G. Matheron

Ecole Nationale Supérieure des Mines, Paris, France

Kriging, or Polynomial Interpolation Procedures?

A contribution to polemics in mathematical geology

Transactions, Volume LXX, 1967, pp. 240-244

ABSTRACT

Scrutinizing polemics between Dr. D. G. Krige and Prof. E. H. T. Whitten, it appears that the kriging procedure is valid for stationary or intrinsic random functions, whereas the polynomial interpolation procedure should be applied only in specific cases [an error e(x) superimposed to a regular phenomenon m(x)]. It is shown, by a numerical example, how striking but fallacious evidence for a real trend may occur as a result of purely random cumulative effects.

IN the proceedings of the Symposium on Mathematical Statistics and Computer Applications (1), we can read of a very instructive controversy between Dr. D. G. Krige and Prof. E. H. T. Whitten. The vigour, and sometimes the acrimony, of the debates, and the methodological importance of the arguments involved, indicate that there is a fundamental underlying problem. We could state it as follows: is it possible for mathematical geology to accede to the true dignity of a really rational and scientific discipline, or shall it be nothing but a purely mechanical application of computers?

The question was the following: in the Rand goldfield, which is the best method to use in order to estimate the gold content of an unexploited panel when knowing only the gold content of samples taken in its neighbourhood? Krige's approach is based on a precise probabilistic background; we must use as an estimator, he says, the conditional expected value of the panel's gold content, relative to the hypothesis that available samples are assuming given values. He stresses the fact that it is necessary to know exactly what a given estimator is estimating (here, the mean gold content of a minimal block measuring 100 by 100 ft), and adds that polynomial (or other) interpolation procedures fail to verify such a condition; with a least-squares fitted polynomial, $P(x)$, we do not know specifically what the value taken at point x is related to.

For practical purposes, instead of using a conditional expected value, Krige adopts a linear regression formula; i.e., the linear expression corresponding to the minimal estimation variance. Denoting the gold content of the sample that could be taken at a given point, x, by $f(x)$, and the panel to be estimated by S, the actual but unknown content of this panel is:

$$\frac{1}{s}\int_S f(x)\,dx \qquad \text{(Eq. I)}$$

We then take an estimator, $E(S)$, of the form:

$$E(s) = \sum_{i=1}^{n} \lambda_i f(x_i) \qquad \text{(Eq. II)}$$

where $x_1\ x_2, \ldots x_n$ denote points actually sampled. The λ_i coefficients minimize the estimation variance; they can be obtained by solving a linear system, the matrix of which depends only on the variogram (4).

Estimator (I) appears as a weighted moving average; indeed, it is not an arbitrary one, but the specific one that minimizes the estimation variance. It is known as the kriging estimator.*

Prof. Whitten did not agree with such an approach, and produced two sorts of arguments. His first objection is that, if calculated on the basis of an area of 100 by 100 ft, and estimator like (II) will tend to suppress all structural features relative to lower scales, and miss important geological information.

The first point does not seem to be founded. Obviously, the kriging estimator (II) may be fitted for estimating the punctual content, $f(x)$, at an arbitrary point x. We obtain:

$$E[\,f(x)\,] = \sum_{i=1}^{n} \lambda_i(x)\, f(x_i) \qquad \text{(Eq. III)}$$

with coefficients $\lambda_i(x)$ depending on x and minimizing the estimation variance for $f(x)$. The kriging estimator (II) itself appears as a moving average of the punctual kriging (III), exactly in the same way as the value of (I) is itself a moving average of $f(x)$. Indeed, it can easily be shown that the λ_i coefficients of Eq. II can be deduced from the $\lambda_i(x)$ of Eq. III by putting:

$$\lambda_i = \frac{1}{s}\int_S \lambda_i(x)\,dx$$

It is then clear that the punctual estimator (III) expresses in the best way all the available information. If microstructures are present, and may be approached on the basis of available samples, estimator (III) answers Prof. Whitten's first objection. For exploitation purposes, however, minimal-size panels are the only interesting features, so that it is practically necessary and theoretically correct to neglect such microstructures by using directly the kriging (II) estimator, which gives us the best estimation of what we precisely want to estimate.

Prof. Whitten's second objection is the following: local fluctuations (i.e., precisely those microstructures that Dr. Krige was formerly reproached to ne-

*This terminology is classical in France, since 1960. We kindly suggest to Dr. Krige's friends and colleagues to use it, instead of the really misleading expression "weighted moving average."

glect) are generally meaningless, and should be eliminated. Only regional trends are of interest. Each value of $f(x)$ may be represented as the sum:

$$f(x) = m(x) + e(x) \quad \text{(Eq. IV)}$$

where $m(x)$ is a very regular and continuous function expressing the trend, and $e(x)$ is a meaningless random fluctuation to be eliminated. Moving-average procedures fail to separate substance or essence — i.e., $m(x)$ — from accidents or appearance — i.e., $e(x)$. According to Prof. Whitten, only interpolation procedures can do this.

As a matter of fact, in physics or information theory, a model like (IV) is often useful, when an interesting phenomenon or a message is altered by a noise. Naturally, such a distinction between message and noise requires a serious theoretical background to be asserted *a priori*. When such a theoretical background is missing (as is generally the case in geology), distinction between $m(x)$ and $e(x)$, which could appear purely empirical, expresses nothing but ingenuous metaphysics, or perceptive illusion (see the example below). In geology, the so-called "trend" $m(x)$ has generally exactly the same stochastical character as noise itself, a distinction between $m(x)$ and $e(x)$ being only a matter of scales. In brief, such a distinction does not appear to be founded on a criterion that could be stated in a rational language.

Furthermore, when criticizing the moving-average procedure, Prof. Whitten did not see that the polynomial interpolation is itself a moving-average procedure.

Indeed, it consists of finding a function, $P(x)$, of the form:

$$P(x) = \sum_r a_r Q_r(x) \quad \text{(Eq. V)}$$

where $Q_r(x)$ are given functions (for instance, monomial functions when using polynomial interpolation). The coefficients a_r are determined by minimizing the quadratic form:

$$\sum_{i=1}^{n} [f(x_i) - \sum_r a_r Q_r(x_i)]^2$$

and can be obtained by solving the linear system:

$$\sum_{i,q} a_r Q_r(x_i) Q_q(x_i) = \sum_i Q_r(x_i) f(x_i)$$

Hence, the a_r coefficients are linearly related to the experimental values, $f(x_i)$. We may write:

$$a_r = \sum_i B_{r,i} f(x_i)$$

and, putting those expressions in Eq. V, we obtain:

$$P(x) = \sum_{i,r} B_{r,i} f(x_i) Q_r(x)$$

or:

$$P(x) = \sum_i \mu_i f(x_i) \quad \text{(Eq. VI)}$$

by putting $\mu_i = \Sigma_r B_{ri} Q_r(x)$. Hence, polynomial interpolation appears itself as a moving-average procedure. However, if, among all the possible choices for the weighting coefficients, kriging leads us to the minimal estimation variance, any other choice, including the interpolation formula (VI), will, of course, give a greater variance, and a less efficient estimation for what we want.

We may add the following remark: if one is given fourteen experimental points, it is always possible to find a 13-degree polynomial fitting them exactly. If, in Eq. V, the number of the Q_r functions equals the number, n, of available data, we will have exactly $P(x_i) = f(x_i)$ at each sampled point x_i. The fit is a perfect one. However, it is well known and universally admitted that such perfection is a fallacy; when increasing the degree for polynomial interpolation, tremendous and meaningless fluctuations are always appearing between interpolation points and, as a result, a pure artefact is obtained. However, the best estimator for gold content $f(x_i)$ at a sampled point x_i remains $f(x_i)$ itself, its true value being known. The impossibility of obtaining such a result in a realistic manner appears to represent a failure for interpolation methods. On the contrary, punctual kriging (III) always gives $E[f(x_i)] = f(x_i)$, as can easily be shown, the corresponding minimal estimation variance being null.

Besides, whereas polynomial interpolation was formerly appearing as a particular moving-average procedure, punctual kriging is now appearing as an exactly fitting interpolation procedure.

We could stop this discussion here. However, it may be useful to scrutinize the conceptual background of the arguments involved, and to give a numerical example to illustrate the perceptive illusions on which is based the common notion of "trend" — a very anthropomorphic term.

Conceptual background

The content, $f(x)$, of the standard sample taken at any point x, considered as an immediate manifestation of a natural phenomenon, constitutes what is called a *regionalized variable* — a neutral term, prior to all probabilistic interpretation — i.e., an ordinary function $f(x)$. Except in certain specific cases — where polynomial interpolation may be very useful — the spatial variability of such a regionalized variable is too complex and too erratic to be studied easily by the usual mathematical functional methods. To rise above such a multi-form chaos, a first step is accomplished by interpreting the regionalized variable as a realization of a random function $F(x)$. It is well known that a random function $F(x)$ may be defined as a random variable with an infinite number of components, each of these components corresponding to the values taken by $F(x)$ at each point x. Random function F and regionalized variable f are related in the same way as an ordinary random variable Y and a numerical value y (for instance, $y = 98$) obtained as a result of a particular experiment. We cannot say that our regionalized variable *is* a random function. Such an expression would be as inadequate as saying, for instance: the number 98 *is* a random variable. Experimental data are not identical to their conceptual model. Just as number 98 is obtained as a result of a single experiment carried out according to the probability law of Y, we must consider the values taken by $f(x)$ at all the points x as the result of a single trial carried out according to the law of the random vector $F(x)$, to which belong an infinite number of components.

This first conceptualizing step cannot be seriously objected to. It is admitted explicitly by Krige, and, more implicitly perhaps, by Whitten. It is not really a hypothesis, that may or may not be supported by experimental evidence, but only a conceptual background. Developing mathematical geology without such a background would be more strenuous.*

*Actually, such a background is convenient for theoretical purposes, but not strictly necessary. In (4), the writer has shown that the main geostatistical results may be deduced without any probablistic interpretation and, *a fortiori*, without any stationary or intrinsic hypothesis.

Real difficulties arise with statistical inference. We need to determine the spatial law of our random function; i.e., the set of all repartition laws:

$$G(f_1, f_2, \ldots; x_1, x_2, \ldots) = P[f(x_1) < f_1, f(x_2) < f_2 \ldots] \quad \text{(Eq.VII)}$$

for all finite sets of points $x_1, x_2, \ldots$ Knowing actually such laws, we could solve all possible problems, including the determination of the conditional expected value for panel contents. This is the real theoretical justification for kriging procedures.

Unfortunately, such a law as (VII) will depend on more than k parameters, when k points $x_1, x_2, \ldots x_k$ are involved — it will at least depend on k order-1 moments and $k(k+1)/2$ order-2 moments, when the data do not exceed the k numerical values $f(x_1), f(x_2) \ldots f(x_k)$, and when samples were actually obtained at points $x_1, x_2, \ldots x_k$. Hence, statistical inference is generally impossible, or indeterminate, unless we introduce some additional hypothesis to reduce the number of parameters involved. Then, however, the ensuing results will not be any more valid than the starting hypothesis itself.

Concerning such an hypothesis, we can at first assume that our random function is a stationary one — i.e., its spatial law is invariant by translation. Because the phenomenon is, then, in a sense, repeating itself throughout the whole spatial field, new opportunities arise for statistical inference. The order-1 moment reduces itself to a constant:

$$E[f(x)] = m$$

and the covariances:

$$K(x_i - x_j) = E[f(x_i)\, f(x_j)] - m^2$$

depend only on the vector $x_i - x_j$ and not on each of the points x_i and x_j separately.

We assume implicitly that our random function has actually an expected value m and a finite variance $K(0)$. Such an assumption may be questioned. In the Rand (3), for instance, size-s samples, in a mineralized area S, have a variance of $\sigma^2 = \alpha \ln(S/s)$, depending on both the sizes s and S. This variance increases indefinitely with S, and does not tend toward a finite limit $K(0)$, as it should do if such an *a priori* finite variance did exist.

It is possible to overcome this difficulty by using the word intrinsic instead of the word stationary, and using the variogram $\gamma(h)$ instead of the covariance $K(h)$. Indeed, $\gamma(h)$ is defined by:

$$\gamma(h) = \frac{1}{2} E[f(x+h) - f(x)]^2$$

and does exist, even if the *a priori* variance does not. Starting from a variogram or from a covariance, we shall obtain the same kriging equations, and the two procedures do not appear to be really distinct (4).

For the Rand goldfield, and many other orebodies and geological phenomena, such an intrinsic or stationary hypothesis seems very well supported by experimental evidence. This is indeed the only justification for it. As a matter of fact, the stationary character needs to be verified locally only. When kriging 100- by 100-ft blocks, it is sufficient to verify that the variogram is not altering itself in a 500- by 500-ft or 1000- by 1000-ft area. Such conditions being satisfied, kriging is really the best possible estimation procedure and polynomial interpolation appears only as an artificial by-product of computer business.

On the other hand, it is possible to introduce another kind of hypothesis, and it seems that Prof. Whitten is doing so. Let $m(x)$ be the order-1 moment:

$$E[f(x)] = m(x)$$

and let us assume that it is a continuous and regular function depending on point x, but not a constant; our random function is no longer stationary. Then, putting

$$e(x) = f(x) - m(x)$$

we again find model (IV), but with a definite interpretation for the trend $m(x)$. This "trend" now appears as the order-1 moment of a non-stationary random function. In order to reduce the number of parameters involved by statistical inference, we assume further that $e(x)$ is a *stationary* random function, with a null mean value and *without auto-correlation.* Polynomial or other interpolation procedures then appear to be fully adequate. Such a procedure may be used, for instance, when $f(x)$ values occur as results of measurements, with a random error $e(x)$ attached to a regular phenomenon $m(x)$. However, without such specific reasons, one could hardly admit that there was a stationary covariance accompanying a non-stationary mean, $m(x)$. Such a dichotomic scheme could find only a limited field of application.

An Example

In order to illustrate the preceding theoretical considerations, a numerical example will now be given. In a famous book by Feller (2), it has been shown that a random-walk (for instance, a Wiener-Lévy) process generates a striking appearance of a systematic trend; such a trend, however, is a fallacious one due only to a cumulative effect of independent random components. Following Feller, we show in *Figure 1* a curve obtained by plotting the cumulative gains of a player in a heads-or-tails game. In other words, denoting by S_n and F_n the number of successes and failures obtained at trial N°n in a series of independent trials, and putting:

$$f(n) = S_n - F_n$$

we are concerned with a realization of such a random function, taken from $n = 1$ to $n = 200$. Such a stochastic process is characterized by its stationary and independent *increments*; it is not a stationary random function, with a finite $K(0)$ variance, but just an intrinsic one. There is no covariance $K(h)$, but just a variogram:

$$\gamma(h) = \tfrac{1}{2}h$$

This variogram sums up all that we need to know about the relatively weak structure of such a phenomenon.

First of all, *Figure 1* shows a striking appearance of heterogeneity, with two different populations — from 1 to 100 and from 101 to 200. The corresponding histograms are given in Table I.

From a geostatistical point of view, the means $m_1 = 2.08$ and $m_2 = -8.26$ are *not* significantly different. With the variogram $\tfrac{1}{2}h$, it can be shown (4) that:

$$E(m_1 - m_2) = 0$$

$$D^2(m_1 - m_2) = \frac{2}{3}h$$

m_1 and m_2 being the observed mean values on two adjacent segments of length h.

With $h = 100$, we therefore have a variance $D^2(m_1 - m_2) = 66.67$, and the observed difference of 10.34 is supported by such a value.

Concerning the variances inside a segment length h, we have demonstrated the theoretical formulae $E(\sigma^2) = 1/6\ h$ and $D^2(\sigma^2) = 1/45\ h^2$. For $h = 100$,

Table I—Histograms for Figure 1

Values	Frequencies from 1 to 100	Frequencies from 101 to 200	Frequencies from 1 to 200
−14		4	4
−13		8	8
−12		8	8
−11		11	11
−10		16	16
− 9		14	14
− 8		7	7
− 7		2	2
− 6		3	3
− 5		7	7
− 4		7	7
− 3	1	3	4
− 2	7	2	9
− 1	10	3	13
− 0	12	2	14
1	15	2	17
2	12	1	13
3	10		10
4	12		12
5	12		12
6	6		6
7	2		2
8	1		1
Total	100	100	200
Mean	+ 2.08	−8.26	−3.09
Variance	6.37	15.15	37.49

Table II—Observed Values for Spacing 10

n	f(n)	n	f(n)	n	f(n)	n	f(n)
0	0	60	0	110	− 2	160	− 8
10	6	70	4	120	− 6	170	−12
20	4	80	4	130	− 6	180	−12
30	2	90	0	140	−10	190	−10
40	0	100	−2	150	−10	200	−10
50	−2						

we obtain E $(\sigma^2) = 16.67$ and $D^2(\sigma^2) = 222$. With such a value for $D^2(\sigma^2)$, the experimental values 6.37 and 15.15 are not in disagreement with the expected value 16.67. For $h = 200$, we obtain E $(\sigma^2) = 33.33$ and $D^2(\sigma^2) = 888$, versus an observed value of 37.49.

Undoubtedly, classical statistical tests of significance would conclude that the two populations are different; such a conclusion is not wrong, because the means, for instance, are really different from 1 to 100 and from 101 to 200. Such a difference, however, does not mean that a real heterogeneity does exist; for, having built it ourselves, we know perfectly well that our phenomenon was generated by the same process from 0 to 200, and hence this phenomenon is a strictly homogeneous one. However, the homogeneous law we have chosen is precisely such that its realizations will necessarily offer all evidences of an apparent heterogeneity.

In the same way, the systematic decrease occurring between point 75 and 150 will perhaps be interpreted as a real negative trend, although we know that such a trend means nothing but a cumulative random effect. In such a case, polynomial interpolation procedures would be very misleading in that they would support metaphysical beliefs concerning a real trend with the aid of fallacious experimental evidence. A real trend is missing, but — as shown by Feller (2) — the law we have chosen is necessarily generating such fallacious random trends for human perception.

For simulation purposes, let us assume that we know only the values taken at points 0, 10, 20, etc..., as shown in Table II.

We want to estimate true mean values between points 0 and 10, 10 and 20, and so on. For instance, between 10 and 20 the mean is:

$$1/10[\tfrac{1}{2}f(10) + f(11) + \ldots + f(19) + \tfrac{1}{2}f(20)] = 5.7$$

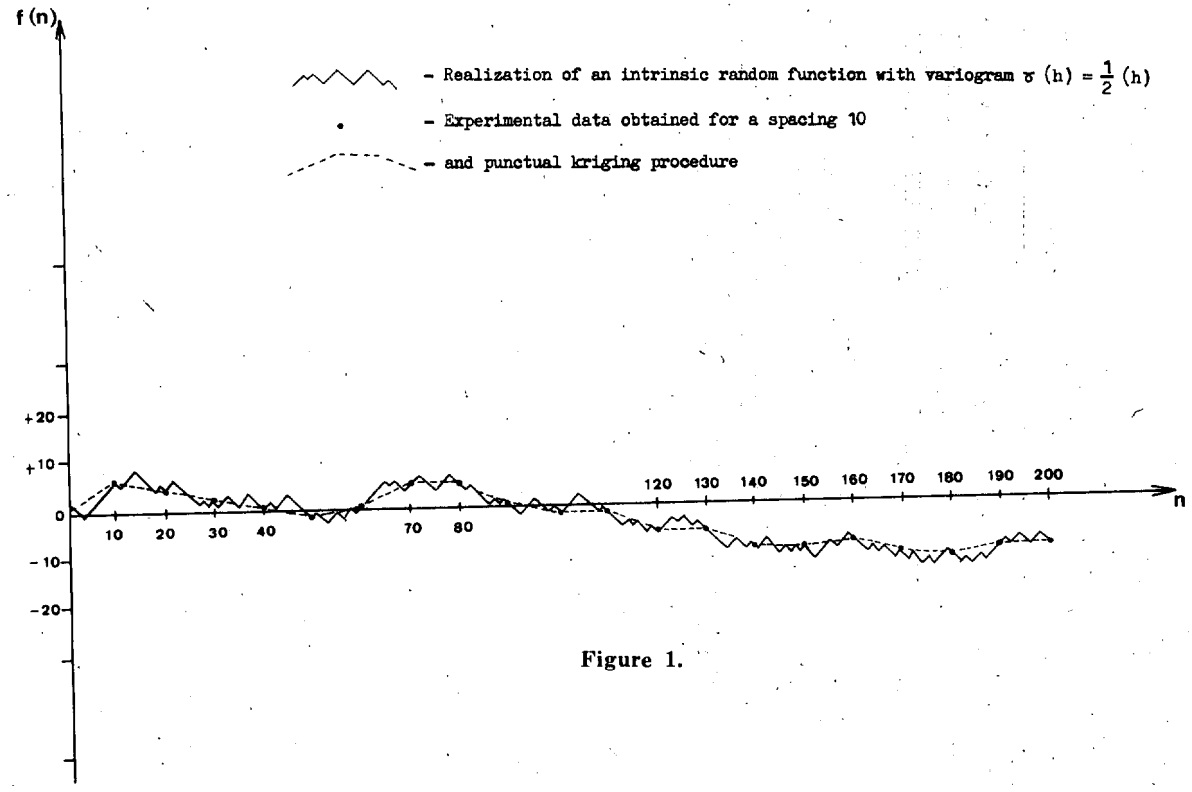

Figure 1.

With a linear variogram $\gamma(h) = \frac{1}{2} h$, punctual kriging is obtained by linear interpolation between the two nearest known values (see *Figure 1*), and kriging over an interval like 10 to 20 is obtained with $E(S) = \frac{1}{2}[f(10 + f(20)] = 5$ (versus a true value of 5.7). Such a procedure, as related to a linear variogram, may also be based on a markovian property of the process: once we know $f(10)$ and $f(20)$, all values $f(11)$, $f(12)$,... taken inside such an interval appear as independent from all outside values. The results of such a procedure are shown in Table III.

The experimental variance for the kriging estimator minus the true value is 0.968. The theoretical value for kriging a segment length h with a variogram $\frac{1}{2}h$ is $1/12h$; i.e., for $h = 10$, the variance is equal to 0.83.

The reader may then confront such an estimation variance with one obtained by any other procedures; for instance, by polynomial interpolation.

Table III—Kriging Procedure

Interval	True Value	Kriging Estimator
0 — 10	1.8	3.0
10 — 20	5.7	5.0
20 — 30	3.1	3.0
30 — 40	1.8	1.0
40 — 50	0.8	— 1.0
50 — 60	— 1.5	— 1.0
60 — 70	3.3	2.0
70 — 80	4.7	4.0
80 — 90	1.7	2.0
90 — 100	— 0.8	— 1.0
100 — 110	— 0.3	— 2.0
110 — 120	— 4.5	— 4.0
120 — 130	— 4.3	— 6.0
130 — 140	— 8.7	— 8.0
140 — 150	— 9.9	—10.0
150 — 160	— 9.6	— 9.0
160 — 170	—10.5	—10.0
170 — 180	—12.7	—12.0
180 — 190	—12.6	—11.0
190 — 200	— 9.1	—10.0

References

(1) *Symposium on Mathematical Statistics and Computer Applications in Ore Valuation*, South African Institute of Mining and Metallurgy, Johannesburg, 1966.

(2) Feller, W., "An Introduction to Probability Theory and its Applications," *J. Wiley and Sons*, Vol. I, 1957; Vol. II, 1966.

(3) Krige, D. G., "A Study of Gold and Uranium Distribution Patterns in the Klerksdorp Goldfield, *Geoexploration*, No. 4, pp. 43-53, 1966.

(4) Matheron, G., Les Variables Régionalisées et leur Estimation," *Masson et Compagnie*, Paris, 1965.

(Preprinted from The Canadian Mining and Metallurgical Bulletin, September, 1967)

Printed in Canada