OPTIMUM MAPPING TECHNIQUES USING REGIONALIZED VARIABLE THEORY

R. A. OLEA

OPTIMUM MAPPING TECHNIQUES

USING REGIONALIZED VARIABLE THEORY

BY

RICARDO A. OLEA[1]

1975

reprinted 1983

[1]*Visiting Research Scientist, 1974-75, Kansas Geological Survey and Staff Scientist, Empresa Nacional del Petrôleo, Santiago, Chile.*

CONTENTS

PAGE

PAGE

PREFACE

Much of the material presented in this report is part of a previous work (Olea, 1972), revised and enlarged during 1974-1975 while the author served as Visiting Research Scientist with the Kansas Geological Survey. This report describes the methods of universal kriging and drift as applied to mapping and map analysis by computers. Based on the theory of regionalized variables developed by G. Matheron, this work contains a detailed derivation of all steps leading from the theory to its practical use. However, a basic knowledge of geologic exploration, numerical analysis, and familiarity with statistics is assumed.

The computer algorithms to perform universal kriging and drift analysis are implemented as FORTRAN IV modules in the SURFACE II graphics system developed by the Kansas Geological Survey. Operational instructions for the programs are contained in the SURFACE II User's Manual by Robert J. Sampson (1975). Inquiries concerning these programs should be sent to the Geologic Research Section, Kansas Geological Survey, 1930 Avenue "A", Campus West, Lawrence, Kansas 66044.

ACKNOWLEDGMENTS

I am grateful to my company, Empresa Nacional del Petróleo, Chile, for releasing the data used as examples and for giving me the opportunity to join the Kansas Geological Survey. A special thanks is due to Dr. John C. Davis, Kansas Geological Survey, for encouragement and help during preparation of the manuscript. Finally, I would like to acknowledge the valuable comments of Drs. John H. Doveton and John C. Tipper of the Kansas Geological Survey.

RICARDO A. OLEA

Lawrence, Kansas

June 1975

ABSTRACT

The theory of regionalized variables has been developed by G. Matheron to allow the drawing of statistical inferences, considering not only sample values but also implicit relationships in the geometry of the sample space. Two aspects of the theory are analyzed in this report; universal kriging and drift. Universal kriging is an estimation procedure presented as an alternative for grid generation in the automatic contouring of point observations. Drift estimations relate to the search for a slowly varying spatial component such as that usually studied by trend surface analysis. Both universal kriging and drift estimation are more general, theoretically satisfying, and reliable than the empirical methods which are now used. In addition, provided the assumptions of the regionalized variable theory are met, the resulting estimates are unbiased with minimum estimation variance. Most importantly, the methods yield the variance of the estimation error at every point in the sample space.

Basic concepts of the regionalized variable theory which pertain to automatic contouring are expressed in this report in terms of classical statistical inference. Because published accounts of the theory are highly compact, it is necessary to provide extensive derivations of the steps in its development. These lead from the fundamental assumptions of the theory to the algorithms required to perform punctual kriging and to estimate the drift, and provide an explanation of the properties of the estimates. The practical utility of these methods is shown in an example using data on a subsurface geologic horizon. Additional proofs are contained in appendices.

OPTIMUM MAPPING TECHNIQUES USING REGIONALIZED VARIABLE THEORY

R. A. OLEA, *Kansas Geological Survey* and *Empresa Nacional del Petróleo*

CHAPTER 1 -- INTRODUCTION

1.1 Review

Earth scientists usually deal with complex phenomena which are the results of the interaction of scores of variables, through relationships which are in part unknown and in part very complex. The final outcomes that are observed in nature are therefore functions whose known values do not follow any simple deterministic function over a wide range. However, the observations exhibit continuity and cannot be regarded as pure random values. Professor Georges Matheron, a French engineer, applied the name "regionalized variables" to these functions. He developed a statistical theory based on a simple assumption of stationarity, which made intensive use of the correlation between spatially non-independent random variables. The theoretical model resembles a stochastic process in the sense used by Wiener (1949). The height of the surface of the Earth, the elevation of a formation top, the amount of precipitation in an area, porosity in a reservoir, and ore content in a mineralized body are all examples of regionalized variables.

The origins of the regionalized variable theory are rooted in ore estimation problems. In the early 1950's in South Africa, D. G. Krige realized that he could not accurately estimate the gold content of mined blocks or stopes without considering the geometrical setting of the samples, such as sample locations and sizes. However, the conclusions of Krige and others remained as isolated empirical results valid only for gold estimation in the Witwatersrand mining district. Matheron expanded Krige's empirical observations into a theory of the behavior of spatially distributed variables which was applicable to any phenomenon satisfying certain basic assumptions, and which was not limited by their physical nature. Today, this theory has been used for ore reserve estimation in more than forty types of ore deposits in ten different countries (Blais and Carlier, 1968). In addition, it has found applications in areas not related to mining, such as studies of gravimetric anomalies (Huijbregts and Matheron, 1971); submarine mapping (Journel, 1969); meteorology (Delhomme and Delfiner, 1973); and forestry (Poissonnet, Millier, and Serra, 1970). Matheron not only generalized Krige's estimation methods but also developed a theoretical understanding of sampling variability associated with the physical size of samples, and a complete theory of estimation error. In 1965, he published the results of more than ten years' work in a highly theoretical (and hence difficult) book which is the cornerstone of the regionalized variable theory (Matheron, 1965). He has later refined and enlarged the theory in several papers and in a collection of textbooks; unfortunately, these works are equally difficult to understand (Matheron, 1969a, 1969b,

1970). Most of Matheron's work and that of his students has been published in French, which has presented an additional translation problem for many researchers.

To honor Krige for this pioneering work and to differentiate his estimation method from others, Matheron refers to his technique as "krigeage" which has been translated into "kriging" in the English-language literature. A later development is called "universal kriging," an estimation method applicable to a variable which is not spatially stationary, or in other words, which contains a trend. Matheron, considering that the term "trend" had been misleadingly used to refer to different phenomena, introduced the word "drift" to denote those slowly varying components which underlie regionalized variables. A more formal definition of drift can be found in Section 2.3.

Mapping is the most appropriate way to represent the spatial variation of a regionalized variable and its drift in two dimensions. In this work I will not present a complete review of this theory, but will consider a reduced and simplified subset concerned with the specific problem of map analysis of punctual samples, such as elevations of the top of geologic horizons, measured in the subsurface. However, I will develop the necessary aspects of the theory from its beginnings because this theory has been largely overlooked and misunderstood in the American literature (Watson, 1971, 1972; Whitten discussion, in Krige, 1966).

In addition to the barrier presented by the French language, the theory has been obscured by an unfamiliar jargon. In the following pages, as necessary, I will introduce many terms which are not part of traditional statistical nomenclature. In each instance, I will provide a definition or an explanation of unfamiliar terms or of certain conventional words that have been given an unusual meaning. The origin of many of these terms is not explained; they are simply embedded in the literature of regionalized variables. In some places, it might be better to use a more conventional word or an approach having equivalent implications, with the aim of reducing the jargon. I have not done so because reading the original references would then become more difficult. I have instead attempted to provide a guide to these other topics and details contained in the original sources cited in the bibliography.

1.2 Regionalized Variables

Definition 1.2.1: A *regionalized variable* is any numerical function with a spatial distribution which varies from one place to another with apparent continuity, but the changes of which cannot be represented by any workable function (Blais and Carlier, 1968). This definition characterizes many variables which describe natural phenomena such as the ore content of a mineralized body or the wheat production per acre of farmland.

Some common characteristics of regionalized variables are:

1) Localization: A regionalized variable is numerically defined by a value which is associated with a sample of specific size, shape, and orientation. These geometric characteristics of the sample are called the *geometric support*. A *geometric field* is that larger volume

from which the samples are drawn. The geometric field and the geometric support do not necessarily comprise volumes, but may instead be areas, lines, or time intervals. When the size of the geometric support tends to zero, we have a point or *punctual sample*, and the geometric support is immaterial.

If our samples are 10 cu. cm. of soil from a farmland and we wish to study variation in soil water saturation, the regionalized variable is soil moisture, the geometric supports are the 10 cu. cm. volumes, and the geometric field is the soil layer of the farm. A classical statistical experiment concerning the toss of coins has results which are independent of whether a penny, dime, franc, or pfennig is flipped. However, it is a common experience that water content of a soil sample depends not only on the location of the sample but also on the sample's size, orientation, and shape. A long, vertical, cylindrical core typically will contain more water than a horizontal sample of the same shape and size taken from the surface. If we test 10 cu. m ., the result will also differ from that obtained from a sample only 10 cu. cm. in volume. This is a very important fact which is often ignored in ordinary statistics, where the shape, size, and orientation of a sample are usually not considered. The theory of regionalized variables does take into account the geometry of the sample. It is no coincidence that this theory has been remarkably successful in ore estimation where small core samples are used to estimate the ore grade in mine blocks having radically different sizes (Krige, 1966; Matheron, 1969b). The theory also has significance for automatic contouring, although localization is immaterial as the samples are points having no volume.

2) Anisotropy: Some regionalized variables are anisotropic; that is, changes in value are gradual in one direction and rapid or irregular in another.

3) Continuity: The spatial variation of a regionalized variable may be extremely large to very small, depending on the phenomenon. Despite the complexity of the fluctuations, an average continuity is generally present.

If we collect soil samples of the same size, shape, and orientation at regular intervals along imaginary lines, we could find two different series, A and B:

A: 5, 10, 15, 20, 25, 20, 15, 10, 5 % water

B: 10, 25, 15, 10, 20, 5, 15, 5, 20 % water

Individual values in the two series are exactly the same. As a result, the mean, the variance, and the frequency histograms of A and B are rigorously identical. Any statistical analysis which does not consider anything other than the mean, variance, and frequency distribution will not be able to differentiate between such sequences. This example emphasizes the importance of measuring the spatial continuity of the regionalized variable. This continuity may be analyzed in several ways, but some kind of correlogram seems to be the most appropriate. In the regionalized variable theory, this role is assigned to the semivariogram.

1.3 Automatic Contouring

Any surface which is single-valued for any coordinate $\vec{x}$ and continuous everywhere may be graphically represented in the same way that a topographic map is a graphic abstraction of a part of the Earth's surface. The main element used in these graphic representations is the *isopleth*, a line connecting points of equal value. A group of isopleths conveniently spaced constitutes a representation of the form of a real surface by a map. To understand the meaning of an isopleth, we will consider a topographic map which uses a special type of isopleth called a contour line. A *contour line* is an imaginary line on the ground which takes any shape necessary to maintain a constant elevation above sea level. Each topographic contour line has a height associated with it. The verb "to contour" therefore has an intrinsic restriction in meaning and should strictly be applied only to topographic maps. However, because the techniques used to draw isopleths are not related to the variable being mapped, the word "contour" is used with the general meaning of construction of isopleth lines. We will use the terms *automatic mapping* or *automatic contouring* to denote the generation of maps by machine.

Maps are the most convenient tool for representing scattered observations of a regionalized variable. Mapping provides a method which simultaneously reduces three-dimensionally distributed observations to two-dimensional pictures, and provides a global view of surfaces which cannot be completely or accurately represented by analytically tractable formulas.

The classic way to produce a topographic map is to send a surveyor to the field to measure latitude, longitude, and altitude in as many points as possible, uniformly distributed across the area to be mapped. Subsequently, a draftsman plots the data points and draws the contour lines at the appropriate positions between the points. A serious disadvantage of this technique is that resolution is lost because it is necessary to limit the sample points to a finite collection out of the infinite number which actually constitutes the surface. Therefore, the draftsman must attempt to "create" information by inferring between control points; the results are more-or-less correct depending on the experience of the draftsman, the smoothness of the real surface and the control-point density. Manual production of contour maps is subjective, laborious, time-consuming, and the results are often inconsistent.

Today, topographic cartography has solved the problem of discrete sampling by using aerial photography. Instead of surveying locations and altitudes, pictures are taken from an airplane. Each picture is a continuous sample or partial view of the actual surface. Surveying is now secondary and not necessary for contouring, but only to present the map features in the correct size relationship to each other and to the Earth's surface. Stereomatic production replaces manual contouring, and the final map is practically error free.

Nevertheless, there are many other fields where it is theoretically or practically impossible to replace discontinuous sampling; the best that can be done is to replace the draftsman by a machine. This is the case in subsurface geology, for example, where the objective is to map the contact between two rock strata which are hundreds or thousands of meters underground.

The best information that can possibly be obtained comes from a finite collection of wells drilled through the contact. In oceanography there is still no way to photograph the bottom of the sea over a wide area. Discontinuous information obtained by sounding, by seismic prospecting, or by sonar is the only way to measure the sea floor, especially under deep water. In other applications, the surface may be a mathematical abstraction with no real existence and mapping of discontinuous sampling points seems to be the only way to visualize the funtion. This is the case when depicting air temperatures, for instance, where the function does not represent anything physically real in the same sense as the surface of the Earth.

Of the many applications of automatic contouring, we will later concentrate on the production of subsurface maps of rock strata. This is an area where maps of high quality are critically important. They are used to monitor oil exploration and exploitation as oil occurrence is closely related to structural features of potentially producing formations. Because this report introduces a methodology that yields improved maps, it seems reasonable to offer an application in a field where improvements are necessary and might be extremely valuable in the future. This does not mean the methods cannot be equally well applied to any surface which can be represented by isopleths.

In most general purpose schemes for automatic contouring from discontinuous samples, the following steps can be distinguished:

1) The computer reads the raw data to be contoured, map control specifications, titles, and other extraneous information to be placed on the map.

2) A rectangular array is automatically generated containing estimated values of the surface at the nodes of a regular grid. This step is usually referred to as "gridding."

3) The regular grid is used to produce the iscpleths, by interpolation between nodes.

4) The isopleths are drawn by a plotter or an output printer.

Use of a rectangular grid is dictated by considerations of convenience in computation and data manipulation and appears in most algorithms. The kernel of any contouring scheme is in the procedure used to generate the grid, as it consumes most of the computing time and the final results depend on its accuracy. There are some general and obvious criteria on which the effectiveness of a gridding algorithm may be judged:

1) The resulting isopleth should not contradict the data, i.e., the isopleth should be exact at least at the control points. Although this point seems obvious, it has been repeatedly violated.

2) Isopleths should be smooth curves, so they are aesthetically appealing and natural in appearance.

3) In some instances the requirement has been that the map should closely resemble that which would be produced by a draftsman using visual interpolation. Such a specification will not only satisfy 1) and 2), but might also make arbitrary changes where erroneous data are suspected.

4) Computational time must be reasonable, so that the cost of an automatically contoured map is competitive with that of a manually contoured map.

Criteria 3) and 4) deserve some elaboration. To try to simulate the human production rather than to model the surface itself is inadvisable, and attempts to do so represent a serious defect in several methods.

The impact of economic considerations on mapping procedures has been overestimated. For some applications, the cheapest is best, regardless of the quality of the final product. However, in some instances, automatically contoured maps two or three times more expensive than those produced by manual contouring may be better for the following reasons:

1) Maps may be produced to standardized specifications, both within a single map and among a series of maps. The variability and inconsistency inherent in manual drafting are avoided.

2) The possibility of error is significantly smaller using a computer for automatic contouring.

3) Under certain circumstances, the production of an automatically contoured map can be significantly faster.

4) It is possible to update maps at a low marginal cost.

5) As will be shown, automatic contouring can be used to evaluate the error or uncertainty which exists in any map made from a finite collection of samples.

1.4 Advantages of Regionalized Variable Methods

Two basic methods will be studied: automatic contouring of the original observations, and drift analysis. The term "original observations" does not necessarily mean the original data as it comes from the field. Some minor algebraic corrections may have been made to prepare the data, such as the subtraction of ground elevation. "Original observations" may also be derived from more basic data. For instance, the difference between elevations of different strata may be used to generate an isopach or thickness map. Similarly, the subtraction of a drift from Bouguer gravity produces a map of gravity residuals (local anomalies).

Methods described in the literature to generate grids for automatic contouring are mostly moving weighted average techniques. However, differences in the manipulation of formulas and data allow a separation of these algorithms into two groups:

1) A "true" weighted average, where a point is estimated as the weighted sum of a number of samples selected according to certain rules.

2) Polynomial interpolation over a reduced neighborhood of data.

Despite the vast amount of effort expended on gridding algorithms (Switzer, Mohr, and Heitman, 1964; Walters, 1969), all such methods have at least one of the following drawbacks:

1) They are highly empirical. There is no theory which forecasts what will happen using the methods under all possible circumstances. "In practice, they [the methods] have been found in a negative manner; that is, from rejection by users on different grounds as each obstacle has been overcome" (Palmer, 1969). In general, it is not possible to predict in

advance whether the results of a given mapping algorithm are going to be "satisfactory" with a given set of data.

2) The methods are not necessarily optimal. Although a given algorithm may work perfectly with some data as far as can be told, it cannot be stated that there is no other method which would work even better.

3) Contouring errors which arise from the use of discrete samples cannot be estimated. These may range from immaterial to substantial. But because contouring methods do not provide a measurement of the error, reality cannot be distinguished from fantasy. A potential advantage over manual contouring is therefore lost.

4) They lack generality. Weighted average techniques, in particular, assume some geometrical distribution of samples used to estimate a grid point (Switzer, Mohr, and Heitman, 1964). Some, for instance, suppose that the point to be estimated is in the center of gravity of an equilateral triangle. This criterion either limits the use of the method or introduces errors because approximations must be made where samples are irregularly distributed or arranged according to another geometric pattern. The converse problem is again a question of optimality. If samples are regularly distributed, there are always several geometric figures which can be chosen for the pattern of moving averages. The point to be estimated might be regarded as in the center of a square, but the point might also be regarded as in the center of two concentric circles.

5) Gridding algorithms are chosen arbitrarily, as are the parameters used in a given method. Suppose some kind of weighted average of nearest points is used. The "nearest points" may be defined, for example, as either all those in a two-km. radius neighborhood or alternatively, the six closest points regardless of distance. The choice is an arbitrary one. The method might work better using a four- or eight-km. radius, or perhaps computation time could be reduced by using only four nearby points without loss of precision. Looking for the most satisfactory parameters involves contouring the same data several times and making comparisons which are usually completely subjective.

6) Most techniques use the data at each individual sample point, but do not consider the implicit relationships among points. This is a waste of potentially valuable information.

7) The methods presume that the samples are points. However, many spatial variables have meanings only when associated with areas (population densities) or volumes (ore value). To ignore the fact that samples may have a physical size and that this size may be different for different groups of samples introduces errors. A direct comparison of maps obtained from samples having different physical characteristics is difficult and inappropriate.

8) Some methods fail to satisfy the elementary property that a grid point and a sample point should be identically the same if they are coincident. This is particularly true with methods using polynomial interpolation.

Universal kriging provides an answer which overcomes all of these handicaps at one time. As the technique of universal kriging is not just an algorithm to generate contour grids but a branch of the theory of regionalized variables, an optimal estimation is guaranteed provided

the data satisfy a set of hypotheses. The estimation will be optimal regardless of the sample distribution in the sense that the estimation is unbiased with a minimum estimation variance. Furthermore, universal kriging is an exact interpolation procedure, provides an estimate of the error at all points, and takes into account the fact that samples can have a volume. However, we will not take advantage of the latter property as we will deal exclusively with point measurements.

The drawbacks of trend surface analysis are equally severe. These include the following points:

1) A trend is the function resulting from the method of least squares; therefore, it is only a mathematical abstraction without physical meaning. The drift is defined as the expected value of the regionalized variable.

2) Trend surface analysis is not optimal. Least squares estimators are linear unbiased estimators with minimum estimation variance *only* if the observations are outcomes of spatially distributed random variables which are uncorrelated, have a common variance, and a mean given by the fitted expression. However, geologic variables characteristically are intercorrelated.

3) Trend surface analysis does not provide a measure of the error of estimation, as it is not in itself a statistical estimation method but rather a technique for obtaining a transformation of the data. This transform will in general vary at any point depending on the distribution of data points.

4) Trend surfaces are very unstable away from control points. There are usually severe edge effects that must be eliminated by visual inspection. However, there may still be misleading features in intermediate areas of poor control. This is a problem related to the lack of error estimation. Drift estimation is more stable in areas with critical control and the estimation variance can be used as a guide for discarding unreliable results.

5) The polynomial degree is the only parameter which must be selected in trend surface analysis. The least-squares surface is obtained by fitting a medium- to high-order polynomial to all sample points. In contrast, Matheron's methods fit a low-order polynomial, taking only those samples inside a circle many times smaller than the entire study area. The exact size of the circular neighborhood is not arbitrary, but is determined by a statistical analysis of the continuity of the data. This additional parameter, the neighborhood size, gives the drift more power and control over the results. The neighborhood size may be regarded as a type of filter which keeps the absolute sizes of local structures within certain limits.

6) Trend surface analysis uses the information contained in each separate observation, but wastes all the implicit information which can be obtained from relationships among sample points. Drift estimation makes use of these relationships.

7) Although we will deal only with point samples, it should be noted that trend surface analysis cannot consider the fact the samples may represent volumes of different shapes, sizes, and orientations. Drift estimation can take such factors into consideration.

1.5 Introductory Examples

We will denote by C the sample space of possible outcomes c_i. C_1, C_2, C_3, etc., will denote subsets of C. Suppose an experiment consists of randomly tossing a coin two times; then the sample space of possible outcomes is $C = \{c: c = HH, HT, TH, TT\}$, where T represents tails and H heads. A subset C_1 would be to obtain a head on the first trial and C_2 to obtain tails.

Because it is tedious to handle the symbology of heads and tails, we will redefine the experiment in terms of real numbers. Let X_1 be a single-valued, real-valued function with domain the sample space C and codomain the space of real numbers A. X_1 takes us from the sample space C to A. Such a function is called a *random variable* (Hogg and Craig, 1970).

In our example, A could be $A = \{x_1 : x_1 = 0, 1, 2\}$ and the assignment table for the random variable could be:

$$X_1(c_1) = X_1(HH) = 0 = x_1^1$$

$$X_1(c_2) = X_1(HT) = 1 = x_1^2$$

$$X_1(c_3) = X_1(TH) = 1 = x_1^2$$

$$X_1(c_4) = X_1(TT) = 2 = x_1^3$$

This random variable X_1 can be used to represent the number of tails occurring in our test sequence, for example.

Another random variable X_2 could be used to indicate whether two consecutive tosses were the same.

$$X_2(c_2) = X_2(c_3) = 0 = x_2^1$$

$$X_2(c_1) = X_2(c_4) = 1 = x_2^2$$

If we consider X_1, X_2 together, A is no longer $(x_1 : x_1 = 1, 2, 3)$ but the ordered pairs:

$$A = \{(x_1,x_2) : (x_1,x_2) = (0,1)(1,0)(2,1)\}$$

mentioning only those pairs which are possible events.

If A is a subset of A, we can define the probability of the event A as $P(A) = P[(X_1,X_2) \in A]$. Consider the subset $A = \{(x_1,x_2) : (x_1,x_2) = (0,1), (2,1)\}$. To compute P(A) we must consider all events in C for which the random variables X_1 and X_2 take values (x_1,x_2) which are elements of A. In our example, $X_1(c_1) = 0$ and $X_2(c_1) = 1$; $X_1(c_4) = 2$ and $X_2(c_4) = 1$. Therefore, P(A) = P(C) when $C = \{c: c = c_1 \text{ or } c_4\}$.

Suppose the probability function $P(c_i)$ assigns a probability 1/4 to each of the four elements of C. Then P(A), which can also be written as $P(X_1 = 1 \text{ or } 2, X_2 = 1)$, is 1/4 + 1/4 = 1/2. In a similar way, P(A) can be computed for each ordered pair (x_1,x_2) of A:

(x_1,x_2)	(0,1)	(1,0)	(2,1)
$P[(X_1,X_2) = (x_1,x_2)]$	1/4	1/2	1/4

We now have a complete notation for the probability distribution over each element of *A*. In this sense, we may speak of the distribution of the random variables X_1, X_2 meaning the distribution of probability. The function which determines the distribution $f(x_1,x_2)$ is called the *probability density function*. In this instance, the probability density function is:

$$f(0,1) = f(2,1) = 1/4$$
$$f(1,0) = 1/2$$
$$f(x_1,x_2) = 0 \text{ elsewhere}$$

As another simple example, suppose we collect three samples of soil along an imaginary line across the ground and measure the soil moisture. The sample space is now $C = \{c: c = c_i, i = 1, 2, 3, \ldots, 8\}$. If W denotes "wet" and D represents "dry," the eight elements of the sample space are c_1 = DDD, c_2 = DDW, c_3 = DWD, c_4 = DWW, c_5 = WDD, c_6 = WDW, c_7 = WWD, and c_8 = WWW.

To specify the outcomes in terms of a space of real numbers *A*, we may define three variables X_i, i = 1, 2, 3, where X_i is the state of the ith sample. The X_i's are single-valued, real-valued functions defined in the sample space *C* which takes us from the sample space to a space of ordered triples according to the following rule:

$$X_i(c_j) = \begin{cases} 0 & \text{if the ith sample of the jth outcome is dry} \\ 1 & \text{if the ith sample of the jth outcome is wet} \end{cases}$$

The following table defines the random variables:

	X_1	X_2	X_3
c_1	0	0	0
c_2	0	0	1
c_3	0	1	0
c_4	0	1	1
c_5	1	0	0
c_6	1	0	1
c_7	1	1	0
c_8	1	1	1

Therefore, the space of ordered triples is:

$$A = \{(x_1,x_2,x_3) : (x_1,x_2,x_3) = (0,0,0),\ (0,0,1),\ (0,1,0),\ (0,1,1),\ (1,0,0),\ (1,0,1),\ (1,1,0),\ (1,1,1)\}$$

We may assume $P(c_i) = 1/8$ for each event in C; under such conditions $P(c) = 1/8$ for any triple in A. Therefore, the probability density function is:

$$f(x_1,x_2,x_3) = \begin{cases} 1/8 & x_1,x_2,x_3 = 0,1 \\ 0 & \text{elsewhere} \end{cases}$$

So far, the example involving coin flipping is essentially the same as the collection of soil samples. However, because of oversimplifications, these examples may be far from reality and the results a mere mathematical game. If the coin is not correctly balanced, for instance, the probability of heads may be 0.4 instead of 0.5 and predictions based on elemental probability laws will be incorrect. When we try to consider refinements like this, we realize our two examples describe very different phenomena which cannot be handled in the same way, using the same model. Whereas classical statistics is adequate for complex specifications on coin tossing, it is insufficient to describe moisture, ore valuation, population density, rock porosity, topography, annual precipitation, or the like. In general, conventional statistical approaches are inadequate for the description of any variable from a natural phenomenon which has a spatial distribution. The theory of regionalized variables is a possible answer for such phenomena. This theory is not a completely new branch of statistics, but rather an extension of conventional statistics, particularly time series analysis.

CHAPTER 2 -- ASPECTS OF THE REGIONALIZED VARIABLE THEORY

2.1 Introduction to the Intrinsic Theory

The theory of regionalized variables is very wide and will not be pursued in all aspects here, since the application of regionalized variables to map analysis requires only a very restricted part of the whole theory. Discussions will be confined to those topics absolutely necessary for this purpose at the simplest level of generalization. Hopefully, this presentation will be a useful introduction to a theory which has been largely ignored in the English-language literature.

The theory of regionalized variables has two branches, the *transitive methods* and the *intrinsic theory* (Matheron, 1965). The first is a highly geometrical abstraction without probabilistic hypotheses and has little practical interest. It will be ignored because it is not necessary for our purposes. The practical counterpart of these geometrical abstractions is the *intrinsic theory*, which is a term for the application of the theory of random variables to regionalized variables.

The initial objective of the intrinsic theory was to make more reliable estimations of ore grade, thus improving mine exploitation. In this application, the geometric support is a volume and complex mathematics are required. In our application, the samples will be point observations and simplifications of the mathematics are possible. For generality, many published discussions of the regionalized variable theory are expressed in terms of the most complex formulas and simplifications do not appear until the end. Since this is not a general presentation of the intrinsic theory, I prefer to sacrifice generality at the beginning in order to gain clarity from the outset. I will intentionally avoid the use of integrals. In addition, I will introduce the very slight limitation that the number of sample points is finite, thus avoiding the use of Hilbert spaces.

Consider again the example of soil samples collected at regular intervals along an imaginary line. The regionalized variable was soil moisture; the outcomes were DDD, DDW, DWD, DWW, WDD, WDW, WWD, and WWW. The probability density function obtained previously was:

$$f(x_1,x_2,x_3) = \begin{cases} 1/8 & x_1,x_2,x_3 = 0,1 \\ 0 & \text{elsewhere} \end{cases}$$

If the probability density function for a regionalized variable is known, random variates can be generated. A basic assumption in the intrinsic theory is that a regionalized variable is a random variate (Matheron, 1969b); that is, the observed values are outcomes following

some probability density function. To simulate heads and tails in a coin-flipping experiment, for instance, simply generate random variates according to the uniform distribution $f(x) = 1/2$, $x = 0,1$. Suppose in the example of the three soil samples we actually collect DDD. Then DDD is a particular realization of the regionalized variable moisture, according to the probability density function $f(x_1,x_2,x_3)$.

Consider now the example of tossing three coins and computing outcomes as ordered triples. If we suppose the coins are balanced, we can verify that the probability density is exactly the same as the one obtained for the example of soil samples. However, if we actually conduct both experiments, theory and reality will match only for coin flipping because the probability density function predicted for the soil samples is incorrect. Our initial analysis in the soil moisture example was an abstraction that was far from reality. Unfortunately, there is no analytical way to take into consideration all the factors, such as soil composition, temperature, weather, surrounding flora, etc., which determine the moisture content of a soil sample. Therefore, a statistical approach is still appropriate, but using the simple model assumed in the example.

This inability to account for all possible variables is not unusual. If the coin is biased and we want to improve the probability density function used in the example of coin tossing, it would be impossible to analyze the moment of inertia of the coin, the weight, and the countless other variables that might be involved. Instead, we run an experiment and flip the coin, testing it repeatedly under constant conditions, until the probability of the outcome stabilizes.

The difficulty of this relative frequency approach with a regionalized variable is that a repeated experiment cannot be run because each outcome is unique. When the moisture content of a soil sample is analyzed, the initial state of the sample is destroyed, so at the end of the assay it is not the same soil sample that was originally collected in the field. There is no other sample in the world with precisely all of the features of this specific sample. Any other would have come from a different location, and would have a different shape, weight, and other characteristics.

As a large number of samples is essential to any statistical inference, we will never be able to determine the probability density function which rules the occurrence of a regionalized variable, although theoretically we can suppose there is one. Fortunately, the impossibility of obtaining the probability density function associated with a regionalized variable is not a serious limitation, as knowledge of the probability function is unnecessary. Most of the properties of interest depend only on the structure of the regionalized variable as specified by its first and second moments. A key assumption is one of some kind of stationarity, which allows statistical inference. Under this assumption, outcomes of different probability density functions can be used in common to estimate moments. The assumption is analogous to that of ergodicity in stochastic processes.

Spatial correlation between samples is a fundamental characteristic of regionalized variables. Imagine we collect soil samples along a line and note that the first one is dry.

The occurrence of a second dry sample some distance away is independent of the occurrence of a dry first sample. This is equivalent to trying to guess if heads or tails are more likely on the twentieth flip of a coin, given that heads were obtained on the first trial. The difference between these two situations appears when we examine the second or third sample along the sequence. In flipping coins, the probability of heads will remain constant regardless of position in the sequence, because all outcomes are mutually independent. However, the probability that a second sample taken one meter away from the first will be dry is very high, and even higher if the second sample is taken only two centimeters away.

We will now try to generalize the example of soil sampling and afterwards demonstrate the aim of stationarity. Suppose we collect three soil samples along a line and specify the location of each sample by the x coordinate of its center of gravity. These will define a line of points if the samples have the same shape, size, and orientation. We have already defined three random variables X_1, X_2, X_3 to specify the moisture content of the respective samples. If we consider the x coordinate of the sample as a fourth variable we may establish the following identities:

$$X_1 = Y(x_1)$$

$$X_2 = Y(x_2)$$

$$X_3 = Y(x_3)$$

where x_1, x_2, x_3 are coordinates of the samples along the line and Y is the random variable moisture. Now, define the random vector $\vec{Y}$ as:

$$\vec{Y} = (X_1, X_2, X_3) = [Y(x_1), Y(x_2), Y(x_3)]$$

The use of vector notation makes it easier to extend the experiment. In general, for k samples, $\vec{Y}$ will be a random vector of the form:

$$\vec{Y} = [Y(x_1), Y(x_2), Y(x_3), \ldots, Y(x_k)]$$

If the geometric field is a farm, we can generalize even further if we suppose that each soil sample has the same size, shape, and orientation but it may come from any point on the farm and from any depth within the soil. The location of the random variable is now (x_i, y_i, z_i) instead of x_i, i = 1, 2, ..., k. Finally, in n-dimensional space, the random vector $\vec{Y}$ will be:

$$\vec{Y} = [Y(\vec{x}_1), Y(\vec{x}_2), \ldots, Y(\vec{x}_k)]$$

where $\vec{x}_i = (x_i^1, x_i^2, \ldots, x_i^n)$ with x_i^j denoting the component in the jth dimension for the ith sample.

The last generalization is to redefine the probability density function for a continuous domain. Then, moisture may be regarded as ranging continuously from zero to 100 percent instead of being simply "dry" or "wet."

To gain insight into the structure of a regionalized variable, we may formulate some hypotheses. A regionalized variable is *stationary* if the statistics of the random variables

$Y(\vec{x}_i)$ and $Y(\vec{x}_j)$ are equal (Mix, 1969). That is, the statistics determined for $Y(\vec{x})$ are equal to those for $Y(\vec{x} + \vec{h})$ for every $\vec{h}$. A regionalized variable with finite moments will be called *stationary to the k order* if, for the joint probability density function $f(\vec{Y})$ of the vector $\vec{Y}$, we have:

$$f[Y_1(\vec{x}_1), Y_2(\vec{x}_2), \ldots, Y_k(\vec{x}_k)] = f[Y_1(\vec{x}_1 + \vec{h}), Y_2(\vec{x}_2 + \vec{h}), \ldots, Y_k(\vec{x}_k + \vec{h})]$$

for all $\vec{h}$ and any collection of points $(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k)$. This condition of stationarity is the same constraint used in the study of random processes. Since there are several types of stationarity, some special terminology has arisen to distinguish these. A regionalized variable is called *strictly stationary* if it is stationary for any order k = 1, 2, ... If k is equal to one, the regionalized variable has first-order stationarity. An interesting case is when k equals two. Under this circumstance, we say the regionalized variable has stationarity of order two or second-order stationarity. But second-order stationarity implies first-order stationarity as well. Therefore, it must be true that the probability density function at location $\vec{x}$ is the same as the probability density function at location $\vec{x} + \vec{h}$, for all values of $\vec{h}$.

$$f[Y(\vec{x})] = f[Y(\vec{x} + \vec{h})] \quad \text{for all } \vec{h}$$

Thus, the first-order density function must be independent of $\vec{x}$. Since the mean value is a function that depends only on the first-order density function, the mean must be independent of the location $\vec{x}$. The mean, or expected value, of X is symbolically represented by E[X]. In general, let X be a random variable having a discrete distribution and let u(X) be a function of X. The sum

$$\sum_x u(x)\, f(x)$$

is called the *expected value* of u(X) and denoted by E[u(X)].

$$E[Y(\vec{x})] = E[Y(\vec{x} + \vec{h})] = \mu$$

From the definition of second-order stationary regionalized variables, it must also be true that:

$$f[Y_1(\vec{x}_1), Y_2(\vec{x}_2)] = f[Y_1(\vec{x}_1 + \vec{h}), Y_2(\vec{x}_2 + \vec{h})] \quad \text{for all } \vec{h},$$

which is a function only of $\vec{x}_2 - \vec{x}_1$. Now, as the mean is a constant, the covariance depends only on the second-order density function. That is, the covariance is a function dependent only on $\vec{x}_2 - \vec{x}_1$, which we will call $\vec{h}$.

$$\text{cov}(\vec{x}_1, \vec{x}_2) = \text{cov}(\vec{x}_2 - \vec{x}_1) = \text{cov}(\vec{h}) = E[Y(\vec{x}_1)\, Y(\vec{x}_2)] - \mu^2$$

In general, for a regionalized variable of stationarity of order k, all the moments of order k or less are invariant under translation. Second-order stationarity is all that is usually

required in the intrinsic theory. Stationarity is a mathematical way to introduce the restriction that the regionalized variable must be homogeneous and permits us to make statistical inferences. By assuming stationarity we can essentially repeat an experiment even though samples must be collected at different points, as all samples are assumed to be drawn from populations having the same moments.

Lemma 2.1.1: For a stationary regionalized variable, the covariance has the following properties:

1) $\text{cov}(0) \geq |\text{cov}(\vec{x}_2 - \vec{x}_1)|$

2) $\lim_{\vec{h} \to \infty} \text{cov}(\vec{h}) = 0$

3) $\text{cov}(0) = \text{var}[Y(\vec{x})]$

4) $\text{cov}(\vec{x}_2 - \vec{x}_1) = \text{cov}(\vec{x}_1 - \vec{x}_2)$

All these properties follow from the definition of $\text{cov}(\vec{h})$ or from elementary statistics.

For second-order stationarity, $\text{var}[Y(\vec{x})]$ must be finite by the definition of stationarity. Then $\text{cov}(0)$ must be finite. However, many phenomena in nature are subject to unlimited dispersion and cannot correctly be described when they are assigned a finite variance. Classic examples include the variance of gold values in South African gold mines (Matheron, 1967a) and some multiplicative Markov chains (Bartlett, 1966, p. 74).

To avoid this restriction, usually immaterial, the intrinsic theory assumes what is called the *intrinsic hypothesis* instead. Let $Z(\vec{x},\vec{h})$ be a random variable defined by:

$$Z(\vec{x},\vec{h}) = Y(\vec{x} + \vec{h}) - Y(\vec{x})$$

We will say the regionalized variable satisfies the intrinsic hypothesis if for all $\vec{h}$ the first and second moments of the difference $Y(\vec{x} + \vec{h}) - Y(\vec{x})$ depend only on the distance between the two points $\vec{x} + \vec{h}$ and $\vec{x}$ and not on their individual locations (Matheron, 1971).

$$E[Z(\vec{x},\vec{h})] = m(\vec{h})$$

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

The function $\gamma(\vec{h})$ is the semivariogram or intrinsic function. This function plays a very important role in the intrinsic theory and it is not used in any other branch of statistics. Therefore, the semivariogram is a symbol of the intrinsic theory. Since the intrinsic hypothesis is slightly more general than second-order stationarity and involves no additional analytical complications, we will usually apply the intrinsic hypothesis although second-order stationarity is all that is usually required in the application to automatic contouring. However, the principal reason for retaining the intrinsic function is not a reason at all, but merely sentiment. Tradition dictates that an intrinsic theory without the intrinsic function simply seems incomplete.

2.2 The Semivariogram

The semivariogram occupies a central position in the intrinsic theory. First we will formally define this function and then we will show some properties that will be useful later.

Definition 2.2.1: Let $Y(\vec{x})$ be a function whose differences $Y(\vec{x} + \vec{h}) - Y(\vec{x})$ have first and second moments depending only on the distance $\vec{h}$ between locations $\vec{x}$ and $\vec{x} + \vec{h}$. The *semivariogram* or *intrinsic function* is denoted by $\gamma(\vec{h})$ and defined by:

$$\gamma(\vec{h}) = 1/2\ E[\{Y(\vec{x} + \vec{h}) - Y(\vec{x}) - E[Y(\vec{x} + \vec{h}) - Y(\vec{x})]\}^2]$$

The following Lemma illustrates important properties of the intrinsic function.

Lemma 2.2.2: Let $\gamma(\vec{h})$ be the intrinsic function for a first-order stationary regionalized variable $Y(\vec{x})$ which satisfies the intrinsic hypothesis. Then

1) $\gamma(\vec{h}) = \gamma(-\vec{h})$
2) $\gamma(0) = 0$

Proof: Because $Y(\vec{x})$ is first-order stationary, $E[Y(\vec{x} + \vec{h}) - Y(\vec{x})] = 0$. Hence,

1) $\gamma(\vec{h}) = \gamma[(\vec{x} + \vec{h}) - \vec{x}]$
$= E[\{Y(\vec{x} + \vec{h}) - Y(\vec{x})\}^2]$
$= E[\{Y(\vec{x}) - Y(\vec{x} + \vec{h})\}^2]$
$= \gamma[\vec{x} - (\vec{x} + \vec{h})]$
$= \gamma(-\vec{h})$
2) $\gamma(0) = \gamma(\vec{x} - \vec{x}) = E[\{Y(\vec{x}) - Y(\vec{x})\}^2] = 0$

Q.E.D.

The only significant difference between second-order stationarity and the intrinsic hypothesis is that cov(0) may be infinite. In fact, if cov(0) is finite, any regionalized variable which is second-order stationary satisfies the intrinsic hypothesis and we can prove the following theorem.

Theorem 2.2.3: Let $Y(\vec{x})$ be a second-order stationary regionalized variable. Then

$$\gamma(\vec{h}) = \text{cov}(0) - \text{cov}(\vec{h})$$

Proof: As $Y(\vec{x})$ is first-order stationary, $E[Y(\vec{x} + \vec{h}) - Y(\vec{x})] = 0$. Hence,

$$\gamma(\vec{h}) = 1/2\ E[\{Y(\vec{x} + \vec{h}) - Y(\vec{x})\}^2]$$
$$= 1/2\ E[\{Y(\vec{x} + \vec{h})\}^2] + 1/2\ E[\{Y(\vec{x})\}^2] - E[Y(\vec{x} + \vec{h})\ Y(\vec{x})]$$

Because the regionalized variable is second-order stationary,

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

and the equality follows by the definition of $cov(\vec{h})$ and from Lemma 2.1.1 (3).

Q.E.D.

If $cov(0)$ is finite, the equality above is a simple linear relationship which allows us to translate any formula from terms of the intrinsic function to covariance and vice versa.

The following Lemma expresses another useful relationship between the intrinsic function and covariance for a second-order stationary regionalized variable.

Lemma 2.2.4: Let $\gamma(\vec{h})$ be the intrinsic function and $cov(\vec{h})$ the covariance for a second-order stationary regionalized variable. Then

$$\lim_{\vec{h} \to \infty} \gamma(\vec{h}) = cov(0)$$

Proof: By Theorem 2.2.3,

$$\lim_{\vec{h} \to \infty} \gamma(\vec{h}) = cov(0) - \lim_{\vec{h} \to \infty} cov(\vec{h})$$

but by Lemma 2.1.1 (2), the last term vanishes and

$$\lim_{\vec{h} \to \infty} \gamma(\vec{h}) = cov(0)$$

Q.E.D.

We will prove one more theorem concerning relationships between the intrinsic function and the covariance which will be useful in the following section.

Theorem 2.2.5: Let $Y'(\vec{u})$ be a random function defined as the difference:

$$Y'(\vec{u}) = Y(\vec{u}) - Y(\vec{u}_0)$$

where $\vec{u}_0$ is constant and $Y(\vec{u})$ is a first-order stationary random function which satisfies the intrinsic hypothesis. Then, the covariance of $Y'(\vec{u})$ is:

$$cov(\vec{u},\vec{v}) = \gamma(\vec{u} - \vec{u}_0) + \gamma(\vec{v} - \vec{u}_0) - \gamma(\vec{u} - \vec{v})$$

Proof: By Definition 2.2.1 and because the mean of $Y'(\vec{u})$ is zero as $Y(\vec{u})$ is first-order stationary,

$$\gamma(\vec{u} - \vec{v}) = 1/2\ E[\{Y'(\vec{u}) - Y'(\vec{v})\}^2]$$

But according to the definition of $Y'(\vec{u})$,

$$\gamma(\vec{u} - \vec{v}) = 1/2\ E[\{(Y(\vec{u}) - Y(\vec{u}_0) - Y(\vec{v}) + Y(\vec{u}_0))\}^2]$$

Expanding the square of the expression within the parentheses and taking the expected value of each term gives

$$\gamma(\vec{u} - \vec{v}) = 1/2\ E[\{Y(\vec{u}) - Y(\vec{u}_0)\}^2] + 1/2\ E[\{Y(\vec{v}) - Y(\vec{u}_0)\}^2] - E[\{Y(\vec{u}) - Y(\vec{u}_0)\}\ \{Y(\vec{v}) - Y(\vec{u}_0)\}]$$

By Definition 2.2.1,

$$\gamma(\vec{u} - \vec{v}) = \gamma(\vec{u} - \vec{u}_0) + \gamma(\vec{u} - \vec{u}_0) - E[\{Y(\vec{u}) - Y(\vec{u}_0)\}\ \{Y(\vec{v}) - Y(\vec{u}_0)\}]$$

The last term is $\text{cov}(\vec{u},\vec{v}) + E[Y'(\vec{u})]\ E[Y'(\vec{v})]$. However, the product vanishes because $E[Y(\vec{u})]$ is constant. This implies that $E[Y'(\vec{u})]$ is null. Hence, replacing and rearranging the expression yields:

$$\text{cov}(\vec{u} - \vec{v}) = \gamma(\vec{u} - \vec{u}_0) + \gamma(\vec{v} - \vec{u}_0) - \gamma(\vec{u} - \vec{v})$$

Q.E.D.

Theorem 2.2.6: Let $Y(\vec{x}_1)$, $Y(\vec{x}_2)$, ..., $Y(\vec{x}_t)$ represent t random variables taken at regular intervals $\vec{a}$ and let $\vec{h}$ be equal to $p\vec{a}$. If all these random variables are first-order stationary, satisfy the intrinsic hypothesis, and have the same semivariogram, the following is an unbiased estimator of $\gamma(\vec{h})$:

$$\overline{\gamma}(\vec{h}) = \frac{1}{2(k-p)} \sum_{j=k'}^{k'+k-p-1} [Y(\vec{x}_j + p\vec{a}) - Y(\vec{x}_j)]^2$$

for $k + k' - 1 \leqslant t$ and $p = 0, 1, \ldots, k-1$.

Proof: If $\overline{\gamma}(\vec{h})$ is unbiased, $E[\overline{\gamma}(\vec{h})]$ must be equal to $\gamma(\vec{h})$.

$$E[\overline{\gamma}(\vec{h})] = E\left[\frac{1}{2(k-p)} \sum_{j=k'}^{k'+k-p-1} [Y(\vec{x}_j + p\vec{a}) - Y(\vec{x}_j)]^2\right]$$

But $1/2(k-p)$ is a constant which can be taken out of the expectation. Interchanging summation and expectation,

$$E[\overline{\gamma}(\vec{h})] = \frac{1}{2(k-p)} \sum_{j=k'}^{k'+k-p-1} E[\{Y(\vec{x}_j + p\vec{a}) - Y(\vec{x}_j)\}^2]$$

However, as $E[Y(\vec{x}_j)]$ is constant,

$$E[\{Y(\vec{x}_j + p\vec{a}) - Y(\vec{x}_j)\}^2] = 2\gamma(p\vec{a}) = 2\gamma(\vec{h})$$

for $j = 1, 2, \ldots, k-1$. Therefore,

$$E[\bar{\gamma}(\vec{h})] = \gamma(\vec{h})$$

Q.E.D.

All terms in the summation have some physical meaning. If all the samples are collected along a line, which is the only case of practical interest, there is a linear relationship between the interval length $\vec{S}$, the number of samples k, and the interval between consecutive samples:

$$\vec{S} = k\vec{a} - \vec{a}$$

The parameter $\vec{S}$ is very important because the estimated semivariogram is not only a function of the relationship among samples but also of $\vec{S}$. Although the intrinsic function is continuous, in a sequence of discontinuous samples taken at regular intervals $\vec{a}$, we can calculate the semivariogram only at the integer multiples of $\vec{a}$ we have called p. Finally, the parameter k' represents the left-most sample in the neighborhood. When the sample sequence is larger than $\vec{S}$, several samples can be taken as the left-most sample in the interval, provided the sequence to the right of the kth sample is not shorter than $\vec{S}$. The semivariogram is usually assumed to be given by the average of all those partial semivariograms of moving origin k'.

The semivariogram has all the structural information needed about a regionalized variable, including the size of the zone of influence around a sample, the isotropic nature of the variable, and the continuity of the variable through space. We will examine each of these properties in turn.

1) Zone of influence: The semivariogram provides a precise meaning to the notion of dependence between samples. Mathematically, the *influence zone* is the n-dimensional sphere whose radius is the smallest distance $\vec{L}$ such that:

$$\text{cov}(0) - \gamma(\vec{L}) \leqslant \varepsilon$$

where ε is any small number (Fig. 2.2.1). $\vec{L}$ is called the *range* of the semivariogram (Matheron, 1963). That part of the semivariogram to the right of the range is referred to as the *sill*.

In the estimation of regionalized variables, $\vec{L}$ divides the samples into two categories. All samples whose distances to the point to be estimated are less than or equal to $\vec{L}$ provide information about the point. All samples outside the neighborhood defined by $\vec{L}$ are independent observations with respect to the point to be estimated and may be disregarded because they do not provide any information about the point.

2) Anisotropy: Anisotropy is revealed by different behavior of the variograms for different directions (Fig. 2.2.2). These differences in the semivariogram are seen mainly in the slope at the origin, in the range, and in the sill, if any. If the regionalized variable is isotropic, the intrinsic function depends only on the magnitude r of vector $\vec{h}$ and not on the direction $\hat{h}$ of the vector $\vec{h}$. If a simple linear transformation of coordinates is sufficient to achieve isotropy, then we say there is a *geometric anisotropy* (Fig. 2.2.3).

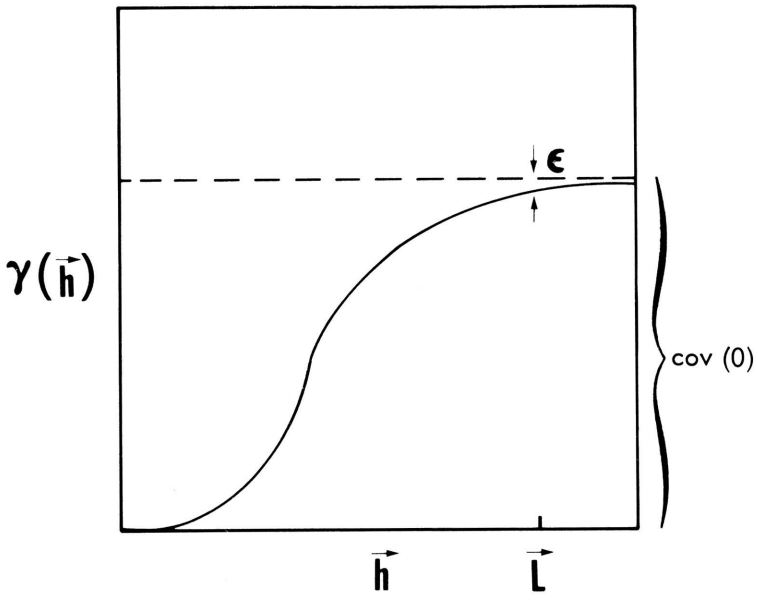


Figure 2.2.1--Semivariogram, or graph of semivariance γ versus distance along vector $\vec{h}$. $\vec{L}$ is the range beyond which the difference, ε, between semivariance and variance (shown by dashed line) is considered negligible.

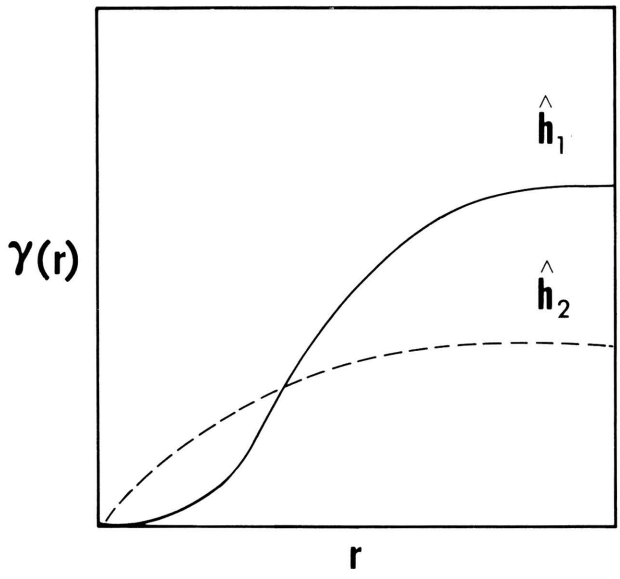


Figure 2.2.2--Anisotropic intrinsic function, in which semivariograms for two different vector directions $\hat{h}_1$ and $\hat{h}_2$ differ in slope, range, and sill.

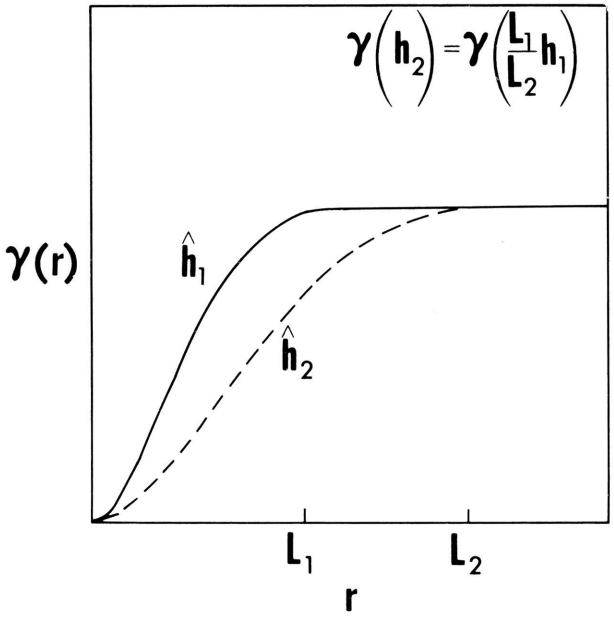


Figure 2.2.3--Geometric anisotropy of the intrinsic function. Semivariograms in vector directions $\hat{h}_1$ and $\hat{h}_2$ are different, but can be made equivalent by a change in scale.

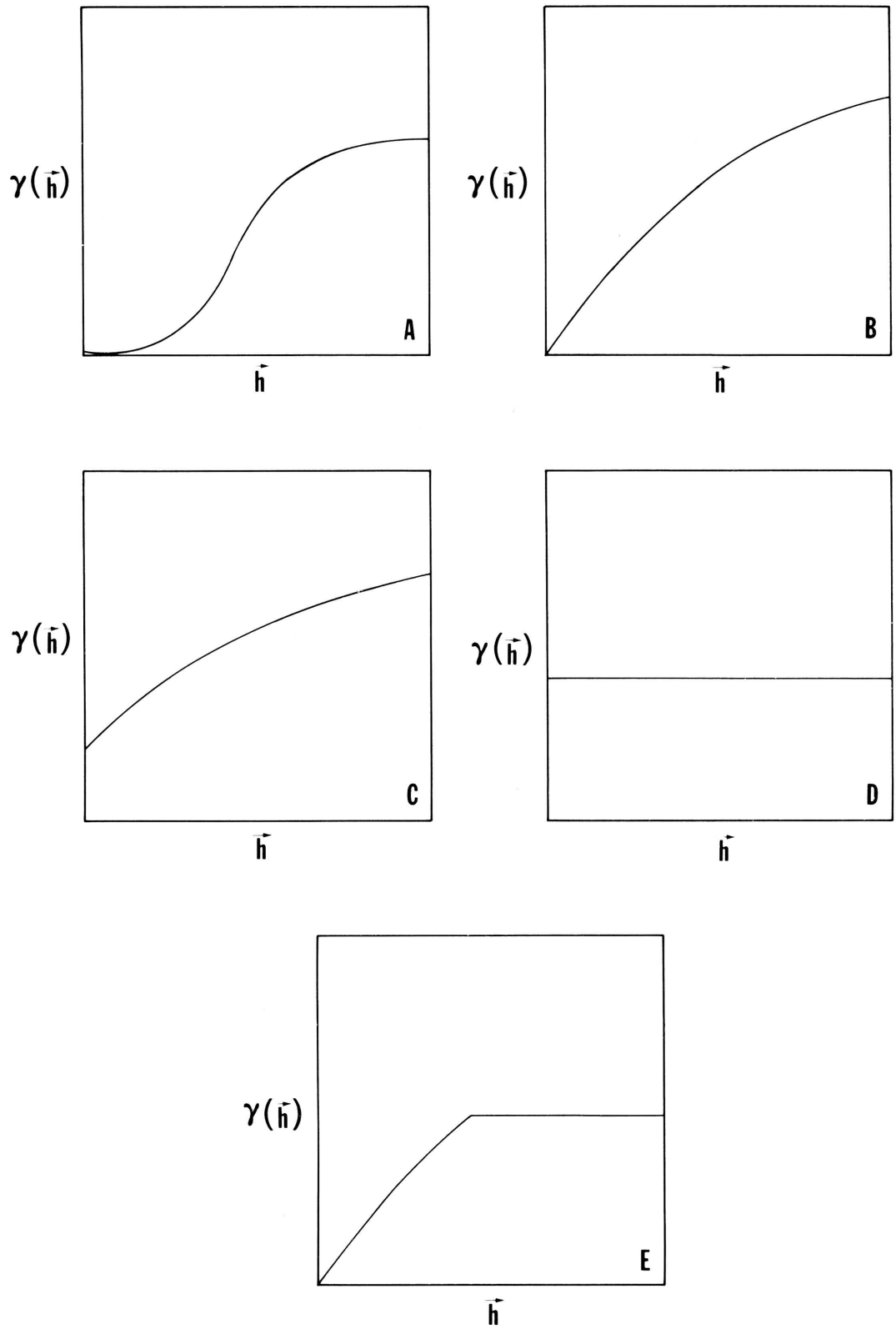


Figure 2.2.4--Idealized semivariograms. *a*. Parabolic form near origin. *b*. Linear form near origin. *c*. "Nugget effect." *d*. Semivariogram of independent random variable. *e*. Transitive type semivariogram.

If $\vec{h}$ is an n-dimensional vector, the semivariogram may not be a function of all n coordinates. For instance, in three-dimensional layered media the regionalized variable may be a function of only the third coordinate, depth. That is, the variable is constant on planes parallel to the other axes. If this is the case, the semivariogram will be a function only of the third component of $\vec{h}$. This is called *zonal anisotropy*, resulting from zonation of the regionalized variable.

3) Continuity: The shape of the semivariogram, and in particular its behavior near the origin, provides information about the continuity and regularity of the regionalized variable. Figure 2.2.4*a* has a parabolic form near the origin. This implies excellent continuity. $\gamma(\vec{h})$ is twice differentiable in $\vec{h} = 0$. Figure 2.2.4*b* shows linear behavior near the origin. This implies a moderate continuity. $\gamma(\vec{h})$ is continuous but not twice differentiable.

The fact that $\gamma(0) = 0$ does not forbid $\lim \gamma(\vec{h}) = C$, $h \to 0$. In this situation, the semivariogram is discontinuous at the origin because of poor continuity of the regionalized variable (Fig. 2.2.4*c*). Discontinuities and sporadic occurrences are characteristic of gold deposits, leading to the expression "nugget constant" for the constant C.

In Figure 2.2.4*d*, the regionalized variable is highly erratic and without discernible pattern. No sample is related to even its closest neighbor. The regionalized variable is a pure random variable and the same results will be obtained from the intrinsic theory as from the classical statistics of independent random variables.

Finally, there is a type called a *transitive semivariogram* (Fig. 2.2.4*e*). This kind of semivariogram is characterized by a finite range; in other words, there is a sill. This semivariogram indicates moderate continuity within reduced neighborhoods and pure random behavior over longer intervals.

2.3 The Drift

In regionalized variable theory or any other approach to spatially distributed variables, the concept of drift is one of the most important properties of the function. Physically the drift represents the trend of the function over the geometric field. In this sense, the drift must represent only the major features, and should present a systematic appearance rather than representing sporadic details. "Major" and "sporadic" are terms relative to the working scale. The same feature which can be regarded as part of the drift at the scale of a county might be a minor, sporadic detail at the scale of a state and perhaps will even disappear when the scale of study is a country. Therefore, the drift is not unique. The idea of drift brings to mind the statistical concept of expected value. As observed values of a regionalized variable are no more than the realizations of a random variable, it seems reasonable to give the following definition for drift.

Definition 2.3.1: Let $Z(\vec{x})$ be a regionalized variable. Then, the *drift* $m(\vec{x})$ is by definition:

$$m(\vec{x}) = E[Z(\vec{x})]$$

That is, the drift at a point $\vec{x}$ is the expected value of the regionalized variable Z at point $\vec{x}$ (Huijbregts and Matheron, 1971).

The concept of drift provides a means for splitting the regionalized variable into two components, the drift as defined above, and the residual $Y(\vec{x})$.

Definition 2.3.2: Let $Z(\vec{x})$ be a regionalized variable with drift $m(\vec{x})$. Then, the *residual* $Y(\vec{x})$ is:

$$Y(\vec{x}) = Z(\vec{x}) - m(\vec{x})$$

An important property of the residuals is that they have zero mean, which in turn suggests that they can be used to calculate a semivariogram according to Theorem 2.2.6.

Lemma 2.3.3: Let $Y(\vec{x})$ be the residual from a regionalized variable. Then,

$$E(Y(\vec{x})] = 0$$

Proof: By the definition of residual,

$$E[Y(\vec{x})] = E[Z(\vec{x}) - m(\vec{x})] = E[Z(\vec{x})] - E[m(\vec{x})]$$

but for a given $\vec{x}$, $m(\vec{x})$ is not a random variable but a constant. Therefore, the expected value of the drift is the drift itself and as the first term is again the drift, the difference is identically zero.

Q.E.D

We will use the drift in two different ways. First, in order to be able to make any estimate of a regionalized variable, it is necessary to know about its continuity. This information will be obtained from the semivariogram of the residuals. Therefore, one important use of the drift will be to provide residuals for semivariogram analysis. In this instance we must obtain an analytical expression for the drift which is valid for an entire neighborhood. The other important use of drift is for map analysis; examination of the smooth changes or general tendencies of the regionalized variable. This is analogous to the objectives of trend surface analysis, the drawbacks of which have already been mentioned in Section 1.4. Drift estimation, in addition to providing a replacement for the ambiguous trend, is an optimal approach.

Point estimates are required to contour the drift. This leads to the major objection to drift calculation, which is the greater complexity involved as compared to trend surface calculations. An optimal estimate of the drift must be obtained for each node of the contouring

grid. Unless the observations are distributed according to a regular pattern, the estimation algorithm must be repeated for every point. This requires much more computation than in trend surface analysis, where the least-squares surface is obtained at once for the entire map. Different algorithms will be devised for estimation of a point and for estimating the coefficients of an analytical expression which is valid inside a restricted neighborhood. In both instances, two cases will be present depending on whether the covariance or the semivariogram for the residuals is the known measure of continuity for the regionalized variable.

From the fact that the drift varies continuously and smoothly, $m(\vec{x})$ can at least locally be approximated by an analytical expression.

Definition 2.3.4: Given a neighborhood of radius r around $\vec{x}_0$, the drift at any $\vec{x}$, $|\vec{x}-\vec{x}_0| \leq r$, can be defined as

$$m(\vec{x}) = \sum_{i=0}^{n} a_i f^i(\vec{x})$$

where a_i are unknown coefficients and $f^i(\vec{x})$ are arbitrary functions of $\vec{x}$.

In practice, $f^i(\vec{x})$ have been restricted to sinusoidal functions or integer powers of $\vec{x}$. The decision about the value of n and the form of the function is arbitrarily made in a first stage. We may choose in conjunction with other considerations the best combination of functions to represent the drift. However, the problem of estimating the a_i's remains. Their estimation is possible under the following assumptions:

1) The expected value of the regionalized variable exists.

2) Either the covariance or the intrinsic function for the residuals is known. Generically, we will call them *correlograms*.

3) There are sufficient samples measured without error and statistically dependent on the point $\vec{x}_0$ where the drift is to be estimated. That is, there are enough samples whose distance to $\vec{x}_0$ is shorter than the radius r.

The second assumption is critical because, as we will see later, it is necessary to know the drift in order to calculate the correlogram. This impasse may be resolved by iteration. We assume a theoretical correlogram and an expression for the drift. We then evaluate the coefficients in the expression of the drift. Using this drift and the samples, we then compute an experimental correlogram for the residuals which must fit the theoretical correlogram for the residuals. If a fit is not achieved, the process is repeated, redefining either a new expression for the drift or a new theoretical correlogram, or both. To redefine the theoretical correlogram usually implies redefining the neighborhood radius r. That is, the number of samples available to determine the coefficients of the drift are either increased or decreased. Hence, the type of expression used for the drift is not arbitrary. It is determined by data, which we will always consider to be point observations.

We will next consider the optimal estimation of an analytical expression for the drift when the covariance exists. We will suppose assumptions 1) and 3) hold and that there is a covariance

for the residuals of the regionalized variable. In this circumstance, we can obtain an analytical expression $m(\vec{x})$ for the drift at a given point $\vec{x}_0$. That is, we can calculate the unknown coefficients a_i in the expression

$$m(\vec{x}) = \sum_{i=0}^{n} a_i f^i(\vec{x})$$

and $m(\vec{x})$ evaluated at $\vec{x} = \vec{x}_0$ will be the drift of the regionalized variable $Z(\vec{x})$ at location $\vec{x}_0$, $|\vec{x}-\vec{x}_0| \leq r$. Both the number of coefficients, n, and the functions $f^i(\vec{x})$ are chosen for convenience. We will assume that it is possible to draw k punctual samples of the regionalized variable $Z(\vec{x})$ at locations $\vec{x}_j$, $j = 1, 2, \ldots, k$.

We will define k statistics, one for each a_i, which are designated as A_i. The statistic for each a_i will be a linear combination of the k samples inside a circle of radius r around $\vec{x}_0$.

Definition 2.3.5: The statistic A_i for a_i is by definition:

$$A_i = \sum_{j=1}^{k} \lambda_i^j Z(\vec{x}_j) \quad \text{for } |\vec{x}_j-\vec{x}_0| \leq r \qquad i = 0, 1, \ldots, n$$

We will say the statistics are optimal if:

$$E[A_i] = a_i \qquad \text{for } i = 0, 1, \ldots, n$$

$$E[(A_i-a_i)^2] \text{ is minimum of } \lambda_i^j \qquad \text{for } i = 0, 1, \ldots, n$$

That is, A_i is an optimal statistic if it is unbiased and minimizes the estimation variance of A_i with respect to all possible weights λ_i^j.

Definition 2.3.6: Let A_i be a statistic of a_i as defined in Definition 2.3.5. The optimal statistic of $m(\vec{x})$ is denoted by $M(\vec{x})$ and is defined by:

$$M(\vec{x}) = \sum_{i=0}^{n} A_i f^i(\vec{x})$$

Lemma 2.3.7: $M(\vec{x})$ is an unbiased statistic for the drift if the statistics A_i in Definition 2.3.6 are unbiased, namely

$$E[M(\vec{x})] = m(\vec{x})$$

Proof: By Definition 2.3.6,

$$E[M(\vec{x})] = E[A_i f^i(\vec{x})]$$

for any $\vec{x}$, such that $|\vec{x}-\vec{x}_0| \leq r$. Interchanging expectation and summation, as $f^i(\vec{x})$ is not a random variable,

$$E[M(\vec{x})] = \sum_{i=0}^{n} f^i(\vec{x})\, E[A_i]$$

But A_i is unbiased. Therefore,

$$E[M(\vec{x})] = \sum_{i=0}^{n} f^i(\vec{x})\, a_i$$

and by Definition 2.3.4,

$$E[M(\vec{x})] = m(\vec{x})$$

Q.E.D.

The problem in drift estimation is the evaluation of the (n+1)k weights λ_i^j. A solution will be valid only inside the neighborhood r around $\vec{x}_0$. Since the choice of r is arbitrary, we must test our selection. This is a problem which will be discussed when we consider the practical application of the theory in Chapter 3.

In order to find the weights λ_i^j, we must consider some further theorems and lemmas.

Definition 2.3.8: We will denote by $\delta(i,s)$ a function with the following properties:

$$\delta(i,s) = \begin{cases} 0 & i \neq s \\ 1 & i = s \end{cases}$$

for i, s = 0, 1, ..., n. This function, among other uses, will be helpful in proving the following theorem.

Theorem 2.3.9: Let λ_s^j be a weight in Definition 2.3.5 for A_s and let $f^i(\vec{x})$ be one of the arbitrary functions in Definition 2.3.4 for $m(\vec{x})$. Then, A_s is unbiased if and only if

$$\sum_{j=1}^{k} \lambda_s^j\, f^i(\vec{x}) = \delta(i,s)$$

for all s and i, s and i being the first n+1 non-negative integers.

Proof: Let us suppose A_s is unbiased. According to the Definition 2.3.5 for A_s, we have

$$A_s = \sum_{j=1}^{k} \lambda_s^j\, Z(\vec{x}_j)$$

Since summation and expectation are interchangeable, and λ_s^j is a constant, we have by Definitions 2.3.1 and 2.3.4

$$E[A_s] = \sum_{j=1}^{k} \lambda_s^j \sum_{i=0}^{n} a_i f^i(\vec{x}_j)$$

As λ_s^j is independent of i, the weights can be introduced in the second summation and we can reverse summations over j and i. Therefore,

$$E[A_s] = \sum_{i=0}^{n} \sum_{j=1}^{k} \lambda_s^j \, a_i f^i(\vec{x}_j)$$

We can take a_i out of the second summation over j because a_i is independent of j. Thus,

$$E[A_s] = \sum_{i=0}^{n} a_i \sum_{j=1}^{k} \lambda_s^j \, f^i(\vec{x}_j)$$

But according to our hypothesis about A_s, it is unbiased. Therefore,

$$E[A_s] = a_s$$

Thus,

$$a_s = \sum_{i=0}^{n} a_i \sum_{j=1}^{k} \lambda_s^j \, f^i(\vec{x}_j)$$

But this is an equality which is also true individually for each a_i. Hence,

$$\sum_{j=1}^{k} \lambda_s^j \, f^i(\vec{x}_j) = \begin{cases} 0 & i \neq s \\ 1 & i = s \end{cases}$$

Hence, if A_s is unbiased,

$$\sum_{j=1}^{k} \lambda_s^j \, f^i(\vec{x}_j) = \delta(i,s)$$

for i = 0, 1, ..., n. Because we did not select any particular A_s, this argument is also true for s = 0, 1, ..., n. Let

$$\sum_{j=1}^{k} \lambda_s^j \, f^i(\vec{x}_j) = \delta(i,s)$$

If we multiply the equality by a_i,

$$a_i \sum_{j=1}^{k} \lambda_s^j \, f^i(\vec{x}_j) = a_i \, \delta(i,s)$$

for all i, i = 0, 1, ..., n. Therefore, the sum is also an equality:

$$\sum_{i=0}^{n} a_i \sum_{j=1}^{k} \lambda_s^j \, f^i(\vec{x}_j) = a_s$$

Introducing a_i under the second summation sign and interchanging summation order gives:

$$\sum_{j=1}^{k} \sum_{i=0}^{n} a_i \lambda_s^j \, f^i(\vec{x}_j) = a_s$$

Taking λ_s^j out of the second summation,

$$\sum_{j=1}^{k} \lambda_s^j \sum_{i=0}^{n} a_i f^i(\vec{x}_j) = a_s$$

But by Definition 2.3.4, the second summation is the drift. Hence,

$$\sum_{j=1}^{k} \lambda_s^j \, E[Z(\vec{x}_j)] = a_s$$

Introducing λ_s^j inside the expectation and interchanging summation and expectation we have:

$$E[\sum_{j=1}^{k} \lambda_s^j \, Z(\vec{x}_j)] = a_s$$

By Definition 2.3.5, this sum is A_s. Therefore,

$$E[A_s] = a_s$$

Hence

$$\sum_{j=1}^{k} \lambda_s^j \, f^i(\vec{x}_j) = \delta(i,s)$$

for i = 0, 1, ..., n, implies that A_s is an unbiased estimate of a_s. The two implications prove Theorem 2.3.9.

Q.E.D.

Therefore, we can use either the relation

$$E[A_s] = a_s$$

or

$$\sum_{j=1}^{k} \lambda_s^j \, f^i(\vec{x}_j) = \delta(i,s)$$

for i = 0, 1, ..., n, to imply that the estimate A_s is unbiased. We will next calculate the variance of A_s which we must minimize. The solution is not straightforward, and some theorems must first be proven to achieve this goal.

Lemma 2.3.10: Let λ_s^j, j = 1, 2, ..., k, be weights in the statistic A_s of a_s and let $\text{cov}(\vec{x}_j, \vec{x}_{jj})$ be the covariance for the residuals. Then, for any two points $\vec{x}_j$ and $\vec{x}_{jj}$,

$$a_i a_s = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_i^j \, \lambda_s^{jj} \, m(\vec{x}_j) m(\vec{x}_{jj})$$

Proof: We will call the second member T until we prove it is actually equal to $a_i a_s$. Substituting Definition 2.3.4 for $m(\vec{x})$,

$$T = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_i^j \, \lambda_s^{jj} \sum_{ii=0}^{n} a_{ii} f^{ii}(\vec{x}_j) \sum_{ss=0}^{n} a_{ss} f^{ss}(\vec{x}_{jj})$$

As the weights are independent of ii and a_{ss}, we may include them within the last two summations.

$$T = \sum_{j=1}^{k} \sum_{jj=1}^{k} \sum_{ii=0}^{n} a_{ii} \lambda_i^j f^{ii}(\vec{x}_j) \sum_{ss=0}^{n} a_{ss} \lambda_s^{jj} f^{ss}(\vec{x}_{jj})$$

Interchanging summations,

$$T = \sum_{ii=0}^{n} a_{ii} \sum_{j=1}^{k} \lambda_i^j f^{ii}(\vec{x}_j) \sum_{ss=0}^{n} a_{ss} \sum_{jj=1}^{k} \lambda_s^{jj} f^{ss}(\vec{x}_{jj})$$

But, according to Theorem 2.3.9,

$$\sum_{j=1}^{k} \lambda_i f^{ii}(\vec{x}_j) = \delta(i,ii)$$

Hence,

$$T = \sum_{ii=0}^{n} a_{ii} \delta(i,ii) \sum_{ii=0}^{k} a_{ss} \delta(s,ss) = a_i a_s$$

Q.E.D.

Lemma 2.3.11: Let $Z(\vec{x})$ be a regionalized variable having drift $m(\vec{x})$ and let $\text{cov}(\vec{x}_j,\vec{x}_{jj})$ be the covariance of the residuals of $Z(\vec{x})$. Then,

$$E[Z(\vec{x}_j)Z(\vec{x}_{jj})] = \text{cov}(\vec{x}_j,\vec{x}_{jj}) + m(\vec{x}_j)m(\vec{x}_{jj})$$

Proof: By Definition 2.3.2,

$$E[Z(\vec{x}_j)Z(\vec{x}_{jj})] = E[\{Y(\vec{x}_j) + m(\vec{x}_j)\}\{Y(\vec{x}_{jj}) + m(\vec{x}_{jj})\}]$$

if we expand the product and take the expectation of each term, we obtain

$$E[Z(\vec{x}_j)Z(\vec{x}_{jj})] = E[Y(\vec{x}_j)Y(\vec{x}_{jj})] + E[m(\vec{x}_j)m(\vec{x}_{jj})] + E[m(\vec{x}_j)Y(\vec{x}_{jj})] + E[m(\vec{x}_{jj})Y(\vec{x}_j)]$$

but the drifts are independent of the probability density function of the regionalized variable and can be taken out of the expectations.

$$E[Z(\vec{x}_j)Z(\vec{x}_{jj})] = E[Y(\vec{x}_j)Y(\vec{x}_{jj})] + m(\vec{x}_{jj})m(\vec{x}_j) + m(\vec{x}_j)E[Y(\vec{x}_{jj})] + m(\vec{x}_{jj})E[Y(\vec{x}_j)]$$

By Lemma 2.3.3 the last two terms vanish. The remaining expectation is equal to the covariance of the residuals as the mean of $Y(\vec{x})$ is zero. Hence,

$$E[Z(\vec{x}_j)Z(\vec{x}_{jj})] = \text{cov}(\vec{x}_j,\vec{x}_{jj}) + m(\vec{x}_j)m(\vec{x}_{jj})$$

Q.E.D.

Theorem 2.3.12: Let λ_s^j be the weights in Definition 2.3.5 for A_s. Then, if A_s is unbiased and there is a covariance for the residuals,

$$\operatorname{cov}(A_s,A_i) = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_s^j \lambda_i^{jj} \operatorname{cov}(\vec{x}_j,\vec{x}_{jj})$$

Proof:

$$\operatorname{cov}(A_s,A_i) = E[(A_s-a_s)(A_i-a_i)] = E[A_sA_i-A_sa_i-A_ia_s+a_sa_i]$$

Instead, we can take the sum of the expectations. As a_s and a_i are constants,

$$\operatorname{cov}(A_s,A_i) = E[A_sA_i] - a_iE[A_s] - a_sE[A_i] + a_ia_s$$

Let us study the terms in the summation separately. By Definition 2.3.5, for A_s,

$$E[A_sA_i] = E[\sum_{j=1}^{k} \lambda_s^j Z(\vec{x}_j) \sum_{jj=1}^{k} \lambda_i^{jj} Z(\vec{x}_{jj})]$$

As we can introduce terms containing j inside the summation over jj, we have

$$E[A_sA_i] = E[\sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_s^j \lambda_i^{jj} Z(\vec{x}_j)Z(\vec{x}_{jj})]$$

Because the expectation of a sum is equal to the sum of the expectations, we can interchange summation and expectation. Hence,

$$E[A_sA_i] = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_s^j \lambda_i^{jj} E[Z(\vec{x}_j)Z(\vec{x}_{jj})]$$

But from Lemma 2.3.11,

$$E[A_sA_i] = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_s^j \lambda_i^{jj} \operatorname{cov}(\vec{x}_j,\vec{x}_{jj}) + \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_s^j \lambda_i^{jj} m(\vec{x}_j)m(\vec{x}_{jj})$$

By Lemma 2.3.10, the second term in the sum is equal to a_sa_i. Therefore,

$$E[A_sA_i] = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_s^j \lambda_i^{jj} \operatorname{cov}(\vec{x}_j,\vec{x}_{jj}) + a_sa_i$$

Let us go back and examine the second and third terms in the expression for $\operatorname{cov}(A_s,A_i)$. As A_s and A_i are unbiased, the second term is equal to

$$a_iE[A_s] = a_ia_s$$

and similarly,

$$a_sE[A_i] = a_ia_s$$

Substituting these last three relationships back into the expression for $\text{cov}(A_s,A_i)$,

$$\text{cov}(A_s,A_i) = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_s^j \lambda_i^{jj} \text{cov}(\vec{x}_j,\vec{x}_{jj})$$

Q.E.D.

Corollary 2.3.13: Let λ_s^j be the weights in Definition 2.3.5 for A_s. Then, if the residuals have a covariance, the variance for A_s is equal to

$$\text{var}(A_s) = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_s^j \lambda_s^{jj} \text{cov}(\vec{x}_j,\vec{x}_{jj})$$

Proof: Making s = i in Theorem 2.3.12:

$$\text{cov}(A_s,A_s) = \text{var}(A_s) = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_s^j \lambda_s^{jj} \text{cov}(\vec{x}_j,\vec{x}_{jj})$$

Q.E.D.

This is a basic relationship which will be used to evaluate all the weights λ_s^j. This also shows why direct evaluation of the drift is theoretically impossible and it can be found only by successive approximations. In fact, it is necessary to know the residuals of the regionalized variable in order to calculate the estimation variance of A_s, because the covariance of these residuals is needed. But in order to obtain the residuals, we must know the drift, and this is what we are attempting to obtain in the first place. The practical implication of this apparent impasse will be discussed in Chapter 3. Here we will suppose that we already know the residual covariance.

We will now return to the problem of optimal estimation of the drift. We have reduced the problem to one of finding (n+1)k weights λ_s^j for the statistics A_s of the coefficients a_s in the analytical drift expression. We will regard a solution as optimal if the A_s minimize the estimation variance, provided the A_s are unbiased. This problem is typical of those which can be solved by Lagrange's method of multipliers (see Appendix A).

Theorem 2.3.9 gives us n+1 relationships between the λ_s^j to express the property of the estimates to be unbiased. Corollary 2.3.13 gives us the expression we must minimize. Therefore, our auxiliary function u is:

$$u = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_s^j \lambda_s^{jj} \text{cov}(\vec{x}_j,\vec{x}_{jj}) - 2 \sum_{i=0}^{n} \mu_s^i \sum_{j=1}^{k} \{\lambda_s^j f^i(\vec{x}_j) - \delta(i,s)\}$$

The partial derivative of u with respect to λ_s^j is:

$$\frac{\partial u}{\partial \lambda_s^j} = 2 \sum_{jj=1}^{k} \lambda_s^{jj} \operatorname{cov}(\vec{x}_j, \vec{x}_{jj}) - 2 \sum_{i=0}^{n} \mu_s^i f^i(\vec{x}_j)$$

for j = 1, 2, ..., k. For a given s, the system of equations will be the k $\partial u/\partial \lambda_s^j$ derivatives set equal to zero, plus the n+1 restrictions. The unknowns are the k weights λ_s^j, j = 1, 2, ... k and the n+1 Lagrangian multipliers.

$$\sum_{j=1}^{k} \lambda_s^j \operatorname{cov}(\vec{x}_1, \vec{x}_j) - \sum_{i=0}^{n} \mu_s^i f^i(\vec{x}_1) = 0$$

$$\sum_{j=1}^{k} \lambda_s^j \operatorname{cov}(\vec{x}_2, \vec{x}_j) - \sum_{i=0}^{n} \mu_s^i f^i(\vec{x}_2) = 0$$

. . . . . .

$$\sum_{j=1}^{k} \lambda_s^j \operatorname{cov}(\vec{x}_{jj}, \vec{x}_j) - \sum_{i=0}^{n} \mu_s^i f^i(\vec{x}_{jj}) = 0$$

. . . . . .

$$\sum_{j=1}^{k} \lambda_s^j \operatorname{cov}(\vec{x}_k, \vec{x}_j) - \sum_{i=0}^{n} \mu_s^i f^i(\vec{x}_k) = 0$$

$$\sum_{j=1}^{k} \lambda_s^j f^0(\vec{x}_j) = \delta(0,s)$$

$$\sum_{j=1}^{k} \lambda_s^j f^1(\vec{x}_j) = \delta(1,s)$$

. . . . . .

$$\sum_{j=1}^{k} \lambda_s^j f^i(\vec{x}_j) = \delta(i,s)$$

. . . . . .

$$\sum_{j=1}^{k} \lambda_s^j f^n(\vec{x}_j) = \delta(n,s)$$

The same system can be expressed more easily in terms of matrices, which can be done by a series of definitions.

Definition 2.3.14: Let $\operatorname{cov}(\vec{x}_j, \vec{x}_{jj})$ be the covariance of the residuals and let $f^i(\vec{x})$ be the arbitrary functions given in Definition 2.3.4 for $m(\vec{x})$. The matrix A is, by definition,

$$A = \begin{bmatrix}
\operatorname{cov}(\vec{x}_1,\vec{x}_1) & \operatorname{cov}(\vec{x}_1,\vec{x}_2) & \dots & \operatorname{cov}(\vec{x}_1,\vec{x}_k) & f^0(\vec{x}_1) & f^1(\vec{x}_1) & \dots & f^n(\vec{x}_1) \\
\operatorname{cov}(\vec{x}_2,\vec{x}_1) & \operatorname{cov}(\vec{x}_2,\vec{x}_2) & \dots & \operatorname{cov}(\vec{x}_2,\vec{x}_k) & f^0(\vec{x}_2) & f^1(\vec{x}_2) & \dots & f^n(\vec{x}_2) \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
\operatorname{cov}(\vec{x}_{jj},\vec{x}_1) & \operatorname{cov}(\vec{x}_{jj},\vec{x}_2) & \dots & \operatorname{cov}(\vec{x}_{jj},\vec{x}_k) & f^0(\vec{x}_{jj}) & f^1(\vec{x}_{jj}) & \dots & f^n(\vec{x}_{jj}) \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
\operatorname{cov}(\vec{x}_k,\vec{x}_1) & \operatorname{cov}(\vec{x}_k,\vec{x}_2) & \dots & \operatorname{cov}(\vec{x}_k,\vec{x}_k) & f^0(\vec{x}_k) & f^1(\vec{x}_k) & \dots & f^n(\vec{x}_k) \\
f^0(\vec{x}_1) & f^0(\vec{x}_2) & \dots & f^0(\vec{x}_k) & 0 & 0 & \dots & 0 \\
f^1(\vec{x}_1) & f^1(\vec{x}_2) & \dots & f^1(\vec{x}_k) & 0 & 0 & \dots & 0 \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
f^i(\vec{x}_1) & f^i(\vec{x}_2) & \dots & f^i(\vec{x}_k) & 0 & 0 & \dots & 0 \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
f^n(\vec{x}_1) & f^n(\vec{x}_2) & \dots & f^n(\vec{x}_k) & 0 & 0 & \dots & 0
\end{bmatrix}$$

Definition 2.3.15: Let λ_s^j be the unknown weights of A_s and let μ_s^i be Lagrange's multipliers. The matrix X_s, by definition, is shown below, left.

$$X_s = \begin{bmatrix}
\lambda_s^1 \\ \lambda_s^2 \\ \dots \\ \lambda_s^{jj} \\ \dots \\ \lambda_s^k \\ -\mu_s^0 \\ -\mu_s^1 \\ \dots \\ -\mu_s^i \\ \dots \\ -\mu_s^n
\end{bmatrix} \qquad B_s = \begin{bmatrix}
0 \\ 0 \\ \dots \\ 0 \\ \dots \\ 0 \\ \delta(0,s) \\ \delta(1,s) \\ \dots \\ \delta(i,s) \\ \dots \\ \delta(n,s)
\end{bmatrix}$$

Definition 2.3.16: The matrix of right-hand parts may be defined as a column of k zeroes and n+1 functions $\delta(i,s)$, as shown above, right.

Notice that although A is the most complicated matrix, it is independent of the coefficient A_s which we wish to calculate. Matrix A is a function only of the geometry given by the location of data points. Therefore, A must be calculated only once for all estimates of A_s. In addition, if the samples are regularly distributed the matrix A is independent of the point $\vec{x}_0$ where we are calculating the drift and depends only on the pattern of the samples, if we discard $\vec{x}_0$ near the map boundaries. Matrix B_s contains n+k zeroes; element k+s+1 is equal to 1.0. To calculate all of the estimates A_s, we must solve the system n+1 times; the only change from one system of equations to the other is the insertion of the unique value 1.0 in matrix B_s.

Algorithm 2.3.17: The following procedure may be used to calculate the statistic $M(\vec{x})$:

Step 1) Calculate the coefficients in matrix A. Set s = 0.

Step 2) Calculate the coefficients of B_s. Increase s by 1.

Step 3) Solve the system of equations

$$AX_s = B_s$$

Step 4) With the weights λ_s^j from X_s, calculate A_s using Definition 2.3.5. If s is less than n, return to step 2); otherwise, continue.

Step 5) The estimated drift $M(\vec{x})$ is

$$M(\vec{x}) = \sum_{s=0}^{n} A_s f^s(\vec{x})$$

END

The solution is valid only inside a neighborhood $(\vec{x}_0, r)$. It can be shown (Matheron, 1969a) that the system always has a unique real solution if and only if the n+1 functions are linearly independent. Let us study some properties of A_s that will be useful.

Theorem 2.3.18: Let A_s be the coefficient in Definition 2.3.5. Then,

$$\text{cov}(A_i, A_s) = \mu_s^i$$

Proof: By Theorem 2.3.12,

$$\text{cov}(A_i, A_s) = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_i^i \lambda_s^{jj} \text{cov}(\vec{x}_j, \vec{x}_{jj})$$

But since each λ_s^j satisfies the equation $\partial u / \partial \lambda_s^j = 0$, where u is the auxiliary function we used in Lagrange's method of multipliers, we can state that:

$$\sum_{jj=1}^{k} \lambda_s^{jj} \text{cov}(\vec{x}_j, \vec{x}_{jj}) = \sum_{ii=0}^{n} \mu_s^{ii} f^{ii}(\vec{x}_j)$$

Therefore,

$$\operatorname{cov}(A_i, A_s) = \sum_{j=1}^{k} \lambda_i^j \sum_{ii=0}^{n} \mu_s^{ii} f^{ii}(\vec{x}_j)$$

As λ_i^j is independent of ii, we can introduce the term under the second summation, interchange the order of summation, and then remove μ_s^{ii} from the second summation because μ_u^{ii} is independent of j. Hence,

$$\operatorname{cov}(A_i, A_s) = \sum_{ii=0}^{n} \mu_s^{ii} \sum_{j=1}^{k} \lambda_i^j f^{ii}(\vec{x}_j)$$

But by Theorem 2.3.9 the second sum is equal to $\delta(i,ii)$. Consequently,

$$\operatorname{cov}(A_i, A_s) = \sum_{ii=0}^{n} \mu_s^{ii} \delta(i,ii) = \mu_s^i$$

Therefore, the covariance of two coefficients A_s, A_i is equal to the multiplier μ_s^i.

Q.E.D.

Corollary 2.3.19: Let A_s be the statistics in Definition 2.3.5. Then

$$\operatorname{var}(A_s) = \mu_s^s$$

Proof: Let s = i in Theorem 2.3.18. Then

$$\operatorname{cov}(A_s, A_s) = \operatorname{var}(A_s) = \mu_s^s$$

Q.E.D.

Next we will consider optimal estimation of an analytical expression for the drift when the intrinsic function exists. If the residuals of the regionalized variables are second-order stationary, the relationship given in Theorem 2.2.3 is true. Then, matrix A may be expressed in terms of the semivariogram because all terms containing covariances can be replaced by

$$\operatorname{cov}(\vec{h}) = \operatorname{cov}(0) - \gamma(\vec{h})$$

Therefore, the system of equations in matrix A holds equally well in terms of the semivariogram.

An important difference arises when the covariance does not exist but the intrinsic hypothesis still holds for residuals of the regionalized variable. In this circumstance, the drift cannot be estimated as perfectly as in the first case. However, by introducing special restrictions, we can devise an estimator for the drift which is the same as the previous one except for an unknown constant over the interval $(\vec{x}_0, r)$. This is not a serious limitation if the

objective of calculating the drift is to subtract it from realizations of the regionalized variable to yield residuals which represent the random component of the regionalized variable. Because we will not use the residuals themselves but their differences, the unknown shift is cancelled out and the limitation is unimportant. Knowledge of the constant becomes crucial if the drift is to be contoured. However, this is a case we will analyze separately as it is not necessary to know an expression for the drift in an entire neighborhood in order to contour the drift.

To make drift estimation possible when only the intrinsic function is known, the function $f^0(\vec{x})$ in the expansion of $M(\vec{x})$ is set identically equal to 1.0. The coefficient a_0 will remain an unknown, allowing us to introduce a new restriction:

$$m(\vec{x}) = a_0 + \sum_{i=1}^{n} a_i f^i(\vec{x})$$

Definition 2.3.20: Suppose there is a semivariogram but not necessarily a covariance for the residuals. Except for a constant error, the following is an analytical expression for the drift in the neighborhood $(\vec{x}_0, r)$:

$$m'(\vec{x}) = \sum_{j=1}^{k} b_i f^i(\vec{x})$$

Now we can again define a statistic B_i as the linear combination of realizations of the regionalized variable.

Definition 2.3.21: We will define B_i as the statistic estimating b_i:

$$B_i = \sum_{j=1}^{k} \lambda_i^j Z(\vec{x}_j)$$

Definition 2.3.22: Let B_i be a statistic for b_i as defined in Definition 2.3.21. The statistic $M'(\vec{x})$ for $m'(\vec{x})$ is, by definition:

$$M'(\vec{x}) = \sum_{i=1}^{n} B_i f^i(\vec{x})$$

In order for B_i to be an optimal estimator, we require that it be unbiased and have minimum estimation variance with respect to all possible λ_i^j. To achieve this, we will introduce an additional restriction, not concerned directly with the B_i's, but with the weights λ_i^j.

Because we have assumed that $Z(\vec{x})$ is not necessarily second-order stationary, any linear combination of realizations does not necessarily have a finite variance for B_i. However, we will show $\text{var}(B_i)$ is finite if the sum of the weights λ_i^j over j is zero. In that case, if $\vec{x}_0$ is constant, it is also true that the drift coefficient B_s for $Z(\vec{x}) - Z(\vec{x}_0)$ and for $Z(\vec{x})$ are the same.

Theorem 2.3.23: Let $\vec{x}_j$, j = 1, 2, ..., k be points used to calculate the estimator B_s as in Definition 2.3.21 and let the sum of all the weights λ_s^j be zero. Then, if the residuals satisfy the intrinsic hypothesis,

$$\operatorname{cov}(B_s,B_i) = -\sum_{j=1}^{k}\sum_{jj=1}^{k} \lambda_s^j \lambda_i^{jj} \gamma(\vec{x}_j-\vec{x}_{jj})$$

Proof: By Theorem 2.3.12,

$$\operatorname{cov}(B_s,B_i) = \sum_{j=1}^{k}\sum_{jj=1}^{k} \lambda_s^j \lambda_i^{jj} \operatorname{cov}(\vec{x}_j,\vec{x}_{jj})$$

when $\sum_{j=1}^{k} \lambda_s^j = 0$ it can be proven that Theorem 2.3.12 also holds if $\operatorname{cov}(\vec{x}_j,\vec{x}_{jj})$ is no longer the covariance of the residuals $Y(\vec{x}_j)$ and $Y(\vec{x}_{jj})$ but the covariance of $Y(\vec{x}_j) - Y(\vec{x}_0)$ and $Y(\vec{x}_{jj}) - Y(\vec{x}_0)$ instead. Therefore, we can replace $\operatorname{cov}(\vec{x}_j,\vec{x}_{jj})$ with the relationship proven in Theorem 2.2.5.

$$\operatorname{cov}(\vec{x}_j,\vec{x}_{jj}) = \gamma(\vec{x}_j-\vec{x}_0) + \gamma(\vec{x}_{jj}-\vec{x}_0) - \gamma(\vec{x}_j-\vec{x}_{jj})$$

Note that $\vec{x}_0$ is a constant. Therefore,

$$\operatorname{cov}(B_s,B_i) = \sum_{j=1}^{k}\sum_{jj=1}^{k} \lambda_s^j \lambda_i^{jj} \gamma(\vec{x}_j-\vec{x}_0) + \sum_{j=1}^{k}\sum_{jj=1}^{k} \lambda_s^j \lambda_i^{jj} \gamma(\vec{x}_{jj}-\vec{x}_0) - \sum_{j=1}^{k}\sum_{jj=1}^{k} \lambda_s^j \lambda_i^{jj} \gamma(\vec{x}_j-\vec{x}_{jj})$$

In the first term, λ_s^j and $\gamma(\vec{x}_j-\vec{x}_0)$ can be taken out of the second summation. Analogously, if we reverse the order of summation in the second term, we can take λ_i^{jj} and $\gamma(\vec{x}_{jj}-\vec{x}_0)$ out of the second summation. Therefore,

$$\operatorname{cov}(B_s,B_i) = \sum_{j=1}^{k} \lambda_s^j \gamma(\vec{x}_j-\vec{x}_0) \sum_{jj=1}^{k} \lambda_i^{jj} + \sum_{jj=1}^{k} \lambda_i^{jj} \gamma(\vec{x}_{jj}-\vec{x}_0) \sum_{j=1}^{k} \lambda_s^j - \sum_{j=1}^{k}\sum_{jj=1}^{k} \lambda_s^j \lambda_i^{jj} \gamma(\vec{x}_j-\vec{x}_{jj})$$

But, by the hypothesis that the sum of the weights is zero,

$$\sum_{jj=1}^{k} \lambda_i^{jj} = \sum_{jj=1}^{k} \lambda_s^j = 0$$

and the first two terms vanish.

$\gamma(\vec{x}_j-\vec{x}_{jj})$ is the semivariogram for the random variables $Y(\vec{x}_j) - Y(\vec{x}_0)$ and $Y(\vec{x}_{jj}) - Y(\vec{x}_0)$. However,

$$\gamma[\{(Y(\vec{x}_j) - Y(\vec{x}_0)) - (Y(\vec{x}_{jj}) - Y(\vec{x}_0))\}^2] = \gamma[\{Y(\vec{x}_j) - Y(\vec{x}_{jj})\}^2]$$

Therefore, $\gamma(\vec{x}_j,\vec{x}_{jj})$ is independent of the third arbitrary point $\vec{x}_0$ and

$$\operatorname{cov}(B_s,B_i) = -\sum_{j=1}^{k}\sum_{jj=1}^{k} \lambda_s^j \lambda_i^{jj} \gamma(\vec{x}_j-\vec{x}_{jj})$$

$\gamma(\vec{x}_j - \vec{x}_{jj})$ is the intrinsic function for the residuals at locations $\vec{x}_j, \vec{x}_{jj}$.

Q.E.D.

Corollary 2.3.24: Let $Z(\vec{x})$ be a regionalized variable whose residuals satisfy the intrinsic hypothesis. Then the estimate B_s has a finite estimation variance if

$$\sum_{j=1}^{k} \lambda_s^j = 0$$

Proof: If we make s = i in Theorem 2.3.23

$$\text{cov}(B_s, B_s) = \text{var}(B_s) = -\sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_s^j \lambda_s^{jj} \gamma(\vec{x}_j - \vec{x}_{jj}) \qquad \text{for } s = 1, 2, \ldots, k$$

Since the residuals verify the intrinsic hypothesis, $\text{var}(B_s)$ is a finite sum of finite terms.

Q.E.D.

Finding the drift $M'(\vec{x})$ is analogous to finding $M(\vec{x})$ when the covariance is known. The main difference is that a slightly different set of restrictions must be considered in the minimization of the estimation of B_s. Our auxiliary function u is now:

$$u = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_s^j \lambda_s^{jj} \gamma(\vec{x}_j, \vec{x}_{jj}) + 2\mu_s^0 \sum_{j=1}^{k} \lambda_s^j + 2 \sum_{i=1}^{n} \mu_s^i \sum_{j=1}^{k} [\lambda_s^j f^i(\vec{x}_j) - \delta(i,s)]$$

and the partial derivatives become

$$\frac{\partial u}{\partial \lambda_s^j} = 2 \sum_{jj=1}^{k} \lambda_s^{jj} \gamma(\vec{x}_j, \vec{x}_{jj}) + 2\mu_s^0 + 2 \sum_{i=1}^{n} \mu_s^i f^i(\vec{x}_j)$$

Therefore, the system that must be solved is:

$$\sum_{j=1}^{k} \lambda_s^j \gamma(\vec{x}_1 - \vec{x}_j) + \mu_s^0 + \sum_{i=1}^{n} \mu_s^i f^i(\vec{x}_1) = 0$$

$$\sum_{j=1}^{k} \lambda_s^j \gamma(\vec{x}_2 - \vec{x}_j) + \mu_s^0 + \sum_{i=1}^{n} \mu_s^i f^i(\vec{x}_2) = 0$$

. . . . . .

$$\sum_{j=1}^{k} \lambda_s^j \gamma(\vec{x}_{jj} - \vec{x}_j) + \mu_s^0 + \sum_{i=1}^{n} \mu_s^i f^i(\vec{x}_{jj}) = 0$$

. . . . . .

$$\sum_{j=1}^{k} \lambda_s^j \gamma(\vec{x}_k - \vec{x}_j) + \mu_s^0 + \sum_{i=1}^{n} \mu_s^i f^i(\vec{x}_k) = 0$$

$$\sum_{j=1}^{k} \lambda_s^j = 0$$

$$\sum_{j=1}^{k} \lambda_s^j f^1(\vec{x}_j) = \delta(1,s)$$

$$\sum_{j=1}^{k} \lambda_s^j f^2(\vec{x}_j) = \delta(2,s)$$

. . . . . .

$$\sum_{j=1}^{k} \lambda_s^j f^i(\vec{x}_j) = \delta(i,s)$$

. . . . . .

$$\sum_{j=1}^{k} \lambda_s^j f^n(\vec{x}_j) = \delta(n,s)$$

If we define three matrices C, W_s and D_s, we can express the same system of equations in matrix notation.

Definition 2.3.25: Let $f^i(\vec{x})$ be any function of $\vec{x}$ and let $\gamma(\vec{x}_j - \vec{x}_{jj})$ be the semivariogram of the residuals. C is a matrix defined as:

$$C = \begin{bmatrix}
\gamma(\vec{x}_1,\vec{x}_1) & \gamma(\vec{x}_1,\vec{x}_2) & \dots & \gamma(\vec{x}_1,\vec{x}_k) & 1 & f^1(\vec{x}_1) & f^2(\vec{x}_1) & \dots & f^n(\vec{x}_1) \\
\gamma(\vec{x}_2,\vec{x}_1) & \gamma(\vec{x}_2,\vec{x}_2) & \dots & \gamma(\vec{x}_2,\vec{x}_k) & 1 & f^1(\vec{x}_2) & f^2(\vec{x}_2) & \dots & f^n(\vec{x}_2) \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
\gamma(\vec{x}_{jj},\vec{x}_1) & \gamma(\vec{x}_{jj},\vec{x}_2) & \dots & \gamma(\vec{x}_{jj},\vec{x}_k) & 1 & f^1(\vec{x}_{jj}) & f^2(\vec{x}_{jj}) & \dots & f^n(\vec{x}_{jj}) \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
\gamma(\vec{x}_k,\vec{x}_1) & \gamma(\vec{x}_k,\vec{x}_2) & \dots & \gamma(\vec{x}_k,\vec{x}_k) & 1 & f^1(\vec{x}_k) & f^2(\vec{x}_k) & \dots & f^n(\vec{x}_k) \\
1 & 1 & \dots & 1 & 0 & 0 & 0 & \dots & 0 \\
f^1(\vec{x}_1) & f^1(\vec{x}_2) & \dots & f^1(\vec{x}_k) & 0 & 0 & 0 & \dots & 0 \\
f^2(\vec{x}_1) & f^2(\vec{x}_2) & \dots & f^2(\vec{x}_k) & 0 & 0 & 0 & \dots & 0 \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
f^i(\vec{x}_1) & f^i(\vec{x}_2) & \dots & f^i(\vec{x}_k) & 0 & 0 & 0 & \dots & 0 \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
f^n(\vec{x}_1) & f^n(\vec{x}_2) & \dots & f^n(\vec{x}_k) & 0 & 0 & 0 & \dots & 0
\end{bmatrix}$$

Definition 2.3.26: Let λ_s^j be the weight of B_s in Definition 2.3.16 and let μ_s^i be Lagrange's multipliers. The matrix W_s is defined (below, left) as:

$$W_s = \begin{bmatrix} \lambda_s^1 \\ \lambda_s^2 \\ \cdots \\ \lambda_s^{jj} \\ \cdots \\ \lambda_s^k \\ \mu_s^0 \\ \mu_s^1 \\ \mu_s^2 \\ \cdots \\ \mu_s^i \\ \cdots \\ \mu_s^n \end{bmatrix} \qquad D_s = \begin{bmatrix} 0 \\ 0 \\ \cdots \\ 0 \\ \cdots \\ 0 \\ 0 \\ \delta(1,s) \\ \delta(2,s) \\ \cdots \\ \delta(i,s) \\ \cdots \\ \delta(n,s) \end{bmatrix}$$

Definition 2.3.27: Let $\delta(i,s)$ be the function in Definition 2.3.8. The matrix D_s (above, right) is a column of k + 1 zeroes followed by n function $\delta(i,s)$. Therefore,

$$CW_s = D_s$$

It can be shown (Matheron, 1969a) that the system always has a unique solution if, and only if, the n functions $f^i(\vec{x})$ are linearly independent. That is, there must not be a linear combination

$$\alpha_1 f^1(\vec{x}) + \alpha_2 f^2(\vec{x}) + \ldots + \alpha_n f^n(\vec{x})$$

which is zero and in which one of the coefficients is not zero. This is nothing more than an expression of the well-known fact that the coefficients of a plane cannot be estimated if all samples are along a line and the coefficients of a second-degree surface cannot be found if all points are on a conic.

The matrices C and D_s are again quite similar to the matrices A and B_s in the case where the covariance rather than the semivariogram was used

Setting the elements from k + 2 to k + n + 1 to 1 in matrix D_s and solving the system of equations n times for each new D_s, we obtain the n statistics B_s; the estimate of the relative drift $M'(\vec{x})$ will be:

$$M'(\vec{x}) = \sum_{i=1}^{n} B_i f^i(\vec{x})$$

which is exactly the same as $M(\vec{x})$ except for the unknown constant a_0.

$$M(\vec{x}) = M'(\vec{x}) + a_0$$

The estimation is valid inside a neighborhood of radius r.

Algorithm 2.3.28: This is an algorithm to calculate $M'(\vec{x})$.

Step 1) Calculate the coefficients in matrix C. Set s = 1.

Step 2) Calculate the coefficients for D_s. Increase s by 1.

Step 3) Solve the system of equations $CW_s = D_s$.

Step 4) With the λ_s^j from W_s, form the sum

$$B_s = \sum_{j=1}^{k} \lambda_s^j Z(\vec{x}_j)$$

If s is less than n, go to step 2) and repeat.

Step 5) $M'(\vec{x})$ is the sum

$$M'(\vec{x}) = \sum_{s=1}^{n} B_s f^s(\vec{x})$$

END

We will prove one more theorem and a corollary which we will need later.

Theorem 2.3.29: Let B_s be an optimal estimator of b_s. If μ_s^i is the Lagrangian multiplier in the matrix W_s in Definition 2.3.26,

$$\operatorname{cov}(B_s, B_i) = \mu_i^s$$

Proof: From Theorem 2.3.23,

$$\operatorname{cov}(B_s, B_i) = -\sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_s^j \lambda_i^{jj} \gamma(\vec{x}_j - \vec{x}_{jj})$$

The λ_s^j were chosen to minimize $\partial u / \partial \lambda_s^j$ in the calculation of B_s using the Lagrange method of multipliers. Hence,

$$\sum_{jj=1}^{k} \lambda_i^{jj} \gamma(\vec{x}_j, \vec{x}_{jj}) = -\mu_i^0 - \sum_{ii=1}^{n} \mu_i^{ii} f^{ii}(\vec{x}_j)$$

and

$$\operatorname{cov}(B_s, B_i) = \sum_{j=1}^{k} \lambda_s^j \mu_i^0 + \sum_{j=1}^{k} \lambda_s^j \sum_{ii=1}^{n} \mu_i^{ii} f^{ii}(\vec{x}_j)$$

The last term vanishes because μ_i^0 can be taken out of the summation and then the sum of the weights is zero. In the first term, as λ_s^j is independent of ii, we can introduce it under the summation, interchange summation order and remove μ_i^{ii} from the second summation because μ_i^{ii} is independent of j. Hence,

$$\text{cov}(B_s,B_i) = \sum_{ii=1}^{n} \mu_i^{ii} \sum_{j=1}^{k} \lambda_s^j f^{ii}(\vec{x}_j)$$

But, by Theorem 2.3.9, the second sum is equal to $\delta(ii,s)$. Therefore,

$$\text{cov}(B_s,B_i) = \sum_{ii=1}^{n} \mu_i^{ii} \delta(ii,s) = \mu_i^s$$

Q.E.D.

Corollary 2.3.30: Let μ_i^i be the multiplier defined in the matrix W_s in the system of equations for B_i. Then,

$$\text{var}(B_i) = \mu_i^i$$

Proof: Making s = i in Theorem 2.3.29,

$$\text{cov}(B_i,B_i) = \text{var}(B_i) = \mu_i^i$$

Q.E.D.

We will now consider optimal estimation of the drift at a point if the covariance exists. Previously presented analyses to estimate the drift inside a neighborhood are time-consuming, as a system of equations must be solved for each coefficient. When it is necessary to obtain an expression valid not only for a point but for a neighborhood, these calculations must be performed. This is the case, for instance, in semivariogram analysis where a unique drift expression for several observations inside a neighborhood is required in order to find the residuals. However, generating a grid in order to contour the drift requires only the values of the drift at points. If attention is restricted to just one point instead of an entire neighborhood it can be proved that the drift can be calculated by solving only one system of equations regardless of the number of coefficients involved.

Assume that there is an expected value of the regionalized variable. Assume also that the covariance for the residuals exists. A minimum number of observations at locations $\vec{x}_j$ around the point $\vec{x}_0$ where the drift is to be estimated is also required.

Definition 2.3.31: Let $Z(\vec{x})$ be a regionalized variable whose residuals have a covariance. The statistics $M^*(\vec{x}_0)$ of $m^*(\vec{x}_0)$ is defined as a linear combination of samples inside a neighborhood

of radius r around $\vec{x}_0$.

$$M^*(\vec{x}_0) = \sum_{j=1}^{k} \lambda_j \, Z(\vec{x}_j)$$

As $M^*(\vec{x}_0)$ is a linear combination of random variables $Z(\vec{x}_j)$, $M^*(\vec{x}_0)$ is also a random variable.

The desired estimator should be unbiased and have a minimum estimation variance with respect to all possible λ_j. In other words, $E[M^*(\vec{x})] = m(\vec{x})$ and $Var[M^*(\vec{x})-m(\vec{x})]$ is a minimum with respect to all possible λ_j. This is a problem of restricted maxima and minima which may be solved using Lagrange's method of multipliers. But first an expression must be found to relate the variance to the weights. Analogously, the fact of being unbiased in terms of the weights must be stated. The following lemma is the first step in this direction.

Lemma 2.3.32: Suppose there is a drift $m(\vec{x}_0)$ which can be represented by

$$m(\vec{x}_0) = \sum_{i=0}^{n} a_i \, f^i(\vec{x}_0)$$

according to Definition 2.3.4. Then, if λ_j are the weights in the statistic $M^*(\vec{x}_0)$,

$$E[M^*(\vec{x}_0)] = \sum_{i=0}^{n} \sum_{j=1}^{k} a_i \, \lambda_j \, f^i(\vec{x}_j)$$

Proof: By Definition 2.3.31

$$E[M^*(\vec{x}_0)] = E[\sum_{j=1}^{k} \lambda_j \, Z(\vec{x}_j)]$$

Interchanging expectation and summation and removing the constant λ_j from the expectation:

$$E[M^*(\vec{x}_0)] = \sum_{j=1}^{k} \lambda_j \, E[Z(\vec{x}_j)]$$

However, the expected value in the right hand side of the equation is the drift at $\vec{x}_j$. Hence,

$$E[M^*(\vec{x}_0)] = \sum_{j=1}^{k} \lambda_j \sum_{i=0}^{n} a_i \, f^i(\vec{x}_j)$$

But λ_j can be replaced in the second summation because it is independent of i. Consequently, if we interchange summation and reverse the product $\lambda_j \, a_i$,

$$E[M^*(\vec{x}_0)] = \sum_{i=0}^{n} \sum_{j=1}^{k} a_i \, \lambda_j \, f^i(\vec{x}_j)$$

Q.E.D.

Theorem 2.3.33: Let the drift $m(\vec{x}_0)$ be continuous so that there is an analytical expression

$$m(\vec{x}_0) = \sum_{j=1}^{n} a_i \, f^i(\vec{x}_0)$$

which represents the drift inside a certain neighborhood. Let

$$M^*(\vec{x}_0) = \sum_{j=1}^{k} \lambda_j \, Z(\vec{x}_j)$$

then,

$$\sum_{j=1}^{k} \lambda_j \, f^i(\vec{x}_j) = f^i(\vec{x}_0) \qquad \text{for } i = 0, 1, \ldots, n$$

if and only if $E[M^*(\vec{x}_0)] = m(\vec{x}_0)$.

Proof: Let

$$\sum_{j=1}^{k} \lambda_j \, f^i(\vec{x}_j) = f^i(\vec{x}_0) \qquad \text{for } i = 0, 1, \ldots, n$$

The equality still holds if each side of the equation is multiplied by a_i. Then, if we add f all possible values of i

$$\sum_{i=0}^{n} a_i \sum_{j=1}^{k} \lambda_j \, f^i(\vec{x}_j) = \sum_{i=1}^{k} a_i \, f^i(\vec{x}_0)$$

But by Lemma 2.3.32, in the left side of the equation we have

$$\sum_{i=0}^{n} \sum_{j=1}^{k} \lambda_j \, f^i(\vec{x}_j) = E[M^*(\vec{x}_0)]$$

and the right side is by Definition 2.3.4 equal to $m(\vec{x}_0)$. Hence,

$$E[M^*(\vec{x}_0)] = m(\vec{x}_0)$$

Let us suppose now $E[M^*(\vec{x}_0)] = m(\vec{x}_0)$. By Lemma 2.3.32

$$E[M^*(\vec{x}_0)] = \sum_{i=0}^{n} \sum_{j=1}^{k} a_i \, \lambda_j \, f^i(\vec{x}_j)$$

and by Definition 2.3.4

$$m(\vec{x}_0) = \sum_{i=0}^{n} a_i \, f^i(\vec{x}_0)$$

Therefore, by the hypothesis,

$$\sum_{i=0}^{n} \sum_{j=1}^{k} a_i \lambda_j f^i(\vec{x}_j) - \sum_{i=0}^{n} a_i f^i(\vec{x}_0) = 0$$

However, a_i can be taken out of the summation over j in the left side of the equation. Hence

$$\sum_{i=0}^{n} a_i \sum_{j=1}^{k} [\lambda_j f^i(\vec{x}_j) - f^i(\vec{x}_0)] = 0$$

But the only way for the left side to always be zero is if:

$$\sum_{j=1}^{k} \lambda_j f^i(\vec{x}_j) = f^i(\vec{x}_0)$$

The two implications thus prove the theorem.

Q.E.D.

Let us prove a Lemma which will be used to introduce the restriction that the estimation must be a minimum with respect to all possible λ_j.

Lemma 2.3.34: Let $m(\vec{x}_0)$ be the drift at $\vec{x}_0$ and let $M^*(\vec{x}_0)$ be the unbiased statistic of $m(\vec{x}_0)$ defined by

$$M^*(\vec{x}_0) = \sum_{j=1}^{k} \lambda_j Z(\vec{x}_j)$$

Then, if there is a covariance $\text{cov}(\vec{x}_j, \vec{x}_{jj})$ for the residuals,

$$E[\{M^*(\vec{x}_0)\}^2] = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj} \text{cov}(\vec{x}_j, \vec{x}_{jj}) + [m(\vec{x}_0)]^2$$

Proof: By Definition 2.3.31,

$$E[\{M^*(\vec{x}_0)\}^2] = E[\sum_{j=1}^{k} \lambda_j Z(\vec{x}_j) \sum_{jj=1}^{k} \lambda_{jj} Z(\vec{x}_{jj})]$$

Introducing the term $\lambda_j Z(\vec{x}_j)$ within the second summation and interchanging summation and expectation,

$$E[\{M^*(\vec{x}_0)\}^2] = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj} E[Z(\vec{x}_j) Z(\vec{x}_{jj})]$$

By Lemma 2.3.11,

$$E[\{M^*(\vec{x}_0)\}^2] = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj} \text{cov}(\vec{x}_j, \vec{x}_{jj}) + \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj} m(\vec{x}_j) m(\vec{x}_{jj})$$

By Definition 2.3.4

$$E[\{M^*(\vec{x}_0)\}^2] = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j\lambda_{jj} \operatorname{cov}(\vec{x}_j,\vec{x}_{jj}) + \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j\lambda_{jj} \sum_{i=0}^{n} a_i f^i(\vec{x}_j) \sum_{ii=0}^{n} a_{ii} f^{ii}(\vec{x}_{jj})$$

In the last equation it is possible to separate terms containing j from those containing jj and reverse the order of summation:

$$E[\{M^*(\vec{x}_0)\}^2] = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j\lambda_{jj} \operatorname{cov}(\vec{x}_j,\vec{x}_{jj}) + \sum_{i=0}^{n} a_i \sum_{j=1}^{k} \lambda_j f^i(\vec{x}_j) \sum_{ii=0}^{n} a_{ii} \sum_{jj=1}^{k} \lambda_{jj} f^{ii}(\vec{x}_{jj})$$

By Theorem 2.3.33

$$\sum_{j=1}^{k} \lambda_j f^i(\vec{x}_j) = f^i(\vec{x}_0)$$

Therefore,

$$E[\{M^*(\vec{x}_0)\}^2] = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j\lambda_{jj} \operatorname{cov}(\vec{x}_j,\vec{x}_{jj}) + \sum_{i=0}^{n} a_i f^i(\vec{x}_0) \sum_{ii=0}^{n} a_{ii} f^{ii}(\vec{x}_0)$$

and because the last term is equal to $[m(\vec{x}_0)]^2$ we finally obtain

$$E[\{M^*(\vec{x}_0)\}^2] = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j\lambda_{jj} \operatorname{cov}(\vec{x}_j,\vec{x}_{jj}) + [m(\vec{x}_0)]^2$$

Q.E.D.

This result will be used in the following theorem.

Theorem 2.3.35: Let $M^*(\vec{x}_0)$ be an unbiased estimator for $m(\vec{x}_0)$ as in Definition 2.3.31. Then, if there is a covariance $\operatorname{cov}(\vec{x}_j,\vec{x}_{jj})$ for the pair of residuals $Y(\vec{x}_j)$, $Y(\vec{x}_{jj})$,

$$\operatorname{var}[M^*(\vec{x}_0) - m(\vec{x}_0)] = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j\lambda_{jj} \operatorname{cov}(\vec{x}_j,\vec{x}_{jj})$$

Proof: As $E[M^*(\vec{x}_0) - m(\vec{x}_0)] = 0$, the expanded expression for the variance is

$$\operatorname{var}[M^*(\vec{x}_0) - m(\vec{x}_0)] = E[\{M(\vec{x}_0)\}^2] + E[\{m(\vec{x}_0)\}^2] - 2E[M^*(\vec{x}_0)\, m(\vec{x}_0)]$$

But as $m(\vec{x}_0)$ is a constant which can be taken out of the expectations and because $M^*(\vec{x}_0)$ is an unbiased estimate of $m(\vec{x}_0)$,

$$\operatorname{var}[M^*(\vec{x}_0) - m(\vec{x}_0)] = E[\{M^*(\vec{x}_0)\}^2] + [m(\vec{x}_0)]^2 - 2m(\vec{x}_0)\, m(\vec{x}_0)$$

But from Lemma 2.3.34

$$\text{var}[M^*(\vec{x}_0) - m(\vec{x}_0)] = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj} \text{ cov}(\vec{x}_j, \vec{x}_{jj})$$

Q.E.D.

Theorems 2.3.33 and 2.3.35 will now be used to find the optimal values of the weights λ_j for the drift estimation. The solution must be such that the λ_j's minimize the estimation variance for the drift, provided the estimator $M^*(\vec{x}_0)$ is unbiased. This conditional optimization can be satisfied using Lagrange's method of multipliers. The auxiliary function for this case is

$$u = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj} \text{ cov}(\vec{x}_j, \vec{x}_{jj}) - 2 \sum_{i=0}^{n} \mu_i \left[\sum_{j=1}^{k} \lambda_j f^i(\vec{x}_j) - f^i(\vec{x}_0) \right]$$

The partial derivatives with respect to the unknowns are:

$$\frac{\partial u}{\partial \lambda_j} = 2 \sum_{jj=1}^{k} \lambda_{jj} \text{ cov}(\vec{x}_j, \vec{x}_{jj}) - 2 \sum_{i=0}^{n} \lambda_i f^i(\vec{x}_j)$$

for j = 1, 2, ..., k. The solution for the k unknowns λ_j plus the additional n + 1 unknown multipliers μ_i will come from the following system of equations.

$$\sum_{j=1}^{k} \lambda_j \text{ cov}(\vec{x}_1, \vec{x}_j) - \sum_{i=0}^{n} \mu_i f^i(\vec{x}_1) = 0$$

$$\sum_{j=1}^{k} \lambda_j \text{ cov}(\vec{x}_2, \vec{x}_j) - \sum_{i=0}^{n} \mu_i f^i(\vec{x}_2) = 0$$

. . . . . .

$$\sum_{j=1}^{k} \lambda_j \text{ cov}(\vec{x}_{jj}, \vec{x}_j) - \sum_{i=0}^{n} \mu_i f^i(\vec{x}_{jj}) = 0$$

. . . . . .

$$\sum_{j=1}^{k} \lambda_j \text{ cov}(\vec{x}_k, \vec{x}_j) - \sum_{i=0}^{n} \mu_i f^i(\vec{x}_k) = 0$$

$$\sum_{j=1}^{k} \lambda_j f^0(\vec{x}_j) = f^0(\vec{x}_0)$$

$$\sum_{j=1}^{k} \lambda_j f^1(\vec{x}_j) = f^1(\vec{x}_0)$$

. . . . . .

$$\sum_{j=1}^{k} \lambda_j \, f^i(\vec{x}_j) = f^i(\vec{x}_0)$$

. . . . . .

$$\sum_{j=1}^{k} \lambda_j \, f^i(\vec{x}_i) = f^n(\vec{x}_0)$$

Definition 2.3.36: Let λ_j be the unknown weights of $M^*(\vec{x}_j)$ in Definition 2.3.31 and let μ_i be Lagrange's multipliers. The matrix X (below, left) is by definition

$$X = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \cdots \\ \lambda_{jj} \\ \cdots \\ \lambda_k \\ \mu_0 \\ \mu_1 \\ \cdots \\ \mu_i \\ \cdots \\ \mu_n \end{bmatrix} \qquad B = \begin{bmatrix} 0 \\ 0 \\ \cdots \\ 0 \\ \cdots \\ 0 \\ f^0(\vec{x}_0) \\ f^1(\vec{x}_0) \\ \cdots \\ f^i(\vec{x}_0) \\ \cdots \\ f^n(\vec{x}_0) \end{bmatrix}$$

The matrix containing the right-hand terms will be defined as B.

Definition 2.3.37: Let $f^i(\vec{x}_0)$ be the functions in Definition 2.3.4. The matrix B (above, right) is by definition a column with k zeroes and the functions $f^i(\vec{x}_0)$. A definition of the coefficient matrix is not required as it is the same as in Definition 2.3.14. Therefore, the system of equations, written as a matrix product, is

$$AX = B$$

The covariances in matrix A must not be negative. The system of equations has a unique solution if and only if the n + 1 functions $f^i(\vec{x})$ are linearly independent (Matheron, 1969b). That is, there cannot be a linear combination

$$\alpha_0 \, f^0(\vec{x}) + \alpha_1 \, f^1(\vec{x}) + \ldots + \alpha_n \, f^n(\vec{x}) = 0$$

except in the trivial case where all coefficients α_i are zero.

The relationships resulting from the derivatives of the system of equations allows a much simpler expression for the variance.

Theorem 2.3.38: Let X and B be the matrices in Definitions 2.3.36 and 2.3.37. Then, the estimation variance for the drift is

$$\text{var}[M^*(\vec{x}_0) - m(\vec{x}_0)] = X^T B$$

where X^T means the *transpose* of X.

Proof: As any λ_j satisfies the equation $\partial u/\partial \lambda_j = 0$, where u is the arbitrary function used to derive the system of equations AX = B, then

$$\sum_{jj=1}^{k} \lambda_{jj} \text{ cov}(\vec{x}_j, \vec{x}_{jj}) = \sum_{i=0}^{n} \mu_i f^i(\vec{x}_j)$$

Therefore, from Theorem 2.3.35

$$\text{var}[M^*(\vec{x}_0) - m(\vec{x}_0)] = \sum_{j=1}^{k} \lambda_j \sum_{i=0}^{n} \mu_i f^i(\vec{x}_j)$$

As λ_j is independent of i, the term can be introduced under the second summation. Interchanging the order of summation and removing μ_i from the second summation because μ_i is independent of j yields

$$\text{var}[M^*(\vec{x}_0) - m(\vec{x}_0)] = \sum_{i=0}^{n} \mu_i \sum_{j=1}^{k} \lambda_j f^i(\vec{x}_j)$$

But from Theorem 2.3.33,

$$\text{var}[M^*(\vec{x}_0) - m(\vec{x}_0)] = \sum_{i=0}^{n} \mu_i f^i(\vec{x}_0)$$

However, by Definition 2.3.36 and 2.3.37,

$$\text{var}[M^*(\vec{x}_0) - m(\vec{x}_0)] = X^T B$$

Q.E.D.

The following algorithm summarizes the practical results obtained for the drift estimation at one point when the covariance for the residuals is known.

Algorithm 2.3.39: This is an algorithm to calculate the drift and its estimation variance at a point $\vec{x}_0$ where the covariance for the residuals exists and the drift can be represented by terms of $f^i(\vec{x})$ inside a restricted neighborhood.

Step 1) Calculate the terms in matrix A in Definition 2.3.14.

Step 2) Calculate the terms in matrix B in Definition 2.3.37.

Step 3) Solve the linear system of equations

$$AX = B$$

Step 4) Using the resulting weights λ_j in matrix X in Definition 2.3.36, calculate the estimate $M^*(\vec{x}_0)$ for the drift as:

$$M^*(\vec{x}_0) = \sum_{j=1}^{k} \lambda_j \, Z(\vec{x}_j)$$

$\vec{x}_j$ being the locations of k samples within the interval $(\vec{x}_0, r)$ where the approximation in Definition 2.3.4 is valid.

Step 5) Perform the product $X^T B$. The result is the estimation variance for the drift.

END

Drift estimation at a point where there is only an intrinsic function leads to certain complications. If the covariance does not exist, the drift is no longer a slowly varying function, but is the realization of a random function. If further complications and assumptions are introduced, the drift can be estimated. However, in physical problems, the drift is meaningful only if it changes smoothly. Therefore, as this complication has no practical consequences and is extremely complex, we will omit this analysis. Refer to Matheron (1971) who solves this particular situation by using "weighting functions." He finds that only the estimation variance depends on these weighting functions and that the drift estimate does not.

Point estimation of the drift for second-order stationary residuals with a linear semivariogram is a special case of the analysis summarized as Algorithm 2.3.39. Let us start by defining three matrices:

Definition 2.3.40: Let $|\vec{x}_j - \vec{x}_{jj}|$ be the distance between two locations $\vec{x}_j$ and $\vec{x}_{jj}$ and let $f^i(\vec{x})$ be the function of $\vec{x}$ in Definition 2.3.4. Denote by G the matrix defined as:

$$G = \begin{bmatrix}
0 & |\vec{x}_1-\vec{x}_2| & \dots & |\vec{x}_1-\vec{x}_k| & 1 & f^1(\vec{x}_1) & f^2(\vec{x}_1) & \dots & f^n(\vec{x}_1) \\
|\vec{x}_2-\vec{x}_1| & 0 & \dots & |\vec{x}_2-\vec{x}_k| & 1 & f^1(\vec{x}_2) & f^2(\vec{x}_2) & \dots & f^n(\vec{x}_2) \\
\dots & & & & & & & & \\
|\vec{x}_{jj}-\vec{x}_1| & |\vec{x}_{jj}-\vec{x}_2| & \dots & |\vec{x}_{jj}-\vec{x}_k| & 1 & f^1(\vec{x}_{jj}) & f^2(\vec{x}_{jj}) & \dots & f^n(\vec{x}_{jj}) \\
\dots & & & & & & & & \\
|\vec{x}_k-\vec{x}_1| & |\vec{x}_k-\vec{x}_2| & \dots & 0 & 1 & f^1(\vec{x}_k) & f^2(\vec{x}_k) & \dots & f^n(\vec{x}_k) \\
1 & 1 & \dots & 1 & 0 & 0 & 0 & \dots & 0 \\
f^1(\vec{x}_1) & f^1(\vec{x}_2) & \dots & f^1(\vec{x}_k) & 0 & 0 & 0 & \dots & 0 \\
f^2(\vec{x}_1) & f^2(\vec{x}_2) & \dots & f^2(\vec{x}_k) & 0 & 0 & 0 & \dots & 0 \\
\dots & & & & & & & & \\
f^i(\vec{x}_1) & f^i(\vec{x}_2) & \dots & f^i(\vec{x}_k) & 0 & 0 & 0 & \dots & 0 \\
\dots & & & & & & & & \\
f^n(\vec{x}_1) & f^n(\vec{x}_2) & \dots & f^n(\vec{x}_k) & 0 & 0 & 0 & \dots & 0
\end{bmatrix}$$

Definition 2.3.41: Let λ_j and μ_i be unknowns and let ω be the slope of a linear semivariogram for the residuals. Then T is the column defined below, left

$$T = \begin{bmatrix}
\lambda_1 \\ \lambda_2 \\ \dots \\ \lambda_{jj} \\ \dots \\ \lambda_k \\ \mu_0/\omega \\ \mu_1/\omega \\ \mu_2/\omega \\ \dots \\ \mu_i/\omega \\ \dots \\ \mu_n/\omega
\end{bmatrix} \qquad Q = \begin{bmatrix}
\text{cov}(0)/\omega \\ \text{cov}(0)/\omega \\ \dots \\ \text{cov}(0)/\omega \\ \dots \\ \text{cov}(0)/\omega \\ 1 \\ f^1(\vec{x}) \\ f^2(\vec{x}) \\ \dots \\ f^i(\vec{x}) \\ \dots \\ f^n(\vec{x})
\end{bmatrix}$$

Definition 2.3.42: Let cov(0) be the covariance for the residuals at lag zero, that is, the variance of the residuals. Then, if $f^i(\vec{x})$ is a function of $\vec{x}$ as in Definition 2.3.4, Q (above right) is defined as a column containing k terms cov(0)/ω, then a 1, and finally the terms $f^i(\vec{x})$.

If the residuals are second-order stationary, Theorem 2.2.3 holds. In addition, assume the intrinsic function is a line through the origin with slope ω. Then,

$$\text{cov}(\vec{x}_j,\vec{x}_{jj}) = \text{cov}(0) - \omega|\vec{x}_j-\vec{x}_{jj}| \qquad \text{for } |\vec{x}_j,\vec{x}_{jj}| < 2r$$

where r is the radius of the neighborhood in which $m(\vec{x})$ can be represented by an analytical expression as in Definition 2.3.4.

Replacing this particular expression for the covariance in the system of equations in Algorithm 2.3.39 and dividing the first k equations by ω yields a new system of equations,

$$GT = Q$$

if it is assumed that $f^0(\vec{x}) = 1$ in Definition 2.3.4. Let us examine some properties of the new system of equations.

Theorem 2.3.43: Let the residuals be second-order stationary with a linear semivariogram. Then if $f^0(\vec{x}) = 1$ in Definition 2.3.4, the statistic $M^*(\vec{x})$ for the drift is independent of the variance of the residuals cov(0) and the semivariogram slope ω.

Proof: By Definition 2.3.31, the value of $M^*(\vec{x})$ depends only on the observations considered and on the weights λ_j. But the weights come from the solution of the equation GT = Q. This system of equations contains the terms cov(0) and ω only on the right side of the first k equations. If any two k different pairs of equations are subtracted among the first k equations, we will have an equivalent system of equations which will yield weights λ_j, and therefore, an estimator $M^*(\vec{x})$ independent of the parameters cov(0) and ω.

Q.E.D.

By Theorem 2.3.3, the residuals have mean zero, therefore

$$\text{cov}(0) = \frac{1}{ns-1} \sum_{j=1}^{ns} Y^2(\vec{x}_j)$$

where ns is the number of observations and $Y(\vec{x}_j)$ are the residuals. However, a more practical approach is to obtain an estimate of cov(0) from the semivariogram. This can be done if we can determine the sill, as we have seen that

$$\gamma(h) = \text{cov}(0)$$

for those values of $\gamma(h)$ larger than the range.

Theorem 2.3.44: Let the residuals be second-order stationary with a linear semivariogram. Then if $f^0(\vec{x}) = 1$ in Definition 2.3.4, the estimation variance for the drift at $\vec{x}_0$ depends on the residual variance cov(0) and the semivariogram slope ω in the following way:

$$\text{var}[M^*(\vec{x}_0) - m(\vec{x}_0)] = \omega T^T Q$$

Proof: If we return to the application of Lagrange's method of multipliers which yielded the system of equations AX = B in Algorithm 2.3.39, there is an auxiliary function u which contains the covariance. The expression for the covariance may be replaced with the equivalent equation for a linear semivariogram when the residuals are second-order stationary, giving:

$$u = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj} [\text{cov}(0) - \omega|\vec{x}_j - \vec{x}_{jj}|] - 2 \sum_{i=0}^{n} \mu_i [\sum_{j=1}^{k} \lambda_j f^i(\vec{x}_j) - f^i(\vec{x}_0)]$$

But $f^0(\vec{x}) = 1$ implies $\sum_{j=1}^{k} \lambda_j = 1$. Therefore, as cov(0) is a constant:

$$u = \text{cov}(0) - \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj} \omega|\vec{x}_j - \vec{x}_{jj}| - 2 \sum_{i=0}^{n} \mu_i [\sum_{j=1}^{k} \lambda_j f^i(\vec{x}_j) - f^i(\vec{x}_0)]$$

But u must be minimal with respect to any λ_j. Therefore

$$\frac{\partial u}{\partial \lambda_j} = -2 \sum_{jj=1}^{k} \lambda_{jj} \omega|\vec{x}_j - \vec{x}_{jj}| - 2 \sum_{i=0}^{n} \mu_i f^i(\vec{x}_j) = 0$$

Theorem 2.3.35 can be written in terms of the semivariogram as

$$\text{var}[M^*(\vec{x}_0) - m(\vec{x}_0)] = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj} [\text{cov}(0) - \omega|\vec{x}_j - \vec{x}_{jj}|]$$

But as $\sum_{j=1}^{k} \lambda_j = 1$ because $f^0(\vec{x}) = 1$, from the relationship shown above for $\partial\mu/\partial\lambda_j$

$$\text{var}[M^*(\vec{x}_0) - m(\vec{x}_0)] = \text{cov}(0) + \sum_{j=1}^{k} \lambda_j \sum_{i=0}^{n} \mu_i f^i(\vec{x}_j) = 0$$

As λ_j is independent of i, the weights can be introduced under the second summation. The order of summation can be interchanged and μ_i taken from the second summation because μ_i is independent of j. Consequently,

$$\text{var}[M^*(\vec{x}_0) - m(\vec{x}_0)] = \text{cov}(0) + \sum_{i=0}^{n} \mu_i \sum_{j=1}^{k} \lambda_j f^i(\vec{x}_j)$$

but from Theorem 2.3.33,

$$\text{var}[M^*(\vec{x}_0) - m(\vec{x}_0)] = \text{cov}(0) + \sum_{i=0}^{n} \mu_i f^i(\vec{x}_0)$$

Finally, from Definitions 2.3.41 and 2.3.42,

$$\text{var}[M^*(\vec{x}_0) - m(\vec{x}_0)] = \omega T^T Q$$

Q.E.D

Therefore, even though the point estimate for the drift $M^*(\vec{x}_0)$ does not depend on either the variance for the residuals or on the slope of the semivariogram, the estimation variance for the point estimate for drift depends on both. The independence of $M^*(\vec{x}_0)$ from the slope of the semivariogram does not mean the drift estimate is independent of the semivariogram. The semivariogram has another parameter, the neighborhood size, which ultimately determines the observations to be used in the estimation.

Algorithm 2.3.45: This is an algorithm for point estimation of the drift and its estimation variance when the residuals are second-order stationary; the semivariogram in any direction is a line through the origin with slope ω for an interval of length 2r; and $f^0(\vec{x}) = 1$ as in Definition 2.3.4. To estimate the drift and its estimation variance at location $\vec{x}_0$, do the following steps:

Step 1) Calculate the terms for matrix G in Definition 2.3.40, using all those samples whose location $\vec{x}_j$ is inside a circle of radius r and center $\vec{x}_0$.

Step 2) If the value of the estimation variance for the drift is needed, introduce the value of cov(0) in Q in Definition 2.3.42. Otherwise set cov(0) equal to zero in Q.

Step 3) Solve the linear system of equations

$$GT = Q$$

where T is the matrix in Definition 2.3.41.

Step 4) Use the weights λ_j in T to calculate the statistic

$$M^*(\vec{x}_0) = \sum_{j=1}^{k} \lambda_j \, Z(\vec{x}_j)$$

using all the samples described in step 2). $M(\vec{x}_0)$ is the point estimate for the drift at location $\vec{x}_0$.

Step 5) If the estimation variance is not needed or cov(0) = 0 in step 2), then terminate. Otherwise, compute the product $\omega T^T Q$. The result is the estimation variance for the difference between the real and the estimated drift.

END

2.4 Universal Kriging

The prediction of values of a regionalized variable in areas of sparse control has preoccupied researchers for many years. There are almost as many proposed solutions as there are persons interested in the problem. Solutions differ widely because, although the problem is universal, practical considerations in different fields of application lead to emphasis on different approaches. In addition, the backgrounds of the researchers are very diverse.

Despite the great amount of effort and money spent on the problem of estimation of regionalized variables, we have already pointed out in section 1.4 that all methods have at least one serious defect. *Universal kriging*, in contrast, is the best linear unbiased method to estimate a regionalized variable. The only objection to universal kriging is that its optimality is accompanied by great mathematical complexity which usually (but not necessarily) means higher implementation and execution costs. The only apparent reason to reject the use of universal kriging is to save money. This may or may not be a critical consideration, depending on the reliability required in the estimations. If high reliability is not essential, universal kriging is probably not the best method from a point of view that considers not only the characteristics of the algorithm but also its operational cost. However, there is no reason to reject the use of universal kriging in a study requiring high precision and accuracy.

The theoretical advantages of universal kriging have been substantiated through statistical tests using different types of data (Walden, 1972). According to Walden, "the Kriging algorithm consistently gives the highest surface correlations and the lowest surface error measures of any method...."

The problem we will solve using universal kriging will be the optimal interpolation or extrapolation of the value of the regionalized variable Z at location $\vec{x}$. It is probably worthwhile to say that universal kriging and the numerical results are, or course, quite different from the estimation of the drift of Z at location $\vec{x}$. Most of the alternative methods to kriging as an estimation procedure in automatic contouring are a form of trend surface applied to reduced neighborhoods. A trend surface is a least squares regression on geographic coordinates. The resulting trend surface polynomials do not have a physical meaning. They are neither drift nor an estimated surface for the regionalized variable. They are simply surfaces which minimize the squared difference between the control values and the values defined by a polynomial expansion of a particular degree. Unless the polynomial equation contains n - 1 terms, where n is the number of control points, surfaces fitted by polynomial interpolation procedures are not exact because they fail to coincide with the sample values at control points. However, it is well known that the perfect fit obtained by such a large number of terms is fallacious. Increasing the degree of the polynomial equation leads to tremendous and meaningless fluctuations in the surface between control points (Matheron, 1967b).

Unlike other estimation procedures, universal kriging does not yield an analytical function which can be evaluated within an interval $\vec{x}_0$. Universal kriging produces an estimate which is a number, the most probable value for the regionalized variable at $\vec{x}_0$. Unless the samples have been collected on a regular pattern, the estimation procedure must be repeated as many times as there are points to be estimated. We will discuss later economizations allowed by regular sampling patterns.

We will study two different cases, depending on whether the covariance of the residuals or the semivariogram of the residuals are known. In both, we will assume we have sufficient samples measured without error and inside a neighborhood of radius r around the point $\vec{x}_0$ to be estimated, 2r being less than or equal to the semivariogram range. We will also assume that the samples are

point observations and that the estimated value has an associated point geometric support. A greater generalization for this application is not necessary, as we will see in Chapter 3. This restricted, although sufficient, subset of kriging is far easier to understand than the more general formulations that are usually presented (Matheron, 1969a; Journel, 1969).

We will now consider universal kriging when the covariance of the residuals is known.

Definition 2.4.1: Let $Z(\vec{x})$ be a non-stationary regionalized variable whose residuals at the control points have a covariance $\text{cov}(\vec{x},\vec{y})$. We define the statistic $Z^*(\vec{x})$ of $Z(\vec{x})$ as a linear combination of the samples inside a neighborhood of radius r around $\vec{x}$.

$$Z^*(\vec{x}) = \sum_{j=1}^{k} \lambda_j \, Z(\vec{x}_j)$$

As $Z^*(\vec{x})$ is a linear combination of random variables $Z(\vec{x}_j)$, $Z^*(\vec{x})$ is also a random variable.

Optimization of the statistic $Z^*(\vec{x})$ will be performed by imposing the constraints

$$E[Z^*(\vec{x}_0) - Z(\vec{x}_0)] = 0$$

$$E[\{Z^*(\vec{x}_0) - Z(\vec{x}_0)\}^2] \quad \text{is a minimum with respect to } \lambda_j.$$

These restrictions imply that the difference $Z^*(\vec{x}_0) - Z(\vec{x}_0)$ is unbiased and the variance of this difference is a minimum. In an attempt to calculate the k unknown weights, we must prove several theorems and lemmas.

Lemma 2.4.2: Let

$$m(\vec{x}) = \sum_{i=0}^{n} a_i \, f^i(\vec{x})$$

be the expression for the drift of the regionalized variable at $\vec{x}_0$. Then, if λ_j are the weights in the statistic $Z^*(\vec{x})$,

$$E[Z^*(\vec{x}_0)] = \sum_{i=0}^{n} \sum_{j=1}^{k} a_i \lambda_j \, f^i(\vec{x}_j)$$

Proof: By Definition 2.4.1,

$$E[Z^*(\vec{x}_0)] = E[\sum_{j=1}^{k} \lambda_j \, Z(\vec{x}_j)]$$

But we can interchange expectation and summation and remove the constant λ_j from the expectation:

$$E[Z^*(\vec{x}_0)] = \sum_{j=1}^{k} \lambda_j \, E[Z(\vec{x}_j)]$$

However, this expected value is the drift at $\vec{x}_j$. Hence,

$$E[Z^*(\vec{x}_0)] = \sum_{j=1}^{k} \lambda_j \sum_{i=0}^{n} a_i f^i(\vec{x}_j)$$

But λ_j can be placed in the second summation because it is independent of i. Therefore, if we interchange summation and reverse the product $\lambda_j a_i$,

$$E[Z^*(\vec{x}_0) = \sum_{i=0}^{n} \sum_{j=1}^{k} a_i \lambda_j f^i(\vec{x}_j)$$

Q.E.D.

Theorem 2.4.3: Let $E[Z(\vec{x}_0)]$ be the drift at $\vec{x}_0$. Let the analytical expression for the drift at $\vec{x}_0$ be:

$$m(\vec{x}_0) = \sum_{i=0}^{n} a_i f^i(\vec{x}_0)$$

and let

$$Z^*(\vec{x}_0) = \sum_{j=1}^{k} \lambda_j Z(\vec{x}_j)$$

Then,

$$\sum_{j=1}^{k} \lambda_j f^i(\vec{x}_j) = f^i(\vec{x}_0) \qquad \text{for } i = 0, 1, \ldots, n$$

if and only if $E[Z^*(\vec{x}_0) - Z(\vec{x}_0)] = 0$.

Proof: Let us suppose $E[Z^*(\vec{x}_0) - Z(\vec{x}_0)] = 0$. By Lemma 2.4.2,

$$E[Z^*(\vec{x}_0)] = \sum_{i=0}^{n} \sum_{j=1}^{k} a_i \lambda_j f^i(\vec{x}_j)$$

We may then interchange summation order and take a_i out of the second summation. By hypothesis,

$$E[Z^*(\vec{x}_0)] = E[Z(\vec{x}_0)]$$

Hence, from Lemma 2.4.2 and Definitions 2.3.1 and 2.3.4,

$$\sum_{i=0}^{n} a_i \sum_{j=1}^{k} \lambda_j f^i(\vec{x}_j) = \sum_{i=0}^{n} a_i f^i(\vec{x}_0)$$

Therefore,

$$\sum_{j=1}^{k} \lambda_j f^i(\vec{x}_j) = f^i(\vec{x}_0) \qquad \text{for } i = 0, 1, \ldots, n$$

if $E[Z^*(\vec{x}_0) - Z(\vec{x}_0)] = 0$.

Let

$$\sum_{j=1}^{k} \lambda_j f^i(\vec{x}_j) = f^i(\vec{x}_0) \qquad \text{for } i = 0, 1, \ldots, n$$

In this case,

$$a_i \sum_{j=1}^{k} \lambda_j f^i(\vec{x}_j) = a_i f^i(\vec{x}_0) \qquad \text{for } i = 0, 1, \ldots, n$$

Hence, the sum of all these terms can be set equal to:

$$\sum_{i=0}^{n} a_i \sum_{j=1}^{k} \lambda_j f^i(\vec{x}_j) = \sum_{i=0}^{k} a_i f^i(\vec{x}_0)$$

Let us examine each term in the equality. On the left side of the equation, by Lemma 2.4.2,

$$\sum_{i=0}^{n} a_i \sum_{j=1}^{k} \lambda_j f^i(\vec{x}_j) = E[Z^*(\vec{x}_0)]$$

On the right side of the initial equation, by Definition 2.3.4,

$$\sum_{i=0}^{n} a_i f^i(\vec{x}_0) = E[Z(\vec{x}_0)]$$

Therefore,

$$E[Z^*(\vec{x}_0)] = E[Z(\vec{x}_0)]$$

or

$$E[Z^*(\vec{x}_0) - Z(\vec{x}_0)] = 0$$

Hence,

$$\sum_{j=1}^{k} \lambda_j f^i(\vec{x}_j) = f^i(\vec{x}_0)$$

implies

$$E[Z^*(\vec{x}_0) - Z(\vec{x}_0)] = 0$$

The two implications prove Theorem 2.4.3.

Q.E.D.

Theorem 2.4.3 is an important result because it provides an alternative expression of the constraint that the difference $Z^*(\vec{x}_0)$ and $Z(\vec{x}_0)$ must be unbiased. Our interest will be in obtaining an expression for the estimation variance. The following three lemmas are intermediate steps in accomplishing this goal.

Lemma 2.4.4: Let $m(\vec{x}_0)$ be the drift at $\vec{x}_0$ and let $Z^*(\vec{x}_0)$ be the statistic of $Z(\vec{x}_0)$ defined by

$$Z^*(\vec{x}_0) = \sum_{j=1}^{k} \lambda_j \, Z(\vec{x}_j)$$

Then, if there is a covariance $\text{cov}(\vec{x},\vec{y})$ for the residuals,

$$E[\{Z^*(\vec{x}_0)\}^2] = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj} \, \text{cov}(\vec{x}_j,\vec{x}_{jj}) + [m(\vec{x}_0)]^2$$

Proof: By Definition 2.4.1,

$$E[\{Z^*(\vec{x}_0)\}^2] = E[\sum_{j=1}^{k} \lambda_j \, Z(\vec{x}_j) \sum_{jj=1}^{k} \lambda_{jj} \, Z(\vec{x}_{jj})]$$

Introducing the term $\lambda_j \, Z(\vec{x}_j)$ within the second summation and interchanging summation and expectation,

$$E[\{Z^*(\vec{x}_0)\}^2] = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj} \, E[Z(\vec{x}_j) \, Z(\vec{x}_{jj})]$$

By Lemma 2.3.11,

$$E[\{Z^*(\vec{x}_0)\}^2] = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj} \, \text{cov}(\vec{x}_j,\vec{x}_{jj}) + \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj} \, m(\vec{x}_j) \, m(\vec{x}_{jj})$$

By Definition 2.3.4,

$$E[\{Z^*(\vec{x}_0)\}^2] = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj} \, \text{cov}(\vec{x}_j,\vec{x}_{jj}) + \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj} \sum_{i=0}^{n} a_i \, f^i(\vec{x}_j) \sum_{ii=0}^{n} a_{ii} \, f^{ii}(\vec{x}_{jj})$$

In the last equation it is possible to separate terms containing j from those containing jj and reverse the order of summation:

$$E[\{Z^*(\vec{x}_0)\}^2] = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj} \, \text{cov}(\vec{x}_j,\vec{x}_{jj}) + \sum_{i=0}^{n} a_i \sum_{j=1}^{k} \lambda_j \, f^i(\vec{x}_j) \sum_{ii=0}^{n} a_{ii} \sum_{jj=1}^{k} \lambda_{jj} \, f^{ii}(\vec{x}_{jj})$$

By Theorem 2.4.3,

$$\sum_{j=1}^{k} \lambda_j \, f^i(\vec{x}_j) = f^i(\vec{x}_0)$$

Therefore,

$$E[\{Z^*(\vec{x}_0)\}^2] = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj} \operatorname{cov}(\vec{x}_j, \vec{x}_{jj}) + \sum_{i=0}^{n} a_i f^i(\vec{x}_0) \sum_{ii=0}^{n} a_{ii} f^{ii}(\vec{x}_0)$$

Because the last term is equal to $[m(x_0)]^2$, we finally obtain

$$E[\{Z^*(\vec{x}_0)\}^2] = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj} \operatorname{cov}(\vec{x}_j, \vec{x}_{jj}) + [m(\vec{x}_0)]^2$$

Q.E.D.

Lemma 2.4.5: Let $m(\vec{x}_0)$ be the drift at location $\vec{x}_0$ and let $\operatorname{cov}(\vec{x},\vec{y})$ be the covariance of the residuals. Then,

$$E[\{Z(\vec{x}_0)\}^2] = \operatorname{cov}(\vec{x}_0, \vec{x}_0) + [m(\vec{x}_0)]^2$$

Proof: From Definition 2.3.2,

$$E[\{Z(\vec{x}_0)\}^2] = E[\{Y(\vec{x}_0) + m(\vec{x}_0)\}^2]$$

Expanding and taking the expectation of each term yields:

$$E[\{Z(\vec{x}_0)\}^2] = E[Y(\vec{x}_0)\ Y(\vec{x}_0)] + 2E[Y(\vec{x}_0)\ m(\vec{x}_0)] + E[\{m(\vec{x}_0)\}^2]$$

But $m(\vec{x}_0)$ is a constant and can be taken out of the expectations:

$$E[\{Z(\vec{x}_0)\}^2] = E[Y(\vec{x}_0)\ Y(\vec{x}_0)] + 2m(\vec{x}_0)\ E[Y(\vec{x}_0)] + m^2(\vec{x}_0)$$

Because the mean of the residuals is zero, the second term vanishes. For the same reason, the first term is the covariance of the residuals.

$$E[\{Z(\vec{x}_0)\}^2] = \operatorname{cov}(\vec{x}_0, \vec{x}_0) + [m(\vec{x}_0)]^2$$

Q.E.D.

Lemma 2.4.6: Let $m(\vec{x}_0)$ be the drift of $Z(\vec{x}_0)$ at $\vec{x}_0$. Let $Z^*(\vec{x}_0)$ be the statistic for $Z(\vec{x}_0)$ as in Definition 2.4.1. Then,

$$E[Z^*(\vec{x}_0)\ Z(\vec{x}_0)] = \sum_{j=1}^{k} \lambda_j \operatorname{cov}(\vec{x}_j, \vec{x}_0) + [m(\vec{x}_0)]^2$$

Proof: Replacing $Z^*(\vec{x}_0)$ by its definition and $Z(\vec{x}_0)$ in terms of Definition 2.3.2 yields:

$$E[Z^*(\vec{x}_0)\ Z(\vec{x}_0)] = E[\{\sum_{j=1}^{k} \lambda_j\ Z(\vec{x}_j)\}\ \{Y(\vec{x}_0) + m(\vec{x}_0)\}]$$

Expanding and replacing $Z(\vec{x}_j)$ by Definition 2.3.2,

$$E[Z^*(\vec{x}_0)\ Z(\vec{x}_0)] = E[\sum_{j=1}^{k} \lambda_j \{Y(\vec{x}_j)\ Y(\vec{x}_0) + m(\vec{x}_0)\ Y(\vec{x}_j) + m(\vec{x}_j)\ Y(\vec{x}_0) + m(\vec{x}_0)\ m(\vec{x}_j)\}]$$

We can interchange expectation and summation and take the expectation of each term, remembering that λ_j and $m(\vec{x})$ are constants and can be removed from the expectations:

$$E[Z^*(\vec{x}_0)\ Z(\vec{x}_0)] = \sum_{j=1}^{k} \lambda_j \{E[Y(\vec{x}_j)\ Y(\vec{x}_0)] + m(\vec{x}_0)\ E[Y(\vec{x}_j)] + m(\vec{x}_j)\ E[Y(\vec{x}_0)] + m(\vec{x}_0)\ m(\vec{x}_j)\}$$

Since the mean of the residuals is zero, the second and third terms vanish, and for the same reason,

$$E[Y(\vec{x}_j)\ Y(\vec{x}_0)] = \mathrm{cov}(\vec{x}_j,\vec{x}_0)$$

Expanding the sum gives:

$$E[Z^*(\vec{x}_0)\ Z(\vec{x}_0)] = \sum_{j=1}^{k} \lambda_j\ \mathrm{cov}(\vec{x}_j,\vec{x}_0) + \sum_{j=1}^{k} \lambda_j\ m(\vec{x}_0)\ m(\vec{x}_j)$$

In the last term, $m(\vec{x}_0)$ can be taken out of the summation and $m(\vec{x}_j)$ can be replaced by Definition 2.3.4:

$$E[Z^*(\vec{x}_0)\ Z(\vec{x}_0)] = \sum_{j=1}^{k} \lambda_j\ \mathrm{cov}(\vec{x}_j,\vec{x}_0) + m(\vec{x}_0) \sum_{j=1}^{k} \lambda_j \sum_{i=0}^{n} a_i\ f^i(\vec{x}_j)$$

By Theorem 2.4.3,

$$\sum_{i=0}^{n} a_i\ f^i(\vec{x}_j) = f^i(\vec{x}_0)$$

Therefore,

$$E[Z^*(\vec{x}_0)\ Z(\vec{x}_0)] = \sum_{j=1}^{k} \lambda_j\ \mathrm{cov}(\vec{x}_j,\vec{x}_0) + m(\vec{x}_0) \sum_{i=0}^{n} a_i\ f^i(\vec{x}_0)$$

But the last summation is equal to $m(\vec{x}_0)$. Hence,

$$E[Z^*(\vec{x}_0)\ Z(\vec{x}_0)] = \sum_{j=1}^{k} \lambda_j\ \mathrm{cov}(\vec{x}_j,\vec{x}_0) + [m(\vec{x}_0)]^2$$

Q.E.D.

Now we will prove a theorem which is basic to the calculation of the unknown weights λ_j used in the statistic $Z^*(\vec{x}_0)$.

Theorem 2.4.7: Let $Z^*(\vec{x}_0)$ be a statistic for $Z(\vec{x}_0)$ as in Definition 2.4.1. Then, if there is a covariance $\text{cov}(\vec{x}_j,\vec{x}_{jj})$ for the residuals $Y(\vec{x}_j)$,

$$\text{var}[Z^*(\vec{x}_0) - Z(\vec{x}_0)] = \text{cov}(\vec{x}_0,\vec{x}_0) - 2\sum_{j=1}^{k} \lambda_j \text{cov}(\vec{x}_j,\vec{x}_0) + \sum_{j=1}^{k}\sum_{jj=1}^{k} \lambda_j\lambda_{jj} \text{cov}(\vec{x}_j,\vec{x}_{jj})$$

Proof: As $E[Z^*(\vec{x}_0) - Z(\vec{x}_0)]$ is zero, if we expand the square in $E[\{Z^*(\vec{x}_0) - Z(\vec{x}_0)\}^2]$ and take the expected value of each term,

$$\text{var}[Z^*(\vec{x}_0) - Z(\vec{x}_0)] = E[\{Z^*(\vec{x}_0)\}^2] + E[\{Z(\vec{x}_0)\}^2] - 2E[Z^*(\vec{x}_0)\, Z(\vec{x}_0)]$$

Replacing appropriate terms from Lemmas 2.4.4, 2.4.5, and 2.4.6,

$$\text{var}[Z^*(\vec{x}_0) - Z(\vec{x}_0)] = \sum_{j=1}^{k}\sum_{jj=1}^{k} \lambda_j\lambda_{jj} \text{cov}(\vec{x}_j,\vec{x}_{jj}) + [m(\vec{x}_0)]^2 + \text{cov}(\vec{x}_0,\vec{x}_0) + [m(\vec{x}_0)]^2$$

$$- 2\sum_{j=1}^{k} \lambda_j \text{cov}(\vec{x}_j,\vec{x}_0) - 2[m(\vec{x}_0)]^2$$

Therefore,

$$\text{var}[Z^*(\vec{x}_0) - Z(\vec{x}_0)] = \text{cov}(\vec{x}_0,\vec{x}_0) - 2\sum_{j=1}^{k} \lambda_j \text{cov}(\vec{x}_j,\vec{x}_0) + \sum_{j=1}^{k}\sum_{jj=1}^{k} \lambda_j\lambda_{jj} \text{cov}(\vec{x}_j,\vec{x}_{jj})$$

Q.E.D.

The optimal estimator $Z^*(\vec{x})$ will be comprised of those λ_j which minimize the expression in Theorem 2.4.7, provided the restriction on Theorem 2.4.3 holds. We can satisfy this conditional optimization using Lagrange's multiplier method. The auxiliary function for this case is:

$$u = \text{cov}(\vec{x}_0,\vec{x}_0) - 2\sum_{j=1}^{k} \lambda_j \text{cov}(\vec{x}_j,\vec{x}_0) + \sum_{j=1}^{k}\sum_{jj=1}^{k} \lambda_j\lambda_{jj} \text{cov}(\vec{x}_j,\vec{x}_{jj})$$

$$- 2\sum_{i=0}^{n} \mu_i \left[\sum_{j=1}^{k} \lambda_j f^i(\vec{x}_j) - f^i(\vec{x}_0)\right]$$

The partial derivatives with respect to each of the unknowns are:

$$\frac{\partial u}{\partial \lambda_j} = 2\,\text{cov}(\vec{x}_j,\vec{x}_0) + 2\sum_{jj=1}^{k} \lambda_{jj} \text{cov}(\vec{x}_j,\vec{x}_{jj}) - 2\sum_{i=0}^{n} \mu_i f^i(\vec{x}_j)$$

for $j = 1, 2, \ldots, k$. Therefore, the solution for the $k\lambda_j$ and $n + 1$ μ_i unknowns will come from the following system of equations.

$$\sum_{j=1}^{k} \lambda_j \operatorname{cov}(\vec{x}_1,\vec{x}_j) - \sum_{i=0}^{n} \mu_i f^i(\vec{x}_1) = \operatorname{cov}(\vec{x}_1,\vec{x}_0)$$

$$\sum_{j=1}^{k} \lambda_j \operatorname{cov}(\vec{x}_2,\vec{x}_j) - \sum_{i=0}^{n} \mu_i f^i(\vec{x}_2) = \operatorname{cov}(\vec{x}_2,\vec{x}_0)$$

. . . . . .

$$\sum_{j=1}^{k} \lambda_j \operatorname{cov}(\vec{x}_{jj},\vec{x}_j) - \sum_{i=0}^{n} \mu_i f^i(\vec{x}_{jj}) = \operatorname{cov}(\vec{x}_{jj},\vec{x}_0)$$

. . . . . .

$$\sum_{j=1}^{k} \lambda_j \operatorname{cov}(\vec{x}_k,\vec{x}_j) - \sum_{i=0}^{n} \mu_i f^i(\vec{x}_k) = \operatorname{cov}(\vec{x}_k,\vec{x}_0)$$

$$\sum_{j=1}^{k} \lambda_j f^0(\vec{x}_j) = f^0(\vec{x}_0)$$

$$\sum_{j=1}^{k} \lambda_j f^1(\vec{x}_j) = f^1(\vec{x}_0)$$

. . . . . .

$$\sum_{j=1}^{k} \lambda_j f^i(\vec{x}_j) = f^i(\vec{x}_0)$$

. . . . . .

$$\sum_{j=1}^{k} \lambda_j f^n(\vec{x}_j) = f^n(\vec{x}_0)$$

This system of equations is easier to handle using matrix notation. Although we need three matrices, we do not need to define all of them because the square coefficient matrix is the same as that used in Algorithm 2.3.17.

Definition 2.4.8: Below (left) we denote by U the matrix containing the unknowns:

$$U = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \cdots \\ \lambda_{jj} \\ \cdots \\ \lambda_k \\ -\mu_0 \\ -\mu_1 \\ \cdots \\ -\mu_i \\ \cdots \\ -\mu_n \end{bmatrix} \qquad E = \begin{bmatrix} \text{cov}(\vec{x}_1, \vec{x}_0) \\ \text{cov}(\vec{x}_2, \vec{x}_0) \\ \cdots \\ \text{cov}(\vec{x}_{jj}, \vec{x}_0) \\ \cdots \\ \text{cov}(\vec{x}_k, \vec{x}_0) \\ f^0(\vec{x}_0) \\ f^1(\vec{x}_0) \\ \cdots \\ f^i(\vec{x}_0) \\ \cdots \\ f^n(\vec{x}_0) \end{bmatrix}$$

Definition 2.4.9: Let $\text{cov}(\vec{x},\vec{y})$ be the covariance for the residuals and let $f^i(\vec{x})$ be arbitrary functions of x. We define the matrix E above right. We have now reduced the universal kriging problem to one of the solution of a system of simultaneous linear equations:

$$AU = E$$

Recall that A is not a function of the location of the k samples. Although E is a function of the individual points $\vec{x}_j$ and $\vec{x}_0$, it can be proved that the solution U to the system of equations remains the same under a change of coordinates. In other words, the same weighting coefficients result if all points $\vec{x}_j$ and $\vec{x}_0$ are translated the same amount $\vec{h}$. This can result in a powerful simplification if the data are collected on a regular pattern. It is not necessary to solve the system of equations for every point to be estimated, but only as many times as there are different patterns of points. This is a significant saving, because it may be necessary to solve the system only three or four times for most of the map area, plus some special cases near the map boundaries.

Instead of repeatedly solving a system of equations, it is necessary to recognize spatial patterns; universal kriging in this form is no more expensive than the most simplified estimation algorithm for regionalized variables. If data are collected on a regular pattern, there is no reason to consider any algorithm other than kriging. However, if points are irregularly spaced, the system of equations must be solved as many times as there are points to be estimated. Computer time and expense then become significant factors in the selection of a contouring algorithm.

Algorithm 2.4.10: This is an algorithm to perform universal kriging when the covariance for the residuals is known.

Step 1) Calculate terms for matrix A in Definition 2.3.14.

Step 2) Calculate terms for matrix E in Definition 2.4.9.

Step 3) Solve the system of equations AU = E.

Step 4) Using the λ_j weights in U in Definition 2.4.8, $Z^*(\vec{x}_0)$ is found by

$$Z^*(\vec{x}_0) = \sum_{j=1}^{k} \lambda_j \, Z(\vec{x}_j)$$

where $\vec{x}_j$ are locations of k samples inside the influence zone around $\vec{x}_0$.

END

The following theorem is not of immediate necessity, but will be useful later.

Theorem 2.4.11: Let $Z^*(\vec{x}_0)$ be the optimal estimator for a regionalized variable at location $\vec{x}_0$ using universal kriging and let that regionalized variable have a covariance $\text{cov}(\vec{x},\vec{y})$ for its residuals. Then,

$$\text{var}[Z^*(\vec{x}_0) - Z(\vec{x}_0)] = \text{cov}(\vec{x}_0,\vec{x}_0) - U^T E$$

Proof: As the λ_j's satisfy the Lagrangian constraint,

$$\sum_{jj=1}^{k} \lambda_{jj} \, \text{cov}(\vec{x}_j,\vec{x}_{jj}) = \text{cov}(\vec{x}_j,\vec{x}_0) + \sum_{i=0}^{n} \mu_i \, f^i(\vec{x}_j)$$

for j = 1, 2, ..., k. We can multiply both sides of this equation by λ_j and sum the k equalities. Then,

$$\sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj} \, \text{cov}(\vec{x}_j,\vec{x}_{jj}) = \sum_{j=1}^{k} \lambda_j \, \text{cov}(\vec{x}_j,\vec{x}_0) + \sum_{j=1}^{k} \lambda_j \sum_{i=0}^{n} \mu_i \, f^i(\vec{x}_j)$$

In the last term, we can introduce λ_j under the second summation, reverse the order of the summation and take μ_i out of the second summation:

$$\sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj} \, \text{cov}(\vec{x}_j,\vec{x}_{jj}) = \sum_{j=1}^{k} \lambda_j \, \text{cov}(\vec{x}_j,\vec{x}_0) + \sum_{i=0}^{n} \mu_i \sum_{j=1}^{k} \lambda_j \, f^i(\vec{x}_j)$$

But, by Theorem 2.4.3 the second summation in the final term is $f^i(\vec{x}_0)$. Therefore,

$$\sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj} \, \text{cov}(\vec{x}_j,\vec{x}_{jj}) = \sum_{j=1}^{k} \lambda_j \, \text{cov}(\vec{x}_j,\vec{x}_0) + \sum_{i=0}^{n} \mu_i \, f^i(\vec{x}_0)$$

From this equality and Theorem 2.4.7, we finally obtain:

$$\text{var}[Z^*(\vec{x}_0) - Z(\vec{x}_0)] = \text{cov}(\vec{x}_0,\vec{x}_0) - \sum_{j=1}^{k} \lambda_j \, \text{cov}(\vec{x}_j,\vec{x}_0) + \sum_{i=0}^{n} \mu_i \, f^i(\vec{x}_0)$$

which, by the definition of the matrices U and E is:

$$\text{var}[Z^*(\vec{x}_0) - Z(\vec{x})] = \text{cov}(\vec{x}_0,\vec{x}_0) - U^T E$$

Q.E.D.

Next we will examine universal kriging with a known intrinsic function for the residuals. If the regionalized variable has second-order stationary residuals, we can replace the covariance by the semivariogram in matrices A and E in Algorithm 2.4.10 and universal kriging is then expressed in terms of the semivariogram. This approach is actually different only if we know just the semivariogram and not the covariance of the residuals. The solution for universal kriging when we have a semivariogram but not necessarily a covariance for the residuals is fairly straightforward, considering previous results, as will be shown now.

Definition 2.4.12: Let $Z(\vec{x})$ be a non-stationary regionalized variable whose residuals satisfy the intrinsic hypothesis. The statistic for $Z_s^*(\vec{x})$ is, by definition,

$$Z_s^*(\vec{x}) = \sum_{j=1}^{k} \lambda_j \, Z(\vec{x}_j)$$

We say $Z_s^*(\vec{x}_0)$ is optimal if:

$$E[Z_s^*(\vec{x}_0) - Z(\vec{x}_0)] = 0$$

$$\text{var}[Z_s^*(\vec{x}_0) - Z(\vec{x}_0)] \quad \text{is a minimum of } \lambda_j.$$

Remember that when we did not know the covariance for the residuals, we assumed in the previous section that the drift would be:

$$m(\vec{x}_0) = a_0 + \sum_{i=1}^{n} b_i \, f^i(\vec{x}_0)$$

a_0 being an unknown and $f^0(\vec{x}_0) = 1$, for all $\vec{x}_0$. This change introduces a slight modification in Theorem 2.4.3 which we will present in the following form:

Theorem 2.4.13: Let the analytical expression for the drift at x_0 be:

$$m(\vec{x}_0) = a_0 + \sum_{i=0}^{n} b_i \, f^i(\vec{x}_0)$$

Then

$$\begin{cases} \sum_{j=1}^{k} \lambda_j \, f^i(\vec{x}_j) = f^i(\vec{x}_0) & \text{for } i = 1, 2, \ldots, n \\ \\ \sum_{j=1}^{k} \lambda_j = 1 \end{cases}$$

if and only if $E[Z_s^*(\vec{x}_0) - Z(\vec{x}_0)]$ is zero.

Proof: Let $E[Z_s^*(\vec{x}_0) - Z(\vec{x}_0)] = 0$. If we rewrite $m(\vec{x}_0)$ as

$$\begin{cases} m(\vec{x}_0) = \sum_{i=0}^{n} b_i \, f^i(\vec{x}_0) \\ f^0(\vec{x}_0) = 1 \quad \text{for all } \vec{x}_0 \end{cases}$$

The proof follows exactly that given for the first part of Theorem 2.4.3, except at the end, where for i = 0 we have:

$$\begin{cases} \sum_{j=1}^{k} \lambda_j \, f^0(\vec{x}_j) = f^0(\vec{x}_0) \quad \text{for } i = 0, 1, 2, \ldots, n \\ f^i(\vec{x}_0) = 1 \quad \text{for all } \vec{x}_0 \end{cases}$$

Therefore, the case of i = 0 reduces to:

$$\sum_{j=1}^{k} \lambda_j = 1$$

The second part of this proof also follows identically to the second part of Theorem 2.4.3, except for the final step. Therefore,

$$\begin{cases} \sum_{j=1}^{k} \lambda_j \, f^i(\vec{x}_j) = f^i(\vec{x}_0) \quad \text{for } i = 1, 2, \ldots, n \\ \sum_{j=1}^{k} \lambda_j = 1 \end{cases}$$

if and only if $E[Z_s^*(x_0) - Z(x)]$ is zero.

Q.E.D.

This slight modification to Theorem 2.4.3 allows some other modifications in the following theorem, which is equivalent to Theorem 2.4.7.

Theorem 2.4.14: Let $Z(\vec{x})$ be a random variable which is not second-order stationary but whose residuals verify the intrinsic hypothesis. Then, if $\sum_{j=1}^{k} \lambda_j = 1$ and the statistic $Z_s^*(\vec{x}_0)$ is unbiased, $\text{var}[Z^*(\vec{x}_0) - Z(\vec{x}_0)]$ is finite.

Proof: As $Z_s^*(\vec{x}_0)$ is unbiased,

$$\text{var}[Z_s^*(\vec{x}_0) - Z(\vec{x}_0)] = E[\{Z_s^*(\vec{x}_0) - Z(\vec{x}_0)\}^2]$$

Replacing the statistic by Definition 2.4.1,

$$\text{var}[Z_s^*(\vec{x}_0) - Z(\vec{x}_0)] = E[\{\sum_{j=1}^{k} \lambda_j Z(\vec{x}_j) - Z(\vec{x}_0)\}\{\sum_{jj=1}^{k} \lambda_{jj} Z(\vec{x}_{jj}) - Z(\vec{x}_0)\}]$$

However, as $\sum_{j=1}^{k} \lambda_j = 1$ and $\vec{x}_0$ is constant,

$$\text{var}[Z_s^*(\vec{x}_0) - Z(\vec{x}_0)] = E[\sum_{j=1}^{k} \lambda_j\{Z(\vec{x}_j) - Z(\vec{x}_0)\} \sum_{jj=1}^{k} \lambda_{jj}\{Z(\vec{x}_{jj}) - Z(\vec{x}_0)\}]$$

Introducing $\lambda_j[Z(\vec{x}_j) - Z(\vec{x}_0)]$ within the second summation and interchanging summation and expectation,

$$\text{var}[Z_s^*(\vec{x}_0) - Z(\vec{x}_0)] = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j\lambda_{jj} E[\{Z(\vec{x}_j) - Z(\vec{x}_0)\}\{Z(\vec{x}_{jj}) - Z(\vec{x}_0)\}]$$

By Lemma 2.3.11,

$$\text{var}[Z_s^*(\vec{x}_0) - Z(\vec{x}_0)] = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j\lambda_{jj} \text{cov}(\vec{x}_j,\vec{x}_{jj}) + \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j\lambda_{jj} [m(\vec{x}_j)$$

$$- m(\vec{x}_0)] [m(\vec{x}_{jj}) - m(\vec{x}_0)]$$

As $\text{cov}(\vec{x}_j,\vec{x}_{jj})$ is the covariance of the residual differences $Y(\vec{x}_j) - Y(\vec{x}_0)$ and $Y(\vec{x}_{jj}) - Y(\vec{x}_0)$, by Theorem 2.2.5,

$$\text{cov}(\vec{x}_j,\vec{x}_{jj}) = \gamma(\vec{x}_j-\vec{x}_0) + \gamma(\vec{x}_{jj}-\vec{x}_0) - \gamma(\vec{x}_j-\vec{x}_{jj})$$

In the second term of the right side of the equality, by Definition 2.3.4, we have

$$\sum_{j=1}^{k} \lambda_j[m(\vec{x}_j) - m(\vec{x}_0)] = \sum_{j=1}^{k} \lambda_j \sum_{i=0}^{n} a_i[f^i(\vec{x}_j) - f^i(\vec{x}_0)]$$

Introducing λ_j in the second summation and reversing summation order,

$$\sum_{j=1}^{k} \lambda_j[m(\vec{x}_j) - m(\vec{x}_0)] = \sum_{i=0}^{n} a_i[\sum_{j=1}^{k} \lambda_j f^i(\vec{x}_j) - \sum_{j=1}^{k} \lambda_j f^i(\vec{x}_0)]$$

By the restriction over λ_j and because $f^i(\vec{x}_0)$ is independent of j, the second summation over j in the right side of the equation reduces to $f^i(\vec{x}_0)$. According to Theorem 2.4.3,

$$\sum_{j=1}^{k} \lambda_j f^i(\vec{x}_j) = f^i(\vec{x}_0)$$

Therefore,

$$\sum_{j=1}^{k} \lambda_j[m(\vec{x}_j) - m(\vec{x}_0)] = 0$$

Analogously,

$$\sum_{j=1}^{k} \lambda_{jj}[m(\vec{x}_{jj}) - m(\vec{x}_0)] = 0$$

Consequently, the expression for the variance reduces to:

$$\text{var}[Z_s^*(\vec{x}_0) - Z(\vec{x}_0)] = \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j\lambda_{jj}[\gamma(\vec{x}_j-\vec{x}_0) + \gamma(\vec{x}_{jj}-\vec{x}_0)$$

$$- \gamma(\vec{x}_j-\vec{x}_{jj})]$$

Separating the right side of the equation,

$$\text{var}[Z_s^*(\vec{x}_0) - Z(\vec{x}_0)] = \sum_{j=1}^{k} \lambda_j \gamma(\vec{x}_j-\vec{x}_0) \sum_{jj=1}^{k} \lambda_{jj} + \sum_{jj=1}^{k} \lambda_{jj} \gamma(\vec{x}_{jj}-\vec{x}_0) \sum_{j=1}^{k} \lambda_j$$

$$- \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j\lambda_{jj} \gamma(\vec{x}_j-\vec{x}_{jj})$$

But, because the sum of the weights is one, the first two terms on the right side of the equation are equal. Therefore,

$$\text{var}[Z_s^*(\vec{x}_0) - Z(\vec{x}_0)] = 2\sum_{j=1}^{k} \lambda_j \gamma(\vec{x}_j-\vec{x}_0) - \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j\lambda_{jj} \gamma(\vec{x}_j-\vec{x}_{jj})$$

which is a linear combination of finite terms. Therefore, $\text{var}[Z_s^*(\vec{x}_0) - Z(\vec{x}_0)]$ is finite.

Q.E.D.

Once more, using Theorem 2.4.13 and Theorem 2.4.14, we can find the optimal weights λ_j in $Z_s^*(\vec{x}_0)$ using Lagrange's method of multipliers. The auxiliary function is now:

$$u = 2\sum_{j=1}^{k} \lambda_j \gamma(\vec{x}_j-\vec{x}_0) - \sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j\lambda_{jj} \gamma(\vec{x}_j-\vec{x}_{jj}) - 2\mu_0\left(\sum_{j=1}^{k} \lambda_j - 1\right)$$

$$- 2\sum_{i=1}^{n} \mu_i\left[\sum_{j=1}^{k} \lambda_j f^i(\vec{x}_j) - f^i(\vec{x}_0)\right]$$

and we have to minimize all the partial derivatives with respect to all the λ_j.

$$\frac{\partial u}{\partial \lambda_j} = 2\gamma(\vec{x}_j-\vec{x}_0) - 2\sum_{jj=1}^{k} \lambda_{jj} \gamma(\vec{x}_j-\vec{x}_{jj}) - 2\mu_0 - 2\sum_{i=1}^{n} \mu_i f^i(\vec{x}_j)$$

for j = 1, 2, ..., k. The k derivatives plus the n+1 restrictions provide a system of equations to compute the k λ_j's, the n μ_i's and μ_0.

$$\sum_{j=1}^{k} \lambda_j \gamma(\vec{x}_1-\vec{x}_j) + \mu_0 + \sum_{i=1}^{n} \mu_i f^i(\vec{x}_1) = \gamma(\vec{x}_1-\vec{x}_0)$$

$$\sum_{j=1}^{k} \lambda_j \gamma(\vec{x}_2-\vec{x}_j) + \mu_0 + \sum_{i=1}^{n} \mu_i f^i(\vec{x}_2) = \gamma(\vec{x}_2-\vec{x}_0)$$

. . . . . .

$$\sum_{j=1}^{k} \lambda_j \gamma(\vec{x}_{jj}-\vec{x}_j) + \mu_0 + \sum_{i=1}^{n} \mu_i f^i(\vec{x}_{jj}) = \gamma(\vec{x}_{jj}-\vec{x}_0)$$

. . . . . .

$$\sum_{j=1}^{k} \lambda_j \gamma(\vec{x}_k-\vec{x}_j) + \mu_0 + \sum_{i=1}^{n} \mu_i f^i(\vec{x}_k) = \gamma(\vec{x}_k-\vec{x}_0)$$

$$\sum_{j=1}^{k} \lambda_j = 1$$

$$\sum_{j=1}^{k} \lambda_j f^1(\vec{x}_j) = f^1(\vec{x}_0)$$

$$\sum_{j=1}^{k} \lambda_j f^2(\vec{x}_j) = f^2(\vec{x}_0)$$

. . . . . .

$$\sum_{j=1}^{k} \lambda_j f^i(\vec{x}_j) = f^i(\vec{x}_0)$$

. . . . . .

$$\sum_{j=1}^{k} \lambda_j f^n(\vec{x}_j) = f^n(\vec{x}_0)$$

This system of equations is easier to handle, visualize, and solve by computers if we define two matrices and use Definition 2.3.25.

Definition 2.4.15: V will be by definition the matrix (below, left) containing the unknowns:

$$V = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \cdots \\ \lambda_{jj} \\ \cdots \\ \lambda_k \\ \mu_0 \\ \mu_1 \\ \mu_2 \\ \cdots \\ \mu_i \\ \cdots \\ \mu_n \end{bmatrix} \qquad F = \begin{bmatrix} \gamma(\vec{x}_1-\vec{x}_0) \\ \gamma(\vec{x}_2-\vec{x}_0) \\ \cdots \\ \gamma(\vec{x}_{jj}-\vec{x}_0) \\ \cdots \\ \gamma(\vec{x}_k-\vec{x}_0) \\ 1 \\ f^1(\vec{x}_0) \\ f^2(\vec{x}_0) \\ \cdots \\ f^i(\vec{x}_0) \\ \cdots \\ f^n(\vec{x}_0) \end{bmatrix}$$

Definition 2.4.16: Let $\gamma(\vec{x}-\vec{y})$ be the semivariogram of the residuals and let $f^i(\vec{x})$ be an arbitrary function of $\vec{x}$. Then, we define the matrix F as above (right). Therefore, the universal kriging problem is to calculate the vector V in the system of equations:

$$CV = F$$

The terms containing $(\vec{x}_j-\vec{x}_{jj})$ depend only on the relative locations of samples and the location of the point to be estimated, $\vec{x}_0$. It can be proved that the entire system is invariant under a translation in $\vec{h}$. This represents a considerable saving in computation for regular patterns of data.

Algorithm 2.4.17: This is an algorithm to perform universal kriging when the semivariogram for residuals is known.

Step 1) Calculate the terms for matrix C in Definition 2.3.25.

Step 2) Calculate the terms for matrix F in Definition 2.4.16.

Step 3) Solve the linear system of equations

$$CV = F$$

Step 4) Using the λ_j weights in V in Definition 2.4.15, the kriged value $Z_s^*(\vec{x}_0)$ is:

$$Z_s^*(\vec{x}_0) = \sum_{j=1}^{k} \lambda_j \, Z(\vec{x}_j)$$

$\vec{x}_j$ being the locations of k samples within an interval $(\vec{x}_0, r)$.

END

Notice that although we used a drift containing an unknown constant, the universal kriging solution is exact as it does not depend on a_0. The universal kriging system of equations is consistent with the condition stated in the introduction that there can be only one sample per location. If more than one observation was provided for a sample point, it would result in a singular matrix in Algorithm 2.4.17. Universal kriging is not a smoothing or averaging procedure: if there is more than one observation taken at any sample point, a unique value must be selected in advance of kriging.

As we will see in the following theorem, universal kriging not only provides an estimation of the regionalized variable at $\vec{x}_0$, as any other method does, but also is able to provide a confidence interval.

Theorem 2.4.18: The estimation variance associated with the random variable $Z_s^*(\vec{x}_0) - Z(\vec{x}_0)$ given by the solution of the system of equations CV = F is:

$$\text{var}[Z_s^*(\vec{x}_0) - Z(\vec{x}_0)] = V^T F$$

Proof: From Lagrange's method,

$$\sum_{jj=1}^{k} \lambda_{jj}\, \gamma(\vec{x}_j - \vec{x}_{jj}) = \gamma(\vec{x}_j - \vec{x}_0) - \mu_0 - \sum_{i=1}^{n} \mu_i\, f^i(\vec{x}_j)$$

for j = 1, 2, ..., k. The k equalities are still true if we multiply them by λ_j. The sum of the resulting terms in the first member must be equal to the sum of the 3k resulting terms on the right-hand side of the equalities. Hence,

$$\sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj}\, \gamma(\vec{x}_j - \vec{x}_{jj}) = \sum_{j=1}^{k} \lambda_j\, \gamma(\vec{x}_j - \vec{x}_0) - \sum_{j=1}^{k} \lambda_j\, \mu_0 - \sum_{j=1}^{k} \lambda_j \sum_{i=1}^{n} \mu_i\, f^i(\vec{x}_j)$$

and μ_0 can be taken out of the summation. In the preceding term, λ_j can be introduced under the second summation, the summation can be reversed, and μ_i taken out of the second summation. Hence,

$$\sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj}\, \gamma(\vec{x}_j - \vec{x}_{jj}) = \sum_{j=1}^{k} \lambda_j\, \gamma(\vec{x}_j - \vec{x}_0) - \mu_0 \sum_{j=1}^{k} \lambda_j - \sum_{i=1}^{n} \mu_i \sum_{j=1}^{k} \lambda_j\, f^i(\vec{x}_j)$$

and by Theorem 2.4.13,

$$\sum_{j=1}^{k} \sum_{jj=1}^{k} \lambda_j \lambda_{jj}\, \gamma(\vec{x}_j - \vec{x}_{jj}) = \sum_{j=1}^{k} \lambda_j\, \gamma(\vec{x}_j - \vec{x}_0) - \mu_0 - \sum_{i=1}^{n} \mu_i\, f^i(\vec{x}_0)$$

From this equality and Theorem 2.4.14,

$$\text{var}[Z_s^*(\vec{x}_0) - Z(\vec{x}_0)] = 2 \sum_{j=1}^{k} \lambda_j \gamma(\vec{x}_j - \vec{x}_0) - \sum_{j=1}^{k} \lambda_j \gamma(\vec{x}_j - \vec{x}_0) + \mu_0 + \sum_{i=1}^{n} \mu_i f^i(\vec{x}_0)$$

Therefore,

$$\text{var}[Z_s^*(\vec{x}_0) - Z(\vec{x}_0)] = \sum_{j=1}^{k} \lambda_j \gamma(\vec{x}_j - \vec{x}_0) + \mu_0 + \sum_{i=1}^{n} \mu_i f^i(\vec{x}_0)$$

But this is exactly the product $V^T F$. Hence,

$$\text{var}[Z_s^*(\vec{x}_0) - Z(\vec{x}_0)] = V^T F$$

Q.E.D.

Although it is generally true that estimation by universal kriging is expensive on a point by point basis, we obtain with no additional effort an estimation of the reliability of the estimates. Although estimating the reliability of an interpolated value is an important matter, no other method provides such a measure of reliability.

Error estimation allows us to estimate confidence intervals about the surface. The statistic $Z_s^*(\vec{x}_0)$ plus the confidence interval are really the final results of the whole universal kriging procedure. This agrees with the definition of a regionalized variable, which is considered to be a continuous random variable. A single, specific value has a probability of occurrence of zero, and hence is meaningless. When considering continuous random variables, only intervals have a finite probability of occurrence.

Universal kriging is an exact interpolation procedure. Interpolation is the process of finding a value of a function among known values by a procedure other than that which created the known values themselves. Interpolation is exact if the procedure reproduces the known values at the control points. The following theorem proves that universal kriging is an exact interpolation procedure.

Theorem 2.4.19: Universal kriging is an exact interpolation procedure.

Proof: By Theorem 2.4.13,

$$\sum_{j=1}^{k} \lambda_j f^i(\vec{x}_j) = f^i(\vec{x}_0) \qquad \text{for } i = 1, 2, \ldots, n.$$

If any $\vec{x}_{jj} = \vec{x}_0$, then:

$$\lambda_1 f^i(\vec{x}_1) + \lambda_2 f^i(\vec{x}_2) + \ldots + (\lambda_{jj} - 1) f^i(\vec{x}_{jj}) + \ldots + \lambda_k f^i(\vec{x}_k) = 0$$

for $i = 1, 2, \ldots, n$. The only way to satisfy this equality for any $f^i(\vec{x}_j)$, k and n is to have:

$$\lambda_1 = \lambda_2 = \ldots = \lambda_{jj} - 1 = \ldots = \lambda_k = 0$$

Therefore, if $\vec{x}_0 = \vec{x}_{jj}$,

$$Z_s^*(\vec{x}_0) = \sum_{j=1}^{k} \lambda_j \, Z(\vec{x}_j) = Z(\vec{x}_{jj})$$

Therefore, universal kriging is an exact interpolation procedure.

Q.E.D.

This theorem expresses another advantage of universal kriging over other regionalized variable estimation procedures used in automatic contouring. Commonly used techniques which employ least-squares polynomial surface fitting over a reduced neighborhood are not exact interpolation procedures and incorrect estimates are obtained even at the control points. However, as no one uses interpolation procedures to estimate points already known, this seems to have gone largely unnoticed. However, when a point is later estimated, the tendency is to consider the estimation error free, the best possible, and forget that it is impossible for this estimate to be optimal if even the estimations at control points are incorrect. According to the most elementary considerations of statistical inference, away from control points the estimation should be even worse, on the average. The problem is especially serious as none of these arbitrary interpolation procedures provides an assessment of the possible error.

Universal kriging may be performed assuming a linear semivariogram. When the intrinsic function for the residuals of a regionalized variable is a linear expression, as

$$\gamma(\vec{x}_j - \vec{x}_{jj}) = \bar{\omega}|\vec{x}_j - \vec{x}_{jj}| \qquad \text{for } |\vec{x}_j - \vec{x}_{jj}| \leq 2r$$

Algorithm 2.4.17 can be rewritten in terms of coefficient matrices which are simpler than C and F. In addition, we will show that the statistic $Z_s^*(\vec{x}_0)$ is independent of the slope $\bar{\omega}$. Let us define matrix H.

Definition 2.4.20: Let $|\vec{x}_j - \vec{x}_0|$ be the distance between location $\vec{x}_j$ and the location $\vec{x}_0$ to be estimated. Let the $f(\vec{x})$'s be the arbitrary function in Definition 2.3.20. Then, the column H is given by the definition at the top of the following page. The system of equations in this particular instance becomes:

$$GT = H$$

with G and T being the matrices in Definitions 2.3.40 and 2.3.41 respectively. The independence of $Z_s^*(\vec{x}_0)$ and $\bar{\omega}$ can be easily shown.

$$H = \begin{bmatrix} |\vec{x}_1-\vec{x}_0| \\ |\vec{x}_2-\vec{x}_0| \\ \cdots \\ |\vec{x}_{jj}-\vec{x}_0| \\ \cdots \\ |\vec{x}_k-\vec{x}_0| \\ 1 \\ f^1(\vec{x}_0) \\ f^2(\vec{x}_0) \\ \cdots \\ f^i(\vec{x}_0) \\ \cdots \\ f^n(\vec{x}_0) \end{bmatrix}$$

Theorem 2.4.21: Let the semivariogram of the residual of a regionalized variable be a linear function of the distance between two points.

$$\gamma(\vec{x}_j-\vec{x}_{jj}) = \bar{\omega}|\vec{x}_j-\vec{x}_{jj}| \qquad \text{for } |\vec{x}_j-\vec{x}_{jj}| \leqslant 2r$$

Then the statistic $Z_s^*(\vec{x}_0)$ is independent of the slope $\bar{\omega}$.

Proof: The solution of the system of equations in Algorithm 2.4.17 does not change if we replace the linear expression for the semivariogram and divide the first k equations by $\bar{\omega}$. It can be easily shown that the new system of equations is:

$$GT = H$$

G and T are the matrices in Definitions 2.3.40 and 2.3.41 respectively. But both G and H depend only on the distances between points and the arbitrary functions $f^i(\vec{x})$. Therefore, the unknown weights λ_j are independent of $\bar{\omega}$. Because the samples $Z(\vec{x}_j)$ are also independent of $\bar{\omega}$, then the statistic $Z_s^*(\vec{x}_0)$ is also independent of $\bar{\omega}$.

Q.E.D.

The fact that the statistic $Z_s^*(\vec{x}_0)$ is independent of the slope of the semivariogram of the residuals does not mean that it is independent of the semivariogram. The semivariogram has a second parameter, which defines a radius r for the neighborhood where the linear approximation of the semivariogram holds. The neighborhood $(\vec{x}_0,r)$ determines the samples $Z(\vec{x}_j)$,

$j = 1, 2, \ldots, k$, which can be used by the statistic $Z_s^*(\vec{x}_0)$. Therefore, $Z_s^*(\vec{x}_0)$ depends on the semivariogram of the residuals through the radius r.

Theorem 2.4.22: Let the semivariogram for the residuals of a regionalized variable be a linear expression with slope $\bar{\omega}$. Then the estimation variance is directly proportional to the slope of the semivariogram for the residuals.

Proof: From Theorem 2.4.18,

$$\text{var}[Z_s^*(\vec{x}_0) - Z(\vec{x}_0)] = V^T F$$

But, as the semivariogram is linear, from Definitions 2.4.15, 2.4.16, 2.3.41, and 2.4.20,

$$V^T F = \bar{\omega} T^T H$$

Therefore,

$$\text{var}[Z_s^*(\vec{x}_0) - Z(\vec{x}_0)] = \bar{\omega} T^T H$$

Hence, the estimation variance is directly proportional to the slope of the semivariogram of the residuals.

Q.E.D.

Algorithm 2.4.23 is a particular case of Algorithm 2.4.17 in which the semivariogram of the residuals is linear. The new algorithm is based on the two preceding theorems and is a fundamental result for the next chapter.

Algorithm 2.4.23: This is an algorithm to perform universal kriging and to evaluate the estimation variance when the semivariogram is a linear function in an interval of length 2r.

Step 1) Calculate the terms for matrix G in Definition 2.3.40, using all those samples whose locations $\vec{x}_j$ are not farther than r from the location $\vec{x}_0$ where we want to estimate the regionalized variable.

Step 2) Calculate the terms for matrix H in Definition 2.4.20, using the same samples as in Step 1).

Step 3) Solve the linear system of equations

$$GT = H$$

T is the matrix in Definition 2.3.41.

Step 4) With the weights λ_j in T, calculate the statistic

$$Z_s^*(\vec{x}_0) = \sum_{j=1}^{k} \lambda_j \, Z(\vec{x}_j)$$

using all the samples whose locations are inside the neighborhood $(\vec{x}_0, r)$. $Z_s^*(\vec{x}_0)$ is the universal kriging estimation of the surface at location $\vec{x}_0$.

Step 5) Compute the product $\bar{\omega}T^{T}H$. The result is the estimation variance for the probable difference between the real and the estimated values of the surface at point $\vec{x}_0$.

END

The point estimate and the estimation variance together provide a way of defining an interval around $Z_s^*(\vec{x}_0)$ which can be assigned a specified probability of containing the true value $Z(\vec{x}_0)$. These two terms, $Z_s^*(\vec{x}_0)$ and $\bar{\omega}T^{T}H$, are the end results of kriging, as they provide interpolated points from which maps of both the estimated surface and the probable error may be constructed.

2.5 The Semivariogram of the Estimated Residuals

By Definition 2.3.2, a residual is the difference between the regionalized variable and the drift:

$$Y(\vec{x}) = Z(\vec{x}) - m(\vec{x})$$

In practice we cannot know the actual drift, but only the statistics $M(\vec{x})$ or $M'(\vec{x})$. Consequently, we never know the residuals, only the estimated residuals.

Definition 2.5.1: Let $Z(\vec{x})$ be a regionalized variable and let its estimated drift be $M(\vec{x})$. Then, by definition, the *estimated residual* $R(\vec{x})$ is:

$$R(\vec{x}) = Z(\vec{x}) - M(\vec{x})$$

as $Z(\vec{x})$ and $M(\vec{x})$ are random variables, $R(\vec{x})$ is a random variable.

Lemma 2.5.2: Let $R(\vec{x})$ be the estimated residual in Definition 2.5.1. Then,

$$E[R(\vec{x})] = 0$$

Proof: From Definition 2.5.1,

$$E[R(\vec{x})] = E[Z(\vec{x}) - M(\vec{x})]$$

Taking the expectation of each term,

$$E[R(\vec{x})] = E[Z(\vec{x})] - E[M(\vec{x})]$$

From Lemma 2.3.7, the right side of the equality is zero because both terms are equal.

Q.E.D.

In this section we will show the relationship existing between the semivariogram of the residuals and the semivariogram of the estimated residuals. First, we must demonstrate some basic lemmas.

Lemma 2.5.3: Let $Z(\vec{x})$ be a regionalized variable with drift $m(\vec{x})$ and residuals which verify the intrinsic hypothesis. Then, if $\gamma(\vec{h})$ is the semivariogram of the residuals,

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

Proof: By Definition 2.3.2,

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

Collecting the residuals and the drifts, expanding the square and taking the expected value of each term yields:

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

As the residuals satisfy the intrinsic hypothesis and have a zero mean, as shown by Lemma 2.3.3, then:

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

Because the drift is not a random variable, it can be taken out of the expectation. Then, the last term vanishes, by Lemma 2.3.3. Therefore,

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

Lemma 2.5.4: Let $M(\vec{x})$ be the statistic for the drift according to Definition 2.3.6. Let μ_s^i be the Lagrangian multiplier in Definition 2.3.26. Then, if both $\vec{x}$ and $\vec{x}+\vec{h}$ are inside the same neighborhood $(\vec{x},r)$,

$$E[\{M(\vec{x}+\vec{h}) - M(\vec{x})\}^2] = \sum_{i=1}^{n}\sum_{s=1}^{n} \mu_s^i [f^i(\vec{x}+\vec{h}) - f^i(\vec{x})][f^s(\vec{x}+\vec{h}) - f^s(\vec{x})] - [m(\vec{x}+\vec{h}) - m(\vec{x})]^2$$

Proof: Since $M(\vec{x})$ and $M'(\vec{x})$ are equal except for a constant, the difference $M(\vec{x}+\vec{h}) - M(\vec{x})$ is exactly equal to $M'(\vec{x}+\vec{h}) - M'(\vec{x})$. From Definition 2.3.22,

$$E[\{M(\vec{x}+\vec{h}) - M(\vec{x})\}^2] = E[\{\sum_{i=1}^{n} B_i f^i(\vec{x}+\vec{h}) - \sum_{s=1}^{n} B_s f^s(\vec{x})\}^2]$$

Expanding the square and taking the expected value of each term yields:

$$E[\{M(\vec{x}+\vec{h}) - M(\vec{x})\}^2] = E[\sum_{i=1}^{n} B_i f^i(\vec{x}+\vec{h}) \sum_{s=1}^{n} B_s f^s(\vec{x}+\vec{h})] - 2E[\sum_{i=1}^{n} B_i f^i(\vec{x}+\vec{h}) \sum_{s=1}^{n} B_s f^s(\vec{x})]$$
$$+ E[\sum_{i=1}^{n} B_i f^i(\vec{x}) \sum_{s=1}^{n} B_s f^s(\vec{x})]$$

All terms $f^i(\vec{x})$ can be taken out of the expected values. So, for example,

$$E[\sum_{i=1}^{n} B_i f^i(\vec{x}) \sum_{s=1}^{n} B_s f^s(\vec{x})] = \sum_{i=1}^{n}\sum_{s=1}^{n} f^i(\vec{x}) f^s(\vec{x}) E[B_i, B_s]$$

But, by Theorem 2.3.29,

$$E[\sum_{i=1}^{n} B_i f^i(\vec{x}) \sum_{s=1}^{n} B_s f^s(\vec{x})] = \sum_{i=1}^{n}\sum_{s=1}^{n} f^i(\vec{x}) f^s(\vec{x})(\mu_s^i + b_i b_s)$$

Performing the same operation on the other two terms in $E[\{M(\vec{x}+\vec{h}) + M(\vec{x})\}^2]$ produces:

$$E[\{M(\vec{x}+\vec{h}) - M(\vec{x})\}^2] = \sum_{i=1}^{n}\sum_{s=1}^{n} (\mu_s^i + b_i b_s)[f^i(\vec{x}+\vec{h}) f^s(\vec{x}+\vec{h}) - 2f^i(\vec{x}+\vec{h}) f^s(\vec{x}) + f^i(\vec{x}) f^s(\vec{x})]$$

Expressing the sum of the $f^i(\vec{x})$ as a product:

$$E[\{M(\vec{x}+\vec{h}) - M(\vec{x})\}^2] = \sum_{i=1}^{n}\sum_{s=1}^{n} (\mu_s^1 + b_i b_s)[f^i(\vec{x}+\vec{h}) - f^i(\vec{x})][f^s(\vec{x}+\vec{h}) - f^s(\vec{x})]$$

Separating $b_i b_s$ from μ_s^i,

$$E[\{M(\vec{x}+\vec{h}) - M(\vec{x})\}^2] = \sum_{i=1}^{n}\sum_{s=1}^{n} \mu_s^i[f^i(\vec{x}+\vec{h}) - f^i(\vec{x})][f^s(\vec{x}+\vec{h}) - f^s(\vec{x})]$$

$$+ \sum_{i=1}^{n} [b_i f^i(\vec{x}+\vec{h}) = b_i f^i(\vec{x})] \sum_{s=1}^{n} [b_s f^s(\vec{x}+\vec{h}) - b_s f^s(\vec{x})]$$

But

$$\sum_{i=1}^{n} b_i f^i(\vec{x}) = m(\vec{x}) - a_0$$

Therefore,

$$E[\{M(\vec{x}+\vec{h}) - M(\vec{x})\}^2] = \sum_{i=1}^{n}\sum_{s=1}^{n} \mu_s^i[f^i(\vec{x}+\vec{h}) - f^i(\vec{x})][f^s(\vec{x}+\vec{h}) - f^s(\vec{x})] + [m(\vec{x}+\vec{h}) - m(\vec{x})]^2$$

Q.E.D.

Lemma 2.5.5: Let the analytical expression for the drift $m(\vec{x})$ be given by Definition 2.3.4. Let the estimator A_s of the coefficient a_s be:

$$A_s = \sum_{j=1}^{k} \lambda_s^j Z(\vec{x}_j)$$

with $\sum_{j=1}^{k} \lambda_s^j = 0$.

Then, if $\vec{x}_0$ is constant and $|\vec{h}|$ is smaller than r in Definition 2.3.4,

$$[m(\vec{x}) - m(\vec{x}_0)][m(\vec{x}+\vec{h}) - m(\vec{x})] = \sum_{s=1}^{n} [f^s(\vec{x}+\vec{h}) - f^s(\vec{x})] \sum_{j=1}^{k} \lambda_s^j [m(\vec{x}) - m(\vec{x}_0)]\, m(\vec{x}_j)$$

Proof: From Definition 2.3.4,

$$[m(\vec{x}) - m(\vec{x}_0)][m(\vec{x}+\vec{h}) - m(\vec{x})] = [m(\vec{x}) - m(\vec{x}_0)] \sum_{s=1}^{n} [a_s\, f^s(\vec{x}+\vec{h}) - a_s\, f^s(\vec{x})]$$

However, by Theorem 2.3.9,

$$[m(\vec{x}) - m(\vec{x}_0)][m(\vec{x}+\vec{h}) - m(\vec{x})] = [m(\vec{x}) - m(\vec{x}_0)] \sum_{s=1}^{n} [f^s(\vec{x}+\vec{h}) - f^s(\vec{x})] \sum_{i=0}^{n} a_i \sum_{j=1}^{k} \lambda_s^j\, f^i(\vec{x}_j)$$

The term a_i can be introduced within the summation over j, the summation order can be reversed for i and j, and λ_s^j can be taken out of the summation over i:

$$[m(\vec{x}) - m(\vec{x}_0)][m(\vec{x}+\vec{h}) - m(\vec{x})] = [m(\vec{x}) - m(\vec{x}_0)] \sum_{s=1}^{n} [f^s(\vec{x}+\vec{h}) - f^s(\vec{x})] \sum_{j=1}^{k} \lambda_s^j \sum_{i=0}^{n} a_i\, f^i(\vec{x}_j)$$

But the summation over i is equal to $m(\vec{x}_j)$. Therefore, introducing $m(\vec{x}) - m(\vec{x}_0)$ into the summation over j:

$$[m(\vec{x}) - m(\vec{x}_0)][m(\vec{x}+\vec{h}) - m(\vec{x})] = \sum_{i=1}^{n} [f^s(\vec{x}+\vec{h}) - f^s(\vec{x})] \sum_{j=1}^{k} \lambda_s^j [m(\vec{x}) - m(\vec{x}_0)]\, m(\vec{x}_j)$$

Q.E.D.

Lemma 2.5.6: Let $M(\vec{x})$ be the estimator of the drift of the non-stationary regionalized variable $Z(\vec{x})$. Then,

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

Proof: Consider the identity

$$E[\{Z(\vec{x}+\vec{h}) - Z(\vec{x})\}\{M(\vec{x}+\vec{h}) - M(\vec{x})\}] = E[\{Z(\vec{x}+\vec{h}) - Z(\vec{x}_0) - Z(\vec{x}) + Z(\vec{x}_0)\}\{M(\vec{x}+\vec{h}) - M(\vec{x})\}]$$

where $\vec{x}_0$ is a constant. Expanding the expression on the right side of the equality and splitting the expectations:

$$E[\{Z(\vec{x}+\vec{h}) - Z(\vec{x})\}\{M(\vec{x}+\vec{h}) - M(\vec{x})\}] = E(\{Z(\vec{x}+\vec{h}) - Z(\vec{x}_0)\}\{M(\vec{x}+\vec{h}) - M(\vec{x})\}]$$

$$- E[\{Z(\vec{x}) - Z(\vec{x}_0)\}\{M(\vec{x}+\vec{h}) - M(\vec{x})\}]$$

Within the second expectation, as $M(\vec{x}+\vec{h}) - M(\vec{x})$ is equal to $M'(\vec{x}+\vec{h}) - M'(\vec{x})$, thus

$$E[\{Z(\vec{x}) - Z(\vec{x}_0)\}\{M(\vec{x}+\vec{h}) - M(\vec{x})\}] = E[\{Z(\vec{x}) - Z(\vec{x}_0)\}\{\sum_{s=1}^{n} B_s f^s(\vec{x}+\vec{h}) - \sum_{s=1}^{n} B_s f^s(\vec{x})\}]$$

Introducing $Z(\vec{x}) - Z(\vec{x}_0)$ within the summations, reversing the order of summation and expectation, removing $f^s(\vec{x})$ from the expectation, and replacing B_s by its Definition 2.3.21 produces the equivalent expression:

$$E[\{Z(\vec{x}) - Z(\vec{x}_0)\}\{M(\vec{x}+\vec{h}) - M(\vec{x})\}] = \sum_{s=1}^{n} [f^s(\vec{x}+\vec{h}) - f^s(\vec{x})]\, E[\{Z(\vec{x}) - Z(\vec{x}_0)\} \sum_{j=1}^{k} \lambda_s^j Z(\vec{x}_j)]$$

However, because the weights λ_s^j are optimal,

$$\sum_{j=1}^{k} \lambda_s^j = 0$$

Hence,

$$\sum_{j=1}^{k} \lambda_s^j Z(\vec{x}_j) = \sum_{j=1}^{k} \lambda_s^j Z(\vec{x}_j) - Z(\vec{x}_0) \sum_{j=1}^{k} \lambda_s^j = \sum_{j=1}^{k} \lambda_s^j [Z(\vec{x}_j) - Z(\vec{x}_0)]$$

Therefore,

$$E[\{Z(\vec{x}) - Z(\vec{x}_0)\}\{M(\vec{x}+\vec{h}) - M(\vec{x})\}] = \sum_{s=1}^{n} [f^s(\vec{x}+\vec{h}) - f^s(\vec{x})] \sum_{j=1}^{k} \lambda_s^j$$

$$E[\{Z(\vec{x}) - Z(\vec{x}_0)\}\{Z(\vec{x}_j) - Z(\vec{x}_0)\}]$$

But from Lemma 2.3.11 and Theorem 2.2.5,

$$E[\{Z(\vec{x}) - Z(\vec{x}_0)\}\{M(\vec{x}+\vec{h}) - M(\vec{x})\}] = \sum_{s=1}^{n} [f^s(\vec{x}+\vec{h}) - f^s(\vec{x})] \sum_{j=1}^{k} \lambda_s^j [\gamma(\vec{x}-\vec{x}_0)$$

$$+ \gamma(\vec{x}_j-\vec{x}_0) - \gamma(\vec{x}-\vec{x}_j) + \{m(\vec{x}) - m(\vec{x}_0)\}\{m(\vec{x}_j) - m(\vec{x}_0)\}]$$

We will consider each term separately. Because $\sum_{j=1}^{k} \lambda_s^j = 0$ and $\gamma(\vec{x}-\vec{x}_0)$ is independent of j,

$$\sum_{j=1}^{k} \lambda_s^j \gamma(\vec{x}-\vec{x}_0) = \gamma(\vec{x}-\vec{x}_0) \sum_{j=1}^{k} \lambda_s^j = 0$$

From Lagrange's method of multipliers leading to Algorithm 2.3.28,

$$\sum_{j=1}^{k} \lambda_s^j \gamma(\vec{x}_j-\vec{x}_0) = -\sum_{i=1}^{n} \mu_s^i f^i(\vec{x}_0) - \mu_0$$

and also

$$\sum_{j=1}^{k} \lambda_s^j \gamma(\vec{x}-\vec{x}_j) = -\sum_{i=1}^{n} \mu_s^i f^i(\vec{x}) - \mu_0$$

Because $\sum_{j=1}^{k} \lambda_s^j = 0$ and $[m(\vec{x}) - m(\vec{x}_0)]\, m(\vec{x}_0)$ is independent of j,

$$\sum_{j=1}^{k} \lambda_s^j [m(\vec{x}) - m(\vec{x}_0)]\, m(\vec{x}_0) = [m(\vec{x}) - m(\vec{x}_0)]\, m(\vec{x}_0) \sum_{j=1}^{k} \lambda_s^j = 0$$

Finally, by Lemma 2.5.5,

$$\sum_{s=1}^{n} [f^s(\vec{x}+\vec{h}) - f^s(\vec{x})] \sum_{j=1}^{k} \lambda_s^j [m(\vec{x}) - m(\vec{x}_0)]\, m(\vec{x}_j) = [m(\vec{x}) - m(\vec{x}_0)][m(\vec{x}+\vec{h}) - m(\vec{x})]$$

Therefore,

$$E[\{Z(\vec{x}) - Z(\vec{x}_0)\}\{M(\vec{x}+\vec{h}) - M(\vec{x})\}] = \sum_{s=1}^{n} [f^s(\vec{x}+\vec{h}) - f^s(\vec{x})][-\sum_{i=1}^{n} \mu_s^i f^i(\vec{x}_0) + \sum_{i=1}^{n} \mu_s^i f^i(\vec{x})]$$

$$+ [m(\vec{x}) - m(\vec{x}_0)][m(\vec{x}+\vec{h}) - m(\vec{x})]$$

Analogously, we can demonstrate that

$$E[\{Z(\vec{x}+\vec{h}) - Z(\vec{x}_0)\}\{M(\vec{x}+\vec{h}) - M(\vec{x})\}] = \sum_{s=1}^{n} [f^s(\vec{x}+\vec{h}) - f^s(\vec{x})][-\sum_{i=1}^{n} \mu_s^i f^i(\vec{x}_0) + \sum_{i=1}^{n} \mu_s^i f^i(\vec{x}+\vec{h})$$

$$+ [m(\vec{x}+\vec{h}) - m(\vec{x}_0)][m(\vec{x}+\vec{h}) - m(\vec{x})]$$

Therefore,

$$E[\{Z(\vec{x}+\vec{h}) - Z(\vec{x})\}\{M(\vec{x}+\vec{h}) - M(\vec{x})\}] = \sum_{s=1}^{n} [f^s(\vec{x}+\vec{h}) - f^s(\vec{x})] \sum_{i=1}^{n} \mu_s^i [f^i(\vec{x}+\vec{h}) - f^i(\vec{x})]$$

$$+ [m(\vec{x}+\vec{h}) - m(\vec{x})]^2$$

As $f^s(\vec{x}+\vec{h}) - f^s(\vec{x})$ is independent of i, we can introduce this expression under the second summation sign and reverse summation order. By Theorem 2.5.4,

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

Q.E.D.

The following theorem is interesting because it provides a useful relationship between the semivariogram of the residuals and the semivariogram of the estimated residuals.

Theorem 2.5.7: Let $\gamma_E(\vec{h})$ be the semivariogram for the estimated residuals $R(\vec{x})$ in Definition 2.5.1 and let $\gamma(\vec{h})$ be the semivariogram of the residuals $Y(\vec{x})$ in Definition 2.3.2. Let μ_s^i be

a Lagrangian multiplier resulting from the system of equations in Algorithm 2.3.28. Then,

$$\gamma_E(\vec{h}) = \gamma(\vec{h}) - \sum_{i=1}^{n}\sum_{s=1}^{n} \mu_s^i [f^i(\vec{x}+\vec{h}) - f^i(\vec{x})][f^s(\vec{x}+\vec{h}) - f^s(\vec{x})]$$

Proof: From the definition for the semivariogram and Lemma 2.5.2,

$$\gamma_E(\vec{h}) = 1/2\ E[\{R(\vec{x}+\vec{h}) - R(\vec{x})\}^2]$$

and by Definition 2.5.1,

$$\gamma_E(\vec{h}) = 1/2\ E[\{Z(\vec{x}+\vec{h}) - M(\vec{x}+\vec{h}) - Z(\vec{x}) + M(\vec{x})\}^2]$$

Be expansion, and taking the expectation of each term,

$$\gamma_E(\vec{h}) = 1/2\ E[\{Z(\vec{x}+\vec{h}) - Z(\vec{x})\}^2] + 1/2\ E[\{M(\vec{x}+\vec{h}) - M(\vec{x})\}^2] - E[\{Z(\vec{x}+\vec{h}) - Z(\vec{x})\}\{M(\vec{x}+\vec{h}) - M(\vec{x})\}]$$

From Lemmas 2.5.3, 2.5.4, and 2.5.6,

$$\gamma_E(\vec{h}) = \gamma(\vec{h}) - 1/2 \sum_{i=1}^{n}\sum_{s=1}^{n} \mu_s^i [f^i(\vec{x}+\vec{h}) - f^i(\vec{x})][f^s(\vec{x}+\vec{h}) - f^s(\vec{x})]$$

The practical implications of this theorem for statistical induction are contained in the following theorem.

Theorem 2.5.8: Let $R(\vec{x}_1)$, $R(\vec{x}_2)$, ..., $R(\vec{x}_t)$ be estimated residuals taken at regular intervals in a neighborhood where the analytical expression for the drift does not change. Then, if the estimated residuals satisfy the intrinsic hypothesis and the residuals have a common semivariogram, the following is an unbiased statistic for $\gamma_E(\vec{h})$:

$$\overline{\gamma}_E(p\vec{a}) = \frac{1}{2(k-p)} \sum_{j=k'}^{k+k'-p-1} [R(\vec{x}_j + p\vec{a}) - R(\vec{x}_j)]^2$$

for $k + k' \leqslant t$ and $p = 0, 1, \ldots, k-1$.

Proof: If $\overline{\gamma}_E(\vec{h})$ is an unbiased statistic for $\gamma_E(\vec{h})$,

$$E[\overline{\gamma}_E(\vec{h})] = \gamma_E(\vec{h})$$

From the definition of $\overline{\gamma}_E(\vec{h})$,

$$E[\overline{\gamma}(\vec{h})] = \frac{1}{2(k-p)} E[\sum_{j=k'}^{k+k'-p-1} \{R(\vec{x}_j+p\vec{a}) - R(\vec{x}_j)\}^2]$$

Interchanging summation and expectation:

$$E[\overline{\gamma}_E(\vec{h})] = \frac{1}{2(k-p)} \sum_{j=k'}^{k+k'-p-1} E[\{R(\vec{x}_j+p\vec{a}) - R(\vec{x}_j)\}^2]$$

But the expectation is twice the semivariogram of the estimated residuals. Hence,

$$E[\bar{\gamma}_E(\vec{h})] = \frac{1}{2(k-p)} \sum_{j=k'}^{k+k'-p-1} 2\gamma_E(\vec{h}) = \gamma_E(\vec{h})$$

Q.E.D.

Theorem 2.5.9: Let $\gamma_E(\vec{h})$ be the semivariogram of the estimated residuals at t locations at regular intervals $\vec{a}$. Then, if $\vec{h}$ is p times the distance $\vec{a}$ between consecutive samples:

$$\gamma_E(\vec{h}) = \gamma(\vec{h}) - \frac{1}{2(k-p)} \sum_{j=k'}^{k+k'-p-1} \sum_{i=1}^{n} \sum_{s=1}^{n} \mu_s^i [f^i(\vec{x}_j+p\vec{a}) - f^i(\vec{x}_j)][f^s(\vec{x}_j+p\vec{a}) - f^s(\vec{x}_j)]$$

for $k + k' \leqslant t$ and $p = 0, 1, \ldots, k-1$.

Proof: Let us consider the identity

$$\gamma_E(\vec{h}) = \frac{1}{(k-p)} \sum_{j=k'}^{k+k'-p-1} \gamma_E(\vec{h})$$

From Theorem 2.5.7,

$$\gamma_E(\vec{h}) = \gamma(\vec{h}) - \frac{1}{2(k-p)} \sum_{j=k'}^{k+k'-p-1} \sum_{i=1}^{n} \sum_{s=1}^{n} \mu_s^i [f^i(\vec{x}_j+\vec{h}) - f^i(\vec{x}_j)][f^s(\vec{x}_j+\vec{h}) - f^s(\vec{x}_j)]$$

Q.E.D.

From the result in Theorem 2.5.9 the semivariogram for the estimated residuals is a biased estimate for the semivariogram of the actual residuals. The bias is represented by those terms in the triple summation. The only exception comes when the drift is constant. In that case the triple summation is zero as all the $f^i(\vec{x})$ are zero. These are considerations we must remember in the next chapter.

CHAPTER 3 -- APPLICATIONS TO MAPPING AND MAP ANALYSIS

In this chapter the results obtained in Chapter 2 will be used in a practical way to emphasize the advantages of regionalized variable theory for contour mapping. In particular, the results contained in Algorithm 2.3.28 will be used to estimate semivariograms; Algorithm 2.3.45 will be used to map the drift, and Algorithm 2.4.23 to perform universal kriging. Although these may be implemented without reference to the preceding mathematical derivations, Chapter 2 contains the foundation material on which these are based. That chapter may also be useful as a reference to develop other algorithms to cover different needs.

The selected example is related to the search for oil by the mapping of subsurface geologic structures. However, those interested in other applications will find this particular example helpful. Although specific values and the nature of the problem will necessarily be different, the steps to be taken remain essentially the same.

3.1 Geologic Example

The Magellan Basin is at the southern end of South America (Fig. 3.1.1); its boundaries are roughly the Andes Mountains on the south and west, parallel 47°S. on the north, and the Atlantic Ocean on the east. Prior to formation of the basin, there was intensive volcanic activity which covered the area where the basin subsequently developed with a thick bed of tuffs, breccias, and ignimbrites known as the Tobifera Series. Marine transgression began at the end of the Jurassic. The Tobifera Series is overlain by the Springhill Group, a sequence of alternating sandstones and shales, with an average thickness of 50 feet. The Springhill Group is continental in origin at its base, representing an intermediate stage between the volcanic and marine deposits. Marine sedimentation ended during the Miocene. All sediments younger than the Springhill Group are impermeable except for sporadic sandstone lenses.

Most petroleum traps in the basin are essentially structural but combined stratigraphic-structural traps occur. However, the stratigraphic component is always secondary, composed of facies changes which define several productive levels or zones around a productive structure. Magellan Basin reservoirs are related to areas of positive relief in the underlying volcanics. Salt domes are unknown. Most faults seem to be younger than the hydrocarbon migration, as there are normal faults but no reservoirs have been found on the lower blocks.

In some areas of the world, all that remains to be done in the search for petroleum reserves after a positive structural anomaly is located is to drill its top. In the Magellan Basin, the only producing formation may be absent from the top of structures. In some places, positive structure is present but the sandstone is absent and vice versa. In intermediate

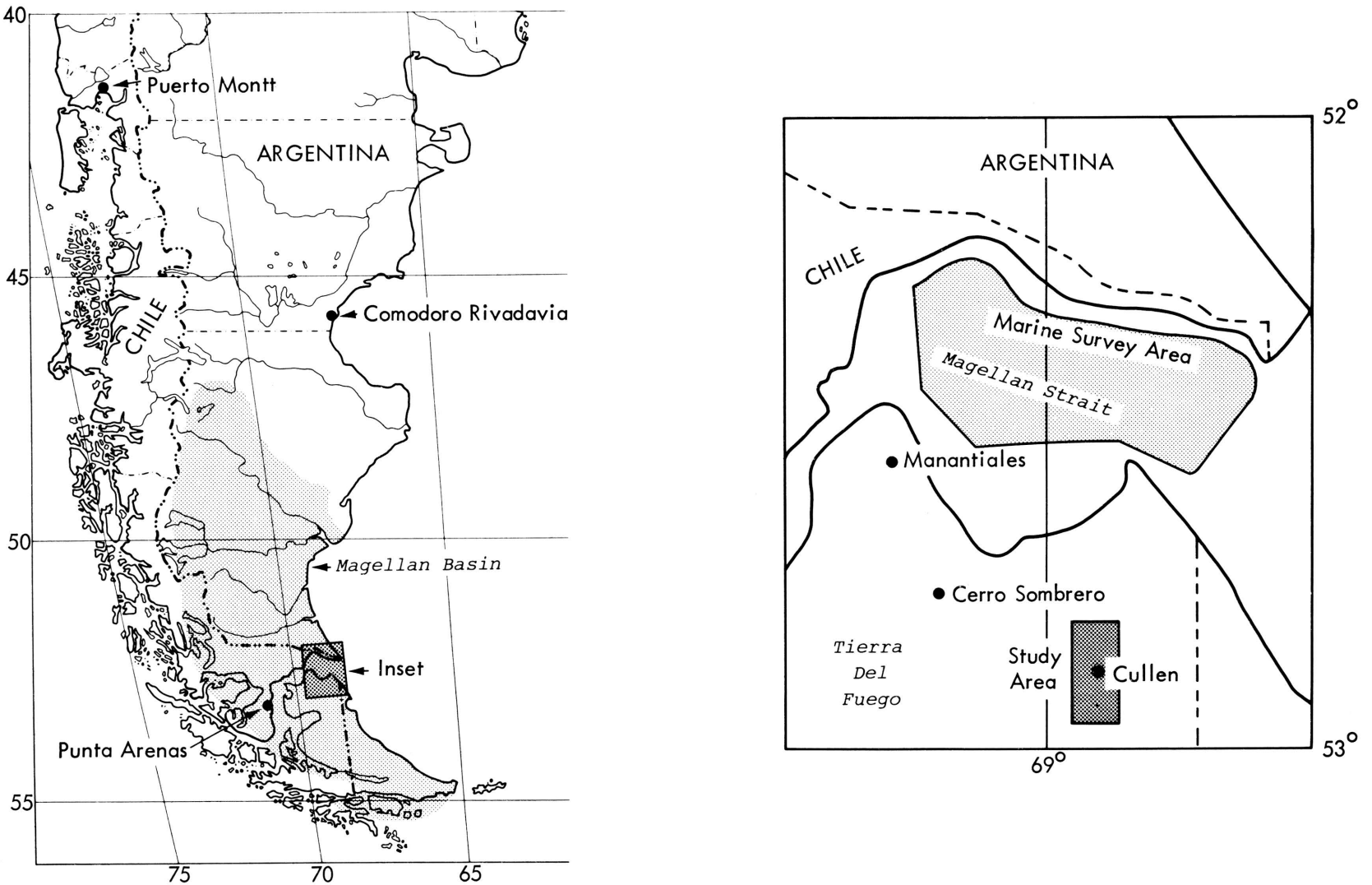


Figure 3.1.1--Index map of study area. Shaded portion indicates Magellan Basin. Inset shows location of marine seismic survey area in Magellan Strait and mapped area around Cullen field.

situations such as the Cullen field, annular reservoirs surround a barren core. This results from the fact that sands were removed or possibly not deposited on the top of the old topographic peaks until the basin deepened so much that sedimentation of shales and marls occurred over their crests.

Cullen is a mature field in the Magellan Basin with a cumulative production of 35 million barrels of oil in 20 years. Data used in this example come from two sources. The main body of information is taken from the 163 wells drilled in the area of the Cullen field. As the well density in the southwest is poor, 63 additional values were taken from seismic reflection profiles to improve control in that area.

3.2 Structural Analysis

Knowledge of the semivariogram for the residuals of the regionalized variable is required for any computation of drift or universal kriging. We have already defined this function and given some expressions for its estimation, but have not presented any specific semivariograms. This intrinsic function cannot be chosen arbitrarily; it is determined by the nature of spatial continuity of the regionalized variable.

In the matrices in the algorithms we also have terms $f^i(\vec{x})$ coming from the analytical expression for the drift. We have not yet found a specific drift for the structural unit to be mapped, which is the top of the Tobifera Series in the Cullen field. We have seen that there is a close relationship between the type of semivariogram and the drift, and therefore some form of conditional analysis is needed to find the most convenient semivariogram given a certain arbitrary drift. This is performed by what is called a *structural analysis*, and is an essential prerequisite for the calculation of any terms in the matrices.

Data Selection for Structural Analysis: The very first step in a structural analysis is to search for data. We need samples at regular intervals along a line in order to calculate a semivariogram. The selection of samples at regular intervals from randomly distributed data is always difficult and requires different approaches according to the particular circumstances. Within certain tolerances, samples can be selected which are not exactly at regular intervals along a line. When that cannot be done some kind of interpolation is necessary, although this is not recommended and is to be avoided if possible. Interpolation methods fail to predict unexpected changes in the regionalized variable in the absence of good control. Therefore, the sequence of interpolated values will vary more gently than the true sequence and the semivariogram will indicate a false higher continuity for the regionalized variable. This means the estimation standard deviation will yield unrealistically narrow confidence intervals for the estimates. On the other hand, a certain amount of statistical redundance is desirable to obtain reliable and stable semivariograms. This means that a few short lines are not sufficient for a good structural analysis. Finally, the regionalized variable must be tested for isotropy by studying the semivariograms along different directions, which again requires a sizeable data set.

Although the data to be mapped may not fit all requirements for a good structural analysis, sometimes it is part of a larger body of information or there is an appropriate data set for the same regionalized variable in the nearby vicinity. In such instances it is much better to use the best data set and to extrapolate results of the structural analysis rather than to extrapolate values in the data set to be mapped. This was done for this study, as the well data from the Cullen field consists of 163 irregularly distributed values. Forty kilometers to the north there are more than 2000 geophysical measurements from a marine survey, taken at regular intervals along lines. The only problem with the marine data is that the measurements are oriented only along two main directions. However, it seems better to obtain an estimate of the semivariogram from this source rather than from the well data itself. In either case, the measurements are estimates of the depth below sea level of a structural horizon, although in the case of the marine survey, this is inferred from seismic return times.

Strategy for an Optimal Semivariogram and Drift Search: Our problem is to look for: a) The optimal analytical expression and parameters for the intrinsic function, in particular, the range. The optimality is partially decided by: b) The analytical expression for the drift; that is, the nature of the functions $f^i(\vec{x})$ in Definition 2.3.20, as well as their number n, and c) The isotropy of the semivariogram of the estimated residuals.

In the previous chapter we have seen that the determination of the intrinsic function and the drift are not independent. More precisely, one must be assumed in order to determine the other. This requires that we check the initial guess against the final result and repeat the procedure if the fit between the assumption and the implication is not satisfactory. This problem is similar to searching for the roots of a polynomial equation of high degree. Because there is no formula whereby a root can be calculated, we guess at a solution, say x_1, and assume it is a root. Then, if it actually is a root, it should be: $P(x_1) = 0$.

We then evaluate the polynomial to see if the expression is really equal to zero. If it is, x_1 is a root of the expression. If it is not, we try a second guess, x_2, and so on, until we obtain a result that, although it may not produce an expression exactly equal to zero, gives a reasonably small residual. For example, if one-decimal precision is all that is required, the residual $P(x_1) = 0.001$ and the root $x_1 = 1.4$ could be sufficient, even though the real root is $\sqrt{2}$. In a similar manner, we may search for a semivariogram.

The following algorithm, used iteratively, provides a method of finding optimal semivariograms.

Algorithm 3.2.1: This is a procedure to calculate the semivariogram for estimated residuals at regular intervals along a line.

Step 1) Define an arbitrary analytical expression for the semivariogram of the actual residuals.

Step 2) Select the analytical expression for the drift in Definition 2.3.20.

Step 3) Set k' = 1 in Theorem 2.2.6. Select a first neighborhood with x_1 as the left-most point. The length of the interval (k-p)a may be at most equal to the distance between the extreme points in the data sequence.

Step 4) Translate the origin so that the coordinate of the left-most sample is zero. Using Algorithm 2.3.28, calculate the drift coefficients.

Step 5) Calculate the estimated residuals using Definition 2.5.1.

Step 6) Use Theorem 2.2.6 to calculate the semivariogram of the estimated residuals for p = 0, 1, ..., k-1. Increase k' by one in the expression of this theorem.

Step 7) Slide the interval by p'a, p' being an arbitrary but normally small integer. If the interval is still contained within the sample sequence, go to Step 4) and calculate a new semivariogram for the new interval.

Step 8) Average all partial semivariograms for the sequence. The result is the average semivariogram for the sequence.

Step 9) If there is another sequence in the same direction, go to Step 3).

Step 10) Average all the average semivariograms. The result is the average semivariogram for the direction.

END

Obtaining coincidence between the assumed and the resulting semivariogram in Algorithm 3.2.1 is more an art than a science. Starting with neighborhoods of four or five points and simple drifts seems the safest way.

The result of a search for the drift and the intrinsic function is not unique. Just as in the problem of finding roots, there is usually more than one distinct solution. There are always several combinations of drift and semivariogram expressions which may be equally satisfactory. Therefore, after completion of Algorithm 3.2.1, experimentation should continue until a collection of solutions is obtained. The final solution or solutions are selected from among all acceptable possibilities on the basis of convenience.

If attention is restricted to a small neighborhood $(\vec{x}_0, r)$ around $\vec{x}_0$ (that is, if we never consider samples whose distance to the point to be kriged is larger than r), $\gamma(\vec{h})$ will never be needed for arguments $|\vec{h}| > 2r$. Then $\gamma(\vec{h})$ can be defined for $|\vec{h}| < 2r$ only and $\gamma(\vec{h})$ beyond that distance can be ignored. In this circumstance, a linear semivariogram of the form:

$$\gamma(\vec{h}) = \bar{\omega}\,\vec{h}, \qquad |\vec{h}| \leqslant 2r$$

is a good approximation for 2r not larger than the range.

With this choice, we can use Algorithm 2.3.45 and 2.4.25 in place of Algorithms 2.3.39 and 2.4.17 for drift and universal kriging respectively.

Because the analysis is made along lines, we do not need more than one coordinate. Therefore, we will work with a semivariogram of the form:

$$\gamma(h) = \bar{\omega}\,h, \qquad h < 2r$$

Because of the simplifications in the calculations, this is the most popular expression for the semivariogram in practical applications. In this example, we will not consider other possible expressions for the semivariogram. To determine the semivariogram in this form, all that is necessary is to determine the slope $\bar{\omega}$ and the boundary r.

Because the drift should represent only the main features of the regionalized variable and not the details, a simple analytical expression is usually enough. Alternatives for the drift might be polynomials no higher than second degree. Again, as our analysis is in one dimension, we need one coordinate. We need those drift estimates which are valid for the entire neighborhood; therefore M(x) will be

$$M(x) = a_0$$

$$M(x) = a_0 + b_1 x$$

or

$$M(x) = a_0 + b_1 x + b_2 x^2$$

So, for a given neighborhood, using Algorithm 3.2.1, we must calculate the statistics B_s in order to obtain estimated residuals and then compare the experimental semivariogram of the estimated residuals with the theoretical result given by Theorem 2.5.9. The constant a_0 is not necessary as it vanishes from the calculations.

From Appendix B, for a linear semivariogram and linear drift, the statistic B_1 for b_1 is:

$$B_1^1 = \frac{Z(x_k) - Z(x_1)}{(k-1)a}$$

Here, a is the distance between two consecutive samples along a line in an interval of length r whose extremes are x_k and x_1.

From Appendix C, for a linear semivariogram and quadratic drift, the expressions for the necessary statistics are:

$$B_2^2 = \frac{-6}{(k-2)(k-1)a^2} [Z_m - 1/2 \{Z(x_k) + Z(x_1)\}]$$

$$B_1^2 = \frac{1}{(k-1)a} [Z(x_k) - Z(x_1)] - (k-1)a\, B_2^2$$

where

$$Z_m = \frac{1}{k} \sum_{j=1}^{k} Z(x_k)$$

and x_k and x_1 are the ends of the interval where the drift is valid. Therefore, the experimentally estimated residuals when the drift is a constant are:

$$R^0(x_j) = Z(x_j) - a_0$$

The experimentally estimated residuals for a linear drift are:

$$R^1(x_j) = Z(x_j) - a_0 - B_1^1 x_j$$

and the experimentally estimated residuals for a quadratic drift are given by:

$$R^2(x_j) = Z(x_j) - a_0 - B_1^2 x_j - B_2^2 x_j^2$$

The *experimental semivariogram* for the estimated residual comes from the following relation, according to Theorem 2.5.8:

$$\gamma_E^i(pa) = \frac{1}{2(k-p)} \sum_{j=k'}^{k+k'-p-1} [R^i(x_j+pa) - R^i(x_j)]^2$$

for k+k' less than or equal to the total number of samples and p = 0, 1, ..., k-1.

But from Theorem 2.5.9 we know that the semivariogram for the estimated residual will be a biased estimator of the semivariogram for the true residuals if there is a drift. Appendices D and E provide the amount of bias for the case of first- and second-degree polynomial drifts. However, to calculate the bias, the actual semivariogram for the residuals must be known. Under the assumption that the semivariogram for the real residuals is a straight line of slope ω, from Appendix D, such a semivariogram $\gamma(h)$ is given by:

$$\gamma(h) = \gamma_E^1(h) + \frac{\omega h^2}{L}$$

for the case of a linear drift. For a quadratic drift and a linear semivariogram, from Appendix E,

$$\gamma(h) = \gamma_E^2(h) + \frac{\omega h^2}{L(L^2-a^2)} [2L^2 + 2aL - a^2 - 2(L+a)h + h^2]$$

The goodness of fit depends on what slope is assumed for the semivariogram for the actual residuals. In practice there are several ways to select the slope. We can take the slope at the origin from the expression for $\gamma_E^i(h)$. Another possible solution is to fit a regression line trying to minimize the difference between both types of semivariogram taking into account that values for small arguments are statistically more reliable than those for large arguments as they are calculated using smaller numbers of pairs. If the experimental semivariogram is indeed a straight line both methods yield very similar results.

The comparison between the semivariogram of the estimated residuals and the semivariogram for the actual residuals is usually shown in the literature as parabolic curves, being biased forms of the semivariograms. Such a presentation tends to confuse readers who are more familiar with the unbiased, monotonically increasing semivariograms such as those in Figure 2.2.4. Using the formulas above it is fairly straightforward either to remove the bias in the experimental semivariograms or to bias the assumed semivariograms. The results of our structural analysis are in terms of unbiased semivariograms.

The last step in the selection of the semivariogram and the drift is to determine how well the semivariogram obtained with the estimated residuals fits the line $\bar{\omega}\, h$. If they are alike, all parameters and assumptions are assumed correct and the particular drift and semivariogram chosen is one possible solution in the structural analysis. Otherwise, at least one parameter or the assumption that the semivariogram can be represented by a straight line through the origin is incorrect. The model must be rejected in such an instance and other alternatives must be

sought. Structural analysis is considerably simplified by the use of computers. Semivariograms for the top of the Tobifera Series were obtained using SEMIVAR, a FORTRAN IV program for automatic structure analysis (Olea, 1975). The length r of the interval, the degree of the drift, and the sequences of samples are entered into the program, producing the average semivariograms for each individual sequence and for parallel sequences.

Semivariograms for the Top of the Tobifera Series: Figures 3.2.1 to 3.2.7 contain the experimental semivariograms for the estimated residuals with the bias removed, and the assumed linear semivariograms. These are given for the various directions and parameters tested, following the strategy outlined above and using as data the elevations of the top of the tuff in an area in

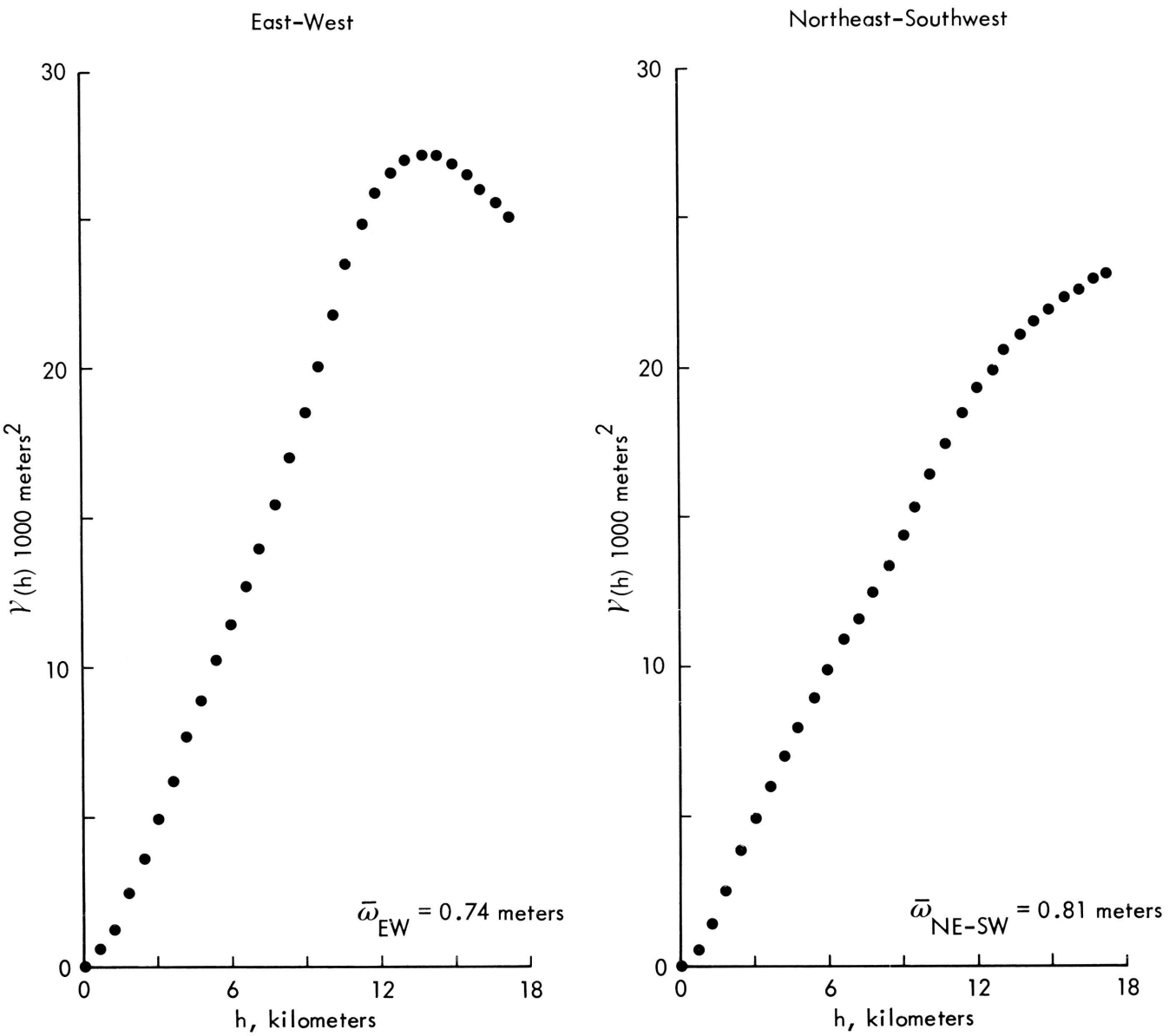


Figure 3.2.1--Experimental semivariograms for the top of the Tobifera Series, assuming no drift. The neighborhood is 17.4 km and the average slope at the origin, ω_A, is 0.77 meters.

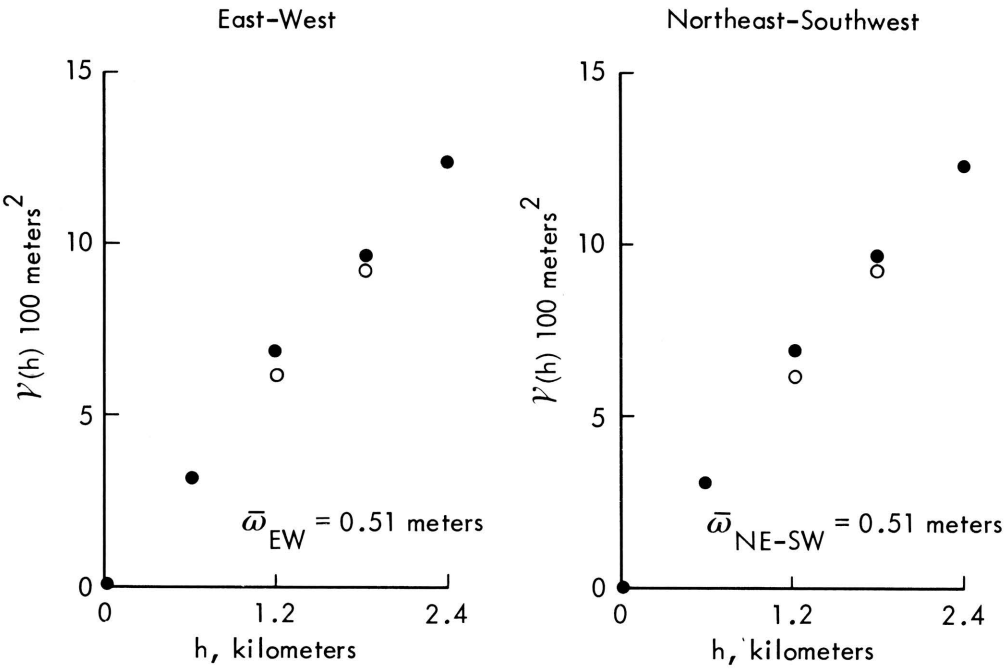


Figure 3.2.2--Experimental semivariograms for the top of the Tobifera Series, assuming a linear drift. The neighborhood is 2.4 km, and the average slope at the origin, ω_A, is 0.51 meters. Open circles are the assumed linear semivariogram, dots are the experimental semivariogram.

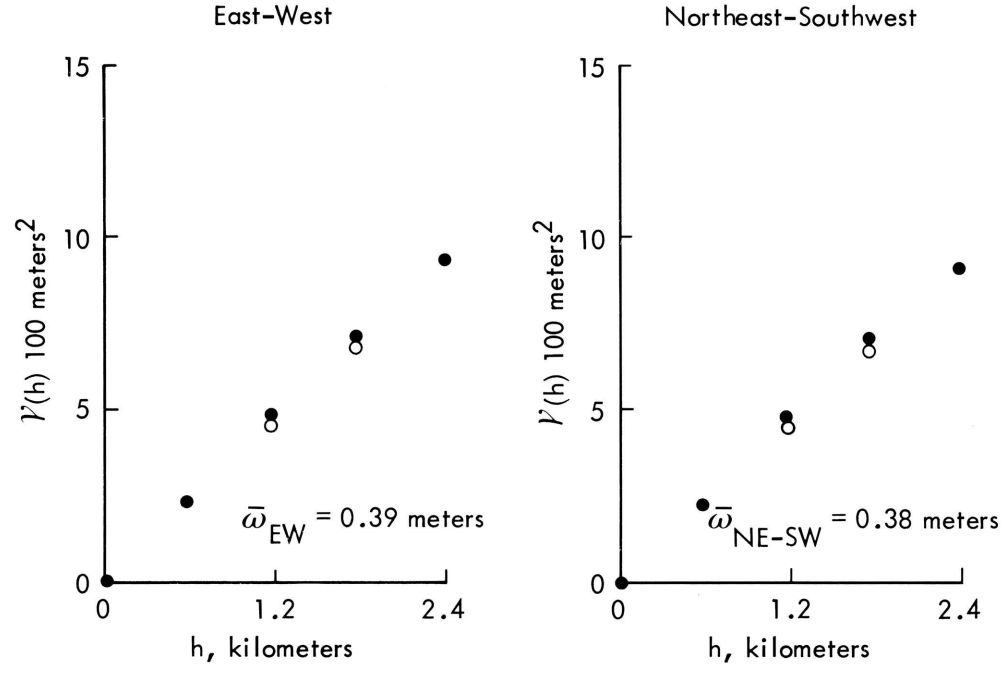


Figure 3.2.3--Experimental semivariograms for the top of the Tobifera Series, assuming a quadratic drift. The neighborhood is 2.4 km, and the average slope at the origin, ω_A is 0.39 meters. Open circles are the assumed linear semivariogram, dots are the experimental semivariogram.

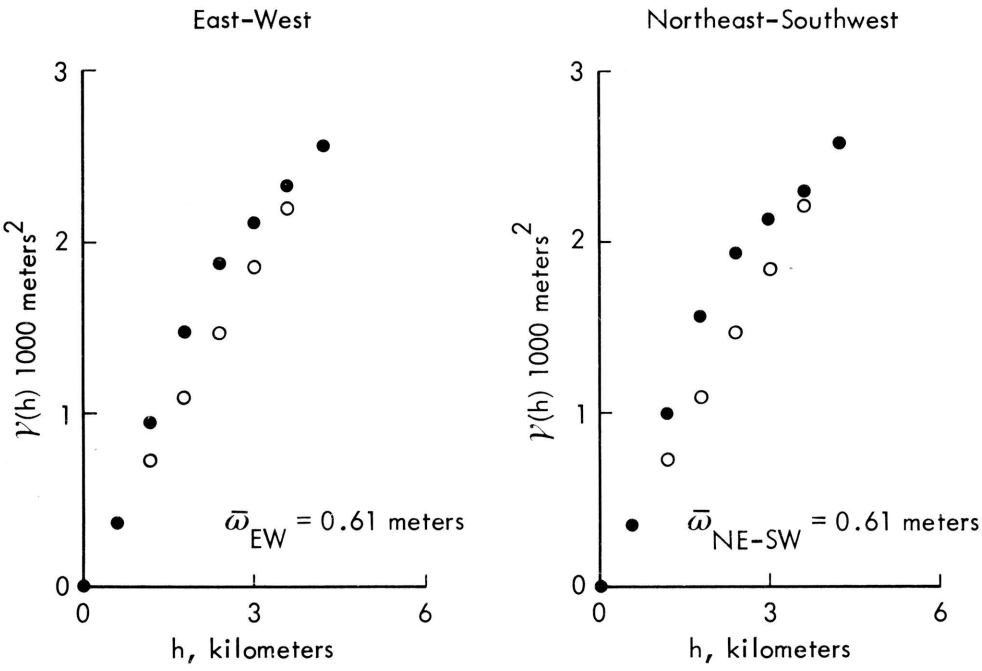


Figure 3.2.4--Experimental semivariograms for the top of the Tobifera Series, assuming a linear drift. The neighborhood is 4.2 km, and the average slope at the origin, ω_A, is 0.61 meters. Open circles are the assumed linear semivariogram, dots are the experimental semivariogram.

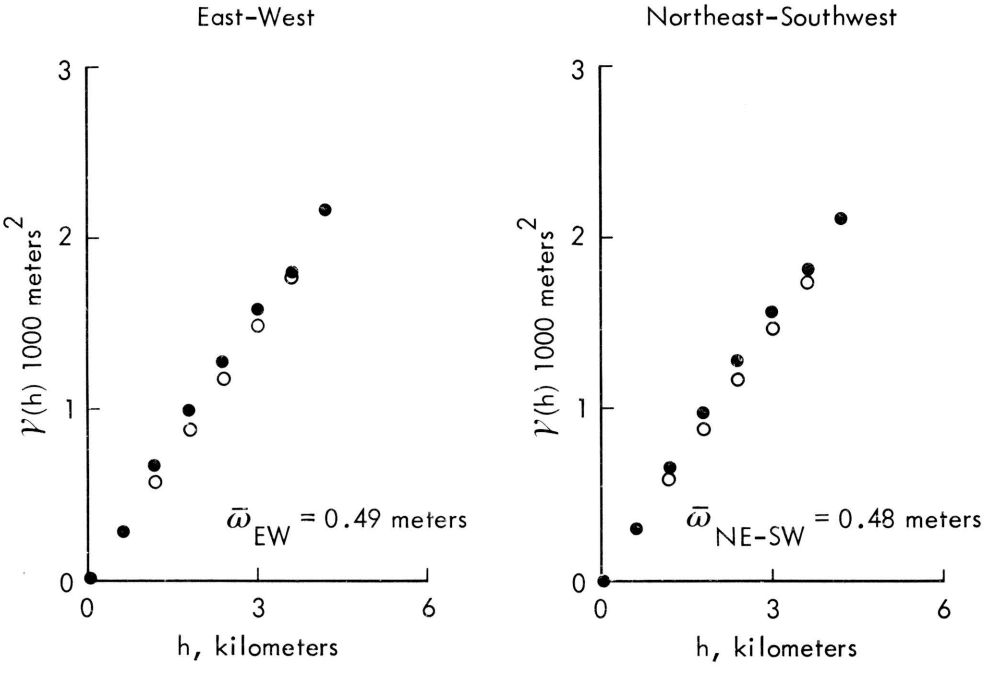


Figure 3.2.5--Experimental semivariograms for the top of the Tobifera Series, assuming a quadratic drift. The neighborhood is 4.2 km, and the average slope at the origin, ω_A, is 0.49 meters. Open circles are the assumed linear semivariogram, dots are the experimental semivariogram.

the eastern Magellan Strait. Figure 3.2.1 shows semivariograms resulting under the assumption that the drift is a constant. The semivariograms are free from any nugget effect with gentle parabolic shape at the origin implying that the top of the tuff has a moderate continuity. This assures us that there is in effect a varying drift underlying the regionalized variable.

The semivariograms indicate that the top of the tuff is isotropic. However, analysis of only two directions is not sufficient to assure isotropy in all remaining directions. If anisotropy is present, the semivariogram is no longer a function of only the distance between two points, but also is a function of the orientation of the line through these points:

$$\gamma(\rho,\theta) = \bar{\omega}(\theta)\ \gamma_0(\rho)$$

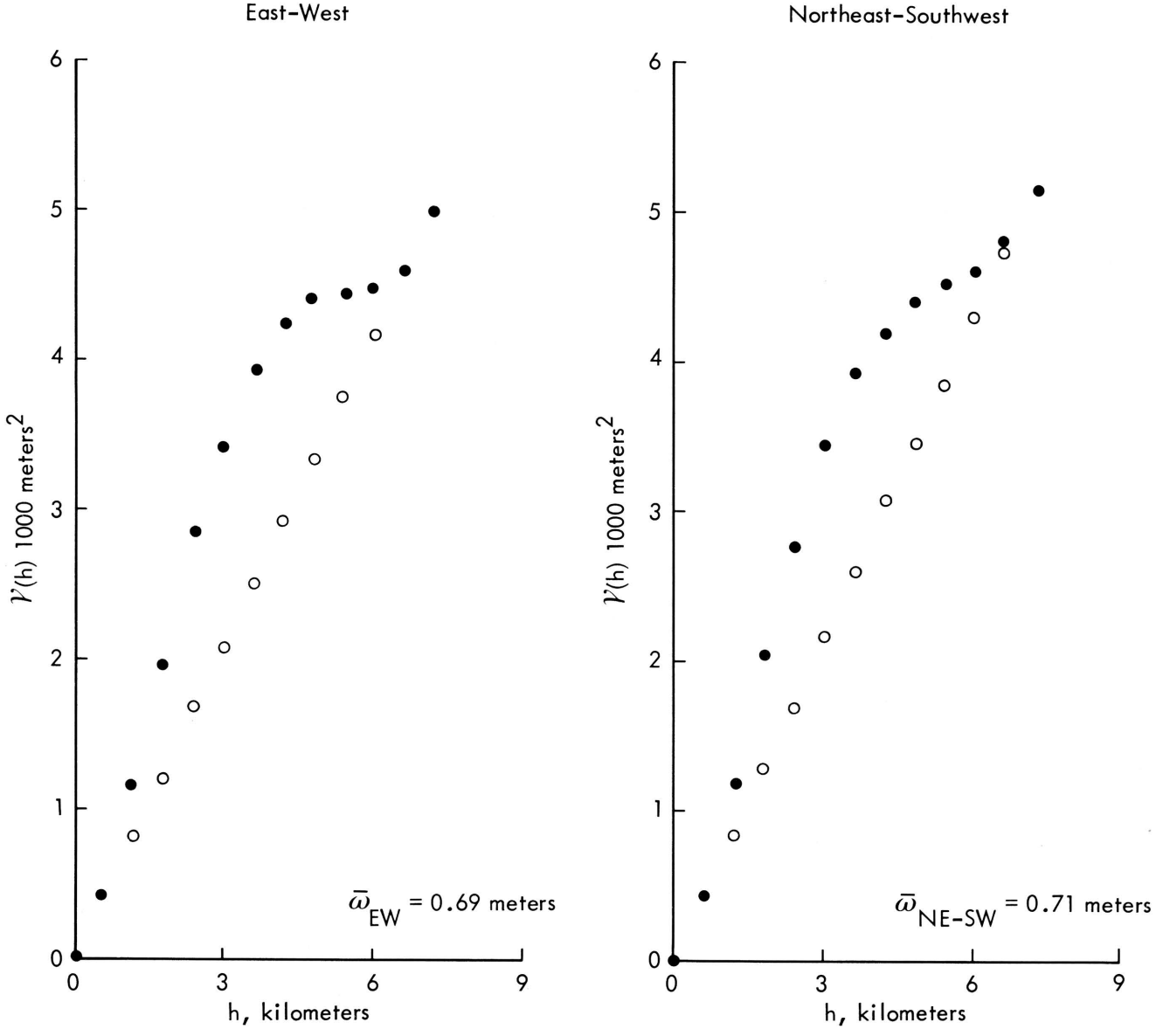


Figure 3.2.6--Experimental semivariograms for the top of the Tobifera Series, assuming a linear drift. The neighborhood is 7.2 km, and the average slope at the origin, ω_A, is 0.7 meters. Open circles are the assumed linear semivariogram, dots are the experimental semivariogram.

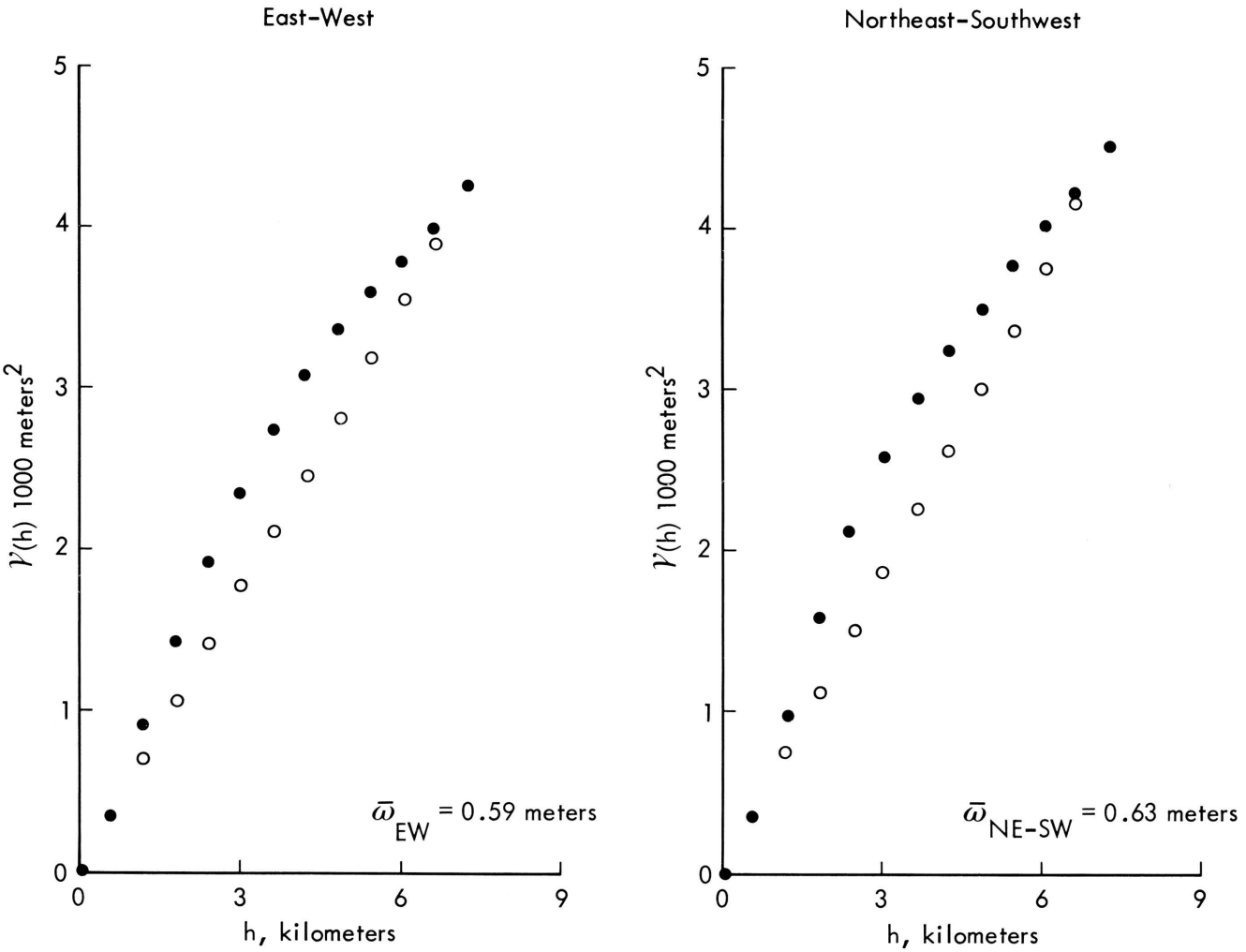


Figure 3.2.7--Experimental semivariograms for the top of the Tobifera Series, assuming a quadratic drift. The neighborhood is 7.2 km, and the average slope at the origin, ω_A, is 0.61 meters. Open circles are the assumed linear semivariogram, dots are the experimental semivariogram.

The general practice under isotropy or moderate anisotropy is to use an average slope $\bar{\omega}_A$ instead of $\bar{\omega}(\theta)$. The semivariograms indicate that for a given neighborhood size, the goodness of fit between the assumed and experimental semivariogram is always better for the case of a second-degree polynomial for the drift. Also, the semivariogram slope at the origin is smaller when the drift is a second-degree polynomial which means we may have greater confidence in the reliability of the estimation. It is also clear that as the neighborhood size increases the goodness of fit deteriorates regardless of the degree of the polynomial for the drift. The optimum neighborhood has a 2.1 km radius, as a smaller neighborhood will not include enough observations to perform the calculations and for a larger one the model will be in error. The second degree of polynomial drift is optimal for universal kriging. For drift estimation we will see

later the best choice is not only restricted by the structural analysis but also by the working scale which yields the best physical interpretation of the results. For a first-degree polynomial for the drift we will use the expression

$$\gamma(h) = 0.61h, \quad h \leqslant 4200 \text{ meters}$$

And for a second-degree polynomial for the drift we will use

$$\gamma(h) = 0.49h, \quad h \leqslant 4200 \text{ meters}$$

3.3 Grid Generation

Now that we have found a semivariogram and established the form of the drift, we are almost ready to perform kriging and drift estimation. However, there is still an important practical consideration concerning the number of samples to be used to krige each point of the regular grid.

The relation between the neighborhood $(\vec{x}_0, r)$ and the collection of samples available to produce a kriged estimate at a point is seldom optimal. Sometimes not enough points are contained within the neighborhood and a specific point cannot be estimated. Sometimes scores of samples are included within the neighborhood of almost any point. Remember from Algorithm 2.4.17, if we use k samples to krige a point, we must solve a system of simultaneous equations with a (k+6) by (k+6) coefficient matrix for quadratic drift. This system of equations must be solved for each of the hundreds or thousands of points in the contouring grid. Any reduction in the number of samples that must be considered is very important. For instance, if we can reduce the samples by one half, computer execution time will drop to one fourth. Although ideally we should keep all the information contained in data points within the neighborhood, there are practical reasons for not doing so. To use 84 points, for instance, means that we must invert a 90 by 90 matrix. Procedures used to discard excess points must be different for universal kriging and drift estimation as the systems of equations which yield the solutions behave differently. For universal kriging, the screen effect (Appendix G) shows that considering all available information or considering only the closest samples is, for practical purposes, equivalent. The contributions of the more distant samples are immaterial.

To implement an efficient sample discarding procedure for universal kriging we may divide the space around the point to be kriged into eight octants and keep the closest points in each octant. This insures a balanced distribution of samples, not only by distance, but also by radial direction. With samples in all octants (Fig. 3.3.1), tests indicate that a limit of 16 samples or two points per octant are adequate (Appendix G). With fewer than nine samples we begin to have trouble with meaningless results because of instability in the solutions for Algorithm 2.4.17. Therefore, we may provide a secondary alternative for those points containing fewer than nine samples per neighborhood by widening the neighborhood to include four samples per octant, which creates a maximum 38 by 38 matrix. If we have fewer than half the maximum number of points possible in a two sample per octant interval, it is unlikely that we will obtain 32 samples within the expanded neighborhood. In fact, most instances of sparse control

occur along the margins of the map where at least half the octants are empty (Fig. 3.3.2). Hence, use of wider neighborhoods for special cases usually increases the number of samples to no more than 16. This alternative restricts us to solving, on an average, a 22 by 22 matrix. To retain four samples per octant in every case would be approximately four times more expensive and would provide no significant improvement in accuracy.

Discarding samples for drift estimation is almost impossible as there is no screen effect and we cannot change the size of the neighborhood. However, as 90 by 90 matrices cannot be efficiently manipulated, some discarding procedure must be implemented. The elimination of samples should result in uniform sample distribution rather than keeping the nearest points to the estimation location as in universal kriging. In our particular example, the number of observations per estimated point is always within reasonable limits.

The location $\vec{x}$ in our example is determined by two coordinates, the latitude x and the longitude y. Consequently, the expression for a quadratic drift would be:

$$m'(x) = a_1x + a_2y + a_3x^2 + a_4xy + a_5y^2$$

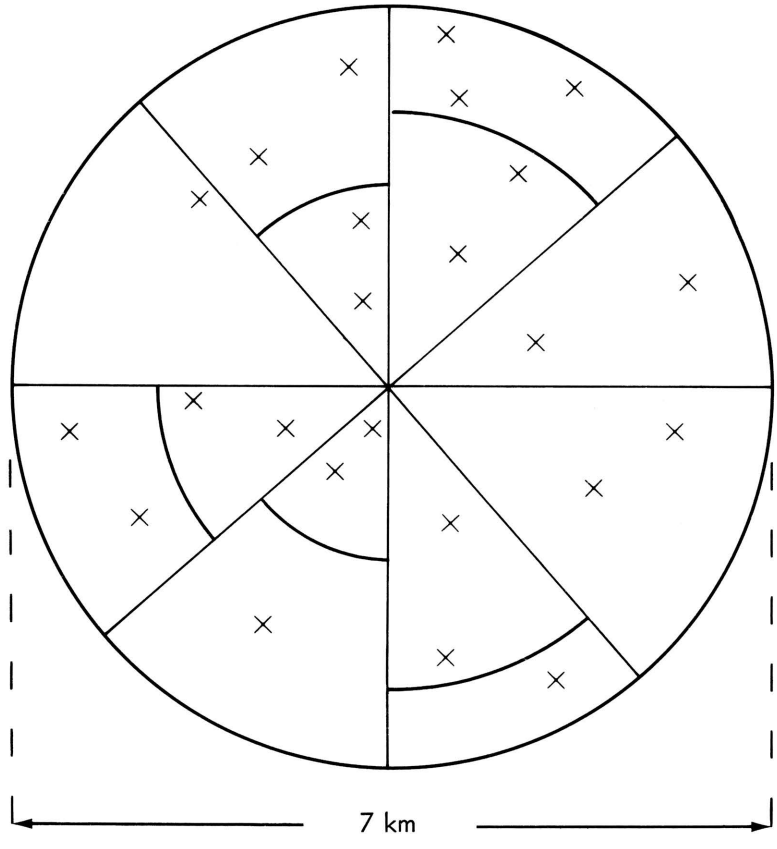


Figure 3.3.1--Seven-kilometer neighborhood divided into octants, with two points to be considered in each octant.

Hence, there will be five functions $f^i(\vec{x})$ in matrices G and H.

$$f^1(x) = x$$

$$f^2(x) = y$$

$$f^3(x) = x^2$$

$$f^4(x) = xy$$

$$f^5(x) = y^2$$

The program used to generate all grids has been incorporated under the module KRIGE of SURFACE II, a graphics system developed by the Kansas Geological Survey (Sampson, 1975). This system also constructed the contour maps of the grids. The grids with kriged values and the estimation variances are stored on a file which is used by the line-drawing modules, which converts the grid values into plotter instructions to produce finished maps. Although contour line construction is a significant aspect of automatic mapping, it is independent of the gridding algorithm and will not be considered here.

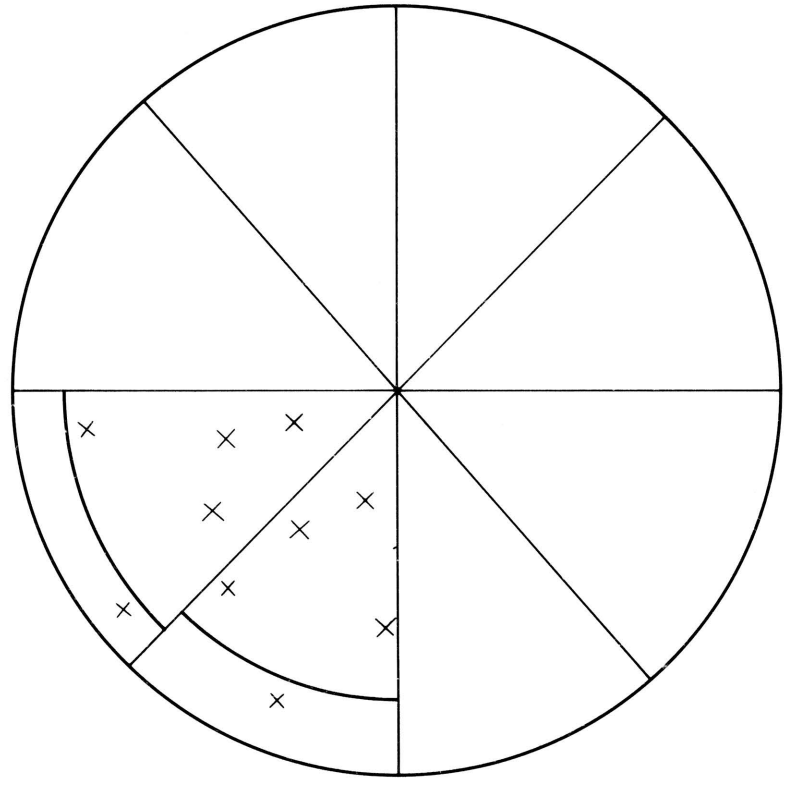


Figure 3.3.2--Neighborhood divided into octants, with four points to be considered in each octant. This illustrates the special case which occurs in the corner of a map where over half the octants are empty.

3.4 Results

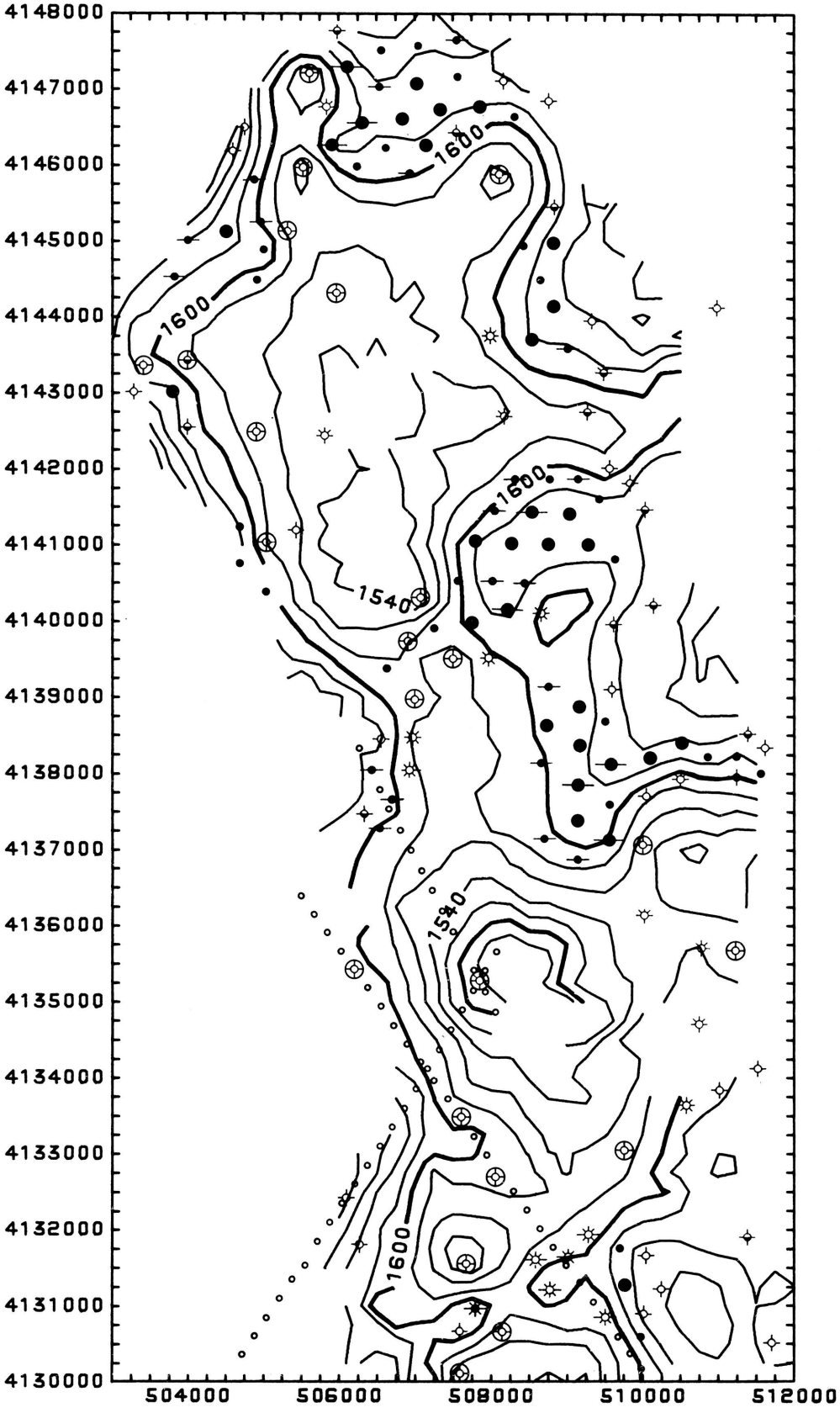
Figure 3.4.1 is the map of the top of the Tobifera Series in the Cullen field of the Magellan Basin, Chile, made by universal kriging using a 37 by 73 grid, second-degree polynomial for the drift, and 2.4 km neighborhood radius. Figure 3.4.2 is the square root of the estimation variance produced in the same manner. Figures 3.4.1 to 3.4.6 are at a scale of 1 cm = 0.75 km or 1:75,000.

On the University of Kansas Honeywell 635 computer, computation expenses to obtain the tape with the commands for the plotter were $14. This includes the commands to generate both the map of the top of the Tobifera Series and the error map. Therefore, the average cost per map is $7. However, if we want to estimate by universal kriging the top of the tuff and not produce the error map, the cost is essentially unchanged, because the cost of the error map is marginal. The cost for the same commands but using the standard estimation procedure in SURFACE II was $8; therefore, the kriging algorithm implemented in this study is extremely economical as compared to previously implemented versions (Olea, 1972).

To interpret these maps, it is useful to work on a light-table so they may be superimposed. Suppose for the same location on both maps we read a 1600-meter elevation on the top of the tuff and 10 meters on the error map. If it is assumed that the random variable $Z_s^*(\vec{x}_0) - Z(\vec{x}_0)$ is normally distributed, the probability is 95 percent that the absolute value of the difference between the estimated depth and the true depth is less than 20 meters, or twice the standard deviation indicated on the error map.

The error map is a very powerful tool which can be used to analyze the reliability of each feature in the map produced by universal kriging. Certain, probable, and fictitious structures can be differentiated in an objective manner. The error map and the semivariogram can also be used to determine where more information is needed to refine the map, and to estimate the number of additional samples needed. In this way, the regionalized variable theory provides criteria on which future sampling can be planned to achieve a specified reliability in areas with insufficient information. The error can always be reduced below a pre-specified limit by taking more samples in these critical areas. As kriging is an optimal method of estimation, no other technique will be able to provide better estimates in areas of poor control; more samples are the only answer.

The purpose of drift estimation is to provide a map which emphasizes local structures that may be especially prospective. In this sense, the drift itself is not usually the most helpful tool; a map of the residuals from the drift usually is more revealing. However, for a given regionalized variable, there will be an infinite number of drifts, one for each working scale. The selection of a neighborhood size and analytical expression to represent the drift are parameters to be adjusted to achieve the final goal, being the choice of the appropriate drift. All may be equally correct in terms of regionalized variable theory if their parameters have been selected properly. Figures 3.4.3 and 3.4.4 are two possible drift representations of the top of the Tobifera Series for a neighborhood radius of 2.1 km. Figure 3.4.3 uses a second-degree polynomial and Figure 3.4.4 uses a first-degree polynomial drift. In terms of a practical exploration problem, some drifts and their associated residuals maps will do better than others.



Well Symbols

- ● Cumulative oil production greater than 400,000 bbls., or initial production above average.
- • Cumulative oil production less than 400,000 bbls., or initial production fair to poor.
- ✲ Gas.
- ✷ Oil and gas.
- ✧ Oil shows.
- ✲ Gas shows.
- ✷ Oil and gas shows.
- ✧ Hydrocarbons not detected.

Modifications to Basic Symbols

- — Abandoned well, as:
  - ⊸ abandoned oil well.
- ○ Springhill Group missing (production, if any, from units adjacent to the unconformity), as:
  - ⊗ Springhill Group missing, gas shows.

Other Information Sources

- ∘ Seismic survey data.

Scale: 1:75,000
Contour interval: 20 m.

Figure 3.4.1--Structure contour map of top of Tobifera Series in Cullen area, estimated by universal kriging. Contours in meters below sea level.

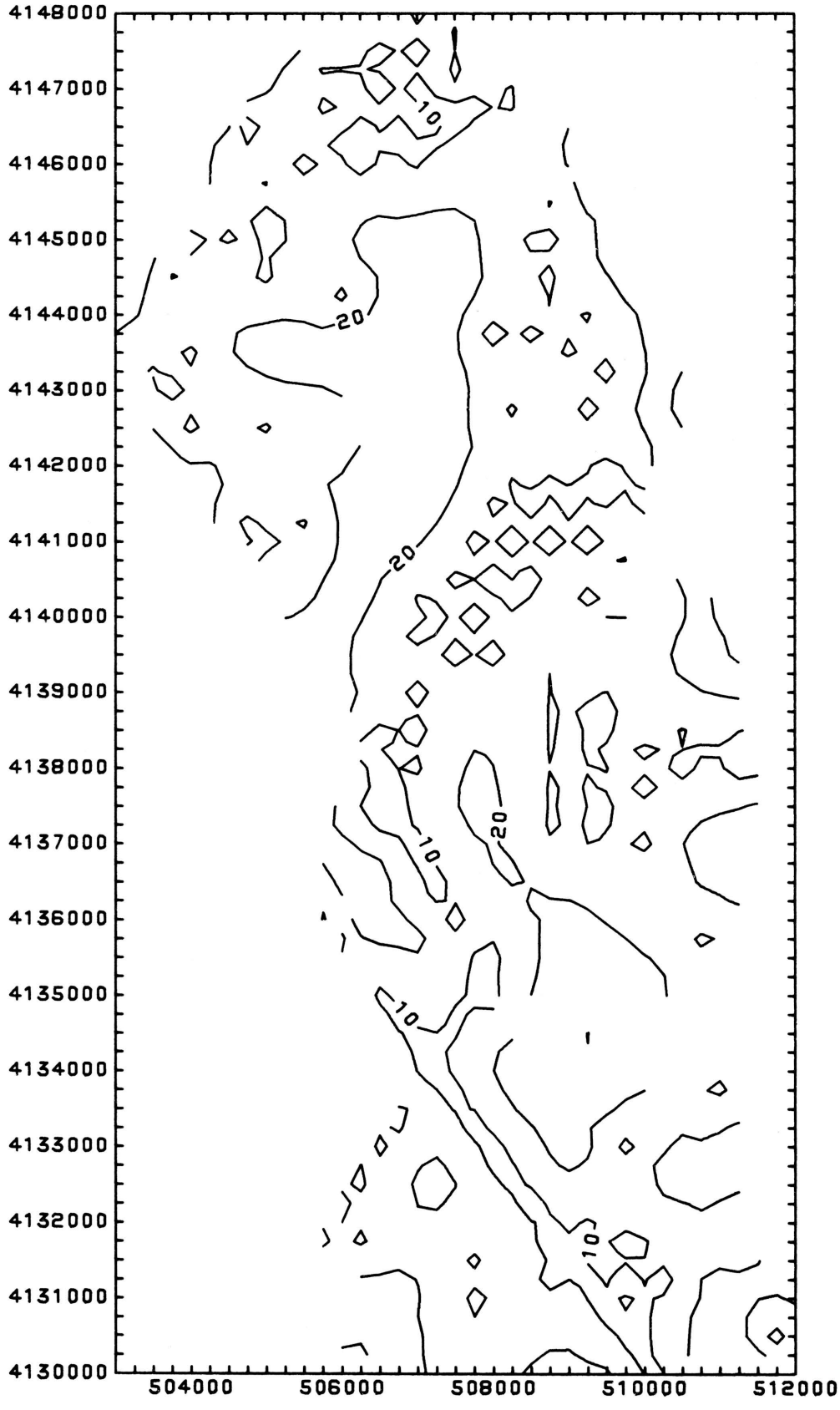


Figure 3.4.2--Error map, giving square root of estimation variance for map of top of Tobifera Series. Contours in meters.

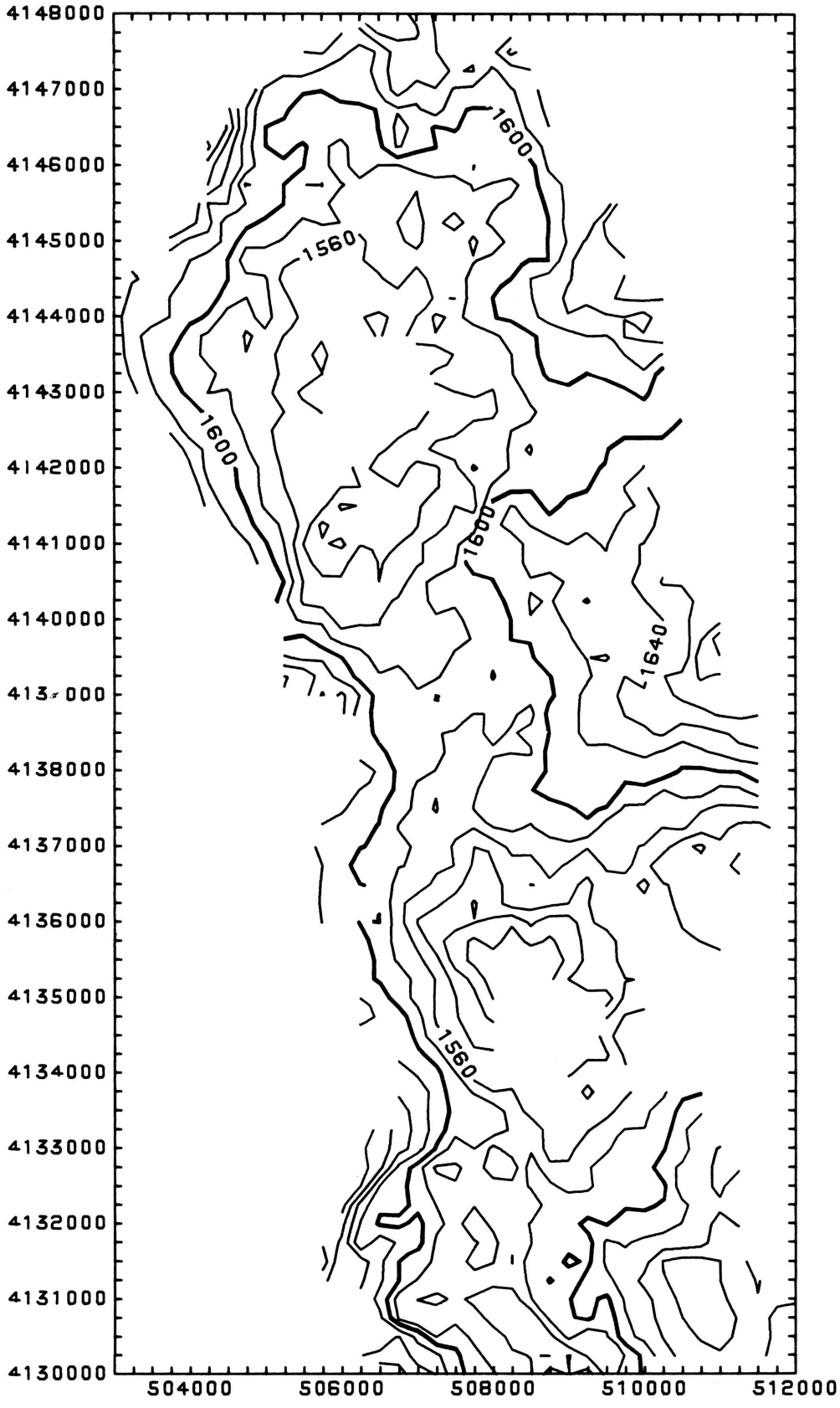


Figure 3.4.3--Second-degree drift of top of Tobifera Series, Cullen area, made using 2.1 km neighborhood radius. Contours in meters below sea level.

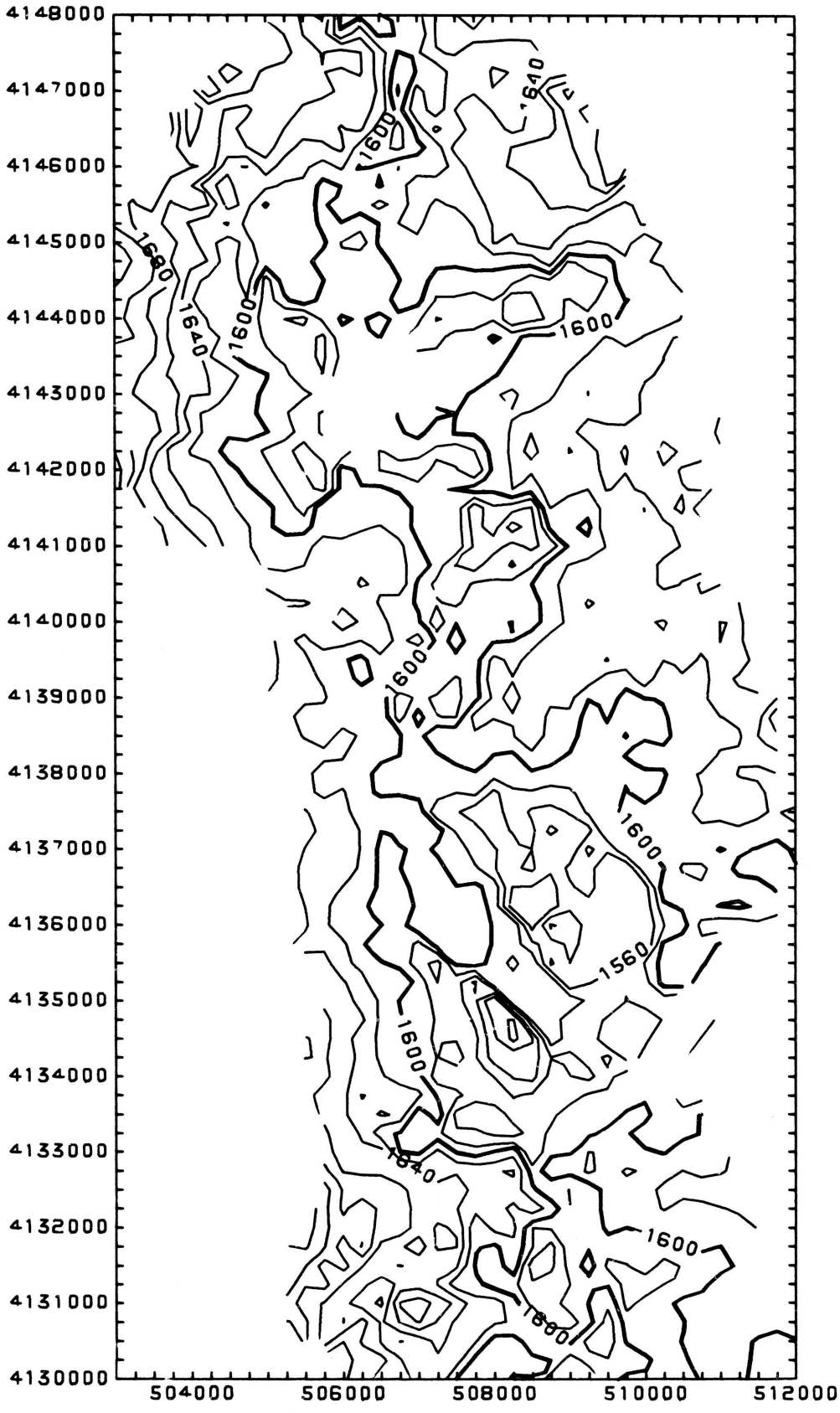


Figure 3.4.4--First-degree drift of top of Tobifera Series, Cullen area, made using 2.1 km neighborhood radius. Contours in meters below sea level.

The choice of the best working scale is beyond what the theory of regionalized variables can forecast and is entirely a matter of experience in the area of application. However, once experience has been gained in an area, the results will be useful in forecasting the same regionalized variable in surrounding areas. In the example, we are interested in a drift yielding a map of residuals whose structures reflect the sedimentary environments at the time of the Springhill Group deposition and which can be used to forecast oil production. In this sense, Figure 3.4.4 is a map of the best drift for the Tobifera Series top in the Cullen field area and Figure 3.4.5 is the corresponding estimation standard deviation map. Figure 3.4.6 is the map of residuals obtained by subtracting from the drift the estimates for the top of the Tobifera Series in Figure 3.4.1. The contour lines in the drift map in Figure 3.4.4 appear very complex, but the complexity is not the result of an extremely variable surface, but rather of the extremely low relief.

The map of residuals in Figure 3.4.6 was subdivided into three classes and two of the resulting subareas shaded to assist in interpretation. Areas of negative residuals are dotted and those with positive residuals greater than 30 meters are shown by dark shading. Intermediate areas of residuals between 0 and 30 meters were left blank, as are areas in which contouring could not be performed. Well location and production symbols are the same as in Figure 3.4.1. The wells have been grouped in four production categories:

1) Wells with cumulative oil production of 400,000 barrels or more, and recently drilled wells with above-average initial production.
2) Wells with cumulative production of less than 400,000 barrels or poor initial production.
3) Dry wells which encountered the Springhill Group.
4) Dry wells where the Springhill Group is missing.

The negative residuals in Figure 3.4.6 appear to be related to the large embayments formed in the northeastern portion of the Cullen field during the transgression at the end of the Jurassic. Marine currents, winds, tides, sediment sources or other geologic factors did not permit sand deposition except in bays open to the north, northeast, or east, and sand deposition was precluded on the promontories between the bays. Fine sediments laid down over the tuff were not deposited contemporaneously with the Springhill sandstone, but at a quiet time when the water depth was sufficient to cover the promontories.

The west side of the Cullen structure appears steeper and more monotonous than the east side. This could be explained as the result of quiet oceanic environment. Calm water would also explain the relative thinning of Lower Cretaceous sediments in the west as compared to the Tobifera Series in the east. Oil production in the west is poor, especially in the southwest where the scattered production is associated with shaly sediments having poor porosity. In the remainder of the Cullen area, oil production and residual magnitude are highly correlated. Figures 3.4.7 and 3.4.8 may be used to visualize this relationship.

Eighty-six percent of all the wells drilled where the Springhill Group is missing are in areas of positive residuals greater than 30 meters. In areas of negative residuals, 81 percent of the wells have high oil production and only 2 percent of the wells encounter a section in which the Springhill Group is missing. The proportion of high production wells is more than six times greater in negative residual areas than in areas of positive residuals of more than

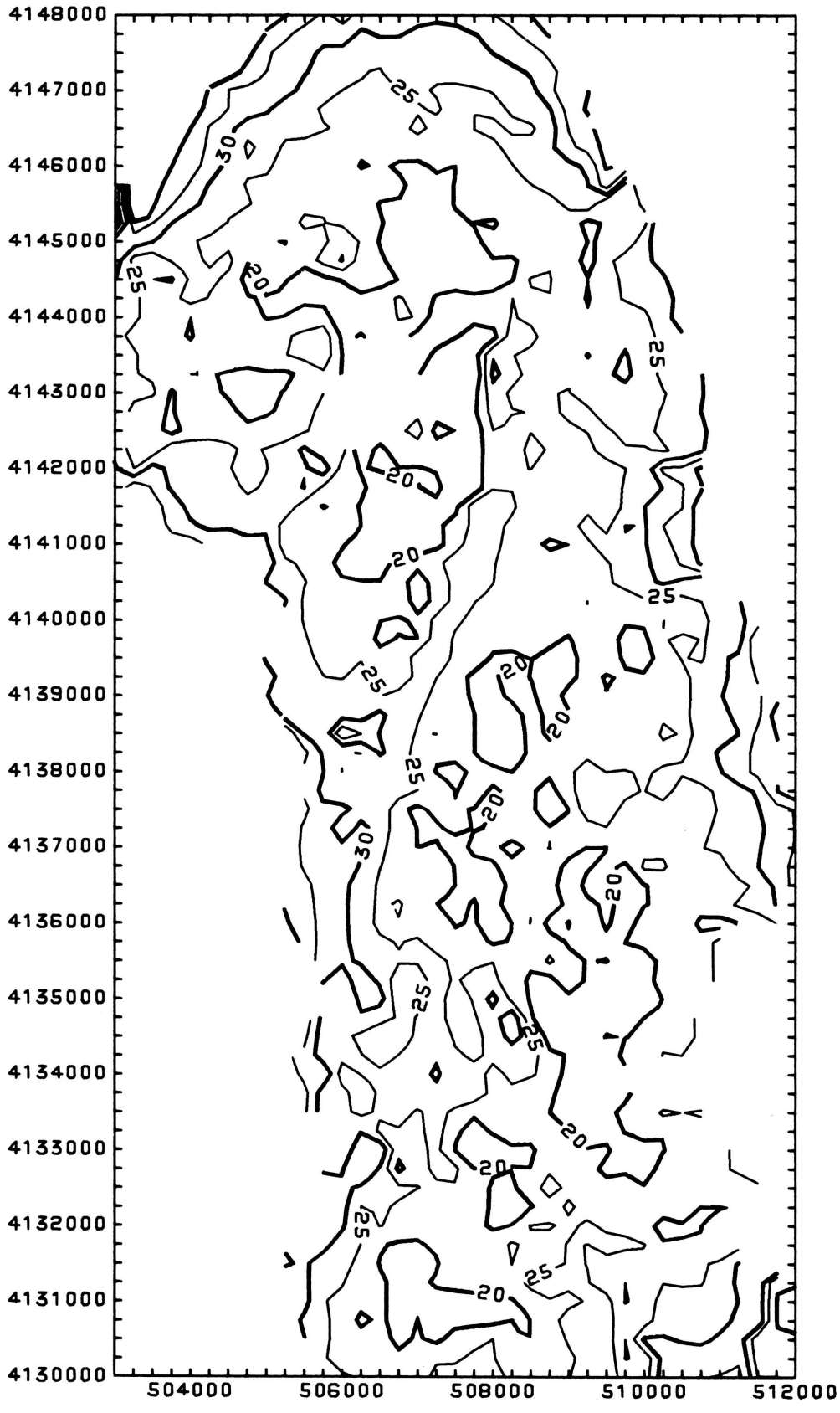


Figure 3.4.5--Error map, giving square root of estimation variance for first-degree drift of top of Tobifera Series. Contours in meters.

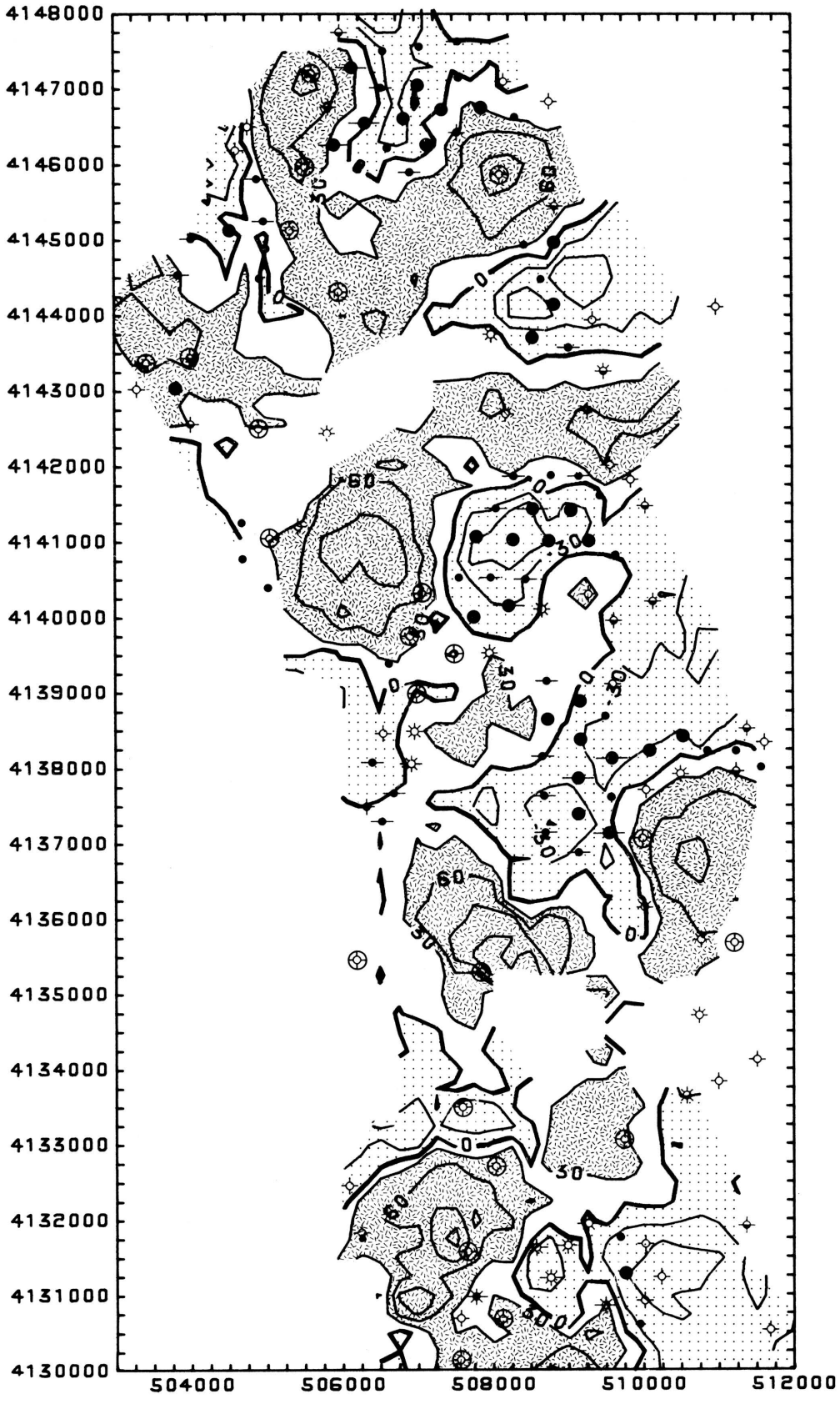


Figure 3.4.6--Residuals from first-degree drift, obtained by subtracting surface shown in Figure 3.4.4 from structure contour map in Figure 3.4.1. Contours in meters.

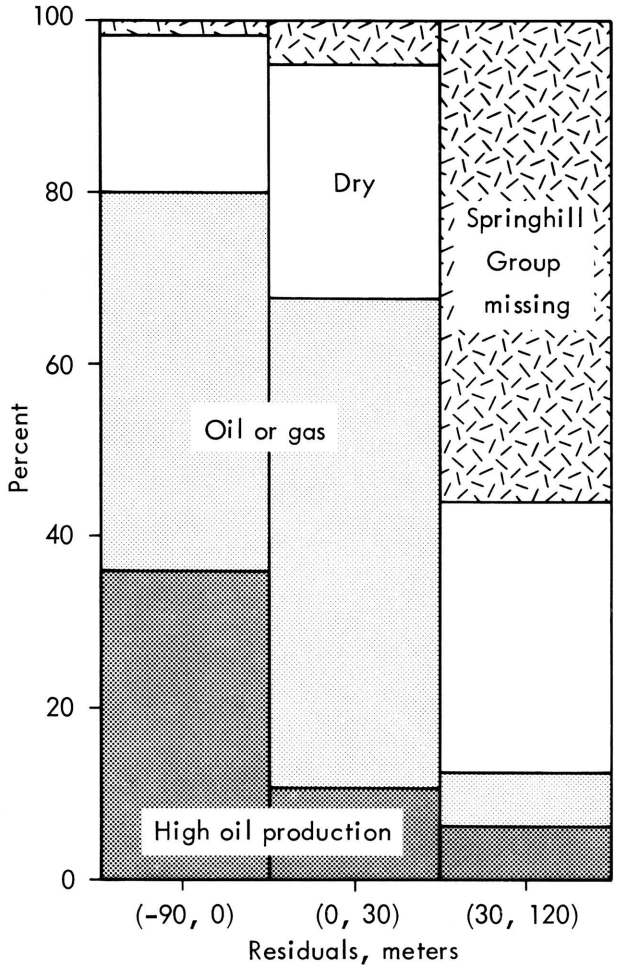


Figure 3.4.7--Distribution of production categories in classes of residuals as shown on Figure 3.4.6.

30 meters. The proportion of all productive wells is almost seven times as great in the areas of negative residuals.

Areas of residuals of intermediate size are also intermediate in production. In such areas, the proportion of wells having high production is less than one sixth of that for the negative residual areas. Only 40 percent of the wells are producers compared to 57 percent productive wells in negative residual areas. The proportion of wells that do not encounter the Springhill Group is twice that of the negative residual areas.

Wells without commercial production of hydrocarbons are almost equally distributed in the three areas. However, where residuals are larger than +30 meters, the producing horizon tends to be impervious; in areas of negative residuals, the sandstone usually has good porosity but is water saturated.

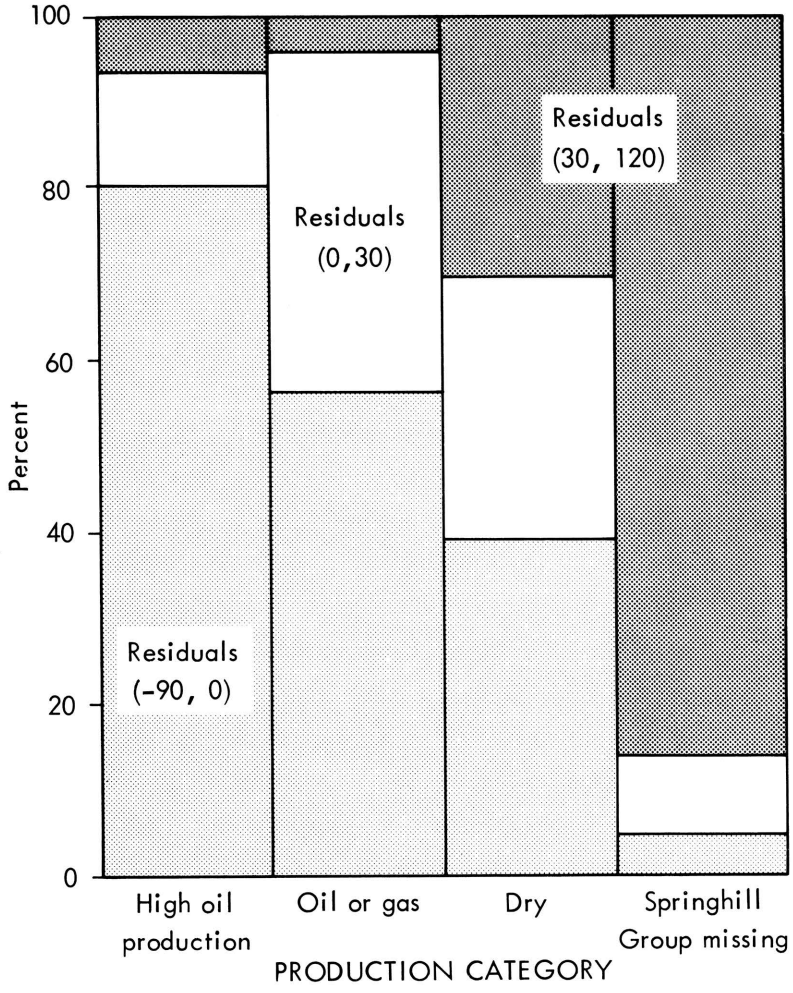


Figure 3.4.8--Proportion of each residual class in each production category.

3.5 Conclusions

The most widely used techniques for map analysis are variations of polynomial surface-fitting procedures. These methods and their simple mathematical formulations yield solutions which may not be valid or optimal. The theory of regionalized variables, although more complex, introduces a rigor that should produce better results, provided the observations can be regarded as outcomes of random variates and the residuals satisfy a particular kind of stationarity.

In contrast to other methods, algorithms based on the theory of regionalized variables are accompanied by guides to the selection of parameters. These parameters are found by a prior study known as a structure analysis. The parameters include measures of the degree of continuity and isotropy of the regionalized variable, the regional drift, and the influence zone of the samples. Relationships among these parameters preclude their independent selection. Computer time for structural analysis is very small; the main expense is usually incurred in the man-hours spent inspecting the data collection.

Automatic contouring methods are techniques to represent a three-dimensional, single-valued function by two-dimensional isopleths drawn using a plotter. The task is performed in

several steps, starting with raw data input to the computer. Succeeding steps differ from one implementation to another, but the basic problem is the estimation of an array of points located at nodes of a regular grid. Universal kriging is used here as an exact interpolation procedure for grid generation.

In map analysis, knowledge of the surface itself sometimes is not as crucial as formation of an impression of the main components of the surface. Drift estimation is presented here as an alternative to trend surface analysis for this purpose. The drift is defined in a statistical way as the expected value of the regionalized variable. The actual drift is relative to the working scale. Therefore, the drift at a point is not unique. The resulting maps differ substantially from trend surface polynomials. Trend and drift, although similar in concept, are different mathematically and in appearance.

Universal kriging and drift estimation are optimal in the sense that they are unbiased and the estimation variance is a minimum. They do equally well with regularly distributed or randomly distributed samples.

Advantages of kriging and drift estimation have been demonstrated theoretically and by means of an example. In particular, maps of residuals were used in a hindsight prediction of paleoenvironments of deposition and hence the occurrence of oil in the Cullen field of Chile.

Because universal kriging and drift estimation are optimal, provided the fundamental assumptions expressed in the intrinsic hypothesis are met, they do not require an external yardstick to measure their effectiveness. However, the methods do provide a statistic which shows the reliability of the estimation. This statistic, the estimation variance, is not an indicator of the effectiveness of the techniques, but rather depends upon the quality of sampling and the continuity of the regionalized variable. The estimation variance shows the reliability of the contour map and reflects the density and distribution of control points and the degree of spatial autocorrelation within the surface. Certainly this measure is a highly desirable feature of an estimation procedure.

The optimal drift for the purpose of revealing a physical phenomenon (such as oil occurrence) is related to the search for the appropriate working scale and must be found by experimentation, combined with additional specialized criteria from the particular area of application. This additional, troublesome, but unavoidable analysis is not necessary for universal kriging itself, which always yields the best possible estimates of the mapped phenomenon.

The error map and structure analysis are useful for planning additional sampling when more accurate results are required. The structural analysis will indicate the maximum permissible spacing between control points, and the error map will show those areas where control is unacceptably poor.

A contour map produced using universal kriging is approximately twice as expensive as one obtained using simple, empirical estimation procedures. The ratio is more nearly equal when comparing universal kriging to more elaborate contouring methods or if the conventional map requires that several parameters be selected and tested; the cost of the final conventional map is then several times the cost of a single computer run. In addition, universal kriging can produce an error map with almost no added computational expense.

Regionalized variable techniques are complex statistical methods which may be unnecessarily refined for preliminary examination of data with an homogeneous distribution or where high precision and error estimation are not necessary. However, universal kriging and drift estimation should be adopted in any map analysis involving significant research, economic decisions, or any kind of analysis where spatial estimation has an important bearing on the decision-making process.

REFERENCES

Bartlett, M. S., 1966, An introduction to stochastic processes with special references to methods and applications: Cambridge Univ. Press, Cambridge, 362 p.

Blais, R. A., and Carlier, P. A., 1967, Applications of geostatistics in ore evaluation: Canadian Inst. of Mining and Metallurgy, Ore Reserve Estimation and Grade Control, Special Vol. 9, p. 41-68.

Delhomme, J. P., and Delfiner, P., 1973, Application du krigeage à l'optimisation d'une campagne pluviométrique en zone aride: Symposium on the Design of Water Resources Projects with Inadequate Data, UNESCO-WHO-IAHS, Madrid, Spain, June 1973, p. 191-210.

Gonzalez, E., et al., 1965, La cuenca petrolífera de Magallanes: Minerales, Año XX, no. 91, Santiago de Chile, p. 42-62.

Hogg, R. V., and Craig, A. T., 1970, Introduction to mathematical statistics: Macmillan Co., New York, 415 p.

Huijbregts, C., and Matheron, G., 1971, Universal kriging (an optimal method for estimating and contouring in trend surface analysis): Canadian Institute of Mining and Metallurgy, Decision-Making in the Mineral Industry, Special Vol. 12, p. 159-169.

Journel, A., 1969, Rapport d'etudes sur l'estimation d'une variable régionalisée: application à la cartographie automatique: Service Hydrographique de la Marine, Paris, 110 p.

Karas, S., and Paccou, Y., 1969, Le krigeage universal et ses applications à la cartographie automatique: Ecole des Mines de Paris, Département Géostatistique, Centre de Morphologie Mathématique, Fontainebleau, 25 p.

Krige, D. G., 1966, Two-dimensional weighted moving average trend surfaces for ore evaluation: Proc. Symposium on Mathematical Statistics and Computer Applications in Ore Valuation, Johannesburg, p. 13-79, 379-380.

Matheron, G., 1963, Principles of geostatistics: Economic Geology, v. 58, p. 1246-1266.

Matheron, G., 1965, Les variables régionalisées et leur estimation. Une application de la théorie des fonctions aléatoires aux sciences de la nature: Masson et Cie., Editeurs, Paris, 305 p.

Matheron, G., 1967a, Presentation des variables régionalisée: "Técnica," Revista dos Alunos do Instituto Superior Técnico, No. 386, Lisboa, p. 401-413.

Matheron, G., 1967b, Kriging or polynomial interpolation procedures? A contribution to polemics in mathematical geology: The Canadian Mining and Metallurgical Bulletin, v. 60, no. 665, p. 1041-1045.

Matheron, G., 1969a, Le krigeage universal: Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau, Fascicule 1, Edité par l'Ecole Nationale Supérieure des Mines de Paris, 83 p.

Matheron, G., 1969b, Cours de geostatistique: Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau, Fascicule 2, Edité par l'Ecole Nationale Supérieure des Mines de Paris, 80 p.

Matheron, G., 1970, Random functions and their applications in geology, in Geostatistics, a Colloquium, D. F. Merriam, ed.: Plenum Press, New York, p. 79-87.

Matheron, G., 1971, The theory of regionalized variables and its applications: Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau, Fascicule 5, Edité par l'Ecole Nationale Supérieure des Mines de Paris, 211 p.

Mix, D. F., 1969, Random Signal Analysis: Addison-Wesley Publ. Co., Reading, Mass., 271 p.

Olea, R. A., 1972, Application of regionalized variable theory to automatic contouring, Spec. Report, Project 131, American Petroleum Institute: Center for Research, Inc., Univ. Kansas, Lawrence, p. 1-191.

Olea, R. A., 1974, Optimal contour mapping using universal kriging: Jour. Geophysical Research, v. 79, no. 5, p. 695-702.

Olea, R. A., 1975, Measuring statistical dependence with semivariograms: Kansas Geological Survey Series in Spatial Analysis, Univ. Kansas, Lawrence, in press.

Palmer, J. A. B., 1969, Automated mapping: Proc. Fourth Computer Conference, Adelaide, South Australia, p. 463-466.

Poissonnet, M., Millier, C., and Serra, J., 1970, Morphologie mathématique et sylviculture: 3ème Conférence du groupe des Statisticiens Forestiers, Paris, p. 287-307.

Sampson, R. J., 1975, SURFACE II User's Manual: Kansas Geological Survey Series in Spatial Analysis, Univ. Kansas, Lawrence, 206 p.

Singleton, R. C., 1969, An efficient algorithm for sorting with minimal storage: Communications of the ACM, v. 12, no. 3, p. 185-187.

Switzer, P. C., Mohr, M., and Heitman, R. E., 1964, Statistical analyses of ocean terrain and contour plotting procedures: Project Trident Technical Report No. 1440464: Arthur D. Little, Inc., Cambridge, Mass., 81 p.

Walden, A., 1972, Quantitative comparison of automatic contouring algorithms, Kansas Oil Exploration Decision System Technical Report: Kansas Geological Survey, Univ. Kansas, Lawrence, 115 p.

Walters, R. F., 1969, Contouring by machine: A user's guide: American Assoc. Petroleum Geologists Bulletin, v. 35, no. 11, p. 2324-2340.

Watson, G. S., 1971, Trend-surface analysis: Jour. International Assoc. for Mathematical Geology, v. 3. no. 3, p. 215-226.

Watson, G. S., 1972, Trend surface analysis and spatial correlation, in Quantitative Geology, Symposium 82nd Ann. Mtg., Geol. Soc. America, Nov. 10, 1969: Geological Society of America, Spec. Paper 146, p. 39-46.

Wiener, N., 1949, Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications: The M.I.T. Press, Cambridge, Mass., 160 p.

APPENDICES

APPENDIX A

LAGRANGE'S METHOD OF MULTIPLIERS

This is a method for finding the maximum and minimum values of a function of several variables when relations between the variables are given. For instance, if it is desired to find the maximum area of a rectangle whose perimeter is a constant k, it is necessary to find the maximum value of xy for $2x + 2y - k = 0$. Lagrange's method of multipliers involves solving the system:

$$\begin{aligned} 2x + 2y - k &= 0 \\ \partial u/\partial y &= 0 \\ \partial u/\partial x &= 0 \end{aligned}$$

where $u = xy + \mu(2x + 2y - k)$.

Therefore, we must find x, y, and μ in the system:

$$\begin{aligned} 2x + 2y &= k \\ y + 2\mu &= 0 \\ x + 2\mu &= 0 \end{aligned}$$

whose solution is:

$$\mu = -k/8, \qquad x = y = k/4$$

In general, we may be given a function $f(x_1, x_2, \ldots, x_n)$ of n variables connected by h distinct relations. To find the values $x_1, x_2, \ldots, x_n$ for which this function may be a maximum or a minimum, equate to zero the partial derivatives of the auxiliary function:

$$u = f(x_1, x_2, \ldots, x_n) + \mu_1\phi_1 + \mu_2\phi_2 + \ldots + \mu_h\phi_h$$

with respect to $x_1, x_2, \ldots, x_n$. Regarding $\mu_1, \mu_2, \ldots, \mu_h$ as constants, solve these equations simultaneously with the given relations, treating the μ's as unknowns to be eliminated in a system of n + h unknowns.

APPENDIX B

Linear Drift Estimation for a Linear Semivariogram

As presented in Section 2.3, we must estimate the drift in order to find the estimated residuals for samples at regular intervals along a line. We will solve this problem for that particular instance with the following restriction: the samples are points distributed along a line in such a way that if jj > j, $(x_{jj}-x_j) = a(jj-j)$ as shown in Figure B.1. The distance x_k-x_1 is L. The linear semivariogram is:

$$\gamma(h) = \bar{\omega}\, h$$

From Definition 2.3.2, the statistic for the drift is:

$$M'(x) = Bx$$

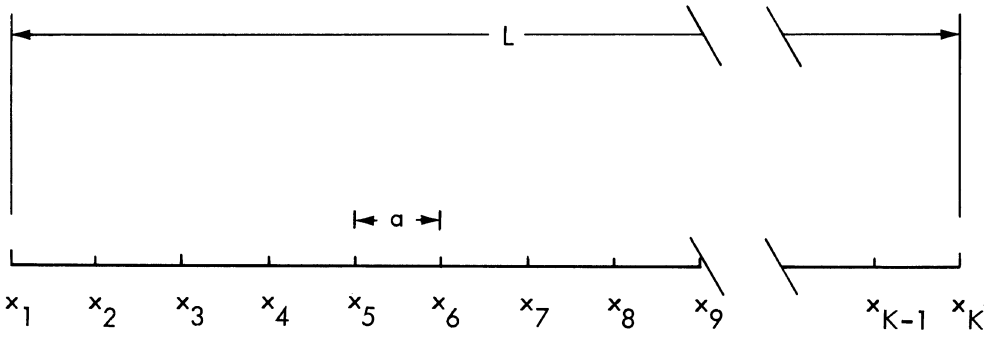


Figure B.1--Location of samples at regular intervals along a line.

Our problem is to determine the statistic B. To find the optimal λ_j, we must apply Algorithm 2.3.28. The system of equations in this case is:

$$\lambda_1\bar{\omega}(x_1-x_1)+\lambda_2\bar{\omega}(x_2-x_1)+\lambda_3\bar{\omega}(x_3-x_1)+\ldots+\lambda_k\bar{\omega}(x_k-x_1)+\mu x_1+\nu = 0$$

$$\lambda_1\bar{\omega}(x_2-x_1)+\lambda_2\bar{\omega}(x_2-x_2)+\lambda_3\bar{\omega}(x_3-x_2)+\ldots+\lambda_k\bar{\omega}(x_k-x_2)+\mu x_2+\nu = 0$$

$$\lambda_1\bar{\omega}(x_3-x_1)+\lambda_2\bar{\omega}(x_3-x_2)+\lambda_3\bar{\omega}(x_3-x_3)+\ldots+\lambda_k\bar{\omega}(x_k-x_3)+\mu x_3+\nu = 0$$

. . . . . .

$$\lambda_1\bar{\omega}(x_j-x_1)+\lambda_2\bar{\omega}(x_j-x_2)+\lambda_3\bar{\omega}(x_j-x_3)+\ldots+\lambda_k\bar{\omega}(x_k-x_j)+\mu x_j+\nu = 0$$

. . . . . .

$$\lambda_1\bar{\omega}(x_k-x_1)+\lambda_2\bar{\omega}(x_k-x_2)+\lambda_3\bar{\omega}(x_k-x_3)+\ldots+\lambda_k\bar{\omega}(x_k-x_k)+\mu x_j+\nu = 0$$

$$\lambda_1x_1 + \lambda_2x_2 + \lambda_3x_3 + \ldots + \lambda_kx_k = 1$$

$$\lambda_1 + \lambda_2 + \lambda_3 + \ldots + \lambda_k = 0$$

The first equation, for $x_j = x_1$, can be rewritten as:

$$\bar{\omega} \sum_{j=1}^{k} \lambda_j x_j - \bar{\omega} x_1 \sum_{j=1}^{k} \lambda_j + \mu x_1 + \nu = 0$$

and similarly for $x_j = x_k$,

$$\bar{\omega} x_k \sum_{j=1}^{k} \lambda_j - \bar{\omega} \sum_{i=1}^{k} \lambda_j x_j + \mu x_k + \nu = 0$$

But, by the last two equalities in the system of equations, these relationships reduce to:

$$\bar{\omega} + \mu x_1 + \nu = 0$$

$$-\bar{\omega} + \mu x_k + \nu = 0$$

Therefore,

$$\mu = \frac{2\bar{\omega}}{L}$$

$$\nu = \frac{-\bar{\omega}(x_k + x_1)}{L}$$

Replacing μ and ν back into the second equation of the system and grouping in a different way yields:

$$-x_2\left(\sum_{j=1}^{k} \lambda_j - 2\lambda_1\right) + \left(\sum_{j=1}^{k} \lambda_j x_j - 2\lambda_1 x_1\right) + \frac{2x_2}{L} - \frac{x_k + x_1}{L} = 0$$

But, since the first summation is zero and the other is one,

$$2x_2\lambda_1 + 1 - 2\lambda_1 x_1 + \frac{x_2 - x_1 - (x_k - x_2)}{L} = 0$$

Therefore,

$$\lambda_1 = -\frac{x_2 - x_1}{L(x_2 - x_1)} = -\frac{1}{L}$$

Now, replacing λ_1, μ, and ν in the third equation of the system,

$$-x_3\left(\sum_{j=1}^{k} \lambda_j + \frac{2}{L} - 2\lambda_2\right) + \left(\sum_{j=1}^{k} \lambda_j x_1 + \frac{2x_1}{L} - 2\lambda_2 x_2\right) + \frac{(x_3 - x_1) - (x_k - x_3)}{L} = 0$$

which can be reduced to:

$$[L + (x_3 - x_1) - x_k - x_3] - 2(x_3 - x_1) = 2L(x_2 - x_3)\,\lambda_2$$

or

$$L + 2h - L + 2h - 4h = L(x_2 - x_3)\,\lambda_2$$

Therefore,

$$\lambda_2 = 0$$

In a similar way, it can be shown that all the remaining weights are zero, except for the last one which is:

$$\lambda_k = \frac{1}{L}$$

Therefore, the statistic B_1 is:

$$B_1 = \sum_{j=1}^{k} \lambda_j \, Z(x_j) = \frac{Z(x_k) - Z(x_1)}{L}$$

APPENDIX C

Quadratic Drift Estimation for a Linear Semivariogram

Let the x_i be the locations of k samples along a line at regular intervals. From Algorithm 2.3.28, the solution of the problem of finding B_1 and B_2 in the estimated drift

$$M'(x) = B_1 x + B_2 x^2$$

is the solution of a system of equations

$$CW_1 = D_1$$

$$CW_2 = D_2$$

We will suppose that the location of the samples is similar to that in Figure B.1 and that $x_0 = 0$.

Remember, from Definition 2.3.21,

$$B_s = \sum_{j=1}^{k} \lambda_s^j \, Z(x_j)$$

By using Definition 2.3.8 for $\delta(i,s)$, we can define two auxiliary functions $p_1(j)$ and $p_2(j)$:

$$p_1(j) = \delta(j,k) - \delta(j,1)$$

$$p_2(j) = a - \frac{L+a}{2} [\delta(j,k) + \delta(j,1)]$$

where a is the distance between consecutive samples and L is the distance between the first and the last. As we will see, a linear combination of $p_1(j)$ and $p_2(j)$ is the solution for λ_s^j:

$$\lambda_s^j = c_s^1 \, p_1(j) + c_s^2 \, p_2(j)$$

Therefore, the estimation of a quadratic drift with a linear semivariogram is obtained by finding the constants c_s^i for λ_1^j and λ_2^j.

We will use the restrictions

$$\sum_{j=1}^{k} \lambda_s^j \, f^i(x_j) = \delta(i,s)$$

in the system $CW_s = D_s$ in Algorithm 2.3.28 to calculate the coefficients c_s^i.

Calculation of c_1^1: Because

$$\sum_{j=1}^{k} \lambda_1^j \, x_j = 1$$

then,

$$c_1^1 \sum_{j=1}^{k} [\delta(j,k) - \delta(j,1)] x_j + c_1^2 \sum_{j=1}^{k} [a - \frac{L+a}{2} \{\delta(j,k) + \delta(j,1)\}] x_j = 1$$

But,

$$\sum_{j=1}^{k} [\delta(j,k) - \delta(j,1)]\, x_j = x_k - x_1 = L$$

and

$$\sum_{j=1}^{k} \left[a - \frac{L+a}{2} \{\delta(j,k) + \delta(j,1)\}\right] x_j = a \sum_{j=1}^{k} x_j - \frac{L+a}{2} (x_k + x_1)$$

$$= \frac{a^2(k-1)k}{2} - \frac{(L+a)}{2} L$$

$$= \frac{L(L+a)}{2} - \frac{L(L+a)}{2} = 0$$

Hence,

$$c_1^1 L = 1$$

and

$$c_1^1 = 1/L$$

Calculation of c_1^2: As

$$\sum_{j=1}^{k} \lambda_1^j x_j^2 = 0$$

then,

$$c_1^1 \sum_{j=1}^{k} [\delta(j,k) - \delta(j,1)]\, x_j^2 + c_1^2 \sum_{j=1}^{k} \left[a - \frac{L+a}{2} \{\delta(j,k) + \delta(j,1)\}\right] x_j^2 = 0$$

Let us examine each term separately:

$$\sum_{j=1}^{k} [\delta(j,k) - \delta(j,1)]\, x_j^2 = x_k^2 - x_1^2$$

$$= L^2$$

$$\sum_{j=1}^{k} \left[a - \frac{L+a}{2} \{\delta(j,k) + \delta(j,1)\}\right] x_j^2 = a \sum_{j=1}^{k} x_j^2 - \frac{L+a}{2} (x_k^2 + x_1^2)$$

$$= a \sum_{j=0}^{k=1} a^2 j^2 - \frac{(L+a)L^2}{2}$$

$$= \frac{a^3(k-1)(k)(2k-1)}{6} - \frac{(L+a)L^2}{2}$$

$$= \frac{L(L+a)(2L+a)}{6} - \frac{3L^3 + 3aL^2}{6}$$

$$= \frac{-L(L^2 - a^2)}{6}$$

Therefore,

$$L - c_1^2 \frac{L(L^2 - a^2)}{6} = 0$$

So,

$$c_1^2 = \frac{6}{L^2 - a^2}$$

Calculation of c_2^1: The restriction we will use now is:

$$\sum_{j=1}^{k} \lambda_2^j x_j = 0$$

But this is exactly the same expression used for c_1^1 except for the right side of the equality. Hence,

$$c_2^1 L = 0$$

which implies that c_2^1 is zero.

Calculation of c_2^2: Now,

$$\sum_{j=1}^{k} \lambda_2^j x_j^2 = 1$$

Using the expansion of c_1^2 for the left side of the equality and remembering that c_2^1 is zero,

$$c_2^2 = - \frac{6}{L(L^2 - a^2)}$$

Replacing all the values c_s^i we have found:

$$\lambda_1^j = \frac{1}{L} [\delta(j,k) - \delta(j,1)] + \frac{6}{L^2 - a^2} [a - \frac{(L+a)}{2} \{\delta(j,k) + \delta(j,1)\}]$$

$$\lambda_2^j = - \frac{6}{L^2 - a^2} [a - \frac{(L+a)}{2} \{\delta(j,k) + \delta(j,1)\}]$$

Therefore,

$$B_1 = \sum_{j=1}^{k} \left\{ \frac{1}{L} [\delta(j,k) - \delta(j,1)] + \frac{6}{L^2 - a^2} \{a - \frac{L+a}{2} [\delta(j,k) + \delta(j,1)]\} \right\} Z(x_j)$$

Expanding the expression,

$$B_1 = \frac{Z(x_k) - Z(x_1)}{L} - \frac{3[Z(x_k) + Z(x_1)]}{L-a} + \frac{6ak}{L^2-a^2} \frac{1}{k} \sum_{j=1}^{k} Z(x_j)$$

or, if we define Z_m as

$$Z_m = \frac{1}{k} \sum_{j=1}^{k} Z(x_j)$$

$$B_1 = \frac{Z(x_k) - Z(x_1)}{L} + \frac{3}{L-a} [2Z_m - Z(x_k) - Z(x_1)]$$

Analogously,

$$B_2 = - \frac{6}{L(L^2-a^2)} \sum_{j=1}^{k} [a - \frac{(L+a)}{2} \{\delta(j,k) + \delta(j,1)\}] Z(x_j)$$

or

$$B_2 = - \frac{3}{L(L-a)} [2Z_m - Z(x_k) - Z(x_1)]$$

Calculation of Lagrange's Multipliers: Here we will calculate the multipliers μ_s^i and prove that our solution of λ_s^j satisfies all equations in the system. The first group of equations in Algorithm 2.3.28 for a linear semivariogram is of the form:

$$\sum_{j=1}^{k} \lambda_s^j \omega|x_{jj}-x_j| + \mu_s^1 x_j + \mu_s^2 x_j^2 + \nu = 0$$

The first term can be rewritten as:

$$\sum_{j=1}^{k} \lambda_s^j \omega|x_{jj}-x_j| = \omega \sum_{j=1}^{jj} \lambda_s^j (x_{jj}-x_j) + \omega \sum_{j=jj+1}^{k} \lambda_s^j (x_j-x_{jj})$$

Replacing λ_s^j by its solution in terms of $p_i(j)$, we have four terms in the equality. We will study each of them separately.

1) $$\sum_{j=1}^{jj} p_1(j)(x_{jj}-x_j) = \sum_{j=1}^{jj} [\delta(j,k) - \delta(j,1)](x_{jj}-x_j)$$

$$= - (x_{jj}-x_1)$$

2) $$\sum_{j=jj+1}^{k} p_1(j)(x_j-x_{jj}) = \sum_{j=jj+1}^{k} [\delta(j,k) - \delta(j,1)](x_j-x_{jj})$$

$$= x_k-x_{jj}$$

3) $$\sum_{j=1}^{jj} p_2(j)(x_{jj}-x_j) = \sum_{j=1}^{jj} [a - \frac{L+a}{2}\{\delta(j,k) + \delta(j,1)\}](x_{jj}-x_j)$$

$$= ajjx_{jj} - a\sum_{j=1}^{jj} a(j-1) - \frac{(L+a)}{2}(x_{jj}-x_1)$$

$$= a(jj-1+1)x_{jj} - a^2\frac{(jj-1)(jj-1+1)}{2} - \frac{(L+a)}{2}(x_{jj}-x_1)$$

$$= (x_{jj}+a)x_{jj} - \frac{x_{jj}(x_{jj}+a)}{2} - \frac{(L+a)}{2}(x_{jj}-x_1)$$

$$= \frac{(x_{jj}+a)x_{jj}}{2} - \frac{(L+a)}{2}(x_{jj}-x_1)$$

4) $$\sum_{j=jj+1}^{k} p_2(j)(x_j-x_{jj}) = \sum_{j=jj+1}^{k} [a - \frac{L+a}{2}\{\delta(j,k) + \delta(j,1)\}](x_j-x_{jj})$$

$$= a\sum_{j=jj+1}^{k} (j-1)a - a(k-jj)x_{jj} - \frac{L+a}{2}(x_k-x_{jj})$$

$$= a^2\frac{(k-1)k}{2} - a^2\frac{(jj-1)jj}{2} - a(k-jj)x_{jj} - \frac{L+a}{2}(x_k-x_{jj})$$

$$= \frac{L(L+a)}{2} - \frac{x_{jj}(x_{jj}+a)}{2} - (L-x_{jj})x_{jj} - \frac{(L+a)}{2}(x_k-x_{jj})$$

Replacing all expressions from 1 to 4 in $\sum_{j=1}^{k} \lambda_s^j \omega|x_{jj}-x_j|$:

$$\sum_{j=1}^{k} \lambda_s^j \omega|x_{jj}-x_j| = c_s^1(x_k-x_{jj}-x_{jj}+x_1) + c_s^2[\frac{L(L+a)}{2} - \frac{L+a}{2}(x_k-x_{jj}+x_{jj}-x_1) - (L-x_{jj})x_{jj}]$$

$$\sum_{j=1}^{k} \lambda_s^j \omega|x_{jj}-x_j| = c_s^1(L-2x_{jj}) + c_s^2(x_{jj}-L)x_{jj}$$

Therefore, for s = 1,

$$\frac{\omega}{L}(L-2x_{jj}) + \frac{6\omega}{L^2-a^2}(x_{jj}-L)x_{jj} + \mu_1^1 x_{jj} + \mu_1^2 x_{jj}^2 + \nu_1 = 0$$

for jj = 1, 2, ..., k.

If jj = 1, $x_1 = 0$. Hence,

$$\frac{\omega L}{L} + \nu_1 = 0$$

Therefore,

$$\nu_1 = -\omega$$

For jj = 2 and 3,

$$\frac{6\omega}{L^2-a^2}(a-L)\,a - \frac{2\omega a}{L} + \mu_1^1 a + \mu_1^2 a^2 = 0$$

$$\frac{6\omega}{L^2-a^2}(2a-L)\,2a - \frac{4\omega a}{L} + 2\,\mu_1^1 a + 4\,\mu_1^2 a^2 = 0$$

Hence,

$$\mu_1^1 = \frac{2\omega(4L^2-a^2)}{L(L^2-a^2)}$$

$$\mu_1^2 = -\frac{6\omega}{L^2-a^2}$$

For s = 2,

$$\frac{6\omega(L-x_{jj})x_{jj}}{L(L^2-a^2)} + \mu_2^1 x_{jj} + \mu_2^2 x_{jj}^2 + \nu_2 = 0$$

Making jj equal to 1, se can see ν_2 is zero.

For jj = 2 and 3,

$$\frac{6\omega(aL-a^2)}{L(L^2-a^2)} + \mu_2^1 a + \mu_2^2 a^2 = 0$$

$$\frac{6\omega(2aL-4a^2)}{L(L^2-a^2)} + 2\mu_2^1 a + 4\mu a^2 = 0$$

Therefore,

$$\mu_2^1 = -\frac{6\omega}{L^2-a^2}$$

$$\mu_2^2 = \frac{6\omega}{L(L^2-a^2)}$$

We still must verify one more relation in the system of equations to see if the expression for λ_s^j is correct. This is the last equation in Algorithm 2.3.28 which states

$$\sum_{j=1}^{k} \lambda_s^j = 0 \qquad \text{for } s = 1, 2, \ldots, k$$

To prove this is true, we must show that

$$c_s^1 \sum_{j=1}^{k} p_1(j) + c_s^2 \sum_{j=1}^{k} p_2(j) = 0$$

But

$$\sum_{j=1}^{k} p_1(j) = \sum_{j=1}^{k} [\delta(j,k) - \delta(j,1)] = 1 - 1 = 0$$

and

$$\sum_{j=1}^{k} p_2(j) = \sum_{j=1}^{k} [a - \frac{L+a}{2} \{\delta(j,k) + \delta(j,1)\}]$$

$$= ka - \frac{L+a}{2} (1+1)$$

$$= L+a - (L+a) = 0$$

Therefore,

$$\sum_{j=1}^{k} \lambda_s^j = \sum_{j=1}^{k} [c_s^1 p_1(j) + c_s^2 p_2(j)] = 0$$

APPENDIX D

SEMIVARIOGRAM OF THE ESTIMATED RESIDUALS FOR A LINEAR SEMIVARIOGRAM AND A LINEAR DRIFT

Let the x_j be the locations of t samples along a line at regular intervals. Let the linear semivariogram of the residuals be:

$$\gamma(h) = \bar{\omega}\, h$$

By Theorem 2.5.9, the semivariogram for estimated residuals is in this case:

$$\gamma_E^1(h) = \gamma(h) - \frac{1}{2(k-p)} \sum_{j=k'}^{k+k'-p-1} \mu_1^1 \left[f^1(x_j+pa) - f^1(x_j)\right]^2$$

By Appendix B,

$$\mu_1^1 = \frac{2\,\bar{\omega}}{L}$$

and

$$\sum_{j=k'}^{k+k'-p-1} \left[f^1(x_j+pa) - f^1(x_j)\right]^2 = \sum_{j=k'}^{k+k'-p-1} p^2a^2 = (k-p)h^2$$

Therefore,

$$\gamma_E^1(h) = \gamma(h) - \frac{\bar{\omega}\, h^2}{L}$$

APPENDIX E

SEMIVARIOGRAM OF THE ESTIMATED RESIDUALS FOR A LINEAR SEMIVARIOGRAM AND QUADRATIC DRIFT

Let us define an auxiliary function $T_s^i(\vec{pa})$ as:

$$T_s^i(\vec{pa}) = \frac{1}{k-p} \sum_{j=k'}^{k+k'-p-1} [f^i(\vec{x}_j+\vec{pa}) - f^i(\vec{x}_j)][f^s(\vec{x}_j+\vec{pa}) - f^s(\vec{x}_j)]$$

We will assume that $h = \vec{pa}$ and that our t samples, $t \geqslant k+k'$, are taken at regular intervals a along a line. Therefore, we can reduce the analysis to one dimension. We can now rewrite Theorem 2.5.9 as:

$$\gamma_E(h) = \gamma(h) - 1/2 \sum_{i=1}^{n} \sum_{s=1}^{n} \mu_s^i \, T_s^i(h)$$

For n = 2, our task is to calculate four weights μ_s^i and four functions T_s^i.

Calculation of T_1^1: By its definition at the beginning of this appendix,

$$T_1^1(h) = \frac{1}{k-p} \sum_{j=k'}^{k+k'-p-1} (x_j+pa-x_j)^2$$

$$= \frac{1}{k-p} (k-p)h^2$$

$$= h^2$$

Calculation of T_1^2 and T_2^1: As $T_i^s = T_s^i$, we need only to calculate either of the two.

$$T_1^2(h) = \frac{1}{k-p} \sum_{j=k'}^{k+k'-p-1} [(x_j+pa) - x_j][(x_j+pa)^2 - x_j^2]$$

$$= \frac{h}{k-p} \sum_{j=k'}^{k+k'-p-1} (2x_j \, h+h^2)$$

$$= \frac{h^2}{k-p} \sum_{j=k'}^{k+k'-p-1} [2(j-k') \, a + ap]$$

We may define a new summation variable jj such that:

$$jj = j - k'$$

Then,

$$\text{if } j = k', \quad jj = 0$$

and

$$\text{if } j = k+k'-p-1, \quad jj = k - p - 1$$

Therefore,

$$T_1^2(h) = \frac{ah^2}{k-p} \sum_{jj=0}^{k-p-1} 2j + \frac{h^2}{k-p} \sum_{j=0}^{k-p-1} ap$$

But ap is independent of j. Hence,

$$T_1^2(h) = \frac{ah^2}{k-p} (k-p-1)(k-p) + h^3$$

$$= h^2 L$$

Calculation of T_2^2:

$$T_2^2(h) = \frac{1}{k-p} \sum_{j=k'}^{k+k'-p-1} [(x_j+h)^2 - x_j^2]^2 = \frac{h^2}{k-p} \sum_{j=k'}^{k+k'-p-1} (2x_j+h)^2$$

$$= \frac{h^2}{k-p} \sum_{j=k'}^{k+k'-p-1} [2(j-k')a+h]^2$$

$$= \frac{a^2h^2}{k-p} \sum_{j=k'}^{k+k'-p-1} [4(j-k')^2 + 4(j-k')p + p^2]$$

Introducing the same change of variable as for T_1^2:

$$T_2^2 = \frac{a^2h^2}{k-p} \left[\frac{4(k-p-1)(k-p)(2k-2p-1)}{6} + \frac{4(k-p-1)(k-p)p}{2} + (k-p)p^2\right]$$

$$= \frac{h^2}{3} [2(L-h)(2L-2h+a+3h) + 3h^2]$$

$$T_2^2 = \frac{h^2}{3} [4L^2-2Lh+h^2+2a(L-h)]$$

The multipliers μ_s^i were calculated in Appendix C. Therefore,

$$\Sigma\mu_s^i T_s^i = \frac{\omega(8L^2-2a^2)h^2}{L(L^2-a^2)} - \frac{12\omega h^2 L}{(L^2-a^2)} + \frac{6\omega}{L(L^2-a^2)} \frac{h^2}{3} [4L^2 - 2Lh + h^2 + 2a(L-h)]$$

$$= \frac{\omega h^2}{L(L^2-a^2)} [(4L^2+4aL-2a^2) - 4(L+a)h + 2h^2]$$

And,

$$\gamma_E^2(h) = \gamma(h) - \frac{\omega h^2}{L(L^2-a^2)} [2L^2+2aL-a^2-a(L+a)h + h^2]$$

for h = 0, a, 2a, ..., (k-1)a.

APPENDIX F

A Numerical Example of Structure Analysis

Here we will show how to numerically obtain a semivariogram and to check our initial hypotheses. We will use as data the depth to the top of a formation in six wells spaced one km apart along a line.

$$Z(x_1) = 1470 \text{ (meters)}$$
$$Z(x_2) = 1482$$
$$Z(x_3) = 1520$$
$$Z(x_4) = 1532$$
$$Z(x_5) = 1544$$
$$Z(x_6) = 1550$$

To solve this problem we will follow the ten-step procedure outlined in Algorithm 3.2.1.

Step 1) We will consider a linear semivariogram and a 4 km interval.

$$\gamma(h) = \bar{\omega}\, h \qquad |h| \leq 4$$

Step 2) We will also consider a linear drift.

$$m'(x) = a_1 x$$

with $(x_j - x_{jj}) \leq 4$ km for any x_j, x_{jj}.

Step 3) Our first interval will be (x_1, x_5).

Step 4) The system of equations we must solve has been developed in Appendix B. Our only coefficient B_1 is:

$$B_1 = \frac{Z(x_5) - Z(x_1)}{4}$$

Hence,

$$B_1 = 18.5$$

Step 5) The estimated residuals are:

$$R(x_1) = Z(x_1) - a_0 - B_1 x_1 = 1470.0 - a_0$$
$$R(x_2) = Z(x_2) - a_0 - B_1 x_2 = 1463.5 - a_0$$
$$R(x_3) = Z(x_3) - a_0 - B_1 x_3 = 1483.0 - a_0$$
$$R(x_4) = Z(x_4) - a_0 - B_1 x_4 = 1476.5 - a_0$$
$$R(x_5) = Z(x_5) - a_0 - B_1 x_5 = 1470.0 - a_0$$

Step 6) By Lemma 2.2.2,

$$\gamma_E(0) = 0$$

$$\gamma_E(a) = \frac{1}{2x4} \sum_{j=1}^{4} [R(x_j + a) - R(x_j)]^2$$

$$= 1/8\ (42.25 + 380.25 + 42.25 + 42.25)$$

$$= 63.375$$

$$\gamma_E(2a) = \frac{1}{2x3} \sum_{j=1}^{3} [R(x_j+2a) - R(x_j)]^2$$

$$= 1/6\ (169 + 169 + 169)$$

$$= 84.5$$

$$\gamma_E(3a) = \frac{1}{2x2} \sum_{j=1}^{2} [R(x_j+3a) - R(x_j)]^2$$

$$= 1/4\ (42.25 + 42.25)$$

$$= 21.125$$

$$\gamma_E(4a) = \frac{1}{2} \sum_{j=1}^{1} [R(x_j+4a) - R(x_j)]^2$$

$$= 0$$

Step 7) For p' = 1, we have a new interval (x_2,x_6) which is still part of the sequence. Therefore, we go back to step 4).

Step 4)

$$B_1 = \frac{Z(x_6) - Z(x_2)}{4} = 17.0$$

Step 5) The new residuals are:

$$R(x_2) = Z(x_2) - a_0 - Bx_2 = 1482 - a_0$$
$$R(x_3) = Z(x_3) - a_0 - Bx_3 = 1503 - a_0$$
$$R(x_4) = Z(x_4) - a_0 - Bx_4 = 1498 - a_0$$
$$R(x_5) = Z(x_5) - a_0 - Bx_5 = 1493 - a_0$$
$$R(x_6) = Z(x_6) - a_0 - Bx_6 = 1482 - a_0$$

Step 6) By Lemma 2.2.2,

$$\gamma(0) = 0$$

$$\gamma_E(a) = \frac{1}{2x4} \sum_{j=2}^{5} [R(x_j+a) - R(x_j)]^2$$

$$= 1/8\ (441 + 25 + 25 + 121)$$

$$= 76.5$$

$$\gamma_E(2a) = \frac{1}{2x3} \sum_{j=2}^{4} [R(x_j+2a) - R(x_j)]^2$$

$$= 1/6\ (256 + 100 + 256)$$

$$= 102.0$$

$$\gamma_E(3a) = \frac{1}{2x2} \sum_{j=2}^{3} [R(x_j+3a) - R(x_j)]^2$$

$$= 1/4\ (121 + 441)$$

$$= 140.5$$

$$\gamma_E(4a) = \frac{1}{2} \sum_{j=2}^{2} [R(x_j+4a) - R(x_j)]^2$$

$$= 0$$

Step 7) If we move one step forward up the interval, we need a seventh sample which we do not have. Therefore, we go to step 8).

Step 8) The average semivariogram for those 6 samples is:

h	$\gamma_E^1(h)$
0	0.00
1	69.94
2	93.25
3	80.81
4	0.00

Step 9) Omit this step as we have one sequence.

Step 10) Omit this step as well. There are no semivariograms to average.

END

By the proof presented in Appendix D, the semivariogram of the estimated residuals for a linear drift and linear semivariogram of the real residuals is:

$$\gamma_E(h) = \gamma(h) - \frac{\bar{\omega}h^2}{L}$$

The experimental semivariograms we found are independent of the value $\bar{\omega}$ chosen for the slope of the theoretical semivariogram. In other words, whatever $\bar{\omega}$ we use, we will end up with the same semivariograms for the estimated residuals when the semivariogram is a line through the origin.

However, to remove the bias $\bar{\omega}h^2/h$, the value for the slope is needed. One way to estimate this slope is by the tangent through the origin of the semivariogram for the estimated results. For a linear semivariogram and a linear drift, this is:

$$\bar{\omega}_A = \frac{L}{L-1} \gamma_1^E(1)$$

Hence, for our instance, $\bar{\omega}_A$ is equal to 92.50 [m^2/km].

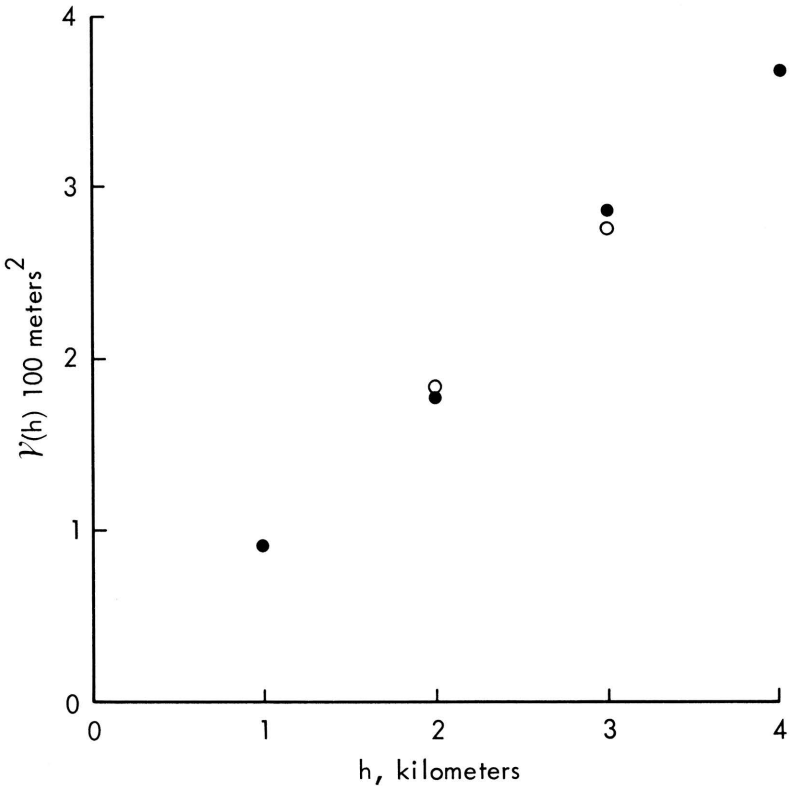


Figure F.1--Comparison between the semivariogram for the estimated residuals with the bias removed (dots) and the assumed linear semivariogram (open circles).

Figure F.1 is a plot showing how close the semivariogram obtained using the observations fits a straight line of slope $\bar{\omega}_A$. Figure F.2 shows the same semivariograms, but here the bias for the semivariogram of the estimated residuals has not been removed and the linear semivariogram has been biased for comparative purposes.

The curves fit fairly well in both graphical displays. Therefore, the choice of a 4 km interval, a linear drift, and a linear semivariogram with a 92.50 [m^2/km] slope is appropriate.

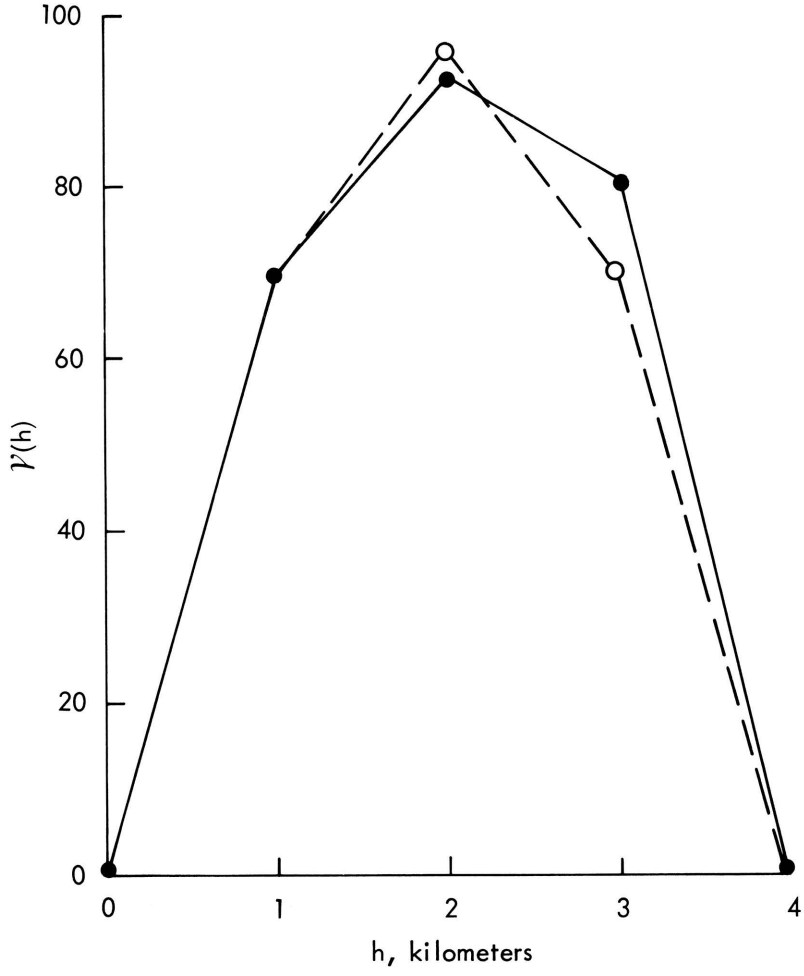


Figure F.2--Comparison between the semivariogram for the estimated residuals (dots) and the assumed linear semivariogram with a bias added (open circles).

APPENDIX G

The Screen Effect

The screen effect in universal kriging refers to the lack of influence of a sample at location $\vec{x}_1$ on the estimation at $\vec{x}_0$ when points $\vec{x}_i$ exist between the two. It is as though the intervening points hide the point $\vec{x}_0$ from point $\vec{x}_1$.

A rigorous proof of the screen effect formulas is not particularly instructive. Instead, we will illustrate the effect with an example. Figures G.1 to G.3 show the weights for 20 points along the sides of a square. Omitted values can be deduced from the symmetry of the arrays. A linear semivariogram was used to calculate the values. The initial 20 samples account for 100 percent of the weightings. In Figure G.2 a single curtain of 12 points immediately reduces the contribution of the original points to 20.8 percent. In Figure G.3 a double curtain of 16 samples further reduces the influence of the original points to only 2.4 percent. In a critical situation, just one sample at the exact location of the point to be kriged is enough to reduce the weights of all other points to zero.

The screen effect is a desirable phenomenon when selecting the optimal number of samples to be used to krige a point. Although theoretically the more samples the better, in practice,

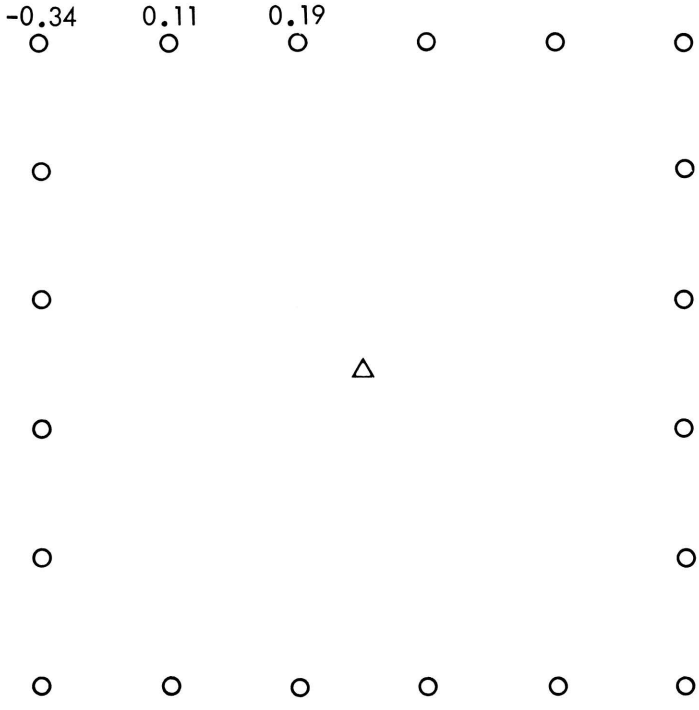


Figure G.1--Weights for universal kriging along the sides of a square. Weighting pattern is symmetrical.

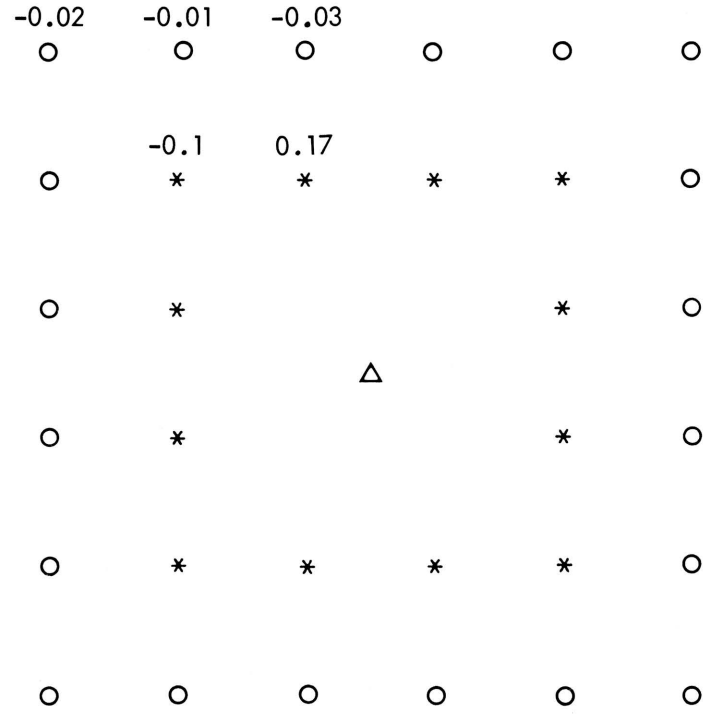


Figure G.2--Screen effect, showing reduction in weights assigned to points on a square by insertion of a curtain of 12 points.

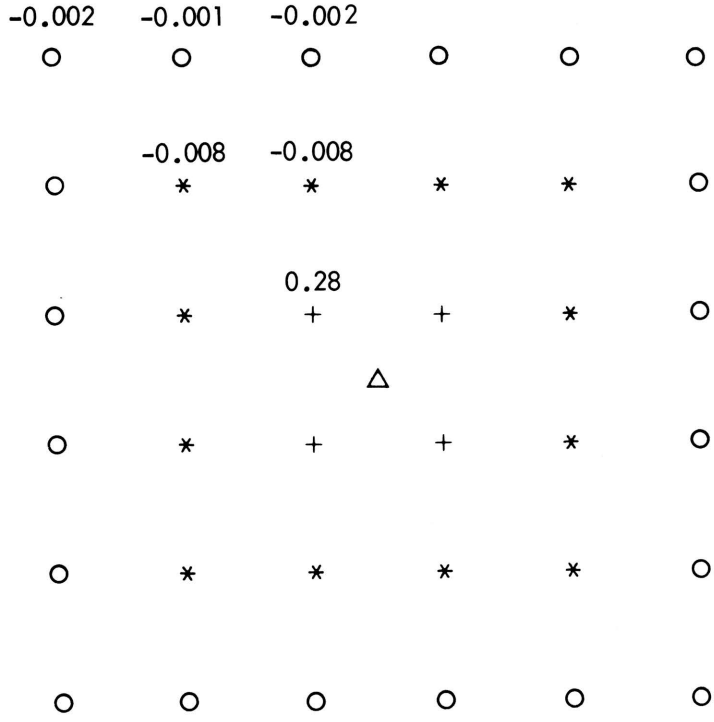
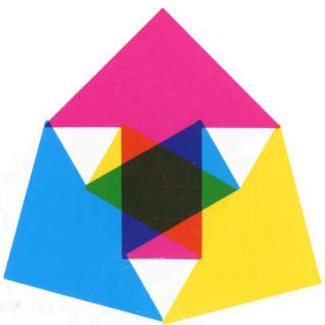


Figure G.3--Screen effect produced by insertion of a double curtain of 16 points.

computational efficiency demands the fewer the better. The screen effect provides a useful criterion for finding an optimal balance. In Figure G.3 the 16 samples nearest the point to be estimated are sufficient to krige that point as they account for more than 97 percent of the total weightings. This represents an average of only two samples per octant.

This conclusion is restricted to a regular and symmetric array of samples. However, experiments with randomly distributed points show that there is no significant improvement in the estimation variance or change in the kriged value for more than two samples per octant. With each doubling of the number of samples, the computation time increases by four times.



NUMBER TWO
SERIES ON SPATIAL ANALYSIS

KANSAS GEOLOGICAL SURVEY
1930 AVENUE A CAMPUS WEST
LAWRENCE, KANSAS U. S. A. 66044