Mathematical Geology, Vol. 16, No. 1, 1984

Problems with Universal Kriging[1]

M. Armstrong[2]

INTRODUCTION

Ordinary linear geostatistics is now well established, both theoretically and practically, as a method for estimating stationary phenomena. Its use is widely accepted in the mining industry. In other fields, however, such as meteorology, bathymetry and reservoir engineering, the variables under study are clearly not stationary.

In 1969, Matheron proposed the theory of universal kriging, which provided linear estimates of a variable even when a trend or drift was present. By 1973, however, this theory had been discarded in favor of the theory of intrinsic random functions of order k (IRF-k). Newcomers to geostatistics often wonder why. Some clues as to the reasons are given in the NATO papers (Guarascio, David, and Huijbregts, 1976). When introducing the then new IRF-k, Delfiner (p. 47) alludes to one of the problems: that of identifying the underlying covariance or variogram. In the same volume, Sabourin (p. 101–109) explores two possible methods for getting at the underlying covariance, but the methods seem rather cumbersome. To get some more detailed answers on this question, one has to go back to the Centre's internal reports (Matheron, 1969, 1971), which are mostly in French and are not easily accessible to outsiders.

The organizers of a recent summer school curriculum felt the need for a more comprehensive account of the problems encountered in UK. This led to the production of an internal note in English (Chauvet and Galli, 1982). The aim of this article is to provide a summary of that note and to indicate to readers other references on the subject.

DESCRIPTION OF UNIVERSAL KRIGING

A study of the sea floor by Journel (1969) provoked Matheron's interest in the problems of nonstationarity and led to the development of universal kriging.

---

[1]Manuscript received 8 December 1982.

[2]Centre de Geostatistique et de Morphologie Mathématique, École des Mines de Paris, 35 Rue St.-Honoré, 77305 Fontainebleau, France.

The approach parallels that of trend surfaces, where the phenomenon of interest is split into two components–a deterministic trend (or drift) plus a random error (or fluctuation, as the geostatisticians prefer to call it). The difference between the two approaches is that the fluctuations used in geostatistics are not assumed to be independent, as they are in trend surfaces.

In mathematical terms, the random function $Z(x)$, representing the variable under study, is split into a deterministic drift $m(x)$ plus a random function $Y(x)$ with zero mean

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

It is assumed that the drift can be expressed locally as a linear combination $\Sigma_l a_l f^l(x)$ of k known basis functions $f^l(x)$ (generally polynomials) with unknown coefficients a_l. The choice of the basis functions is important: the uniqueness of the solution to kriging systems depends on their being linearly independent on the sample points. For example, the linearly dependent functions (1, $\cos x$, $\sin x$, $\cos^2 x$, $\sin^2 x$) cannot be used, nor could the monomials x^n if the data are linear and if the degree of the drift is greater than or equal to 1. (See p. 29-30 of Chauvet and Galli, 1982, for more details).

In order to make any statistical inferences about $Z(x)$, some hypotheses must be made about the stationarity of $Y(x)$. Two cases are considered

1. *Stationary Hypothesis.* $Y(x)$ is assumed to be weakly stationary with covariance function K_{xy}

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

This is then assumed to depend only on $(x - y)$.

2. *Intrinsic Hypothesis.* The increments $Y(y) - Y(x)$ are assumed to be weakly stationary. Their variogram γ_{xy} which is defined as

$$2\gamma_{xy} = E[Y(y) - Y(x)]^2$$

is assumed to depend only on $x - y$.

For each of these two cases three estimation problems can be considered

(i) estimating the value $Z(x)$ at certain points
(ii) estimating the drift $m(x)$ at certain points
(iii) estimating the drift coefficient a_l at certain points

Case 1: Estimation Under the Stationary Hypothesis

Estimating the Value $Z(x)$

The objective is to find the linear combination $\Sigma\, \lambda_\alpha z(x_\alpha)$ of the sample values $z(x_1), \ldots, z(x_n)$, which is unbiased and which minimizes the estimation variance. Since the estimator has to be unbiased whatever the value of the coefficients, Matheron called it a "universal" estimator. A set of sufficient conditions for $E[\Sigma \lambda_\alpha z(x_\alpha)]$ to be equal to $E[z(x)]$ is that

$$\sum \lambda_\alpha f^l(x_\alpha) = f^l(x) \qquad \text{for} \quad l = 0, 1, \ldots, k$$

As in ordinary kriging, we use the Lagrange multipliers, μ_l ($l = 0, 1, \ldots, k$) to minimize the estimation variance

$$E\left[\sum \lambda_\alpha Z(x_\alpha) - Z(x)\right]^2 = \sum\sum \lambda_\alpha \lambda_\beta K_{\alpha\beta} - 2\sum \lambda_\alpha K_{\alpha x} + K_{xx}$$

subject to the preceding universality conditions. Setting the partial derivatives with respect to λ_α to zero gives rise to the following n conditions

$$\sum \lambda_\beta K_{\alpha\beta} + \sum \mu_l f^l(x_\alpha) = K_{\alpha x} \qquad \text{for} \quad \alpha = 1, \ldots, n$$

If we shorten $f^l(x_\alpha)$ to f^l_α, these n equations plus the k universality conditions can be written in matrix form as:

$$\begin{bmatrix} K_{\alpha\beta} & f^l_\alpha \\ f^l_\beta & 0 \end{bmatrix} \begin{bmatrix} \lambda_\beta \\ \mu_l \end{bmatrix} = \begin{bmatrix} K_{\alpha x} \\ f^l_x \end{bmatrix}$$

When the solution of this system is substituted back into the expression for the estimation variance, the kriging variance is obtained: $K_{xx} - \Sigma \lambda_\alpha K_{\alpha x} - \Sigma \mu_l f^l_x$. Provided that $k_{\alpha\beta}$ is strictly positive definite (which excludes the case of repeated sampling at sample points) the system is nonsingular and has a unique solution.

Estimating the Drift $m(x)$

The equations for estimating the drift are obtained in the same way. The kriging system obtained is

$$\begin{bmatrix} K_{\alpha\beta} & f^l_\alpha \\ f^l_\beta & 0 \end{bmatrix} \begin{bmatrix} \lambda^\beta \\ \mu_l \end{bmatrix} = \begin{bmatrix} 0 \\ f^l_x \end{bmatrix}$$

Estimating the Drift Coefficient a_s

In this case the kriging system becomes

$$\begin{bmatrix} K_{\alpha\beta} & f^l_\alpha \\ f^l_\beta & 0 \end{bmatrix} \begin{bmatrix} \lambda_{\beta,s} \\ \mu_{l,s} \end{bmatrix} = \begin{bmatrix} 0 \\ \delta^l_s \end{bmatrix}$$

Case 2: Estimation Under the Intrinsic Hypothesis

In this context, the term "intrinsic" is to be taken in the narrow sense: that is, intrinsic but not stationary. In that case, we can handle only admissible linear combinations ALC, $\Sigma \lambda_\alpha Z(x_\alpha)$ with $\Sigma \lambda_\alpha = 0$.

The same procedure is used as before, except that the covariance K_{xy} is replaced by $-\gamma_{xy}$.

Estimating the Value $E(x)$

The kriging system is now

$$\begin{bmatrix} -\gamma_{\alpha\beta} & f_\alpha^l \\ f_\beta^l & 0 \end{bmatrix} \begin{bmatrix} \lambda_\beta \\ \mu_l \end{bmatrix} = \begin{bmatrix} -\gamma_{\alpha x} \\ f_x^l \end{bmatrix}$$

Estimating the Drift $m(x)$

Since the drift itself is not an *ALC*, we can only estimate drift increments $m(x) - m(y_o)$. If we let ϕ_x^l denote $f_x^l - f_{y_o}^l$, the system for estimating the drift increments is

$$\begin{bmatrix} -\gamma_{\alpha\beta} & \phi_\alpha^l \\ \phi_\beta^l & 0 \end{bmatrix} \begin{bmatrix} \lambda_\beta \\ \mu_l \end{bmatrix} = \begin{bmatrix} 0 \\ \phi_x^l \end{bmatrix}$$

It is instructive to expand the drift increment

$$m(x) - m(y_o) = \sum_o^k a_l f^l(x) - \sum_o^k a_l f^l(y_o) = \sum_1^k a_l [f^l(x) - f^l(y_o)]$$

The constant term a_o is filtered out.

Estimating the Drift Coefficient a_s

The system for estimating any of the a_l except the constant term a_o is

$$\begin{bmatrix} -\gamma_{\alpha\beta} & f_\alpha^l \\ f_\beta^l & 0 \end{bmatrix} \begin{bmatrix} \lambda_{\beta,s} \\ \mu_{l,s} \end{bmatrix} = \begin{bmatrix} 0 \\ \delta_s^l \end{bmatrix}$$

Remarks

The properties of universal kriging (tensorial invariance and the additivity theorem) are presented in detail in Matheron (1969) and Chauvet and Galli (1982). Comparisons with least-squares and maximum-likelihood estimators are also discussed in these two references, as is universal cokriging. An interesting example of cokriging is given in Chauvet (1977).

We now go on to discuss two major problems in universal kriging.

FIRST PROBLEM: THE INDETERMINANCY IN THE DRIFT

As was mentioned in the preceding section, the constant term a_0 cannot be estimated when only the intrinsic hypothesis holds. We can ask if this problem can be overcome by some minor modification of the system or whether it is inherent in the methodology.

The universal kriging method is based on a decomposition of the phenomenon under study into a deterministic drift (or trend) plus correlated fluctuations. In Section 2 we defined the drift to be the mean of $Z(x)$. This definition was chosen to provide a rigorous meaning of the idea of drift and to allow us to develop a mathematical theory. It is neither true nor false. It is merely a choice that may prove to be fruitful or otherwise. What is more, it is not the only possible choice.

We also assumed that the drift could be expanded as $\Sigma a_l f^l(x)$. This hypothesis is designed to bring this definition closer to our intuitive idea of a drift by constraining it to be regular. But this time the hypothesis can be rejected experimentally (for example, if the basis functions were badly chosen or if there were not enough of them).

In other words, if we want to develop a new representation for the drift that would allow us to estimate the constant term a_o, we can change the definition of the drift but we will still require some sort of regularity condition.

In trend-surface analysis in classical statistics, the coefficients of the trend can be considered either as deterministic or as random variables. We can try to replace the deterministic drift coefficients by random variables. Clearly, some additional covariances between the various drift coefficients and between a given drift coefficient and the random function $Y(x)$ are needed. Because of the uniqueness of the data set, these cannot be estimated. However, they can be eliminated from the equations by introducing some additional universality conditions, but this brings us back to the same system of equations as before. See Matheron (1971, p. 173-183) or Chauvet and Galli (1982, p. 78-86) for more details.

The indeterminacy in the drift coefficient a_0 in the intrinsic case is one of the inherent indeterminacies of universal kriging. Working with linear combinations of differences, which we must do in the intrinsic case, effectively filters out the constant. This can be compared to the indeterminacy in integration. If we know only the derivative, the corresponding integral is known up to an additive constant. Going further, this comparison suggests that the higher powers in x might be filtered out by higher order differences, which is precisely the principle behind the IRF-k.

SECOND PROBLEM: INDETERMINACY IN THE UNDERLYING VARIOGRAM

The problem described in the preceding section is important from a theoretical point of view but does not condem the use of universal kriging. The problem discussed in this section has far more wide-reaching implications. It means that universal kriging can be used only in cases where the underlying variogram is known a priori.

When the universal kriging system was developed earlier under the section on universal kriging the variogram (or covariance) was assumed to be known. But this is rarely the case in practice. One exception to this was a case study on sea floor estimation done by Chilès (1977). As there was no drift in one direction, the variogram for this direction was taken to be the underlying variogram in the other directions too.

However if the underlying variogram is not known, we are faced with a chicken-and-egg problem:

(i) We need the (unknown) underlying variogram (or covariance) for the universal kriging system.

(ii) If we use instead the variogram of residuals

$$\gamma_R(h) = \tfrac{1}{2}E\{[Z(x+h) - m^*(x+h) - [Z(x) - m^{8}(x)]\}^2$$

we then need to know the estimated drift $m^*(x)$.

(iii) In order to calculate $m^*(x)$ we have to know the universal kriging system—back to square one.

At first sight it would seem to be possible to proceed iteratively, successively refining the estimates of both the variogram and the trend. Sabourin (Guarascio, David and Huijbregts, 1976) has presented two methods of doing this. But two more difficulties then arise.

First, the variogram of residuals gives a very biased estimate (an underestimate) of the true underlying variogram. Figure 1 (which is taken from Chilès' (1977) thesis) shows how bad the situation is. Equations for the bias term are given in Chauvet and Galli (1982, p. 90).

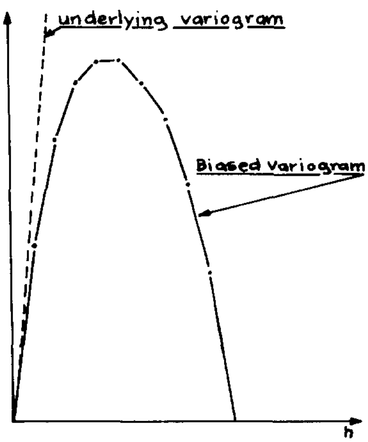

Fig. 1. Underlying variogram and the variogram of residuals, which is seriously biased.

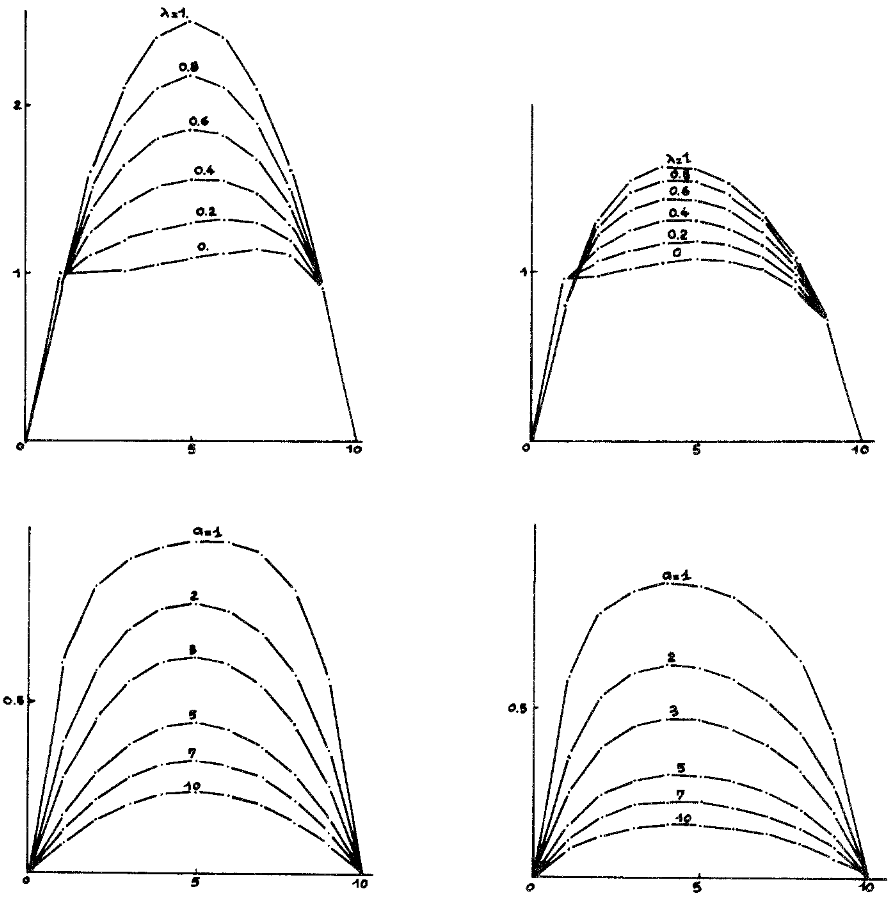

Fig. 2. Variogram of residuals corresponding to either linear or quadratic drifts and other exponential or power function variograms.

Second, it is extremely difficult to deduce either the degree of the drift or the type of the underlying variogram from the variogram of residuals. Figure 2 (taken from Chilès, 1977) shows the variogram of residuals for two different underlying variograms (exponential and power function) for both linear and quadratic drifts. Seeing these, we can appreciate the truth of Sabourin's comment that "this procedure can lead to false interpretations of $\gamma(h)$."

CONCLUSION

Although the theory of universal kriging is mathematically correct, the difficulty in estimating the variogram and the drift at the same time makes it unworkable in practice, except for those rare cases where the variogram is known a priori. The other problem (that of estimating the drift coefficient a_o) shows that there are inherent difficulties estimating some terms under the intrinsic hypothesis, which can only be overcome by changing our definition of a "drift." On the positive side, it suggests that it might be better to filter out the drift by taking higher order differences.

REFERENCES

Chauvet, P., 1977, Exemple d'application de la géostatistique non stationnaire: le cokrigeage: Proceedings, Symposium, the Mining Pribram in Science and Technique.

Chauvet, P. and Galli, A., 1982, Universal Kriging: Internal Report n° C-96, CGMM., 94 p.

Chiles, J. P., 1977, Géostatistique des Phénomènes non-stationnaires: Thèse de Docteur-Ingénieur, Université de Nancy I.

Guarascio, M., David, M., and Huijbregts, C., 1976, Advanced Geostatistics in the Mining Industry: NATO A.S.I., Rome, 1975. D. Reidel Publishing Co., Dordrecht, Holland, 461 p.

Journel, A. G., 1969, Rapport d'etude sur l'estimation d'une variable régionalisée: Internal Report n° N-156, CGMM.

Matheron, G., 1969, Le Krigeage Universel: Fascicule 1, Cahiers du CMM., 82 p.

Matheron, G., 1971, The Theory of Regionalized Variables and Its Applications (English Version): Fascicule 5, Cahiers du CMM., 211 p.