Kriging

Introduction

Kriging, at its most fundamental level, is an interpolation method used to convert partial observations of a spatial field to predictions of that field at unobserved locations. In geostatistics, the field may represent permeability of soil and the observations from physical measurements in a well-bore. In another example, of particular importance for environmental problems, the field may represent the concentration of pollutants in a contaminated site and the observations from remote sensing equipment. In this article we give a short account of kriging under some basic modeling assumptions and observation scenarios. We give a number of equivalent derivations of the kriging prediction, each shedding light on a particular aspect of kriging. Taken as a whole, these derivations provide a broader perspective than any single derivation of kriging can provide.

Kriging is named after the South African mining engineer, D. G. Krige, who during the 1950s developed statistical techniques for predicting ore-grade distributions from ground samples [1]. The term *kriging* was coined by Matheron [2] who is one of the most influential contributors to the subject. By now, kriging has become a predominant prediction tool in spatial statistics (see Ref. 3 for a historical account) and can be shown to be related to spline interpolation [4], generalized least squares (GLS) regression, Wiener filtering, and objective analysis in meteorology [5]. Book length accounts of kriging can be found in Refs 2, 5–13.

One of the fundamental assumptions used in kriging is that the spatial field of interest is a realization of a **random field** or a **stochastic process**. The randomness can be a consequence of some physical mechanism generating the spatial field (e.g., diffusion process driven by random forcing) or may simply be used to account for uncertainty. Either way, the random field model is used to define the kriging predictor as the optimal linear unbiased predictor, where optimality and unbiasedness are defined through the randomness inherent in the spatial field model. In this article, we explore two alternatives for characterizing the kriging prediction: using GLS regression and splines. Using GLS regression is arguably the cleanest and easiest to interpret formula for kriging. In contrast, the spline characterization is useful for relating all three characterizations. The spline characterization also allows one to easily generalize from the point-wise predictions discussed in this article to the prediction of other linear functionals (integrals or derivatives for example).

Based in part on the article "Kriging" by Paul Switzer, which appeared in the *Encyclopedia of Environmetrics*.

The Basic Model

In this section, we present the basic modeling assumptions which provide the main ingredients for kriging and serve as the basis for the generalizations discussed in section titled Extensions to the Basic Model. We start by letting $Y(x)$ denote the spatial field of interest, where x ranges in $\mathbb{R}^d$. The basic kriging model stipulates the following form for Y

$$Y(x) = \sum_{p=1}^{N} \beta_p f_p(x) + Z(x) \qquad (1)$$

where Z is a centered (i.e., mean zero) correlated random field; the functions f_p are known basis functions; and β_p are unknown coefficients. We assume the field Y is observed at spatial locations $x_1, \ldots, x_n$ with some additive noise corrupting the observations. In particular, our observations constitute n measurements $y_1, \ldots, y_n$ where each y_k has the form

$$y_k = Y(x_k) + \sigma\epsilon_k$$

where $\epsilon_1, \ldots, \epsilon_n$ are independent mean zero random variables (also independent of Y) with standard deviation 1. The goal is to then predict $Y(x_0)$ at some spatial location $x_0 \in \mathbb{R}^d$. Kriging produces such a prediction, denoted $\widehat{Y}(x_0)$, which is a linear function of the data $y_1, \ldots, y_n$.

The additive errors $\sigma\epsilon_k$ constitute what is called a *nugget effect*. The nugget effect can model instrumental noise or the presence of a microscale process, possibly discontinuous, which has correlation length scales much smaller than Z. The nomenclature comes from the geostatistics literature (see Ref. 8) which uses a nugget effect to model the presence of small nuggets which contribute a spatially uncorrelated (or

Encyclopedia of Environmetrics, Online © 2006 John Wiley & Sons, Ltd.

This article was published in *Encyclopedia of Environmetrics Second Edition* in 2012 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470057339.vak003.pub2

nearly so) discontinuous field added to the larger scale concentration variations modeled with Y. When there is no nugget effect (i.e., $\sigma = 0$) kriging becomes an exact interpolator.

To fix notation for the remainder of the article we let $K(x, y) \equiv \mathrm{cov}(Z(x), Z(y))$ denote the covariance function of Z. Let $y \equiv (y_1, \ldots, y_n)^t$ denote the vector of observations. Let $\beta \equiv (\beta_1, \ldots, \beta_n)^t$ denote the vector of unknown coefficients and $X = \Big(f_p(x_k)\Big)_{k,p}$ denote the matrix of covariates at the observations, where the rows index the observations and the columns index the covariates. Also let $\mathrm{x} \equiv (f_1(x_0), \ldots, f_N(x_0))$ be the row vector of the covariates at the prediction location x_0. Finally, the matrix $\Sigma = \Big(K(x_i, x_j)\Big)_{i,j} + \sigma^2 I$ denotes the covariance matrix for the observation vector y where I is the identity matrix and $\Sigma_0 = \big(K(x_0, x_1), \ldots, K(x_0, x_n)\big)$ denotes the row vector of the covariance of $Y(x_0)$ with the observations in y.

Kriging with Generalized Least Squares Regression

The most intuitive definition of kriging is easiest to derive when additionally assuming both Z and the observations errors $\epsilon_1, \ldots, \epsilon_n$ are Gaussian. In actuality, Gaussianity is unnecessary but it makes the exposition somewhat cleaner. Under the Gaussian assumption one has the familiar regression characterization of the observations: $y \sim \mathcal{N}[X\beta, \Sigma]$ (i.e., y is a Gaussian vector with mean $X\beta$ and covariance matrix Σ). Even more is true, by Gaussianity of Y, the joint distribution of y and $Y(x_0)$ is given by

$$\begin{pmatrix} Y(x_0) \\ y \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mathrm{x}\beta \\ X\beta \end{pmatrix}, \begin{pmatrix} K(x_0, x_0) & \Sigma_0 \\ \Sigma_0^t & \Sigma \end{pmatrix}\right] \quad (2)$$

If β were known, one would simply use the conditional expectation $E(Y(x_0)|y)$ to predict $Y(x_0)$, which can be computed as follows

$$E(Y(x_0)|y) = \Sigma_0 \Sigma^{-1}(y - X\beta) + \mathrm{x}\beta \quad (3)$$

However, the kriging model (1) assumes β is unknown. To account for this uncertainty it is natural to use the regression format of the observations $y \sim \mathcal{N}[X\beta, \Sigma]$ to estimate β. Indeed the GLS estimate of β is given by

$$\hat{\beta} = (X^t \Sigma^{-1} X)^{-1} X^t \Sigma^{-1} y$$

Now, by replacing β in (3) with the GLS estimate $\hat{\beta}$, one obtains the following kriging prediction $\widehat{Y}(x_0)$ (even if Gaussianity failed to hold):

$$\widehat{Y}(x_0) = \Sigma_0 \Sigma^{-1}(y - X\hat{\beta}) + \mathrm{x}\hat{\beta} \quad (4)$$

This is arguably the easiest way to construct the kriging prediction and it makes clear how one can extend the basic kriging model to more complex observation scenarios and predictions. For example, if one observed linear functionals of Y, then the basic structure of (2) does not change. The only difference is a change in the design matrices X and the covariance matrices Σ and Σ_0. Once these changes are made, (4) gives the appropriate kriging estimate for these generalizations.

In the simplified case when β is known, the kriging estimate $\widehat{Y}(x_0)$ generated from (3) is often referred to as *simple kriging*. When $N = 1$ and the covariate $f_1(x)$ is a constant, $\widehat{Y}(x_0)$ from (4) is referred to as *ordinary kriging*. Finally, the prediction $\widehat{Y}(x_0)$ with no simplified assumptions on the size of N or the covariates f_p is generally referred to as *universal kriging*.

Kriging as a Best Linear Unbiased Predictor

Our second characterization of kriging is arguably the most popular. It defines kriging as optimal linear unbiased prediction. One clear advantage of this viewpoint is that it makes it clear why one should use kriging: it is unbiased and has the smallest expected (squared) prediction error among all linear unbiased estimates of $Y(x_0)$.

To derive this characterization first consider the set of linear combinations of the data $\lambda^t y = \sum_{k=1}^n \lambda_k y_k$, where $\lambda \equiv (\lambda_1, \ldots, \lambda_n)^t$. One of these linear combinations is the kriging prediction $\widehat{Y}(x_0)$. To isolate which one, first restrict λ by requiring $\lambda^t y$ and $Y(x_0)$ have the same expected value, irrespective of the unknown coefficient vector β. We call this condition unbiasedness. To enforce this condition notice that $E\lambda^t y = \lambda^t X\beta$ and $EY(x_0) = \mathrm{x}\beta$. Therefore it is clear that $\lambda^t X = \mathrm{x}$ (or equivalently $X^t\lambda = \mathrm{x}^t$) is a sufficient condition for unbiasedness. Of course, the vector x^t must be a member of the span of the columns of X^t for such a vector λ to exist. Assuming there exists at least one such λ, we can isolate the optimal unbiased linear predictor—the kriging predictor—by minimizing the *mean square prediction*

Encyclopedia of Environmetrics, Online © 2006 John Wiley & Sons, Ltd.

This article was published in *Encyclopedia of Environmetrics Second Edition* in 2012 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470057339.vak003.pub2

error $E\left[Y(x_0) - \lambda^t y\right]^2$ over the set of all λ which satisfy $X^t\lambda = \mathrm{x}^t$ (the unbiasedness constraint). Using the joint distribution (2) one gets

$$E\left[Y(x_0) - \lambda^t y\right]^2 = \sigma^2 \sum_{k=1}^{n} \lambda_k^2 + \sum_{k,j=0}^{n} \lambda_k \lambda_j K(x_k, x_j), \quad \text{where } \lambda_0 = -1 \tag{5}$$

Then using Lagrange multipliers, the minimizer of the above quantity, over the set of λ which satisfies $X^t\lambda = \mathrm{x}^t$, is characterized as follows:

$$\widehat{Y}(x_0) = \lambda^t y, \quad \text{where } \begin{pmatrix} \Sigma & X \\ X^t & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ \gamma \end{pmatrix} = \begin{pmatrix} \Sigma_0^t \\ \mathrm{x}^t \end{pmatrix} \tag{6}$$

Note that (5) also allows one to quantify *prediction error* of the kriging prediction $\widehat{Y}(x_0)$. At this point it is not entirely clear how this estimate relates to the one given in (4). It can be seen in the next section that the easiest way to relate (6) and (4) is through the third characterization: spline smoothers.

Remark A sufficient condition for a minimizer of (5) to exist over the set of λ which satisfy $X^t\lambda = \mathrm{x}^t$ is that X^tX be nonsingular and $\sum_{k,j=0}^{n} \lambda_k \lambda_j K(x_k, x_j) \geq 0$ for any λ. If not, and there exists a λ such that $\sum_{k,j=0}^{n} \lambda_k \lambda_j K(x_k, x_j) < 0$ for example, then the right hand side of (5) may not have a finite minimizer (the right hand side may be unbounded from below). Indeed, this exposes the need for the covariance function $K(x, y)$ to be *positive definite* when constructing the kriging prediction.

Kriging as a Spline Interpolator

There are a number of advantages to a spline characterization of the kriging prediction, one of which is that the estimate is written as a function of x_0. This allows easy computation of $\widehat{Y}(x_0)$ at a large number of points x_0, simultaneously. The spline characterization of $\widehat{Y}(x_0)$ is given by

$$\widehat{Y}(x_0) = \sum_{k=1}^{n} c_k K(x_0, x_k) + \sum_{p=1}^{N} b_p f_p(x_0), \quad \text{where } \begin{pmatrix} \Sigma & X \\ X^t & 0 \end{pmatrix} \begin{pmatrix} c \\ b \end{pmatrix} = \begin{pmatrix} y \\ 0 \end{pmatrix} \tag{7}$$

This is the typical form of a spline estimate where, in the spline literature, K is referred to as a *reproducing kernel* and the locations x_k are called the *spline knots* (see Ref. 4, for example).

At this point it is easy to establish the equivalence of all three characterizations: (4), (6), and (7). We start by showing the equivalence of (6) and (7)

$$\begin{aligned} \lambda^t y &= (\lambda^t, \gamma^t) \begin{pmatrix} y \\ 0 \end{pmatrix} \\ &\overset{\text{by (6)}}{=} (\Sigma_0, \mathrm{x}) \begin{pmatrix} \Sigma & X \\ X^t & 0 \end{pmatrix}^{-1} \begin{pmatrix} y \\ 0 \end{pmatrix} \\ &\overset{\text{by (7)}}{=} (\Sigma_0, \mathrm{x}) \begin{pmatrix} c \\ b \end{pmatrix} \\ &= \sum_{k=1}^{n} c_k K(x_0, x_k) + \sum_{p=1}^{N} b_p f_p(x_0) \end{aligned}$$

Note that we are tacitly assuming the above matrix inverse exists. A sufficient condition for this existence is that Σ be strictly positive definite and X^tX be nonsingular (see Section 3.4.1 of Ref. 7 for details). To see why the spline interpolator is equivalent to the GLS characterization (4) simply notice that $b = \hat{\beta}$ and $c = \Sigma^{-1}(y - X\hat{\beta})$ satisfy (7). This establishes the equivalence of all three characterizations.

The Variogram and Generalized Covariance Functions

In this section, we investigate an important fact about kriging: often one does not need to specify the full covariance structure of the process Z to compute the kriging prediction and the mean square prediction error. In particular, depending on what covariates are present in the kriging model (1), there are different covariance functions K which lead to the same prediction and mean square prediction error. This leads to simplified modeling assumptions, the theory of intrinsic random functions and generalized covariance functions developed by Matheron [14].

The classic example of a modeling simplification occurs when one of the covariates $f_p(x)$ is a constant in the kriging model (1). In this case, it can be shown that the kriging prediction and mean square prediction error only depend on what is called the **variogram**, defined as $\mathrm{var}(Z(x) - Z(y)) = K(x, x) + K(y, y) - 2K(x, y)$, in contrast to the full covariance

Encyclopedia of Environmetrics, Online © 2006 John Wiley & Sons, Ltd.

This article was published in *Encyclopedia of Environmetrics Second Edition* in 2012 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470057339.vak003.pub2

function $K(x, y)$ (see Refs 7 and 8). To be explicit, when computing the kriging prediction in (4), (6), or (7) and the mean square prediction error (5) one can replace all occurrences of $K(x, y)$ with the function $-\frac{1}{2}\mathrm{var}(Z(x) - Z(y))$ and obtain the same quantity. This is important since multiple covariance functions may have the same variogram. For example, the two covariance functions $(1 - |x - y|)^+$ and $|x| + |y| - |x - y|$ both have the same variogram when $x, y \in (0, 1)$. The advantage, then, is that one does not need to spend time deciding between two competing covariance models if they have the same variogram, for they yield the same predictions and mean square prediction errors. Another advantage is that variograms must satisfy a somewhat less stringent requirement (they are conditionally negative definite) than the positive definite condition required for covariance functions (see Section 2.3.6 in Ref. 8). The ubiquity of the variogram in the kriging literature can therefore be partially understood through the simple fact that it is common for kriging models to have at least one of the covariates be a constant function.

The simplification provided by the variogram can be generalized further. In particular, suppose the covariates $\{f_p(x)\colon p = 1, \ldots, N\}$ contain all monomials with degree less than or equal to some nonnegative integer k_0. In addition, suppose there exists a function $G(x, y)$ which is related to $K(x, y)$ as follows

$$K(x, y) = G(x, y) + \sum_{0 \le |\boldsymbol{i}| \le \boldsymbol{k_0}} a_{\boldsymbol{i}}(y)x^{\boldsymbol{i}} + \sum_{0 \le |\boldsymbol{i}| \le k_0} b_{\boldsymbol{i}}(x)y^{\boldsymbol{i}} + \sum_{0 \le |\boldsymbol{i}| + |\boldsymbol{j}| \le \boldsymbol{2k_0}} c_{\boldsymbol{i},\boldsymbol{j}} x^{\boldsymbol{i}} y^{\boldsymbol{j}} \quad (8)$$

where $\boldsymbol{i}, \boldsymbol{j}$ are d-dimensional multiindices; $a_{\boldsymbol{i}}(y)$ and $b_{\boldsymbol{i}}(x)$ are arbitrary functions; and $c_{\boldsymbol{i},\boldsymbol{j}}$ is an arbitrary real number for each $\boldsymbol{i}, \boldsymbol{j}$. Then to compute (4), (5), (6), or (7) one can simply replace all occurrences of $K(x, y)$ with $G(x, y)$ to obtain the same quantity. A derivation is beyond the scope of this article, however, we remark that the natural way to show this fact is through the kriging characterization given by (6) along with the unbiasedness condition $X^t\lambda = \mathrm{x}^t$. Any function $G(x, y)$ which satisfies (8) for some positive definite function $K(x, y)$ is said to be a *generalized covariance function of order* k_0 [7, 12, 14]. A detailed study of the theory of generalized covariance functions can be found in Refs [7 and 14].

Although the form of (8) seems complicated, there are rich classes of functions $G(x, y)$ which satisfy (8). For example, any covariance function $K(x, y)$ can be written as in (8) when $k_0 = 0$ and $G(x, y)$ is set to $-\frac{1}{2}\mathrm{var}(Z(x) - Z(y))$. This recovers the aforementioned fact that if one of the covariates $f_p(x)$ is constant then the kriging prediction only depends on the variogram. Another important class of generalized covariance functions are indexed by a single parameter $\alpha > 0$ and given by

$$G_\alpha(x, y) = \begin{cases} (-1)^{1+\lfloor \alpha/2 \rfloor}|x - y|^\alpha, & \text{when } \alpha/2 \notin \mathbb{Z} \\ (-1)^{1+\alpha/2}|x - y|^\alpha \log|x - y|, & \text{when } \alpha/2 \in \mathbb{Z} \end{cases}$$

where $\mathbb{Z}$ denotes the set of integers. It can be shown that $G_\alpha(x, y)$ is a generalized covariance function of order $k_0 = \lfloor \alpha/2 \rfloor$ (see Sections 4.5.5 and 4.5.7 in Ref. 7). Therefore, one can use $G_\alpha(x, y)$ in place of $K(x, y)$ in any of the formulas (4), (6), (7), or (5) so long as the set of covariates $\{f_p(x)\colon p = 1, \ldots, N\}$ contain all monomials of order less than or equal to $\lfloor \alpha/2 \rfloor$. Note that when $0 < \alpha < 2$ the function $G_\alpha(x, y)$ corresponds to a generalized covariance function of order $k_0 = 0$ for the fractional Brownian covariance $K(x, y) = |x|^\alpha + |y|^\alpha - |x - y|^\alpha$. Another important case occurs when $d = 2$, $\alpha = 2$, and the set of covariates $\{f_p(x)\colon p = 1, \ldots, N\}$ are *exactly* the set of monomials of order less than or equal to $k_0 = 1$. In this case, kriging with the generalized covariance $G_2(x, y)$ yields the classic thin-plate spline smoother (see Section 2.4 of Ref. 4).

One advantage provided by the decomposition given in (8) is that one needs not worry about modeling $a_{\boldsymbol{j}}(y), b_{\boldsymbol{j}}(x)$, or $c_{\boldsymbol{i},\boldsymbol{j}}$, when the set of covariates $\{f_p(x)\colon p = 1, \ldots, N\}$ contain all monomials of order less than or equal to k_0 since it has no impact on prediction or prediction mean square error. A particularly cogent example is the invariance of kriging to adding a random, mean zero, polynomial of order k_0 to Z. If the random coefficients have finite second moments, then the only effect on the covariance structure of Z is through $a_{\boldsymbol{j}}(y), b_{\boldsymbol{j}}(x)$, and $c_{\boldsymbol{i},\boldsymbol{j}}$ in (8). Therefore, it is unnecessary to model any such additive random polynomial in Z since it is inconsequential to kriging (so long

Encyclopedia of Environmetrics, Online © 2006 John Wiley & Sons, Ltd.

This article was published in *Encyclopedia of Environmetrics Second Edition* in 2012 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470057339.vak003.pub2

as the covariates f_p contain a sufficient number of monomials).

Estimating the Covariance Structure

Throughout the above exposition it has been assumed that the covariance function $K(x, y)$ and the standard deviation σ are both known. In practice, however, this is almost never the case. There are a multitude of different techniques for estimating K and σ. Examples include variogram model fitting, spectral density estimation, maximum likelihood (or REML) estimation, Bayesian techniques, and cross validation. In this section, we present a short discussion of two popular techniques: REML estimation and cross validation.

If $K(x, y)$ is known up to some parameter vector θ, then maximum likelihood estimation (MLE) based on the data y provides a natural approach for estimating θ and σ. One of the difficulties with the MLE, in this case, is the presence of the additional (unknown) coefficients β in the data model $y \sim \mathcal{N}[X\beta, \Sigma_{\theta,\sigma}]$, where $\Sigma_{\theta,\sigma}$ denotes the covariance matrix of y as it depends on the unknown parameters θ and σ. REML estimation circumvents this problem by restricting the data to linear combinations of the data which do not depend on β. For example, assume there exists a matrix M such that the rows form linearly independent vectors which are orthogonal to the columns of X. Notice this implies that $MX = 0$ and therefore $My \sim \mathcal{N}[0, M\Sigma_{\theta,\sigma}M^t]$ which does not depend on β. The REML estimate is now defined as the maximizer of the likelihood for My over θ and σ. REML estimation is particularly well suited for the generalized covariance functions discussed in section titled The Variogram and Generalized Covariance Functions. In particular, if the covariates $\{f_p(x): p = 1, \ldots, N\}$ contain all monomials with degree less than or equal to some nonnegative integer k_0 and $K(x, y)$ is decomposed as in (8), then the covariance matrix $M\Sigma_{\theta,\sigma}M^t$ only depends on $G(x, y)$. This fact along with the discussion presented in section titled The Variogram and Generalized Covariance Functions provides a unified methodology for estimation and kriging within the class of generalized covariance functions.

Cross validation is another approach which can be used to estimate σ and a covariance parameter θ. Like REML, cross validation only requires modeling the generalized covariance functions when the covariates are sufficiently rich. The simplest form of cross validation divides the data into a test set and a training set. The training set is used to predict (with kriging) the test set values. Prediction squared error is measured then subsequently minimized to find the optimal values of θ and σ. This is the basic cross validation technique (with many generalizations). Since kriging only depends on the generalized covariance functions $G(x, y)$ (when the covariates contain enough monomials), the corresponding cross validation estimates of θ and σ only depend on $G(x, y)$. We mention a minor simplification that occurs when performing cross validation which can reduce the number of kriging parameters by one. When the covariance structure of Z is known up to a proportionality constant, so that $\mathrm{cov}(Z(x), Z(y)) = \theta^2 K_0(x, y)$, for example, then the kriging prediction only depends on σ^2 and θ^2 through the ratio σ^2/θ^2. In particular, one can simply replace $K(x, y)$ with $K_0(x, y)$ and σ^2 with σ^2/θ^2 in (4), (6), or (7) to obtain the same value. In doing so, the number of parameters which needs to be estimated is reduced by one. Note that a derivation of this fact follows simply from the GLS characterization of kriging given in (4).

Extensions to the Basic Model

The kriging methodology presented here constitutes only a small portion of what appears in the current literature on kriging. In this section, we discuss some of the generalizations on the basic kriging model presented in section titled The Basic Model, with the understanding that this is far from an exhaustive review. We loosely characterize generalizations by those which preserve some facet of linearity and those which do not.

There are a whole class of kriging problems devoted to linear generalizations of the basic kriging methodology. For example, one can generalize the stipulation, given in section The Basic Model, that the observations and predictions are based on pointwise observations of Y to the case that one observes or predicts linear functionals of Y. For example, in *block kriging* the problem is to predict a block average $\frac{1}{|B|}\int_B Y(x)dx$ where B is some d-dimensional region and $|B|$ is the volume. In another situation, the observations or predictions are derivatives of Y. Another linear generalization, called *cokriging*, deals

Encyclopedia of Environmetrics, Online © 2006 John Wiley & Sons, Ltd.

This article was published in *Encyclopedia of Environmetrics Second Edition* in 2012 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470057339.vak003.pub2

with the multidimensional setting where Y is a random vector field, mapping $\mathbb{R}^d$ to $\mathbb{R}^m$ for some $m > 1$. Many of the results found in sections titled Introduction and Estimating the Covariance Structure extend easily to these examples.

In contrast to the linear case, some generalizations introduce nonlinearity into the kriging methodology. Consider, for example, the situation that the observations or predictions are nonlinear functions of Y (indicators, maximum values, etc.). Another example postulates nonlinearity in the mean function of Y. Yet another considers predictors $\widehat{Y}(x_0)$ which are nonlinear functions of the data $y_1, \ldots, y_n$. This is often used when the data is highly non-Gaussian so that restricting to linear combinations of the data may result in poor performance. Two cogent examples are *robust kriging* and *Bayesian kriging*. In robust kriging, one attempts to mitigate the effects of outliers or highly skewed/non-Gaussian distributions. In *Bayesian kriging*, one incorporates a prior distribution on the mean function of $Y(x)$ which can produce predictors which are nonlinear functions of the data $y_1, \ldots, y_n$.

References

[1] Krige, D. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand, *Journal of the Chemical, Metallurgical and Mining Society of South Africa* **52**, 119–139.

[2] Matheron, G. (1963). Principles of geostatistics, *Economic Geology* **58**, 1246–1266.

[3] Cressie, N. (1990). The origins of Kriging, *Mathematical Geology* **22**, 239–252.

[4] Wahba, G. (1990). *Spline Models for Observational Data*, SIAM, Philadelphia.

[5] Gandin, L. S. (1965). *Objective analysis of meteorological fields*, Israel Program for Scientific Translations, Jerusalem.

[6] Armstrong, M. (1998). *Basic Linear Geostatistics*, Springer-Verlag, Berlin.

[7] Chiles, J-P. & Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*, John Wiley & Sons, New York.

[8] Cressie, N. (1991). *Statistics for Spatial Data*, John Wiley & Sons, New York.

[9] Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*, Oxford University Press, Oxford.

[10] Kitanidis, P. K. (1997). *Introduction to Geostatistics: Applications to Hydrogeology*, Cambridge University Press, Cambridge.

[11] Olea, R. (1999). *Geostatistics for Engineers and Earth Scientists*, Kluwer, Dordrecht.

[12] Stein, M. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*, Springer, New York.

[13] Wackernagel, H.(1998). *Multivariate Geostatistics*, 2nd Edition, Springer-Verlag, Berlin.

[14] Matheron, G. (1973). The intrinsic random functions and their applications, *Advances in Applied Probability* **5**, 439–468.

(*See also* **Kriging, asymptotic theory**; **Covariogram, examples of**; **Kriging for functional data**; **Multivariate kriging**; **Random field, Gaussian**; **Random fields**; **Screening effect**)

ETHAN B. ANDERES

Encyclopedia of Environmetrics, Online © 2006 John Wiley & Sons, Ltd.

This article was published in *Encyclopedia of Environmetrics Second Edition* in 2012 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470057339.vak003.pub2