Blackbox Kriging: Spatial Prediction Without Specifying Variogram Models

Ronald Paul BARRY and Jay M. VER HOEF

This article proposes a new approach to kriging, where a flexible family of variograms is used in lieu of one of the traditionally used parametric models. This nonparametric approach minimizes the problems of misspecifying the variogram model. The flexible variogram family is developed using the idea of a moving average function composed of many small rectangles for the one-dimensional case and many small boxes for the two-dimensional case. Through simulation, we show that the use of flexible piecewise-linear models can result in lower mean squared prediction errors than the use of traditional models. We then use a flexible piecewise-planar variogram model as a step in kriging the two-dimensional Wolfcamp Aquifer data, without the need to assume that the underlying process is isotropic. We prove that, in one dimension, any continuous variogram with a sill can be approximated arbitrarily close by piecewise-linear variograms. We discuss ways in which the piecewise-linear variogram models can be modified to improve the fit of the variogram estimate near the origin.

Key Words: Anisotropy; Geostatistics; Moving averages.

1. KRIGING AND THE CHOICE OF VALID VARIOGRAM FAMILIES

Ordinary kriging is a method for predicting the value of a random process at a specific location in a region, given a dataset consisting of measurements of the random process at a variety of locations in the region. Specifically, let $Z(s_i)|i = 1, \ldots, n$ be a set of measurements at locations $s_1, \ldots, s_n$ in an m-dimensional region D. These measurements are assumed to be one realization of a random process $Z(\cdot)$ with the following properties:

1. $E[Z(\cdot)] = \mu$
2. $2\gamma(h) = var\left(Z(s) - Z(s-h)\right)$, for all $s, s-h$ in D, exists and only depends on h

(Cressie 1993, p. 40). These two assumptions form the intrinsic stationarity hypothesis (Cressie, p. 60), and the variogram is defined to be the function $2\gamma(h) = \text{var}\big(Z(s) -$

Ronald Paul Barry is Associate Professor, Department of Mathematical Sciences, University of Alaska Fairbanks, Fairbanks, AK 99775-6660, ffrpb@aurora.alaska.edu. Jay M. Ver Hoef is a Biometrician, Alaska Department of Fish and Game, Division of Wildlife Conservation, Fairbanks, AK 99701, ffjmv@aurora.alaska.edu.

Journal of Agricultural, Biological, and Environmental Statistics, Volume 1, Number 3, Pages 297–322

$Z(s-h)\big)$. Additionally, *isotropy* is often assumed, where $2\gamma(h) = 2\gamma(\|h\|)$; that is, the variogram depends only on the length of h and not its direction. Ordinary kriging is the best linear, unbiased predictor (Cressie 1993) for the random process at a specific location. To predict at the location s_0 in D, use $\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i z(s_i)$, where $\sum \lambda_i = 1$, which guarantees unbiasedness so that $\mathrm{E}(\hat{Z}(s_0) - Z(s_0))^2$, the mean squared prediction error, is minimized.

It is straightforward to find λ_i if $2\gamma(h) = \mathrm{var}(Z(s_i) - Z(s_j) | s_i - s_j = h)$ is known. Unfortunately, we will almost never know this, but will have to estimate it from the one realization of the random process that we have sampled. Using the data to estimate $2\gamma(h)$ will present two problems. First, if care is not taken in the estimation of $2\gamma(h)$, negative estimates of the mean squared prediction error $\mathrm{E}(\hat{Z}(s_0) - Z(s_0))^2$ may result. Second, the error in the estimation of $2\gamma(h)$ will increase the imprecision of the spatial prediction.

Fortunately, users of kriging have found ways of dealing with these problems. The first problem is handled by assuming that the true variogram is a member of a parameterized set of valid variograms. By valid, we mean that we are assured of getting nonnegative estimates of the mean squared prediction errors. To ensure this, a family of variogram models is chosen where all functions in the family are conditionally negative definite so that

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j 2\gamma(s_i - s_j) \leq 0$$

for any whole number n, any locations $s_1, \ldots, s_n$, and any set of real numbers $a_1, \ldots, a_n$ such that $\sum_{i=1}^{n} a_i = 0$. The second problem has been analyzed by several researchers, including Stein and Handcock (1989), but most users simply assume that the fitted variogram model is the true variogram. Thus, ordinary kriging is carried out in the following stages:

1. Compute the empirical (also called experimental) variogram (see Cressie 1993, p. 69), using the data $\{Z(s_j)\}_{j=1}^{N}$.
2. Choose a family of valid variogram models, such as the exponential models (see Cressie 1993, p. 61), where all of the functions in the family are conditionally negative definite.
3. Estimate the variogram using some technique to fit the valid variogram to the empirical variogram.
4. Assume that the fitted, valid variogram is the true variogram $2\gamma(h)$.
5. Calculate the best linear predictor and its mean squared prediction error. The minimum prediction error estimator is $\hat{Z}(s_0) = \sum_{j=1}^{N} \lambda_j Z(s_j)$, where

$$(\lambda_1, \cdots, \lambda_N) = \left(\gamma + \mathbf{1}\frac{1 - \mathbf{1}'\Gamma^{-1}\gamma}{\mathbf{1}'\Gamma^{-1}\mathbf{1}}\right)' \Gamma^{-1}. \tag{1.1}$$

Here, $\gamma = \big(\gamma(s_0 - s_1), \cdots, \gamma(s_0 - s_N)\big)'$, and Γ is the $N \times N$ matrix with $\gamma(s_i - s_j)$

as the (i, j)th entry. The mean squared prediction error is

$$\sigma^2(s_0) = \boldsymbol{\gamma}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma} - \frac{(1 - \mathbf{1}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma})^2}{\mathbf{1}'\boldsymbol{\Gamma}^{-1}\mathbf{1}}. \tag{1.2}$$

This article is primarily concerned with the selection of a family of valid variograms to fit to the empirical variogram. Obviously, if the true variogram of the random process is not closely approximated by a member of the selected variogram family, the predictions and estimated prediction errors might be seriously biased. Practitioners will often "eyeball" a plot of the empirical variogram to help them select a variogram family. Unfortunately, this will not work very well for data drawn from a two- or three-dimensional domain. For this reason, practitioners often assume isotropy, which makes modeling the variogram much easier. This further restricts the choice of variogram families so that the selected model may be far from the true variogram. For a list of isotropic models, see Cressie (1993, p. 63). Further, the empirical variogram may not look like one of the known variogram families. One solution is to use families of variograms that are sufficiently flexible to fit almost all variograms, isotropic or not—the nonparametric approach.

This article will introduce flexible families of valid variograms for use with one- and two-dimensional data. We will show that arbitrary variograms can be approximated by members of these new flexible families. We will also show how special variogram models can be constructed that are more accommodating near the origin of the variogram than farther from the origin—because the quality of the approximation to the variogram near the origin is of the greatest importance (Stein 1988), flexibility of the nonparametric variograms near the origin is desirable.

Shapiro and Botha (1991) introduced the idea of nonparametric variogram fitting. They introduced a family of variograms that is flexible enough to approximate almost any variogram so that the misspecification of the variogram model and its attendant bias is no longer a worry. As in any nonparametric regression, there is a problem of overfitting, where the variogram model fits the data too well, as though there was no random error in the empirical variogram. Conditions can be put on the variogram estimate to damp out oscillatory behavior and avoid overfitting. Shapiro and Botha discovered a family of cosine series variograms that is sufficiently flexible to fit arbitrary variograms. Parameterized by the coefficients $(c_0, a_1, \ldots, a_m)$, where m is less than the number of data, the variogram family is

$$2\gamma(h|c_0, a_1, \ldots, a_m) = c_0 + \sum_{j=1}^{m} a_j - \sum_{j=1}^{m} a_j \cos(\delta j h),$$

where $\delta > 0$ and $a_j > 0$. Shapiro and Botha recommended using m one less than the number of data. Because this family is very flexible for large amounts of data (large n), they prevent overfitting by explicitly bounding the slope of the fitted variogram, by ensuring that the fitted variogram is convex, or by ensuring that the fitted variogram is monotone increasing. One feature of their method is that the parameter estimates can be obtained by quadratic programming.

Shapiro and Botha (1991) introduced a family of valid two-dimensional variograms based on linear combinations of Bessel functions. The parameters of the best-fitting

variogram can be obtained by quadratic programming, and overfitting is avoided by again bounding the derivative of the fitted variogram, by forcing the fitted variogram to be convex, or by forcing it to be monotone. Unfortunately, all of these variograms are isotropic, which means that $2\gamma(h)$ depends only on the magnitude $\|h\|$ of h, and not on the direction. These flexible variogram families are all amenable to solution by quadratic programming, but they cannot model anisotropy and cannot be made more flexible near the origin.

2. FLEXIBLE VARIOGRAM MODELS AND MOVING AVERAGES

Matern (1986) and Webster (1985) have shown how moving average functions can be used to construct valid variogram families. The moving average idea can be used to define new flexible variogram families that complement the flexible variogram family of Shapiro and Botha (1991) and can accommodate anisotropy and have greater flexibility near the origin.

To obtain a valid variogram in m dimensions, a function $f : \mathcal{R}^m \to \mathcal{R}$ is selected such that

$$2\gamma(\mathbf{h}) = \int_{\mathcal{R}^m} \big(f(\mathbf{x}) - f(\mathbf{x}-\mathbf{h})\big)^2 \, \mathbf{dx} < \infty \quad \text{for all} \quad \mathbf{h} \in \mathcal{R}^m. \tag{2.1}$$

As proven in Appendix A, we can use such a function to construct a random process with the variogram $2\gamma(\mathbf{h}) + 2c_0$ (where $2c_0$, the nugget effect of the process, is nonnegative) for $h \neq 0$. Thus, any function f that satisfies Equation (2.1) yields a valid variogram, which can be described explicitly as long as we can carry out the integration.

To obtain a family of parameterized variograms, we need to start with a family of functions $f(\mathbf{x}|\theta) : \mathcal{R}^m \to \mathcal{R}$, parameterized by θ, where

$$2\gamma(\mathbf{h}|\theta) = \int_{\mathcal{R}^m} (f(\mathbf{x}|\theta) - f(\mathbf{x}-\mathbf{h}|\theta))^2 \mathbf{dx} < \infty \quad \text{for all} \quad \mathbf{h} \in \mathcal{R}^m \tag{2.2}$$

for every choice of θ from a set of possible parameters. For example, consider the family of functions $f(x|a,b) = a\mathcal{I}(0 \leq x \leq b)$, where $a, b > 0$ (here, $\mathcal{I}$ is the indicator function). First, plugging the function into (2.2), we see that

$$\begin{aligned} 2\gamma(\mathbf{h}) &= \int_{\mathcal{R}} \big(f(\mathbf{x}) - f(\mathbf{x}-\mathbf{h})\big)^2 \mathbf{dx} \\ &= \int_{\mathcal{R}} \big(a\mathcal{I}(0 \leq x \leq b) - a\mathcal{I}(h \leq x \leq h+b)\big)^2 \mathbf{dx} < \infty \quad \text{for all} \quad \mathbf{h} \in \mathcal{R}. \end{aligned}$$

Essentially we are overlaying two rectangles, both of height a and width b, with one of the rectangles offset by h with respect to the other. Where the rectangles overlap, they will cancel. Where they do not overlap, we square the height and take the resulting area. Clearly, when $h = 0$ the rectangles cancel and $2\gamma(0|a,b) = 0$. As h increases, the variogram increases linearly, until, when $h = b$, the rectangles no longer overlap. For all $a, b > 0$, the integral is finite, so we know without need of further checking that the

resulting variograms will be valid. In fact, the valid variogram family obtained is

$$2\gamma(h|a,b) = \int_{-\infty}^{\infty} [f(x|a,b) - f(x-h|a,b)]^2 dx = \begin{cases} 0 & \text{if } h = 0; \\ 2a^2|h| & \text{if } 0 < |h| \le b; \\ 2a^2 b & \text{if } |h| > b, \end{cases}$$

the linear-with-sill variograms. The construction described in Appendix A gives us, in addition, a nice physical interpretation of the random process. The moving averaging corresponds to a smoothed version of a large number of small independent and random phenomena.

The moving average approach can help find new variogram families, especially when the smoothing concept matches our physical intuition. In general, however, we would prefer to have the data determine the shape of the variogram as much as possible. Imagine that the true (but unknown) variogram is merely continuous for $h \neq 0$. What we will do is truncate the variogram outside the range of the data. This truncated variogram will turn out always to have a moving average function. We can approximate the moving average function by a piecewise constant moving average function to obtain a variogram that is close to the true variogram. The family of valid variograms derived from piecewise-constant moving average functions will then be a flexible variogram family, flexible in the sense that they can approximate any continuous variogram with sill.

3. ONE-DIMENSIONAL FLEXIBLE VARIOGRAM MODELS

We will now introduce a new flexible variogram family based on moving averages. For any positive integer k and range $c > 0$, define the piecewise constant function,

$$f(x|a_1,\ldots,a_k,c,k) = \sum_{j=1}^{k} a_j \mathcal{I}\left(\frac{(j-1)c}{k} < x \le \frac{jc}{k}\right).$$

This function has support $(0,c]$, and consists of k steps of equal width c/k and heights $a_1,\ldots,a_k$. The resulting family of valid variograms can be readily found from (2.1). If h is an integer multiple m of the width of each piecewise constant step so that $|h| = mc/k$, then the breaks line up and the variogram is very simple to compute:

$$2\gamma(h) = \frac{c}{k}\sum_{i=1}^{m} a_i^2 + \frac{c}{k}\sum_{i=m+1}^{k} (a_i - a_{i-m})^2 + \frac{c}{k}\sum_{i=k+1}^{k+m} a_{i-m}^2.$$

The first and third terms involve parts of the moving average functions that do not overlap, while the second term involves the overlap. Multiplying this out, we get

$$2\gamma(h) = \frac{2c}{k}\sum_{i=1}^{k} a_i^2 - \frac{2c}{k}\sum_{i=m+1}^{k} a_i a_{i-m}.$$

When h is not an integer multiple of c/k, we can use the fact that the variogram is piecewise-linear (the integral of a piecewise constant function is always piecewise-linear), and obtain the variogram value through interpolation. For h when $0 < h < c$,

we find the integers m_l and m_u such that $m_l c/k$ is lower and $m_u c/k$ is higher than h: $m_l = \lfloor (hk/c) \rfloor$ is the nearest integer less than hc/k and $m_u = \lceil (hk/c) \rceil$ is the nearest integer greater than hc/k. Then $V = \big(h - (m_l c/k)\big)/(c/k)$ is the fraction of the way that h is from m_l to m_u. Then

$$\begin{aligned} 2\gamma(h) &= (1-V)2\gamma(m_l c/k) + V 2\gamma(m_u c/k) \\ &= \frac{2c}{k}\sum_{i=1}^{k} a_i^2 - \frac{2c}{k} V \sum_{i=1}^{k} a_i a_{i-\lceil \frac{|h|k}{c} \rceil} - \frac{2c}{k}(1-V)\sum_{i=1}^{k} a_i a_{i-\lfloor \frac{|h|k}{c} \rfloor} \end{aligned} \quad (3.1)$$

for $|h| < c$, where $a_i = 0$ for $i < 1$, and

$$2\gamma(h|a_1, \ldots, a_k, c, k) = \frac{2c}{k}\sum_{i=1}^{k} a_i^2$$

for $|h| > c$. These variograms are piecewise-linear, continuous and linear on each piece $\big(c(i-1)/k, ci/k\big)$, for $i = -k+1, \ldots, k$, and elsewhere constant. In Appendix B we prove that, as k becomes large, the family of piecewise-linear variograms approaches arbitrarily close to any continuous variogram with a sill.

4. TWO-DIMENSIONAL FLEXIBLE VARIOGRAM MODELS

If we cut a $c \times d$ rectangle m times in the y direction and n times in the x direction and pick a function that is constant on each of the resulting $m \times n$ subrectangles, we get the analog in two dimensions of a piecewise constant function. Formally, for a set of parameters $a_{i,j}, i = 1, \ldots, n, j = 1, \ldots, m$, we get

$$f(x_1, x_2) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{i,j} \mathcal{I}\left[\big(c(i-1)/m < x_1 < ci/m\big)\big(d(j-1)/n < x_2 < dj/n\big)\right],$$

which is a step function with height $a_{i,j}$ in the subrectangle in the ith row and the jth column.

We can find the corresponding variogram much as we did in the one-dimensional case. First, for any lag (h_1, h_2) in two dimensions, where h_1 and h_2 are integer multiples of c/m and d/n, respectively, the variogram is easy to compute, because the subrectangles of the moving average function and the lagged moving average function line up:

$$2\gamma(h_1, h_2) = \frac{2cd}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} a_{i,j}^2 - \frac{2cd}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} a_{i,j} a_{i-\lceil \frac{|h_1|m}{c} \rceil, j-\lceil \frac{|h_2|n}{d} \rceil}.$$

For the general case, we can use the fact that the integral of a function that is constant on each subrectangle in a grid is piecewise-planar. Then we can obtain the value of the variogram at an arbitrary lag (h_1, h_2) by making a planar interpolation of the four variogram values on the corners of the subrectangle containing the lag (h_1, h_2).

Let V be the decimal part of $|h_2|n/d$ and W be the decimal part of $|h_1|m/c$. Then the interpolated variogram value is

$$\begin{aligned}
2\gamma(h_1,h_2) \;=\; & \frac{2cd}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}a_{i,j}^2 \\
& -\frac{2cd}{mn}V\left\{W\sum_{i=1}^{m}\sum_{j=1}^{n}a_{i,j}a_{i-\lceil\frac{|h_1|m}{c}\rceil,j-\lceil\frac{|h_2|n}{d}\rceil}\right. \\
& \qquad \left.+(1-W)\sum_{i=1}^{m}\sum_{j=1}^{n}a_{i,j}a_{i-\lfloor\frac{|h_1|m}{c}\rfloor,j-\lceil\frac{|h_2|n}{d}\rceil}\right\} \\
& -\frac{2cd}{mn}(1-V)\left\{W\sum_{i=1}^{m}\sum_{j=1}^{n}a_{i,j}a_{i-\lceil\frac{|h_1|m}{c}\rceil,j-\lfloor\frac{|h_2|n}{d}\rfloor}\right. \\
& \qquad \left.+(1-W)\sum_{i=1}^{m}\sum_{j=1}^{n}a_{i,j}a_{i-\lfloor\frac{|h_1|m}{c}\rfloor,j-\lfloor\frac{|h_2|n}{d}\rfloor}\right\}, \qquad (4.1)
\end{aligned}$$

for $0 < |h_1| \le c, 0 < |h_2| \le d$, and

$$2\gamma(h_1,h_2|a_{i,j},i,j=1,\ldots,k) = \frac{2c}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}a_{i,j}^2$$

for $|h_1| > c$ or $|h_2| > d$. This variogram is continuous, and planar over each rectangle $\{h_1,h_2 : (i-1)c/m \le h_1 \le ic/m, (j-1)d/n \le h_2 \le jd/n, i,j = -k+1,\ldots,k\}$. This function is also a quadratic form that will be used in the fitting of this variogram model in Section 7.

5. FITTING FLEXIBLE VARIOGRAM MODELS

Fitting a one- or two-dimensional flexible variogram model to an empirical variogram is simply a special case of nonlinear minimization. In this section we will discuss the details of fitting the flexible variogram models.

Fitting a one-dimensional flexible model is done in three steps:

1. The empirical variogram is produced,
2. The nugget effect is estimated,
3. The parameters of the valid variogram are estimated.

The empirical variogram is defined as

$$2\hat{\gamma}_{\text{emp}}(h) = \frac{1}{N(h)}\sum_{|s_i-s_j|=h}\big(z(s_i)-z(s_j)\big)^2, \qquad (5.1)$$

where $N(h)$ is the number of pairs of $z(s_i), z(s_j)$ such that $|s_i - s_j| = h$. Note that the empirical variogram is not necessarily a valid variogram, because we cannot be assured

that it is conditionally negative definite (Cressie 1993, p. 90). Each value of the empirical variogram $2\hat{\gamma}_{\text{emp}}(h)$ is associated with a weight $N(h)$, reflecting the number of pairs of locations averaged to get that value.

If the locations are irregularly spaced along the transect, $N(h)$ will tend to never be bigger than one. Researchers often respond by binning the data, essentially pretending that the locations are regular. The disadvantages of this are the loss of information about the variogram at small lags and an artificial inflation of the size of the estimated nugget. If we cannot assume that the nugget effect is negligible, we will have to estimate it. We have found that simultaneously estimating the magnitude of the nugget effect and the variogram parameters leads to overestimation of the nugget effect, so we recommend estimating the nugget effect before estimating the other model parameters.

The underestimation of the nugget effect when the nugget effect is estimated simultaneously with the parameters $a_1, \ldots, a_k$ is a result of overfitting. When the nugget effect is large, the realizations of the process are noisy, so the empirical variogram tends to be quite rough. Because the sill is the value of the variogram as h exceeds the range of the process, it is just $(2c/k)\sum_{i=1}^{k} a_i^2 + 2c_0$. For a fixed sill, there is a trade-off between the size of the nugget $2c_0$ and the sum of the squares of the parameters $a_1, \ldots, a_k$. If the nugget is overestimated, the $|a_i|$ must be small. This limits the flexibility of the variogram estimate. Thus, the minimization algorithm attempts to fit the rough empirical variogram at the expense of the estimate of the nugget effect. By estimating the nugget effect separately from the rest of the parameters, these difficulties are avoided. Figure 1 shows the true variogram, the empirical variogram, and the fitted 15-piece variogram with a range of 50, for a single realization from the process that gave the results in Table 1. In Figure 1a, the estimation of the nugget effect and the moving average parameters was done simultaneously. Here, notice that the fitted variogram was able to closely approximate the roughness of the empirical variogram, but at the cost of an underestimated nugget effect. In Figure 1b, the nugget effect was estimated by fitting a line, via weighted least squares, to the first five lags prior to estimating the moving average parameters. In this case, the estimate of the nugget effect is larger, in fact quite close to the true value of 20.0; as a result the fitted variogram is smoother.

Next, the parameters of the flexible model are estimated. This is done by minimizing either the weighted or nonweighted sum of squared errors (SSE). If the locations were regular, we would minimize the weighted least squares formula:

$$\text{SSE}(a_1, \ldots, a_k, c, k) = \sum_h \left(N(h)/\hat{\gamma}^2_{\text{emp}}(h)\right) [\hat{\gamma}(h|a_1, \ldots, a_k, c, k) + 2c_0 - \hat{\gamma}_{\text{emp}}(h)]^2$$

(Cressie 1993, p. 99). Note that the previously estimated nugget effect $2c_0$ is added to the variogram estimate $\hat{\gamma}(h|a_1, \ldots, a_k, c, k)$.

When the locations are irregularly spaced and the data has not been binned, the empirical variogram may be a very poor estimator of the true variogram at any particular lag h. In this case we forego the weighting and minimize the nonweighted sum of squared error:

$$\text{SSE}(a_1, \ldots, a_k, c, k) = \sum_h [\hat{\gamma}(h|a_1, \ldots, a_k, c, k) + 2c_0 - \hat{\gamma}_{\text{emp}}(h)]^2.$$

Cressie discusses several other possible formulas that can be minimized to obtain variogram estimates, including generalized least squares (Cressie 1993, p. 95) and robust formulas (Cressie, p. 97).

In the simulation described in Section 6, we minimized the weighted least squares formula using the multidimensional simplex method for nonlinear minimization (Nelder and Mead 1965). In this simulation we fixed the range c of the flexible model, and the number of steps k, and found $a_1, \ldots, a_k$ that minimized the weighted least squares formula.

Estimating the parameters of a two-dimensional flexible model involves straightforward extensions of the methods used in the one-dimensional case. In the analysis we present in Section 7, we fitted a two-dimensional flexible model to the empirical variogram from the Wolfcamp Aquifer dataset. In this analysis we decided that the nugget effect was negligible. If we had decided to estimate the nugget effect, we would have regressed the empirical variogram $2\hat{\gamma}_{\text{emp}}(h_1, h_2)$ against the magnitude of the lag $\|(h_1, h_2)\|$. We would have interpreted the intercept as the estimate $2c_0$ of the nugget effect.

We then sought the parameters $\mathbf{a} = (a_{1,1}, \ldots, a_{5,4})$ so that the function

$$\begin{aligned}\text{SSE}(a_{1,1}, \ldots, a_{5,4}) &= \sum_{i=1}^{84} \sum_{j=i+1}^{85} \Big[\big(z(x_i, y_i) - z(x_j, y_j)\big)^2 \\ &\qquad - 2\gamma(x_i - x_j, y_i - y_j | a_{1,1}, \ldots, a_{5,4}) \Big]^2 \quad (5.2)\end{aligned}$$

was minimized. Because the equation for the two-dimensional variogram [see Equation (4.1)] is a quadratic form in the parameters $\mathbf{a} = (a_{1,1}, \ldots, a_{5,4})$, the preceding equation can be written in the form

$$\text{SSE}(\mathbf{a}) = \sum_{i=1}^{84} \sum_{j=1}^{85-i} (\mathbf{a}' \mathbf{Q}_{\mathbf{i},\mathbf{j}} \mathbf{a} - Y_{i,j})^2. \quad (5.3)$$

We could have used a nonlinear minimization method that did not require knowing the gradient of the SSE, but for this analysis we used the fact that the gradient of (5.3) can be made explicit.

Let $a^{(n)}$ be the nth iterative approximation to the parameters $(a_{1,1}, \ldots, a_{5,4})$. The iterated equation was

$$\begin{aligned}\hat{\mathbf{a}}^{(\mathbf{n})} &= \hat{\mathbf{a}}^{(\mathbf{n}-1)} - s^{(n)} \sum_i \sum_j \Big(\hat{\mathbf{a}}^{(\mathbf{n}-1)'} \mathbf{Q}_{\mathbf{i},\mathbf{j}} \hat{\mathbf{a}}^{(\mathbf{n}-1)} \\ &\qquad - \big(z(x_i, y_i) - z(x_j, y_j)\big)^2 \Big) \hat{\mathbf{a}}^{(\mathbf{n}-1)'} \mathbf{Q}_{\mathbf{i},\mathbf{j}},\end{aligned}$$

where $s^{(n)}$, the step size, was increased every time the iteration led to a set of parameter estimates with a better least squares fit. When the least squares fit worsened, the parameter estimates were reset to the previous values, and the step size was reduced. This equation is just a "walk downhill" minimization of the form

$$\hat{\mathbf{a}}^{(\mathbf{n})} = \hat{\mathbf{a}}^{(\mathbf{n}-1)} - (\text{step}) * \left(\text{gradSSE}(\hat{\mathbf{a}}^{(\mathbf{n})}) \right).$$

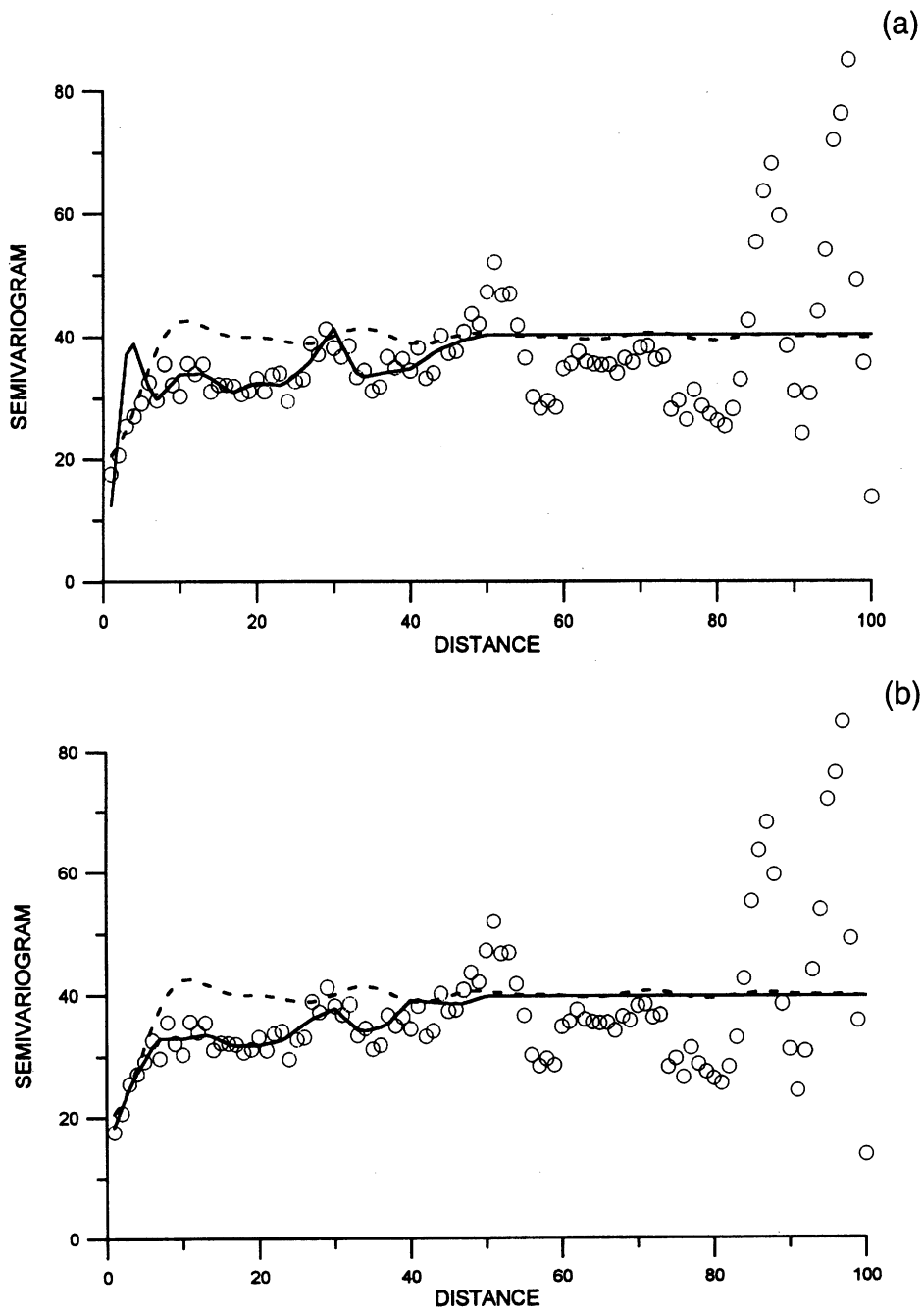

Figure 1. The Empirical Variogram and Fitted 15-Part Piecewise-Linear Variogram With a Range of 50, for One Realization of a Process With True Variogram $2\gamma(h) = 20 + 10(1-2\sin(h/2)/h)+10(1-3\sin(h/3)/h)$. *(a) Simultaneous estimation of nugget effect and moving average parameters. (b) Separate estimation of nugget effect and moving average parameters.* ◯ *= empirical semivariogram;* —— *= fitted flexible model;* - - - *= true semivariogram.*

Table 1. Results of 1000 Simulations Using the Semivariogram Model $2\gamma(h) = 20+10(1-2\sin(h/2)/h)+10(1-3\sin(h/3)/h)$ for $h\neq 0$, Predicting $Z(51)$

	True	*Spherical*	*k5c100*	*k9c100*	*k15c100*	*k15c50*
Mean$((\hat{Z}(51)-Z(51))^2)$	22.9	24.3	29.6	24.8	24.9	24.6
Mean$(\hat{\sigma}^2(51))$	23.2	23.9	20.8	21.9	22.5	22.6
Coverage	.948	.945	.885	.921	.938	.937

We have also fit two-dimensional variogram models to data by minimizing the SSE directly, using minimization techniques that do not require the gradient (such as the function NLMIN in S-plus), and these gradient-free methods give parameter estimates comparable to those obtained when using the gradient. Because constructing $Q_{i,j}$ is difficult, we would generally recommend minimizing Equation (5.2) directly, using a gradient-free method.

6. SIMULATION COMPARISONS FOR A ONE-DIMENSIONAL PROCESS

The flexible variogram model should be able to fit empirical variograms that are unlike any of the commonly used variogram models. In practice, most researchers use linear, linear-with-sill, spherical, and exponential variograms in fitting one-dimensional variograms. With the flexible variogram models the practitioner must set the number of pieces k and the range c. Another question to be answered is whether the mean squared prediction error is lower when using a flexible variogram model than for a nonflexible model.

We simulated random processes with variograms that were linear combinations of two different wave-effect variograms (Cressie 1993, pp. 62, 85), and computed the prediction by fitting a spherical variogram model (a reasonable choice given the shape of the resulting empirical variograms) and then by fitting piecewise-linear variograms for various values of pieces k and ranges c. This was repeated many times, giving us an estimate of the mean squared prediction error.

6.1 THE SIMULATION

Using the Cholesky Decomposition Method described by Cressie (1993, p. 210), 1,000 realizations of a random process on the integers $Z(1), \ldots, Z(101)$ were simulated with a semivariogram that was a linear combination of two wave-effect semivariograms:

$$\gamma(h) = 0.2 + 15(1 - 2\sin(h/2)/h) + 25(1 - 3\sin(h/3)/h)$$

for $h \neq 0$. Another 1,000 realizations were generated using the semivariogram

$$\gamma(h) = 20 + 10(1 - 2\sin(h/2)/h) + 10(1 - 3\sin(h/3)/h)$$

for $h \neq 0$. We choose these variograms because they yield realized empirical variograms that could arguably come from a process with a spherical variogram. The first variogram

has relatively little nugget effect, with a nugget that is .2/40.2 or .5% of the sill, which results in smoother empirical variograms. The second variogram has a nugget that is 20/40 or half of the sill.

In each realization, $Z(51)$ was held out, and the empirical variogram $2\hat{\gamma}_{\text{emp}}(h)$ was computed from the values $Z(1), \ldots, Z(50), Z(52), \ldots, Z(101)$, using Equation (5.1). The kriging predictor $\hat{Z}_{\text{true}}(51)$ was calculated using the true variogram in (1.1) and (1.2). A spherical variogram with nugget,

$$2\hat{\gamma}_{\text{sph}}(h|a,b,c_0) = \begin{cases} 0 & h = 0, \\ c_0 + b\left((3/2)(|h|/a) - (1/2)(|h|/a)^3\right) & h \le a, \\ c_0 + b & |h| \ge a \end{cases}$$

(Cressie 1993, p. 61), was fitted to each empirical variogram, and the resulting fitted variogram used in (1.1) and (1.2) to compute the ordinary kriging predictor $\hat{Z}_{\text{sph}}(51)$. The shape of the empirical variograms for the first few lags was suggestive of a spherical or exponential variogram model, and one of these would undoubtedly be used by most practitioners on any of the realized empirical variograms. The fitting was implemented using the multidimensional simplex method for minimization (Nelder and Mead 1965), yielding a, b, c_0 that minimized the weighted least squares formula:

$$\text{SSE}(a,b,c_0) = \sum_{h=1}^{100} \left(N(h)/\hat{\gamma}^2_{\text{emp}}(h)\right) [\hat{\gamma}_{\text{sph}}(h|a,b,c_0) - \hat{\gamma}_{\text{emp}}(h)]^2.$$

A flexible variogram was fitted to each realized empirical variogram, as discussed in Section 5. With a range of $c = 100$, we fitted, successively, 5-, 9-, and 15-piece variograms, and, with a range of $c = 50$, a 15-piece variogram. In all cases we assumed a nonnegligible nugget effect. Using the fitted variograms in (1.1) and (1.2), the resulting predictors were $\hat{Z}_{\text{k5c100}}(51)$, $\hat{Z}_{\text{k9c100}}(51)$, $\hat{Z}_{\text{k15c100}}(51)$, and $\hat{Z}_{\text{k15c50}}(51)$. We also computed the true mean squared prediction error $\left(\hat{Z}(51) - Z(51)\right)^2$ and the estimated mean squared prediction error $\hat{\sigma}^2(51)$ for each of the four flexible models.

For each realization and each type of predictor, the 95% prediction interval $\hat{Z}(51) \pm 1.96\hat{\sigma}(51)$ was calculated, and whether or not the true value $Z(51)$ was inside the interval was noted. The coverage was the proportion of realizations in which the nominal 95% prediction interval actually contained the true value.

6.2 SIMULATION RESULTS

The results obtained when the true variogram had a small nugget effect are displayed in Table 2. As expected, using the true variogram gave the correct inference. Use of the spherical variogram caused an extreme overestimation of the mean squared prediction error $\hat{\sigma}^2(51)$, resulting in very conservative confidence intervals. k5c100 was not flexible enough, underestimating the mean squared prediction error and not estimating $Z(51)$ well, as evidenced by the large true prediction errors. The models k9c100 and k15c100 did better, but still overestimated $\hat{\sigma}^2(51)$, resulting in conservative prediction intervals. Using k15c50 gave, by far, the lowest level of prediction error, though the resulting prediction intervals were still somewhat conservative. Using the flexible variogram model, with

Table 2. Results of 1000 Simulations Using the Semivariogram Model $2\gamma(h) = .2+15(1-2\sin(h/2)/h)+25(1-3\sin(h/3)/h)$ for $h \neq 0$, Predicting $Z(51)$

	True	*Spherical*	*k5c100*	*k9c100*	*k15c100*	*k15c50*
Mean$((\hat{Z}(51)-Z(51))^2)$	.237	.693	.720	.712	.677	.541
Mean$(\hat{\sigma}^2(51))$	.241	5.209	.374	1.696	2.185	1.043
Coverage	.956	1.000	.727	.984	.995	.975

$k = 15$ pieces and a range of $c = 50$, was the superior method. Figure 2 shows boxplots of the actual prediction error $(\hat{Z}(51) - Z(51))$ for each of the five methods. Note that they all are approximately unbiased. Figure 3 shows the boxplots of the fourth root of the estimated mean squared prediction errors, $\sqrt{\hat{\sigma}(51)}$. Note that the rigid spherical model was very prone to occasional wild values, leading to its poorer performance. Figure 4 shows boxplots of the fourth root of the actual mean squared prediction errors $(\hat{Z}(51) - Z(51))^2$. Of course, knowing the true model lowered the actual prediction error and resulted in a constant estimated mean square prediction error.

The results obtained when the true variogram had a large nugget effect are displayed in Table 1. Using the true variogram gave the correct inferences. The spherical model did well, only slightly underestimating the mean squared prediction error. Again, k5c100 was too inflexible. k9c100, k15c100, and k15c50 performed essentially as well as the spherical model. The large nugget effect masked differences between the models, having made the empirical variograms much rougher.

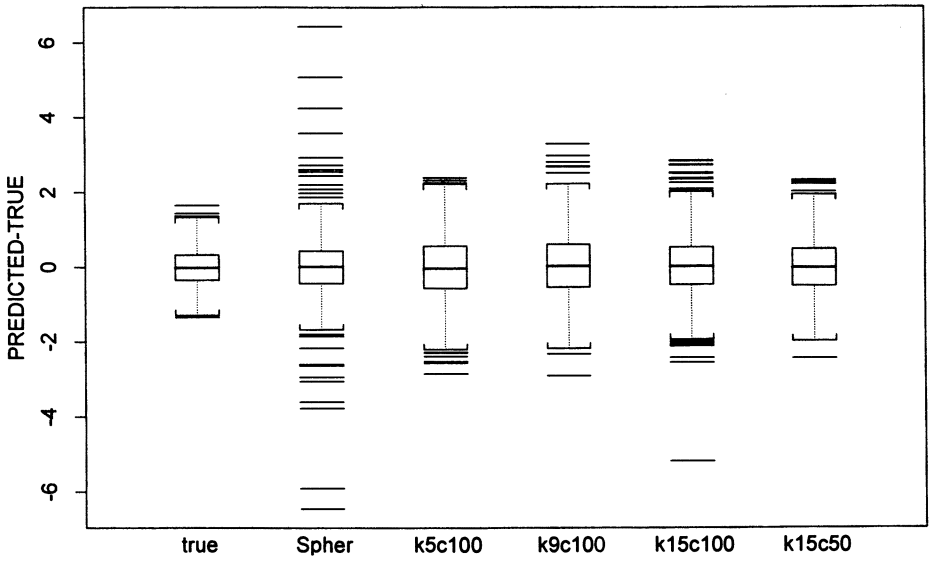

Figure 2. For Small Nugget Effects, Boxplots of the Actual Prediction Errors $(\hat{Z}(51)-Z(51))$ When the True Variogram is Known and Under Five Variogram Estimation Techniques.

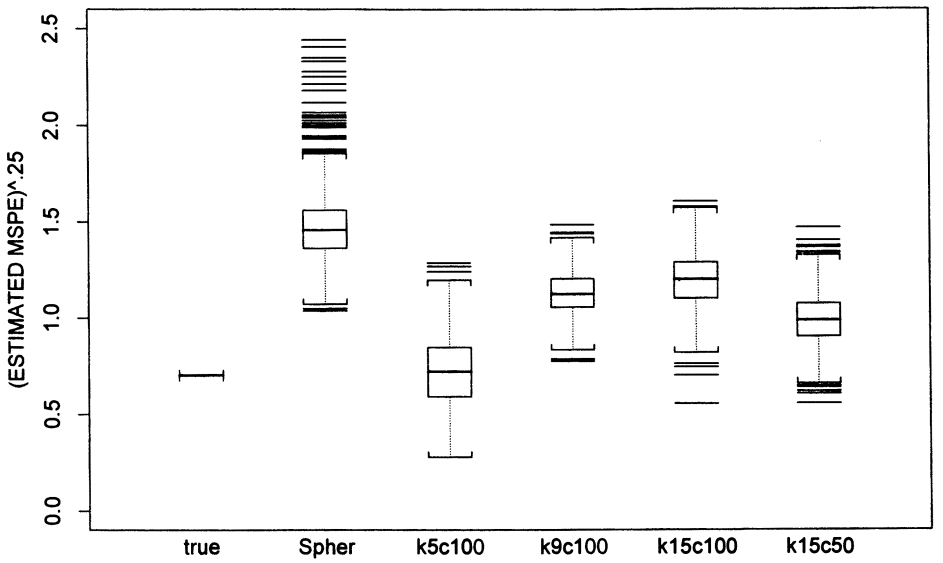

Figure 3. For Small Nugget Effects, Boxplots of the Fourth Root of the Estimated Mean Squared Prediction Error, $\sqrt{\hat{\sigma}(51)}$, When the True Variogram is Known and Under Five Variogram Estimation Methods.

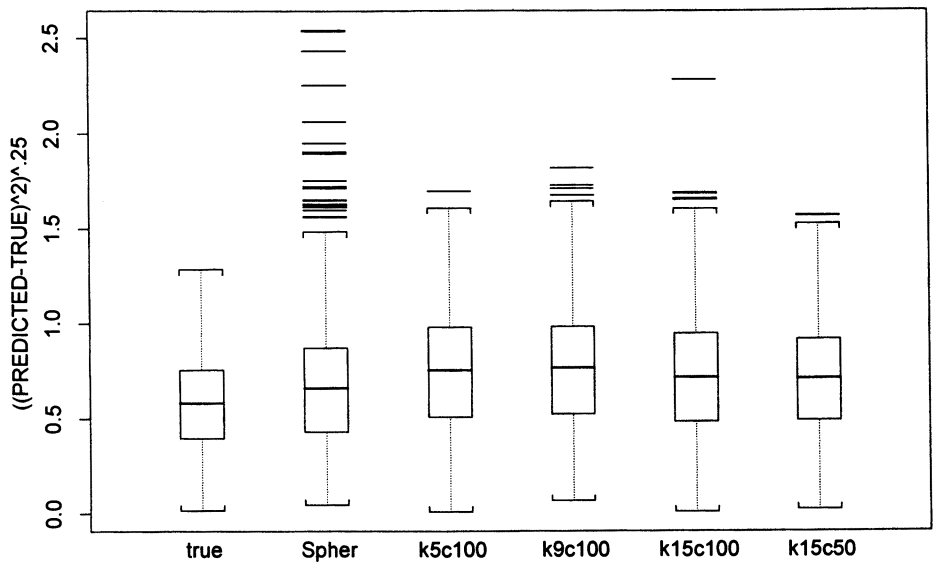

Figure 4. For Small Nugget Effects, Boxplots of the Fourth Root of the True Mean Squared Prediction Error, When the True Variogram is Known and Under Five Variogram Estimation Methods.

Table 3. Results of 1000 Simulations Using the Semivariogram Model $2\gamma(h) = .2+15(1-2\sin(h/2)/h)+25(1-3\sin(h/3)/h)$ for $h \neq 0$, Predicting $Z(91)$

	True	*Spherical*	*k5c100*	*k9c100*	*k15c100*	*k15c50*
Mean$((\hat{Z}(91)-Z(91))^2)$	.244	.881	1.988	.714	.538	.624
Mean$(\hat{\sigma}^2(91))$	.246	5.345	.898	2.536	2.643	1.259
Coverage	.947	.999	.775	1.000	1.000	.977

Figure 5 compares boxplots of the estimation error $\hat{z}(51) - z(51)$, Figure 6, boxplots of the transformed estimated mean squared prediction errors $\sqrt{\hat{\sigma}(51)}$, and Figure 7, boxplots of the fourth root of the true mean squared prediction errors $\left(\hat{Z}(51) - Z(51)\right)^2$. With a large nugget effect, all of the methods perform similarly, as the influence of model misselection is masked by noise.

We recommend the use of a fairly flexible variogram model for kriging in one dimension, with $k = 15$. The 15-part linear model with the range restricted to 50 fit especially well, as it had the most flexibility near the origin. Note especially that the choice of variogram model does have a big effect on the size of the actual mean squared prediction error and the estimate of the mean squared prediction error.

In order to determine whether the results would depend on the location at which we wanted to predict, we also ran the simulation, predicting $Z(91)$ from the values $Z(1)$, $\ldots$, $Z(90)$, $Z(92)$, $\ldots$, $Z(101)$. We only considered the case where the nugget effect was .2. The results were qualitatively the same as when we predicted $Z(51)$, except that

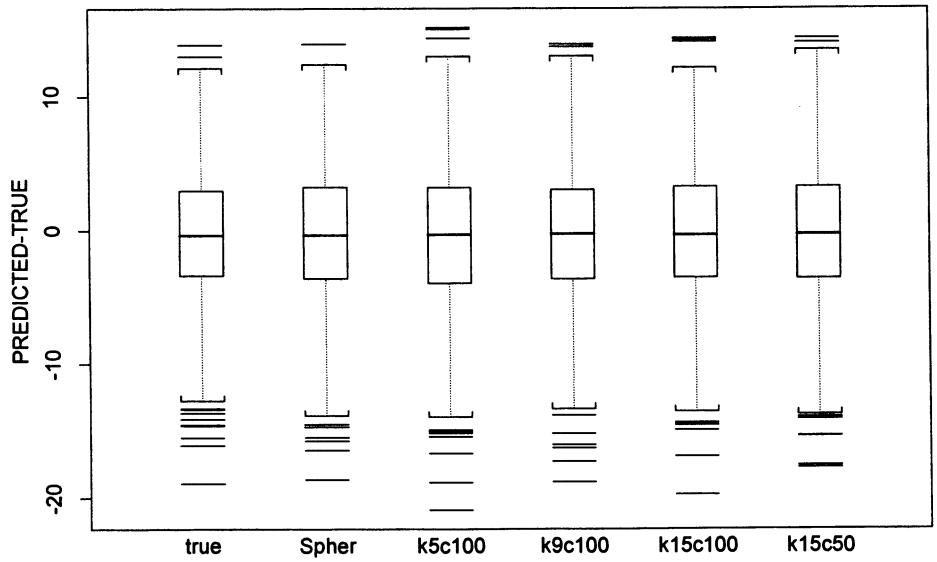

Figure 5. For Large Nugget Effects, Boxplots of the Actual Prediction Errors ($\hat{Z}(51)-Z(51)$) When the True Variogram is Known and Under Five Variogram Estimation Techniques.

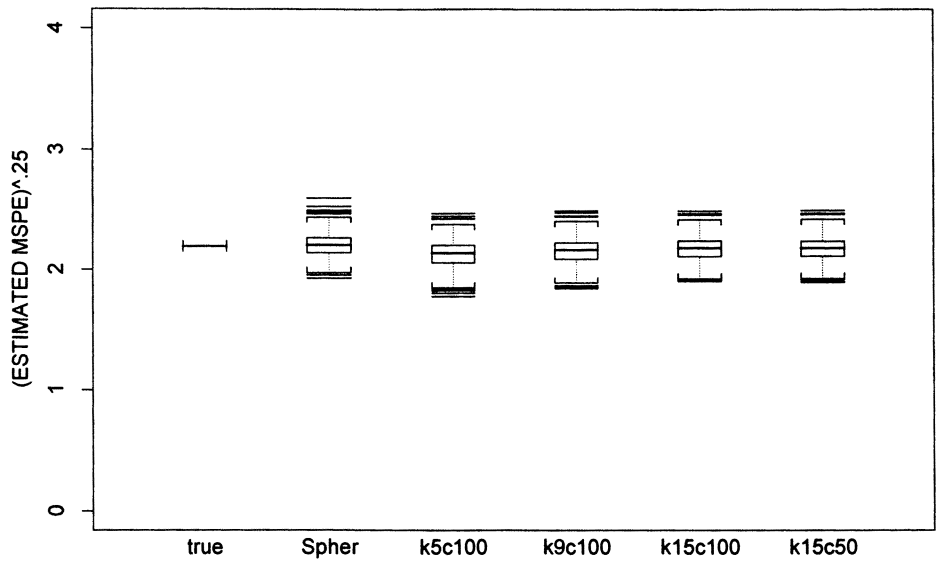

Figure 6. For Small Nugget Effects, Boxplots of the Fourth Root of the Estimated Mean Squared Prediction Error, when the True Variogram is Known, and Under Five Variogram Estimation Methods.

the use of a range of 100 led to marginally better prediction than the use of the range 50. The results are shown in Table 3. Here, k15c100 did best with respect to the true mean squared prediction error, but it still overestimated its variance, $\hat{\sigma}(91)$, whereas k15c50 didn't do quite as well at prediction, but gave better estimates of the mean squared prediction error.

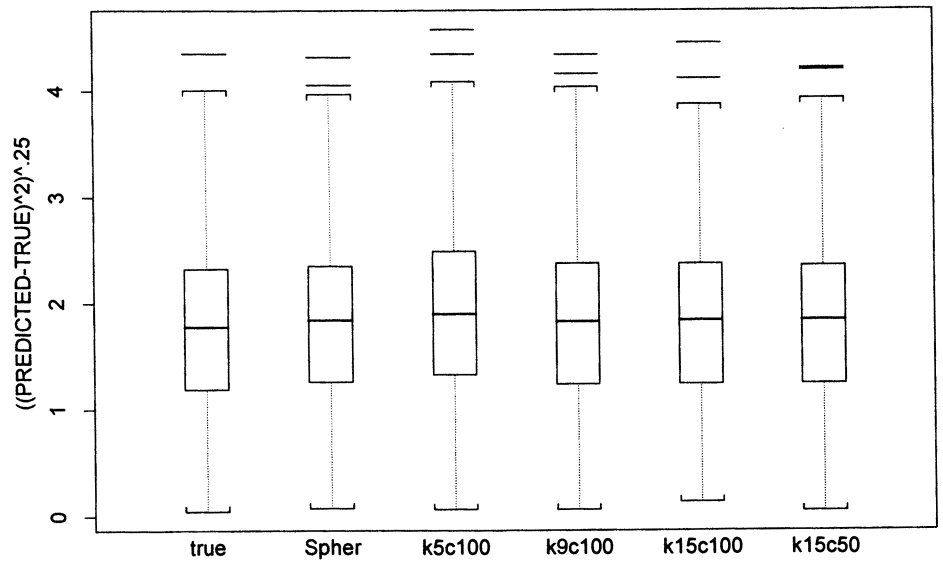

Figure 7. For Small Nugget Effects, Boxplots of the Fourth Root of the True Mean Squared Prediction Error, When the True Variogram is Known and Under Five Variogram Estimation Methods.

7. FITTING TWO-DIMENSIONAL ANISOTROPIC DATA

We fit the Wolfcamp Aquifer data (Cressie 1993, p. 214) with a 20-piece piecewise-planar variogram. The data consist of 85 measurements, each with an x coordinate, y coordinate, and piezometric head $Z(x, y)$. For each of the 3,570 pairs of measurements, we calculated lags $h_1 = |x_i - x_j|$ and $h_2 = |y_i - y_j|$, and the variance estimate $Y_{i,j} = \big(Z(x_i, y_i) - Z(x_j, y_j)\big)^2$. The piecewise-planar variogram described by (4.1) was fit to the 3,570 variance estimates, where the moving average function had a rectangular support of $c = 259$ miles by $d = 176$ miles and was constant over each of 20 equal regions delimited by a five by four grid. The choice of $m = 5$ and $n = 4$ was motivated by the desire to make the fit as flexible as possible, without the convergence problems that would result from fitting too many parameters. As discussed in detail in Section 5, we assumed that the nugget effect was negligible, and minimized the nonweighted sum of squared errors (5.2) using the explicit form of the gradient (5.4) to find the best fitting parameters. The estimated parameters of the piecewise constant moving average function were as in Table 4.

The fitted variogram, obtained by using the parameters in Table 4, is shown in Figure 8. From this fitted variogram, we obtained the matrix Γ and the vector γ in (1.1) and (1.2). These yielded the kriged prediction surface and the prediction standard deviations shown in Figure 9 and 10.

These estimates are in good agreement with the prediction and error surfaces obtained for the Wolfcamp Aquifer data by Cressie using ordinary kriging with the assumption of anisotropy and power model variogram estimators (Cressie 1993, p. 218). This can be seen in Figure 11, which shows the difference between the predictions from the flexible variogram model and predictions from the ordinary kriging done by Cressie. Similarly, Figure 12 shows the differences between the mean squared prediction errors calculated using a flexible variogram model, and those obtained by ordinary kriging. The blackbox approach implicitly allows for anisotropy and does not require the specification of a specific parametric family of variograms from which the variogram estimate is to be selected. Finding a good variogram model in two dimensions is much more difficult than in the case of one dimension. If the empirical variogram appears to be anisotropic, the usual approach is to apply a linear transformation to the data, or to calculate different one-dimensional variograms in two perpendicular directions, assuming geometric anisotropy (e.g., see Cressie 1993, p. 215). This is typically done by eye. The use of a flexible model is attractive in these circumstances.

Table 4. Estimated Parameters of the Piecewise Constant Moving Average Function

EAST →		*NORTH* ↑							
a_{11}	2.819	a_{12}	3.234	a_{13}	3.396	a_{14}	3.234	a_{15}	2.819
a_{21}	3.198	a_{22}	3.626	a_{23}	3.802	a_{24}	3.626	a_{25}	3.198
a_{31}	3.198	a_{32}	3.626	a_{33}	3.802	a_{34}	3.626	a_{35}	3.198
a_{41}	2.819	a_{42}	3.234	a_{43}	3.398	a_{44}	3.234	a_{45}	2.819

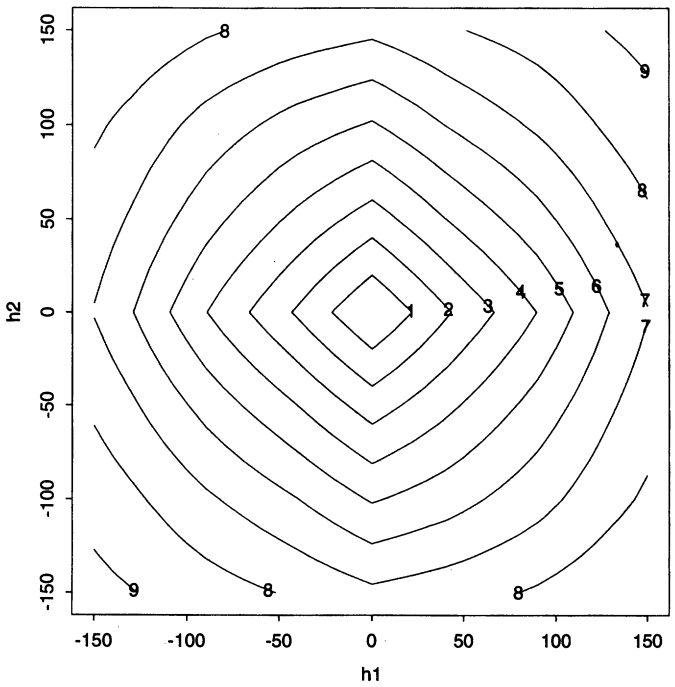

Figure 8. The 20-Part Piecewise-Planar Variogram That Best Fits the Wolfcamp Aquifer Empirical Variogram (in units of $10^5 ft^2$). Here, h1 is the distance in the easterly direction from one location to the other, and h2 is the distance in the northerly direction from one location to the other. Note the anisotropy of the fitted variogram, with the variogram increasing more rapidly in the northeast direction than in the southeast direction.

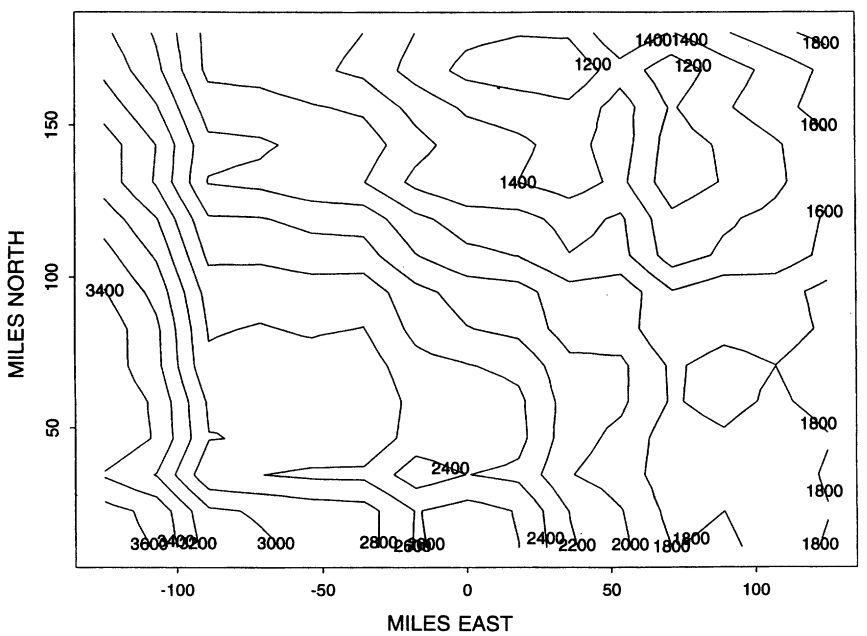

Figure 9. The Kriged Prediction Surface of Piezometric Head for the Wolfcamp Aquifer, Using a 20-Part Piecewise-Planar Variogram.

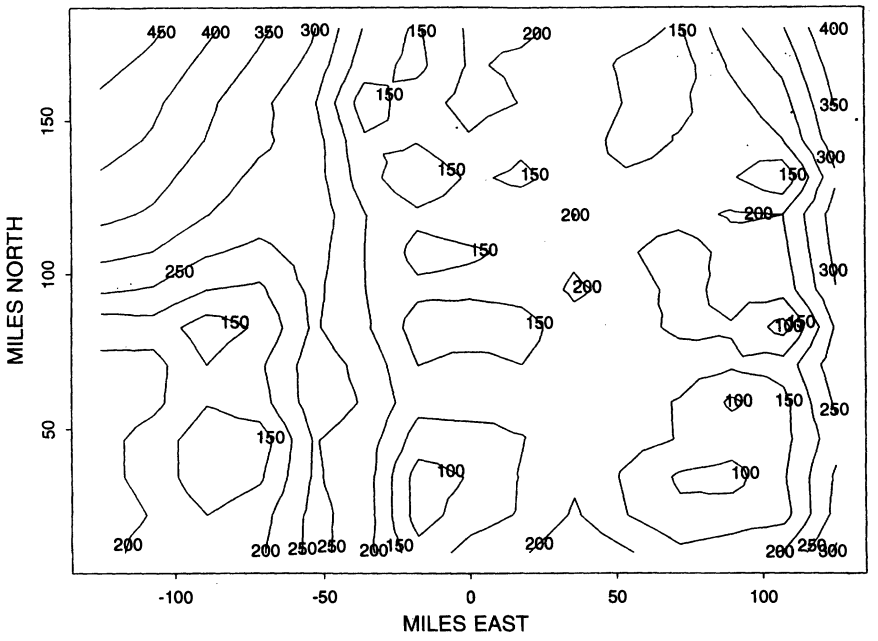

Figure 10. The Kriged Standard Error Surface Corresponding to the Prediction Surface Shown in Figure 3.

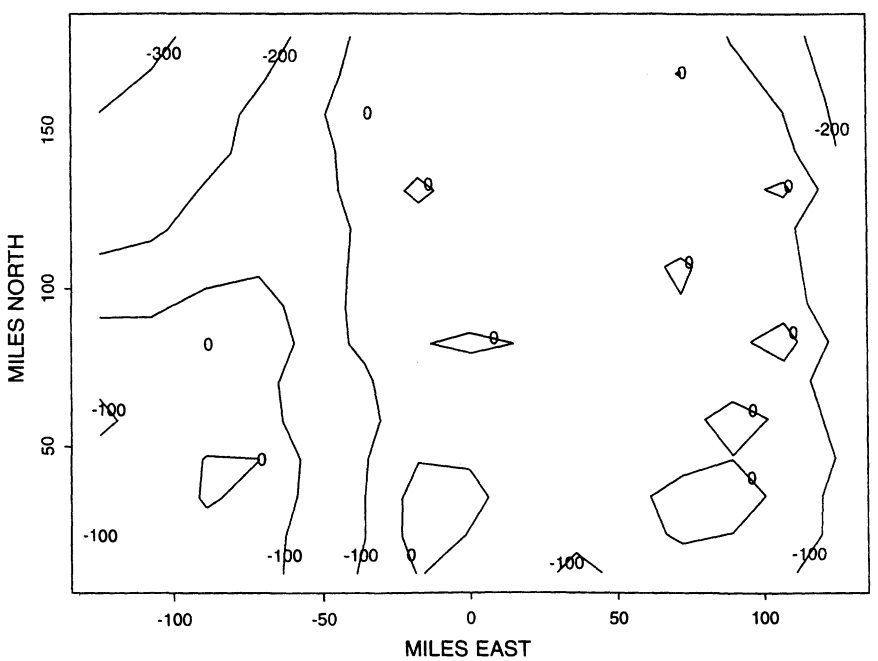

Figure 11. The Difference Between Predicted Values of Piezometric Head Predicted by the Ordinary Kriging Model Using a Power Variogram and the Piezometric Head Predicted Using Flexible Kriging.

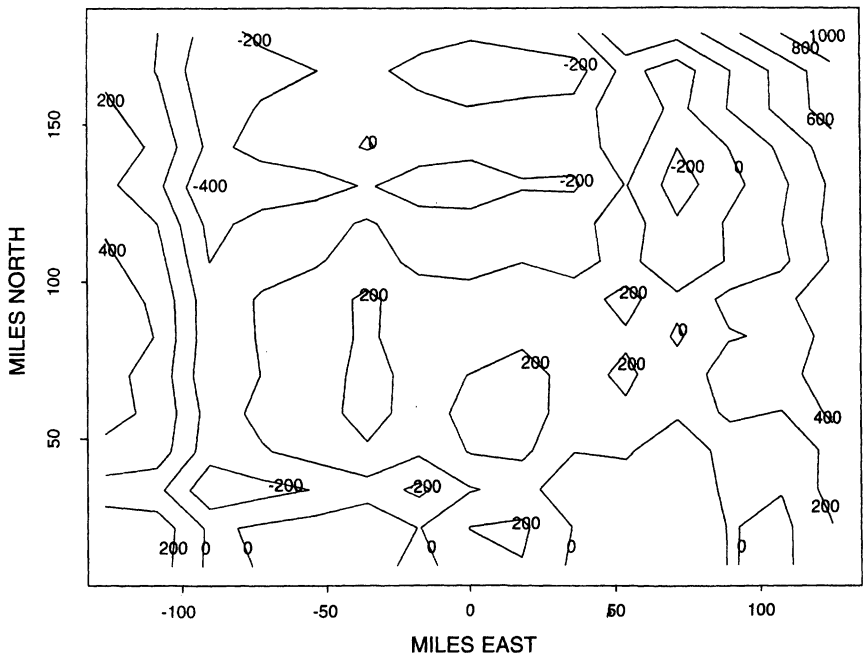

Figure 12. The Difference Between Mean Squared Prediction Errors Given by Ordinary Kriging Using a Power Model and Mean Squared Prediction Error Using Flexible Kriging.

8. PIECEWISE-LINEAR VARIOGRAMS WITH LINEAR PARTS OF UNEQUAL LENGTH

The most informative portion of the variogram is the part nearest the origin (Stein 1988). Most variogram families [including the cosine variograms of Shapiro and Botha (1991)] do not fit the variograms more flexibly near the origin. It is possible to modify the piecewise-linear variogram models so as to obtain higher resolution in regions of interest. To do so we assume that some of the parameters always have the same value. For example, consider the five-part moving average function

$$f(x|a_1, a_2, a_3, a_4, a_5) = \sum_{i=1}^{5} a_i \mathcal{I}\left(c(i-1)/5 < x \leq ci/5\right).$$

This will yield a piecewise-linear variogram that has five equally spaced joints at $ci/5$ for $i = 1, 2, 3, 4, 5$. Now consider

$$f(x|a_1, \ldots, a_{25}) = \sum_{i=1}^{25} a_i \mathcal{I}\left(c(i-1)/25 < x \leq ci/25\right)$$

with constraints $a_5 = a_6 = \cdots = a_{25}$. This piecewise-linear model still has five free parameters, but it now has nine joints at $ci/25$ for $i = 1, 2, 3, 4, 21, 22, 23, 24, 25$. For example, Figure 13 shows the variogram that occurs when $a_1 = 10, a_2 = -10, a_3 = 20, a_4 = -20, a_5 = \cdots = a_{25} = 4$. Notice that the joints are no longer equally spaced, and that this variogram is more flexible near the origin. The same principle extends to the

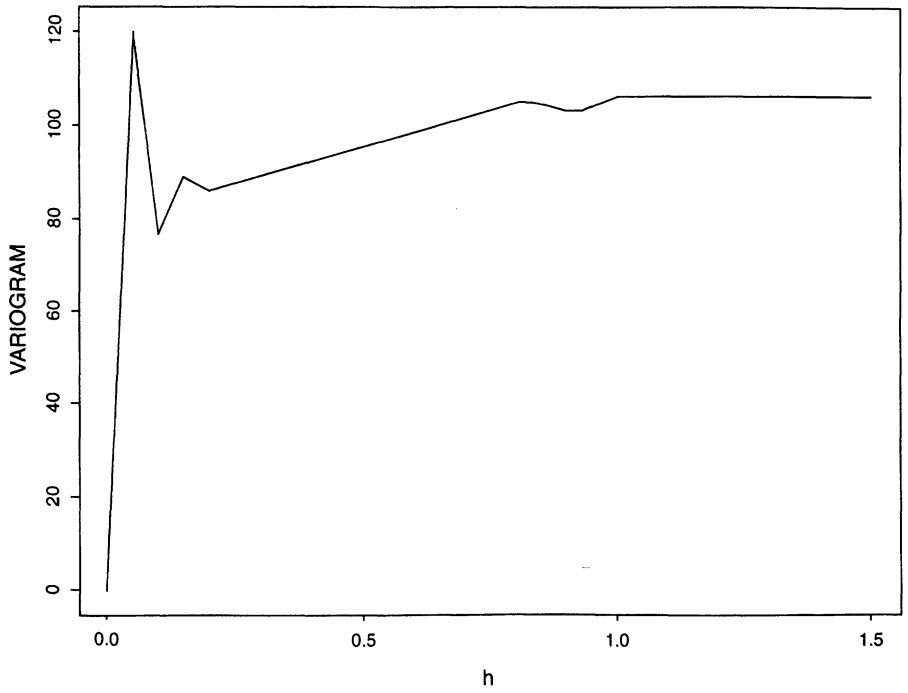

Figure 13. An Example of a Piecewise-Linear Variogram With Pieces of Unequal Length.

two-dimensional case: By forcing some of the many parameters of the piecewise-planar model to be equal, the number of parameters can be greatly reduced while still allowing a good fit to the empirical variogram near the origin.

9. DISCUSSION

Kriging using a flexible variogram model frees the geostatistician from the task of selecting an appropriate family of variograms and hence can allow a "blackbox" approach to spatial prediction. The variogram families introduced in this article and in the article of Shapiro and Botha (1991) have the required flexibility. Using the latter, the empirical variogram is fitted by sums of cosines (in one dimension), and by Bessel and Sinc functions (in two and three dimensions) with enforced convexity, enforced monotonicity, or bounded derivatives used to avoid overfitting. The parameters of these families can be fit using quadratic programming, ensuring that the fit obtained is optimal. Using the variogram families introduced in this paper, the empirical variogram is fitted by piecewise-linear variograms with the linear parts not necessarily of equal length (in one dimension), or by piecewise-planar variograms with the planar parts not necessarily of equal area (in two dimensions).

The simulations indicated that the predictive power of the model increases with increasing flexibility (larger k). This perhaps should be expected, as greater flexibility in the variogram family allows the fitted variogram to approximate a possibly irregular true variogram. The tendency to underestimate the nugget effect needs to be investigated

further, though the underestimation can be prevented by estimating the nugget effect separately from the rest of the parameters. Although it is possible to estimate the nugget effect with the two-dimensional model, we estimated the two-dimensional variogram under the assumption of negligible nugget. An area for further research would be the accurate estimation of nugget effects under the two-dimensional model.

A critical consideration in any analysis involving the minimization of a function of many parameters is the rapidity and stability of the minimization technique. In the one-dimensional simulation, we used the multidimensional simplex method of Nelder and Mead (1965). In the two-dimensional case, we used an explicit computation of the gradient in a "sliding-downhill" algorithm. Both techniques worked well. The multidimensional simplex method, because it does not require derivatives, is very easy to use but slower than methods that use the gradient. As the number of parameters increases, as happens for the two-dimensional example, gradient methods are preferred to the multidimensional simplex method, as the former exploits our ability to explicitly write the gradient of the SSE.

The parameters of these fitted variograms are obtained iteratively, which, unlike quadratic programming, may not always give the best possible fit. On the other hand, the piecewise-planar families can be used in situations where the assumption of isotropy is unsupportable. Piecewise-linear or -planar variogram families can also be chosen that do the best job of fitting the empirical variogram near the origin, which is usually the area of greatest interest to geostatisticians. In any event, these two classes of variogram families should satisfy the needs of geostatisticians who would prefer a blackbox approach to kriging that obviates the need to specify a variogram family, yet should yield estimates close to the true variogram. Because the modeling of the variogram can be eliminated, kriging may be made automatic by substituting the fitted variogram into the kriging Equations (1.1) and (1.2). Of course, the fitted variogram itself is often interesting, and it is still available for examination (e.g., Figures 1 and 8).

APPENDIX A: CONSTRUCTION OF RANDOM PROCESSES FROM MOVING AVERAGE FUNCTIONS

The limiting version of smoothing or filtering a large number of independent, small random variables is integration with respect to a white-noise process.

Define the stochastic process

$$Z(\mathbf{s}|\theta) = \mu + \int_{\mathcal{R}^m} f(\mathbf{x} - \mathbf{s}|\theta)W(\mathbf{x})\mathbf{dx} + X(\mathbf{s}),$$

where $\mathbf{s}, \mathbf{x} \in \mathcal{R}^m$; $f : \mathcal{R}^m \to \mathcal{R}$ is square-integrable; $W(\mathbf{s})$ is a zero-mean unit variance white-noise process [an ideal continuous process with a constant spectral density, and the formal derivative of Brownian motion (see Yaglom 1987, pp. 117–120)] on $\mathcal{R}^m$; $X(\mathbf{s})$ is independent of $X(\mathbf{t})$ for $\mathbf{t} \neq \mathbf{s}$; $X(.)$ is independent of $W(.)$; $EX(\mathbf{s}) = 0$; and $\text{var}(X(\mathbf{s})) = c_0$. The process clearly has stationary mean μ. We will refer to f as a moving average function for the process $Z(\mathbf{s})$. The variogram of this process is easy to compute using the following lemma:

Lemma 1. *Let $g : \mathcal{R}^m \to \mathcal{R}$ such that $\int_{\mathcal{R}^m} g^2(x)dx$ is finite. Then if $W(x)$ is the white-noise process on $\mathcal{R}^m$,*

$$\mathrm{var}\Big(\int_{\mathcal{R}^m} g(x)W(x)dx\Big) = \int_{\mathcal{R}^m} g^2(x)dx$$

(e.g., see Yaglom 1987, p. 46).

Then, for each θ and h,

$$\begin{aligned}
2\gamma(\mathbf{h}|\theta) &= \mathrm{var}\left(Z(\mathbf{s}) - Z(\mathbf{s}-\mathbf{h})\right) \\
&= \mathrm{var}\Big[\mu + \int_{\mathcal{R}^m} f(\mathbf{x}-\mathbf{s}|\theta)W(\mathbf{x})\mathbf{dx} + X(\mathbf{s}) - \mu \\
&\qquad - \int_{\mathcal{R}^m} f(\mathbf{x}-\mathbf{s}-\mathbf{h}|\theta)W(\mathbf{x})\mathbf{dx} - X(\mathbf{s}-\mathbf{h})\Big] \\
&= \mathrm{var}\left[\int_{\mathcal{R}^m} f(\mathbf{x}-\mathbf{s}|\theta)W(\mathbf{x})\mathbf{dx} - \int_{\mathcal{R}^m} f(\mathbf{x}-\mathbf{s}-\mathbf{h}|\theta)W(\mathbf{x})\mathbf{dx}\right] \\
&\qquad + \mathrm{var}\left(X(\mathbf{s}-\mathbf{h}) - X(\mathbf{s})\right) \\
&= \mathrm{var}\left[\int_{\mathcal{R}^m} \left(f(\mathbf{x}-\mathbf{s}|\theta) - f(\mathbf{x}-\mathbf{s}-\mathbf{h}|\theta)\right) W(\mathbf{x})\mathbf{dx}\right] \\
&\qquad + \mathrm{var}\left(X(\mathbf{s}-\mathbf{h}) - X(\mathbf{s})\right) \\
&= \int_{\mathcal{R}^m} \left(f(\mathbf{x}) - f(\mathbf{x}-\mathbf{h})\right)^2 \mathbf{dx} + 2c_0.
\end{aligned}$$

In the case where f is square-integrable, the resulting variogram always has a sill. The more general formulation only requires that Equation (2.1) holds, and not that f be square-integrable. In this case, we must define the process as

$$Z(\mathbf{s}|\theta) = \mu + \int_{\mathcal{R}^m \setminus [-\infty,0]^m} f(\mathbf{x}-\mathbf{s}|\theta)W(\mathbf{x})\mathbf{dx} + X(\mathbf{s}),$$

in order for the stochastic integral to be well-defined. This process can be shown to have the variogram

$$\int_{\mathcal{R}^m} (f(\mathbf{x}) - f(\mathbf{x}-\mathbf{h}))^2 \mathbf{dx} + 2c_0$$

similar to the proof for f square-integrable. Of course, the flexible variograms we are working with always have sills.

APPENDIX B: A VALID VARIOGRAM WITH SILL CAN BE APPROXIMATED BY PIECEWISE-LINEAR VARIOGRAMS

For continuous variograms, the quality of an approximation is arguably best measured in the sup norm, where the distance between variograms f and g is $\sup_{x \in D} |f(x) - g(x)|$. The size of the sup norm difference between a true variogram and an approximating variogram directly bounds the largest possible error in computing a variance using the approximating variogram.

For any variogram γ, weights $a_1, \ldots, a_m$ summing to 0, and locations $s_1, \ldots, s_m$, the following holds:

$$\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j \gamma(s_i - s_j) = -\text{var}\left(\sum_{i=1}^{m} a_i Z(s_i)\right)$$

(Cressie 1993, p. 87).

If we compare the effects of using either of two variograms, γ_{true} or γ_{approx}, we obtain

$$\begin{aligned}
&\left|\text{var}_{\text{true}}\left(\sum_{i=1}^{m} a_i Z(s_i)\right) - \text{var}_{\text{approx}}\left(\sum_{i=1}^{m} a_i Z(s_i)\right)\right| \\
&= \left|\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j \left(\gamma_{\text{true}}(s_i - s_j) - \gamma_{\text{approx}}(s_i - s_j)\right)\right| \\
&\leq \sum_{i=1}^{m}\sum_{j=1}^{m} |a_i a_j| \sup_{i,j} |\gamma_{\text{true}}(s_i - s_j) - \gamma_{\text{approx}}(s_i - s_j)|.
\end{aligned}$$

Thus, a small difference between variogram models under the sup norm ensures little error in variance estimates.

It is reassuring that, under the sup norm, any well-behaved one-dimensional variogram with a sill can be fit arbitrarily well by a piecewise-linear variogram.

Definition 1. *A function $f : \mathcal{R}^m \to \mathcal{R}$ is said to be Lipschitz (p, q), or $f \in Lip(p, q)$, where p is an integer greater than 0, and $q \in [0,1)$, if the pth derivative of f satisfies the Hölder condition*

$$\sup_{|x_1 - x_2| < h} |f^{(p)}(x_1) - f^{(p)}(x_2)| < C(f^{(p)})h^q,$$

where C depends only on $f^{(p)}$, and not on h.

Theorem A.1. *Let $2\gamma(h)$ be a variogram with a range of c and let $\gamma \in Lip(p, q)$, where p is whole number, and $q \in [0,1)$. Then the k-part piecewise-linear function $g(h)$ that interpolates the knots $[\frac{ci}{k}, 2\gamma(\frac{ci}{k})]_{i=-k}^{k}$ is a valid variogram, for any integers c and k.*

Proof: Let

$$t_i = \begin{cases} 2\gamma(\frac{ci}{k}) & \text{if } i \in \mathcal{Z}, |i| \leq k; \\ 2\gamma(c) & \text{if } i \in \mathcal{Z}, |i| > k. \end{cases}$$

Because the sill of the variogram is $2\gamma(c)$, all of these points lie on the graph of the variogram. Thus these points must be conditionally negative definite, and $(t_i - t_k)/2 = \text{cov}(Z(s), Z(s + (k - i)c/k))$ for $|i| < k$. But then the series $\{(t_k - t_i)/2\}_{i=-\infty}^{\infty}$ is positive definite with support on $i = -k, \ldots, k$. A well-known result in time series analysis (Brockwell and Davis 1987, p. 89) tells us that this series is the autocorrelation function of a moving average process, and that there exist parameters $\theta_0, \ldots, \theta_k$ such that

$$(t_k - t_i)/2 = \sum_{j=0}^{k-i} \theta_j \theta_{j+i}.$$

Define the moving average function

$$\phi(x) = \sum_{i=1}^{k} \theta_i \mathcal{I}((ic/k) < x \leq (i+1)c/k).$$

A simple integration shows that the variogram corresponding to this moving average function is piecewise-linear with knots $\{\frac{di}{k}, 2\gamma(\frac{di}{k})\}_{i=-k}^{k}$, and thus is $g(h)$. Therefore $g(h)$ is a valid variogram. □

Corollary 1. *Assuming the preceding hypotheses, there exists $M > 0$ depending only on γ such that*

$$\inf_{g \in G} \sup_{h \in [-c,c]} |2\gamma(h) - g(h)| \leq Mk^{-p-q},$$

where G is the set of piecewise-linear variograms with $2k-1$ equally spaced joints on $(-c, c)$.

Proof: From deBoor (1978), a piecewise-linear, continuous interpolant $g(h)$ to a function $2\gamma(h)$ that is $\text{Lip}(p, q)$ satisfies

$$\sup_h |g(h) - 2\gamma(h)| \leq r(\gamma(h))c^{p+q}k^{-p-q}.$$

Let $M := rc^{p+q}$, and note that the piecewise-linear, continuous interpolant to $2\gamma(h)$ at knots $\{ic/k\}_{i=-k}^{k}$ is also a valid variogram. □

Thus, as $k \to \infty$, the family of k-part piecewise-linear variograms includes variograms that are arbitrarily close, in the sup norm, to any continuous variogram with sill.

ACKNOWLEDGMENTS

Support for this work was provided by Federal Aid in Wildlife Restoration to the Alaska Department of Fish and Game and the Eli-Lander Foundation. We thank an associate editor and two anonymous reviewers for their comments on earlier versions of the manuscript.

[Received January 1995. Revised January 1996.]

REFERENCES

Brockwell, P. J., and Davis, R. A. (1987), *Time Series: Theory and Methods,* New York: Springer-Verlag.

Cressie, Noel A. C. (1993), *Statistics for Spatial Data, Revised Edition,* New York: John Wiley.

De Boor, C. (1978), *A Practical Guide to Splines,* New York: Springer-Verlag, p. 42.

Matern, B. (1986), *Spatial Variation* (2nd ed.), Lecture Notes in Statistics, New York: Springer-Verlag.

Nelder, J. A., and Mead, R. (1965), "A Simplex Method for Function Minimization" *Computer Journal,* 7, 308–313.

Shapiro, A., and Botha, J. D. (1991), "Variogram Fitting With a General Class of Conditionally Nonnegative Definite Functions," *Computational Statistics and Data Analysis,* 11, 87–96.

Stein, M. L. (1988), "Asymptotically Efficient Prediction of a Random Field With a Misspecified Covariance Function," *The Annals of Statistics,* 16, 55–63.

Stein, M. L., and Handcock, M. S. (1989), "Some Asymptotic Properties of Kriging When the Covariance Function is Misspecified," *Mathematical Geology,* 21, 171–190.

Webster, R. (1985), "Quantitative Spatial Analysis of Soil in the Field," in *Advances in Soil Science* (Vol. 3), New York: Springer-Verlag.

Yaglom, A. M. (1987), *Correlation Theory of Stationary and Related Random Functions. Volume I: Basic Results*, New York: Springer-Verlag.