UNIVERSAL KRIGING

Pierre CHAUVET, Alain GALLI

Cours C–96

Septembre 1982

CENTRE DE GÉOSTATISTIQUE

35 rue Saint–Honoré, 77305 FONTAINEBLEAU (France)

ECOLE DES MINES
DE PARIS

CENTRE DE GÉOSTATISTIQUE,
Ecole Nationale Supérieure des Mines de Paris,
35 rue Saint-Honoré, 77305 FONTAINEBLEAU, France

Pierre CHAUVET, Alain GALLI. *Universal kriging*. Ecole d'Eté 1982. 94 p.
C-96

U N I V E R S A L K R I G I N G

Table of Contents

CHAPTER 1 - INTRODUCTION TO UNIVERSAL KRIGING

1 - INTRODUCTION TO UNIVERSAL KRIGING

1.1 PRELIMINARY REMARKS

Universal Kriging (UK) forms the link between stationary geostatistics and the theory of intrinsic random functions (IRF-k): it takes over from the stationary geostatistics when the assumptions needed for this are clearly too severe, but on the other hand, it tries to reformulate the problem so that the techniques already developed for the stationary case can be applied. This is done by splitting the phenomenon under study into two components - one structured and the other random, and by treating these two separately.

This is the natural way to approach the problem. It is exactly what is done in trend surface analysis. If the phenomenon under study seems to be too complicated to be treated as is, it seems natural to decompose it into two "sub-phenomena" with structural properties which operate at different distance scales. But this decomposition is by no means as innocent as it seems at first glance. The mathematical operations which are used in this, almost inevitably have no direct link with the physical reality. Of course, it is possible to imagine cases (e.g. studies on stresses and strains in blocks, or mapping electric fields) where the physics of the problem makes it possible to postulate a particular model a priori ; but most of the time (and this is particularly true of the earth sciences) this is simply not the case. Variables such as the depth of a sedimentary seam, the grade of metal in a deposit or the general behaviour observed in gravimetric studies, cannot be described by deterministic equations. This remark should in no way be taken as condemning attempts to seek simplifications or generalizations. But it is nevertheless vital to keep in mind that formulae - even when rigourously correct - are only the expression of one (possibly arbitrary) way of tackling the problem, and consequently

that it is essential to verify the hypotheses by comparing them with the reality. Here, the reality means the raw data, before any decomposition or interpretation has been done.

Unfortunately, it is often tempting to attribute a physical significance to a model, but this can have disastrous consequences. The danger is all the more serious since the model is generally chosen for its "good" properties from the mathematical point of view, and this can lead to the use of methods having no relation to the reality. It is easy to be carried away by a model which, though mathematically correct, still has nothing in common with the actual phenomenon.

> As an example of this, consider a field of gravimetry anomalies. One "general behaviour" could be defined as being the component of the anomaly due to some well-defined phenomenon (e.g. isostasy, or the presence of a mountain range) which has its own properties (local extrema, gradient, etc..)
>
> Clearly, there is absolutely no way of measuring the general behaviour directly. The only way of finding a representation for it is to choose a model. For example, we could say that the general behaviour can be described by an nth degree surface. But then we do not have the right to say that the local extrema (or the gradient) of this surface is that of the general behaviour. Nor should we say that the general behaviour is an nth degree surface.

The worst type of confusion between the model and the reality is to try to give a physical explanation for each of the parameters of the model. Some typical examples of this are "the general behaviour is an nth degree surface because" or "the mineralization has a particular number of nested structures because..." which are then followed by a naturalistic explanation. People who use this type of reasoning forget that the parameters of the model (degree of the function, number of elementary variogram models...) are the result of a choice and that another equally well qualified person

could come to quite different conclusions. However the reality always remains the same.

> In any case, the model should not be identified with the reality. The grades in a deposit were never allocated at random according to a probability distribution. Nor was the general behaviour developed by evaluating a polynomial expression (no matter what the degree) etc.. The use of a particular formalism (e.g. of random functions in geostatistics) is a constitutive choice which is neither true nor false of itself. Moreover the appropriateness of the choice can only be judged by the comparison with the experimental results.

One interesting but also difficult aspect of Universal Kriging (U.K.) is to explain these points to the user. Even though the approach taken is the most natural, several sources of ambiguity and indeterminacy arise. Although these may prove to be instructive in coming to an understanding of non-stationary phenomena, they also pose certain intractable technical problems.

The presentation of U.K. and later of the generalized intrinsic random functions (GIRF) has two principal objectives :

a) To analyze the inherent ambiguities, to see how they arise and what they mean, and to determine their consequences from the practical point of view.

b) To highlight the relationship between U.K. and GIRF, to show how the theory of the GIRF overcomes the difficulties encountered in U.K. and finally to give an intuitive explanation of this new mathematical tool.

More emphasis will be placed on understanding the physical significance of the problems and on interpreting the results than on the details of the proofs.

1.2 EMPIRICAL STATIONARITY AND THE MODEL OF STATIONARITY

1.2.1 Stationarity from the empirical point of view.

At first, the idea of a stationary phenomenon may seem intuitively clear. Broadly speaking, a variable will be considered to be stationary if "it behaves in the same way throughout the whole of the region under consideration" ; that is, if it acts as if there is some sort of force which always pulls the variable back to its average value. Numerically, a stationary phenomenon can be said to be "self-regulating" (this should not be taken as implying or denying the existence of a physical self-regulating mechanism).

This apparently simple notion turns out to be fraught with difficulties...

- This empirical definition can only be true on average. Except for a variable which always takes the same value, no phenomenon ever respects this definition strictly. This behaviour can only be "more or less" verified in practice.
- To be more precise, to test the stationarity of a variable, we must first choose a certain zone of investigation (with a fixed size and shape) which can subsequently be moved throughout the whole of the region under study. For any given position, the investigation zone will contain a certain number of sample points. If, for each such position, the statistical characteristics of these points are the same, the variable can be said to be stationary.
- It is clear that stationarity is by no means absolute - not even in the most intuitive sense. It depends very much on the investigation zone (or sliding neighbourhood, as it is called), and is,

to a large extent, arbitrary. In particular, the size selected for the sliding neighbourhood determines the scale at which the variable can be said to be stationary.

- The property of stationarity is therefore fundamentally dependent on the scale considered, and it would be a serious methodological error to try to extrapolate stationarity to different distance scales (either larger or smaller) than the one for which it was verified.

- Definition : We will describe a property as being local if it is true at the scale of the sliding neighbourhood, and global if it holds for the whole of the region under study.

Simplistic Example.

We shall now study a one-dimensional case : the function $y = \sin x$ where x ranges from 0 to (say) 200π. Can y be considered to be stationary in this interval ?

Fig. 1

In its present form the question is meaningless because the scale has not been defined.

. If the length of the sliding neighbourhood is a multiple of 2π, the statistics calculated within these intervals will all be the same, and the stationarity will be verified exactly.

. As the length of the neighbourhood increases, the difference between the statistics lessens and so it would be possible to accept the hypothesis of stationarity.

. For neighbourhoods of length π, the statistics vary considerably, and the hypothesis of stationarity could not be accepted. In any one of these intervals the curve could be concave or convex, or could have a point of inflexion. It could quite well be represented by a 3rd degree polynomial.

- If the interval becomes much smaller (e.g. $\pi/20$, the curve can still not be considered as stationary, but this time can be represented locally by a straight line.

It should be remembered that in practice, we do not need complete stationarity. That is, the histograms of values within these neighbourhoods do not have to be identical. All that is required is that the first two moments (i.e. the local mean and the dispersion around this local mean) are constant.

1.2.2 The Theoretical point of view. Notation.

The constitutive choice made in geostatistics is that of considering the phenomenon as a realization of a random process (See [1]). The definition of second order stationarity (i.e. the stationarity of the first two moments) does not pose any theoretical problems.

We now define the notation that will be used subsequently.

- We shall use small letters to denote deterministic variables - in particular, experimental values. For example, $z(x)$ represents the value of the function z measured at point x.
- Capital letters will be used to denote random variables and random functions. These quantities exist only in the model and not in reality. For example, $Z(x)$ is a random function used in the model to describe the phenomenon observed.
- Most of the time the spatial coordinates will be written as if for a 1-D space : e.g. $z(x)$. It is not difficult to mentally replace x by a vector when dealing with 2 or 3-D spaces.

. Finally the usual statistical notation will be used to denote the model :

$E[Z(x)]$: the mathematical expectation (mean) of the R.V. Z at the point x.

$\left.\begin{matrix} Var[Z(x)] \\ D^2[Z(x)] \end{matrix}\right\}$: the variance of the R.V. Z at point x

$Cov\,[Z(x),Z(y)]$: the centred covariance between the values of Z at points x and y.

The usual definition of second order stationarity can now be given :

1.2.2.1 The 1st order moment.

If Z(x) is stationary then $E[Z(x)]$ is invariant under translation, that is, it has the same value throughout the space.

1.2.2.2 The 2nd order moment.

To say that $Cov(Z(x),Z(y))$ is invariant under translation means that it depends only on the relative position of x and y and not on their particular location in space. It is, therefore, a function of only one variable $h = y-x$ and is usually denoted by $C(h)$.

1.2.3 Link between the two points of view.

Stationary geostatistics should be used only after the hypothesis of stationarity has been verified. It is therefore essential to check whether the experimental data can reasonably be considered as being a realization of a random function satisfying the conditions given in the preceding paragraph.

In practice, we have only one set of data (that is, only one realization of the random function that we are trying to model), so it is not possible to develop a model of the R.F. without making additional assumptions - no more than it would be possible to find a probabilistic model for roulette given the result of only one toss.

The additional assumption made is the following : we decide to equate empirical stationarity with theoretical stationarity. So we consider that the hypothesis of stationarity can be accepted if the experimental data are statistically stationary. Without going into the technical details, it should be noted that this choice should be tested subsequently whenever possible.

- In fact it does not seem reasonable to demand more from the model than can be observed in the data. Conversely, it is easy to imagine cases (such as heterogenous data) where the data are empirically stationary (second order) but where the use of a stationary geostatistical model would be quite inappropriate.
- In order to test the empirical stationarity, we have to test that the values calculated over the sliding neighbourhood do not fluctuate "too much" (see § 2-1). To be quite rigourous, the permissible limits for these fluctuations can only be calculated once the model has been fixed. So we are obliged to be satisfied with approximations. The best we can do is to check that the final model is internally consistent.

1.2.4 Conclusion.

It does not seem to be possible to propose a model at the beginning of a study, which will be valid until the end.

Even in the most favourable cases a model must be "alive" and keep changing to fit reality.

- If the model becomes internally inconsistent or is not in agreement with the data, it must be modified.
- If this not the case, we can go on to the next step in the study and to additional checks.

1.3 MODELLING NON-STATIONARY PHENOMENA.

1.3.1 Examples of non-stationary phenomena.

The hypothesis of stationarity is a very strong one and as such is often at variance with the facts.

1.3.1.1 Non-stationary first order moment.

The most common cause of non-stationarity is the presence of a "drift" or trend (i.e. when the mean of the phenomenon observed in the sliding neighbourhoods is not constant). For example consider the following sequence of pollution data which were taken in an urban area over a period of 2 1/2 years. They show an alternance of high values (in winter) followed by low ones (in summer).

Fig. 2

It is clear that a stationary model could not account for this sequence of highs and lows. In fact, an overall drift corresponding to the seasonal fluctuations is apparent. But we can still consider (as was done in the stationary case) that there is a "self-regulating" force which tends to keep the variable close to the drift. Looked at from this point of view, we may be tempted to try to split the raw phenomenon into two components so as to bring the problem back to the stationary case

(at least partially). This idea, which seems natural enough, mixes two radically different requirements :

- On one hand, the seasonal component or drift seems to correspond to a real phenomenon. It would be nice to attribute a physical significance to the "fluctuations" (i.e. the differences between the measured value and the drift). This is a naturalistic sort of requirement.

- On the other hand, to be able to use stationary geostatistics, the fluctuations have to be stationary (of second order) or intrinsic. In contrast to the preceding one, this is a methodological requirement.

There is no reason why these two viewpoints have to converge. In order to be able to use this method, the second requirement is the more important but in this case, it is essential to avoid extrapolating the results obtained without making additional checks, and also to avoid attributing a physical significance to them. A critical review of the results is indispensable when using these mathematical methods.

1.3.1.2 Non-Stationary second order moment.

This cause of non-stationarity is not as easy to visualize as the preceding one was. The figure below - which, as it happens, is the curve of daily increments from the preceding example - is fairly level overall. That is, the first moment is empirically stationary.

Fig. 3

However, the fluctuations around this average level vary with time. We know that this variability will not affect the type of estimator used, although it certainly affects their

quality (i.e. the estimation variance). Depending upon the type of problem to be solved, we could use methods similar to those employed in ordinary geostatistics (e.g. related to the proportional effect) or else develop some more specialized techniques (e.g. related to anamorphosis).

1.3.1.3 No second order moment at all.

This problem led to the development of the intrinsic hypothesis ([1]). Although it may seem surprising at first, it is possible to find sets of experimental values which appear to have an "infinite variance". Of course in practice we only have a finite number of values in the sliding neighbourhoods and their variance is always finite.

But the experimental variance sometimes increases without ever reaching a limit, as the size of the sliding neighbourhoods is increased. In physical terms this means that the variable has no self-regulating mechanism. Under those circumstances it would be foolish to try to use the experimental variance in a stationary model. The results would be incompatible with reality.

It is worth noting that the only way of revealing this type of non-stationarity is in "dynamic" studies on the data (i.e. ones where the working scale is changed). The user can test for it by comparing the experimental variogram with the experimental variance.

1.3.2 Modelling the drift.

1.3.2.1 A simplistic example.

Suppose that we want to use least squares to fit a polynomial curve to the 4 sample values shown below :

Fig. 4

If we decide to fit a linear curve, the "solution" obtained is :

Fig. 5

Fitting a parabola gives us :

Fig. 6

Note that this parabola is quite close to the straight line shown in Fig. 5 in the zone considered. If for some reason we decided to ignore one of the two points on the left when fitting the parabola, we would get one of the two parabolas shown in Fig. 7 :

Fig. 7

The results obtained when using these curves to extrapolate would clearly diverge rapidly.

Lastly, fitting a cubic curve gives quite different results :

Fig. 8

1.3.2.2 Which approach is correct ?

Before going into this question, a few preliminary comments must be made. Undoubtedly no one would hesitate in rejecting the curve shown in Figure 8 since it is quite unrealistic. However from a mathematical point of view it is just as correct as the others (and incidentally is statistically a "better" fit). In rejecting this curve we are introducing a naturalistic argument into the method.

In the introductory remarks, we noted how dangerous this can be. It is not difficult to think of cases where there are physical reasons for requiring the trend to be monotonic in the zone considered, which would effectively exclude Figure 8. It is also possible to imagine cases where one or the other of the two sample points on the left would be suspect - and this would eliminate one of the parabolas on Figure 7. But in eliminating curves in this way we are slowly letting naturalistic considerations slide in between the fitted curves(which are the end product of a mathematical method, and as such rigourously correct) and the phenomenon which they are supposed to represent.

1.3.2.3 General Hypotheses of Universal Kriging (U.K.)

The correct way of attacking the problem is to develop a coherent and viable mathematical model and at least initially, to consider the results that come from it as the end product of a mathematical operation. Only after comparing the results with the facts can one start to attribute a physical meaning to them.

To start with, the drift of R.F. $Z(x)$ is defined as the expectation :

$$m(x) = E[Z(x)] \qquad (1.1)$$

In contrast to ordinary (stationary) geostatistics, this function $m(x)$ will not be assumed to be stationary

(i.e. constant). Consequently it is really a function of the location $\underset{\sim}{x}$ in space.

The covariance of Z(x) is defined as

$$C_{xy} = \mathrm{Cov}[Z(x), Z(y)]$$
$$= E[Z(x)\ Z(y)] - m(x)\ m(y) \qquad (1.2)$$

To be of any use, universal kriging must be able to solve the following two problems :

a) to estimate the drift (i.e. the function m(x)) at any given point.

b) to estimate the value z(x) of the variable at a particular point (or alternatively a linear transformation of z(x)).

As in ordinary geostatistics, the idea is to use linear estimators which will then have a corresponding estimation variance. The optimality criterion chosen is that of minimum estimation variance.

1.3.2.4 Simplifying the hypotheses.

In its present form, the problem of U.K. is insoluble because it is too vague. The definition of the drift is quite clear, but it still does not allow us to develop a workable method. No statistical estimates can be obtained until some additional hypotheses are made.

These hypotheses follow from the ideas expressed in paragraph 2.1. The drift will be constrained to be fairly regular and continuous at the scale considered. By doing this, we are anticipating the possibility of equating this drift to "general behaviour" or "trend" ; but for the moment, the constraint is purely mathematical. The idea behind it is that the drift m(x) can be

represented locally by the first few terms of its development in terms of the basis functions f_x^ℓ . Let

$$m(x) = a_\ell \, f_x^\ell \qquad (\ell = 0,1,\ldots,k) \tag{1.3}$$

The number k of functions is known and so are the functions themselves but the coefficients a_ℓ (which are different in different neighbourhoods) are not. So these will have to be estimated.

Summarizing, we consider the reality z(x) as a realization of a non-stationary random function Z(x) having the following properties :

$$\left.\begin{aligned} Z(x) &= Y(x) + a_\ell \, f_x^\ell \\ E[Z(x)] &= a_\ell \, f_x^\ell = m(x) \\ Cov[Z(x),Z(y)] &= E[Y(x)\ Y(y)] = K_{xy} \end{aligned}\right\} \tag{1.4}$$

1.3.3. The Intrinsic Model.

1.3.3.1 Review of the Intrinsic Hypothesis.

In section 1.3.1.3 we mentioned that the intrinsic hypothesis had originally been developed to deal with cases where the second order moment did not exist (See [1], 2.1). A R.F. Z(x) is said to be intrinsic if its increments Z(x+h) - Z(x) are weakly stationary (i.e. of order 2). This hypothesis is much less restrictive than stationarity, and allows us to apply the methods developed in ordinary geostatistics to a much wider range of phenomena.

The decomposition of Z(x) given in (1.4) is more closely linked to this than it seems at first. We will show that the intrinsic model corresponds to a linear drift. The most striking example of this is the random walk : - a fair coin is

tossed and the outcome of the ith toss is either $X_i = +1$ or $X_i = -1$. The cumulative sum up to the nth toss, $S_n = X_1 + \ldots + X_n$ has a mean of 0. We also know that this R.F. satisfies the intrinsic hypothesis. Even though the mathematical expectation is 0, actual realizations of this always have a local linear drift. (cf. Delfiner [2]).

This shows the close links that exist between these two types on non-stationarity. In fact, in practice it is impossible to tell whether a drift observed in the data ought to be considered as coming from a deterministic model or a random one. So it is hardly surprising to find that one method can be used to treat the two types of non-stationarity. However, as far as the model is concerned the user has to decide whether the drift is to be considered as being random or deterministic. Consequently we should not be surprised to find that this one method leads to a certain indeterminacy.

1.3.3.2 The meaning of the intrinsic hypothesis.

There is an important difference between the intrinsic model and the stationary one : in the intrinsic model, the object under study is the increments of the R.V. rather than the R.V. itself. Of course from a pratical point of view, our attention is still centred on the R.V. $z(x)$ but from a mathematical standpoint, the random function $Z(x)$ is completely forgotten.

Since the only link between the model and the reality is through the increments, the model is more general. The variogram is consequently more general than the covariance and can therefore take account of a wider range of phenomena. But, on the other hand, the operational methods which can be used are more restricted, particularly in so far as the type of estimator is concerned. This approach makes no distinction between two R.F. with the same increments (i.e. differing by a constant). The whole question is to know whether this indeterminacy is troublesome in practice.

It is intuitively clear that the relation between a R.F. and its increments is the same as that between a function and its derivative. Using the intrinsic model to study a phenomenon is like studying the derivative of a curve. This introduces an indeterminacy which is similar to the additive constant seen in integration.

1.3.3.3 Admissible linear combinations.

Since it is more convenient to express the model in terms of the R.F. Z(x) rather than its increments, we reformulate the method accordingly. In the stationary case, we had the right to work with any linear combination $\lambda^i Z_i$. (This is the shortened way of writing $\sum \lambda^i Z(x_i)$). This time, to ensure that the variance is finite we have to restrict ourselves to linear combinations of increments ; that is, to linear combinations $\lambda^i Z_i$ where $\sum \lambda_i = 0$. Linear combinations of this type are said to be admissible. This also highlights the manner in which the field of possible estimators has been restricted in going from the stationary model to the intrinsic one. As a compensation, the range of possible applications is much wider because the admissible linear combinations (A.L.C.) effectively filter out the constants.

It is well known in ordinary geostatistics [1] that the expression for the variance of an A.L.C. $\lambda^i Z_i$ can be obtained by replacing the covariance K in the formula for the stationary case by the variogram $-\gamma$. This gives

$$\left.\begin{aligned} &\sum \lambda_i = 0 \\ &D^2[\lambda^i Z_i] = -\lambda^i \lambda^j \gamma_{ij} \end{aligned}\right\} \qquad (1.5)$$

This formula is very important because it simplifies the calculations considerably. In addition to this, it is interesting since it highlights the similarity between the function K which operates on the space of all linear combinations and the

function $-\gamma$ which operates on the subspace of admissible linear combinations. This analogy arises again in the presentation of Generalized Intrinsic Random Functions.

1.4 ORGANIZATION OF THE FOLLOWING CHAPTERS

CHAPTER 2 : Case of an Underlying Stationary Model.

In this Chapter the equations for U.K. are developed for the case where the covariance of $Z(x)$ exists, without dwelling on the proofs. Special attention is focused on the following topics :

- Point kriging
- The estimation of the drift and of its coefficients
- The properties of the U.K. estimators.

CHAPTER 3 : The Case of an Underlying Intrinsic Model.

The additional problems encountered when the covariance does not exist are treated in this Chapter, without going back over the development of the equations.

CHAPTER 4 : The Indeterminacy of Universal Kriging.

This Chapter considers the meaning of the indeterminacy and then goes on to see to what extent this is irreducible and how it affects the structural analysis.

A theoretical presentation of U.K. is given in [1].

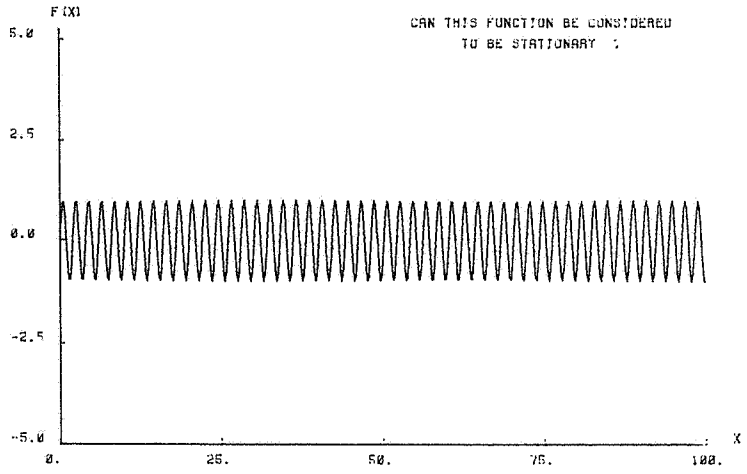


FIG. 1

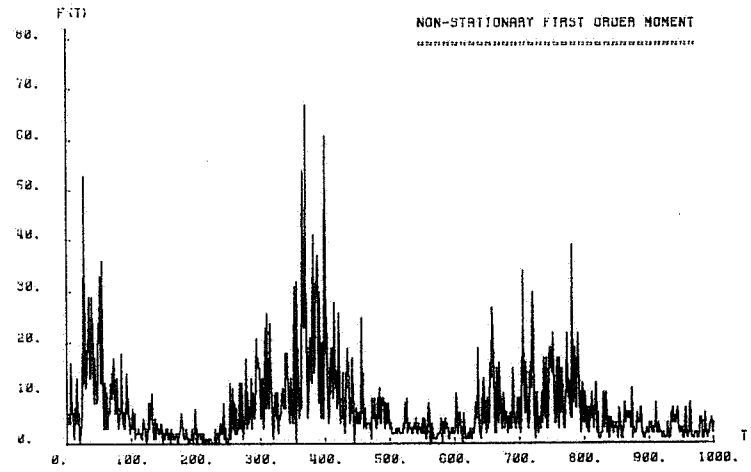


FIG. 2

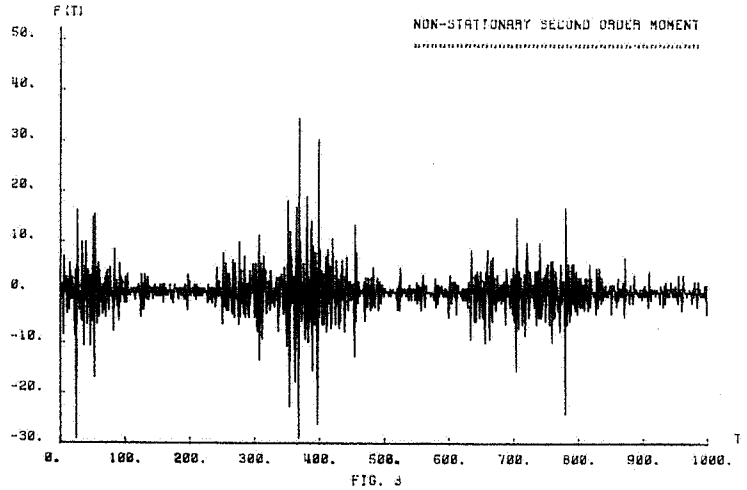


FIG. 3

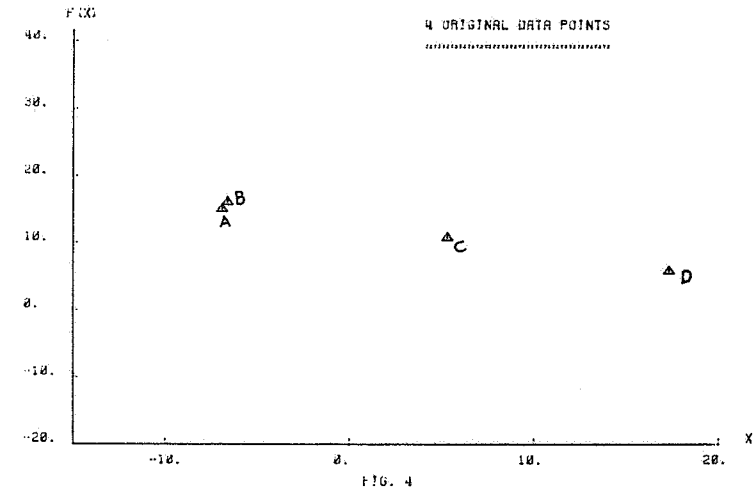


FIG. 4

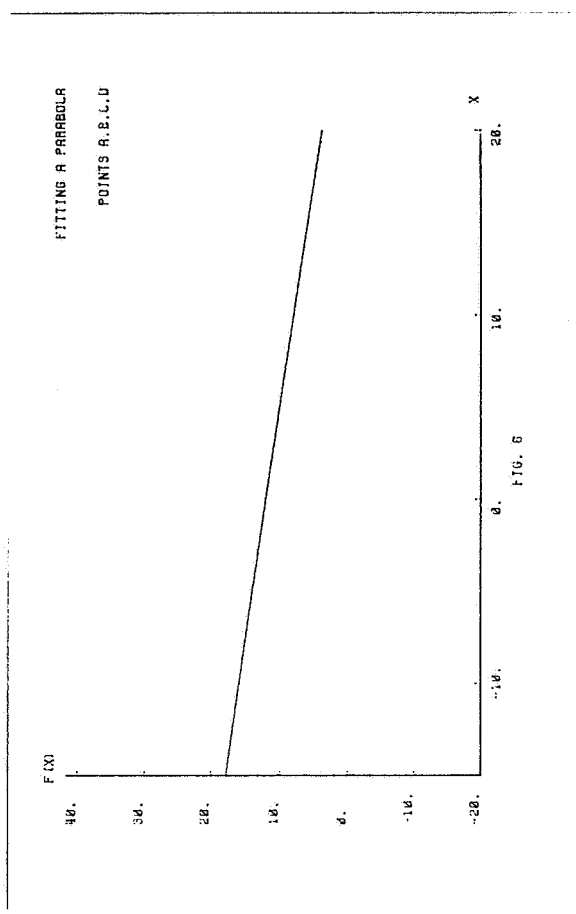


FIG. 6

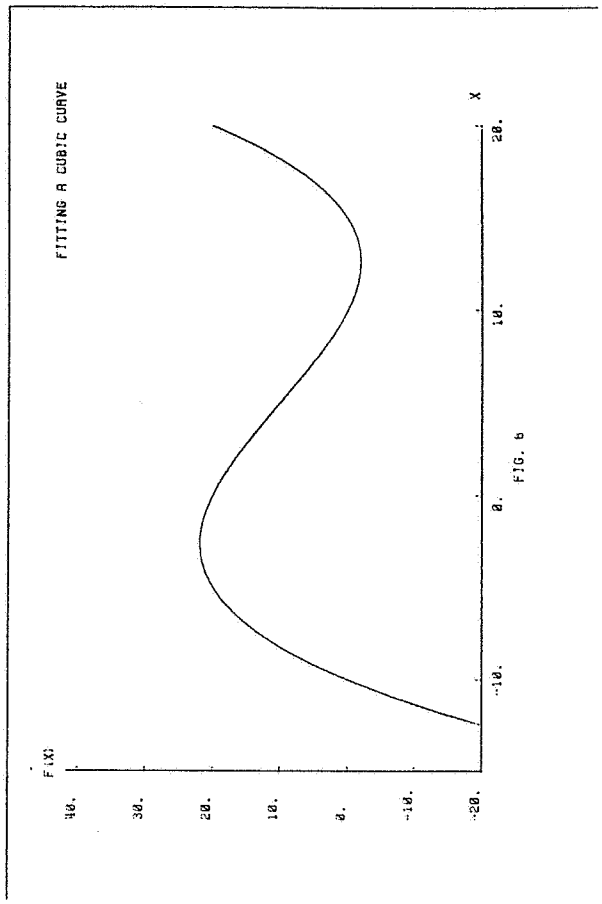


FIG. 6

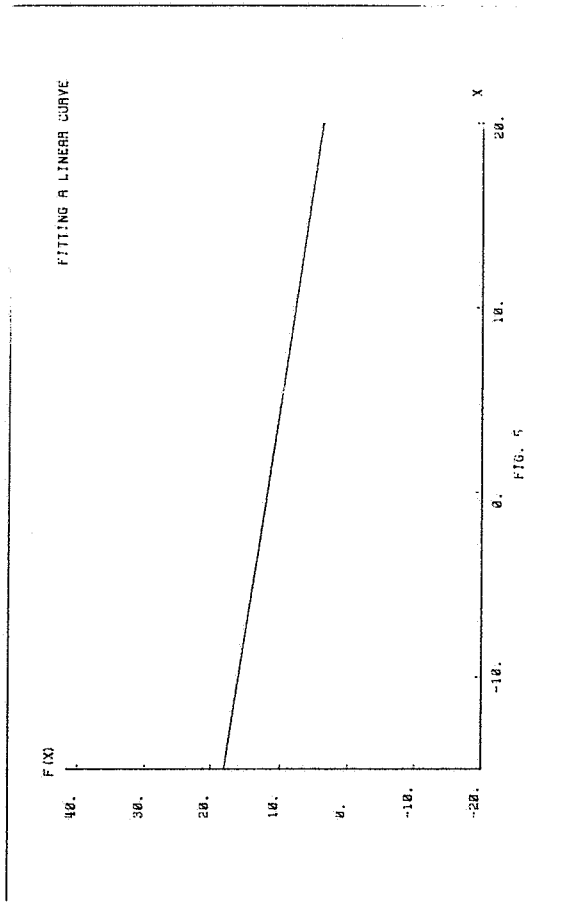


FIG. 5

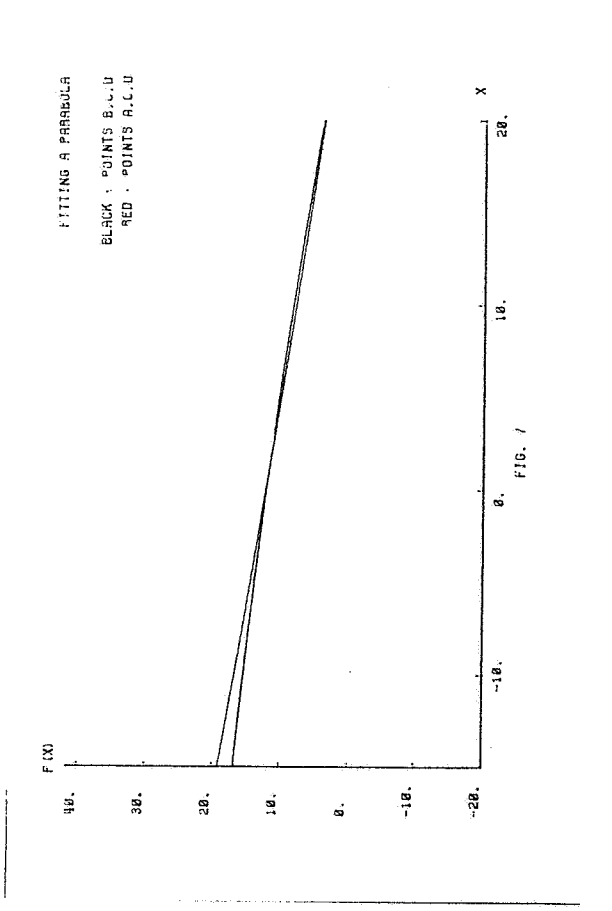


FIG. 7

CHAPTER 2 - CASE WHERE THE UNDERLYING MODEL IS STATIONARY

2 - CASE WHERE THE UNDERLYING MODEL IS STATIONARY

2.1 GENERAL PRESENTATION

In this chapter we assume that the conditions described in 1.4 are satisfied ; that is, the behaviour of the R.F. Z(x) can be split into two components :

$$Z(x) = Y(x) + a_\ell f_x^\ell \qquad (2.1)$$

where Y(x) is a R.F. with mean 0 and covariance K.

$$E(Y(x)) = 0 \qquad (2.2)$$

$$E[Y(x)\ Y(y)] = K_{xy} \qquad (2.3)$$

The mean E[Z(x)] is, by definition, called the drift of Z. We suppose that it can be expressed as

$$E[Z(x)] = a_\ell f_x^\ell \qquad (2.4)$$

where the f_x are the basis functions (for $\ell = 0$ to k fixed) and where the a_ℓ are unknown numerical coefficients.

Once the decomposition given in (2.1) has been accepted, two types of problems may be treated :

- the estimation of values of z(x) at certain points, or of average values. In this case, the problem is one of kriging in the strict sense of the word.
- The estimation of the characteristics of the drift. Here it is important to distinguish the estimation of the drift itself from that of its coefficients a_ℓ .

In any case, it should be remembered that all linear combinations have a finite variance and that there is therefore no

need, a priori, to impose any conditions on the λ^i. The covariance function K_{xy} will be called the underlying covariance model, and will be assumed to be known and to be stationary. For the time being, the question of knowing how this model has been determined will be kept in abeyance.

> It should be noted that the U.K. equations do not necessarily require K_{xy} to be stationary. But if this were not the case, any statistical inference concerning K would be impossible. Since this is unrealistic, K will be assumed to be stationary.

2.2 KRIGING.

2.2.1 The Equations for U.K. of point values.

2.2.1.1 The problem.

We have a certain number of experimental values of z recorded at the points α ($\alpha = 1,\ldots,n$), at our disposal. The problem is to estimate the value z_x at a point x using a linear combination $\lambda^\alpha z_\alpha$ of the data values. The estimator must satisfy two conditions :

1) it must be "universal", i.e. it must be unbiased no matter what the drift is.
2) it must be optimal, in the sense that the estimation variance is to be minimized.

So, in the model, we are looking for a set of coefficients λ^α such that

$$\lambda^\alpha Z_\alpha = Z_x^* \qquad \text{(estimator of } z_x\text{)} \qquad (2.5)$$

$$E[Z_x^*] = E[Z_x] \qquad \text{(universal)} \qquad (2.6)$$

$$D^2[Z_x^* - Z_x] \quad \text{minimum} \quad \text{(optimal)} \qquad (2.7)$$

2.2.1.2 Universality Conditions.

Replacing Z_x^* in equation (2.6) by its expression (2.5) and then expanding the expressions for the expectations of Z_x^* and Z_x using (2.4) gives :

$$\lambda^\alpha a_\ell f_\alpha^\ell = a_\ell f_x^\ell$$

or

$$a_\ell (\lambda^\alpha f_\alpha^\ell - f_x^\ell) = 0$$

This equation must be satisfied for all possible values of the (unknown) coefficients a_ℓ. Since any of the a_ℓ coefficients must be 0, we obtain the following set U of universality conditions :

$$\lambda^\alpha f_\alpha^\ell = f_x^\ell \qquad [U] \qquad (2.8)$$

which must hold for all $\ell = 0,1,\ldots,k$.

NOTE : The following equations hold for any linear combinations $\lambda^\alpha Z_\alpha$ which satisfy the universality conditions U :

$$\begin{aligned} Z_x - \lambda^\alpha Z_\alpha \qquad & \text{(estimation error)} \\ &= Y_x + a_\ell f_x^\ell - \lambda^\alpha (Y_\alpha + a_\ell f_x^\ell) \qquad \text{from (2.1)} \\ &= Y_x - \lambda^\alpha Y_\alpha + a_\ell (f_x^\ell - \lambda^\alpha f_\alpha^\ell) \\ &\equiv Y_x - \lambda^\alpha Y_\alpha \qquad \text{(from the condition U)} \end{aligned}$$

So for any universal estimator $\lambda^\alpha Z_\alpha$ (i.e. one satisfying U) we have

$$D^2[Z_x - \lambda^\alpha Z_\alpha] = D^2[Y_x - \lambda^\alpha Y_\alpha] = \overline{E[Y_x - \lambda^\alpha Y_x]^2}$$

since the mean of Y_x and Y_α are, by definition, 0.

2.2.1.3 Optimality Conditions.

From the preceding comment and also from the definition of the covariance (2.3), it is clear that the variance of the estimation error $Z_x^* - Z_x$ is in the form :

$$D^2[Z_x^* - Z_x] = \lambda^\alpha \lambda^\beta K_{\alpha\beta} - 2 \lambda^\alpha K_{\alpha x} + K_{xx} \qquad (2.9)$$

In order to minimize this quadratic form subject to the constraints (2.8), we introduce a series of Lagrange multipliers μ_ℓ (each associated with one of the universality conditions). We obtain the optimum of (2.9) under these constraints by minimizing the expression

$$(\lambda^\alpha \lambda^\beta K_{\alpha\beta} - 2 \lambda^\alpha K_{\alpha x} + K_{xx}) + 2 \mu_\ell (\lambda^\alpha f_\alpha^\ell - f_x^\ell) \qquad (2.10)$$

where the λ^α and the μ_ℓ are all unknown.

Setting the partial derivatives with respect to μ_ℓ to zero gives us back the universality consitions. Setting the partial derivatives with respect to λ^α to zero gives us the optimality conditions :

$$\lambda^\beta K_{\alpha\beta} + \mu_\ell f_\alpha^\ell = K_{\alpha x} \qquad [O] \qquad (2.11)$$

which must hold for all $\alpha = 1,2,\ldots,n$.

2.2.1.4 The U.K. System.

The whole U.K. system is obtained by grouping [U] and [O] :

$$\begin{aligned} \lambda^\beta K_{\alpha\beta} + \mu_\ell f_\alpha^\ell &= K_{\alpha x} \qquad [O] \\ \lambda^\alpha f_\alpha^\ell &= f_x^\ell \qquad [U] \end{aligned} \qquad (2.12)$$

This can be written in matrix form as :

$$\left[\begin{array}{c|c} K_{\alpha\beta} & f_\alpha^\ell \\ \hline f_\beta^\ell & 0 \end{array}\right] \left[\begin{array}{c} \lambda^\beta \\ \hline \mu_\ell \end{array}\right] = \left[\begin{array}{c} K_{\alpha x} \\ \hline f_x^\ell \end{array}\right]$$

When the solution of this system is substituted back into the expression for the estimation variance (2.9), we obtain the kriging variance :

$$D^2[Z_x^* - Z_x] = K_{xx} - \lambda^\alpha K_{\alpha x} - \mu_\ell f_x^\ell \qquad (2.13)$$

If we assume that the covariance matrix $K_{\alpha\beta}$ is always strictly positive definite, we exclude the case of repeated sampling at the same point since, in this case, two of the rows in $K_{\alpha\beta}$ would be identical. If this condition is satisfied it can be shown that the system (2.12) is non singular, and therefore has a unique solution if and only if the basis functions are linearly independent on the set of data points.

Mathematically speaking, this condition can be expressed as follows :

"If there exist a set of coefficients c_ℓ ($\ell = 0, 1, \ldots k$) such that $c_\ell f_\alpha^\ell = 0$ for any sample point α, then the c_ℓ must all be zero".

As an example of linearly dependent functions, consider the following family :

$$1, \cos\alpha, \sin\alpha, \cos^2\alpha, \sin^2\alpha$$

It is easy to find a non-zero set of coefficients (e.g. in order 1, 0, 0, -1, -1) for which the linear combination is identically zero.

This example is extreme because the basis functions are dependent no matter what sample points are chosen. If

we take the monomials x^n as basis functions in $\mathbb{R}^2$, these will be dependent if the sample points are in a straight line and if the degree of the drift is greater than or equal to 1. This can quite easily arise in practice ; for example if seismic data happen to lie on straight lines.

If the drift is quadratic, the problem of dependence arises if the data lie on a circle, or along two straight lines, etc.. Intuitively this means that a straight line is not enough to define a plane ; that a circle (or two straight lines) is not sufficient for a quadratic

It is easy to see that the kriging system can become degenerate if the data are unsuitably laid out relative to the model proposed for the drift. For example, in 2 dimensions, it is possible to determine a linear drift (and to solve the system) if the data lie on two intersecting lines, but this would not be sufficient if the drift is quadratic.

Let us suppose that this minor difficulty has been overcome. Under these circumstances, once a solution to the kriging system is obtained, it must be unique and therefore is the right solution. This can be used to show that kriging is an <u>exact interpolator</u>. If a point to be estimated coincides with one of the sample points α, the values $\lambda^\alpha = 1$, $\lambda^\beta = 0$ $(\beta \neq \alpha)$ and $\mu = 0$ is a solution to the system and therefore <u>the</u> solution. The best estimator of a known Z_α is then Z_α itself (which is reassuring). It is easy to show that the estimation variance in this case is zero.

2.2.2 <u>Universal Kriging of a Weighted Average</u>.

This case presents no additional difficulties to those encountered in the preceding section (i.e. kriging point values) and will therefore not be discussed separately in future chapters.

The problem here is to estimate a quantity of the form $\int p(dx)\ Z(x)$ where $p(dx)$ is a weighted function and $\int p(dx) = 1$.

The problem can be solved either directly from the definition or alternatively by invoking the linearity of the formulae given in the preceding section. Either way there are no problems deriving the formulae :

$$\left[\int p(dx)\ Z(x)\right]^* = \lambda^\alpha\ Z_\alpha$$

with
$$\begin{cases} \lambda^\beta\ K_{\alpha\beta} + \mu\ f_\alpha = \int p(dx)\ K_{\alpha x} \\ \lambda^\alpha\ f_\alpha = \int p(dx)\ f_x \end{cases}$$

and the kriging variance is

$$\iint p(dx)\ p(dy) K_{xy} - \lambda^\alpha \int p(dx)\ K_{\alpha x} - \mu \int p(dx)\ f_x$$

Once the weighting function p has been chosen, this formalism is quite unambiguous. However p must be chosen carefully. In mining geostatistics it is always a simple average over a volume and its meaning in physical terms is quite clear. In fields like contour mapping, geophysics, meteorology etc.. the situation may be much more complex and p might have to represent a convolution, for example, or even a linear operator such as a gradient or a "Laplacian". In any case, putting this formalism into operation does not introduce any additional indeterminacy, and the final results will be as meaningful as the function p is. Compared to what happens when the drift is being estimated, this situation is very healthy.

2.3 OPTIMAL ESTIMATION OF THE DRIFT.

2.3.1 The problem.

It is important to remember that the drift has been defined in the model as the mathematical expectation of the

function Z(x). We have also assumed that the drift can be represented locally by a function of the form :

$$m(x) = E[Z(x)] = a_\ell \, f_x^\ell \qquad (2.4)$$

Giving this definition a physical meaning is a dangerous "extrapolation" which can lead to serious errors in the interpretation of the results. Consequently, in contrast to the estimation of weighted averages, we shall not consider that the estimation of the drift has any physical significance. However, it is important from the methodological point of view. In the preceding paragraph, the covariance K_{xy} was assumed to be known. In contrast to what happens in stationary geostatistics, the covariance function which is by definition

$$K_{xy} = E[Y(x) \; Y(y)]$$

has to be calculated from the residuals and not from the raw data values themselves.

We can now go ahead to develop a method for estimating the drift. As usual, we are looking for the weighting factors λ^α of the linear combination which minimizes the estimation variance, that is, such that

$$\lambda^\alpha \, Z_\alpha = M_x^* \qquad \text{(estimator of } m_x\text{)} \qquad (2.14)$$

$$E[M_x^*] = m_x \qquad \text{(universality)} \qquad (2.15)$$

$$\text{minimize } D^2[M_x^* - m_x] \qquad \text{(optimality)} \qquad (2.16)$$

2.3.2 Universality Conditions.

In equation (2.15), we replace M_x^* by the expression (2.14), then m_x and the terms m_α using formula (2.4). This gives us

$$\lambda^\alpha \, a_\ell \, f_\alpha^\ell = a_\ell \, f_x^\ell$$

which leads to the same system of universality conditions U as before :

$$\lambda^{\alpha} f_{\alpha}^{\ell} = f_{x}^{\ell} \qquad [U] \qquad (2.8)$$

for all C = 0,1,...,k.

2.3.3 Optimality Conditions.

Since m_x is deterministic in the model, the variance when estimating the drift (given by (2.16)) reduces to $D^2[M_{\alpha}^{*}]$. After expanding the expression (2.14) this becomes

$$D^2[M_x^{*}] = \lambda^{\alpha} \lambda^{\beta} K_{\alpha\beta} \qquad (2.17)$$

As before, we use Lagrange multipliers to optimize $D^2[M_x^{*}]$ under the universality conditions (2.8). We therefore minimize

$$\lambda^{\alpha} \lambda^{\beta} K_{\alpha\beta} + 2 \mu_{\ell} (\lambda^{\alpha} f_{\alpha}^{\ell} - f_{x}^{\ell}) \qquad (2.18)$$

by differentiating with respect to λ^{α} and μ_{ℓ}.

Setting the partial derivatives with respect to λ^{α} to zero gives the optimality conditions :

$$\lambda^{\alpha} K_{\alpha\beta} + \mu_{\ell} f_{\alpha}^{\ell} = 0 \qquad [O] \qquad (2.19)$$

2.3.4 The System for Estimating the Drift.

Combining the equations [U] and [O] gives us the complete system of equations for estimating the drift at a point x :

$$\left.\begin{aligned} \lambda^{\alpha} K_{\alpha\beta} + \mu_{\ell} f_{\alpha}^{\ell} &= 0 \qquad [O] \\ \lambda^{\alpha} f_{\alpha}^{\ell} &= f_{x}^{\ell} \qquad [U] \end{aligned}\right\} \qquad (2.20)$$

This can be written in matrix form as

$$\left[\begin{array}{c|c} K_{\alpha\beta} & f_\alpha^\ell \\ \hline f_\beta^\ell & 0 \end{array}\right] \left[\begin{array}{c} \lambda^\beta \\ \hline \mu_\ell \end{array}\right] = \left[\begin{array}{c} 0 \\ \hline f_x^\ell \end{array}\right]$$

When the solution to this system is substituted into the expression for the variance of M_x^*, we finally obtain

$$D^2[M_x^*] = - \mu_\ell \, f_x^\ell \qquad (2.21)$$

2.3.5 Remark.

It is interesting to note that the left-hand side of this system (2.20) is the same as was obtained for universal kriging (2.12). However the right-hand side is different. This means that the remarks about the non singularity of the system which were made for U.K. also hold in this case (§ 2.2.1.4).

It should also be noted that the question of estimating the drift (when it is to be used in structural analysis) has not been completely solved. This problem can well be called the vicious circle of U.K.

a) For U.K., we need to have done the structural analysis i.e. to have obtained the covariance function.

b) To calculate the experimental covariance function, we have to subtract the drift m(x) from the raw phenomenon Z(x).

c) The "true" drift is unattainable (In fact, this notion is only meaningful within the model, and has nothing in common with the experimental drift).

d) Lastly, the optimal estimator of the drift (which we might wish to substitute for m_x in order to

make statistical inferences) can only be obtained after the covariance function is already known ; that is, after the inference has been made).

This raises one of the major methodological problems in U.K., one which seriously limits the applicability of this method. This difficulty will be re-examined in a later section. For the moment we can already say that this is the price that has to be paid for the model chosen at the outset (2.1). Although this approach is mathematically rigourous, it places too much emphasis on the dichotomy "drift vs residual" which is so intuitively appealing. Some of the advantages of the IRF-k are that they will consciously reject this approach, and that they treat the study of non-stationary phenomena from a purely operational point of view.

For the rest of this chapter and in the following one we shall assume that this obstacle has been overcome, that is, that the covariance K_{xy} is known. This fundamental methodological question will be treated in chapter 4. For the time being, the implications of adopting the intrinsic hypothesis (as opposed to stationarity) will be considered.

2.4 ESTIMATING THE COEFFICIENTS OF THE DRIFT.

2.4.1 The problem.

Suppose that the covariance K_{xy} is known. We have just found the system for estimating the drift m_x at a given point. Now in the model, the drift is of the form :

$$m_x = a_\ell \, f_x^\ell \qquad (2.4)$$

where the f_x are fixed and the coefficients a_ℓ are unknown. Our new problem consists of estimating these coefficients.

2.4.2. The system.

Looking at the system (2.20) for estimating the drift at a given point x, it is clear that the solutions λ^β and μ_ℓ depend linearly on the f_x. Naturally when a different point x is to be estimated, the solutions λ^β and μ_ℓ vary, and are therefore a function of x. So it would be better to call them $\lambda^\beta(x)$ and $\mu_\ell(x)$. The preceding remark shows that there exist matrices λ^α_ℓ and $\mu_{\ell s}$ such that :

$$\lambda^\alpha(x) = \lambda^\alpha_\ell \; f^\ell_x \quad \text{and} \quad \mu_\ell(x) = \mu_{\ell s} \; f^s_x \tag{2.22}$$

Since we have assumed that the f^ℓ_x are linearly independent, the matrices λ^α_ℓ and $\mu_{\ell s}$ are unique. If we substitute the expressions in (2.22) back into the system for estimating the drift (2.20), we obtain the following system :

$$\left.\begin{aligned} \lambda^\beta_s \; f^s_x \; K_{\alpha\beta} + \mu_{\ell s} \; f^s_x \; f^\ell_\alpha &= 0 \\ \lambda^\alpha_s \; f^s_x \; f^\ell_\alpha &= f^\ell_x \end{aligned}\right\} \tag{2.23}$$

which must hold identically for all x. For each term f^s_x, we must therefore have :

$$\left.\begin{aligned} \lambda^\beta_s \; K_{\alpha\beta} + \mu_{\ell s} \; f^\ell_\alpha &= 0 \\ \lambda^\alpha_s \; f^\ell_\alpha &= \delta^\ell_s \end{aligned}\right\} \tag{2.24}$$

These are the equations that have to be satisfied by the coefficients λ^α_ℓ and $\mu_{\ell s}$.

2.4.3 Estimating the Coefficients a_ℓ.

The optimal estimator of the drift can be written as

$$M^*_x = \lambda^\alpha(x) \; Z_\alpha = \lambda^\alpha_\ell \; f^\ell_x \; Z_\alpha = A^*_\ell \; Z_\alpha$$

where $$A^*_\ell = \lambda^\alpha_\ell \, Z_\alpha \qquad (2.25)$$

We note that, from the system (2.24)

$$E[A^*_\ell] = \lambda^\alpha_\ell \; E[Z_\alpha] = \lambda^\alpha_\ell \; a_s \, f^s_\alpha = a_\ell$$

This holds no matter what the true values of a_ℓ are. Moreover, the first part of system (2.24) shows that provided $E[A^*_\ell] = a_\ell$, the variance $D^2[A^*_\ell]$ is a minimum.

Summarizing, the quantity A^*_ℓ defined by 2-25 is the optimal (linear) universal estimator of the coefficient a_ℓ, for any value of ℓ.

Using the system (2.24) it is easy to show that

$$\text{Cov}\,[A^*_\ell, A^*_s] = -\mu_{\ell s} \qquad (2.26)$$

2.5 PROPERTIES OF THE U.K. ESTIMATORS.

From now on we shall assume that the reader understands the mechanism for finding the universal estimators (i.e. defining the universality condition and then minimizing a quadratic form (the variance) under certain constraints). So we shall not go into the details of the calculations in the subsequent sections. In general these follow from elementary theorems in linear algebra. Details on this aspect can be found in [1], chapter 4. We shall now focus our attention on the conclusions that can be drawn from these calculations and on the meaning that should be given to them.

2.5.1 Tensorial Invariance.

2.5.1.1 The Problem.

The question is whether the estimation of the

drift (or, if you prefer, the dichoyomy "drift vs residual") depends on the coordinate system used.

Initially we assume that we are working in a unique neighbourhood (and not a moving one). The representation $m(x) = a_\ell f_x^\ell$ is valid for all the available data, and all the data are included in the system for estimating the drift (i.e. indices α and β)

$$\left.\begin{aligned} \lambda^\beta K_{\alpha\beta} + \mu_\ell f_\alpha^\ell &= 0 \\ \lambda^\alpha f_\alpha^\ell &= f_x^\ell \end{aligned}\right\} \qquad (2.20)$$

2.5.1.2 Changing the coordinate system.

Suppose that a invertible linear transformation is applied to the basis functions f^ℓ. Let B be a non singular matrix representing the transformation. Its inverse is B'. Let $\varphi_x^\ell = B_s^\ell f_x^s$. Hence $f_x^\ell = B'^\ell_s \varphi_x^s$

Then

- the true drift m(x) can be written as

$$m(x) = a_\ell f_x^\ell = a_\ell B'^\ell_s \varphi_x^s = a'_s \varphi_x^s \qquad (2.27)$$

where $a'_s = a_\ell B'^\ell_s$

- the optimal estimator of m(x) in the new coordinate system is :

$$m^{*\prime}(x) = \lambda'^\alpha Z_\alpha$$

where

$$\left.\begin{aligned} \lambda'^\beta K_{\alpha\beta} + \mu'_\ell \varphi_\alpha^\ell &= 0 \\ \lambda'^\alpha \varphi_\alpha^\ell &= \varphi_x^\ell \end{aligned}\right\} \qquad (2.28)$$

This is just the system (2.20) rewritten in the terms of the new coordinates. If we replace φ_s^ℓ by the expression

$\beta_s^{\ell} f_{\alpha}^{s}$, we see that

a) $\mu'_{\ell} \varphi_{\alpha}^{\ell}$ is of the form $\mu''_{\ell} f_{\alpha}^{\ell}$,

b) the equations $\lambda'^{\alpha} \varphi_{\alpha}^{\ell} = \varphi_x^{\ell}$ can be written as $\lambda'^{\alpha} f_{\alpha}^{\ell} = f_x^{\ell}$ (because the matrix B is non singular).

The system (2.28) then becomes

$$\left.\begin{aligned} \lambda'^{\beta} K_{\alpha\beta} + \mu''_{\ell} f_{\alpha}^{\ell} &= 0 \\ \lambda'^{\alpha} f_{\alpha}^{\ell} &= f_x^{\ell} \end{aligned}\right\} \qquad (2.29)$$

But this is just the system (2.20) expressed in terms of the initial coordinate system. It has a unique solution. Consequently

$$\lambda'^{\alpha} = \lambda^{\alpha}$$

$$\mu''_{\ell} = \mu_{\ell}$$

and

$$M'^{*}_x = M^{*}_x$$

So the optimal estimator of the drift remains unchanged when a invertible linear transformation is applied to the basis functions.

2.5.1.3 Transformation of the coefficients A^{*}_{ℓ} .

From the preceding result it is clear that

$$A'^{*}_{\ell} \varphi_x^{\ell} = M'^{*}_x = M^{*}_x = A^{*}_{\ell} f_x^{\ell}$$

Substituting for φ_x^{ℓ} gives

$$A'^{*}_{\ell} B_s^{\ell} f_x^{s} = A^{*}_{\ell} f_x^{\ell}$$

and hence

$$A'^{*}_{\ell} = B'^{s}_{\ell} A^{*}_{s}$$

that is, the optimal estimators are transformed in the same way as the real coefficients (cf. formula (2.27)).

2.5.1.4 Important practical comment.

The results given above were possible only because the transformation applied to the basis functions was linear and non-singular.

In practice it often happens that the region under study is too large for one representation $m(x) = a_\ell f_x^\ell$ to be valid throughout. We therefore have to work in sliding neighbourhoods each with a different form for the drift, and with different coefficients a_ℓ. In general in each neighbourhood the coordinates are transformed by a translation, so the basis functions are now $\varphi^\ell(x) = f^\ell(x-x_o)$. It can be shown that the necessary and sufficient condition for $\varphi^\ell(x)$ to be of the form $B_s^\ell f_x^s$ where B_s^ℓ is non-singular, is that the basis functions f^ℓ be "polynomial exponentials" (that is, of the form $P(x)\, e^{<\lambda,x>}$ where $P(x)$ is a polynomial and $<\lambda,x>$ is a linear form of the components of the vector x.).

The polynomial exponentials include the trigonometric functions, and also the ordinary polynomials which play an important role in describing the drift. In fact, unless special physical considerations focus attention on a particular point in space which would then become the origin, it would be normal practice to use the polynomials x^n for $n \leq$ some value m as the basis functions. This paragraph shows that it is not just for the sake of convenience that the polynomials were chosen as basis vectors. There are very strong theoretical grounds for this decision. This situation also reappears when dealing with the IRF-k.

2.5.2 Comparison with the method of Maximum Likelihood.

In the case where the set of data is finite and where the R.F. $Z(x)$ is normally distributed, it can be shown (see [1] § 4.2.5) that the optimal estimator of the drift as defined in

section 2.3 is identical with the maximum likelihood estimator.

To be more precise, if we assume that the drift is of the form $m_\alpha = a_\ell f_\alpha^\ell$ at the sample points and if we are looking for the coefficients a_ℓ which maximize the likelihood (i.e. the joint probability) of the R.V. Z_α at the sample points, we obtain the same solution for the A_ℓ as was obtained in the formulae (2.24) and (2.25).

2.5.3 Comparison with least squares.
(cf [1], § 4.2.7)

Fitting the drift $m(x) = b_\ell f_x^\ell$ by least squares consists of determining the coefficients b_ℓ which minimize the integral

$$I = \int_S [Z(x) - b_\ell f_x^\ell]^2 \, dx$$

where S represents the domain where Z(x) has been measured.

It can be shown that the estimator $B_\ell f_x^\ell$ obtained in this way is a universal estimator but that it is, in general, not optimal. The necessary and sufficient condition for it to be optimal is that the (k+1) basis functions f_x^ℓ must be linear combinations of the k+1 distinct eigen-functions of the kernel K_{xy} considered as an operator acting on the functions with their support in S.

This condition is fairly complex, since it involves the choice of the basis functions f_x^ℓ, the form of the covariance function K_{xy} and the geometry of the region where the realization is known. Fortunately it can be demonstrated that this is satisfied, for any region S and for any continuous f_x^ℓ, when K_{xy} is a pure nugget effect covariance ; that is, when Z(x) can be split into a drift plus a completely unstructured residual.

As a first approximation, when the distinction between the drift and the residual is unclear (e.g. very short range for

K_{xy}), the least squares estimator turns out to be quite close to the optimal estimator. However, when the residual includes a substantial part of the structure of the phenomenon, the two estimators are significantly different.

2.5.4. The Additivity Theorem.. (cf [1], § 4.3.2)

We now go back to U.K. (§ 2.2.1). In this section we shall use U to denote the weighting factors which are the solution to (2.12). Let Z_u denote the U.K. estimator. Formula (2.5) then becomes

$$Z_u = \lambda_u^\alpha Z_\alpha \tag{2.30}$$

If the drift m(x) was really known, there would be no need for the U.K. method, since the residuals Y(x) would be known and would be stationary. All that would be needed, is to apply the usual method for simple kriging ([1], § 3.4.1). This would give

$$Z^* = \lambda_K^\alpha (Z_\alpha - m_\alpha) + m_x \tag{2.31}$$

where the λ_K^α denote the weighting factors for simple kriging.

The additivity theorem which is easy to prove (see [1]) leads to the following relation

$$Z_u = \lambda_K^\alpha (Z_\alpha - M_\alpha^*) + M_x^* \tag{2.32}$$

This shows that the U.K estimator can be obtained by kriging (in the sense of simple kriging) the optimally estimated residuals $Z_\alpha - M_\alpha^*$ as if they were the true residuals, and then adding the optimal estimator of the drift.

Of course, this proof is essentially based on the fact that M_x^* is the optimal universal estimator of the drift.

One other remark about the additivity theorem has to be made. Since $\lambda_K^\alpha Z_\alpha$ is just the estimator given by simple kriging (and denoted here for convenience by Z_K), the preceding equation can be rewritten as

$$Z_u = Z_K + Z_D \tag{2.33}$$

where Z_D is the correction term for the drift and is equal to $M_x^* - \lambda_K^\alpha M_\alpha^*$. From this it is easy to show that the variances are additive

$$D^2[Z_u] = D^2[Z_K] + D^2[Z_D] \tag{2.34}$$

This shows that the variance in U.K. is equal to the simple kriging variance plus the variance of the correction term.

Since the term $D^2[Z_D]$ is clearly either positive or zero, the U.K. variance is always greater than or equal to that of simple kriging. This can be understood by considering kriging in terms of projections (cf [3]). In simple kriging, the solution belongs to the space of linear combinations whereas in U.K., the solution belongs to that of authorized linear combinations, (i.e. those satisfying the universality conditions. This new space is a subspace of the preceding one, and the variance of the new estimator is necessarily larger than in the first case.

2.6 KRIGING CONSIDERED AS AN INTERPOLATOR.

2.6.1 The Problem.

The idea behind geostatistics is to treat the available data as a realization z(x) of a R.F. Z(x) which is known at the sample points x_α, $\alpha = 1,2,\ldots,n$. Provided that certain hypotheses are

satisfied, this model can be used to determine the covariance K(x,y) and hence to estimate the value at a point as a linear combination of the available data. This is done by minimizing the estimation variance. So the estimator $z^*(x)$ is

$$z^*(x) = \lambda^\alpha \, z_\alpha = \lambda^\alpha(x) \; z(x_\alpha)$$

As can be seen from the form of the kriging system

$$\left[\begin{array}{c|c} K & F \\ \hline F^t & 0 \end{array}\right] \left[\begin{array}{c} \lambda \\ \hline \mu \end{array}\right] = \left[\begin{array}{c} K_x \\ \hline f_x \end{array}\right]$$

the weighting factors λ^α depend on the location of the point to be estimated, and can therefore be written as $\lambda^\alpha(x)$. They also depend on the relative positions of the sample points and on the choice of the covariance function and the drift functions.

If we consider the estimator as a function of x, the procedure can be described in the following way :

For a given function z(x) (which is a realization of Z(x) known at the points x_α), we have found an interpolating function $z^*(x) = \lambda^\alpha(x) \; z(x_\alpha)$ which has the property of being an exact interpolator at sample points.

The main objective in this section is to study the properties of the kriging estimator viewed as an interpolator. Firstly we shall see that it is a linear combination of the functions $K_{\alpha x}$ and $f^\ell(x)$. At the same time we shall also find the linear system satisfied by the coefficients in this linear combination. We shall call this system, the dual system [1] , [4].

2.6.2 The Notation to be used.

Let z denote the vector whose components are $z_1, z_2, \ldots, z_n$. Using matrix notation we can write

$$z^*(x) = z^t \lambda(x)$$

Now suppose that the kriging matrix is non-singular. Let its inverse be

$$\left[\begin{array}{c|c} U & V \\ \hline V^t & W \end{array}\right]$$

Clearly this satisfies

$$\begin{array}{c} \\ n \\ k \end{array} \begin{array}{c} \begin{array}{cc} n & k \end{array} \\ \left[\begin{array}{c|c} K & F \\ \hline F^t & 0 \end{array}\right] \end{array} \begin{array}{c} \begin{array}{cc} n & k \end{array} \\ \left[\begin{array}{c|c} U & V \\ \hline V^t & W \end{array}\right] \end{array} = \left[\begin{array}{c|c} I & 0 \\ \hline 0 & I \end{array}\right] \qquad (2.35)$$

Using this notation

$$\left[\begin{array}{c} \lambda \\ \hline \mu \end{array}\right] = \left[\begin{array}{c|c} U & V \\ \hline V^t & W \end{array}\right] \left[\begin{array}{c} K_x \\ \hline f_x \end{array}\right] \qquad (2.36)$$

Hence

$$\lambda(x) = U K_x + V f_x$$

If we substitute this expression for λ into the equation for $z^*(x)$, we find that

$$z^*(x) = z^t U K_x + z^t V f_x \qquad (2.37)$$

From (2.35) we note that U is symmetric. Put

$$b = Uz \qquad (2.38)$$

$$c = V^t z \qquad (2.39)$$

$$z^*(x) = b^t K_x + c^t f_x$$

This can be rewritten using Einstein's summation

convention as :

$$z^*(x) = b^\alpha K_{\alpha x} + c_s f^s(x) \qquad (2.40)$$

So $z^*(x)$ is a linear combination of $K_{\alpha x}$ and $f^\ell(x)$. This can be incorporated into a single matrix equation by adding k zeros at the end of the vector z. Then

$$\left[\begin{array}{c|c} U & A \\ \hline V^t & B \end{array}\right]\left[\begin{array}{c} z \\ \hline 0 \end{array}\right] = \left[\begin{array}{c} b \\ \hline c \end{array}\right]$$

Since A and B can be chosen arbitrarily, we can take A = V and B = W. This gives

$$\left[\begin{array}{c|c} U & V \\ \hline V^t & W \end{array}\right]\left[\begin{array}{c} z \\ \hline 0 \end{array}\right] = \left[\begin{array}{c} b \\ \hline c \end{array}\right]$$

Since we recognize the inverse of the kriging matrix in the above expression, we also have

$$\left[\begin{array}{c|c} K & F \\ \hline F^t & 0 \end{array}\right]\left[\begin{array}{c} b \\ \hline c \end{array}\right] = \left[\begin{array}{c} z \\ \hline 0 \end{array}\right] \qquad (2.41)$$

By definition this is the dual system [4]. If we compare this with usual kriging systems we see that the kriging system uses the values of $K_{\alpha x}$ and f^ℓ_x whereas this one only involves the experimental values.

From this system we note that

$$F^t b = 0$$

i.e. $\quad b^\alpha f^\ell_\alpha = 0 \qquad \forall \ell = 1, 2, \ldots, k$

That is, in the usual notation, b is an admissible linear combination.

It should be noted that the dual system cannot be used to calculate the estimation variance which has no meaning in this framework. Nor is it possible to obtain the estimation error by replacing $z(x)$ by $z^*(x)$ in this system. Moreover, there is no way of using this to choose the covariance function $K(x,y)$, at least without going back to the kriging methodology. This system only serves to define an interpolator of the form $b^\alpha K_{\alpha x} + c_\ell f^\ell(x)$ by defining the set of coefficients which guarantee that it is an exact interpolator at sample points. This leads to n conditions. To this we have to add the conditions to ensure that the b^α are ALC's.

2.6.3 Properties of the Estimator.

We shall now go on to study the properties of this estimator.

a) Let $z_\beta = \delta_{\alpha\beta}$. We know that in this case

$$z^*(x) = \lambda_\alpha(x)$$

that is, $\lambda^\alpha(x)$ is of the form $b^\alpha K_{\alpha x} + c_\ell f^\ell(x)$. Since this takes the value 1 in x_α and 0 elsewhere, it is easy to show the $\lambda^\alpha(x)$ are independent since $a_\alpha \lambda^\alpha(x) \equiv 0$ implies that $a_\alpha \lambda^\alpha(x_\beta) = 0$ for all $\beta = 1,\ldots,n$ and hence that $a_\beta = 0$ for all β.

b) We shall now consider the case where $z(x) = f^\ell(x)$

$$z^*(x) = \lambda^\alpha(x)\ f^\ell(x_\alpha)$$

but by construction, the λ^α satisfy

$$\lambda^\alpha\ f^\ell(x_\alpha) = f^\ell(x)$$

hence

$$z^*(x) = f^\ell(x) = z(x)$$

that is, the interpolator is exact for the basis functions.

c) It can also be shown [1] that this interpolator is also exact for the functions $f(x) = d^\alpha K_{\alpha x}$ where d belongs to the space orthogonal to the one that is spanned by the drift functions.

$$d^\alpha f^\ell(x_\alpha) = 0 \qquad \forall\, \ell = 1,\ldots,k$$

2.6.4 Examples in $\mathbb{R}^1$.

A. Let $K(h) = |h^3|$. This function is not a covariance because it is positive-definite only if λ is an ALC of order 1. (For the definition of an ALC of order n see [9].

In a subsequent chapter we shall see that this function is a covariance of order 1 and can therefore be used to krige IRF-1. We shall see that the system is exactly the same if we replace the covariance matrix by a generalized covariance matrix.

We can therefore write

$$z^*(x) = b^\alpha |x-x_\alpha|^3 + c_1 + c_2 x$$

where b^α satisfies

$$\sum b^\alpha = 0 \quad , \sum b^\alpha x_\alpha = 0$$

We can now assume that with $x_1 < x_2 < \ldots < x_n$:

1/ If $x \geq x_n$

$$z^*(x) = b^\alpha (x-x_\alpha)^3 + c_1 + c_2 x$$

$$= b^\alpha x^3 - 3 b^\alpha x^2 x_\alpha + 3 b^\alpha x x_\alpha^2 - b^\alpha x_\alpha^3 + c_1 + c_2 x$$

Since b^α is an ALC of order 1, $b^\alpha x^3$ and $3 b^\alpha x^2 x_\alpha$ disappear, leaving

$$z^*(x) = c_1 - b^\alpha x_\alpha^3 + (c_2 - 3\, b^\alpha x_\alpha^2)\, x$$

Consequently outside the interval $]x_1, x_n[$, $z^*(x)$ is a first order polynomial.

3/ Inside the interval $[x_i, x_{i+1}[$

$$z^*(x) = \sum_{\alpha \le i} b^\alpha (x - x_\alpha)^3 + \sum_{\alpha > i} b^\alpha (x_\alpha - x)^3 + c_1 + c_2\, x_2$$

that is, it is a cubic. Since the function K(h) can be differentiated twice, so is $z^*(x)$. We therefore find that

$$z^{*\prime\prime}(x) = 6\, b^\alpha |x - x_\alpha|$$

In the interval $[x_i, x_{i+1}[$

$$z^{*\prime\prime}(x) = 6 \sum_{\alpha \le i} b^\alpha (x - x_\alpha) - 6 \sum_{d > i} b_\alpha (x - x_\alpha)$$

Hence

$$z^{*\prime\prime\prime}(x) = 12 \sum_{\alpha \le i} b^\alpha$$

This step follows from the condition $\sum b^\alpha = 0$. So

$$\lim_{x \uparrow x_{i+1}} z^{*\prime\prime\prime} = 12 \sum_{\alpha \le i} b^\alpha$$

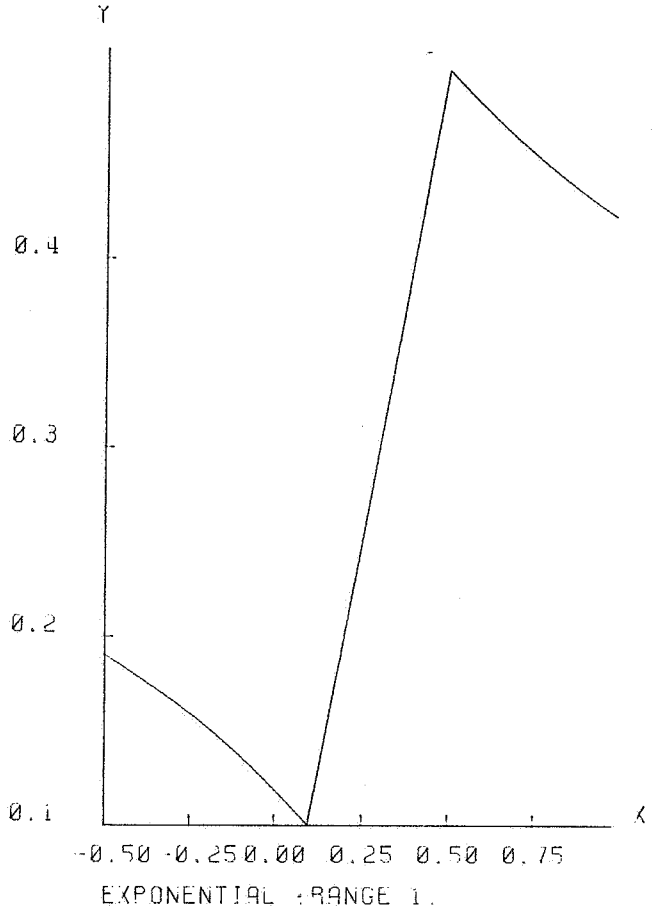
Similarly in the interval $[x_{i+1}, x_{i+2}[$

$$z^{*\prime\prime\prime}(x) = 12 \sum_{\alpha + i + 1} b^\alpha$$

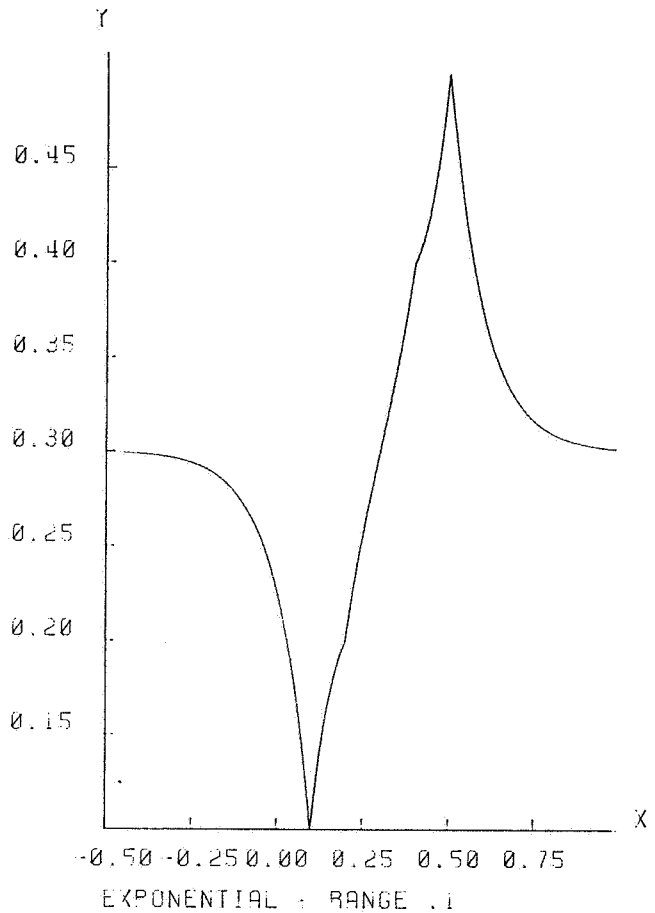
Consequently the jump in the 3rd order derivative at the point x_{i+1} is $12\, b_{i+1}$. The condition $\sum b_\alpha = 0$ implies that on average these jumps have different signs.

B. Interpolation with an exponential covariance of the five points :

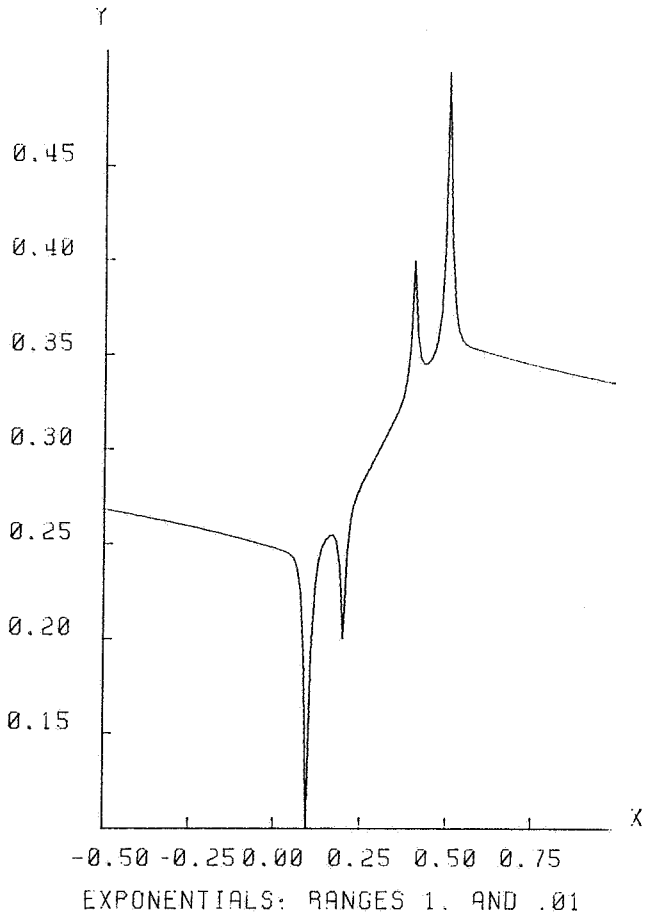
$x_1 = (.1, .1)$ $x_2 = (.2, .2)$ $x_3 = .3, .3)$ $x_4 = (.4, .4)$ $x_5 = (.5,.5)$



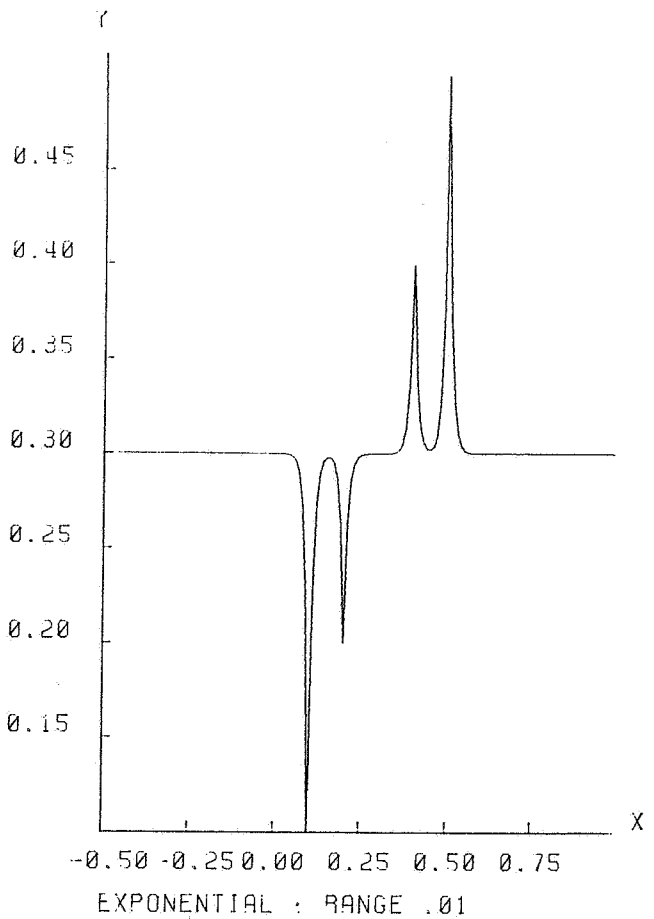
EXPONENTIAL : RANGE 1.



EXPONENTIAL : RANGE .1



EXPONENTIALS: RANGES 1. AND .01



EXPONENTIAL : RANGE .01

2.6.5 The Effect of certain Linear Operators.

Now that we have an interpolator z*, we can carry out a certain number of operations (such as differentiation, integration, convolution etc.) provided that these are not prohibited because of the properties of K(h) and of $f^{\ell}(x)$. We can ask whether the interpolator (after being transformed) is the same as the kriged estimator of the transformed variable.

We shall first consider the case where the operation is a moving average over a certain volume. Let Z_V be the kriged estimate of the average value of Z over V. Then

$$(Z^*)_V = \frac{1}{V} \int_V Z^*(x)\, dx$$

$$= \frac{b^{\alpha}}{V} \int_V K_{\alpha x}\, dx + \frac{c_{\ell}}{V} \int_V f^{\ell}(x)\, dx$$

In the usual way let

$$f_V = \frac{1}{V} \int_V f^{\ell}(x)\, dx \quad \text{and} \quad K_{\alpha v} = \frac{1}{V} \int_V K_{\alpha x}\, dx$$

so

$$(Z^*)_V = b^{\alpha} K_{\alpha v} + c_{\ell} f_V^{\ell}$$

From (2.38) and (2.39) this can be written

$$(Z^*)_V = Z^t U K_V + Z^t V f_V$$

$$= Z^t (U K_V + V f_V)$$

The term in brackets is just the values of the weighting factors used to krige the average. So

$$(Z^*)_V = Z_V$$

This property holds more generally. It is due to the fact that kriging is a linear operator. Consequently, kriging a variable that is obtained from Z by a linear transformation merely involves a change in the vector part (i.e. the right hand side) of the kriging system. In this case, $K_{\alpha x}$ is replaced by $K_{\alpha \mathcal{L}} = \text{Cov}\ (Z_\alpha, \mathcal{L}(Z))$ where $\mathcal{L}$ is the transformation in question. The property, therefore, holds whenever

$$\text{Cov}\ (Z_\alpha, \mathcal{L}(Z)) = \mathcal{L}(\text{Cov}\ (Z_\alpha, Z))$$

which is the case for most linear operators such as integration, differentiation and convolution.

2.6.6 Parallel with the common methods of numerical integration.

Suppose that we want to calculate

$$I(f) = \frac{1}{V} \int_V f(x)\ dx$$

There are lots of methods in numerical analysis for calculating I(f) as a linear combination of the (point) values

$$I_n(f) = \lambda^\alpha\ f(x_\alpha)$$

The weighting factors are often calculated so that the formula for the numerical integration will be correct over a vector space of functions D, i.e.

$$I(f) = I_n(f) \qquad \forall\ f \in D$$

One of the most common choices for D is the space of polynomials. There are many reasons for this : convenience, the fact that these methods are well known, or because there is no reason for choosing another one.

The form of $z^*(x)$ suggests that it may be more interesting to choose the weighting factors so as to integrate the functions $K_{\alpha x}$ and $f^{\ell}(x)$ exactly. (This would automatically lead to an exact numerical integration of $z^*(x)$).

This would give us

$$I(K_{\alpha x}) = K_{\alpha v} = I_n(K_{\alpha x}) = \lambda^{\alpha} K_{\alpha x}$$

$$I(f^{\ell}(x)) = f^{\ell}_v = I_n(f^{\ell}(x)) = \lambda^{\alpha} f^{\ell}(x_{\alpha})$$

That is, it leads to a system of n+k equations to determine n unknowns, which in general has no solution.

However in the stationary case with a known mean, the last k equations disappear, so it can be solved.

In fact, what we really want to do is to integrate $z^*(x)$ numerically in an exact way. Remembering the conditions

$$b^{\alpha} f^{\ell}(x_{\alpha}) = 0 ,$$

we see that it is no longer necessary that the error in integrating $K_{\alpha x}$ be zero. It need only be of the form $\mu_{\ell} f^{\ell}(x_{\alpha})$.

1/ $I(K_{\alpha x}) - I_n(K_{\alpha x}) = \mu_{\ell} f^{\ell}(x_{\alpha})$

Writing this in the usual way gives us

$$\lambda^{\beta} K_{\alpha\beta} + \mu_{\ell} f^{\ell}(x_{\alpha}) = K_{\alpha v}$$

2/ Moreover, the $f^{\ell}(x)$ have been integrated exactly (i.e. $\lambda^{\alpha} f^{\ell}(x_{\alpha}) = f^{\ell}_v$).

It is easy to check that conditions 1/ and 2/ give the system for kriging a mean value.

It should be noted that even though this formalism leads to the same system as kriging does, it does not allow us to evaluate the error variance since there is no way of interpreting the μ_ℓ using this method of numerical integration.

2.6.7 The Link between Kriging and Spline Interpolation.

The dual system provides an operational definition of an interpolating function which is exact over a vector space of functions and passes through the experimental points. It can be shown that this interpolator has some interesting properties as a function.

To start with, it is a spline function (i.e. an interpolating function which minimizes the norm of a suitably chosen operator in a certain Hilbert space. Conversely, any spline function can be interpreted as a kriging interpolator. The only problem with this approach is to choose the appropriate operator and a suitable Hilbert space. For more details about this see [4], [6] and [7], [8].

2.7 COKRIGING

2.7.1 The Problem.

Cokriging was developed to treat cases where several regionalizations are being studied simultaneously. This situation often arises in practice and can occur in many different ways. For example, the grades in a polymetallic deposit ; the simultaneous study of windspeed and pressure in meteorology or of altitude and gradient in topography. In other cases it may be necessary to take into account the accuracy of the data values as well as their actual values.

No fundamentally new theoretical problems arise in cokriging (cf. § 2.7.3). But the problems involving the notation are by no means trivial. From the practical point of view it has to be used with care, or else it is easy to end up having to solve very large systems of linear equations which give rise to only a limited improvement in the precision.

Rather than going into all the details of the Universal Cokriging which can be found in § 4.5 of [1], we shall limit ourselves to considering a minor theoretical complication concerning the singularity of the kriging matrix. The rest of the section can be omitted in the first reading.

2.7.2 Notation.

At first it seems wise to use a slightly different system of notation from that used earlier.

- Instead of a single R.F. $Z(x)$, we now have a family of R.F. $Z_i(x)$ where the index i denotes the variable under study and takes the values $1,2,\ldots,d$. Let D denote the set $\{1,2,\ldots,d\}$ of indices. The d R.F. Z_i are supposed to be defined on $\mathbb{R}^n$.

- Each of these R.F. has a drift $m_i(x)$ which is defined by

$$m_i(x) = E[Z_i(x)] \tag{2.42}$$

This drift can be developed in terms of a family of basis functions :

$$m_i(x) = a_\ell(i)\ f_i^\ell(x) \tag{2.43}$$

where the index ℓ goes from 0 to k_i.

Lastly the cross-covariance is defined by

$$K_{ij}(x,y) = \text{cov}\ [Z_i(x),\ Z_j(y)] \tag{2.44}$$

The method described in the preceding sections can now be re-applied here. The only difference is that there is an extra index. The usual properties for kriging (exact interpolation and the additivity theorem) hold here too.

However, even though it may seem to be of little practical use now, it is worthwhile considering another way of representing cokriging. We consider the d values of the R.F. $Z_i(x)$ as a single R.F. $Z(x,i)$ which is defined on the space $E = \mathbb{R}^n \times D$.

From this point of view, Z can be considered as a vector R.F. having $Z(x_o,1),\ldots,Z(x_o,d)$ as its components. Similarly, Z has a drift $m(x,i)$ and a covariance cov $[Z(x,i),\ Z(y,j)]$ which can be denoted by $K_{ij}(x,y)$. Except for the fact that the space E has no obvious physical meaning, this change of notation brings us back to the problem described in the first section § 2.1. Consequently, cokriging can be considered as an example of U.K. defined on a rather special space. This point of view will be used in the following section.

2.7.3 Linear Independence of the Basis Functions.

Let S_i ($i \in D$) denote the domain within $\mathbb{R}^n$ where the ith component of $Z(x,i)$ has been measured. (Intuitively, where we know the realisation of the ith variable $Z_i(x)$).

The set of sample points of the R.F. Z is therefore a subset S of $\mathbb{R}^n \times D$, which is defined by the pairs (x,i) where $i \in D$ and $x \in S_i$. Using this notation (which is just that of U.K.) we assume that the drift can be expanded as

$$m(x,i) = a_\ell \, f^\ell(x,i) \qquad (2.45)$$

where the f^ℓ is a family of basis functions defined on $E = \mathbb{R}^n \times D$. It should be noted that in this presentation there is only 1 variable ; i is merely a coordinate in the space E and a_ℓ does not depend on it.

The reader will recall that the necessary and sufficient condition for the U.K. matrix to be non-singular (provided that the matrix of covariances $K_{ij}(x,y)$ is strictly positive definite) is that : "If there exists a set of coefficients c_ℓ such that

$$c_\ell \, f_\alpha^\ell = 0$$

for all data points α, then the c_ℓ must necessarily be zero (cf. § 2.2.1.4). In the present case, the necessary and sufficient condition is that :

If there exist a set of coefficients c_ℓ such that

$$c_\ell \, f_\alpha^\ell(x,i) = 0 \qquad (2.46)$$

for all couples $(x,i) \in S$, then the c_ℓ must all be zero.

2.7.4 EXAMPLE : Two random functions.

In the simple case where $D = \{1,2\}$, that is, when there are only two R.F. $Z_1(x)$ and $Z_2(x)$, it is easy to write the equations in full.

2.7.4.1 Algebraically independent drifts.

Suppose that there are $k_1 + k_2$ functions $f^{\ell}(x,i)$ (appearing in formula (2.45))and that they satisfy

$$f^{\ell}(x,2) = 0 \qquad \text{for} \quad 1 \leq \ell \leq k_1$$
$$f^{\ell}(x,1) = 0 \qquad \text{for} \quad k_1+1 \leq \ell \leq k_1 + k_2 \tag{2.47}$$

Intuitively this means that there is no "interaction" between the two drifts. More exactly, there is no (functional) algebraic relations between them, This is reasonable when dealing with grades in a polymetallic deposit, or for windspeed and pressure in meteorology, etc..

Suppose that $S_1 = \phi$ (intuitively, that the first R.F. $Z_1(x)$ has not been measured anywhere). It is clear that because of the algebraic independence, knowing $Z_2(x)$ does not provide any information about the drift $m(x,1)$ and that cokriging is therefore impossible.

The set S mentioned in condition (2.46) reduces to $S_2 \times \{2\}$ since S is empty. If we let $c_{\ell} = 0$ for $k_1 < \ell \leq k_1 + k_2$ (2.48) and let c_{ℓ} take any value for $1 \leq \ell \leq k_1$, we find that because of (2.47), $c_{\ell} f^{\ell}(x,i) = 0$ for all $(x,i) \in S$. The basis functions $f^{\ell}(x,i)$ are not linearly independent on S and the cokriging system is singular.

Of course, the same thing happens if S_2 is empty.

The implication of this result is quite clear : there is no point in accumulating information about one variable when there is no available information about the other and when the two are not linked algebraically. Moreover, experience shows that this conclusion can be extended further. When we want to use cokriging to estimate the ith R.F., the data for the other R.F.

Z_j ($j \neq i$) should be considered as additional information. When the cokriging system is being set up, the stress must be put on the data for Z_i itself. All the major practical problems in cokriging stem from this important but rather vague idea.

2.7.4.2 Algebraically linked drifts.

As an example of this, take the case where the drift $m(x,2)$ is the derivative of $m(x,1)$. For simplicity, suppose that we are working in $\mathbb{R}^1$.

The functions $m(x,1)$ and $m(x,2)$ are now linked algebraically. To take account of this, we choose the basis functions such that

$$\begin{cases} f^{\ell}(x,1) = \varphi^{\ell}(x) & \text{for } 0 \leq \ell \leq k \\ f^{\ell}(x,2) = \frac{d}{dx}\varphi^{\ell}(x) & \text{for } 0 \leq \ell \leq k \end{cases}$$

Suppose that in our 1-D example, the $\varphi^{\ell}(x)$ are monomials of degree ℓ (the importance of these functions was discussed in § 2.5.1.4). In this case, the basis functions satisfy

$$\left.\begin{aligned} f^{\ell}(x,1) &= x^{\ell} \\ f^{\ell}(x,2) &= \ell x^{\ell-1} \end{aligned}\right\} \quad \text{for } 0 \leq \ell \leq k \qquad (2.49)$$

Using the same notation as before, it is clear that even if $S_2 = \phi$, the drifts can be linearly independent if the set S is chosen appropriately. However, since $f^{0}(x,2) = 0$, if S_1 is empty, it is possible to find a set of coefficients which are not all zero ($c_0 \neq 0$) for which

$$c_{\ell} f^{\ell}(x,i) = 0 \qquad \text{for all } (x,i) \in S$$

The drifts are linearly dependent and the kriging matrix is singular.

From an intuitive point of view this means that it is possible to cokrige the derivative given the values of the variable itself, but that the converse is not possible.

2.7.5 Some Comments of a Practical Nature.

The preceding examples may seem too abstract and too adademic. However the question of the independence of the drift actually arises in practice. Often it occurs in the following way : the physics of the problem suggests that the basis functions are linked algebraically (e.g. the gamma log is a convolution of the uranium grade, the relation between depth and inclination, or between atmospheric pressure and wind speed). Should this be taken into account when the cokriging system is being set up ? Both answers are possible. It is clear that the choice made will have a marked effect on universality conditions and that this will have serious repercussions on the numerical stability of the results.

Over and above the problems of notation, considerations of this type are in fact the real problem in using cokriging, and this can only be overcome case by case in practice.

CHAPTER 3 - THE INTRINSIC HYPOTHESIS

CHAPTER 3 - THE INTRINSIC HYPOTHESIS

3.1 GENERAL PRESENTATION

In this chapter, the word "intrinsic" will be used in the narrower sense ; that is, it will be taken to mean "intrinsic but not stationary".

In this case, only admissible combinations can be used, because, otherwise, the mean and the variance need not exist. The definition of these combinations is given in [1]. Going back to the decomposition proposed in the preceding chapter :

$$Z(x) = Y(x) + a_\ell \, f^\ell(x) \qquad (3.1)$$

we can no longer say that $a_\ell \, f^\ell(x)$ is the drift of $Z(x)$ since in the intrinsic model the mathematical expectation of $Z(x)$ may be infinite.

In reality, when using the intrinsic model, the object under study is no longer the R.F. $Z(x)$ as it was in the stationary case. Rather, our attention is focused on a new mathematical entity - the admissible linear combinations (ALC), that is, expressions of the form $\lambda^i \, Z(x_i)$ where $\sum \lambda_i = 0$. The object of our study, $Z(x)$, is now seen via an operator and all the new problems encountered using this formalism are due to the fact that there is not a 1-1 correspondence between $Z(x)$ and its transformed values. If $Z(x)$ is known, there are no problems obtaining the ALC, but the converse is not true. Even if all possible ALC are known, there is no way of reconstituting Z, which is only known up to an additive constant.

Consider two R.F. $Z_1(x)$ and $Z_2(x)$ such that

$$Z_1(x) = Z_2(x) + A$$

where A is an arbitrary random variable. Then for all

λ^i such that $\sum \lambda^i = 0$

$$\lambda^i Z_1(x_i) = \lambda^i Z_2(x_i) + A \sum \lambda^i$$

$$= \lambda^i Z_2(x_i)$$

so any admissible linear combination formed using $Z_1(x)$ always has the same value as the same combination formed using $Z_2(x)$.

Conversely, suppose that $\lambda^i Z_1(x_i) = \lambda^i Z_2(x_i)$ for all possible ALC. Let $Y(x) = Z_1(x) - Z_2(x)$. Then $\lambda^i Y(x_i) = 0$ for all ALC, and in particular for any choice of x, $Y(x) - Y(o) = 0$, or $Y(x) = Y(o)$. Consequently $Z_1(x) = Z_2(x) + A$.

This additive constant is going to play an important role in what follows in this chapter. In order to incorporate it into the overall formalism, we are going to let the first basis function f_x^o be unity, and so the additive constant will just be the coefficient a_o.

Rather than going back over all the calculations given in the preceding chapter, we shall limit our attention to the new problems due to this indeterminacy and their consequences on the final results.

3.2 KRIGING

3.2.1 The Problem.

As in the stationary case, we want to find the optimal linear universal estimator :

$$Z_x^* = \lambda^\alpha Z_\alpha \qquad (3.2)$$

But this time, before attempting to minimize the variance $D^2[Z_x^* - Z_x]$ we must check that $Z_x^* - Z_x$ is in fact an ALC. Similarly the universality condition can no longer be expressed as in equation (2.6) :

$$E[Z_x^*] = E[Z_x]$$

because, in the present model, Z_x <u>is not</u> an ALC and so we cannot speak about its mean. The universality condition can only be written as

$$E[Z_x^* - Z_x] = 0 \qquad (3.3)$$

and then only provided that $Z_x^* - Z_x$ is an ALC.

Summarizing, we want to find the set of coefficients λ^α such that

(i) $Z_x^* = \lambda^\alpha Z_\alpha$ (3.2)

(ii) the estimation error $Z_x^* - Z_x$ is an ALC (3.4)

(iii) $E[Z_x^* - Z_x] = 0$ (universality) (3.3)

(iv) $D^2[Z_x^* - Z_x]$ is a minimum (optimality) (3.5)

3.2.2 <u>Setting up the System</u>.

The four conditions listed above are going to be considered in order.

The estimation error can be written as

$$\lambda^\alpha Z_\alpha - Z_x$$

The necessary and sufficient condition for this to be an ALC is that it must be of the form $\lambda^i Z(x_i)$ where $\sum \lambda^i = 0$. That is, $\sum \lambda^i - 1 = 0$. Consequently the condition for admissibility is that

$$\Sigma \lambda^{\alpha} = 1 \qquad [A] \qquad (3.6)$$

After taking account of [A], the universality conditions become

$$E[\lambda^{\alpha} Z_{\alpha} - Z_x] = 0$$

Applying the expansion given in (3.1) gives

$$a_{\ell} (\lambda^{\alpha} f_{\alpha}^{\ell} - f_x^{\ell}) = 0$$

which must hold for all possible values of a_{ℓ} . In fact since $f_x^0 = 1$, it is clear that the admissibility condition [A] guarantees that this holds for $\ell = 0$. The remaining constraints for $\ell \geq 1$ are then

$$\lambda^{\alpha} f_{\alpha} = f_x^{\ell} \qquad [U] \quad \text{for } \ell \geq 1 \qquad (3.7)$$

Finally, because of [A], the variance $D^2[Z_x^* - Z_x]$ is calculated in the same way as before except that the variogramme for Z, $-\gamma$, takes the place of the covariance function K. Thus

$$D^2[Z_x^* - Z_x] = - \lambda^{\alpha} \lambda^{\beta} \gamma_{\alpha\beta} + 2 \lambda^{\alpha} \gamma_{\alpha x} - \gamma_{xx} \qquad (3.8)$$

As in the stationary case, we use Lagrange multipliers to minimize (3.8) subject to the constraints [A] and [U]. Since the conditions [A] and [U] can both be written as :

$$\forall \ell \quad : \quad \lambda^{\alpha} f_{\alpha}^{\ell} = f_x^{\ell} \qquad (3.9)$$

the sought-after solution for λ^{α} and μ_{ℓ} minimizes the expression :

$$(-\lambda^{\alpha} \lambda^{\beta} \gamma_{\alpha\beta} + 2 \lambda^{\alpha} \gamma_{\alpha x} - \gamma_{xx}) + 2 \mu_{\ell} (\lambda^{\alpha} f_{\alpha}^{\ell} - f_x^{\ell}) \ell \geq 0 \qquad (3.10)$$

However it should be remembered that the condition for $\ell = 0$ is fundamentally different from those for $\ell \geq 1$. In what follows, the parameter μ_o will have a very special status (cf infra). But for the moment, it is simpler to write all the conditions in the same way.

3.2.3 Solving the System.

The UK system is obtained in the same way as for the stationary case. This gives :

$$\left.\begin{aligned} -\lambda^{\beta}\,\gamma_{\alpha\beta} + \mu_{\ell}\,f_{\alpha}^{\ell} &= -\gamma_{\alpha x} && [O] \\ \lambda^{\alpha}\,f_{\alpha}^{o} &= f_{x}^{o} && [A] \\ \lambda^{\alpha}\,f_{\alpha}^{\ell} &= f_{x}^{\ell} && [U] \quad \ell \geq 1 \end{aligned}\right\} \qquad (3.11)$$

When the solution to this is substituted into the expression for the estimation variance (3.8), we obtain

$$D^2[Z_x^* - Z_x] = \lambda^{\alpha}\,\gamma_{\alpha x} - \mu_{\ell}\,f_x^{\ell} \qquad (3.12)$$

which, incidentally, can be deduced from the formula for the stationary case (2.13) by replacing K by $-\gamma$ and on noting that $\gamma_{xx} = 0$.

3.2.4 Comments.

The properties which held in the stationary case are also true here, viz.

- the necessary and sufficient condition for the system to be non-singular is that the basis function be linearly independent over the data points (cf § 2.2.1.4).
- kriging is an exact interpolator (cf § 2.2.1.4).
- the linearity of the expressions obtained makes it possible to go from kriging point values to kriging weighted averages as was done earlier in section 2.2.2.

3.3 OPTIMAL ESTIMATION OF THE DRIFT

3.3.1 The Problem.

In contrast to the stationary case, the decomposition

$$Z(x) = Y(x) + a_\ell f_x^\ell \qquad (3.1)$$

goes further than is possible with the intrinsic model. In fact, the formalism being developed in this chapter deals only with ALC, or if you prefer, with increments and combinations of increments of Z. We therefore systematically filter out the additive constants, in particular the term a_0 .

So right from the outset it must be said that the term a_0 which appears in the decomposition (3.1) cannot be estimated. This spotlights the difficulties which were raised in § 1.1, of trying to reconcile the methodological requirements with the intuitive ones of the naturalist. In practice, it is shocking having to renounce estimating the additive term, particularly when the available data Z_α are not subject to any incertitude of this sort. However the criticism is attenuated in some respects by the fact that this does not cause any trouble as far as kriging is concerned. It only becomes a problem when we try to get more out of the results than is permissible from the theoretical point of view.

We will see later that this indeterminacy can be partly removed, but only at the price of an additional arbitrary hypothesis. This shows that there are two types of properties : those which are fundamentally linked to the methodology and can be called inherent properties, and those which are due to arbitrary choices made by the user. This distinction will be developed further in the theory of the IRF-k.

3.3.2 Setting up the System.

The preceding remarks lead to the conclusion that it is impossible to estimate the drift $a_\ell f_x^\ell$ itself. This is also confirmed by looking at the formulae obtained in the stationary case and noting that the estimation obtained has $\sum \lambda^\alpha = 1$ and is therefore not an ALC.

However we can still estimate the increment in the drift between any two points. Let y_o be an arbitrary point and let

$$Z'(x) = Z(x) - Z(y_o)$$

The R.F. $Z'(x)$ is an ALC and therefore has a mean. Its drift is

$$m'(x) = E[Z'(x)] = a_\ell (f_x^\ell - f_{y_o}^\ell) = a_\ell \varphi_x^\ell \qquad (3.14)$$

where $\varphi_x^\ell = f_x^\ell - f_{y_o}^\ell$. Note that the term in a_o has disappeared from the formula.

In addition, $Z'(x)$ has a covariance

$$\mathrm{Cov}[Z'(x), Z'(y)] = -\gamma_{xy} + \gamma_{xy_o} + \gamma_{yy_o} \qquad (3.15)$$

This brings us back to the situation in Chapter 2, except that the estimator of the drift m'_x has to be an ALC of the Z (from condition [A]). So we have to find an estimator satisfying the following conditions :

$$M'^*_x = \lambda^\alpha Z_\alpha \qquad (3.16)$$

$$\sum \lambda^\alpha = 0 \qquad [A] \qquad (3.17)$$

Because of [A], the formula (3.16) can be written as

$$M_x'^{*} = \lambda^\alpha Z_\alpha' \tag{3.18}$$

We then want the estimator to be universal ; that is

$$E[\lambda^\alpha Z_\alpha'] = m_x' = a_\ell \varphi_x^\ell \qquad (\ell \geq 1)$$

When $\lambda^\alpha Z_\alpha'$ is expanded, this becomes

$$\lambda^\alpha a_\ell \varphi_\alpha^\ell = a_\ell \varphi_x^\ell \qquad (\ell \geq 1)$$

which leads to the universality conditions

$$\lambda^\alpha \varphi_\alpha^\ell = \varphi_x^\ell \qquad (\ell \geq 1) \qquad [U] \tag{3.19}$$

Lastly we want to minimize the variance $D^2[M_x'^{*}]$ which is equal to

$$D^2[M_x'^{*}] = D^2[\lambda^\alpha Z_\alpha'] = \lambda^\alpha \lambda^\beta [- \gamma_{\alpha\beta} + \gamma_{\alpha y_o} + \gamma_{\beta y_o}]$$

Now since $\sum \lambda^\alpha = 0$, it is easy to show that

$$D^2[M_x'^{*}] = - \lambda^\alpha \lambda^\beta \gamma_{\alpha\beta} \tag{3.20}$$

The system of equations can be found using Lagrange multipliers :

$$\left.\begin{aligned} - \lambda^\beta \gamma_{\alpha\beta} + \mu_o' + \mu^\ell \varphi_\alpha^\ell &= 0 \qquad (\ell \geq 0) \quad [O] \\ \sum \lambda^\alpha &= 0 \qquad [A] \\ \lambda^\alpha \varphi_\alpha^\ell &= \varphi_x^\ell \qquad (\ell \geq 1) \quad [U] \end{aligned}\right\} \tag{3.21}$$

If we replace φ_x^ℓ by $f_x^\ell - f_{y_o}^\ell$ and let us denote $\mu_o' - \mu^\ell f_{y_o}^\ell$ by μ_o , this system becomes :

$$\left.\begin{aligned} - \lambda^\beta \gamma_{\alpha\beta} + \mu_o + \mu_\ell f_\alpha^\ell &= 0 \qquad [O] \\ \sum \lambda^\alpha &= 0 \qquad [A] \\ \lambda^\alpha f_\alpha^\ell &= f_x^\ell - f_{yo}^\ell \qquad (\ell \geq 1) \quad [U] \end{aligned}\right\} \tag{3.22}$$

These equations make it possible to estimate the <u>increment</u> $m'(x) = m(x) - m(y_o)$. It is easy to show that the variance of $M_x'^* = \lambda^\alpha Z_\alpha$ is equal to

$$D^2[M_x'^*] = - \mu_\ell [f_x^\ell - f_{y_o}^\ell] \qquad (3.23)$$

3.4 ESTIMATING THE COEFFICIENTS OF THE DRIFT

3.4.1 Preliminary Remark.

The system for estimating the drift was rather unsatisfactory in that it was not possible to estimate the drift itself. This was because one of the components of the drift, a_o, could not be estimated. We now wish to check that a_o is the only such component and that the others can be estimated without any additional problems.

3.4.2 Setting up the System.

As in the stationary case (cf § 2.4.2) the form of the system shows that the solutions for λ^α and μ_ℓ are linear combinations of a family of functions in x. In the present case, these functions are the differences $f_x^\ell - f_{y_o}^\ell$ (for $\ell \geq 1$).

So there are two unique matrices λ_ℓ^α and $\mu_{\ell s}$ which satisfy

$$\left.\begin{aligned} \lambda^\alpha &= \lambda_\ell^\alpha (f_x^\ell - f_{y_o}^\ell) && \ell \geq 1 \\ \mu_s &= \mu_{s\ell} (f_x^\ell - f_{y_o}^\ell) && s \geq 0 \end{aligned}\right\} \qquad (3.24)$$

It is easy to obtain the equations that the λ_ℓ^α and the $\mu_{\ell s}$ satisfy (by substituting the results (3.24) into (3.21) and identifying the various functions in x, as was done in § 2.4.2).

This gives :

$$\left.\begin{aligned} -\lambda_{\ell}^{\beta}\,\gamma_{\alpha\beta} + \mu_{\ell s}\, f_{\alpha}^{s} &= 0 \qquad && \ell \geq 1 \\ \lambda_{\ell}^{\alpha}\, f_{\alpha}^{s} &= \delta_{\ell}^{s} && s \geq 0 \end{aligned}\right\} \qquad (3.25)$$

3.4.3 Estimating the coefficients a_ℓ.

If the objective is to estimate the coefficients of $m'(x)$, the problem is exactly analogous to the stationary case, and it is easy to show (as in § 2.4.3) that

$$A_{\ell}^{*} = \lambda_{\ell}^{\alpha}\, Z_{\alpha} \qquad (3.26)$$

is the optimal universal estimator of the coefficient a_ℓ for $\ell \geq 1$. All the coefficients of $m'(x) = a_\ell\,(f_x^\ell - f_{y_o}^\ell)$ can be estimated using this method since a_o does not appear in the expression for $m'(x)$.

But in fact we are interested in the drift $m(x)$ itself and not in its increments $m'(x)$. Now

$$m(x) = a_\ell\, f_x^\ell \qquad \ell \geq 0$$

and the case $\ell = 0$ is not taken into account in equation (3.25) . It is always possible to "complete" the equation by adding the index $\ell = 0$. The system will definitely be non-singular and will thus produce a set of coefficients λ_o^α . Unfortunately, if we let

$$A_{o}^{*} = \lambda_{o}^{\alpha}\, Z_{\alpha} \qquad (3.27)$$

then A_o^* is not an ALC. Consequently, in the model we have no right to talk about its mean and even less about its variance. So it is meaningless to call A_o^* the estimator of a_o. However we will see later the meaning that can be attributed to this variable.

3.5 SEVERAL PROPERTIES AND THE CONCLUDING REMARKS

3.5.1 Generalization of the results in the stationary case.

Except for the difficulties involving the term A_o , all the properties found for the stationary case also hold here. In particular this is true for

- the tensorial invariance (see § 2.5.1)
- the additivity theorem (see § 2.5.4)

There are no new theoretical problems in developing the equations for cokriging for the case where only the variogram exists.

3.5.2 Conclusion.

In this chapter we have stressed the insoluble problems of indeterminacy in the estimation of the coefficient a_o when the intrinsic hypothesis (in the narrower sense) holds. In many ways, this indeterminacy is the price that has to be paid in order to be able to develop the UK equations under a less restrictive hypothesis than the stationary one.

However, from the practical point of view, this is disturbing, even if it affects only the estimation of the drift and not the UK itself. The drift (taken in the intuitive sense) is just as well defined at sample points as are the sample values themselves ; in any case, a user would be appalled to have to accept that the "trend" or the "general tendency" is only known up to an additive constant.

This difficulty comes from the fact that the mathematical definition of the drift that has been given here (which is, of course, the thing being estimated) is not identical with the intui-

- CHAPTER 4 -

THE INDETERMINACY IN UNIVERSAL KRIGING

4 - THE INDETERMINACY IN UNIVERSAL KRIGING

4.1 PRELIMINARY REMARK

In the last two chapters we touched on two problems which have to be solved if universal kriging is to be put into practice.

- Up till now we have always assumed that the covariance function (or the variogram) was known. In section 2.3.5 we discussed the apparently intractable problem of structural analysis.

  In addition to this, in chapter 3, we raised another problem : in the intrinsic case, even if the drift could be estimated before γ, it would be known only up to an additive constant. We can ask what consequences this would have on the calculation of the experimental variogram.

  In any case the "real" drift (the one in the model) will never be known. We need to know if substituting an estimate for the real drift is going to cause any additional problems.

- When the intrinsic hypothesis holds, the estimation of the drift is subject to an unsurmountable problem of indeterminacy. What meaning can be attributed to this ? Or, in other words, does this indeterminacy cast any light on the relationship between the mathematically defined drift and its intuitive equivalent ? The drift in the model has the advantage of being rigourously defined, but does not necessarily correspond to anything real ; the user's concept of drift is clearly related to reality but cannot be used as a rigourous formalism.

In this chapter we intend to treat the second point first, and then to extend the meaning of the universality conditions. Afterwards we shall develop a possible method of structural analysis, which will highlight the inherent indeterminacy in the model and the consequences of this indeterminacy on the estimators found in the preceding two chapters.

4.2 THE MEANING OF THE INDETERMINACY IN THE DRIFT

4.2.1 Preliminary Remark.

In section 2.1 we defined the drift as being the mean of $Z(x)$

$$m(x) = E[Z(x)] \tag{4.1}$$

In addition we assumed that it could be expanded in the form

$$m(x) = a_\ell \, f_x^\ell \tag{4.2}$$

These two steps have totally different objectives :

- The definition is there to provide a rigourous meaning to the idea of a drift and to allow us to use it in the theory. It is neither true nor false. It is merely a choice which may prove to be fruitful or otherwise. What is more, it is certainly not the only possible one.

- The hypothesis (4.2) is designed to bring the definition (4.1) closer to our intuitive idea of a drift (which is generally associated with the idea of regularity, of trend, of low frequency, etc..), by constraining it to be regular. But this time the hypothesis can be rejected experimentally (e.g. if the basis functions f_x^ℓ are badly chosen, or if there are not enough of them, etc..).

In other words, if we want to develop a new presentation for the drift, we can change the definition (4.1), provided that this turns out to be more useful, but we will always need some sort of condition like (4.2) if we want the new drift to be closely related to our intuitive ideas.

4.2.2 Hypothesis of a random drift.

The preceding remark suggests an idea for a new model. This time the R.F. $Z(x)$ is split into two components :

$$Z(x) = Y(x) + M(x) \tag{4.3}$$

where $E[Y(x)] = 0$ and $M(x)$ is itself a random function.

In the first instance, we shall not worry about the practical problems of fitting a model of this type. We shall limit ourselves to the theoretical consequences of this choice. In fact, nothing prevents us from considering the drift as a realization of a random process. This choice may turn out to be unproductive but it cannot be proved to be true or false. In any case, to maintain the link with the intuitive idea of a drift, we need to introduce a regularity condition of the form :

$$M(x) = A_\ell \, f_x^\ell \tag{4.4}$$

where the A_ℓ are random variables this time. This condition embodies the idea that the decomposition into components (4.3) is not arbitrary. On the contrary, the components $Y(x)$ and $M(x)$ incorporate structures which operate at significantly different scales.

To complete the model we let

$$\left.\begin{aligned} E[Y(x)\ Y(y)] &= K(x,y) \\ E[m(x)\ m(y)] &= D(x,y) \\ E[Y(x)\ m(y)] &= R(x,y) \end{aligned}\right\} \tag{4.5}$$

We assume that the functions D and R vary slowly enough in space so that each can be expressed by the sum of the first few terms of its development in terms of the basis functions f_ℓ. Let

$$\left.\begin{aligned} D(x,y) &\simeq D_{\ell s}\, f_x^\ell\, f_y^s \\ R(x,y) &\simeq R_{\ell s}\, f_x^\ell\, f_y^s \end{aligned}\right\} \qquad (4.6)$$

In fact, the first of these approximations follows from (4.4), if we put $D_{\ell s} = E[A_\ell\, A_s]$. The second is a slightly stronger type of regularity condition.

Naturally, these hypotheses have an influence on the form of the functions D and R, but the value of the coefficients $D_{\ell s}$ and $R_{\ell s}$ is still unknown, as is the distribution of the random variables A_ℓ.

4.2.3 Estimation of a Random Drift.

Suppose that we want to estimate the drift M(x) by a linear combination of the data :

$$M^*(x) = \lambda^\alpha\, Z_\alpha = \lambda^\alpha\, Y_\alpha + \lambda^\alpha\, M_\alpha \qquad (4.7)$$

The mean of the square of the error is equal to

$$E[M^*(x) - M(x)]^2 = E[\lambda^\alpha\, Y_\alpha + \lambda^\alpha\, M_\alpha - M_x]^2$$

We expand this using equation (4.5) and then apply the approximation (4.6) where possible. This gives :

$$\begin{aligned} E[M^*(x) - M(x)]^2 = \lambda^\alpha\, \lambda^\beta\, K_{\alpha\beta} &+ 2\, \lambda^\alpha\, f_\alpha^\ell\, R_{\ell s}(\lambda^\beta\, f_\beta^s - f_x^s) \\ &+ D_{\ell s}(\lambda^\alpha\, f_\alpha^\ell - f_x^\ell)(\lambda^\beta\, f_\beta^s - f_x^s) \end{aligned} \qquad (4.8)$$

This expression cannot be minimized for arbitrary values of λ^α because the coefficients R_{ℓ_s} and D_{ℓ_s} are not known. The minimization can only be done by adding constraints which effectively eliminate the unknown coefficients. These constraints are just

$$\lambda^\alpha f_\alpha^\ell = f_\alpha^\ell \tag{4.9}$$

that is, the usual universality conditions.

Given these conditions, the expression (4.8) becomes

$$E[M^*(x) - M(x)]^2 = \lambda^\alpha \lambda^\beta K_{\alpha\beta} \tag{4.10}$$

Applying these conditions to the formula for the mean gives

$$E[M^*(x) - M(x)] = 0 \tag{4.11}$$

So, the system for estimating a random drift is exactly the same as in chapter 2. Given the hypothesis (4.6), a random drift can be treated in exactly the same way as a deterministic (functional) drift.

4.2.4 Estimating a moving average.

In this section, we shall not go into the details of the development which is not difficult (see [1], § 4.4.3). We shall concentrate on trying to understand the results.

4.2.4.1 Setting up the system.

Let Z be an intrinsic random function having γ for its variogram. Let Z have a zero drift ; that is, $\sum \lambda^i = 0 \Rightarrow E[\lambda^i Z(x_i)] = 0$.

Let p be a very regular weighting function with its sum equal to 1. Suppose that its range is large compared

to the zone under study. We define

$$\left.\begin{aligned} M(x) &= Z * \check{p} = \int Z(x+y)\ p(y)\ dy \\ Y(x) &= Z(x) - M(x) \end{aligned}\right\} \qquad (4.12)$$

M(x) is a weighted average of Z(x) and, as such, can be estimated by UK. We know that the UK equations do not present any problems of indeterminacy (see § 2.3.2 and 3.2.4). This can also be proved directly from the decomposition

$$Z(x) = Y(x) + M(x) \qquad (4.13)$$

where M(x) is a random drift. The function p satisfies the required regularity conditions (from their definition) so we can apply the results from the preceding section.

In particular, the random function Y is an ALC with a stationary covariance function $K_{xy} = K(x-y)$. Applying the formulae for calculating a regularized variogram and taking account of the regularity of p, it is easy to show that K_{xy} is of the form :

$$K_{xy} = -\gamma_{xy} + P_{\ell s}\ f_x^\ell\ f_y^\ell \qquad (4.14)$$

The expression to be minimized in order to estimate M(x) is equal to $\lambda^\alpha\ \lambda^\beta\ K_{\alpha\beta}$. After applying the universality conditions, this can be written as

$$\lambda^\alpha\ \lambda^\beta\ K_{\alpha\beta} = K(o) - \lambda^\alpha\ \lambda^\beta\ \gamma_{\alpha\beta} \qquad (4.15)$$

Summarizing, the system for estimating M(x) is

$$\left.\begin{aligned} -\lambda^\beta\ \gamma_{\alpha\beta} + \mu_\ell\ f_x^\ell &= 0 \\ \lambda^\alpha\ f_\alpha^\ell &= f_x^\ell \end{aligned}\right\} \qquad (4.16)$$

The corresponding variance is

$$D^2[M_x^* - M_x] = K(o) - \mu_\ell f_x^\ell \qquad (4.17)$$

4.2.4.2 Comparison with kriging.

We have seen that M(x) can be estimated directly by UK, since it is a weighted average of Z(x). The UK system is then :

$$\left.\begin{aligned} -\lambda^\beta \gamma_{\alpha\beta} + \mu_o &= \int \gamma(y)\, p(y-x+x_\alpha)\, dy \\ \Sigma \lambda^\alpha &= 1 \end{aligned}\right\} \qquad (4.18)$$

There is no universality condition here because Z has been assumed to satisfy the intrinsic hypothesis with zero mean.

Comparing the two systems (4.16) and (4.18) is particularly instructive :

- The second point of view does not involve any simplifying approximations. The system (4.18) is true for any weighting function p. But if this function p satisfies the hypothesis stated in § 4.2.4.1 (in particular if it has a large range), then the values of γ for large distances will be needed to estimate M(x) since $\gamma(y)$ appears in the right hand side of the system. In general the variogram values are known reasonably accurately only for short distances. Consequently, although the system (4.18) is quite satisfactory from the theoretical point of view, it turns out to be impracticable.

- In contrast to this, we only need the values of $\gamma_{\alpha\beta}$ (i.e. the variogram values for short distances) for the random drift model. Considering the data which is actually available, this is probably a better estimator of a large moving average. Consequently the system

(4.16) requires certain simplifying hypotheses but nevertheless allows us to make the estimates under more reasonable conditions from a practical point of view.

4.2.4.3 Properties of the system.

Going back to the system (4.16) for estimating a random drift M(x) :

$$\left.\begin{aligned} -\lambda^{\beta}\,\gamma_{\alpha\beta} + \mu_{\ell}\, f_x^{\ell} &= 0 \\ \lambda^{\alpha}\, f_{\alpha}^{\ell} &= f_x^{\ell} \end{aligned}\right\} \qquad (4.16)$$

This system does not depend upon the weighting function p. It is therefore the same as that for estimating $M_x^* = \lambda^{\alpha} Z_{\alpha}$. So this estimator will be the same for all moving averages provided that the weighting function p satisfies the conditions listed in § 4.2.4.1. However the variance (4.17) of this estimator does depend upon p, since the term K(o) appears in the formula.

Even if there is no "real" drift (i.e. in the model) the optimal estimator of the drift is still meaningful : it gives an estimate of a large-scale moving average. But it does this without specifying precisely which average is being considered since they all lead to the same estimator.

It can be proved that the variance of this estimator increases as the zone for the moving average enlarges.

If we put

$$M(x) = A_{\ell}\, f_x^{\ell} \quad ; \quad M^*(x) = A_{\ell}^*\, f_x^{\ell}$$

then it is easy to show that $A_{\ell}^* = \lambda_{\ell}^{\alpha} Z_{\alpha}$ where the λ_{ℓ}^{α} are solutions of the system :

$$\left.\begin{aligned} -\lambda_{\ell}^{\beta}\,\gamma_{\alpha\beta} + \mu_{\ s}\, f_{\alpha}^{\ell} &= 0 \\ \lambda_{\ell}^{\alpha}\, f_{\alpha}^{s} &= \delta_{\ell}^{s} \end{aligned}\right\} \qquad (4.19)$$

(this result can be obtained in exactly the same way as was used in 2.4 and 3.4)

This ultimately leads to

$$\left.\begin{aligned} \mathrm{Cov}[A_{\ell}^{*} - A_{\ell},\ A_{o}^{*} - A_{s}] &= \mu_{\ell s} \qquad (\ell \neq 0 \text{ or } s \neq 0) \\ D^{2}[A_{o}^{*} - A_{o}] &= \mu_{oo} + K(o) \end{aligned}\right\} \qquad (4.20)$$

Except for μ_{oo}, the Lagrange multipliers represent the covariances between $(A_{\ell}^{*} - A_{\ell})$ and $(A_{s}^{*} - A_{s})$. The variance of $A_{o}^{*} - A_{o}$ is the only one which depends on the weighting function p.

4.2.5. The Constant a_o.

Consider the R.F.

$$Z(x) = Y_1(x) + m_1(x)$$

where $m_1(x)$ is a drift (either random or deterministic) and where the covariance of Y_1 does not exist. Let p be a weighting function which is resonably regular and has a large enough range. (Its sum is taken as 1).

Z(x) can be expressed in the form :

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

where

$$Y = Y_1 - Y_1 * p \quad ; \quad M = m_1 + Y_1 * p$$

and $M(x)$ can be expanded as $A_\ell \, f_x^\ell$.

Given this new decomposition, Y has a covariance because it is an ALC. So by construction, we can apply the formulae from section 4.2.4. The optimal estimator of the random drift M is $M^*(x) = A_\ell^* \, f_x^\ell$ where $A_\ell^* = \lambda_\ell^\alpha \, Z_\alpha$ and the λ_ℓ^α are the solutions of the system (4.19). In particular, this gives an estimator A_o^* which does not depend on the function p.

Now the system (4.19) is no other than the one obtained for the intrinsic hypothesis with a deterministic drift after the addition of the equations for $\ell = 0$ (see (3.25)). We can now justify the manipulations carried out there, which lead to a non-admissible estimator. Lastly, the expression $A_\ell^* \, f_x^\ell$ gives an estimate of the true drift after correction for a certain moving average. Its variance depends on the choice of the moving average. But since the estimator itself does not depend on this, it is reasonable to estimate all the terms A_ℓ^* using this set of equations (4.19).

4.3 INDETERMINACY IN THE UNDERLYING VARIOGRAM.

4.3.1 The Problem.

In the preceding two chapters we assumed that the variogram (or the covariance function) was known. The methodological problems involving the estimation of the variogram have already been raised in section 2.3.5 but in reality the problem is more fundamental. The variogram which appears in the system for the estimators is the variogram of the true residuals of Z(x) (i.e. the variogram of $Z(x) - E[Z(x)]$. This will be called the underlying variogram in the following sections. Clearly, there is no way of calculating this from the experimantal values since the true values of $E[Z(x)]$ are unknown. However if we know how to estimate the drift

$M(x)$, then the estimated residuals $Z(x) = M^*(x)$ can be calculated, and hence the variogram. The question is to know what is the relation between the underlying variogram and the variogram of residuals. We then need to know how to get from one to the other given only a single realization.

In this section, we shall concentrate more on understanding the results than on the details of the proofs.

4.3.2 The nature of the indeterminacy.

Consider a R.F. $Z(x) = Y(x) + a_\ell f_x^\ell$ where $a_\ell f_x^\ell$ is a deterministic (functional) drift. Let σ_{xy} be its covariance function. Let $M(x)$ be a very regular random function which is independent of Y and which has a covariance function K_{xy} that can be expanded as follows :

$$K_{xy} = K_{\ell s} f_x^\ell f_y^s$$

Then $Z'(x) = Z(x) + M(x)$ has an underlying covariance σ'_{xy} :

$$\sigma'_{xy} = \sigma_{xy} + K_{\ell s} f_x^\ell f_y^s$$

and a drift

$$a_\ell f_x^\ell + M(x) = (a_\ell + B\ell) f_x^\ell$$

The a^ℓ are fixed unknown coefficients. The $B\ell$ are random variables having a covariance matrix $K_{\ell s}$.

Since only one realization of the R.F. Z' is available, there is no way of estimating the distribution of the $B\ell$. All we have are the numerical values so it would be futile to consider them as random variables with an unknown distribution. Consequently there is no way of calculating the $K_{\ell s}$ from the experimental values ;

nor is there any way of separating the a_ℓ and the B. So the B_ℓ must also be treated as numbers. But then the covariance of $Z'(x)$ in the model is σ_{xy} and not σ'_{xy}. Consequently, given only one realization of $Z'(x)$ there is no way of choosing between the two covariances.

4.3.3 The form of the indeterminacy and its consequences.

In the case where an underlying covariance exists, it can be shown that any function τ satisfying

$$\tau_{\alpha\beta} = \sigma_{\alpha\beta} + T_{\ell s} f^{\ell}_{\alpha} f^{s}_{\beta} \qquad (4.21)$$

where $T_{\ell s}$ is an arbitrary symmetric matrix, is an admissible covariance (see [1], § 4.6.3 for proof).

The indeterminacy is fundamental. It means that a particular set of experimental results could come from any one of a wide range of underlying models and that it is impossible to choose among them unless additional hypotheses are made.

Moreover, it can be shown that the estimates obtained are not affected by this indeterminacy. The supplementary hypothesis made to circumvent this problem only influences the estimation variance.

The equation (4.21) can be adapted for the case where there is only an underlying variogram. This then provides the following problem in structural analysis. Given the available numerical values, how can we fit the essential parameters for the underlying model ? Secondly, which additional hypotheses would allow us to resolve the problems of indeterminacy ?

4.3.4 The Variogram of Residuals.

We now put ourselves in the usual situation of a real case study. Not only is the variogram of residuals unknown but so are the optimal estimates of the residuals since we would need to know the underlying variogram to find these.

So, in practice, the only thing that we can use is a universal estimator of the drift (but unfortunately not the optimal one).

Let this estimator be

$$M^*(x) = A_\ell^* \, f_x^\ell \tag{4.22}$$

We can then calculate the (non-optimal) estimated residuals

$$\begin{aligned} R^*(x) &= Z(x) - M^*(x) \\ &= Z(x) - A_\ell^* \, f_x^\ell \end{aligned} \tag{4.23}$$

It should be noted that the awkward problem of estimating A_o does not arise since this term disappears when the variogram of residuals is being calculated.

Of course the form of the variogram depends on the type of estimator chosen for the drift. However the relationship between this variogram and the underlying one can be established quite generally.

Let S be the set of sample points (i.e. the points where it is possible to evaluate the residuals experimentally). Let $k(x)$ be the indicator function for S (i.e. $k(x) = 1$ if $x \in S$; $k(x) = 0$ if $x \notin S$). Lastly let

$$K(h) = \int_S k(x) \; k(x+h) \; dx$$

be the measure of the intersection between S and S after a translation of h. The experimental variogram of residuals is then

$$2\,\gamma^*(h) = \frac{1}{K(h)} \int_S k(x)\, k(x+h)\, [R^*(x+h) - R^*(x)]^2\, dx \qquad (4.24)$$

(This formula merely shows how to evaluate an experimental variogram when the data are in a continuous 1-D space). The problem is to know whether this experimental variogram is an estimator of the underlying variogram γ, or not. To test this, we see whether its mean is γ. We substitute the expression for the residuals (4.23) into (4.24) and then take the square. This leads to a 3-term expression :

$$E[2\,\gamma^*(h)] = T_o - 2\,T_1 + T_2 \qquad (4.25)$$

where
$$T_o = \frac{1}{K(h)} \int_S k(x)\, k(x+h)\, E[Z(x+h)-Z(x)]^2\, dx$$

$$T_1 = \frac{1}{K(x)} \int_S k(x)\, k(x+h)\, E[A^*_\ell\, (Z(x+h)-Z(x))(f^\ell(x+h)-f^\ell(x))\, dx$$

$$T_2 = \frac{1}{K(h)} \int_S k(x)\, k(x+h)\, [f^\ell(x+h)-f^\ell(x)][f^s(x+h)-f^s(x)]\, Cov[A^*_\ell, A^*_s]\, dx$$

The term T_o is just $2\,\gamma(h)$. The other two terms can be expressed in terms of γ if we take account of the form of the estimator of the drift. They represent an unavoidable bias. The mean of the experimental variogram of residual is therefore not equal to the underlying variogram.

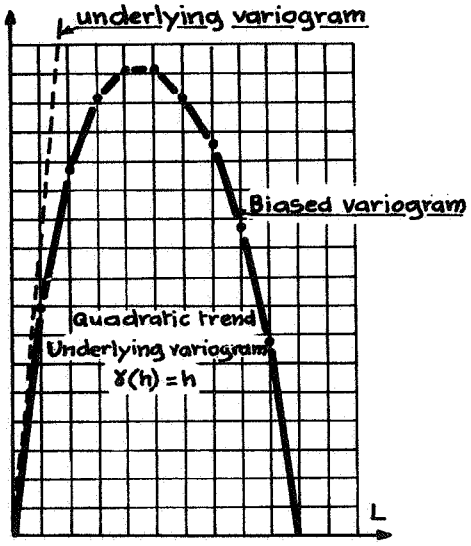
This situation is not caused by the fact that a non-optimal estimator of the drift is being used, since it is easy to show that even if this was available (which is unrealistic) equation (4.25) would be

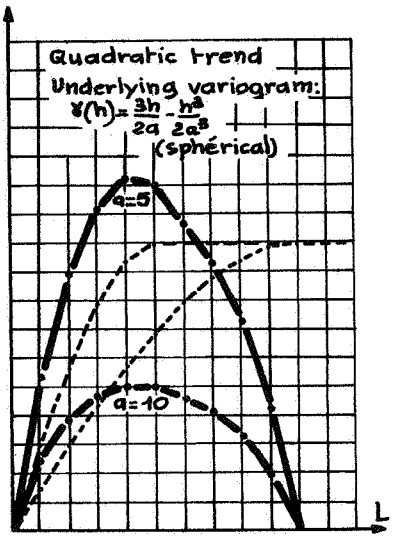
$$E[2\,\gamma^*(h)] = T_o - T_2 \qquad (4.26)$$

where T_2 is always positive.

4.3.5 Practical aspects of the Bias and the Indeterminacy.

Applying equation (4.25) to the common variogram models shows the importance of the bias term. In fact, the variogram of residuals bears no resemblance to the underlying variogram. The figures below show this for several simple cases :





This effect is particularly treacherous since an unsuspecting user could easily equate the variogram of residuals with the underlying variogram and thereby come to the entirely false conclusion that the residuals were independent. The real nature of the phenomenon under study would be completely hidden.

For simplicity's sake, suppose that the data set that we are working with, is discrete and on a regular grid. Suppose that the variogram of residuals $\gamma^*(h)$ has been calculated for a great many sliding neighbourhoods, for the integer values of h from 1 to N. The average value $\Gamma(h)$ of all these variograms could reasonably be considered as being equal to $E[\gamma^*(h)]$, and we therefore know the relation between $\Gamma(h)$ and the underlying $\gamma(h)$. Since the distance h must be an integer, the relation (4.26) can be written as

b) Once the underlying variogram has been selected, we can go on to find one optimal variogram of residuals for this model from the experimantal values. The optimal theoretical variogram of residuals can then be obtained using equation (4.26).

The test proposed in b) is very important since the first step is numerically unstable. The existing charts of the theoretical variograms of residuals show how similar these are. It is therefore difficult to make a choice between them.

The choice made by the user is decisive here as it was for the estimation of a_0. The user knows that whole classes of underlying variograms give rise to the same variogram of residuals. It is up to him to decide which model from these classes should be tested. Fortunately, the choice only affects the variance of the estimators and not their values. The only way to make this choice is to use arguments based upon a knowledge of the physics of the phenomenon.

REFERENCES

[1] MATHERON, G.- The Theory of Regionalized Variables and its Applications.- Cahiers du Centre de Morphologie Mathématique n° 5 (1970).

[2] CHILES, J.P., DELFINER, P., MATHERON, G.- Specialized Techniques of Geostatistics - Selected Papers (C-79, 1981).

[4] MATHERON, G.- Splines and Kriging : their formal equivalence. (N-667, 1980).

[5] " " Remarques sur le Krigeage et son dual. (N-695, 1981).

[6] " " Splines et Krigeage : Le cas fini. (N-698, 1981).

[7] DUBRULE, O.- Krigeage et Splines en Cartographie Automatique. (Thèse de Docteur-Ingénieur, Décembre 1981).

[8] " " Comparing "Plaque Mince" splines and kriging. (N-697, 1981).

[9] DELFINER, P.- The Intrinsic Model of Order k. (C-97, 1982).