Mathematical Geology, Vol. 22, No. 3, 1990

The Origins of Kriging[1]

Noel Cressie[2]

In this article, kriging is equated with spatial optimal linear prediction, where the unknown random-process mean is estimated with the best linear unbiased estimator. This allows early appearances of (spatial) prediction techniques to be assessed in terms of how close they came to kriging.

KEY WORDS: blue, blup, covariance function, geodesy, homogeneous structure function, meteorology, mining, optimum interpolation, spatial blup, statistics, variogram.

INTRODUCTION

A rapidly developing subject such as geostatistics often does not take the time to examine its origins. Such an examination can prove enriching, since nuggets of knowledge hitherto passed over can become apparent.

The purpose of this article is to discuss the origins of kriging, not only within geostatistics but also within other scientific disciplines. I hope that readers of this article will inform me of any omissions, remembering that I have restricted myself here to "public knowledge" obtained from *published* books and papers.

The use of the word "kriging" in spatial statistics has come to be synonymous with "optimally predicting" or "optimal prediction" in space, using observations taken at known nearby locations. I shall attempt to document how that word arose, and where else the same ideas might have appeared. Hence, I shall restrict attention to what is now known as simple kriging and ordinary kriging, and shall avoid the multitude of other "krigings" that have appeared since the mid 1960s.

Ord (1983), in an entry of the *Encyclopedia of Statistical Sciences*, describes kriging as "a method of interpolation for random spatial processes," and presents predictors that are linear in the observations, although he mentions nonlinear possibilities. According to Hemyari and Nofziger (1987), in an article for soil scientists, "Kriging is a form of weighted averaging in which the weights

[1]Manuscript received 26 July 1989; accepted 7 September 1989.

[2]Department of Statistics, Iowa State University, Ames, Iowa 50011.

are chosen such that the error associated with the [predictor] is less than for any other linear sum. . . . The weights depend upon the location of the points used in the [prediction] process and upon the [covariation] . . . reflected in the semivariogram."

Matheron (1963b) defines kriging as follows: "It consists [of predicting] the grade of a panel by computing the weighted average of available samples. . . . The suitable weights a_i . . . are determined by. . . . $\Sigma a_i = 1$ . . . [and the prediction] variance . . . should take the smallest possible value." Notice that the grade of the panel is obtained by averaging over a volume (or support) that is, in general, different from that of the samples.

Krige (1978, p. 24) has his own perception of the word, as "the name given . . . by Matheron to: the *multiple regression* procedure for arriving at the *best linear unbiased* [*predictor*] or best linear *weighted moving average* [predictor] of the ore grade of an ore block (of any size) by assigning an optimum set of weights to all the available and relevant data inside and outside the ore block" (emphases are Krige's). It is without doubt that Matheron (1963b) brought the word "kriging" into Anglo-Saxon mining terminology, but interestingly the original French term "krigeage," was coined by Pierre Carlier. (G. Matheron states this in a letter to me dated April 12, 1989.) Apparently, it was first used in the late 1950s at the French *Commissariat à l'Energie Atomique* (see the Preface to Matheron, 1962, written by F. Blondel).

Clearly, kriging was originally a linear predictor, and it is in this context that it will be used here. In more recent developments in geostatistics, methods of optimal nonlinear spatial prediction have become part of the "kriging family."

The organization of this article is as follows. The next section looks at various versions of kriging as assumptions about the mean function change. The original formulation of kriging, now known as ordinary kriging (e.g., Journel and Huijbregts, 1978, p. 563), is seen to consist of three key components, and the definitions quoted above should be assessed in this light. This also permits a guided review of the origins of kriging in mining, in meteorology, in statistics, and in other disciplines. Finally, a historical map of kriging (up to 1963) is given, which attempts to summarize and compare the contributions of various individuals.

BEST LINEAR UNBIASED PREDICTION (BLUP) AND KRIGING

Suppose that data $\mathbf{Z} \equiv (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}))'$ can be thought of as observations from a random (or stochastic) process

$$\{Z(\mathbf{s}): \mathbf{s} \in D\}; D \subset \mathbb{R}^d \tag{1}$$

at known locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$. Assume furthermore that

$$Z(\mathbf{s}) = \mu + \delta(\mathbf{s}); \mathbf{s} \in D \tag{2}$$

where $\delta(\cdot)$ is a zero-mean stochastic process with known covariance function,

$$C(\mathbf{s}, \mathbf{u}) \equiv \operatorname{cov}\big(Z(\mathbf{s}), Z(\mathbf{u})\big);\ \mathbf{s}, \mathbf{u} \in D$$

It is desired to predict (or krige) $Z(\mathbf{s}_0)$ at a known location $\mathbf{s}_0$, based on data $\mathbf{Z}$; notice that this is *not* a filtering or state-estimation problem where the goal is to predict $S(\mathbf{s}_0)$, a noiseless version of $Z(\mathbf{s}_0)$. Should some of the data be defined on a larger support (e.g., $\int_{B_i} Z(\mathbf{s})\, d\mathbf{s} / \int_{B_i} d\mathbf{s}$ instead of $Z(\mathbf{s}_i)$), the equations that follow require only slight modification, and the conclusions are unchanged. For simplicity, kriging will be presented in terms of spatial prediction from data all with point support.

The calculations that follow are necessarily brief. Details of the algebra can be found in Cressie (1991, Ch. 3); although it is not claimed that any of these calculations are new, I have not seen them brought together and interpreted in this way before.

Simple Kriging

Assume μ is known. The *simple* kriging predictor (e.g., Journel and Huijbregts, 1978, p. 561) can be obtained as the linear predictor,

$$\sum_{i=1}^{n} l_i Z(\mathbf{s}_i) + k$$

of $Z(\mathbf{s}_0)$ that minimizes the mean-squared prediction error:

$$E\left(Z(\mathbf{s}_0) - \sum_{i=1}^{n} l_i Z(\mathbf{s}_i) - k\right)^2 \tag{3}$$

It can be shown that minimizing Eq. (3) with respect to $\boldsymbol{l} \equiv (l_1, \ldots, l_n)'$ and k yields optimal values,

$$\boldsymbol{l}' = \mathbf{c}'C^{-1}; \qquad k = (1 - \mathbf{c}'C^{-1}\mathbf{1})\mu \tag{4}$$

where

$$\mathbf{c} \equiv \big(C(\mathbf{s}_0, \mathbf{s}_1), \ldots, C(\mathbf{s}_0, \mathbf{s}_n)\big)' \tag{5}$$

$$C \equiv \big(C(\mathbf{s}_i, \mathbf{s}_j)\big) \tag{6}$$

the $n \times n$ matrix whose $(i, j)^{\text{th}}$ element is $C(\mathbf{s}_i, \mathbf{s}_j)$, and $\mathbf{1}$ is an $n \times 1$ vector of 1s.

Thus the optimal predictor in the case where μ *is known* is,

$$Z^*(\mathbf{s}_0) \equiv \mathbf{c}'C^{-1}\mathbf{Z} + (1 - \mathbf{c}'C^{-1}\mathbf{1})\mu \tag{7}$$

the simple kriging predictor. The mean-squared prediction error is,

$$E\big(Z(\mathbf{s}_0) - Z^*(\mathbf{s}_0)\big)^2 = C(\mathbf{s}_0, \mathbf{s}_0) - \mathbf{c}'C^{-1}\mathbf{c} \tag{8}$$

The optimal predictor (7) was derived in a temporal setting by Wold (1938, p. 77), Kolmogorov (1941a), and Wiener (1949, Ch. II), and later by others in mining, meteorology, statistics, etc. (see below for more details).

Ordinary Kriging

Assume μ is unknown. Then Eq. (7) is no longer a predictor. There are at least two possible approaches to take.

First, one could restrict the class of linear predictors to the so-called homogeneous linear predictors

$$\sum_{i=1}^{n} \lambda_i Z(\mathbf{s}_i)$$

A further restriction of uniform unbiasedness yields the condition

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

Thus, one could look for the *best linear unbiased predictor (blup)*, obtained by minimizing

$$E\left(Z(\mathbf{s}_0) - \sum_{i=1}^{n} \lambda_i Z(\mathbf{s}_i)\right)^2 \tag{9}$$

over $\lambda_1, \ldots, \lambda_n$, subject to

$$\sum_{i=1}^{n} \lambda_i = 1 \tag{10}$$

By the method of Lagrange multipliers, the optimal values are

$$\boldsymbol{\lambda}' = \left(\mathbf{c} + (1 - \mathbf{c}'C^{-1}\mathbf{1})(\mathbf{1}'C^{-1}\mathbf{1})^{-1}\mathbf{1}\right)'C^{-1} \tag{11}$$

where $\mathbf{c}$ and C are given by (5) and (6). Thus, the optimal linear predictor of $Z(\mathbf{s}_0)$ in the case where μ *is unknown* is,

$$\hat{Z}(\mathbf{s}_0) = \mathbf{c}'C^{-1}\mathbf{Z} + (1 - \mathbf{c}'C^{-1}\mathbf{1})(\mathbf{1}'C^{-1}\mathbf{1})^{-1}(\mathbf{1}'C^{-1}\mathbf{Z}) \tag{12}$$

the ordinary kriging predictor. The mean-squared prediction error is,

$$E\left(Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0)\right)^2 = C(\mathbf{s}_0, \mathbf{s}_0) - \mathbf{c}'C^{-1}\mathbf{c} + (1 - \mathbf{c}'C^{-1}\mathbf{1})^2(1'C^{-1}\mathbf{1})^{-1} \tag{13}$$

A second approach, more *ad hoc*, is to make (7) a (linear) predictor by substituting in a (linear) estimator of μ. The obvious choice is the best linear unbiased estimator (blue),

$$\hat{\mu} = (\mathbf{1}'C^{-1}\mathbf{1})^{-1}\mathbf{1}'C^{-1}\mathbf{Z} \tag{14}$$

Upon substituting $\hat{\mu}$ for μ in (7), it is clear that the blup (12) is obtained (Goldberger, 1962). Thus, it appears that optimal linear estimation of mean parameters yields optimal linear prediction; these ideas are made more precise in Cressie (1991, Section 3.4e). And, with reference to inference about the unknown random variable $Z(\mathbf{s}_0)$, it should also be clear now why I have used the word "prediction" and avoided using the word "estimation."

For those who do not recognize Eqs. (12) and (13) as the ordinary kriging predictor and the kriging variance, respectively, it is straightforward to show that minimizing (3) subject to (10) equally yields,

$$\hat{Z}(\mathbf{s}_0) = \boldsymbol{\gamma}'\boldsymbol{\Gamma}^{-1}\mathbf{Z} + (1 - \boldsymbol{\gamma}'\boldsymbol{\Gamma}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{\Gamma}^{-1}\mathbf{1})^{-1}(\mathbf{1}'\boldsymbol{\Gamma}^{-1}\mathbf{Z}) \tag{15}$$

$$E\big(Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0)\big)^2 = \boldsymbol{\gamma}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma} - (1 - \boldsymbol{\gamma}'\boldsymbol{\Gamma}^{-1}\mathbf{1})^2(\mathbf{1}'\boldsymbol{\Gamma}^{-1}\mathbf{1})^{-1} \tag{16}$$

where $\boldsymbol{\gamma} \equiv (\gamma(\mathbf{s}_0, \mathbf{s}_1), \ldots, \gamma(\mathbf{s}_0, \mathbf{s}_n))'$, $\boldsymbol{\Gamma}$ is an $n \times n$ matrix whose $(i, j)^{\text{th}}$ element is $\gamma(\mathbf{s}_i, \mathbf{s}_j)$, and

$$2\gamma(\mathbf{s}_i, \mathbf{s}_j) \equiv C(\mathbf{s}_i, \mathbf{s}_i) + C(\mathbf{s}_j, \mathbf{s}_j) - 2C(\mathbf{s}_i, \mathbf{s}_j) \tag{17}$$

is called the *variogram* [$\gamma(\mathbf{s}_i, \mathbf{s}_j)$ is called the semivariogram]. In contrast, the simple kriging predictor (7) has no analogous expression in terms of just the variogram.

There are versions of the kriging equations that involve spatial prediction of an *average* (or regularization) of the process, $Z(B) = \int_B Z(\mathbf{s})\, d\mathbf{s}/|B|$, where $B \subset D$ and $|B| \equiv \int_B d\mathbf{s}$. Inference on the average value $Z(B)$ from the data $\{Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)\}$ has a change-of-support aspect to it, in that the predictand $Z(B)$ is defined on the support B and the data have point support.

In this case, the optimal predictor $\hat{Z}(B)$ and its mean-squared prediction error are given by Eqs. (12) and (13), respectively, with $\mathbf{c}$ replaced by $\mathbf{c}(B) \equiv (C(B, \mathbf{s}_1), \ldots, C(B, \mathbf{s}_n))'$, and $C(\mathbf{s}_0, \mathbf{s}_0)$ replaced by $C(B, B)$. Here $C(B, \mathbf{s}) \equiv \int_B C(\mathbf{u}, \mathbf{s})\, d\mathbf{u}/|B|$, and $C(B, B) \equiv \int_B \int_B C(\mathbf{u}, \mathbf{v})\, d\mathbf{u}\, d\mathbf{v}/|B|^2$. Similar expressions for these kriging equations in terms of the variogram can be found, for example, in Journel and Huijbregts (1978, p. 306).

The result given above, that ordinary kriging is just simple kriging using a blue, can be traced back to Goldberger (1962). However, Goldberger couched his results in a multivariate-analysis setting, and did not write about their potential for spatial prediction. To be more specific, kriging is concerned with prediction of one part of a stochastic process from observations on other parts. Thus, although Goldberger's treatment is in a sense more general (he considered linear prediction of the variable Y from other variables $Z_1, \ldots, Z_n$ in a general-linear-model setting), he needs to specify covariances not only between the data but also between the data and the predictand. In a spatial (and temporal) setting one needs only knowledge of the random process' covariance function or variogram to perform kriging via (12) or (15).

As a consequence of the development just given, I propose the following three key components of (ordinary) kriging:

Component 1: Use of covariances $\mathbf{c}$, C in (12) [or of variograms γ, Γ in Eq. (15)] rather than some other weights $\mathbf{w}$, W.

Component 2: Use of the blue $\hat{\mu}$ given by Eq. (14) in the predictor (7), rather than assuming μ known.

Component 3: Use of spatial locations $\{\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_n\}$ to define covariances (or semivariograms) of the random process Z, rather than using covariances from a general multivariate-analysis setting.

In other words, prediction based on (ordinary) kriging is equivalent to *spatial blup*. The predictor minimizes the mean-squared prediction error over all linear unbiased predictors, for a given covariance function $C(\cdot, \cdot)$.

SPATIAL BLUP (KRIGING) IN APPLICATIONS-ORIENTED AREAS

The need to obtain better predictions from observed data can be found in all scientific disciplines. Those that have embraced statistical notions of random variation are able to do this by exploiting the statistical dependence in the data and the random variable to be predicted. It is not surprising then that best linear unbiased prediction (blup) has often been developed independently in different subject-matter areas, to solve the particular problems at hand. This section discusses the independent development of blup in a spatial setting, in various disciplines. All components of the spatial blup were present in mining and meteorology in the early 1960s, but to my knowledge its development in other disciplines took longer.

Mining

Forty years ago, this discipline was using only the most rudimentary tools of statistics. The statistical component of the decision to establish a mine was typically based on biased sampling prompted by surface exposure of the ore deposit. This might result in a hundred or so assay values. Clearly, the sample mean of the assays multiplied by the estimated ore-body volume, is not a very good estimate of total recoverable ore. Nor does the sample standard deviation give a very good picture of local (i.e., small-scale spatial) variability throughout the ore body. And it is this local variability that makes or breaks a mining venture; concentrations of high-grade ore are easier and more profitable to mine, the ore body. And it is this local variability that makes or breaks a mining venture; concentrations of high-grade ore are easier and more profitable to mine, while regions of low-grade ore should be passed by.

In the 1940s, the gold fields of South Africa were using precisely the sample mean of nearby core-sample assays to estimate the average grade in a prospective mining block. Those estimates were then used to mine selectively. In the early 1950s, D. G. Krige made a great contribution to the subject by noting three fundamental flaws in this technique. First, the gold-assay data are log-normal, necessitating them to be transformed, and the block-grade estimate to be back-transformed. Second, the variability of block grade is considerably less than the variability of (averaged) core-sample grade (i.e., there is an effect due to the change of support). Third, the block grade and the core-sample grade are correlated, a fact that should be exploited for better prediction. Krige (1951) gives a nice discussion of how regression can improve prediction, in a mining context. At that time, mining engineers were not only failing to regress toward the mean, but by not recognizing the difference in variation between blocks and core samples their predictors actually progressed *away* from the mean in the wrong direction.

Although this contribution of Krige's, upon which he embellished in later articles in the 1950s is outstanding in its treatment of change of support, I shall discuss how spatial blup (kriging) did not actually appear until Matheron's contributions in the early 1960s. Indeed Krige said in his 1951 Masters Thesis from the University of the Witwatersrand, "Weighting [of samples around the entire periphery of the block] based on the so-called 'distance or area of influence of a sample' cannot be justified scientifically and will not cancel any bias due to the disproportionate location of sample sections round the periphery." A shorter version of this is found in Krige (1951, p. 125).

From the perspective of the 1980s, Krige was taking both sides of the same argument. By only considering samples on the periphery, he was saying that samples at a greater distance should be given *zero* weight. Yet of those samples he chose, he proposed that they should all be given equal weight, regardless of their relative locations to the block. In a letter to me dated April 12, 1989, D. G. Krige embellished on the statement of his Masters Thesis: "This practice [of weighting by distance or area of influence] was therefore equivalent in a sense to the now discredited polygonal weighting of data and it was this type of weighting I objected to. . . ." He also commented that the equal weights he was proposing would not be far from the optimal (kriging) weights, in the circumstances for which he was using them.

G. Matheron (1963b, p. 1265), in his development of spatial blup, says that in the presence of a large nugget effect his predictor is "not different from those [predictors] proposed formerly by D. G. Krige himself, in connection with the gold deposit of the Rand, in which the nugget effect is probably very strong." It appears that Matheron saw Krige's statement about equal weights and reasoned (wrongly) that this recommendation was a practical one to be used in the gold fields of South Africa; the importance of Krige's taking only the closest samples was missed by him. In fact, as will be seen below, *the nugget effect on the Rand is rather small, about 20%.*

By the 1960s, Krige saw the value of using data from not only the nearest neighbors, but also the second nearest neighbors, and third, etc. Krige (1962a, p. 362) actually contains the beginnings of blup, there in the context of an advancing mining face. An unweighted average of (log-transformed) samples of the nearest face and an unweighted average of (log-transformed) samples of the second-nearest face were used to predict the next block grade using multiple-regression techniques. Although the unweighted averaging of samples is not strictly optimal, another source of nonoptimality arises when Krige estimates the unknown means, not with generalized least-squares estimators (shown above to be optimum for squared error loss) but with ordinary least-squares estimators. *In the problems studied by Krige* these means were based on masses of data, making the practical difference between the two types of estimates negligible.

Transforming the data and back-transforming the predictor present special problems that will not be discussed in this article. Apart from these considerations, Krige's predictions of block ore grades were made with large amounts of data, strong positive correlation (0.8) between faces at small distances (several feet), and almost-zero correlation at 200 feet (see Krige, 1962b, p. 366). Thus, in the problems studied by Krige, use of nonoptimal weights in Eq. (12) is not likely to affect the results very much.

Faced with a large (in the early 1960s) computational burden of inverting $n \times n$ matrices (where n is the number of faces used in the prediction), and wondering how to incorporate not just nearby sample values but also recovery-grade values from nearby recently mined blocks, Krige and Ueckermann (1963, p. 442) proposed "an engineering approach to the principles of regression" by predicting using a two-dimensional weighted moving average of (log-transformed) values within some predetermined radius. The weights were chosen in an *ad hoc* way (Krige, 1966) and are certainly not optimal, even for the case in which means are assumed known. Krige and Ueckemann (1963) showed that *for the problem they were considering*, the predictions from the moving average predictor were almost identical to the predictor based on multivariate regression. However, computing prediction standard errors using Krige's "engineering approach" is not possible.

In the early 1960s, the important spatial component (Component 3) was still missing from Krige's work [i.e., the vector $\mathbf{c}$ in (12), or γ in (15) was obtained independently from the matrices C or Γ]. Matheron's fundamental contribution to this area was to add Component 3, yet still retain the change-of-support aspect introduced by Krige.

In a two-volume work of 504 pages (in French), Matheron (1962, 1963a) published his *Treatise of Applied Geostatistics*, making available to the profession a comprehensive theory to handle the spatial aspects of mining problems. In particular, Volume II (Matheron, 1963a) deals entirely with kriging, but there is also some discussion of it in Volume I (Matheron, 1962, Ch. VIII). A

very much abbreviated version (in English) appeared in 1963 (Matheron, 1963b), its goal being to call attention to the two-volume treatise.

The spatial blup (kriging) in mining truly belongs to Matheron (1962, Ch. VIII), who derives the equations for the optimal coefficients [see Eq. (12) or (15)], and applies them to a particular stochastic model he calls the "scheme of de Wijs." Kriging was not without some opposition from geologists and mining engineers, however, most notably E. H. T. Whitten in the U.S. (who preferred polynomial interpolation); see Whitten (1966) and a rejoinder by Matheron (1967).

Meteorology

While Matheron was developing a theory of spatial blup in France, the meteorologist L. S. Gandin in the Soviet Union was doing remarkably similar work. His book of 238 pages (Gandin, 1963), translated into English in 1965, is notable for its depth of treatment of spatial prediction and design, for its clarity of exposition, and for its attractive blend of both theory and applications. There the variogram is called a *homogeneous structure function* (Ch. 2, Sec. 1), simple kriging is called *optimum interpolation* (Ch. 3, Sec. 2), ordinary kriging is called *optimum interpolation with normalization of weighting factors* (Ch. 3, Sec. 5), and simple cokriging is called *optimum matching of fields* (Ch. 5, Sec. 2).

Meteorologists are not only interested in interpolation; indexes for the circulation of the atmosphere, total currents, and snow-cover depth are all *average* characteristics of meteorological fields. Gandin (1963, Ch. 7, Sec. 3) gives the version of Eq. (12) for optimally predicting $Z(B) = \int_B Z(\mathbf{s})\, d\mathbf{s}/|B|$, and hence also addresses the change-of-support aspect of spatial blup.

Statistics

Yaglom (1962, Part 2) develops in detail the theory of optimal linear extrapolation and filtering of time-stationary random functions with known mean, expanding on the earlier work of Wold (1938), Kolmogorov (1941a), and Wiener (1949). However, as the following discussion shows, blup does not necessarily need a temporal (or spatial) setting.

Suppose the data $\mathbf{Z} \equiv (Z_1, \ldots, Z_n)'$ can be written as

$$\mathbf{Z} = \mu\mathbf{1} + \boldsymbol{\delta} \tag{18}$$

where μ is an unknown scalar, $\mathbf{1}$ is an $n \times 1$ vector of 1s, and $\boldsymbol{\delta}$ is a zero-mean error vector with $\text{var}(\boldsymbol{\delta}) = \Sigma_{ZZ}$, an $n \times n$ positive-definite matrix. Furthermore, suppose that the random variable Y can be written as

$$Y = \mu + \omega \tag{19}$$

where μ is the same unknown scalar as in Eq. (18), and ω is a zero mean error such that

$$\text{var}\left((\boldsymbol{\delta}'\omega)'\right) = \begin{bmatrix} \Sigma_{ZZ} & \boldsymbol{\sigma}_{ZY} \\ \boldsymbol{\sigma}_{ZY}' & \sigma_Y^2 \end{bmatrix} \tag{20}$$

where $\sigma_Y^2 > 0$.

The problem is to find an optimal predictor of Y (i.e., a predictor with minimum mean-squared prediction error) that is linear in the data (i.e., $\hat{Y} = \boldsymbol{\lambda}'\mathbf{Z}$) and is unbiased [i.e., $E(\hat{Y}) = E(Y) = \mu$, for all $\mu \in (-\infty, \infty)$]. A special case of a result due to Goldberger (1962), yields

$$\hat{Y} = \boldsymbol{\lambda}'\mathbf{Z} \tag{21}$$

where $\boldsymbol{\lambda}'$ is given by Eq. (11) with $\boldsymbol{\sigma}_{ZY}$ replacing $\mathbf{c}$, and Σ_{ZZ} replacing C. The same replacement in Eq. (13), along with σ_Y^2 replacing $C(\mathbf{s}_0, \mathbf{s}_0)$, yields the minimum mean-squared prediction error.

Goldberger's model is more general in that the constant mean effects in Eqs. (18) and (19) are generalized to $X\boldsymbol{\beta}$ and $\mathbf{X}_0\boldsymbol{\beta}$, respectively, expressing mean effects that are linear in the p parameters $\boldsymbol{\beta} \equiv (\beta_1, \ldots, \beta_p)'$. Put into a spatial setting, his results would give the universal kriging predictor (e.g., Journel and Huijbregts, 1978, p. 318).

Notice that, to compute the optimal $\boldsymbol{\lambda}$ in Eq. (21), knowledge of both $\boldsymbol{\sigma}_{ZY}$ and Σ_{ZZ} is needed. In a spatial setting these would be immediate from the covariance function $C(\mathbf{s}, \mathbf{u})$. So, although Eq. (21) is a blup, it is not a spatial blup.

Whittle (1963) also assumes a general linear model, but in a temporal setting; thus he obtains the (universal) kriging predictor of $Y \equiv Z_{n+L}$ from the time series data $\{Z_1, \ldots, Z_n\}$. Interestingly, his criterion is different from the usual one. In the context of the simpler model (18), (19), he looks for a predictor $\hat{Y} = \boldsymbol{\lambda}'\mathbf{Z}$ that minimizes

$$\max_{\mu} E(Y - \boldsymbol{\lambda}'\mathbf{Z})^2 \tag{22}$$

It is a consequence of the mathematics that

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

and hence the optimal predictor according to the minimax criterion (22) is also the blup (21).

Other Disciplines

Perhaps one might have expected the spatial blup to have occurred in *forestry*, given the highly original and important publication of Matern (1960). In

144 pages, B. Matern developed the framework for modeling spatial processes (whose index varies continuously over a spatial domain), and applied these ideas to spatial sampling. However, throughout his book he always uses

$$\overline{Z} \equiv \sum_{i=1}^{n} Z(\mathbf{s}_i)/n$$

as his predictor. Although $\overline{Z}$ is linear and unbiased, it is not optimal unless the spatial correlation is white noise.

Within *physics*, the theory of turbulence attracted contributions from A. N. Kolmogorov in the 1940s; in some of that work (e.g., Kolmogorov, 1941b) he assumed the existence of a variogram (in four dimensions, space and time) to characterize the local structure of turbulence. [Stochastic processes possessing a variogram were later called locally homogeneous by Yaglom (1957, p. 284), where the variogram was called a structure function.] However, when Kolmogorov wrote about optimum prediction, he considered an equally spaced time series, and assumed the mean was known (Kolmogorov, 1941a). His interest was more in the limits of mean-squared prediction errors for extrapolation and interpolation, and he references Wold's (1938, p. 77) solution to the optimal prediction of time series with known mean (simple kriging).

During World War II, N. Wiener developed similar (simple) kriging equations to Kolmogorov, also in a temporal setting. The purpose was to predict enemy aircraft movements from known radar measurements, and hence the data were $\{Z(t): t \leq t^*\}$, rather than $\{Z_1, \ldots, Z_n\}$. The theory was later published in Wiener (1949). Thompson (1956) has an extension of Wiener's approach, to a spatial setting. None of their results adapt well to the sparse, irregular data configurations often found in mining and geology applications.

A similar development to that described for the statistics discipline occurred in *plant and animal breeding*. Here, a multivariate-analysis approach was taken to predict the genotypic value of a plant or an animal from measured characteristics (e.g., for wheat-selection, these might be yield of grain, baking quality of flour, etc.). Using the sample mean to estimate the unknown means, Fairfield Smith (1936), in plant breeding, considered best linear prediction based on the covariance structure of the mean-corrected data and the cross-covariance between the mean-corrected data and the predictand. Apart from the mean corrections, the predictor was essentially simple kriging given by Eq. (7). In animal breeding, similar results were derived by Hazel (1943). Henderson (1963) extended these results by considering efficient mean corrections that led to the blup (21) and its multiparameter version.

Geodesists, interested in drawing spatial maps, have also modeled their measurements as coming from a random process. This began with Gauss (1809), who assumed independent errors and developed the method of least squares to deal with them (actually, the method of least squares appeared earlier, in Legendre, 1805, but without a probabilistic justification). Dependent-error models

came much later; Moritz (1963), assuming known mean and stationary covariances in Eqs. (5) and (6), developed the simple kriging equations (7).

CONCLUSIONS

It is clear that many researchers in many disciplines have contributed to the problem of optimal spatial prediction. The previous section proposes that both G. Matheron and L. S. Gandin were the first to publish a definitive development of spatial blup (kriging). D. G. Krige's contributions in mining engineering were considerable but he did not discover kriging, illustrating once again Stigler's Law of Eponymy (Stigler, 1980). The other researchers mentioned above all came close in their own way, but did not show each of the three key components. To summarize the important contributions, Table 1 shows a historical map of kriging, giving an index for each of the components defined earlier.

An index of "1" for Component 1 means that all possible covariances were used to define weights ("1/2" means only some covariances were used; "0" means some other weights were used). An index of "1" for Component 2 means that the weighted least-squares estimator of μ was used ("1/2" means some other estimator was used; "0" means μ was assumed known). An index of "1" for Component 3 means that prediction was performed in a spatial setting ("1/2" means a temporal setting was used; "0" means that neither a spatial nor a temporal setting was considered).

I have received suggestions from colleagues to add further embellishments to the discussion above; these include kriging's practical implementation (from

Table 1. Historical Map of Researchers' Contributions to Spatial Blup (Kriging)[a]

Researcher	1938	1941–1949	1951	1956	1962	1963
Wold	(1, 0, 1/2)					
Kolmogorov		(1, 0, 1/2)				
Wiener		(1, 0, 1/2)				
Krige			(1/2, 1/2, 0)		(1, 1/2, 0)	
Thompson				(1, 0, 1)		
Goldberger					(1, 1, 0)	
Matheron					(1, 1, 1)	
Gandin						(1, 1, 1)
Whittle						(1, 1, 1/2)
Moritz						(1, 0, 1)
Henderson						(1, 1, 0)

[a]Entries (x_1, x_2, x_3) denote indices on the three key components (see text); briefly, $x_1 = 1$ corresponds to optimal weights (known mean μ), $x_2 = 1$ corresponds to optimal estimation of μ, and $x_3 = 1$ corresponds to a spatial setting.

variogram estimation to the use of moving neighborhoods), its connection with splines, its interpretation as a projection, and its change-of-support properties. Change-of-support here refers to predicting parts of the process whose support is different from that of the data. This was Krige's (1951) original goal, namely to convince mining engineers to use multiple regression to obtain more accurate and precise predictors of mining blocks from assay values. Wold, Kolmogorov, Wiener, and Whittle did not consider this problem, although both Matheron and Gandin did. In writing this article, I have considered the choice of optimal weights in Eq. (12) or (15) as defining kriging; however, its role in dealing with change-of-support is seen as more important by G. Matheron (in a letter to me dated April 12, 1989).

Why is the term "kriging" used to describe spatial optimal linear prediction? Because Matheron (1963b) chose to honor D. G. Krige's contributions to related areas of mining. The third main section of this article gives the details, and shows that G. Matheron and L. S. Gandin independently developed ordinary kriging, as we know it today.

ACKNOWLEDGMENT

I have benefitted considerably from the comments of numerous people, to whom I sent a first draft of this paper. I shall not name the many who generously gave me their reactions, but simply thank them collectively. Any mistakes and omissions then remain truly my own. This research was partially supported by the NSF under grant DMS-8703083.

REFERENCES

Cressie, N., 1991, Statistics for spatial data: Wiley, New York (forthcoming).

Fairfield Smith, H., 1936, A discriminant function for plant selection: Annals of Eugenics, v. 7, p. 240–250.

Gandin, L. S., 1963, Objective analysis of meteorological fields: Gidrometeorologicheskoe Izdatel'stvo (GIMIZ), Leningrad (translated by Israel Program for Scientific Translations, Jerusalem, 1965, 238 pp.).

Gauss, C. F., 1809, Theoria motus corporum celestium: Perthes et Besser, Hamburg (translated as Theory of motion of the heavenly bodies moving about the sun in conic sections, trans. C. H. Davis: Little, Brown, Boston, 1857).

Goldberg, A. S., 1962, Best linear unbiased prediction in the generalized linear regression model: J. Am. Stat. Assoc., v. 57, p. 369–375.

Hazel, L. N., 1943. The genetic basis for constructing selection indexes: Genetics, v. 28, p. 476–490.

Hemyari, P., and Nofziger, D. L., 1987, Analytical solution for punctual kriging in one dimension: Soil Sci. Soc. Am. J., 51, p. 268–269.

Henderson, C. R., 1963, Selection index and expected genetic advance, *in* Statistical genetics and plant breeding, W. D. Hanson and H. F. Robinson (Eds.): National Research Council Publication 982, National Academy of Sciences, Washington, D.C., p. 141–163.

Journel, A. G., and Huijbregts, C. J., 1978, Mining geostatics: Academic Press, London, 600 pp.

Kolmogorov, A. N., 1941a, Interpolation and extrapolation of stationary random sequences: Izvestiia Akademii Nauk SSR, Seriia Matematicheskaia, v. 5, p. 3–14 (translation, Memo RM-3090-PR, Rand Corporation, Santa Monica, California, 1962, 14 pp.).

Kolmogorov, A. N., 1941b, The local structure of turbulence in an incompressible fluid at very large Reynolds numbers: Doklady Akademii Nauk SSR, v. 30, p. 301–305 (reprinted in Turbulence: classic papers of statistical theory, S. K. Friedlander and L. Topping (Ed.): Interscience Publishers, New York, 1961, p. 151–155).

Krige, D. G., 1951, A statistical approach to some basic mine valuation problems on the Witwatersrand: J. Chem. Metal. Min. Soc. South Africa, v. 52, p. 119–139.

Krige, D. G., 1962a, Effective pay limits for selective mining: J. South Africa Inst. Min. Metal., v. 62, p. 345–363.

Krige, D. G., 1962b, Economic aspects of stoping through unpayable ore: J. South African Inst. Min. Metal., v. 62, p. 364–374.

Krige, D. G., 1966, Two-dimensional weighted moving average trend surfaces for ore valuation, *in* Symposium on mathematical statistics and computer applications in ore valuation: Johannesberg, South Africa, 1966. Special issue of J. South African Inst. Min. Metal., p. 13–79.

Krige, D. G., 1978, Lognormal-de Wijsian geostatistics for ore evaluation: South African Institute of Mining and Metallurgy Monograph Series, Johannesburg, 50 pp.

Krige, D. G., and Ueckermann, H. J., 1963, Value contours and improved regression techniques for ore reserve valuations: J. South African Inst. Min. Metal., v. 63, p. 429–452.

Legendre, A. M., 1805, Nouvelles méthodes pour la détermination des orbites des cométes: Courcier, Paris, 80 pp.

Matern, B., 1960, Spatial variation: Meddelanden fran Statens Skogsforskningsinstitute, v. 49, 144 pp.

Matheron, G., 1962, Traité de geostatisque appliquée, vol. I: Memoires du Bureau de Recherches Géologiques et Miniéres, no. 14, Editions Technip, Paris, 333 pp.

Matheron, G., 1963a, Traité de geostatistique appliquée, vol. II, Le krigeage: Memoires du Bureau de Recherches Géologiques et Miniéres, no. 24, Editions Bureau de Recherche Géologiques et Miniéres, Paris, 171, pp.

Matheron, G., 1963b, Principles of geostatistics: Economic Geol., v. 58, p. 1246–1266.

Matheron, G., 1967, Kriging or polynomial interpolation procedures?: Trans. Canad. Inst. Min. Metal., v. 70, p. 240–244.

Moritz, H., 1963, Statistische methoden in der gravimetrischen geodasie: Zeitschrift fur Vermessungswesen, v. 88, p. 409–416.

Ord, J. K., 1983, Kriging, entry, *in* Encyclopedia of statistical sciences, vol. 4, S. Kotz and N. Johnson (Eds.): Wiley, New York, p. 411–413.

Stigler, S. M., 1980, Stigler's law of eponymy: Trans. New York Acad. Sci. Series 2, v. 39 (Merton Festschrift Vol., ed. by T. Gieryn), p. 147–158.

Thompson, P. D., 1956, Optimum smoothing of two-dimensional fields: Tellus, v. 8, p. 384–393.

Whitten, E. H. T., 1966, The general linear equation in prediction of gold content in Witwatersrand rocks, South Africa, *in* Symposium on mathematical statistics and computer applications in ore valuation: Johannesberg, South Africa, 1966; special issue of J. South African Inst. Min. Metal., p. 124–147.

Whittle, P., 1963, Prediction and regulation by linear least-square methods: van Nostrand, Princeton, 141 pp.

Wiener, N., 1949, Extrapolation, interpolation and smoothing of stationary time series: MIT Press, Cambridge, Massachusetts, 158 pp.

Wold, H., 1938, A. study in the analysis of stationary time series: Almqvist and Wiksells, Uppsala, 211 pp.

Yaglom, A. M., 1957, Some classes of random fields in n-dimensional space, related to stationary random processes: Theory Prob. Appl., v. 2, p. 273–320.

Yaglom, A. M., 1962, An introduction to the theory of stationary random functions: Prentice-Hall (as of 1973, published by Dover, New York, 235 pp.).