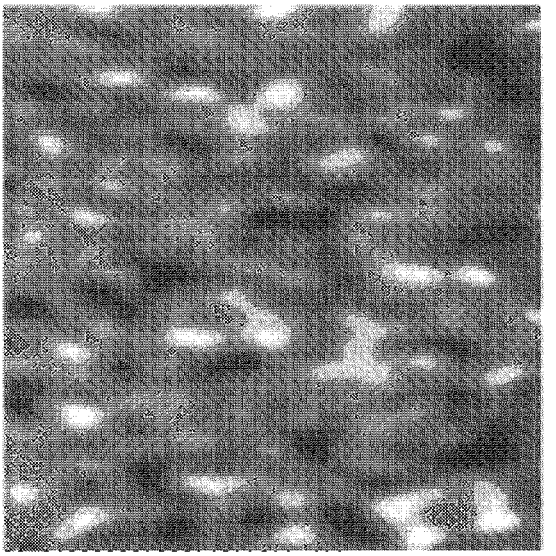


Fig 3.15. Simulation of a variable having a gaussian variogram

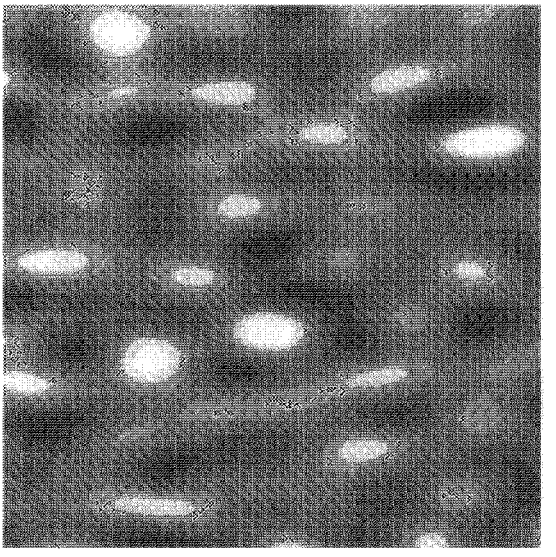


Fig 3.16. Simulation of a variable having a cardinal sine variogram

Ex 3.2 Exponential model. Write down the equation for the exponential model with a scale parameter of 100m and a sill of 2. What is its practical range? Plot its shape for distances up to 500m. Compare this model with the spherical model in the first exercise.

Ex 3.3 Tangents at the origin. Find the slopes of the spherical and the exponential models given in 3.1 and 3.2 by differentiating their equations with respect to h. Find the distance where the tangents at the origin cut the sill.

Ex 3.4 Gaussian model. Write down the equation for the gaussian model with a scale parameter of 100m and a sill of 2. What is its practical range? Plot its shape for distances from h= –500m to h=+500m. As before, the left hand side is a mirror image of the right hand side. What is the slope of its tangent at the origin? The curve is continuous there but what about its derivative? What about higher order derivatives?

Ex 3.5 Factorizable covariances. Write down the equation for the gaussian covariance model with a scale parameter of 1 and a unit sill, remembering that

$$\gamma(h) = C(0) - C(h)$$

Use the fact that in 2D h^2 equals $x^2 + y^2$, to split C(h) into the product of two factors: $\exp(-x^2)$ and $\exp(-y^2)$. Could this be extended to 3D? When we get to the chapter on kriging, we will see that factorizable covariances like this lead to a strange "perpendicular " screen effect.

Ex 3.6 Variograms that are linear near origin. Both the spherical and the exponential models are linear at the origin. Write down their slopes at the origin in terms of the sill, C, and the parameter, a.

Suppose that an experimental variogram is linear with a slope of 5.0 for distances up to 10m. Find suitable values of a and C for a spherical model and for an exponential one that have this slope. Show that choosing C=50 and a=15 gives a slope of 5 at the origin for a spherical. By plotting the corresponding variogram determine whether it is effectively linear up to 10m. In fact larger values of both a and C are needed. Knowing the slopes at the origin can prove helpful later for fitting experimental variograms.

Calculating the variance of linear combinations. Most estimators used in the earth sciences can be written in terms of linear combinations. This includes estimators based on polygons or on inverse distance and inverse distance squared. The following exercises are designed to give readers practice calculating these variances.

Ex 3.7 Suppose that two sample points, x_1 and x_2, are 100m apart. Calculate the variance of the linear combination: $Z^* = Z(x_1) + Z(x_2)$ where Z(x) is a stationary variable with a spherical variogram with a range of 250m and a sill of 3.

What would the variance be if the range is 25m instead of 250m?

What would the variance be if the variogram was a pure nugget effect of 3.0? Why is the value the same as when the range is 25m?

Ex 3.8 Consider a linear combination:

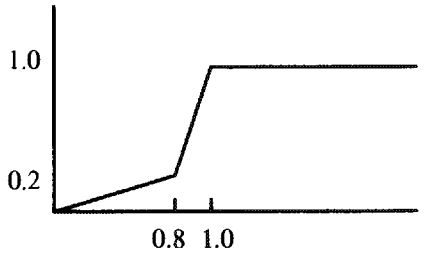
$$Z^* = \lambda_1 Z(x_1) + \lambda_2 Z(x_2) + \lambda_3 Z(x_3) + \lambda_4 Z(x_4)$$

where λ_1, λ_2, λ_3 and λ_4 are constants and x_1, x_2, x_3 and x_4 are the corners of a rectangle with sides of 30m and 40m. The variogram of Z(x) is a spherical with a range of 100m and a sill of 4.0 plus a nugget effect of 1.0. Calculate the variance of Z* for the case where $\lambda_1 = \lambda_2 = 1.0$ and $\lambda_3 = \lambda_4 = -1.0$.

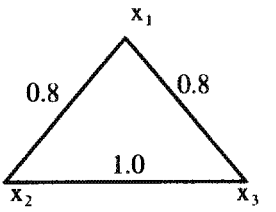
Would it be possible to evaluate the variance of the linear combination if the variogram was linear? Explain why. Would this still be true if the weights were all equal to 0.25?

Ex 3.9 To highlight the necessity of using only admissible models, here is a case where a non-standard model has been used. The model and the data layout are shown below. The three points form an isosceles triangle with sides of 0.8, 0.8 and 1.0. After noting that the sum of the weights equals 0, calculate the variance of the linear combination:

$$Z^* = Z(x_1) - 0.5\, Z(x_2) - 0.5\, Z(x_3)$$



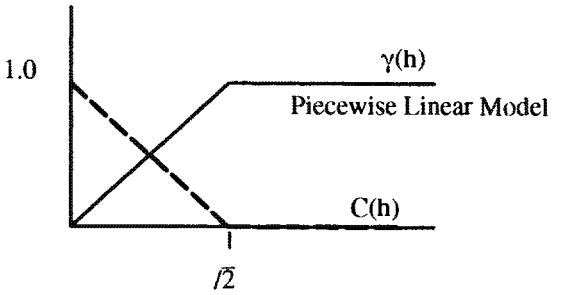
Non-standard variogram



Lay-out of points

The variance of this linear combination is negative because the function used as a variogram model is not positive definite or conditionally negative definite (even though its shape resembles a gaussian variogram model).

Ex 3.10 Piecewise linear model in 2D. This exercise is designed to highlight another subtle feature of positive definite functions. They can be positive definite in a 1D space but not in 2D and higher order spaces. We use the piecewise linear model to illustrate this point. Exercise 3.11 presents a construction for generating this variogram in 1D.

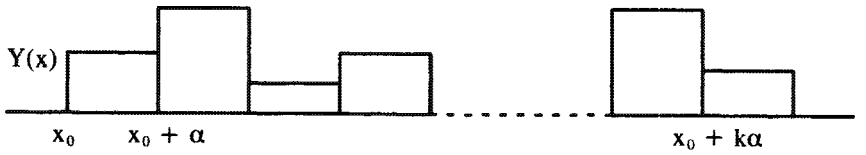


Set up a regular 1m x 1m grid containing 8 x 6 nodes (48 nodes). Let the weights be alternatively +1 and −1 so that neighbouring points always have weights with opposed signs. Check that the sum of the weights is zero. The variogram and the spatial covariance for the piece–wise linear model with a unit sill and a range of $\sqrt{2}$ are shown above. Calculate the variance of this linear combination. Hint: the calculation is much simpler if one works with covariances because they are all zero for distances greater than or equal to $\sqrt{2}$.

The fact that there is at least one linear combination with a negative variance proves that this model is not authorized in 1D. There is no way of "proving" that a model is positive definite by testing the variances of specific combinations because there is no way of testing them all. Much more general methods are required.

Constructing random functions to obtain new variogram models. As it is very difficult to test functions for positive definiteness, most new variogram models are obtained by a suitable construction. These exercises are designed to present some of the simpler ones, starting out from the piece–wise linear model in 1D.

Ex 3.11 Piecewise linear model in 1D. The aim of this exercise is to show that the piecewise linear model is an admissible variogram in 1D. After drawing an origin, x_0, at random in the interval $[0, \alpha]$, divide the line into segments of length α. A random function, $Y(x)$, is constructed by drawing a value for each segment from a distribution with mean m and variance σ^2. The values are independent from one segment to another.

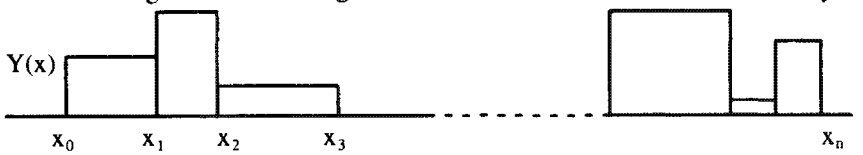


The probability that two points, x and x+h, chosen at random belong to the same interval depends on the distance between them. Show that the probability is 0 if $|h| > \alpha$ and that otherwise it equals: $1 - |h|/\alpha$.

Show that $E[Y(x + h) - Y(x)]^2 = \sigma^2$ if $|h| > \alpha$, whereas it is zero otherwise.

Hence show that the variogram of $Y(x)$ is a piecewise linear model with a sill of σ^2/α and a range of α.

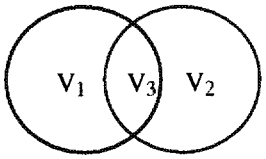
Ex 3.12 Exponential model in 1D. This exercise is a simple extension of the previous one. We construct a random function having an exponential variogram. As before, an origin x_0 is drawn at random in the interval $[0, \alpha]$ but this time we divide the line into segments whose length is a Poisson random variable with intensity λ.



As before, a random function Y(x) is constructed by drawing a value for each segment from a distribution with mean m and variance σ^2. The values are independent from one segment to another. Show that the variogram is an exponential with sill σ^2 and scale parameter λ.

Ex 3.13 Spherical model in 3D. In this exercise we construct a random function in 3D having a spherical variogram. Let the space be filled with Poisson points with intensity λ. So the number of points falling in a volume V is a Poisson random variable with parameter λV. Its mean and variance are both λV. Moreover the numbers of points in two volumes V and V' are independent if the volumes are disjoint. Let Y(x) be the number of Poisson points falling in a sphere of radius D centered on point x. Show that its variogram is a spherical model.

Hint: When the points x and x+h are further apart than D, the spheres do not intersect and so Y(x) and Y(x+h) are independent. When the points are closer together, split the two spheres into three disjoint parts as shown below:



Then if N(V) denotes the number of Poisson points in V,

$$E\left(Y(x+h) - Y(x)\right)^2 = E\left[\left(N(V_1) + N(V_2)\right) - \left(N(V_2) + N(V_3)\right)\right]^2$$

To complete the proof, it suffices to calculate the volume V_2 as a solid of rotation.

$$\text{Volume of } V_2 = \frac{\pi}{6} D^3 \left[1 - \frac{3h}{2D} + \frac{h^3}{2D^3}\right]$$

Ex 3.14 Linear model in 1D. This exercise is designed to construct a random function having a linear variogram. Let W_i be a set of independent random variables that take the values +1 and −1 with equal probability. A regionalized variable Y(n) is constructed for positive integral values of n, by summing the W_i up to n.

$$Y(n) = \sum_{i=0}^{n} W_i$$

Show that Y(n) is not second order stationary because its variance depends on the value of n, but that it satisfies the intrinsic hypothesis. Show that its variogram is linear.

4 Experimental Variograms

4.1 Summary

Like the preceding chapter, this one is on the variogram. The reader is shown how to calculate experimental variograms in 1D, 2D and 3D, and how to fit models to them. Several exercises are provided. The practical problems encountered with troublesome experimental variograms are discussed. These include outliers, almost regularly spaced data, and so on.

4.2 How to calculate experimental variograms

The experimental variogram can be calculated using the following formula:

$$\gamma^*(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i + h) - Z(x_i)]^2 \qquad [4.1]$$

where x_i are the locations of the samples, $Z(x_i)$ are their values and $N(h)$ is the number of pairs $(x_i, x_i + h)$ separated by a distance h – those actually used in the calculation. It is very easy to apply this formula when the samples are regularly spaced in 1D such as down a drillhole, along an underground gallery, a transect or a seismic profile. Example 1 illustrates the procedure.

If samples are missing from a regular pattern, their values should not be interpolated by averaging the neighbouring values, nor should a zero be inserted in its place because these distort the true variability. The square differences are calculated for all available pairs.

If data are not regular, the variograms are calculated for distance classes with an associated tolerance, usually 50% because this covers all possible distances. Going further, when the data are irregularly spaced in 2D, variograms are calculated for angular as well as distance classes.

4.3 In the plane

When data are two dimensional, the variograms should be calculated in at least four directions to check for anisotropies. The first step is to choose the variogram lag and its tolerance, then the principal angles and the angular tolerance.

The procedure then goes as follows. Taking each point x_i in turn, the program calculates the difference $x_i - x_j$ to find out which angular class and which distance class the pair belongs to. The corresponding square difference $(Z(x_i) - Z(x_i + h))^2$ is then calculated and added to the subtotal for the appropriate class. The count of elements in that class is also increased by 1. When all possible pairs have been treated for a given point x_i, the program moves onto the next one. At the end of the process the subtotals are divided by twice the number of pairs in which the successive origins are chosen. A flow-chart for this is shown in Fig. 4.2.

4.4 In three dimensions

The procedure given above could theoretically be generalized to three dimensions by considering classes of solid angles. But in practice the third dimension usually plays a special role. There is often much more variability in the vertical direction than in the horizontal, because of the stratification of many natural phenomena.

Consequently it is usually much more meaningful to calculate the variograms in the plane of the strata using the methods described in the last section and then calculate those perpendicular to this plane. Typically the vertical variograms are calculated using the data down each drillhole then the horizontal variograms are calculated in several horizontal directions. If the orebody has been tilted due to tectonic action, the variograms are calculated in the plane of the deposit and perpendicular to it.

4.5 Example 1: regular 1D data

Use formula [4.1] to calculate the experimental variogram for the first three distance classes for the data given below. The samples are regularly spaced every 5m in 1D.

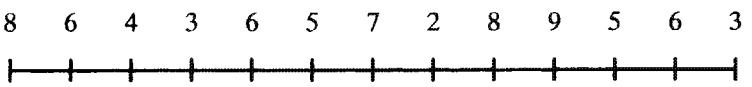


Fig. 4.1. Sample data spaced every 5m along a line, with grades shown

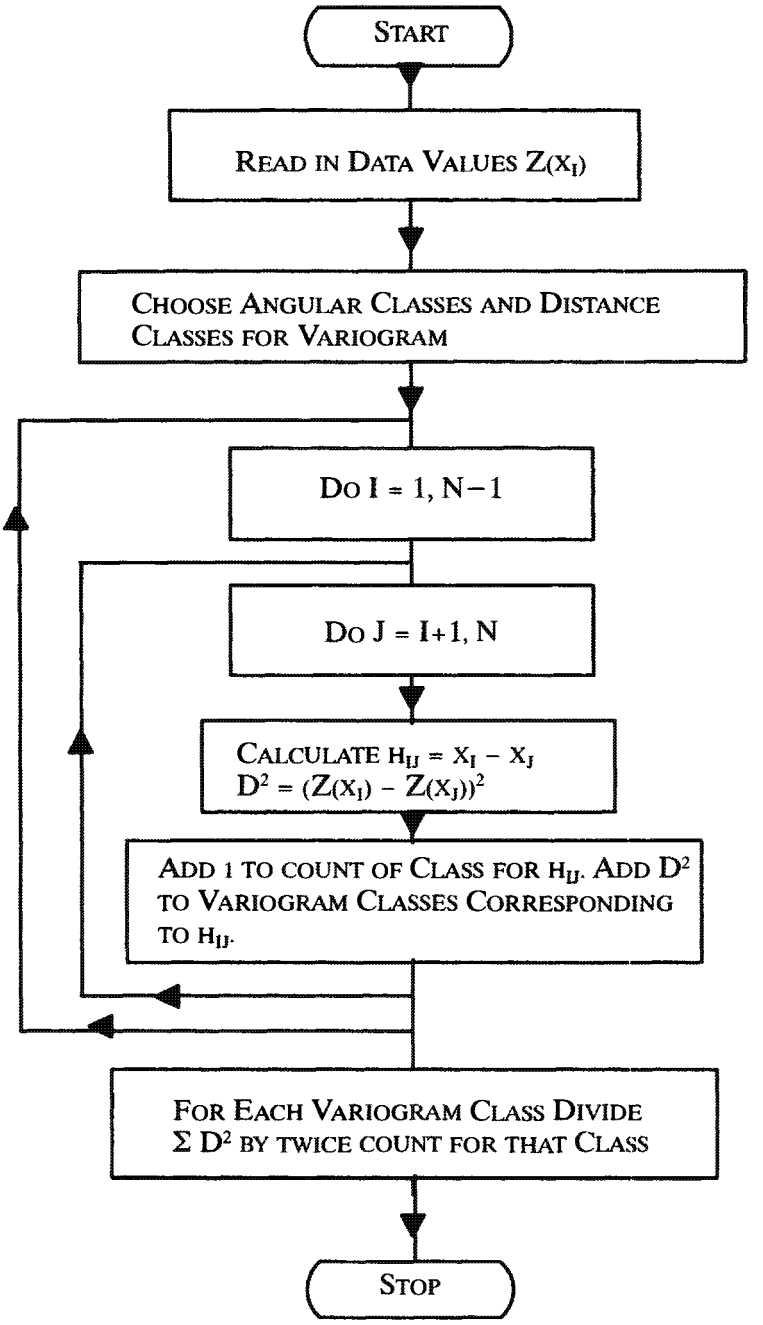


Fig. 4.2. Flowchart showing how to calculate experimental variograms

For the first variogram class (5m), there are twelve square differences:

$$\gamma^*(5) = \frac{1}{2x12}[2^2 + 2^2 + 1^2 + 3^2 + 1^2 + 2^2 + 5^2 + 6^2 + 1^2 + 4^2 + 1^2 + 3^2]$$

$$= 4.625 \qquad [4.2]$$

Show that the values of $\gamma^*(10)$ and $\gamma^*(15)$ are 4.82 and 6.00 respectively. Plotting these as a function of distance gives the experimental variogram (Fig. 4.3.). A solid line has been used to connect known values; a dotted one, to extrapolate back to the origin. The nugget effect could not be higher than 4.0 but could well be less if there were a micro–structure with a range of less than 5m.

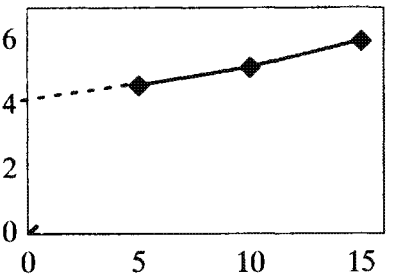


Fig. 4.3. Experimental variogram for samples spaced 5m apart. The solid line connects experimental values; the dotted one extrapolates back to the origin

When sample values have a skew distribution (with a long tail to one side) or when there are outliers in the data (i.e. unusually high or low values), the presence of just a few extreme values can cause trouble. Imagine what would have happened in Example 1, if the value of 7% had been 17% or 70% instead. In the first distance class, the square differences $(5–7)^2$ and $(7–2)^2$ would have become $(5–17)^2$ and $(17–2)^2$, or even worse, $(5–70)^2$ and $(70–2)^2$. These two terms would completely dominate the experimental variogram, making it difficult to interpret it or to fit a model to it. This point will be dealt with at the end of this chapter.

4.6 Example 2: Calculating experimental variograms in 2D

Table 4.1. shows 56 grades arranged in a unit square grid. Use them to calculate the experimental variograms in the four main directions for distances up to 4 lags. For each distance, note the number of pairs of points used. All pairs of points a particular distance apart in a specified direction should be used, not just those in the same row or column. Remember that the distance along the diagonals is a multiple of $\sqrt{2}$.

Table 4.1. Regular 2D data for calculating the experimental variograms

26	22	19	14	16	19	16	14
23	20	17	20	14	23	21	17
22	17	18	19	18	25	20	19
21	15	20	18	20	20	18	13
19	18	15	15	18	23	22	20
18	16	10	16	14	18	20	18
17	14	10	13	13	15	14	17
15	13	11	10	17	16	15	11

Table 4.2. Values of the experimental variograms in the four main directions for data given above, together with the number of pairs of points

	$\gamma(1)$	N(1)	$\gamma(2)$	N(2)	$\gamma(3)$	N(3)	$\gamma(4)$	N(4)
E–W	4.74	56	8.49	48	10.28	40	13.27	32
N–S	5.88	56	9.11	48	9.13	40	10.77	32
NE–SW	7.69	49	12.24	36	18.36	25	18.16	16
NW–SE	7.55	49	12.02	36	10.00	25	14.00	16

Table 4.2. gives the values of the experimental variograms in the four main directions and also the number of pairs of points used in their calculation. The variogram is not reliable for distances greater than half the field length. This is why it has been calculated for only four lags. Figure 4.4. shows the four directional variograms.

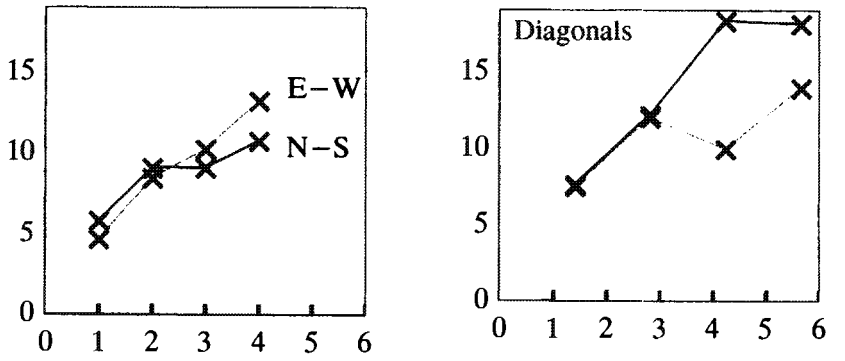


Fig. 4.4. Experimental variograms in 4 directions

As there is little difference between them, we consider them to be isotropic. So the average experimental variogram for all directions was calculated (Fig. 4.5.). A linear model with a slope of about 3 plus a nugget effect of about 3 gives a good fit. This example was designed to show how to calculate variograms. In practice, variograms usually turn out to be more erratic than these.

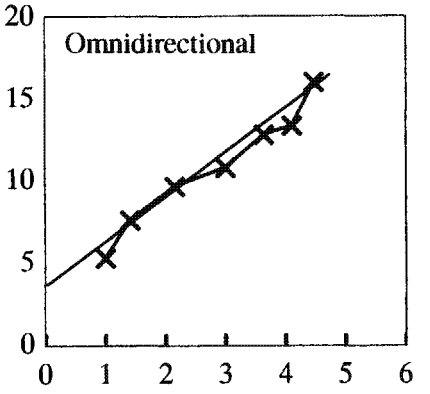


Fig. 4.5. Omnidirectional variogram calculated from the regular 2D data, and the fitted linear model with a nugget effect of 3 and a slope of 3

4.7 Variogram cloud

There are two ways of plotting variograms: the standard way shown above where the average square difference is plotted against distance or alternatively as a cloud of square differences each plotted against its distance. Chauvet (1982) called this the variogram cloud. The advantage of the standard plot is that it synthesizes all the information into one point per distance class, but in so doing, the detail is lost. And this detail can be very helpful in understanding the variogram.

To illustrate the concept, the variogram cloud has been calculated for the data in Example 1. For the first distance class there were 12 square differences: 4, 4, 1, 9, 1, 4, 25, 36, 1, 16, 1 and 9. For the next two distance classes, there were 11 and 10 pairs respectively. To be compatible with the variogram, these values were halved before being plotted. Figure 4.6.**a** shows the resulting variogram cloud. The experimental points sit in vertical columns because the data are regularly spaced every 5m. Only 6 crosses are apparent in the first distance class because there are only 6 distinct values for the half squares. While most of these squares are small, a few are quite large. So their histograms for each distance class are skew.

To illustrate the impact of outliers, the cloud was recalculated for the case mentioned earlier where a value of 17% occurs instead of 7% (Fig. 4.6.**b**). Compared to the original cloud on the left, the vertical scale on the right is four times

larger. For each distance class, the largest two square differences (i.e. the highest two crosses) correspond to pairs involving the outlier.

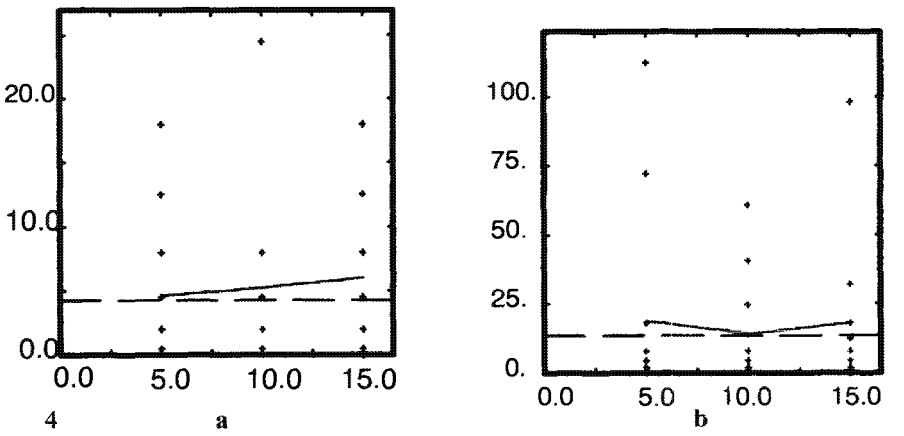


Fig. 4.6. Variogram clouds calculated using the 13 regularly spaced data (**a**) from Example 1 and (**b**) the same data when a value of 7% has been replaced by an outlier of 17%. Note the change in the vertical scale between the two figures. In both cases some crosses represent several values (e.g. 1.0 for a lag of 5m)

4.8 Fitting a variogram model

This is best seen from practical examples. Experience has shown that the analytic form of the model does not matter very much as long as its major features of the phenomenon are respected. In order of importance 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; the nugget effect can be estimated by extrapolating back to the origin. The first variogram value is often obtained from too few pairs of points to be reliable. Additional drillholes at short distances can be helpful so as to get a better idea of the nugget effect.

The range can usually be assessed visually. The sill is set at the value where the variogram stabilizes. For stationary variables this should coincide with the overall

variance but sometimes this is not true because of the presence of long range trends. If there is more than one range the intermediate ranges can be distinguished visually because the variogram has a change in curvature at these points. Modelling anisotropies requires more experience. 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, using a graphic terminal.

People often ask why we do not use least squares or other automatic methods to fit variogram models. There are three reasons for this. Firstly the model must be a positive definite function (otherwise the variance could turn out to be negative). Polynomials obtained by least squares regression would rarely satisfy these conditions. Secondly least squares assumes that sample points are independent observations, which is clearly not true of the experimental variogram. It consists of squares differences of combinations of values. Thirdly the behaviour of the variogram very close to the origin (i.e. for distances shorter than the first lag) is not known and yet it is vital and least squares does not take account of this. Experience and judgment are required. The first problem could be overcome by fitting only positive definite models, but this still does not solve the other two problems.

4.9 Troublesome variograms

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. Armstrong (1984) shows some of the more common ones. Here are some more examples.

4.9.1 Outliers

As was seen in the exercises on calculating experimental variograms, the presence of even one outlier can lead to a highly erratic variogram. In a study on coal from two seams in the Bowen Basin in Australia (Armstrong, 1980) the variograms for three of the variables (seam thickness, ash content and FSI) were very similar for both seams but the sulphur variograms were totally different. (Fig. 4.7.). This was rather surprising. A closer examination revealed that the data from the top seam contained two extremely high sulphur values among a group of 207 values. (These can be seen quite clearly on a histogram). The first step was to check whether these values were correct, by asking the geologist to inspect the remaining half–core. In fact, the samples come from a high sulphur area and the cores showed visible pyrite. So the values are real. After these two abnormal 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.

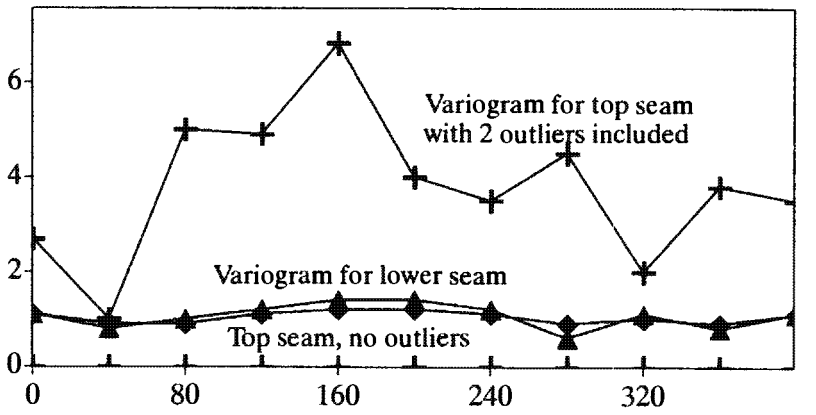


Fig 4.7. Three variograms for sulphur content in two coal seams The top one came from top seam with 2 outliers included, the other variograms came from the lower seam and the top seam after removing the outliers

In this case the solution was quite simple: eliminate the outliers. This was possible for two reasons: firstly because the outliers lie in a geographically distinct area that should be treated separately from the rest, and secondly because high sulphur coal is less valuable than low sulphur coal.

However 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 often makes the difference between opening the mine and not doing so. Eliminating the outliers or cutting them back to an arbitrary value is not a good solution.

Several "robust" methods of calculating the variogram have been suggested (Cressie and Hawkins, 1980; Armstrong and Delfiner, 1980). This subject was one of the major topics of discussion at the NATO geostatistics workshop held at Lake Tahoe in September, 1983 (See Verly, 1984).

4.9.2 Pseudo–periodic hiccups

The ash variogram calculated for a lag of 40m for coal from the Bowen Basin in Australia illustrates the problem of "pseudo–periodic hiccups" (Fig. 4.8.**a**). At first, the two rather strange peaks at h=150m and h=280m might seem to be a sign of some sort of periodicity in the coal but this is not physically likely given the nature of coal. Although it is not obvious initially, the samples lie on an almost regular grid. Plotting the histogram of the distances between pairs of points for each distance class shows that this was clearly the case. (Fig 4.8.**b**.) In this case the solution is quite simple. All we need do is to change the step length. Calculating the variograms for 100m distance classes smoothed out the bumps.

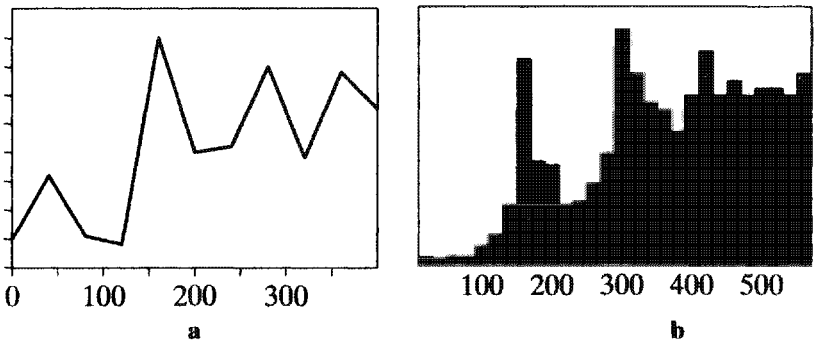


Fig 4.8. (**a**) Erratic experimental variogram for ash content in coal, (**b**) the histogram showing the number of pairs of points for each variogram lag on right Note the peaks in the histogram (160m & 320m)

4.9.3 Artefacts

The next two examples have been included to point out that many of the problems with variograms are due to "operator error" rather statistical problems with the data. Figure 4.9. shows a remarkable "saw-tooth" variogram. The variable under study was highly skew, like gold or uranium. A substantial percentage of values below the recording threshold had been recorded as zero. Since the data looked lognormal it seemed 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 (here to base 10). 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. In this case, the solution was simple: set the below cutoff values to 0.001 rather than 0.00001, in which case their log is –3.0.

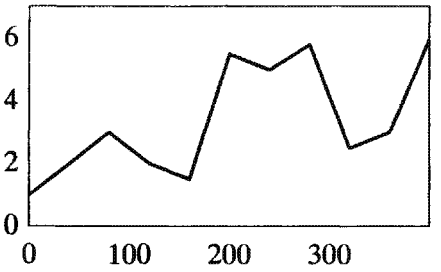


Fig 4.9. : Erratic saw tooth variogram

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 seam thickness. These could indicate that the seam had pinched out in this region but as this seems unlikely, the data were re-examined. This revealed that the zero seam thickness corresponded to missing values for both the top and bottom of the seam. As a value of –1 had been assigned to all missing values, and as the student doing the study forgot to include a test for missing values in his program to calculate seam thickness, the subtraction of –1 from –1 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 more common than most of us care to admit. What is more the only way to find out their cause and treat them is by carefully investigating the data. It would be absolutely fatal to apply a "robust" variogram method to the data to sort out the problems with the variogram. Fortunately computer technology now allows us to work with linked windows. So it is possible to visualize several graphics (such as the base map, the histogram and the variogram) simultaneously to find out what is causing the erratic behaviour.

4.10 Exercises

Ex 4.1 Missing values. Figure 4.10. shows 13 sample locations spaced 5m apart. The grades are available at 12 out of 13 points, the other grade is missing. Show that there are only 10 pairs of points for the first lag, and that the value of the $\gamma^*(5) = 4.7$. Calculate the experimental variogram for the next two lags and plot it. Sometimes people mistakenly put a zero in the place of the missing value. Calculate the first three lags of the experimental variogram and compare it to the previous one.

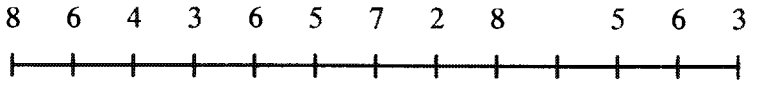


Fig 4.10. Regular 1D data with 1 value missing

Ex 4.2 Outliers. The figure below shows 13 sample locations spaced 5m apart, with one grade (90) much larger than the others. Calculate the experimental variogram for the first three lags and plot it. To see the impact of the outlier, compare this variogram to the one obtained earlier (Fig. 4.3.)

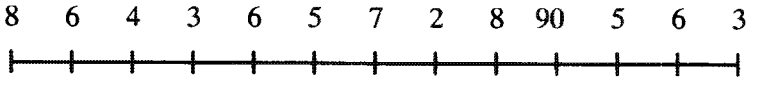


Fig 4.11. Regular 1D data containing an outlier

Ex 4.3 Variograms in 2D. The table below gives 35 grades on a regular 100m x 100m grid. Calculate the experimental variograms in the two principal directions for up to 300m. Is this variogram isotropic?

Table 4.3. Thirty–five grades on a regular 100m x 100m grid

3.6	2.8	4.1	4.6	5.3	3.8	4.2
3.6	3.5	5.2	4.5	6.1	4.4	4.0
5.1	3.8	4.9	3.3	5.7	6.2	6.3
4.2	4.0	5.6	4.2	4.9	5.3	4.4
4.6	5.7	6.1	5.4	4.7	5.2	6.0

5 Structural Analysis

5.1 Summary

Several case studies showing how to carry out a structural analysis are presented in this chapter. Firstly, the principal decisions that have to be made by the geostatistician are reviewed. Are the data stationary? Are they isotropic? Should we work with the variables themselves or their accumulations? Should the study be carried out in 2D or 3D?

The first case study is a relatively simple 3D study of an iron ore deposit. As the horizontal and vertical variograms are well structured, it provides a clear illustration of a straightforward variographic analysis. The second study concerns an Archaean gold deposit that is being mined by opencut methods. So the grades of closely spaced blastholes are available in addition to the more widely spaced exploration drillholes. The third deposit presented is a sedimentary gold deposit with a periodic variogram in one direction because the gold was deposited by stream action. In contrast to the other two deposits, this one is only about 1m thick and so the study was carried out in 2D rather than 3D.

5.2 Steps in a case study

The first phase of any geostatistical study 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 variogram and thirdly fitting a mathematical model to the experimental variogram. As the second and third steps were dealt with in the previous chapter, we will concentrate on the first one. Box 4 summarizes the three main steps in a structural analysis and shows how it fits into the broader framework of a case study.

5.2.1 Step 1: Collect and check the data

After the data has been collected and put into the computer, it has to be checked thoroughly to see that it is correct and representative of what is being studied. This means that any numerical errors in the data or in the coordinates have to be corrected, and some elementary statistics should be calculated. But more importantly, the geostatistician has to familiarize 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, if you have not been involved in the project since its inception, it is important to find a geologist or an engineer who worked with the sampling program and find out:

1. what type of sampling procedure (or procedures) was used,
2. what sample volumes were collected, what types of analyses were carried out and in which laboratories,
3. whether there were any changes in the procedures used during the exploration campaigns. For example, were different drilling companies used at different times? Were different types of gamma logging devices used at different times?
4. whether the area is geologically homogeneous, whether it contains major faults,
5. whether there has been preferential sampling of high grade areas.

If any of these factors is missed at the beginning of a study, the work may well have to be repeated when they are discovered. These basic checks have always been vital at the outset of any ore reserve calculation, but they are even more important nowadays because in many countries, people carrying out these tasks are legally responsible for any errors and omissions in their work. Geostatisticians working in companies or as independent consultants should think carefully about the implications of "due diligence" to ensure that their work is in compliance with the changing legal situation. A brief discussion of this is given in Appendix 2.

5.2.2 The decisions to be made

Initially a series of decisions has to be made which guide the whole of the study. Firstly the variables and the geographical zones to be studied must be defined. Then the geostatistician has to decide:

1. whether the variables are stationary,
2. what their support is,
3. whether they are additive,
4. whether to work with the variables themselves or their accumulations, and
5. whether to carry out the study in 2D or 3D.

Stationarity. In Chapter 2 the question of determining whether a regionalized variable could be considered stationary or not was discussed. So we will not go over it here.

BOX No 4 : Steps in a case–study.

Step 1: Collect and check data.

If you were not actively involved from the start of the project, find those who were and ask them about

- the types of sampling and analyses used and any changes in procedure,
- different geological zones, faulting etc,
- preferential sampling, etc.

At the outset a series of major decisions has to be made.

- Whether to work with grades in 3D or with accumulations in 2D.
- The limits of area to study, the support of the variables and whether they are stationary,

Basic statistics (means, variances, correlations, histograms and scatter diagrams) are calculated. Look for

- outliers or abnormal values
- nonhomogeneous data (mixed populations)

Step 2: Calculate experimental variograms.

Step 3: Fit a variogram model.

Step 4: Kriging or simulation

Support. The geostatistical term "support" refers to the size and shape of a volume. Diamond drillholes and mining blocks have quite different weights and volumes (kilograms compared to hundreds of tons of ore). So although the mean grades should be the same, their variances are quite different. Diamond drillholes and percussion holes could have the same diameter, and yet the statistical characteristics of the data need not be the same.

Additivity and accumulations. In almost all applications in geostatistics the variables studied have to be additive. (One of the rare exceptions is when the objectives of the study are limited to contour mapping). Otherwise, the variables must be additive; that is, the mean over a certain zone must be the arithmetic mean of all the values inside it. This point can best be seen from an example.

Suppose we wanted to find the average reef thickness and the gold grade from two cores, one with a 2m thick intersection and with a grade of 5 g/t, the other with a 3m thick intersection with a grade of 10 g/t.Clearly the average thickness is 2.5m, but the average gold grade is not the arithmetic mean

$$\frac{5 + 10}{2} = 7.5 \text{ g/t} \qquad [5.1]$$

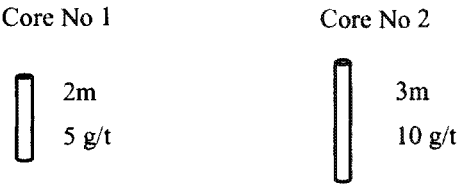


Fig 5.1. Two cores of different lengths

Rather it is the weighted average:

$$\frac{2 \times 5 + 3 \times 10}{2 + 3} = 8 \text{ g/t} \qquad [5.2]$$

The arithmetic average of gold grades (7.5 g/t) would give an entirely false estimate of the grade that could be mined from this area. Hence the need to use the product of the thickness times the grade. This is called the accumulation.

A geostatistician would normally carry out the study on the accumulation and the thickness, and at the end make the change back to the ordinary variables by dividing the kriged accumulation by the kriged estimate of the thickness. Alternatively he might cokrige the accumulations and the thickness and then divide to get grade estimates. Another point to note is that if the density of the ore varies from place to place, it would be wiser to use the accumulation products: grade by thickness by density and thickness by density.

Working in 2D or 3D. Deposits can be split into two broad categories depending on their geometry and the mining method. The first one consists of relatively thin deposits such as coal seams or the gold reefs while the second consists of thicker, more massive deposits which are divided into blocks of constant height for mining. In the first case, the whole mineralized thickness is extracted so there is no vertical selectivity. Consequently the study is carried out in 2D using accumulations rather than grades. In the second case, as the mining blocks have a constant height, the grades themselves are additive and so the study is carried out in 3D on the grades, using information from the levels above and below the one of interest. Open cut mines are a good example of this. In between these two extremes there is a range of orebodies such as thick seams where vertical selectivity is possible, and underground mines where stopes have irregular shapes. Here we limit ourselves to the basic cases. Examples of more complicated ones can be found in journal articles.

5.2.3 Standard statistics

As geostatistics assumes that the data come from a homogeneous population, it is important to apply a few simple statistical tests before calculating 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 locally high or low values. Sometimes abnormal 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.

5.3 Case studies

Three case studies will be discussed; one on iron ore and two on precious metals. The first concerns an iron ore deposit with about 40 vertical drillholes. The cores were cut into 15m high sections which were analysed for several quality variables (silica, alumina, loss on ignition and sometimes manganese content) as well as iron content. As this book deals with univariate geostatistics, the case–study will focus on the primary variable, iron content. Those who are interested in multivariate estimation can consult the book by Wackernagel (1995). This study was carried out in 3D for two reasons: firstly because it is massive rather than seam–like and secondly because the deposit will be mined as an openpit using 15 m high benches.

The second study which was carried out by M. Harley as his CFSG project.[1] It deals with an Archaean gold deposit that is being mined by opencut methods. Whereas the first deposit was relatively easy to estimate, the second one was more complex. Its geometry is more complicated and the distribution of the grades (their histogram) is quite skew. Another difference between these two studies is that in this case two sets of data were available: the initial exploration drillholes and the blastholes. The third study was carried out by M. Thurston as his CFSG project. It is also on a gold deposit. Unlike the previous studies, this one was done in 2D. As the gold in the Witwatersrand was laid down by stream action, the variograms in the directions parallel and perpendicular to the current are different. The latter one shows periodicity.

5.4 An iron ore deposit

The deposit under study is still at the feasibility stage but will be mined as an openpit with 15m high benches. The area of interest contains about 40 vertical drillholes; their layout is shown in Fig. 5.2.**a**. A total of 485 core sections 15m long were analysed for several quality parameters such as silica content and alumina content, as well as their iron ore grade. Here we will consider only the Fe grade.

[1] The CFSG (Cycle de Formation Specialisee en Geostatistique) is a 9 month postgraduate course that trains geologists and engineers to be specialists in mining or petroleum geostatistics.

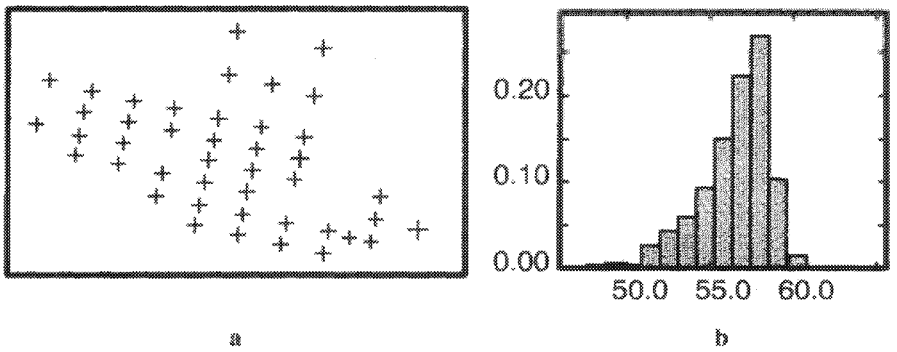


Fig 5.2. (**a**) Base map showing the layout of drillholes; (**b**) histogram of 485 Fe grades from 15m core sections. As is usual for iron ore, the grade histogram is negatively skew with a tail of small values to the left

5.4.1 Vertical variogram

The vertical variograms were calculated down the holes using a lag of 15m for distances up to 135m. Table 5.1. lists the variogram values and the number of pairs of points for each distance class. As expected the number of couples decreases with increasing distance, and secondly the variogram is well structured (Fig 5.3.**a**).

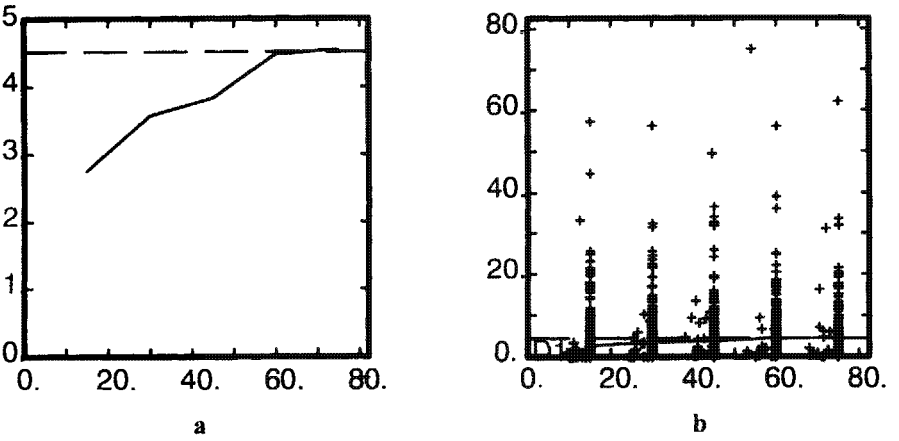


Fig 5.3. (**a**) Vertical variogram and (**b**) the corresponding variogram cloud. Note the change of scale on the vertical axis

Table 5.1. Experimental values of the vertical variogram

Lag	15m	30m	45m	60m	75m	90m	105m	120m	135m
Pairs	388	329	284	247	209	187	169	153	139
γ	2.75	3.57	3.82	4.49	4.56	4.97	4.77	4.80	5.13

5.4.2 Variogram cloud

Before fitting a model to this, the variogram cloud was calculated (Fig 5.3.**b**). The dotted line indicates the sample variance (4.5); the continuous line is just the classical variogram shown on the left. The (half) squares differences range in value from virtually zero to 80 (i.e. nearly 20 times the variance). Although the vast majority of couples occur at distances that are multiples of 15m, a few do not. In those cases, one of the samples making up the pair was shorter than the standard 15m; in fact, these come from the bottom of a drillhole where it hit bedrock.

5.4.3 Fitting a model to the vertical variogram

Now to fit a model to the vertical variogram. Figure 5.3.**a** shows that it stabilizes at a height of about 4.5 at a distance of 65–70m, which gives the total sill and the range. Extrapolating back to the origin gives a nugget effect of at most 2. This suggests trying a nugget effect of 2.0 plus either a spherical with a sill of 2.5 and a range of about 65m or the equivalent exponential. Figure 5.4.**a** shows that a good fit is obtained with a nugget effect of 1.8 plus a spherical with a range of 65m and a sill of 2.7, but that the corresponding exponential fits badly (Fig. 5.4.**b**).

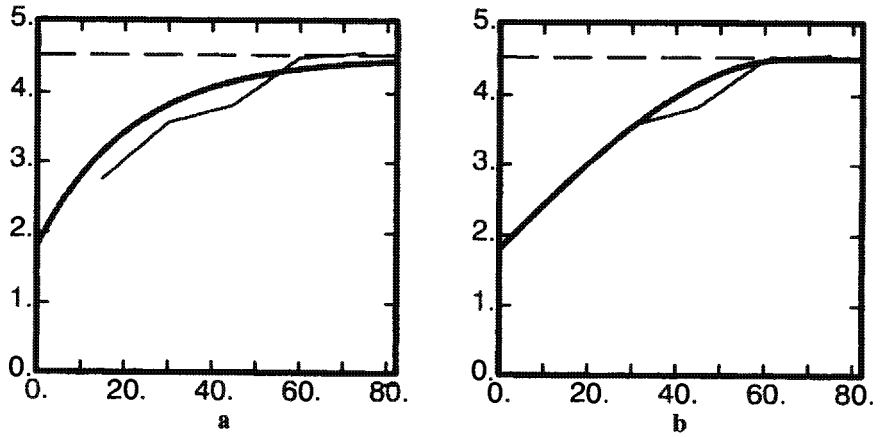


Fig 5.4. Two attempts at fitting the vertical variogram; **b** a nugget effect of 1.8 plus a spherical with a range of 65m and a sill of 2.7, and **a**, the equivalent exponential

This combination of a spherical plus a nugget effect is not the only model that would fit properly. The nugget effect could be replaced by any structure having the same sill and a range of less than 15m. To illustrate this, Fig. 5.5. compares a model consisting of two sphericals with ranges of 10m and 65m and sills of 1.8 and 2.7 respectively with the spherical model earlier. It would, of course, be possible to use a short range structure plus a nugget effect provided their sills sum to 1.8.

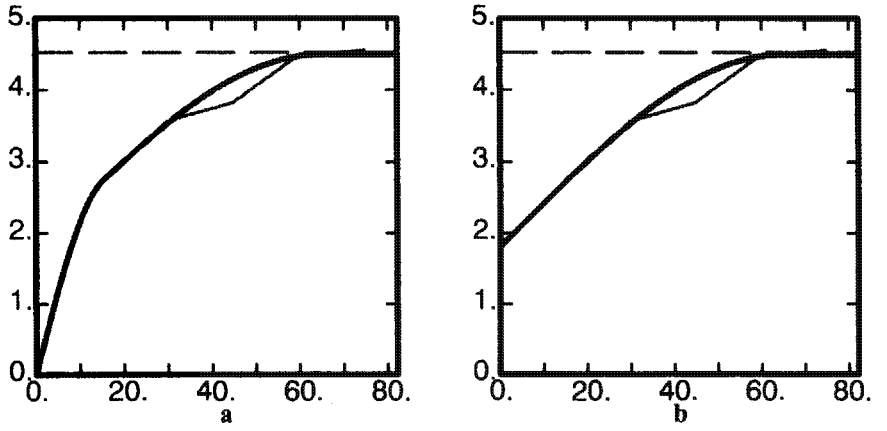


Fig 5.5.: Two models which both fit the experimental variogram; (**a**) two sphericals with ranges of 10m and 65m and sills of 1.8 and 2.7 respectively and (**b**) a nugget effect of 1.8 plus a spherical with a range of 65m and a sill of 2.7

5.4.4 Horizontal variograms

The next step is to choose the parameters for calculating the horizontal variograms. The base map (Fig 5.2.) shows that the spacing between drillholes is approximately 80m. In order not to miss any anisotropies, four variograms were calculated along the directions, E–W, N–S, NE–SW and NW–SE, with an angular tolerance of 22.5° so as to give complete coverage. One more parameter, the vertical slicing height, also has to be chosen. A value of 15m ensures that only horizontal couples are included. Figure 5.6.**a** shows that the four directional variograms are isotropic and can be grouped into a single variogram (Fig 5.6.**b**). Not surprisingly the latter is better structured.

5.4.5 3D variogram model

When the variogram model is used in kriging, we will require its value for oblique vectors, that is, for distances with horizontal and vertical components (not just a horizontal component or a vertical one). Consequently we need a 3D variogram model not just separate horizontal and vertical ones.

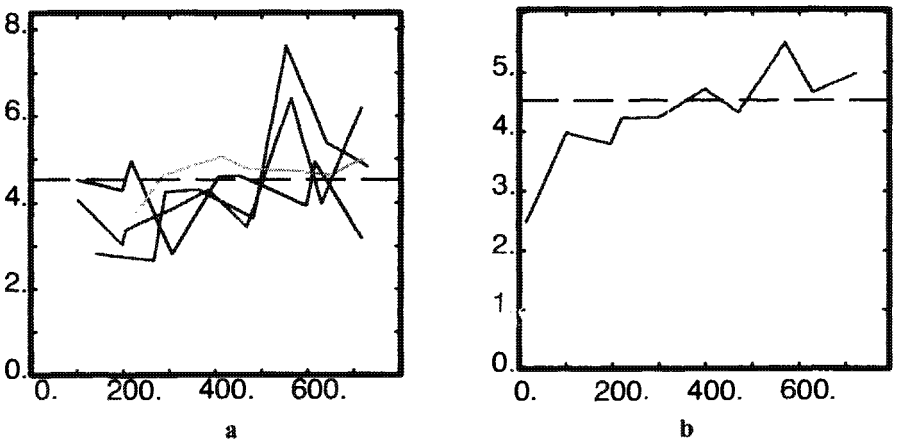


Fig. 5.6. (**a**) Four directional variograms in the horizontal plane. As there is no anisotropy these were averaged (**b**) and a model was later fitted

As the sills are approximately the same (Fig. 5.7.), a model with geometric anisotropy can be used. As a nugget effect of 1.8 seems appropriate horizontally as well as vertically, the vertical model consisting of this nugget effect plus a spherical with a sill of 2.7 and a range of 65m was used as the starting point. Several attempts were made to fit this to the horizontal variogram just by varying the anisotropy ratio but the curvature was not right. However a good fit was obtained by splitting the spherical into two components with different anisotropy ratios. Figure 5.7. shows the experimental variograms in the horizontal and vertical directions together with the fitted model. Table 5.2. shows the parameters of the fitted model, which will be used in the case study on point and block kriging in Chapter 9.

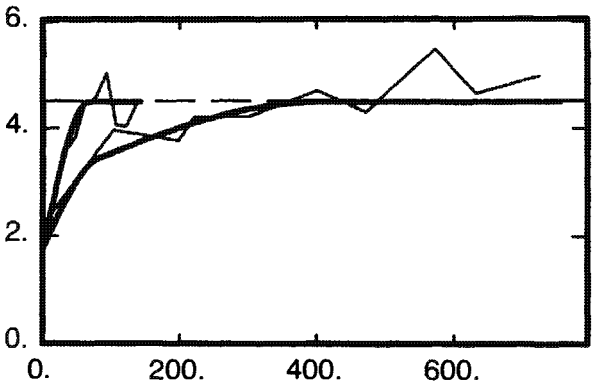


Fig. 5.7. Experimental variograms in the horizontal and vertical directions and the fitted 3D model

Table 5.2. Parameters of fitted variogram model which is isotropic in the horizontal direction but not in the vertical

	Sill	Horizontal Range	Vertical Range
1st Spherical	1.2	80m	65m
2nd Spherical	1.5	400m	65m

5.5 Second case-study: an archaean gold deposit (M. Harley)

Archaean gold orebodies are common in many regions: Western Australia, Central and West Africa, Brazil, Southern India and Guyana. Like most such deposits, this one is being mined by open pit methods. The orebody strikes almost north–south, dipping to the east at about 70°. It ranges in thickness from 10m to 60m.

During the exploration campaign, about 170 holes were drilled and samples 1m long were taken for analysis. As the mining benches are 5 m high, the sample grades were regularized over this height. Even after regularization, the histogram of gold grades was quite skew with the maximum value being about 20 times the average. Difficulties could therefore be expected when interpreting and modelling the variograms. The experimental variograms calculated in the principal directions (down hole, down dip and along strike) shown in Fig. 5.9.**d, e, f** confirm this.

As the mine is now in production, blasthole samples are available on a 3 m x 5 m grid. Their length is also 5 m. Figure 5.8. shows the histogram of about 7000 blasthole grades. Their coefficient of variation is 1.13. Their variograms were calculated in the three principal directions. Whereas the drillhole variograms were highly erratic, these are much better structured (Fig 5.9.**a, b, c**). This is because the spacing is close enough to reveal the short range structures.

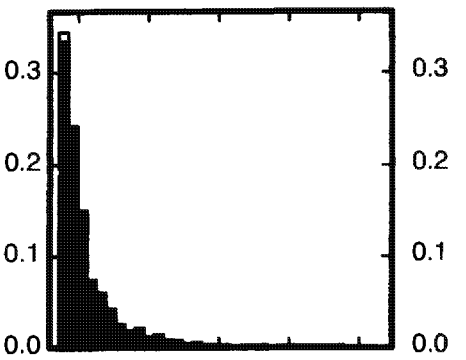


Fig. 5.8. Histogram of blasthole grades

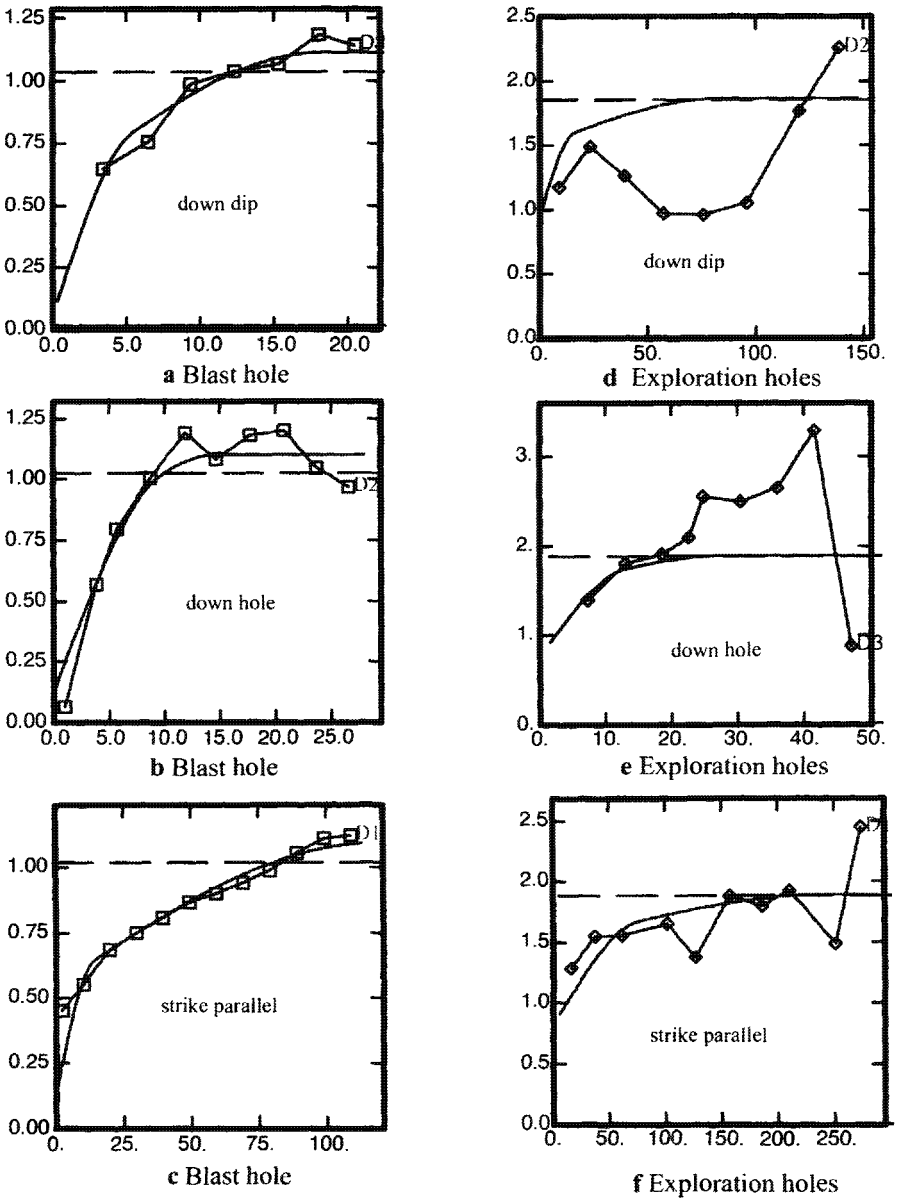


Fig. 5.9. Experimental variograms for an Archaean gold deposit: from closely spaced blastholes (**a, b** and **c**) and from exploration drillholes (**d, e** and **f**)

Table 5.3. Short and long ranges of the blasthole variogram model in the three main directions

Direction	Down Hole	Down Dip	Along Strike
Short Range	5m	10m	30m
Long Range	18m	15m	120m

The ranges of the variograms are different in the various directions, so a model with geometric anisotropy was fitted. It consisted of 10% nugget effect, 40% short range spherical plus 50% long range spherical. The ranges of these two spherical structures are listed below for the three main directions.

5.6 Third study: a Witwatersrand gold deposit (M. Thurston)

The data for this study come from a sedimentary gold deposit. They are the gold accumulations in cmg/t corresponding to 15m x 15m blocks; that is, the original channel sample data have been averaged for each of the blocks. This reduces the quantity of data to be handled from several thousand values to several hundred and also smoothes it out. As is common with precious metal deposits, the distribution of the values is skew (Fig. 5.10.).

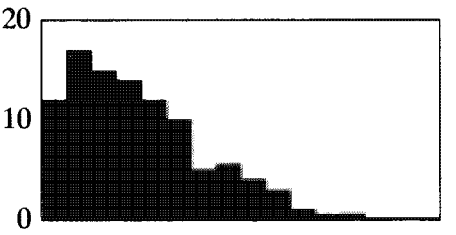


Fig. 5.10. Histogram of gold accumulations

The experimental variograms of the gold accumulations were then calculated in four directions. The principal two (parallel to the current and perpendicular to it) are shown in Fig. 5.11. The NE-SW one has a longer range (about 220m), and so indicates the direction of greatest continuity for the deposit. The other one has a shorter range (about 75m) and also dips down at 150 to 160m and at 300 to 320m. As the gold in this deposit was laid down by the action of water, this indicates that the current flowed NE–SW. This confirms what the mine geologists already knew.

Another interesting feature is the periodicity in the direction NW–SE (the one with the shorter range). The variogram reaches a maximum at a distance of about 75m, drops to a minimum at 150m, rises again and drops to a second minimum at

about 300m. This is due to parallel channels in the streams that deposited the gold. The direction NW–SE cuts across these channels which are parallel and occur regularly every 150m in this area. Whereas the geologists knew the flow direction, they had not realized that the channels were 150m apart on average. So the detailed variogram study added to their knowledge.

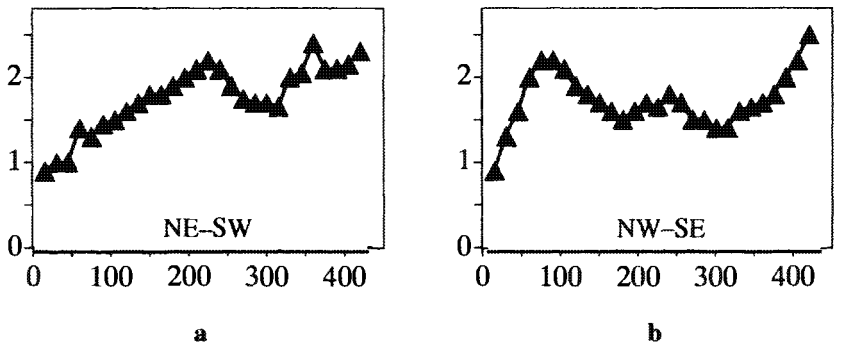


Fig. 5.11. Directional variograms of the gold accumulations; (**a**) parallel with the direction of the current and (**b**) perpendicular to it

Having said that, the next question is to decide whether to choose a variogram model that incorporates the periodicity. For kriging, we only use data up to about 100m from the block to be estimated. So the variogram model only has to fit up to this distance. A model with geometric anisotropy with a range of 80m in the direction NW-SE, and a range of 200m NE-SW was appropriate. A spherical model with these ranges and a nugget effect equal to about 35% of the total sill gave a good fit. So as far as kriging is concerned, there was no real point in modelling long range features like the periodicity.

Lastly as the data were approximately three–parameter lognormally distributed, the experimental variograms of the log were also calculated for comparison with the raw variograms (Fig. 5.12.). The shapes are generally similar to the raw variograms but the fluctuations are less accentuated. The variogram of the logs is more stable numerically, and shows the range and the sill more clearly.

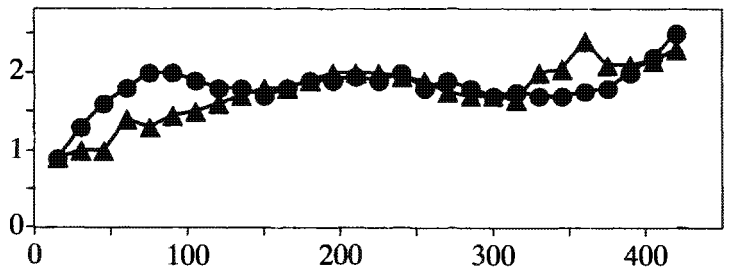


Fig 5.12. Variogram of the logs of the gold accumulations

6 Dispersion as a Function of Block Size

6.1 Summary

This chapter deals with the effect of the support of a regionalized variable (i.e. its physical shape and volume) on its histogram and its variogram. In an introductory exercise, the histograms and the basic statistics are calculated for two support sizes: 1m x 1m blocks and 2m x 2m blocks. Although the means of the two distributions are identical, the variance of the larger support is much smaller and its histogram isalmost bell shaped whereas the other one is skewed.

The formulas for the variance of a point within a block and of one support v inside another V are then given. Krige's additivity relation is proved. Then we see how the regularized variogram is related to the point support variogram. An exercise illustrates the effect that regularizing has on the variogram.

6.2 The support of a regionalized variable

In many practical situations a regionalized variable is measured as the average over a certain volume or surface rather than at a point. The basic volume on which a regionalized variable is measured is called its support. Changing its support leads to a new regionalized variable which is related to the preceding one but which has different structural characteristics. For example the grades measured on 2 inch cores (i.e. with a 50 mm diameter) have a higher variance than those measured on larger diameter cores, or on blocks or bulk samples. 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.

6.2.1 Dispersion versus block size

To illustrate the effect of support, we consider the data from the introductory example in Chapter 1. Table 6.1 gives the grades of the 64 adjoining 1m x 1m blocks and also the average yields for the sixteen 2m x 2m blocks obtained by averaging 4 adjacent 1m x 1m blocks. These values came from about 1000 millet yield values studied by Sandjivy (1980). As expected, the means of the data are the same (201) except for differences due to roundoff. But the variances are not. The variance for the 2m x 2m blocks is 16,641 which is smaller than that of the 1m x 1m blocks (27,592). If the values were statistically independent the variance for the larger support would be ¼ of the other one. Because of the correlations it is higher.

Figs. 6.1. and 6.2. show that the shape of the histogram has changed too. The second one is less skew. The implications of this change are very important in mining. In selective mining only those blocks with a grade above a cutoff can be mined profitably. So it is vital to be able to predict the proportion of ore above a cutoff. As was shown in Chapter 1 when the polygonal method is used for reserve estimation, the grade of the sample inside the polygon is taken as the estimate for the whole polygon. This leads to equating the histogram of core grades to that of blocks, and hence to serious errors in estimating the recoverable reserves because the histograms are quite different, as can be seen by comparing Figs. 6.1. and 6.2.

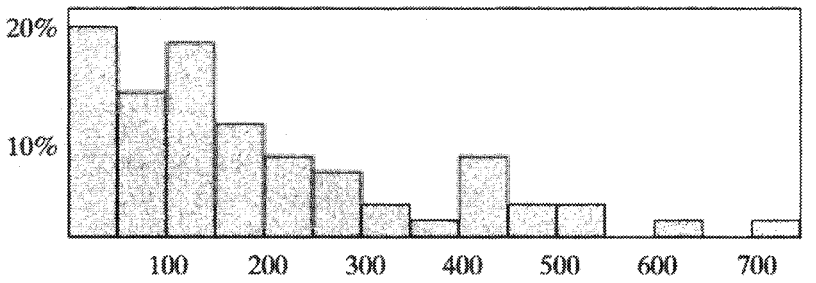


Fig. 6.1. Histogram of the grades of the 64 1m x 1m blocks

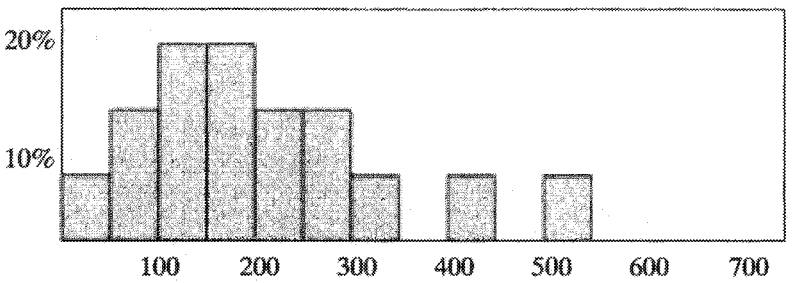


Fig. 6.2. Histogram of the grades of the 16 2m x 2m blocks

Table 6.1. Grades of (**a**) the sixty-four 1m x 1m blocks and (**b**) the sixteen 2m x 2m ones

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

a 1m x 1m blocks

505	143	88	207
270	328	171	411
102	220	154	263
101	54	44	155

b 2m x 2m blocks

6.3 Variance of a point within a volume

We now go on to see how to evaluate the variance of blocks given the variogram of samples. For the sake of generality, the supports will be called v and V. If the data are 2-D, these would be areas rather than volumes. In our model, the variable under study is considered as a realization z(x) of a random function Z(x). If all the values within the volume V were available it would be possible to find the mean over this volume and also the variance of the values within this volume. The mean is

$$m_V = \frac{1}{V}\int_V z(x)\, dx \qquad [6.1]$$

Similarly the variance of the values within the volume V is given by

$$s^2\,(0|V) = \frac{1}{V}\int_V (z(x) - m_V)^2\, dx \qquad [6.2]$$

Here 0 denotes a point i.e. something with a zero volume. If we let the realization vary, the variance of z(x) within V can be obtained as the expected value of s^2 (0|V) over all possible realizations:

$$\sigma^2\,(0|V) = E\,[s^2\,(0|V)] \qquad [6.3]$$

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

$$\sigma^2\,(0|V) = \frac{1}{V^2}\int\int \gamma(x - y)\, dx\, dy \qquad [6.4]$$

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

$$\sigma^2\,(0|V) = \overline{\gamma}(V, V) \qquad [6.5]$$

In practice $\overline{\gamma}(V, V)$ is calculated by discretizing the block V. Exercise 6.1 at the end of the chapter shows readers how to program the calculation.

6.4 Variance of v within V

We now consider a new random function defined as the spatial average inside a volume v:

$$Z_v(x) = \frac{1}{v}\int_v Z(x + t)\, dt \qquad [6.6]$$

The aim is to find the dispersion of this new variable $Z_v(x)$ 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 selective mining unit and V could be the whole deposit.

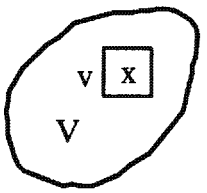


Fig. 6.3. Small block v centered on point x inside the volume V

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

$$\sigma^2(v|V) = E\left(\frac{1}{V}\int_v (Z_v(x) - m_V)^2\, dx\right) \qquad [6.7]$$

Expanding this gives:

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

$$= \bar{\gamma}(V|V) - \bar{\gamma}(v,v) \qquad [6.8]$$

6.5 Krige's additivity relation

Combining the results [6.5] and [6.8] gives an equation called Krige's additivity relation.

$$\sigma^2(v|V) = \bar{\gamma}(V,V) - \bar{\gamma}(v,v) = \sigma^2(0|V) - \sigma^2(0|v) \qquad [6.9]$$

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

$$\sigma^2(v|V') = \sigma^2(v|V) + \sigma^2(V|V') \qquad [6.10]$$

For example, v could be a core section, V a block and V' a large panel or the whole deposit. In that 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". We now check this experimentally for the millet data given earlier. Here v corresponds to a 1m x 1m block while V corresponds to a 2m x 2m block.

From before we have:

$$\sigma^2 (v|V') = \sigma^2_{1x1} = 27,592 \quad [6.11]$$

$$\sigma^2 (V|V') = \sigma^2_{2x2} = 16,641.1 \quad [6.12]$$

The value of $\sigma^2(v|V)$ can be calculated experimentally as the variance of the four small blocks with each larger one. This gives 10,951. It is easy to verify that this value equals 27,592 – 16,641 and hence satisfies the additivity relation. In fact this is true for all cases where the small blocks v exactly fill the next size block up, here V.

6.6 Exercise: stockpiles to homogenize coal production

Often the grade of the run of mine coal arriving at the entry to a coal wash plant or a power station fluctuates too much. The problem is to decide whether it would be economically worthwhile to blend the coal in stockpiles so as to homogenize its quality. Linear geostatistics can be used to calculate the variability (the variance) of the average value of blocks of certain sizes. This assumes perfect mixing.

Suppose that the ash content of coal has a spherical variogram with a range of 300m and a sill of 5.0. Each day the company mines a block 60m x 100m (denoted by v); each week six adjoining blocks are extracted by strip mining. The width of the strip is determined by the length of the boom on the dragline and is 60m. So V is 60m x 600m. Evaluate $\bar{\gamma}(v, v)$ and $\bar{\gamma}(V, V)$, and hence work out the variability of one day's production in that of a 6 day working week.

6.6.1 Solution

The first step is to calculate $\bar{\gamma}(v, v)$ and $\bar{\gamma}(V, V)$. There are two ways of doing this, either by writing a small computer program or by reading the appropriate values from standardized charts. Exercises 6.1 and 6.2 indicate how to apply these methods. The results are

$$\bar{\gamma}(v,v) = 1.05 \quad \text{and} \quad \bar{\gamma}(V,V) = 3.40 \quad [6.13]$$

Now it is easy to calculate the dispersion variance $\sigma^2(v|V)$.

$$\sigma^2(v|V) = \bar{\gamma}(V,V) - \bar{\gamma}(v,v) = 2.35 \quad [6.14]$$

The corresponding standard deviation (the square root) is 1.53. Using $m \pm 2\sigma$ as an approximate confidence interval, the daily averages will usually be within about ±3 units (i.e. 2 x 1.53) of the weekly average. A more precise answer could be obtained by conditionally simulating the deposit and carrying out a mining scenario on the numerical model. See for example, Chica–Olmo and Laille (1984) and Deraisme and de Fouquet (1984). But this is outside the scope of this book.

6.7 Change of support: regularization

Let Z(x) denote a random function defined on a point support. Its average over a volume V defines a new random function $Z_v(x)$ with support V. It can be shown that the variogram of this new regularized variable is:

$$\gamma_v(h) = \bar{\gamma}(V, V_h) - \bar{\gamma}(V, V) \qquad [6.15]$$

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

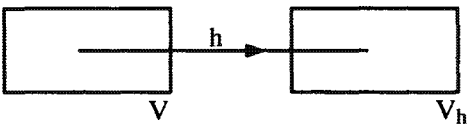


Fig. 6.4. Volume V translated through a vector h to volume V_h

When the distance h is small compared to the size of V, the distances from an arbitrary point in V to an arbitrary point in V_h can vary considerably. For example, if V is a rectangle of length l, then the horizontal distances go from h – l to h + l. However when the distance h is large compared to the size of V, the distances are very close to h. Consequently the mean variogram value $\bar{\gamma}(V,V_h)$ is approximately equal to $\gamma(h)$. So we obtain the relation:

$$\gamma_v(h) = \gamma(h) - \bar{\gamma}(V, V) \qquad [6.16]$$

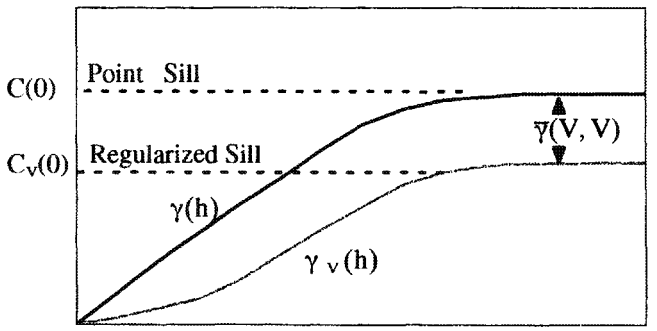


Fig. 6.5. Point support variogram and the regularized variogram

6.8 Exercise: calculating regularized variograms

The data available for reserve calculations do not always all have the same support size. For example, some drillholes may be 8" in diameter whereas others are 2" in

diameter. It is then important to be able to calculate the variograms corresponding to the two supports. As an exercise, we calculate the variogram models for 2m x 2m and 3m x 3m blocks given that $\bar{\gamma}(V,V) = 11,150$ for 2m x 2m blocks and $\bar{\gamma}(V,V) = 13,900$ for 3m x 3m blocks.

6.8.1 Solution

For 2m x 2m blocks

$$\gamma_R(h) = \gamma_{1x1}(h) - 11,150 \qquad [6.17]$$

For 3m x 3m plots

$$\gamma_R(h) = \gamma_{1x1}(h) - 13,900 \qquad [6.18]$$

Consequently the theoretical sills of the regularized variograms are 16,450 and 13,700 respectively. Fig. 6.6. shows the three experimental variograms. Their sills are in close agreement with those calculated theoretically.

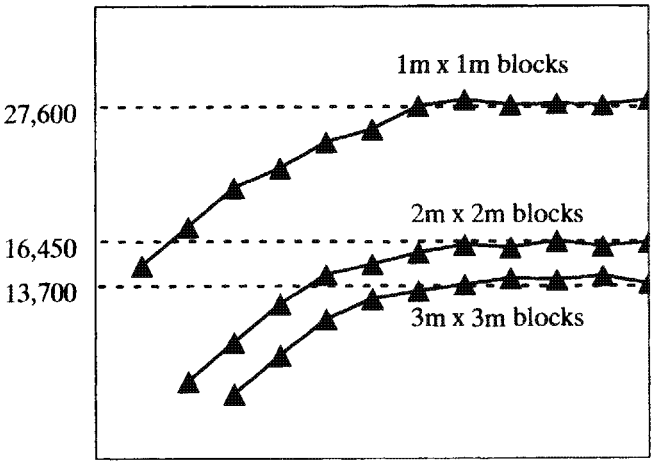
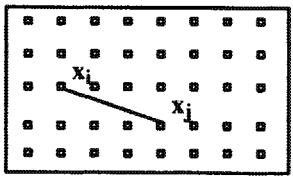


Fig. 6.6. Experimental variograms for all three support sizes

6.9 Exercises

To do these exercises, one has to be able to calculate the values of $\bar{\gamma}(V,V)$. There are two ways of doing this: by writing a small computer program to discretize V or by using standardized tables such as those given by Journel and Huijbregts (1978) pp125–147. The first exercise outlines a suitable computing procedure while the second one shows how to use the tables.

Ex 6.1 Calculating $\bar{\gamma}(V,V)$ by discretizing the block. This exercise outlines a procedure for calculating the average variogram value by discretizing the block V. First the number of grid nodes along each axis has to be chosen because this determines the coordinates of discretized points.



The procedure consists of a double do–loop which takes every pair of points in turn, calculates the vector distance between them, evaluates the corresponding variogram value and sums these. At the end the total is divided by the square of the number of discretized grid nodes to obtain the average variogram value.

The choice of the number of grid nodes is critical. Too few nodes and the average will not be accurate; too many nodes and the computing time explodes. For example, if a square is discretized into 100 x 100 points; there are 10^4 points and 10^8 variogram terms to calculate. Typically using between 25 and 100 points suffices in 2D, and up to 200–300 in 3D.

Write a computer routine to calculate $\bar{\gamma}(V,V)$ and $\bar{\gamma}(v,v)$ where V is a 100m x 100m block and v is a 10m x 10m block and where the variogram is
(a) a spherical with a range of 100m and a sill of 3.0,
(b) an exponential with a practical range of 50m and a sill of 3.0,
(c) a pure nugget effect with a sill of 3.0.

Start the calculations using a 2x2 discretization and work up to 10x10 noting the stabilization in the value of $\bar{\gamma}(V,V)$ and the increase in the computer time.

Ex 6.2 Calculating $\bar{\gamma}(V,V)$ by using the tables. This exercise is designed to illustrate the use of standardized tables that can be found in texts like Journel and Huijbregts (1978). Charts 4 and 5 (p128–9) give the standardized values of $\bar{\gamma}(V,V)$ for the spherical with unit range and unit sill, for 2D and 3D blocks respectively. Similarly for Charts 14 and 15 (p138–9) for the exponential. The key to using these charts is to convert the block size into multiples of the range (or the scale parameter "a" for the exponential) and then read the values off the table.

For example, if the block is 100m x 100m and the variogram is a spherical with a range of 200m and a sill of 3.0, then the block is equivalent to 0.5 x 0.5 ranges. The chart gives a value of 0.375. Multiplying by the sill gives

$\bar{\gamma}(V,V) = 3.0 \times 0.375 = 1.125$

Use the charts to calculate $\bar{\gamma}(V,V)$ and $\bar{\gamma}(v,v)$ where V is a 100m x 100m block and v is a 10m x 10m block and where the variogram is
(a) a spherical with a range of 100m and a sill of 3.0,
(b) an exponential with a practical range of 50m and a sill of 3.0,
(c) a pure nugget effect with a sill of 3.0. (Hint : range = 0).

Compare your results with those obtained in the previous exercise.

Ex 6.3 Calculating $\bar{\gamma}(V,V)$ theoretically. In some very simple cases $\bar{\gamma}(V,V)$ can be calculated by integration, using equation [6.4]. Let V be a core section of length d (in 1D).
(a) If the variogram is linear with slope C, show that $\bar{\gamma}(d,d) = C\,d/3$.
Remember that $\gamma(h) = C|h|$.
(b) Calculate $\bar{\gamma}(d,d)$ for the case where the variogram is an exponential with a sill of C and a practical range of 3.
(c) Calculate $\bar{\gamma}(d,d)$ for the case where the variogram is a pure nugget effect with a sill of C.

Ex 6.4 Dispersion variance. Use the values of $\bar{\gamma}(V,V)$ and $\bar{\gamma}(v,v)$ found above to calculate the dispersion variance of 10m x 10m blocks inside a zone of size 100m x 100m for the cases where the variogram is
(a) a spherical with a range of 100m and a sill of 3.0,
(b) an exponential with a practical range of 50m and a sill of 3.0,
(c) a pure nugget effect with a sill of 3.0. (Hint : range = 0).

Ex 6.5 Dispersion variance. A small mining company extracts 4 blocks of size 10m x 10m x 5m from its open pit each day. The mine manager can choose to take adjoining blocks (i.e. 40m x 10m x 5m) or blocks from four different parts of the mine which are far enough apart to be considered independent. The problem is to predict the daily variance of the mean grades for the two methods. The material being mined has a spherical variogram with a range of 100m and a sill of 3.0.

Ex 6.6 Regularized variograms. The copper grade for 5m long core sections has a spherical variogram with a vertical range of 50m, a sill of 0.1 and a nugget effect of 0.05. As the bench height during mining will be 15m, the data will be regularized over this height. We want to calculate the vertical variogram for this new variable. Firstly calculate $\bar{\gamma}(d,d)$ where d is a 15m core section, then sketch the variograms for 5m and 15m core sections on the same graph.

7 The Theory of Kriging

7.1 Summary

This chapter presents the theory of kriging. Kriging is an estimation method that gives the best unbiased linear estimates of point values or of block averages. Here "best" means minimum variance. Three types of kriging estimators are discussed: ordinary kriging (OK) used when the mean is unknown, kriging the unknown mean value and simple kriging (SK) used when the mean is known.

The equations for these three estimators are derived for the stationary case, and are extended to the case of intrinsic variables for ordinary kriging. The additivity theorem which gives the links between the OK and SK estimators is proved. For ordinary kriging, the formula for the slope of the linear regression of the true grade on its estimate is given, and its importance in relation to conditional unbiasedness is discussed. Lastly, kriging is shown to be an exact interpolator.

7.2 The purpose of kriging

Sampling provides accurate information at 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. The accuracy of the estimates depends on several factors:

1. the number of samples and the quality of the data at each point.
2. the positions of the samples within the deposit. Evenly spaced samples achieve better coverage and thus give more information about the deposit than clustered samples do.
3. the distance between the samples and the point or block to be estimated. It is natural to rely more heavily on neighbouring samples, rather than on more distant ones. Similarly we expect the accuracy to be best in the vicinity of the samples

and to deteriorate further away. The dangers of extrapolating outside the zone sampled need not be spelled out here.

4. the spatial continuity of the variable under consideration. It is easier to estimate the value of a fairly regular variable than one which fluctuates wildly. For example, for a given sample layout, the estimates of the copper grade are more precise than those of gold.

"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 first developed a moving average technique for estimating gold grades to remove the regression effect. Prof. G. Matheron improved on this and the new method was called kriging. In essence, it is a way of finding the best linear unbiased estimator (in the sense of least variance). That is, we choose the weighted average of the sample values which has the minimum variance.

7.3 Deriving the kriging equations

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

$$z_V = \frac{1}{V} \int_V z(x)\, dx \qquad [7.1]$$

The volume V could be the whole deposit, or a mining block, or it could be as small as a single point in the case of point estimation. It could even be an irregular shape. See Box 5 for more information on kriging irregular shapes. To estimate Z(V), we consider a weighted average of the data:

$$z^*_V = \sum \lambda_i z(x_i) \qquad [7.2]$$

where λ_i are the weighting factors. By convention the star will be used to denote the estimated value as opposed to the real but unknown value. The problem is to choose the weighting factors in the best way. This is where we make use of the geostatistical model. We consider the regionalized variable:

$$Z^*_V = \sum \lambda_i Z(x_i) \qquad [7.3]$$

The weights are chosen so that the estimator is:

1. unbiased: $E[Z^*_V - Z_V] = 0$

2. minimum variance: $\mathrm{Var}[Z^{*}_{v} - Z_{v}]$ is a minimum.

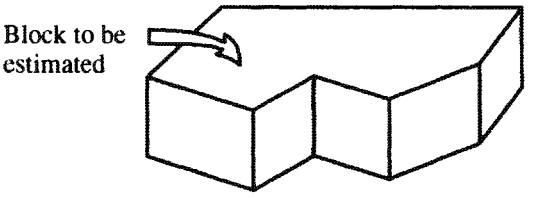
This variance will be called the kriging variance.

7.4 Different kriging estimators

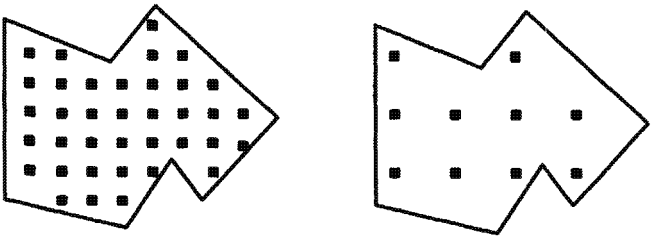
In the first instance we assume that the regionalized variable $Z(x)$ is stationary and that its mean, m, is unknown. Kriging with an unknown mean is called ordinary kriging, which is abbreviated to OK. We first derive the system of equations for ordinary kriging for the stationary case in terms of the variogram and then the covariance, before indicating how to extend these results to the intrinsic case.

BOX No 5 : Can irregular blocks be kriged?

Some people think that only regular blocks can be kriged. But this is completely incorrect. The kriging equations are quite general. The target "V" could be as small as a point or as large as the whole deposit. Most often it is a regularly shaped block, but it could be an irregular shape such as an area marked out for blasting.



The only problems arise when discretizing V in order to calculate $\bar{\gamma}(V,V)$ and $\bar{\gamma}(x,V)$. With a regular shape it is easy to choose a grid size that guarantees a reasonable number of discretized points inside the zone to be estimated. This is more difficult with irregular shapes. As can be seen from the two figures below, a slight change in the grid spacing dramatically changes the number of nodes inside the zone.



The next step is to see how to estimate the unknown mean m. After that we will see what happens to the kriging estimator if the mean, m, is known. This is called simple kriging and is abbreviated to SK. In all these cases a set of linear equations called the kriging system has to be solved to obtain the kriging weights and the kriging variance.

7.5 Ordinary kriging

Unbiasedness. The variable Z(x) is assumed to be stationary with mean m. Its mean at every point is equal to m and so is the mean of any block. That is,

$$E[Z(x)] = m = E[Z_V] \qquad [7.4]$$

Most estimators are weighted moving averages of the surrounding data values; that is, they are linear combinations of the data:

$$Z_V^* = \sum \lambda_i Z(x_i) \qquad [7.5]$$

The mean of the estimation error $[Z_V^* - Z_V]$ is just

$$E\left[\sum \lambda_i Z(x_i) - Z_V\right] = \sum \lambda_i m - m = m\left[\sum \lambda_i - 1\right] \qquad [7.6]$$

In order to be unbiased, the expected error must be zero, so either m = 0 or the kriging weights must add up to 1. In the first case the mean is known. (This leads to simple kriging). If m is unknown then the weights must sum to 1.[2]

Minimum variance. The variance of the error $[Z_V^* - Z_V]$ can be expressed in terms of either the covariance or the variogram:

$$\sigma^2 = \sum\sum \lambda_i \lambda_j C(x_i, x_j) + \bar{C}(V, V) - 2 \sum \lambda_i \bar{C}(x_i, V)$$

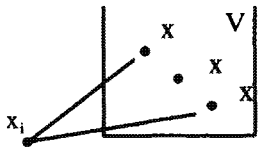
$$= 2 \sum \lambda_i \bar{\gamma}(x_i, V) - \sum\sum \lambda_i \lambda_j \gamma(x_i, x_j) - \bar{\gamma}(V, V) \qquad [7.7]$$

where $\bar{\gamma}(x_i, V)$ is the average of the variogram between x_i and the volume V, i.e.

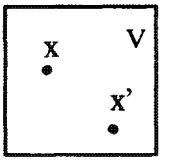
$$\bar{\gamma}(x_i, V) = \frac{1}{V} \int_V \gamma(x_i - x)\, dx$$

---

[2] The idea that the kriging weights must sum to 1 rather than 0 sometimes causes confusion. The sum of all the weights is still zero because there is the weight of −1 in front of Z(V) in the expression for the estimation error.



As was seen in Chapter 6, $\bar{\gamma}(V,V)$ is the average of the variogram between any two points x and x' sweeping independently throughout the volume V.



$$\bar{\gamma}(V,V) = \frac{1}{V^2} \int\int \gamma(x - x')\, dx\, dx'$$

In the same way $\bar{C}(x_i,V)$ and $\bar{C}(V,V)$ are the averages for the covariance. In order to minimize the estimation variance under the constraint that the sum of the kriging weights must be equal to 1, we introduce a Lagrange multiplier μ into the expression to be minimized. Since the sum of the weights must be 1.0, adding the term in μ does not change the value of the expression.

$$\phi = \mathrm{Var}(Z^*_v - Z_v) - 2\mu \left(\sum \lambda_i - 1 \right) \qquad [7.8]$$

The partial derivatives of the quantity are then set to zero. This leads to a set of N+1 linear equations called the kriging system. Box 6 shows the details of the differentiation step. When written in terms of the variogram model the kriging system is:

$$\begin{aligned} \sum_{j=1}^{N} \lambda_j\, \gamma(x_i,x_j) + \mu &= \bar{\gamma}(x_i,V) \qquad i = 1, 2, \ldots N \\ \sum_{i}^{N} \lambda_i &= 1 \end{aligned} \qquad [7.9]$$

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

$$\sigma^2_K = \sum \lambda_i\, \bar{\gamma}(x_i,V) - \bar{\gamma}(V,V) + \mu \qquad [7.10]$$

Clearly the equations could also have been obtained in terms of the covariance by minimizing the first form of [7.7]. The kriging system is then:

$$\begin{aligned} \sum_{j=1} \lambda_j\, C(x_i,x_j) + \mu' &= \bar{C}(x_i,V) \qquad i = 1, 2, \ldots N \\ \sum_{i} \lambda_i &= 1 \end{aligned} \qquad [7.11]$$

BOX No 6 : Deriving the ordinary kriging equations

The esssential step in deriving the kriging equations is minimizing the expression for the estimation variance:

$$\phi = 2\sum \lambda_i \, \bar{\gamma}(x_i, V) - \sum\sum \lambda_i \, \lambda_j \, \gamma(x_i, x_j) - \bar{\gamma}(V, V) + 2\mu(1 - \sum \lambda_i)$$

This is done by differentiating with respect to each of the unknowns and setting the partial derivatives to zero. Here we show this in detail for the case where there are 3 samples. The procedure is the same in the general case where there are N samples. If we let $\gamma_{ij} = \gamma(x_i, x_j)$ and $\gamma_{iV} = \bar{\gamma}(x_i, V)$, then

$$\begin{aligned}\phi = {} & 2\lambda_1 \bar{\gamma}_{1V} + 2\lambda_2 \bar{\gamma}_{2V} + 2\lambda_3 \bar{\gamma}_{3V} - (\lambda_1^2 \gamma_{11} + \lambda_2^2 \gamma_{22} + \lambda_3^2 \gamma_{33} \\ & + 2\lambda_1\lambda_2\gamma_{12} + 2\lambda_1\lambda_3\gamma_{13} + 2\lambda_2\lambda_3\gamma_{23}) - \bar{\gamma}(V, V) \\ & + 2\mu \, (1 - \lambda_1 - \lambda_2 - \lambda_3)\end{aligned}$$

Differentiating with respect to λ_1 gives

$$\frac{\partial\phi}{\partial\lambda_1} = 2\,\bar{\gamma}_{1V} - [\, 2\lambda_1 \gamma_{11} + 2\lambda_2 \gamma_{12} + 2\lambda_3 \gamma_{13} + 2\mu \,] = 0$$

Hence

$$\lambda_1 \gamma_{11} + \lambda_2 \gamma_{12} + \lambda_3 \gamma_{13} + \mu = \bar{\gamma}_{1V}$$

Similarly differentiating with respect to λ_2 and λ_3 gives:

$$\lambda_1 \gamma_{12} + \lambda_2 \gamma_{22} + \lambda_3 \gamma_{23} + \mu = \bar{\gamma}_{2V}$$

$$\lambda_1 \gamma_{13} + \lambda_2 \gamma_{23} + \lambda_3 \gamma_{33} + \mu = \bar{\gamma}_{3V}$$

Lastly differentiating with respect to μ gives

$$\lambda_1 + \lambda_2 + \lambda_3 = 1$$

Consequently the kriging system is:

$$\sum_{j=1}^{3} \lambda_j \, \gamma(x_i, x_j) + \mu = \bar{\gamma}(x_i, V) \qquad i = 1, 2, 3$$

$$\sum_{i}^{3} \lambda_i = 1$$

In the general case, one would have to differentiate with respect to each of the N unknown weights, and the sums in the kriging system would go from 1 to N rather than from 1 to 3. Otherwise the principles are just the same.

The two Lagrange multipliers are related by $\mu' = -\mu$. The corresponding kriging variance is given by:

$$\sigma_K^2 = \bar{C}(V,V) - \mu' - \sum \lambda_i \bar{C}(x_i, V) \qquad [7.12]$$

To solve the system, it is convenient to write it in matrix form: AX = B.

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{1N} & 1 \\ \gamma_{21} & \gamma_{22} & \gamma_{2N} & 1 \\ & & & \\ \gamma_{N1} & \gamma_{N2} & \gamma_{NN} & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \\ \lambda_N \\ \mu \end{bmatrix} = \begin{bmatrix} \bar{\gamma}(x_1, V) \\ \bar{\gamma}(x_2, V) \\ \\ \bar{\gamma}(x_N, V) \\ 1 \end{bmatrix} \qquad [7.13]$$

If γ is an admissible model and if there are no multiple points, the matrix A is always non singular. Its inverse A^{-1} exists. So a solution exists and it can be proved that it is unique. The uniqueness is important because it is used later to link the different types of kriging. The kriging variance can be written:

$$\sigma_K^2 = X^T B - \bar{\gamma}(V,V) \quad (X^T = X \text{ transposed}) \qquad [7.14]$$

Beware the matrix A itself is not positive definite.

7.6 The OK equations for intrinsic regionalized variables

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

$$E[Z(x+h) - Z(x)] = 0 \qquad [7.15]$$

$$\text{Var}[Z(x+h) - Z(x)] = 2\gamma(h) \qquad [7.16]$$

where $\gamma(h)$ depends on h but not on x. So under this hypothesis the estimation error $[Z_V^* - Z_V]$ 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 and is minimized. This leads to the same OK system in terms of the variogram as before. This is one of the reasons why we use the intrinsic hypothesis rather than just stationarity.

7.7 Exercise: Ordinary kriging of a block

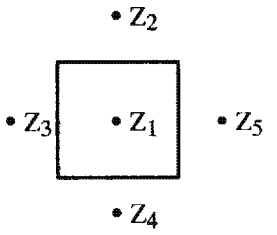


Fig. 7.1. Data configuration with the block to be estimated

The shaded block (200m x 200m) is to be kriged using 5 samples on a regular 200m grid. Suppose that the regionalized variable is stationary with an isotropic spherical variogram with a sill of 2.0 and a range of 250m. To make it possible to do the calculations with a pocket calculator, the values of $\bar{\gamma}(V,V)$ and $\bar{\gamma}(V,x)$ are given.

$$\bar{\gamma}(x_1, V) = 0.88 \quad \bar{\gamma}(x_2, V) = 1.86 \quad \bar{\gamma}(V, V) = 1.13 \qquad [7.17]$$

7.7.1 Solution

The first step is to write down the kriging system. As there are 5 samples, it is 6 x 6 system.

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} & \gamma_{15} & 1 \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} & \gamma_{25} & 1 \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} & \gamma_{35} & 1 \\ \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} & \gamma_{45} & 1 \\ \gamma_{51} & \gamma_{52} & \gamma_{53} & \gamma_{54} & \gamma_{55} & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma(V,x_1) \\ \gamma(V,x_2) \\ \gamma(V,x_3) \\ \gamma(V,x_4) \\ \gamma(V,x_5) \\ 1 \end{bmatrix} \qquad [7.18]$$

The terms in the matrix are found by calculating the distances between the points and then evaluating the corresponding variogram values. For example, for $\gamma(x_2, x_3)$, the distance between the points is $200\sqrt{2}$. As this is greater than the range, the value equals the sill. The resulting system is:

$$\begin{bmatrix} 0 & 1.89 & 1.89 & 1.89 & 1.89 & 1 \\ 1.89 & 0 & 2 & 2 & 2 & 1 \\ 1.89 & 2 & 0 & 2 & 2 & 1 \\ 1.89 & 2 & 2 & 0 & 2 & 1 \\ 1.89 & 2 & 2 & 2 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \mu \end{bmatrix} = \begin{bmatrix} 0.88 \\ 1.86 \\ 1.86 \\ 1.86 \\ 1.86 \\ 1 \end{bmatrix} \qquad [7.19]$$

This can easily be solved to give:

$$\lambda_1 = 0.60$$

$$\lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = 0.10 \qquad [7.20]$$

$$\mu = 0.12$$

So the estimate of the average value over the square is:

$$Z^* = 0.60\, Z_1 + 0.10\, (Z_2 + Z_3 + Z_4 + Z_5) \qquad [7.21]$$

The estimation variance is given by:

$$\sigma_K^2 = \sum \lambda_i\, \bar{\gamma}(V, x_i) + \mu - \bar{\gamma}(V, V) = 0.26 \qquad [7.22]$$

The only tricky point when setting up kriging matrices comes from the nugget effect. In the case above, if there had been a nugget effect of 1.5 in addition to the spherical structure, all the nondiagonal terms would be increased by 1.5 but the diagonal terms remain 0. Conversely when the system is written in terms of covariances, the diagonal terms equal the total sill including the nugget component but this is absent from nondiagonal terms.

7.8 Kriging the value of the mean

In ordinary kriging the objective was to estimate a linear function of the regionalized variable such as the grade at a point or the average grade over a block. Here the objective is to estimate the value of the unknown mean m. If we use the index m to distinguish the weights in this estimator from those in the previous one, then the estimator can be written as

$$m^* = \sum_{i=1}^{N} \lambda_{mi} Z(x_i) \qquad [7.23]$$

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,

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

As the mean of Z(x) is m, this implies that

$$\sum_{i} \lambda_{mi} = 1 \qquad [7.24]$$

The variance of the estimation error is

$$\mathrm{Var}\,[m^* - m] = \mathrm{Var}\left[\sum_{i=1}^{N} \lambda_{mi}\, Z(x_i) - m\right]$$

$$= \sum_{j} \sum_{i} \lambda_{mi}\, \lambda_{mj}\, C\,(x_i, x_j) \qquad [7.25]$$

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

$$\sum_{j=1} \lambda_{mj}\, C\,(x_i, x_j) = \mu_m \qquad i = 1, 2, \ldots N$$

$$\sum \lambda_{mi} = 1 \qquad [7.26]$$

The corresponding kriging variance can be calculated. Interestingly, this gives a meaning for the Lagrange multiplier in this case.

$$\sigma_K^{\,2} = \mathrm{Var}\,(m^*) = \mu_m \qquad [7.27]$$

7.9 Simple kriging

We are now going to derive the kriging system when the mean m of the regionalized variable is known. Firstly we consider a regionalized variable Y(x) with zero mean. Clearly the initial regionalized variable is obtained from this as Z(x) = Y(x) + m. Our estimator of Y(x) is going to be

$$Y_V^* = \sum_{i=1}^{N} \lambda_i' \, Y(x_i) \qquad [7.28]$$

We use primes to distinguish the simple kriging weights from the ones for ordinary kriging and from those for kriging the mean. 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,

$$E\,[Y_V^* - Y_V] = E\left[\sum_{i=1}^{N} \lambda_i' \, Y(x_i) - Y_V\right] = 0 \qquad [7.29]$$

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

$$\text{Var}\,[Y_V^* - Y_V] = E\,[\sum \lambda_i' \, Y(x_i) - Y_V]^2$$

$$= \sum_j \sum_i \lambda_i' \, \lambda_j' \, C(x_i, x_j) + \bar{C}(V,V) - 2\sum_i \lambda_i' \, \bar{C}(x_i, V) \qquad [7.30]$$

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

$$\boxed{\sum_{j=1} \lambda_j' \, C(x_i, x_j) = \bar{C}(x_i, V) \quad i = 1, 2, \ldots N} \qquad [7.31]$$

The corresponding kriging variance, is given by:

$$\boxed{\sigma_{SK}^2 = \bar{C}(V,V) - \sum \lambda_i' \, \bar{C}(x_i, V)} \qquad [7.32]$$

Solving the kriging system [7.32] gives the kriging weights and hence the estimator of Y_V. The estimator of Z_V can be deduced from this by replacing Y(x) by Z(x) – m. This gives

$$Z_V^* = Y_V^* + m = \sum \lambda'_i \, [Z(x_i) - m] + m$$

$$= \sum \lambda_i' \, Z(x_i) + m\left[1 - \sum \lambda_i'\right] = \sum \lambda_i' \, Z(x_i) + m \, \lambda_M \qquad [7.33]$$

The term λ_M is called the weight of the mean in simple kriging.

Simple kriging is rarely used in day to day practical applications because the mean is rarely known. It is sometimes used in large mines such as in South Africa where the mean of each area is known because the region has been mined for many years. It is also used when kriging transformed data (e.g. after a gaussian anamorphosis) when the mean has been set by the transformation to a known value, usually zero; for example, in disjunctive kriging. But one of the most important reasons for studying simple kriging is that the weight of the mean provides one of the best criteria for testing the quality of a kriging. More information on these quality criteria will be given in Chapter 8.

Looking at the estimator [7.33] it is clear that the form of the estimator has changed. Compared to ordinary kriging and kriging the mean, it is no longer just a linear combination of the data. A constant term has been added. This is important when kriging is considered in terms of projections (Journel, 1977).

7.10 The additivity theorem

In the preceding sections we saw how to estimate variables when the mean was known (simple kriging) and when it was unknown (ordinary kriging). We also saw how to estimate the value of the mean in the second case. It is interesting to see how these three estimators are linked. It turns out that substituting the kriged estimator for the mean m into the expression for the SK estimator gives the OK estimator. The proof is given in Box 7. As part of the proof two interesting results appear. These are

$$\lambda_M \mu_m = \mu \qquad [7.34]$$

$$\sigma^2_{OK} = \sigma^2_{SK} + (\lambda_M)^2 \text{ Var } (m^*) \qquad [7.35]$$

The first of these provides an interpretation of the Lagrange multiplier for OK in terms of the weight of the mean in SK and the Lagrange multiplier for kriging the mean. The second equation shows that the ordinary kriging variance can be split into two parts: the first is the simple kriging variance when the mean is known, the second is the variance of the estimator of the mean multiplied by the square of the weighting factor of the mean in simple kriging. The second term gives a measure of the loss of accuracy caused by not knowing the true mean.

BOX No 7 : Proof of the additivity theorem

Starting out from the SK estimator

$$Z_V^* = \sum \lambda_i' Z(x_i) + m \lambda_M$$

we replace m by its kriged estimator m*. If λ_M denotes the weight of the mean in simple kriging

$$Z_V^* = \sum Z(x_i) [\lambda_i' + \lambda_M \lambda_{mi}] \qquad [7.36]$$

At first this does not look like the OK estimator. But as the OK estimator is unique, if this satisfies the OK equations, it must just be another expression for it. We now show that [7.36] does satisfy these equations. Firstly we check that the sum of the weights is 1.0. Summing these gives

$$\sum [\lambda_i' + \lambda_M \lambda_{mi}] = \sum \lambda_i' + \lambda_M = 1 \qquad [7.37]$$

since $\sum \lambda_{mi} = 1$. Now we show that the equations [7.11]

$$\sum_{i=1} \lambda_i C(x_i, x_j) = \overline{C}(x_j, V) - \mu' \quad j = 1, 2, ..N \qquad [7.38]$$

are also satisfied. After substituting $\lambda'_i + \lambda_M \lambda_{mi}$ in place of λ_i the first term becomes:

$$\sum (\lambda_i' + \lambda_M \lambda_{mi}) C(x_i, x_j)$$

$$= \sum \lambda'_i C(x_i, x_j) + \sum \lambda_M \lambda_{mi} C(x_i, x_j) \qquad [7.39]$$

From the SK system, the first term is $\overline{C}(x_j, V)$. Similarly from the kriging system for the mean [7.26] the other term is $\lambda_M \mu_m$. Consequently

$$\sum_{i=1} [\lambda_i' + \lambda_M \lambda_{mi}] C(x_i, x_j) = \overline{C}(x_j, V) + \lambda_M \mu_m \qquad [7.40]$$

So by setting $\lambda_M \mu_m = -\mu' = \mu$, it is clear that these weighting factors do satisfy the equations. So expression [7.36] satisfies all the OK equations. Lastly by substituting $\lambda_i = \lambda_i' + \lambda_M \lambda_{mi}$ into the expression for the kriging variance, a new expression is obtained for the OK kriging variance:

$$\sigma^2_{OK} = \sigma^2_{SK} + (\lambda_M)^2 \text{ Var } (m^*) \qquad [7.41]$$

7.11 Slope of the linear regression

In the introductory exercise at the beginning of the book, the true block grades were plotted against the estimated grades for several estimation methods including the polygonal method and kriging. For a perfect estimator, Z_v^* would always equal Z_v but this is not possible in practice. The next best situation would be to have estimators that are conditionally unbiased; that is,

$$E\,[Z_v \mid Z_v^*] = Z_v^* \qquad [7.42]$$

This means that the regression of Z_v on Z_v^* must be linear with a slope of 1.0.

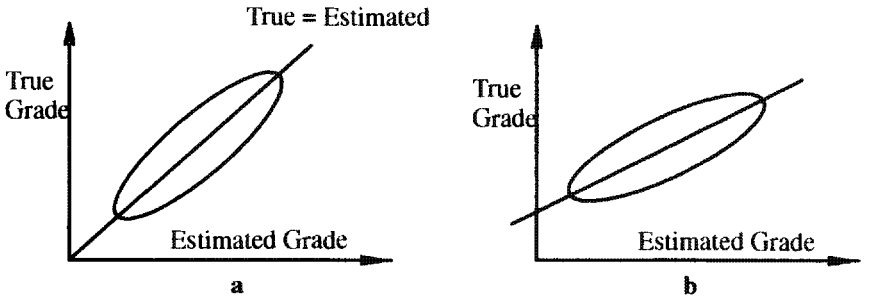


Fig. 7.2. Regressions of the true grades against the estimated ones, (**a**) conditionally unbiased and (**b**) conditionally biased

It is important to note that although kriging is by definition globally unbiased since $E[Z_v^* - Z_v] = 0$, it is not necessarily conditionally unbiased. In this section we will see that assuming that the regression is linear, simple kriging is always conditionally unbiased but ordinary kriging is not.

The slope of the linear regression of Z_v on Z_v^* will be calculated for the OK estimator. In practical cases the distributions of Z_v and Z_v^* are rarely known; so the true shape of the curve $E[Z_v \mid Z_v^*]$ considered as a function of Z_v^* is unknown. The linear regression slope can nevertheless be used to see how far the OK estimator is from conditional unbiasedness. It is wellknown that the slope, p, of the linear regression is given by

$$p = \text{Cov}\,[Z_v,\ Z_v^*] \,/\, \text{Var}\,[Z_v^*] \qquad [7.43]$$

For simple kriging,

$$Z_v^* = \sum \lambda'_i\, Z(x_i) + m\,[1 - \sum \lambda'_i] \qquad [7.44]$$

and consequently

$$\text{Cov}\ [Z_v,\ Z_v^*] = \sum \lambda'_j\ \overline{C}\ (x_j, V)$$

$$\text{Var}\ [Z_v^*] = \sum\sum \lambda'_i\ \lambda'_j\ C(x_i, x_j) \qquad [7.45]$$

From the SK equations, these two terms are equal. Consequently the slope is 1.0.[3] Similarly for ordinary kriging

$$\text{Cov}\ [\ Z_v,\ Z_v^*\ ] = \sum \lambda_i\ \overline{C}(x_i, V)$$

$$\text{Var}\ [\ Z_v^*\ ] = \sum\sum \lambda_i\ \lambda_j\ C(x_i, x_j) \qquad [7.46]$$

$$\text{But as } \sum \lambda_j\ \overline{C}(x_i, V) = \sum\sum \lambda_i\ \lambda_j\ C(x_i, x_j) + \mu$$

$$\text{Cov}\ [Z_v,\ Z_v^*] + \mu = \text{Var}\ [Z_v^*] \qquad [7.47]$$

and hence the slope p of the linear regression of Z_v on Z_v^* is given by

$$p = \frac{\text{Cov}\ [\ Z_v,\ Z_v^*\ ]}{(\text{Cov}\ [\ Z_v,\ Z_v^*\ ] - \mu)} \qquad [7.48]$$

Here the value of the Lagrange parameter has been calculated from the kriging system written in terms of covariances. The sign reverses if the variogram form of the equations is used. In general the slope is less than 1.0. This result concerning the slope of the linear regression of the true grade on the estimate will be used to guide us in the next chapter in choosing how large a kriging neighbourhood to use.

7.12 Kriging is an exact interpolator

When some estimation methods (e.g. trend surfaces) are used to estimate the value of a regionalized variable at a data point, the resulting estimate is not necessarily equal to the sample value. Methods which always return the sample value as the estimate at sample points are said to be exact estimators. The simplest way to show that kriging is an exact interpolator is via an example.

Continuing the OK example in which 5 data points were used to krige a 200m x 200m block, suppose we now want to estimate the value at one of the sample points (say the central one) from the available data including that point. It is easy to see that the matrix on the left hand side is exactly the same as before. Only the vector on the right hand side changes. The new system is

---

[3] This also shows that the kriging error $Z_V^* - Z_V$ is orthogonal to the estimator Z_V^*. This result is needed later in order to condition simulations. Secondly it is important when kriging is considered in terms of projections.

$$\begin{bmatrix} 0 & 1.89 & 1.89 & 1.89 & 1.89 & 1 \\ 1.89 & 0 & 2 & 2 & 2 & 1 \\ 1.89 & 2 & 0 & 2 & 2 & 1 \\ 1.89 & 2 & 2 & 0 & 2 & 1 \\ 1.89 & 2 & 2 & 2 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \mu \end{bmatrix} = \begin{bmatrix} 0 \\ 1.89 \\ 1.89 \\ 1.89 \\ 1.89 \\ 1 \end{bmatrix} \quad [7.49]$$

This can easily be solved to give:

$$\lambda_1 = 1.0$$

$$\lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = 0 \quad [7.50]$$

$$\mu = 0$$

The corresponding kriging variance is zero. This result should not be surprising. It is intuitively clear that the estimate that minimizes the estimation variance is just the sample value itself. This property relies on the fact that the terms in the first row and column are equal to the corresponding terms in the right hand vector.

If a constant (e.g. a nugget effect representing measurement error) had been added to all the variogram terms in the matrix and in the vector, it could be filtered out of them all by using the row of 1s in the last row of the matrix.

Sometimes people confuse this property of the kriging estimator with the variogram crossvalidation procedure. Please note that in the case considered here the point to be estimated is included in the data set, whereas when kriging is used to crossvalidate the variogram model, the data point of interest is dropped out of the data set while its value is re–estimated. The crossvalidation technique is discussed in more detail in the next chapter.

7.13 Geometric exercise showing the minimization procedure

The aim of this exercise is to provide a geometric illustration of the relationship between ordinary and simple kriging. To keep things simple, we assume that only two samples are available in order to krige the grade of a block V. The reason for taking only two samples is that it is easy to plot functions of two unknowns. For simplicity, the regionalized variable representing the grade is considered to be stationary and its covariance C(h) has a unit sill.

7.13.1 Quadratic form to be minimized

The weights for any type of kriging are obtained by minimizing the estimation variance under the appropriate conditions. As estimation variances are quadratic functions of the weights, they are called *quadratic forms*. The general estimation variance that has to be minimized when there are two samples, is

$$\sigma^2_{EST} = \lambda_1^2\, C(1,1) + \lambda_2^2\, C(2,2) + 2\lambda_1\, \lambda_2\, C(1,2) + \overline{C}(V,V) - 2\lambda_1\, \overline{C}(1,V) - 2\lambda_2\, \overline{C}(2,V) \quad [7.51]$$

where $C(1,1) = C(2,2) = 1$. This can be written as

$$\sigma^2_{EST} = (\lambda_1 - a)^2 + (\lambda_2 - b)^2 + 2c\,(\lambda_1 - a)\,(\lambda_2 - b) + d \quad [7.52]$$

where

$$a = \frac{\overline{C}(1,V) - C(1,2)\,\overline{C}(2,V)}{1 - \overline{C}(1,2)^2} \quad [7.53]$$

$$b = \frac{\overline{C}(2,V) - C(1,2)\,\overline{C}(1,V)}{1 - \overline{C}(1,2)^2} \quad [7.54]$$

$$c = C(1,2) \quad [7.55]$$

and where d is a suitably chosen value. Fig. 7.3. shows two ways of visualizing the estimation variance as a function of the weights. The 3D graphic shows that it is a basin shaped surface. From the equation, the minimum obviously occurs when the weights take the values a and b respectively, and then its value is d. Readers can solve the SK systems to check that a and b are just the SK weights and that d is the SK variance. The iso–variance contour lines (on the left) are a projection of the "basin".

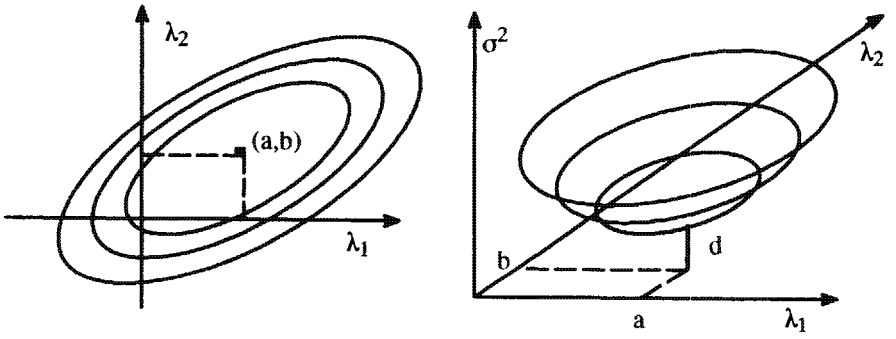


Fig. 7.3. On the right, a 3D representation of the estimation variance as a function of the weights; on the left, the corresponding isovariance contour lines

Fig. 7.3. On the right, a 3D representation of the estimation variance as a function of the weights; on the left, the corresponding isovariance contour lines

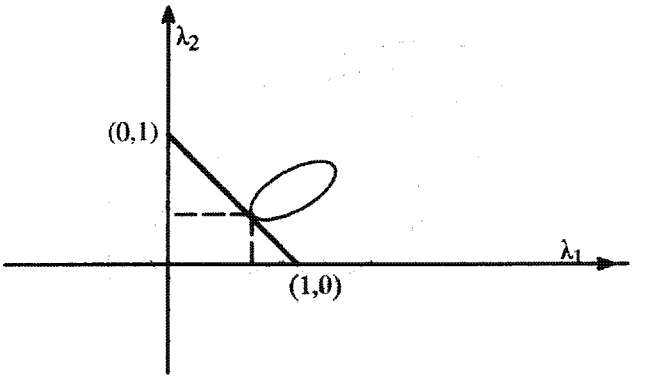


Fig. 7.4. Line representing the OK constraint $\lambda_1 + \lambda_2 = 1$ has been superimposed on the isovariance contour lines. The coordinates of the point where it tangents one of the ellipses are the OK weights.

Having visualized the minimization for simple kriging we now go on to ordinary kriging. Because of the unbiasedness condition the weights must now sum to 1, i.e. $\lambda_1 + \lambda_2 = 1$. Fig. 7.4. shows the line $\lambda_2 = 1 - \lambda_1$ joining the points (0,1) and (1,0) which represents this constraint. The minimum occurs at the point where this line tangents one of the ellipses. The corresponding values of the weights are just the OK weights, and the height at that point is the OK variance. As this is not usually the bottom of the basin, the height at that point is higher than the bottom. In other words, the OK variance is equal to or greater than the SK variance. This can also be seen via the equation. Substituting $\lambda_2 = 1 - \lambda_1$ into the quadratic form gives

$$2\lambda_1^2 (1 - c) - 2\lambda_1 (a - b + 1 - c) + d - 2 \qquad [7.56]$$

This can be rewritten as

$$\phi = 2 (1 - c)\left[\lambda_1 - \left(\frac{a - b}{2 (1 - c)} + 0.5\right)\right]^2 + K \qquad [7.57]$$

From this, it is clear that the minimum estimation variance occurs when the first weight is:

$$\lambda_1 = \left(\frac{a - b}{2 (1 - c)} + 0.5\right) \qquad [7.58]$$

Readers can check that this is just the first OK weight written in another form.

7.14 Exercises

Ex 7.1 Pure nugget effect. Suppose that we have samples at two points (1,0) and (2,0) and want to estimate the value of the regionalized variable at an arbitrary point (x,y). The variogram is a pure nugget effect with a sill of 1.0. Calculate the ordinary kriging weights and the kriging variance. Deduce what the weights would be if there were N samples.

The weights would all be equal. So the kriged estimate would just be the arithmetic average of the sample values. Kriging cannot provide any more detail because even adjacent points are uncorrelated. The data are only used to estimate the overall mean of the regionalized variable.

What would happen in this case if the value of the mean was known and simple kriging was used?

Ex 7.2 Spherical variogram. As before we have two samples at the points (1,0) and (2,0) and we want to estimate the value at an arbitrary point (x,y). But this time the variogram is a spherical with a sill of 2.0 and a range of 0.75. Show that the ordinary kriging weights are given by

$$\lambda_1 = \frac{1}{2} + \frac{\gamma_{20} - \gamma_{10}}{2\gamma_{12}} \quad \text{and} \quad \lambda_2 = \frac{1}{2} + \frac{\gamma_{10} - \gamma_{20}}{2\gamma_{12}} \qquad [7.59]$$

where γ_{20} and γ_{10} denote the variogram values between one of the sample points and the target point and γ_{12} denotes the variogram value between the sample points.

Draw a circle of radius 0.75 (i.e. the range) around each sample point. Outside that zone there is no further correlation between the target and the samples, and both weights are equal to 0.5. The results are the same as for a pure nugget effect. All that can be estimated is the overall mean.

Ex 7.3 Exponential variogram. As before there are two samples at the points (1,0) and (2,0) and we want to estimate the value at an arbitrary point (x,y). This time the variogram is an exponential with a sill of 2.0 and a scale parameter of 1. Calculate the ordinary kriging weights and the kriging variance. Repeat the calculation for the case where the sill equals 4, keeping the scale parameter the same. Although doubling the sill doubles the kriging variance, the kriging weights remain the same.

Ex 7.4 Linear variogram – Markovian lack of memory. As before there are two samples at the points (1,0) and (2,0) but the variogram is a linear with an arbitrary slope. The objective is to estimate the value of the regionalized variable at a point (x,0) lying along the x axis. Show that if the target point is to the left of the first point, (1,0), its ordinary kriging weight is 1.0 and the other one is zero, and conversely to the right of the second one.

This can be extended to the case where there are many samples in a line. If the target point lies to the left of the first point, the weight of the first point is 1 and the others are zero. Similarly to the right of the last one. Only the closest point has a nonzero weight. It is as if the others were "forgotten". All the available information is

condensed into the closest sample. See for yourself what happens if the target point lies in between the samples. Note: these effects only happen in 1D.

This property is called the Markovian property. It is well known in time series, particularly in finance. For example, the last quoted share price is considered to reflect all previous information. The same type of effect is used in option pricing in the Black & Scholes model.

Ex 7.5 Factorizable variograms – perpendicular screen effect. As before there are two samples at the points (1,0) and (2,0). But this time, the variogram is a gaussian with a unit sill and a unit scale parameter. The objective is to estimate the value of a regionalized variable at a point (1,y) lying vertically above or below the first sample. Show that the simple kriging weight of the second point is zero. Only the point directly below/above the target gets any weight. This result depends on the fact that the gaussian can be factorized into two components.

$$C(h) = \exp\left[-\frac{x^2}{a^2}\right] \times \exp\left[-\frac{y^2}{a^2}\right] \quad \text{where} \quad h^2 = x^2 + y^2 \qquad [7.60]$$

Show that the result is also true for the factorized exponential covariance:

$$C(h) = \exp\left[-\frac{|x|}{a}\right] \times \exp\left[-\frac{|y|}{a}\right] \quad \text{where} \quad h^2 = x^2 + y^2 \qquad [7.61]$$

8 Practical Aspects of Kriging

8.1 Summary

This chapter is designed to give an overview of the practical aspects of kriging: negative kriging weights, the impact of the choice of the variogram model on the kriging weights, crossvalidation, the screen effect and last but not least, some criteria for testing the quality of a kriging.

8.2 Introduction

This chapter concentrates on the practical aspects of kriging. Some theoretical results are introduced (e.g. the idea that kriging weights can be negative) but the thrust of the chapter is of a more practical nature. Most of the concepts treated are presented through examples.

The first section deals with the question of negative kriging weights which can lead to negative kriged estimates. The two common situations when these negative weights tend to arise are when points are clustered or when highly structured variogram models are used (e.g. a gaussian model with no nugget effect or a power model with an exponent greater than 1.0). Examples showing both cases are presented.

In the second section, the influence of the variogram model on kriging is discussed. The shape of the variogram near the origin is shown to be critical. Variograms with different proportions of nugget effect (i.e. as a percentage of the total sill) or with different shapes (linear vs quadratic) give rise to quite different kriging weights and kriging variances, and hence to different looking kriged maps. Having said this, it is important to note that superposable models give rise to (virtually) identical kriging systems and hence to similar kriging weights and variances. So kriging is, in this way, stable against minor differences in the variogram models.

The third section deals with the screen effect. When the variogram model is well structured (long range, small nugget effect), information near the target effectively screens out the influence of more distant data. The kriging weights of outer points are zero or negligible. However with a poorly structured model the screen effect is "lifted", and so a much larger kriging neighbourhood is required. Increasing the size of the neighbourhood leads to a drop in the kriging variance and to a significant improvement in the estimator.

The fourth section is on symmetry in the kriging configuration. In practice, data are often on a regular grid. If a symmetrical layout of the data relative to the target is used, some of the weights will be equal. Recognising this at the outset can allow geostatisticians to regroup sets of kriging weights, thus reducing the size of the kriging system and hence the computer time required.

The fifth section presents some criteria for testing the quality of a kriging configuration. The most obvious one, the kriging variance, often proves to be relatively insensitive. Two other parameters, the weight of the mean in simple kriging and the slope of the linear regression, are more helpful, particularly when selecting the size of the kriging neighbourhood.

The last section treats the question of cross–validating variogram models. Points are removed one by one from the data set. The absent point is rekriged. If the variogram model is in good agreement with the data, the kriged estimates should be close to the true values. Several statistics for quantifying "closeness" are discussed.

8.3 Negative weights

It is important to realize that while kriging variances should never be negative, kriging weights can be. (Negative kriging variances can result from using a variogram model that is not positive definite or from programming errors e.g. in discretising blocks). The following example shows two simple cases where negative weights arise. The first involves two highly structured variograms (a gaussian and a power model with an exponent of 1.5) while the second involves a cluster of points.

Example 1. Suppose that samples have been taken at 4 points P_1, P_2, P_3, and P_4 that are regularly spaced 1m apart on a line. The value at their midpoint P_o is to be estimated. The locations of the points are shown in Fig. 8.1.

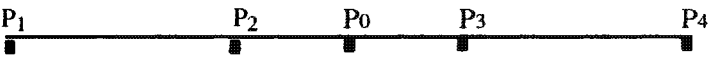


Fig 8.1. Location of the four samples P_1, P_2, P_3, and P_4 and the target P_o

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

1. a power function model with exponent 1.5 (i.e. $\gamma(h) = |h|^{1.5}$),
2. a gaussian with a distance parameter a = 0.8 and a sill of 1.0 and no nugget effect,
3. a gaussian with a distance parameter a = 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 a gaussian is $\sqrt{3}$ times its range; here 1.38. The spherical has the same range but is linear at the origin instead of quadratic. Table 8.1. gives the kriging weights corresponding to these variograms. By symmetry the weights for P_2 and P_3 are identical, as are those for P_1 and P_4.

Table 8.1. Kriging weights corresponding to the four variogram models The numbering of the weights is the same as in Fig. 8.1.

Kriging weight	Model to power 1.5	Gaussian + Nugget	Gaussian No Nugget	Spherical No Nugget
λ_1	−0.047	−0.083	0.008	0.010
λ_2	0.547	0.583	0.492	0.490
Kriging var.	0.201	0.227	0.563	0.590

The weights for the outer points are negative 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 less well structured and consequently do not result in 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 one that is linear at the origin. Consequently it is more likely to give rise to negative weights and hence to negative kriged grades (which are not desirable in mining).

Example 2. In this example we consider clustered points. Having some closely spaced points can help in estimating the behaviour of the variogram near the origin but these points can cause problems of numerical stability when inverting the kriging matrix.

Suppose we want to estimate a square 100 x 100m given 5 samples, one at each of the corners of the block and one in the centre of the block. Later a sixth sample is added close to one of the corners. Let the variogram be a spherical model with a range of 200m and a sill of 2.0. Note that this model is linear at the origin and is therefore less likely to produce negative kriging weights than would a gaussian, a cubic or a power model with an exponent above 1.0.

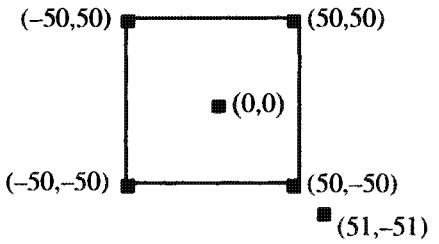


Fig 8.2. Layout of sample points relative to the target block

Table 8.2. gives the kriging weights for the 5 and 6 point data configurations. As there is no nugget effect the kriging weights for 4 of the points change very little from the 5 point to the 6 point configuration. The sum of the weights for the fifth and sixth points is (roughly) equal to the fifth weight in the 5 point configuration. The sixth point even receives a negative weight. There is also little change in the kriging variance in this case from the 5 to the 6 point configuration. So we see that for structures with no nugget effect, the results obtained using 6 points are very similar to those obtained by regrouping the last two points. One advantage of regrouping the points is that it avoids having an extra row and column in the matrix which can sometimes lead to numerical instabilities.

Table 8.2. Kriging weights and kriging variance for a spherical model for a 5 point and a 6 point configuration

Kriging Weight	5 points	6 points
(0,0)	0.436	0.436
(–50,–50)	0.141	0.141
(–50,50)	0.141	0.141
(50,–50)	0.141	0.141
(50,50)	0.141	0.152
(51,51)	–	–0.011
Kriging Variance	0.085	0.085

The results would have been quite different if there had been a significant nugget effect. The weighting would be spread more evenly over all available data points. Adding an extra point would also have led to a marked drop in the kriging variance.

These simple examples show two cases where negative weights occur: when the model is highly structured (e.g. quadratic at the origin) and when the points are clustered. We see that when a less structured model such as a spherical, is chosen or when some nugget effect is introduced the negative weights tend to disappear or at least be attenuated. Those who are interested in finding out more about this problem can consult Barnes (1984) and Chauvet (1988).

8.4 How the choice of the variogram model affects kriging

8.4.1 Similar looking variograms

In the section on fitting variograms, the importance of the choice of the nugget effect and of the shape of the variogram near the origin was stressed. No sophisticated statistical techniques were proposed for fitting models because it is possible to obtain several (visually) similar models that give equally good fits to the experimental variogram. Provided these models all have the same behaviour near the origin the resulting kriged estimates will be very similar, and so will their kriging variances. This is because the rows and columns in the kriging system are virtually identical.

To illustrate this consider the following two visually similar variogram models: firstly an exponential with a sill of 2.06 and a distance parameter a=30m (and hence a practical range of about 100m) and secondly the sum of two sphericals both with sills of 1.0 and with ranges of 40m and 100m respectively. Suppose that the objective is to krige a square block 100m by 100m from 5 data points (the four corners of the square and the centre). Table 8.3. shows how similar the kriging weights and the kriging variances are for the two models.

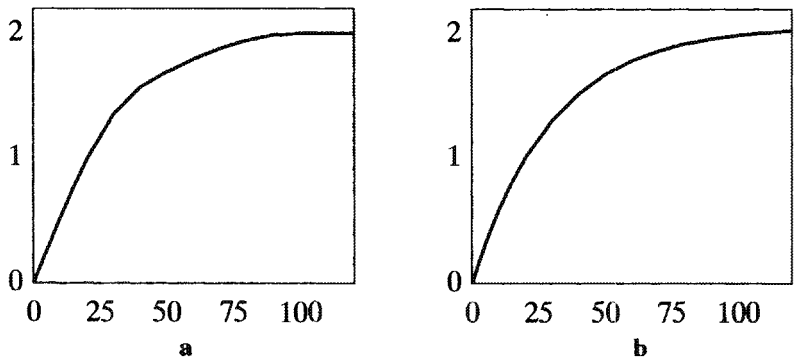


Fig. 8.3. Two visually similar models; (**a**), the sum of two sphericals with ranges of 40m and 100m and sills of 1.0, (**b**), a single exponential with a distance parameter of 30m and a sill of 2.06

Table 8.3. Kriging weights and kriging variance for two visually identical models, when used to krige a 100m by 100m block

Kriging Weights	Exponential	Sum of 2 Sphericals
Center Point	0.338	0.339
Each Corner	0.165	0.165
Kriging Variance	0.285	0.299

8.4.2 The effect of the choice of the nugget effect

The choice of the value of the nugget effect is extremely important since it has a marked effect on both the kriging weights and the kriging variance. The problem when choosing the nugget effect is that there is often no way of knowing the behaviour of the variogram at the origin, or at least for distances less than the first point on the experimental variogram. Unless some additional closely spaced holes are available, the geostatistician must guess the shape of the variogram near the origin. So it is important to understand the impact that the choice of the model has on the results of the kriging. This example is designed to illustrate the effect the model has on the kriging weights and the kriging variance.

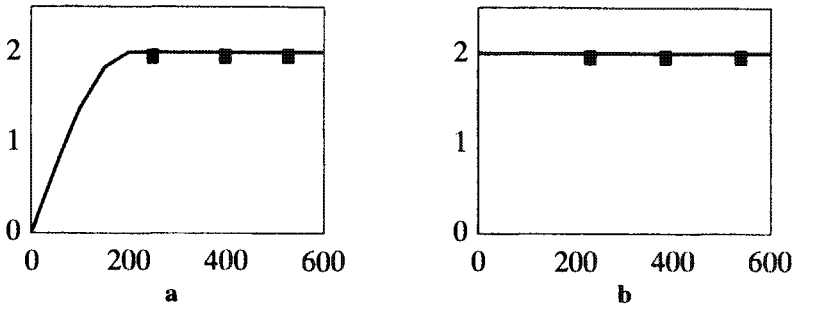


Fig. 8.4. Two models respecting the experimental variogram values but with different short scale behaviour; (**a**) a spherical with a range of 200m and a sill of 2.0, and (**b**) a pure nugget effect of 2.0

Fig. 8.4 shows two possible models fitted to an experimental variogram which had already reached its sill by the first point. Taking the extreme cases, it could be modelled by (1) a spherical variogram with a sill of 2.0 and a range of 200m (zero nugget effect) or (2) a pure nugget effect with a sill of 2.0 Unless we have some prior knowledge about this type of variable there is no way of knowing whether to use the pure nugget effect model or the more structured one, or anything in between.

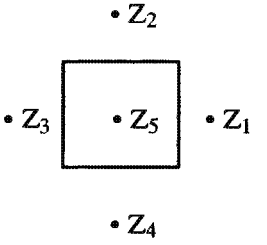


Fig 8.5. A 200 x 200m block to be kriged from 5 samples

Suppose that we want to estimate a 200 x 200m block using the central sample plus the next 4 data on a regular 200 x 200m grid. (N.B. The kriging exercise in the previous chapter had the same data configuration but the range of the variogram was 250m instead of 200m).

Table 8.4. Kriging weights and kriging variances for 5 point configuration

Kriging Weights	Pure Nugget Effect	Spherical
Center Point	0.20	0.540
Each Outer Point	0.20	0.115
Kriging Variance	0.40	0.290

The kriging weights and the kriging variances are shown in Table 8.4. The difference between the two kriging variances is striking. With a pure nugget effect it is much higher than for the other model. The effect on the kriging weights is more subtle. The pure nugget effect model gives equal weight to all points and hence less to the central sample and more to the peripheral ones whereas the structured model attributes a relatively high weight to the central sample. Since the lower nugget effect model gives less weight to the central sample, it leads to smoother contour maps, which is not intuitively obvious. This will be illustrated in Chapter 9.

The 5–point configuration used here is unrealistically small. In practice a much larger neighbourhood would be chosen. It is interesting to see what happens as this neighbourhood is enlarged. The next step up would be to 9 points on a regular 200 x 200m grid. Table 8.5. shows the kriging weights and variances for this new configuration.

Table 8.5. Kriging weights and kriging variances for 9 point configuration

Kriging Weights	Pure Nugget Effect	Spherical
Center Point	0.11	0.51
Each Inner Point	0.11	0.08
Each Outer Point	0.11	0.04
Kriging Variance	0.22	0.26

Compared to the 5–point neighbourhood, there are marked changes in the kriging weights and the kriging variance for the enlarged neighbourhood, for the pure nugget effect model but not for the spherical model. As will be seen in the next section on the screen effect, the points close to the target effectively screen out more distant ones when the variogram is well structured but not for poorly structured models with a high nugget effect or with a short range.

8.5 Screen Effect

Sampling programs often produce hundreds or thousands of data values. From a computational point of view it would clearly be prohibitive to use them all to estimate each block. Common sense suggests that the estimates will be almost as

precise if only the neighbouring data are taken into account. The question is to know how many points to include. A general rule is to take only the first few aureoles (i.e. rings) around the target if the variogram is well structured, that is, if the nugget effect is relatively small. The reason for this is that the first few aureoles screen out the effect of more distant samples. This can best be seen from an example.

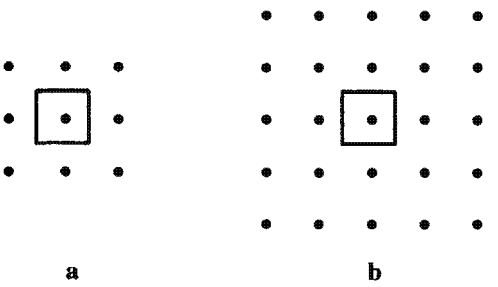


Fig. 8.6. Block with 1 aureole of data (**a**), 2 aureoles of data (**b**)

Suppose that we want to estimate a 200m x 200m block centered on a sample, using data on a 200m x 200m grid. The number of samples can be increased from 1 (the central one), to 9 (1 aureole) and then to 25 (2 aureoles). 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: spherical variograms with a sill of 2.0 and ranges of 250m and 100 m respectively, and thirdly a pure nugget effect of 2.0 (which could be thought of as a spherical with a zero range). Figure 8.7. shows the kriging weights for the 9 point and 25 point configurations for the three variogram models.

For the well structured spherical model (range 250m, at the top of the page) most of the weight is concentrated on the centre point and the four closest points. Consequently increasing the number of points to more than 25 does not lead to any significant improvement in the kriging variance. As the weights do not change much, nor does the estimated value. So there seems 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 can have nonzero OK weights. They are not necessarily zero.

Spherical Range = 250m, Sill =2.

2.6	8.2	2.6
8.2	56.7	8.2
2.6	8.2	2.6

$\sigma_k^2 = 0.22$

0.8	0.8	0.5	0.8	0.8
0.8	1.2	7.0	1.2	0.8
0.5	7.0	56.0	7.0	0.5
0.8	1.2	7.0	1.2	0.8
0.8	0.8	0.5	0.8	0.8

$\sigma_k^2 = 0.20$

Spherical Range = 100m, Sill =2.

9.4	9.4	9.4
9.4	25.0	9.4
9.4	9.4	9.4

$\sigma_k^2 = 0.35$

3.3	3.3	3.3	3.3	3.3
3.3	3.3	3.3	3.3	3.3
3.3	3.3	19.0	3.3	3.3
3.3	3.3	3.3	3.3	3.3
3.3	3.3	3.3	3.3	3.3

$\sigma_k^2 = 0.25$

Pure Nugget Effect, Sill =2.

11.1	11.1	11.1
11.1	11.1	11.1
11.1	11.1	11.1

$\sigma_k^2 = 0.22$

4.0	4.0	4.0	4.0	4.0
4.0	4.0	4.0	4.0	4.0
4.0	4.0	4.0	4.0	4.0
4.0	4.0	4.0	4.0	4.0
4.0	4.0	4.0	4.0	4.0

$\sigma_k^2 = 0.08$

Fig. 8.7. Kriging weights and kriging variance for configurations with 1 or 2 aureoles of data, for three variogram models

8.6 Symmetry in the equations

When kriging was first developed, computer facilities were poor. Inverting large matrices in order to solve 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 symmetry in the system. For example, the exercise on ordinary kriging given in the previous chapter involved kriging a block using 5 data, four of which are set symmetrically outside the block. These four weighting factors are clearly identical.

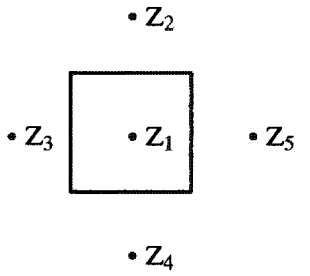


Fig. 8.8. A symmetric data configuration

These four 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. If the first kriging weight is associated with the central point Z_1, and the second weight with the group S, the kriging system can be rewritten as:

$$\begin{bmatrix} \bar{\gamma}(Z_1,Z_1) & \bar{\gamma}(Z_1,S) & 1 \\ \bar{\gamma}(Z_1,S) & \bar{\gamma}(S,S) & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \mu \end{bmatrix} = \begin{bmatrix} \bar{\gamma}(Z_1,V) \\ \bar{\gamma}(S,V) \\ 1 \end{bmatrix} \qquad [8.1]$$

As it is not obvious how to calculate the various variogram values, this will be presented in detail.

$$\bar{\gamma}(Z_1, Z_1) = 0\ ,\ \bar{\gamma}(Z_1, V) = 0.88 \qquad [8.2]$$

$$\bar{\gamma}(Z_1, S) = \frac{1}{4}\left[\gamma(Z_1, Z_2) + \ldots + \gamma(Z_1, Z_5)\right] = 1.89 \qquad [8.3]$$

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

$$= \frac{1}{4} [\, \gamma(0) + 2\, \gamma(100\sqrt{2}) + \gamma(200)] = 1.5 \qquad [8.4]$$

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

$$= \bar{\gamma}(Z_2, V) = 1.86 \qquad [8.5]$$

This leads to solving a 3 x 3 system instead of a 6 x 6 one:

$$\begin{bmatrix} 0 & 1.89 & 1 \\ 1.89 & 1.50 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_S \\ \mu \end{bmatrix} = \begin{bmatrix} 0.88 \\ 1.86 \\ 1.00 \end{bmatrix} \qquad [8.6]$$

The solution is $\lambda_1 = 0.60$, $\lambda_S = 0.40$ and $\mu = 0.12$ and hence:

$$z^* = 0.60 z_1 + 0.40\, (z_2 + z_3 + z_4 + z_5) / 4. \qquad [8.7]$$

Of course, $\sigma_K{}^2$ is the same as before. Here we have succeeded in reducing the system from 6 x 6 to 3 x 3 without any loss of precision. Since the time taken to invert a matrix is roughly proportional to the cube of the size of the system, halving the size effectively reduces the time and hence the cost to about one eight of its original value. This clearly represents a considerable saving.

Having seen that the size of the kriging system can be significantly reduced in some cases without loss of precision, it is important to be able to distinguish symmetric configurations from those which might at first glance appear to be symmetric. For the weights to be identical, the data points must be symmetric with respect to each other and to the point or block to be kriged. In the next configuration the data points are symmetric with respect to each other but not to the block.

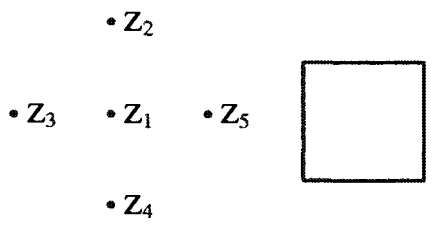


Fig 8.9. A symmetric data configuration with the block off center

Another case that often leads to mistakes is when the data and the block are symmetric relative to each other but when the variogram model is anisotropic.

Suppose that the data configuration is as shown on the left of Fig. 8.10., but that the variogram model has a range of 500m in the NS direction compared to 250m in the EW direction. After taking account of the anisotropy the dimensions in the EW direction are twice those in the NS direction. Consequently, the weights for Z_3 and Z_5 are identical and so are those for Z_2 and Z_4. But the four weights are no longer identical. This has to be kept in mind when regrouping data points.

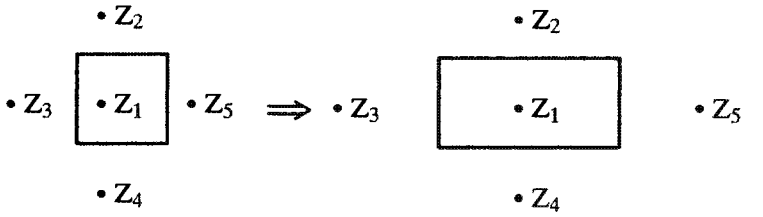


Fig. 8.10. The kriging configuration before and after taking account of the anisotropy

8.7 Testing the quality of a kriging configuration

Looking at how the kriging weights and the kriging variance evolve as the number of points is increased gives us some idea of what size neighbourhood is optimal. But experience has shown that the kriging variance is a fairly insensitive parameter for testing the quality of a kriging configuration. The weight of the mean in simple kriging and the slope of the linear regression of the true grade on the estimated value turn out to be far more sensitive quality control parameters and hence more useful. The additivity theorem in Chapter 7 proved that the OK estimator could be written as

$$Z_V^* = \sum \lambda'_i Z(x_i) + m^* \lambda_M \qquad [8.8]$$

where m* is the kriged estimate of the mean, and that the OK variance can be split up into the SK variance plus a term that depends on the weight of the mean in SK and on the kriging variance of the estimate of the mean:

$$\sigma^2_{OK} = \sigma^2_{SK} + (\lambda_M)^2 \text{ Var } (m^*) \qquad [8.9]$$

The first equation shows that when the weight of the mean in simple kriging λ_M is low (near 0), the estimated grade depends mainly on the local values of Z(x), i.e. the data in the kriging configuration and not on the estimate of m. So the degree of stationarity required is less. In addition the OK estimator and the SK estimator are closer.

The second equation shows that when the weight of the mean in SK is low the OK variance is close to the SK variance. Little precision is lost in having to estimate

the mean. Conversely when it is high, the OK variance is markedly higher than the SK variance. Rivoirard (1984, 1987) has shown that 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.

8.7.1 Example: Adding extra samples improves the quality of the estimate

In South African gold mines, blocks of about 20 x 5m are estimated from channel samples on an approximately 5 x 5m grid. In general the closest three rows of samples are used. Sometimes an extra row of samples next to the block is available. Clearly using additional samples close to the block will improve the quality of the estimate but by how much? Figure 8.11. shows the two possible layouts.

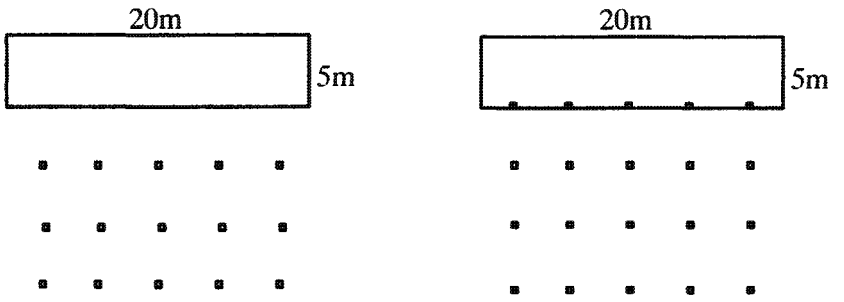


Fig. 8.11. The two sample layouts considered

Suppose that the variogram is spherical with a sill of 1.0 and a range of 20m. The kriging variance for the first layout is 0.368 compared to 0.114 for the second one. The slope of the linear regression rises from 0.531 to 0.863, while the weight of the mean in SK drops from 0.623 (which is high) to 0.255. All three quality parameters show that the inclusion of an extra row of samples leads to a marked improvement in the quality of the estimator.

8.8 Cross-validation

As several different variogram models can often be fitted to an experimental variogram, one would like to know which is the "best". Cross-validation is often used for this. The procedure consists of eliminating one data point from the set temporarily and then kriging its value using the remaining samples as data. If this

is repeated for all the points (or for a representative subset of them) we obtain a series of estimation errors for each data point and for each variogram model. If the variogram suits the data, the mean of the estimation errors and the mean of the standardised estimation errors should be zero, and the variance of the standardised estimation errors should be 1.0.

Let Z^*_{ij} denote the kriged estimate of the i^{th} point obtained using the j^{th} variogram model and let σ^2_{ij} be the kriging variance. If its true value is Z_i, then the corresponding estimation error $Z^*_{ij} - Z_i$. That is, we would expect:

$$E\ (Z^*_{ij} - Z_i\ ) = 0 \qquad [8.10]$$

$$E\left(\frac{Z^*_{ij} - Z_i}{\sigma_{ij}} \right) = 0 \qquad [8.11]$$

$$Var\left(\frac{Z^*_{ij} - Z_i}{\sigma_{ij}} \right) = 1 \qquad [8.12]$$

This leads us to calculate the following statistics:

$$\sum (\ Z^*_{ij} - Z_i\ ) \qquad [8.13]$$

$$\sum \frac{(Z^*_{ij} - Z_i\ )}{\sigma_{ij}} \qquad [8.14]$$

$$\sum \frac{(Z^*_{ij} - Z_i\ )^2}{\sigma^2_{ij}} \qquad [8.15]$$

As all these statistics are strongly affected by any extreme values (i.e. by outliers) it might be preferable to use robust forms of these. But no matter whether robust or ordinary statistics are used, it is not common for all three statistics to show the same model as being "the best". So a choice has to be made.

There are practical problems in using this technique with drillhole data because when one sample is removed and re-estimated, the resulting kriged estimate depends mainly on the nearest samples (i.e. those vertically above and below). Consequently cross-validation only tests the goodness of fit of the vertical component of the variogram and not the rest of the model.

9 Case Study using Kriging

9.1 Summary

The case study in Chapter 5 presented the structural analysis for an iron ore deposit. We now show how to use the fitted 3D variogram model to krige point values then block grades. As the model has a high nugget effect, a large kriging neighbourhood is required. The fourth section shows what happens when smaller neighbourhoods are used. The last section illustrates why it is not advisable to krige small blocks from sparse samples, in order to calculate the recoverable reserves.

9.2 Iron ore deposit

The data available to krige this deposit consists of nearly 500 15m core sections coming from about 40 vertical drillholes. The fitted variogram model consists of a nugget effect of 1.8 plus two anisotropic spherical structures. Their sills and their ranges in the horizontal and vertical directions are shown in Table 9.1.

In the mining indusry block estimation is more common than point estimation which is used sometimes as the input for contouring packages. In this chapter we will illustrate the use of both point and block kriging, Comparing the results highlights the effect of the size of the support. The first step in kriging is to choose the grid size.

Table 9.1 Parameters of fitted variogram model which is isotropic in the horizontal direction but not in the vertical

	Sill	Horizontal Range	Vertical Range
Nugget effect	1.8	–	–
1st Spherical	1.2	80m	65m
2nd Spherical	1.5	400m	65m

9.2.1 Grid size for kriging

As the drillholes are about 80m apart, the horizontal dimensions of the grid to be kriged were set at 100m x 100m. The height of grid cells has been set to 15m because this is the length of the core sections. The resulting grid consists of 19 cells (EW) x 12 cells (NS) x 36 vertical levels. It is not advisable to krige very small blocks (say 10m x 10m horizontally) or blocks which are shorter than the core section length (15m). The reasons for this will be given in detail later in the chapter.

9.3 Point kriging using a large neighbourhood

Having fixed the dimensions of the kriging grid, the next step is to choose the parameters for the kriging neighbourhood. As the model has 40% nugget effect, it is not well structured and so a large neighbourhood containing many points is required. After some preliminary testing it was decided to use a neighbourhood containing a minimum of 8 samples and an optimum of 80 samples. The search ellipse was limited to a horizontal radius of 500m and a vertical one of 300m. In section 9.5 we will show what happens when a small neighbourhood with only a few samples is used for kriging.

Figure 9.1. shows the contour lines for one level (number 14) which is near the middle of the deposit. The outer blocks that are hatched have not been kriged at all because insufficient data were found in the neighbourhood. Figure 9.2. shows the kriging standard deviation map. No labels have been put on the isolines because they were difficult to read but it is clear that the kriging standard deviation is high at the edges of the area and low where the samples are more dense. Table 9.2 gives the basic statistics of the kriged estimates and the corresponding kriging standard deviation.

9.4 Block kriging using a large neighbourhood

The same size neighbourhood was used to krige blocks of size 100m x 100m as for point kriging. One difference between kriging points and blocks is that the blocks have to be discretized in order to calculate terms like $\bar{\gamma}(V,x)$ and $\bar{\gamma}(V,V)$. Here a 6 x 6 x 1 discretization was used. As the core length equals the block height, the vertical discretization had to be 1.

Figures 9.3. and 9.4. show the kriged estimates and the corresponding kriging standard deviations while Table 9.3 gives the basic statistics. The overall shape of the contour lines is the same as for point kriging. Comparing the statistics of the kriged values for blocks with those for points we see that the average grades for the whole area are identical but that the minimum value for the kriged points is lower than for the blocks. Conversely the maximum for the points is higher than for the blocks. That is, the histogram of the kriged block estimates is tighter around the mean than the one for points.

Table 9.2 Basic statistics of kriged point estimates and the corresponding kriging standard deviations

	Number	Minimum	Maximum	Mean	St Dev
Point Estimates	7501	52.91	59.12	56.84	0.704
Standard Deviation	7501	1.655	2.380	2.127	0.103

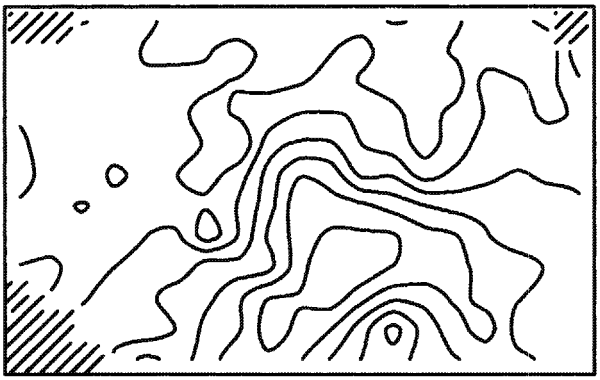


Fig. 9.1. Contour lines for level 14 obtained for point kriging using 8 angular sectors with an optimum of 80 samples, and a minimum of 8 samples

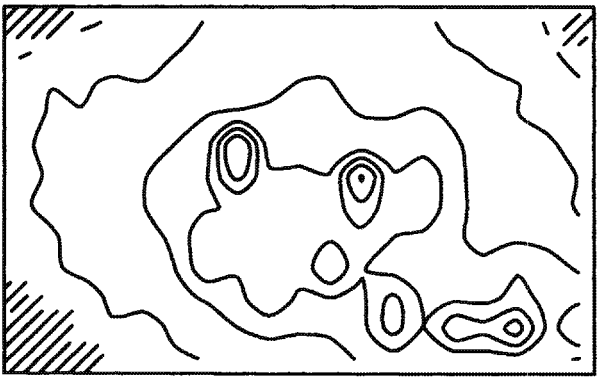


Fig. 9.2. Contour lines for the kriging standard deviation for level 14 obtained for point kriging

Table 9.3 Basic statistics of kriged blocks and the corresponding kriging standard deviations

	Number	Minimum	Maximum	Mean	St Dev
Block Estimates	7501	53.27	58.97	56.84	0.605
Standard Deviation	7501	0.720	1.636	1.238	0.166

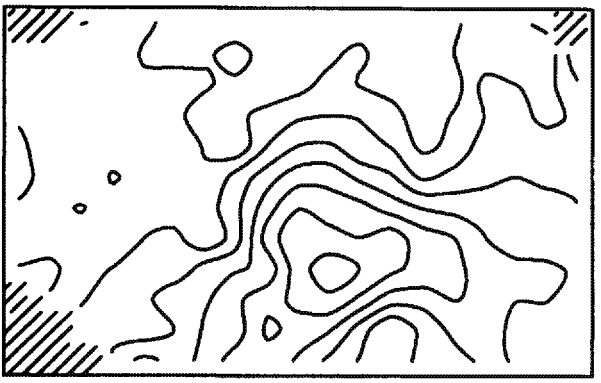


Fig. 9.3. Contour lines for level 14 obtained for block kriging using 8 angular sectors with an optimum of 80 samples, and a minimum of 8 samples

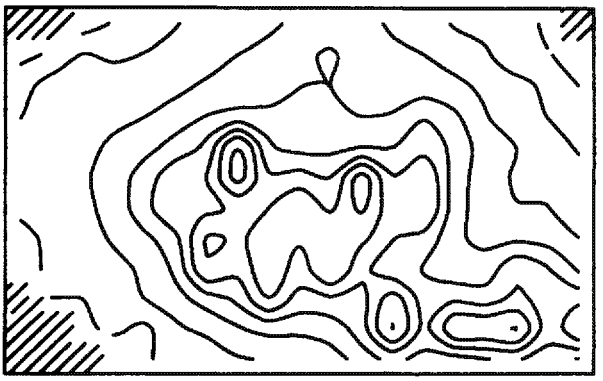


Fig. 9.4. Contour lines for the kriging standard deviation for level 14 for block kriging

The difference between the standard deviations for the points and the blocks is quite startling. The best estimates for points have a standard deviation of 1.655 whereas it is 1.636 for the worst blocks. This is because it is much easier to estimate the average value over a large volume accurately than a very small one.

9.5 Point kriging using smaller neighbourhoods

In section 9.3 a large neighbourhood containing many points was used for point kriging. It is instructive to see what happens when a smaller neighbourhood containing only a few points is used. Figure 9.5. shows the contour lines obtained using only the closest three samples. Clearly something has gone wrong, but what? Before going into detail, we should note that there are fewer hatched blocks in the corners. More blocks have been kriged this time because fewer samples are required in a kriging neighbourhood before the grid node could be estimated.

Fig. 9.5. Contour lines obtained for point kriging using only three samples

9.5.1 What is causing the ugly concentrations of lines?

In order to work out what has gone wrong, we focus on four adjoining grid cells in an area where there is an ugly concentration of isolines. These are rows 2 to 5 in column 12 of level 14. Table 9.4 gives the coordinates of the three samples used in kriging each cell, together with their values and the weighting factors. In three out of the four cases, all the samples come from a single drillhole. The exception occurs where the closest drillhole stops just above level 14. The bottom sample from it is used together with two samples from the next closest drillhole. The kriged estimate and the corresponding standard deviation are also given for each grid node. Note how erratic the kriged grades are.

Table 9.4 Coordinates, kriging weights and values of the samples effectively used for kriging four selected grid nodes

Block (12,5,14): Kriged estimate = 54.83, kriging standard deviation = 2.14

	X coord	Y coord	Z coord	Weight	Value
1st Sample	−1450	−995	17.25	35.1%	56.6
2nd Sample	−1450	−995	18.75	32.4%	56.9
3rd Sample	−1450	−995	15.75	32.4%	51.05

Block (12,4,14): Kriged estimate = 53.68, kriging standard deviation = 2.27

	X coord	Y coord	Z coord	Weight	Value
1st Sample	−1475	−1092	17.25	32.3%	55.0
2nd Sample	−1475	−1092	15.75	33.8%	54.25
3rd Sample	−1475	−1092	18.75	33.8%	51.85

Block (12,3,14): Kriged estimate = 55.05, kriging standard deviation = 2.27

	X coord	Y coord	Z coord	Weight	Value
1st Sample	−1494	−1195	20.25	35.3%	51.6
2nd Sample	−1295	−1236	17.25	36.6%	57.0
3rd Sample	−1295	−1236	18.75	28.1%	56.5

Block (12,2,14): Kriged estimate = 58.75, kriging standard deviation = 2.33

	X coord	Y coord	Z coord	Weight	Value
1st Sample	−1319	−1330	17.25	44.1%	58.9
2nd Sample	−1319	−1330	18.75	27.9%	58.6
3rd Sample	−1319	−1330	20.25	28.0%	58.65

Table 9.5 summarizes the numerical values of the kriged estimates for the grid nodes in level 14 from rows 2 to 5, from columns 10 to 14. The four target cells lie in the central column. Looking at these values it is easy to see why there is such a dense concentration of isolines in the region. As the target grid node moves, the drillhole effectively being used as data for kriging changes. These jumps cause abrupt changes in the kriged estimate and hence the unsightly concentrations of isolines. Problems of this type were first identified by Renard and Yancey (1984) when they were kriging the top of an oil reservoir using seismic data lying on straight line profiles. They realized that the erratic jumps in output values occurred as "noisy" sample points moved into and out of the moving kriging neighbourhoods.

Table 9.5 Kriged grid node values for rows 2 to 5, columns 10 to 14 in level 14. They were obtained using only 3 samples

55.38	54.83	54.83	53.95	54.03
55.66	53.82	53.68	55.47	55.29
57.41	53.34	55.05	56.61	56.57
53.72	55.03	58.75	58.76	55.70

9.5.2 How to eliminate these concentrations of contour lines

In order to eliminate these patterns in the contour lines, more sample points have to be included in the kriging system, but just increasing the total number of samples is not enough. For example if the number were increased to 5 or 9 samples, the closest points would still come from a single drillhole and the problem would not be solved. We need to ensure that samples from several drillholes are incorporated into the kriging system. One simple and convenient way to do this is by using angular sectors. Figure 9.6. shows eight angular sectors centered on grid cell (12, 4, 14).

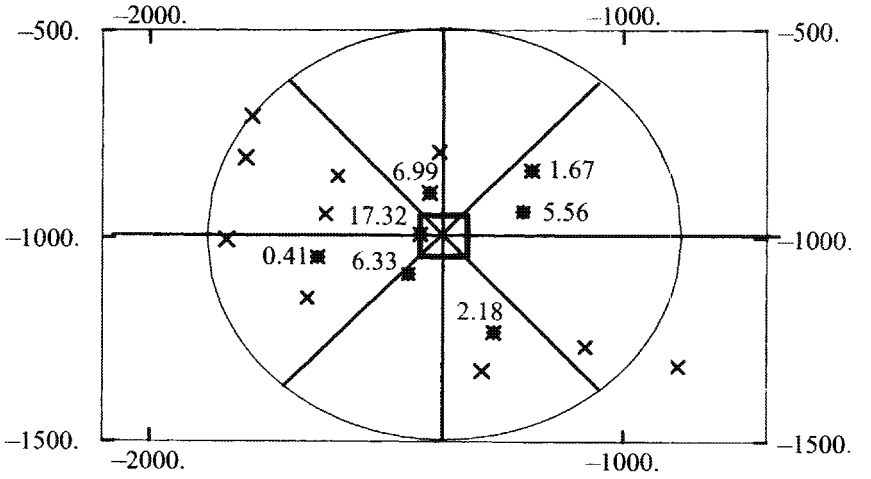


Fig. 9.6. Target grid cell (12, 4, 14) with the area around it split into 8 angular sectors

As the maximum search radius was set to 500m earlier, samples outside this are not taken into account. The circle shows how big this area is. The crosses in the figure mark the position of samples on the same level. The numbers represent the kriging weights, expressed as a percentage, for the samples that were effectively taken into account. Samples on other levels would also have been used.

After the number of sectors has been specified, the kriging routine locates and uses the closest sample or samples in each sector and uses these. The user can usually specify how many empty sectors can be tolerated. A second attempt at kriging the grid nodes was made using 4 angular sectors each containing at least 3 samples (Fig. 9.7.).

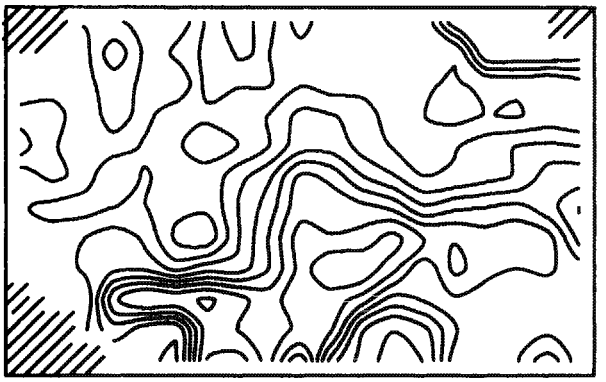


Fig. 9.7. Contour lines obtained for point kriging using 4 angular sectors each containing at least 3 samples.

These results are clearly an improvement over those obtained using only three points but are not good enough. There are still some concentrations of contour lines. The contour map given initially (Fig.9.1.) was obtained using 8 angular sectors, which is a common choice. It is much more satisfactory. This example shows how important it is to include enough samples in the kriging system.

9.6 Kriging small blocks from a sparse grid

One of the main uses of geostatistics in the mining industry has been to estimate the block grades during the feasibility and prefeasibility stages of project evaluation. At this stage the available data is usually widely spaced. In order to design the mine, planners often krige very small blocks. It is then tempting to count up the blocks above various cutoff grades so as to estimate the recoverable reserves. The aim of this section is to show how dangerous and misleading this can be.

The problems of kriging small blocks have been well known for many years. Many authors including Journel and Huijbgrets (1978), David (1977, 1988), Royle (1979) and Clark (1982) have pointed out that the kriged grades are much smoother

than the real grades of the small blocks would be. Ravenscroft and Armstrong (1990) illustrated this for a Witwatersrand–type gold deposit. Blocks of size 2m x 2m were kriged using data on a 10m x 10m grid. For reasons of confidentiality, the grades were scaled to have a mean of 1. Figures 9.8.**a** and **b** show the histograms of the true block grades and the corresponding kriged grades taken from that paper.

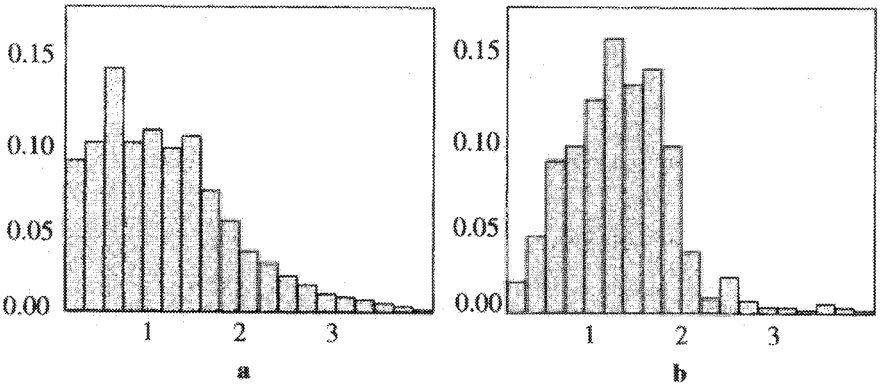


Fig 9.8. Histograms of (**a**) the true block grades and (**b**) the kriged estimates obtained from widely spaced samples, from Ravenscroft and Armstrong (1990)

Although both histograms have the same mean (1.0), they have different shapes and different variances. As expected, the variance of the kriged estimates was smaller than that of the true block grades. Consequently the grade tonnage curve calculated from the kriged block grades is quite different from the real one. This shows that kriging should not be used for estimating the grades of small blocks from widely spaced data. In fact linear estimators in general should not be used for this. That includes inverse distance and inverse distance squared methods.

More sophisticated methods are required for estimating recoverable reserves. These can be split into two broad categories:

1. conditional simulations, and
2. nonlinear estimators such as indicator kriging and disjunctive kriging.

As the scope of this book is limited to linear geostatistics, we shall not go into these methods. Readers who are interested in nonlinear methods such as disjunctive kriging can consult Matheron (1976), Marechal (1975) or Rivoirard (1994). Those interested in indicator kriging can consult Isaaks and Srivastava (1989) or Journel (1983).

Over the past 15 years the use of simulation methods has taken off in the geostatistical community. Applications to many different fields can be found in the proceedings of recent conferences. See for example the volumes edited by Baafi and Schofield (1997), Armstrong and Dowd (1994), Dimitrakopoulos (1994) and Soares (1993).

9.6.1 What size blocks can be kriged?

Having stressed that it is not advisable to krige small blocks from sparse data, it is natural to want to know how small the blocks can be. Experience has shown that it is best to keep the blocks approximately the same size as the separation between samples. At the outside, the horizontal dimensions of the blocks can be half those of the sample grid. In the iron ore case study it would lead to a minimum size of blocks of 40m x 40m x 15m or 50m x 50m x 15m. That is, four times as many blocks would be kriged as there are data.

Another point to note when kriging blocks of this size is their location relative to the samples. Figure 9.9.**a** shows a plan view of samples on a regular 100m x 100m grid together with blocks half that size. Each block has one corner touching a sample. So if the same kriging neighbourhood is used for all blocks, they would all have the same kriging variance. This would not be true of the layout shown in Fig. 9.9.**b**. Blocks containing a sample in the centre would have a low kriging variance; those marked with (2) would have a higher variance and those marked (3) would have a very high variance because they are far from samples. A person looking at a map of kriged grades obtained using the second layout would not necessarily guess that some of the estimates were much better than others. So it is important to look carefully not only at the block size but also the layout.

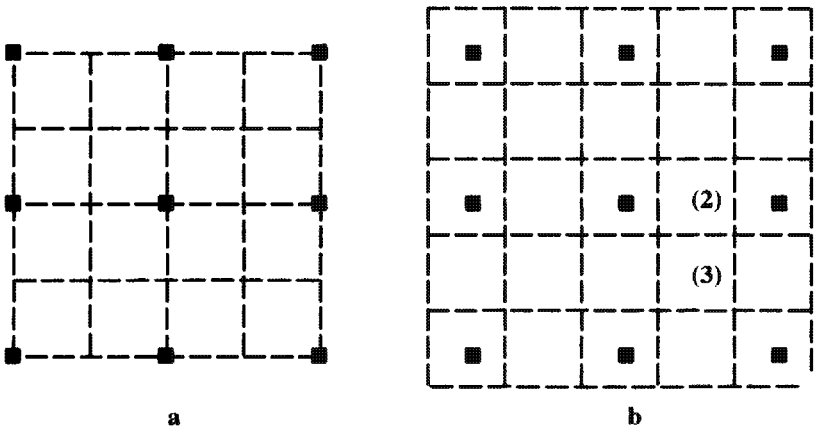


Fig. 9.9. Two possible layouts for blocks of size 50m x 50m relative to samples on a 100m x 100m grid. In (**a**) all blocks have the same kriging variance whereas in (**b**) those marked (2) and (3) have a higher variance than the ones containing a central sample. Layout (**a**) is therefore preferable

10 Estimating the Total Reserves

10.1 Summary

Once a suitable prospect has been found, an exploratory drilling campaign is carried out to establish the limits of the mineralization (if they are not already known), and to determine the total ore tonnage and the average grade. As well as knowing the total reserves, it is very important to know how accurate the estimates are. Provided there are not too many samples, kriging can be used to estimate the reserves and the kriging variance will give a measure of its accuracy. However when there are too many points to invert the kriging system a different approach is required.

This chapter presents several approximations for estimating the variance associated with the total reserves when kriging cannot be used. The variance depends on whether the limits of the orebody are known a priori or not. After presenting the concept of extension variance, the first part of the chapter treats different approximations for evaluating the global estimation variance while the second half considers the question of "optimal" drilling grids.

10.2 Can kriging be used to estimate global reserves?

Up to this point we have seen how kriging can be used to give local estimates of point or block values. So it is natural to ask whether it can be used to estimate global reserves; that is, the reserves contained in the whole deposit or a large part of it. In general the term "global estimation" refers to the estimates made early in a project during the feasibility study when only widely spaced samples are available. During production many more samples are available, for example from blastholes or channel samples.

Two different situations have to be distinguished: (1) when there are relatively few data and (2) when there are a large number of points (say more than several

hundred points). In the first case kriging can be used but the problems of inverting the matrix preclude it in the second.

One way of attacking the problem in the second case would be to divide the region into zones containing less than say 100 to 200 points and krige the zones separately. This would give an accurate estimate of the mean grade. However it still leaves the problem of how to calculate the global variance unsolved. It would be tempting to sum (or average) the individual kriging variances but this gives the wrong answer. David (1973) describes a theoretically sound way of recombining the variances but it is very complicated. At the feasibility study stage a simpler method is required. If the samples are evenly spaced, the global reserves can be estimated by using the arithmetic mean of the grades.

This chapter presents several methods for assessing the associated estimation variance. Two approximations (direct composition of terms, and the use of line and slice terms) will be discussed. The problem of estimating the reserves within an area known to be mineralized will be treated before going on to discuss the case where the geometry of the deposit is also unknown. But first we show how to calculate the extension variance of a sample since this is used in what follows.

10.3 Extension variance

Suppose that we want to estimate the average grade inside a region V; that is, we are interested in the integral:

$$Z(V) = \frac{1}{V} \int_V Z(x)\, dx \qquad [10.1]$$

Suppose that the only information available is the average value for a small volume v. Typically V is a mining block or a panel and v is a drillhole or some other type of sample. So we have to estimate Z(V) from Z(v) where:

$$Z(v) = \frac{1}{v} \int_v Z(x)\, dx \qquad [10.2]$$

It seems natural to take the value of Z(v) as the estimated value of Z(V). What error is made in doing this? First of all, if Z(x) satisfies the stationary or the intrinsic hypotheses, Z(v) is an unbiased estimator of Z(V). We need to be able to calculate the variance of extending the grade of v into V. It is sometimes denoted by $\sigma_E^2(v, V)$ or σ_E^2 for short.

Conceptually it is simply the variance of estimating Z(V) by Z(v). In geostatistics, the term "extension variance" is usually reserved for the case where a block is being estimated from its central sample. The more general term "estimation variance" is used in more complicated situations where several samples are taken into account. The theoretical value of the extension variance is given by:

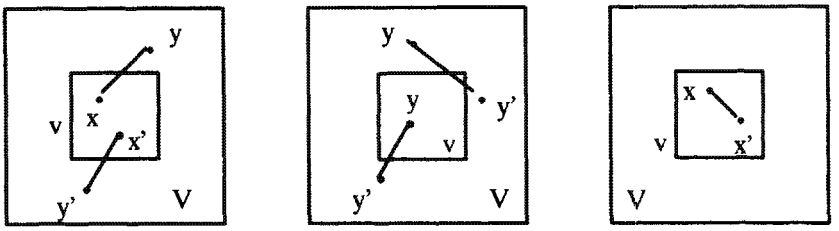
$$\sigma_E^2(v, V) = \frac{2}{vV}\int_v \int_V \gamma\,(x - y)\,dx\,dy$$

$$- \frac{1}{v^2}\int_v \int_v \gamma\,(x - x')\,dx\,dx' - \frac{1}{V^2}\int_V \int_V \gamma\,(y - y')\,dy\,dy' \qquad [10.3]$$

Consequently

$$\sigma_E^2(v, V) = 2\bar{\gamma}(v, V) - \bar{\gamma}(v, v) - \bar{\gamma}(V, V) \qquad [10.4]$$

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 respectively.



$\bar{\gamma}(v, V)$ $\bar{\gamma}(V, V)$ $\bar{\gamma}(v, v)$

Fig. 10.1. Meaning of the average variogram terms

The formula [10.4] holds for any shape of v and V. In particular v does not have to be included in V. The factors influencing the extension variance are:
- the regularity of the variable (through γ),
- the geometry of V,
- the geometry of v,
- the location of v relative to V.

This formula can be rewritten as:

$$\sigma_E^2(v, V) = \left[\bar{\gamma}(v, V) - \bar{\gamma}(V, V)\right] - \left[\bar{\gamma}(v, V) - \bar{\gamma}(v, v)\right] \qquad [10.5]$$

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.

Another obvious but nevertheless important property of the extension variance is that it involves the variogram and the geometry of the area to be estimated but not the actual sample values. This was also seen with the kriging variances and the kriging weights.

10.4 Relationship to the dispersion variance

People tend to confuse the dispersion variance σ^2 [v|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 when the sample v takes all possible positions within V.

We now go on to calculate the total reserves (and the average grade) for large regions known to be mineralized.

10.5 Area known to be mineralized

Suppose that the area is known to be mineralized throughout and that the deposit is two dimensional (e.g. a coal seam or a gold reef). For simplicity's sake it is represented as a rectangle but the same procedure can be applied to any shaped region.

10.5.1 Direct composition of terms

In the first instance, we assume that the samples are on a regular or a nearly regular grid. To estimate the total ore tonnage we multiply the area of the region by the average seam or reef thickness (here average means the arithmetic average). Tests involving kriging using a large number of samples on a regular grid have shown that the weighting factors are very close to 1/N where N is the number of samples. So our estimator is just:

$$Z^*(V) = \sum \frac{1}{N} Z_i \qquad [10.6]$$

The corresponding estimation variance is

$$\frac{2}{NV}\sum_i \int \gamma(x_i - y)dy - \frac{1}{V^2}\int\int \gamma(y - y')dy\ dy' - \frac{1}{N^2}\sum\sum \gamma(x_i - x_j) \qquad [10.7]$$

If N is large, this formula becomes very unwieldy. So an approximation to it is made based on the assumption that the sum of the covariance terms between extension errors is zero. In Fig. 10.2. there are N squares each with a sample at its center. Let $Z(v_i)$ be the true but unknown average over the i^{th} square.

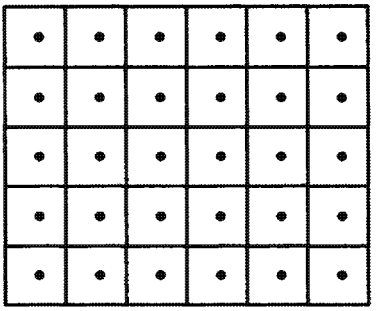


Fig. 10.2. Grid containing regular samples

As the average over the whole area is just the average of the individual squares, the error made in using the average of the samples as the estimator is then:

$$\sum \frac{1}{N} [Z(v_i) - Z_i] \qquad [10.8]$$

That is, it is the average of the partial errors. To simplify the calculations we now assume that the sum of the covariance terms between these is zero. Checks made by David (1973, 1977 p201) have shown that this is quite a good approximation. Consequently the estimation variance is

$$\sigma^2_{EST} = \text{Var}\left(\frac{\sum(Z_i - Z(v_i))}{N}\right) = \frac{1}{N^2}\sum \text{Var}(Z_i - Z(v_i)) \qquad [10.9]$$

Now Var $(Z_i - Z(v_i))$ is just the variance of extending the central value to the whole square v_i. As all the squares are the same size,

$$\sigma^2_{EST} = \frac{1}{N}\,\sigma^2_E(0, v) \qquad [10.10]$$

This provides a simple way of estimating the total reserves and of evaluating the estimation variance in terms of the variogram function. (Note: the same method can also be applied when the blocks v_i are not all the same size. See Journel and Huijbregts (1978) pp 415–417 for details).

This method is based on the direct composition of terms. It is valid only when blocks are roughly square. If the ratio of the length to the breadth is more than about 3:1 the composition by line and slice terms described in the next section 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 in terms of the variogram range rather than distance units. For example in Fig.10.3. if the samples are actually on a square grid (**a**) but the variogram has an anisotropy ratio of 2.1 the sample configuration after taking account of the anisotropy is elongated as is shown in (**b**).

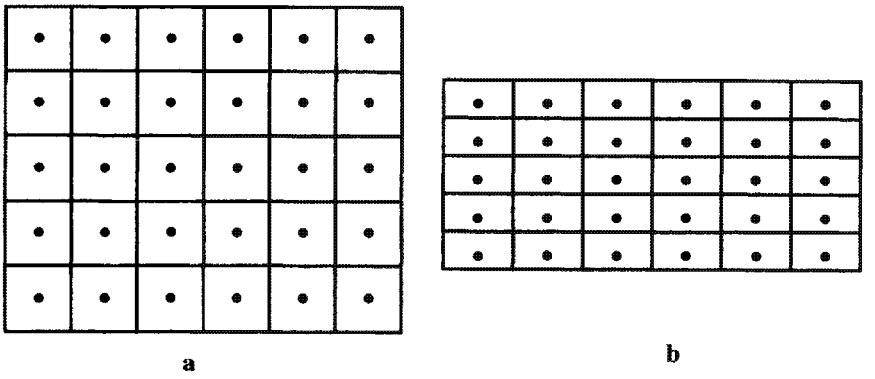


Fig. 10.3. (**a**) Original grid, (**b**) grid after taking account of anisotropy

10.5.2 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 in the other, as for example happens in seismic surveys, sonar measurements of fish stock or sometimes in underground development sampling. 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.

Fig. 10.4. illustrates the case for two profiles. The area under study has been sampled along two profiles d_1 and d_2. If the distance between samples along the lines is s, there are $n_1 = d_1/s$ samples in the first line. If there are more than two profiles, the separation between them is assumed to be constant. Slices of that width are drawn centered on the profiles. Here they are denoted by v_1 and v_2. So the whole area V is the union of the slices.

As usual we let Z(V) be the true but unknown grade of the whole area V. Let $Z(v_i)$ and $Z(d_i)$ be the true grades of the slice v_i and the line section d_i. At first we assume that the line sections 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 area v_i. As the distance between profiles is constant, $v_1/d_1 = v_2/d_2$ and we have

$$Z^*(V) = \frac{\sum_i v_i \, Z(d_i)}{\sum_i v_i} = \frac{\sum_i d_i \, Z(d_i)}{\sum d_i} \qquad [10.11]$$

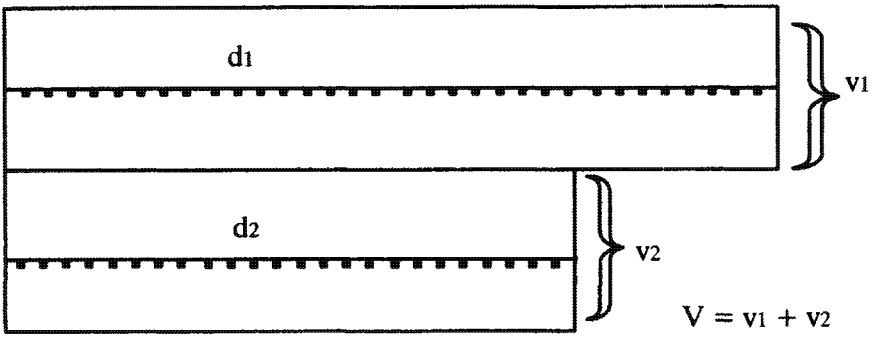


Fig. 10.4. Two profiles (slices) of length d_1 and d_2

In the same way the true but unknown grade Z(V) is

$$Z(V) = \frac{\sum_i v_i \, Z(v_i)}{\sum_i v_i} = \frac{\sum_i d_i \, Z(v_i)}{\sum d_i} \qquad [10.12]$$

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

$$Z^*(V) - Z(V) = \frac{\sum_i d_i \, [\, Z(d_i) - Z(v_i) \,]}{\sum d_i} \qquad [10.13]$$

By the approximation principle the estimation variance is

$$\sigma_E^2 = \mathrm{Var}[Z^*(V) - Z(V)] = \frac{\sum_i d_i^2 \, \sigma_{Ei}^2}{(\sum d_i)^2} \qquad [10.14]$$

where σ_{Ei}^2 is the elementary extension variance of the central line section to its slice of influence. Note that these are weighted by the squares of the lengths d_i.

More realistically the line sections are obtained by averaging the sample grades along the line section. As the samples are equally spaced, the average of the n_i samples is used to estimate $Z^*(d_i)$:

$$Z^*(d_i) = \sum_k Z(s_k) \; / \; n_i \qquad [10.15]$$

The total estimation error $Z^*(V) - Z(V)$ can be split into two terms.

$$\frac{\sum_i d_i \left[Z(d_i) - Z(v_i)\right]}{\sum d_i} + \frac{\sum_i d_i \left[Z^*(d_i) - Z(d_i)\right]}{\sum d_i} \qquad [10.16]$$

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

$$\frac{\sum_i d_i / n_i \sum_k \left[Z^*(s_{ik}) - Z(s_{ik})\right]}{\sum d_i} \qquad [10.17]$$

As usual the sum of the covariances is considered to be zero so we can sum the variances. We finally get:

$$\sigma_E^2 = \frac{\sum_i d_i^2 \, \sigma_{Ei}^2}{[\sum d_i]^2} + \frac{\sum_i d_i^2 \, \sigma^2(O,s)/n_i}{[\sum d_i]^2} \qquad [10.18]$$

If N denotes the total number of samples ($N = \sum n_i$) then

$$\sigma_E^2 = \frac{\sum_i d_i^2 \, \sigma_{Ei}^2}{(\sum d_i)^2} + \frac{\sigma^2(O,s)}{N} \qquad [10.19]$$

So the total estimation variance is the composition of a slice term that accounts for the extension of the line sections to the slice and a line term

$$\frac{\sigma^2(O,s)}{N} \qquad [10.20]$$

that accounts for the error made when extending the samples to the line section. We shall now go on to describe how to calculate the estimation variance when the whole of the zone is not known to be mineralized.

10.6 When the limits of orebody are not known a priori

In some cases the limits of the orebody are not known beforehand from the geology. So they have to be determined from the drillhole data as information becomes available. This uncertainty about the geometry of the orebody introduces a second source of error, called the geometric error, which must then be added to the one described previously.

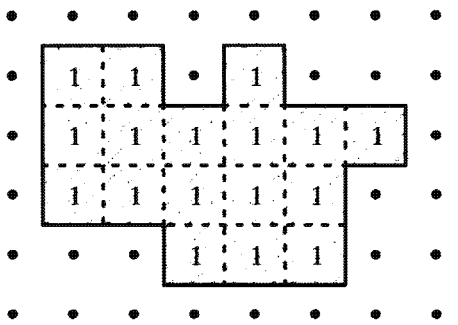


Fig. 10.5. Data layout showing mineralized holes and waste ones

Fig. 10.5. shows a region containing 48 drillholes. Seventeen hit the orebody (indicated by a 1), the rest did not. The simplest way to define the extent of the orebody is as the union of all the grid squares with positive results.

Matheron (1971 a and b) developed a formula for the ratio of the geometric error to the square of the mineralized area A:

$$\frac{\sigma_A^2}{A^2} = \frac{1}{N^2}\left[\frac{N_2}{6} + 0.061\,\frac{(N_1)^2}{N_2}\right] \qquad [10.21]$$

where N is the number of positive samples (here 17). The parameters N_1 and N_2 are found by counting the vertical and horizontal sides in terms of grid squares. To be more exact N_1 and N_2 are obtained by dividing the total number of grid squares in each direction by 2. Note N_1 must be greater than or equal to N_2. In the example there are 12 horizontal sides and 10 vertical ones. So $N_1 = 6$ and $N_2 = 5$, giving

$$\frac{\sigma_A^2}{A^2} = \frac{1}{17^2}\left[\frac{5}{6} + 0.06\,\frac{(6)^2}{5}\right] \qquad [10.22]$$

This turns out to be $(6.6\%)^2$, which is not very high.

One point to note when counting N_1 and N_2, is that all the indentations in the border must be counted, including the perimeter of internal holes. Fig. 10.6. illustrates this point. Lastly the geometric error must be incorporated into the estimation variance. Some care is required when doing this. For details see the worked examples given in Journel and Huijbregts (1978) pp428 – 438. A similar formula exists for 3D cases.

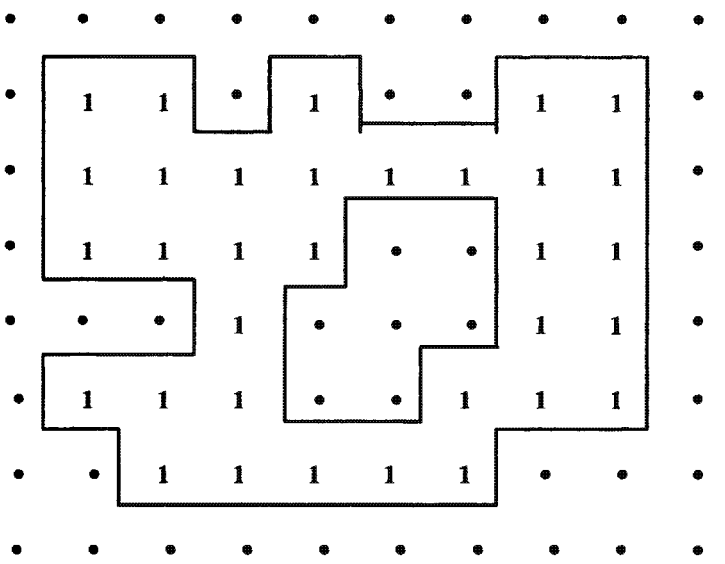


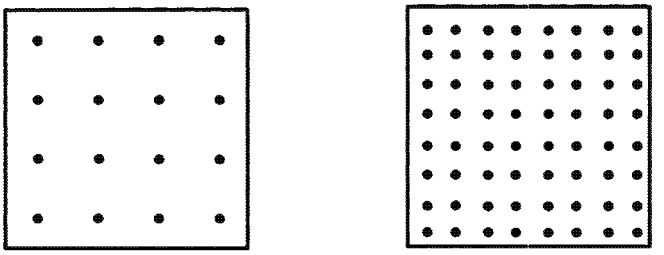
Fig. 10.6. Data layout showing mineralized holes and waste ones

Fig. 10.6. shows a configuration with a waste zone in the center of the ore. This complicates the counting of N_1 and N_2. Working in the vertical direction first, the total number of grid squares in the outer border is 16 whilst that of the hole is 6, giving a total of 22. Similarly parallel to the horizontal direction, the total perimeter is 26 (20 + 6). This means that N_1 is 13 and N_2 is 11. The rest of the calculation follows as before.

10.7 Optimal sampling grids

Once we know how to calculate the estimation variance we can work out the optimal drilling grid for a particular variable. Here "optimal" means the grid which gives the required estimation variance for the fewest drillholes (or samples), and hence at the lowest cost. For an isotropic deposit, a regular square grid should be used. When there is anisotropy, the ratio of the spacing between samples along the axes should be in proportion to the ratio of anisotropy.

Suppose that we want to estimate the average seam thickness and the average grade (here we will consider the sulphur content in coal) in a 4 km x 4 km area. For argument's sake the seam is assumed to exist throughout the area and the regionalization is isotropic. It would be possible to drill 16 holes on a 1km grid, or 64 on a 500m grid and so on, as shown by Fig. 10.7.



1000m grid 500m grid

Fig. 10.7. Two possible drilling grids

Suppose that the thickness variogram is spherical with a range of 1500m and a sill of 1.0 and that the sulphur variogram is also spherical with a range of 200m and a sill of 0.4.

10.7.1 For the 1km grid

The extension variances can be worked out using tables or a small program. Here the values were found to be:

$\sigma_E^2(0|v) = 0.27$ for thickness
$= 0.38$ for sulphur

So the estimation variance is:

$\sigma_E^2 = 0.27/16 = 0.0169$ for thickness
$= 0.38 / 16 = 0.0238$ for sulphur

10.7.2 For the 500m grid

These calculations were repeated for the 500m grid giving:

$\sigma_E^2 = 0.0020$ for thickness
$= 0.0055$ for sulphur

To make meaningful comparisons between the estimation variances for two variables, the coefficients of variation were calculated. That is, the standard deviations were divided by the mean. It is often helpful to plot the coefficient of variation against the number of drillholes on a bi–logarithmic scale.

To get a rough idea of the accuracy, the interval, the mean ± twice the standard deviation, can be taken as an approximate 95% confidence interval. For the 1000 m grid, the thickness values are accurate to ± 6.6% whereas the sulphur values are only accurate to ± 61.2%. These values reflect the ranges of the two variogram models (1500m compared to 300m). Consequently even if this sampling grid was optimal for thickness, it would not be anywhere near precise enough for sulphur.

Table 10.6. Relative precision for sulphur content and seam thickness for the two possible drilling grids

	Grid Size	
	1 000 m	500 m
Thickness	0.033	0.011
Suplhur Content	0.309	0.148

The optimal drilling pattern also depends on the size and, to a lesser extent, on the shape of the region to be estimated. Above all, the dominant factor determining the estimation variance is the total number of drillholes (samples) rather than the distance between them. Consequently although a 1000 m grid might be optimal for one variable for a certain area, it would not be dense enough for a smaller area and would be unnecessarily expensive for a larger one.

When the limits of the orebody are not known at the outset of the campaign, it is best to drill on a fairly wide grid at first and to infill as the limits become known. In that case the estimation variance and the optimal spacing cannot be determined at the outset since the effect of the geometric error depends on the size and shape of the mineralized area.

10.8 Exercises

Ex 10.1 Fifty vertical holes were drilled on a regular 200m x 200m grid to estimate the global reserves for a coal seam in a zone 1000m x 2000m that is known to be mineralized. The variogram for seam thickness is a spherical with a range of 500m and a sill of 3.25. Which approximation principle should be used to calculate the global estimation variance? Calculate its value.

The variogram for ash content consists of a spherical with a range of 250m and a sill of 4.65 plus a nuggest effect of 1.35. Calculate the global estimation variance for ash content too.

As only 50 drillholes are involved, it would have been possible to have used kriging. You can recalculate the estimation variances in this way and compare the values obtained with those given by the approximation principles.

Ex 10.2 Eighty vertical holes were drilled on a regular 200m x 200m grid to estimate the global reserves for an alluvial gold deposit in a zone 1600m x 2000m that is known to be mineralized. The variogram for the gold accumulation is an anisotropic spherical model with a sill of 5. The anisotropy is geometric with the longest range of 500m in the NS direction and with the shortest range of 100m in the EW direction. Which approximation principle should be used to calculate the global estimation variance? Calculate its value.

Ex 10.3 In a seismic survey of a square area 4km x 4km, the variable of interest was measured every 10m along 4 profiles. Calculate the global estimation variance, given that the variogram is an exponential model with a range of 500m and a sill of 10.1 plus a nugget effect of 3.9.

Ex 10.4 A mining company is carrying out an exploration campaign to determine the limits of mineralization for an alluvial deposit. They have dug 80 pits on a regular 50m x 50m grid (Fig. 10.8.). The 1s indicate pits that hit mineralization while the 0s shows nonmineralized ones. The geologist in charge of exploration wants to calculate the variance of the geometric error. Work out the values of N, N_1 and N_2, and hence calculate the variance.

Note that this figure is almost the same as Figure 10.6., except that the central waste zone has been removed. Compare the value obtained here for the geometric variance with the one obtained earlier.

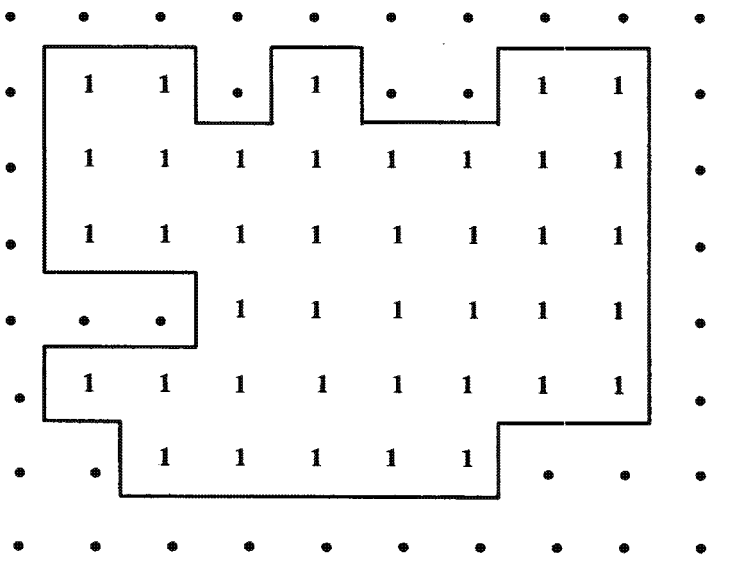


Fig. 10.8. Location of exploration pits. The 1s show pits that hit mineralization

Appendix 1: Review of Basic Maths Concepts

A1 What maths skills are required in linear geostatistics?

Four types of basic mathematical skills are needed in linear geostatistics (i.e. up to variograms and kriging):

1. being able to calculate the means and variances of random variables (and later of regionalized variables),
2. being able to use single and double summations rather than write out long lists of variables
3. being able to differentiate in order to find the minimum of a function, and
4. being familiar with matrix notation for simultaneous linear equations.

As the third and fourth topics are common in mathematics, readers will have no trouble finding suitable textbooks on these. So we will only review the first two points.

A1.1 Means and variances

In geostatistics we use linear combinations (i.e. weighted moving averages) of the data to estimate the values of the variable at points or the averages over blocks. For example, a typical linear combination is

$$Z_0^* = \lambda_1 Z_1 + \lambda_2 Z_2 + \ldots + \lambda_{10} Z_{10}$$

Kriging optimizes the choice of the weights by minimimzing the estimation variance. So we need to be able to express variances as a function of the weights, $\lambda_1, \lambda_2, \ldots \lambda_{10}$, and later, of the variogram model as well. To start with you need to know the mean and variance of the linear combination:

$$E[Z_0^*] = \lambda_1 E[Z_1] + \lambda_2 E[Z_2] + \ldots + \lambda_{10} E[Z_{10}]$$

$$\begin{aligned} Var[Z_0^*] = {} & \lambda_1^2 \, Var[Z_1] + \lambda_2^2 \, Var[Z_2] + \ldots + \lambda_{10}^2 \, Var[Z_{10}] + \\ & + 2\lambda_1 \, \lambda_2 \, Cov[Z_1, Z_2] + \ldots\ldots + 2\lambda_9 \, \lambda_{10} \, Cov[Z_9, Z_{10}] \end{aligned}$$

These results are well known in statistics. The proofs can be found in statistics textbooks. From there we go on to calculating the mean and variance of the estimation error. If our aim is to use the linear combination given earlier to estimate the value of the variable at point x_0, then the estimation error is

$$Z_0^* - Z_0 = \lambda_1 Z_1 + \lambda_2 Z_2 + \ldots + \lambda_{10} Z_{10} - Z_0$$

This is just the previous linear combination with an extra term (with a weight equal to –1) added. So its mean and variance are

$$E[Z_0^* - Z_0] = \lambda_1 E[Z_1] + \lambda_2 E[Z_2] + \ldots + \lambda_{10} E[Z_{10}] - E[Z_0]$$

$$\begin{aligned} \mathrm{Var}[Z_0^* - Z_0] = {} & \lambda_1^2\ \mathrm{Var}[Z_1] + \lambda_2^2\ \mathrm{Var}[Z_2] + \ldots + \lambda_{10}^2\ \mathrm{Var}[Z_{10}] + 1^2 \mathrm{Var}[Z_0] \\ & + 2\lambda_1\ \lambda_2\ \mathrm{Cov}[Z_1, Z_2] + \ldots\ldots + 2\lambda_9\ \lambda_{10}\ \mathrm{Cov}[Z_9, Z_{10}] \\ & - 2\lambda_0\ \lambda_1\ \mathrm{Cov}[Z_0, Z_1] - \ldots\ldots - 2\lambda_0\ \lambda_{10}\ \mathrm{Cov}[Z_0, Z_{10}] \end{aligned}$$

In Chapters 2 and 3 we will see how to evaluate these variances for regionalized variables rather than random variables by using the variogram or the spatial covariance to take account of their layout in space. Then in Chapter 7 we will see how to minimize these variances.

A1.2 Single and double summations

The equations for the expected value (mean) and the variance are long and unwieldy. It would be much better to have a shorthand way of writing them without having to list all the terms. Summations were invented to do this. The Greek letter Σ (pronounced sigma) is used to denote a summation. Using this convention, the expected value can be written as a single summation:

$$\begin{aligned} E[Z_0^*] &= \sum_{i=1}^{10} \lambda_i\ E[Z_i] \\ &= \lambda_1 E[Z_1] + \lambda_2 E[Z_2] + \ldots + \lambda_{10} E[Z_{10}] \end{aligned}$$

In a similar way, the sum of the variance terms can be written as

$$\sum_{i=1}^{10} \lambda_i^2\ \mathrm{Var}[Z_i] = \lambda_1^2\ \mathrm{Var}[Z_1] + \lambda_2^2\ \mathrm{Var}[Z_2] \ldots + \lambda_{10}^2\ \mathrm{Var}[Z_{10}]$$

Lastly we need an abbreviated way of writing the sum of all the covariance terms. As each one contains two subscripts, a double summation over two variables is required:

$$\sum_{i=1}^{9} \sum_{j>i}^{10} \lambda_i\ \lambda_j\ \mathrm{Cov}(Z_i, Z_j) = 2(\lambda_1\ \lambda_2\ \mathrm{Cov}[Z_1, Z_2] + \ldots\ldots + \lambda_9\ \lambda_{10}\ \mathrm{Cov}[Z_9, Z_{10}])$$

More care is required with the double summation. Some exercises on single and double summations are provided to allow you to get familiar with them.

A1.3 Exercises using summations

Ex 1. Expand the summations given below:

$$\sum_{i=1}^{5} \lambda_i Z_i \quad , \quad \sum_{i=1}^{N} \lambda_i^2 \, \mathrm{Var}[Z_i] \quad , \quad 1/N \sum_{i=1}^{N} Z_i \quad , \quad \left[\sum_{i=1}^{N} \lambda_i \right]^2 \quad , \quad \sum_{n=1}^{4} \phi^n \, H_n(y)$$

Ex 2. Expand the following double summations:

$$\sum_{i=1}^{5} \sum_{j=1}^{5} \lambda_i \lambda_j \gamma_{ij} \quad , \quad \left[\sum_{i=1}^{N} \lambda_i \right] \left[\sum_{i=1}^{N} \lambda_j \right] \quad , \quad \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j$$

Ex 3. Show that:

$$\left[\sum_{i=1}^{N} \lambda_i \right]^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j = \sum_{i=1}^{N} \lambda_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j>i}^{N} \lambda_i \lambda_j$$

Ex 4. Show that:

$$\mathrm{Var} \left[\sum_{i=1}^{N} \lambda_i Z_i \right] = \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j \, \mathrm{Cov}[Z_i, Z_j]$$

Appendix 2: Due Diligence and its Implications

A2.1 Stricter controls on ore evaluation

Over the recent years many mining projects have failed, and the lending institutions and shareholders have lost their money. So the stock exchange regulations and in some countries the corporate law have been made much stricter. One of the common causes of project failure has been inaccurate estimations of the reserves, generally overly optimistic appraisals. In some cases those who have lost their money have sued the initial company to get their money back. This can lead to the expert who has done the ore evaluation being called to justify his results. As can be seen from the article by Williamson–Noble and Lawrence (1994), in Australia errors or misrepresentations by an expert in public statements such as in share prospectuses are severely punished: a fine of up to $20,000 plus 5 years jail for the individual. The only legally acceptable defence is "due diligence".

A2.2 Due diligence

Due diligence means appropriately attentive care. Experts have to be able to show that they carefully checked the input data and used background information (such as geological interpretations) in a suitable way, and that they carried out the study in accordance with a suitable set system of procedures with appropriate quality controls. But what does this mean for geostatisticians?

Before giving some guidelines as to this, we should remind everyone that when a mining project fails, the control of the company generally passes from the directors to legally appointed receivers who are on the lookout for scapegoats. The inability to provide clear documentary evidence as to how the study was carried would make it very difficult for a consultant to argue that he worked with "due diligence". This is why we insist that students at Fontainebleau maintain a **logbook** for their case–studies.

A2.3 The logbook

(1) The logbook should be a bound volume not a spiral bound exercise book. Preferably the pages should be numbered. You should write in black ink rather than blue ink or pencil.

(2) Never tear pages out of your logbook. If you make a mistake, cross out the relevant section in red, put the date of the change and write down the reasons.

(3) Each day as you start work you should write the date (including the year) at the point where you start work. Then you note each computer operation that you carry out. For example when calculating an experimental variogram, note the exact set of data used including the selection name, the variogram lag, the angles chosen, the tolerances (angular and distance), etc. so that you could exactly duplicate the variogram calculation at a later date – even 5 years later. It is not uncommon for the Centre to get requests from mining companies to provide them with duplicate copies of data files, working notes or reports which they themselves have lost. Up until now we have succeeded in doing this.

(4) At the outset of the study you will have been given a lot of information about the orebody (maps showing drillhole collar locations, geological sections, reports on the geologists' interpretation,...). Carefully note exactly what information was provided in your logbook and again at the beginning of your report; e.g. the maps sheet numbers, their dates etc. The geological interpretation of the deposit changes with time, so you should note what it was when you start the study and any subsequent modifications during the study.

(5) Sometimes during the study, changes will be made to the data set. Points can be eliminated or new ones added. Note these in your logbook and send the client a fax or a memo confirming what you have done. Phone messages are easily forgotten or lost. Take the precaution of confirming in writing. Glue a photocopy of the fax/ memo into your logbook. The original (or a copy) should be retained in the standard filing system for outgoing letters.

(6) During a study you will accumulate lots of computer printouts; some are useless, others are very important. Some of these (the statistics on data sets) should be glued into your logbook because this helps specify exactly which data were actually used. The other important ones should be carefully stored (with their dates and maybe a note in the corner) in a file. Be sure they do not get crushed so that you can photocopy them for the final report.

These are just some suggestions on how to set up and organize a logbook. The list is by no means exahaustive. If you come across any points that have been missed, I would appreciate hearing about them.

References

Armstrong, M. (1980) Applications de la géostatistique aux problèmes de l'estimation du charbon, Thèse de Doct–Ing ès Sciences et Techniques Minières, Option Géostatistique ENSMP, 148pp.

Armstrong, M. (1984) Common problems seen in variograms. Maths Geol. 16(3) pp305–313.

Armstrong, M. and Delfiner, P. (1980). Towards a more robust variogram: a case study on coal. Unpublished Note N–671, CGMM, Fontainebleau, 49 pp.

Armstrong, M. and Diamond, P. (1984) Testing variograms for positive–definiteness. Maths Geol 16 (4), pp407–421.

Armstrong, M. and Dowd, P. (1994) Geostatistical simulations, Proceedings of a workshop, Fontainebleau, France, 27–28 May, 1993. Kluwer Academic Press, Dordrecht, Holland 255pp.

Baafi, E.Y. and Schofield, N.A. (1997) Geostatistics Wollongong 96, Kluwer Academic Press, Dordrecht, Holland 1333pp.

Barnes, R.J. and Johnson, T.B. (1984) Positive kriging, in Geostatistics for natural resources characterization, ed G Verly et al, proceedings of NATO ASI Lake Tahoe September 1983, pp231–244

Blackwell G.H. & Johnston T.G. (1986) Short and long term open pit grade control, in Ore reserve estimation: methods, models and reality, ed M David, R. Froidevaux, A.J. Sinclair and M. Vallee, CIMM Montreal Canada pp108–129.

Cressie, N. and Hawkins, D.H. (1980). Robust estimators of the variogram: I Maths Geol Vol 12, 2 pp 115 – 126.

Chauvet, P. (1982) The variogram cloud. In 17th APCOM Symposium, pp757–764.

Chauvet, P. (1988) Réflexions sur les ponderateurs négatifs du krigeage. Sci Terre, Inf. Géologique 28, pp65–113.

Chica–Olmo, M. and Laille, J–P. (1984). Constructing a numerical model to simulate a multi–seam lignite deposit, in Geostatistics for natural resources characterization, ed G Verly et al, proceedings of NATO ASI Lake Tahoe September 1983, pp1001–1013.

Clark, I. (1982) Practical geostatistics, Applied Science Publishers, 129pp.

David M. (1973) Variance d'estimation exacte et approchée d'une ligne par N points en schéma sphérique, Bol Geostadistica 5:1–4.

David M. (1977) Geostatistical ore reserve estimation, Elsevier, Amsterdam 364pp

David, M. (1988) Handbook of applied advanced geostatistical ore reserve estimation, Elsevier, Amsterdam. 216pp.

Delfiner, P. (1976). Linear estimation of non-stationary phenomena. Proc NATO ASI Rome 1975 "Advanced geostatistics in the mining industry" ed M. Guarascio et al. Reidel Pub. Co., Dordrecht, Holland pp 49 – 68.

Deraisme J. and de Fouquet C. (1984) Recent and future developments in down–stream geostatistics, in Geostatistics for natural resources characterization, ed G Verly et al, Proceedings of NATO ASI Lake Tahoe September 1983. pp979–999.

Dimitrakopoulos, R. (1994) Geostatistics for the next century, Kluwer Academic Press, 497pp.

Gomez, M. and Hazen, K. (1970) Evaluation of sulphur and ash distribution in coal seams by statistical response surface regression analysis. Report of Investigation 7377, Washington U.S. Bureau of Mines. 120 pp.

Hu, L.Y., Joseph, P. & Dubrule O. (1994) Random genetic simulation of the internal geometry of deltaic sandstone bodies, SPE "Formation evaluation" SPE 24714, Dec 1994.

Isaaks, E.H. and Srivastava, R.M. (1989) An introduction to applied geostatistics, Oxford University Press, New York 561pp.

Jacod, J. and Joathon, P. (1970a). The use of random genetic models in the study of sedimentary processes. Internal Note CGMM N–202.

Jacod, J. and Joathon, P. (1970b). Estimation des paramètres dans les modèles aléatoires–génétiques représentant des processus de sédimentation. Internal Note N206, CGMM, Fontainebleau.

Journel, A.G. (1977) Kriging in terms of projections, Maths Geol Vol 9 No 6 pp563–586.

Journel, A.G. and Huijbregts, C.J. (1978). Mining geostatistics. Academic Press 600 pp.

Journel, A.G. (1983) Nonparametric estimation of spatial distributions, Math Geol, Vol 15,3, pp445–468.

Krige, D.G. (1978). Lognormal and de wijsian geostatistics for ore evaluation Geostatistics Series S.A.I.M.M.

Krige D.G. and Magri E.J. (1982) Geostatistical case studies of the advantages of lognormal–de wijsian kriging with mean for a base metal mine and a gold mine, J Math Geol, Vol 14, no 6, pp 547–555.

Lantuéjoul C. (1994) Nonconditional simulation of stationary isotropic multi–gaussian random functions, in Geostatistical simulations ed M. Armstrong and P. Dowd, Kluwer Academic Publishers Dordrecht pp147–177.

Marechal, A. (1975) Forecasting a grade–tonnage distribution for various panel sizes, in Proceedings 13th APCOM Symposium, Technische Universitat Claustal, Germany.

Matheron, G. (1963). Principles of geostatistics, Economic Geology, Vol 58, pp 1246–66.

Matheron, G. (1965) Les variables régionalisées et leur estimation. Paris, Masson. 306pp.

Matheron, G. (1971a). La théorie des variables régionalisées. Les Cahiers du Centre de Morphologie Mathématique, Fasc 5, Paris ENSMP 212pp.

Matheron, G. (1971b). The theory of regionalized variables (English translation) Les Cahiers du Centre de Morphologie Mathématique, Fasc 5, Paris ENSMP 212pp.

Matheron, G. (1972). Leçons sur les fonctions aléatoires d'ordre 2, Cours de l'Ecole des Mines C–53 Mars 72.

Matheron, G. (1973). The intrinsic random functions and their applications. Adv. in applied Prob. Vol 5, pp 439 – 468.

Matheron, G. (1976) A simple substitute for conditional expectation: disjunctive kriging, In Proceedings of the First NATO Study Inst, Advanced geostatistics in the mining industry, D. Reidel Publishing Co, pp221–236.

Ravenscroft, P.J. and Armstrong, M. (1985) Kriging of block models, the dangers re–emphasized, In Proceedings of APCOM, Berlin pp577–587.

Raymond, G.F. and Armstrong W.P. (1986). Short and long term open pit grade control, in Ore reserve estimation: methods, models and reality, ed M David, R. Froidevaux, A.J. Sinclair and M. Vallee, CIMM Montreal Canada pp 65–107.

Renard, D. and Yancey, J. (1984) Smoothing discontinuities when extrapolating using moving neighbourhoods, in Geostatistics for natural resources characterization, ed G Verly et al, proceedings of NATO ASI Lake Tahoe September 1983. pp679–690.

Rendu, J.M. (1979) Kriging, logarithmic kriging and conditional estimation: comparison of theory and actual results, 16th APCOM, ed. T.J. O'Neil, Society of Mining Engineers of AIME pp109–212.

Rivoirard, J. (1984) Le comportement des poids de krigeage. Thèse de Doct–Ing ès Sciences et Techniques Minières, ENSMP, 72pp.

Rivoirard, J. (1987) Two key parameters when choosing the kriging neighbourhood, Math. Geol. 19 (8) pp851–856.

Rivoirard, J. (1994) Introduction to disjunctive kriging and non–linear geostatistics, Oxford University Press, 180pp.

Royle, A.G. (1979) Estimating small blocks of ore, how to do it with confidence, World Mining Congress.

Sabourin, R. (1975). Geostatistical evaluation of sulphur content in Lingan Coal Mine, Cape Breton. Proc 13th Int APCOM Symposium, Claustal pp 1 – 11, 1 – 16.

Sandjivy L. (1980). Etude géostatistique des données recueillies sur le mil au Niger. CGMM Internal Note N–656, Fontainebleau.

Serra, J. (1968). Morphologie mathématique et génèse des concrétions carbonates des minerais de fer de Lorraine. Sedimentology, 10, pp 183 – 208.

Soares, A. (1993) Geostatistics Troia 92, Kluwer Academic Press, Dordrecht, Holland, 1079pp.

Wackernagel, H (1995) Multivariate geostatistics, Springer Verlag Berlin 256pp.

Williamson–Noble, J.C. and Lawrence, M.J. (1994) Individual personal liability risk management in openpit mining, In Proceeding of 4th Large Openpit Mining Conference, Perth, Australia, 5–9 Sept, AusIMM pp273–280.

Wood, I.D. (1976). Geostatistical investigation of no. 2 seam in the Witbank area. Report No. 2, South African Chamber of Mines, Report no 99/76, 15 pp.

Index

Author Index