Incorporating Parameter Uncertainty into Prediction Intervals for Spatial Data Modeled via a Parametric Variogram

Author(s): Fujun Wang and Melanie M. Wall

Source: *Journal of Agricultural, Biological, and Environmental Statistics*, Sep., 2003, Vol. 8, No. 3 (Sep., 2003), pp. 296-309

Published by: Springer

Stable URL: https://www.jstor.org/stable/1400549

Incorporating Parameter Uncertainty into Prediction Intervals for Spatial Data Modeled via a Parametric Variogram

Fujun WANG and Melanie M. WALL

In spatial predictions, researchers usually treat the estimated theoretical variogram parameters as known without error and ignore the variability of the parameter estimators. Although the prediction is still unbiased, the prediction error is usually underestimated. Therefore, the coverage probability of the prediction interval usually is lower than the nominal probability. A simulation study is performed to show how the coverage probability for prediction relates to the true range and sill of an exponential variogram. This article proposes two parametric bootstrap methods to incorporate the variability of the corresponding parameter estimators. A simulation study is performed to evaluate the coverage probability of these proposed methods. Finally, we apply the parametric bootstrap methods to a real dataset and compare the results with those from naive (i.e., treating estimated parameters as known) and Bayesian methods.

Key Words: Bootstrap; Coverage probability; Kriging; Mean squared prediction error.

1. INTRODUCTION

From its origin in the mining industry in the early fifties, the field of geostatistics has been a rapidly evolving branch of statistics. The term kriging under the model-based geostatistical framework (Matheron 1971, chap. 1) refers to spatial prediction usually based on the best linear unbiased predictor (BLUP). The practical nature of kriging, that is, making predictions based on one sample of n observations from a continuous spatial process, has made it very popular in many other areas of science and industry, and now it is readily available in many major statistical software packages.

Despite its popularity, the statistical properties of the prediction error associated with kriging are only well understood for the case where the covariance structure of the spatial process is assumed to be known. In practice the true covariance structure of the process is

Fujun Wang is Senior Statistician, Eli Lilly and Company, Indianapolis, IN 46285. Melanie M. Wall is Assistant Professor, Division of Biostatistics, School of Public Health at the University of Minnesota, A460 Mayo Building, MMC 303, 420 Delaware Street SE, Minneapolis, MN, 55455 (E-mail: melanie@biostat.umn.edu).

Journal of Agricultural, Biological, and Environmental Statistics, Volume 8, Number 3, Pages 296–309
DOI: 10.1198/1085711031661

almost never known and instead has to be estimated from the same data. The fact that the covariance structure is estimated means that quantifying the correct mean squared prediction error of the spatial predictor is problematic. Thus, forming prediction intervals with accurate coverage probability becomes a problem.

Work has been done in an attempt to approximate this mean squared prediction error. Zimmerman and Cressie (1992) applied the work of Kackar and Harville (1984) for prediction in linear mixed effects models to obtain an approximation to the variance in the spatial prediction setting, and, under a different set of assumptions, Prasad and Rao (1986) and Harville and Jeske (1992) proposed a slightly different approximation. Nevertheless these approximations are only valid in certain special cases under strong assumptions about the process.

Bayesian methods are considered to "provide a convenient way of incorporating the parameter uncertainty into predictive inferences" (Diggle, Tawn, and Moyeed 1998). There have been substantial advances in the computational methods for Bayesian analysis including specifically for spatial data, for example, the GeoBUGS software (Spiegelhalter, Thomas, and Best 1999), yet the amount of time it takes to obtain results can be appreciably longer than traditional frequentist methods. Thus, in this article—which focuses on frequentist technique—we consider the Bayesian method only in the example rather than in the simulations.

In practice it is common to ignore the fact that the covariance structure has been estimated from the data and instead treat the estimated covariance structure as if it were known without error. Putter and Young (2001), in a general spatial prediction setting, showed that the effect of estimation of the covariance structure on predictions and prediction interval coverage probability is negligible asymptotically. Nevertheless, it is not uncommon to encounter small datasets for spatial processes and in such cases the error in the coverage probability of the prediction interval can be appreciable (as will be demonstrated later in this article).

A widely supported technique for estimating the covariance structure is to first form the empirical variogram and then to fit a parametric second order stationary variogram to it via least squares. Recently, Lahiri, Lee, and Cressie (2002) developed the large sample distributions of the least squares estimators of parameters in spatial variograms. Despite what is known about these estimators, the estimates of the parameters in the variogram function are most often treated as if they have no error when used for kriging, that is, their sampling variability is ignored. Thus, the main goal of this article is to incorporate this variability into the prediction intervals via a parametric bootstrap. The performance of the proposed techniques for creating prediction intervals are evaluated and compared with the naive method that ignores the variability by a simulation study.

2. AN EXAMPLE

Acid rain is an important environmental issue in the United States and formation of accurate prediction intervals for acid rain deposition at various spatial locations is important

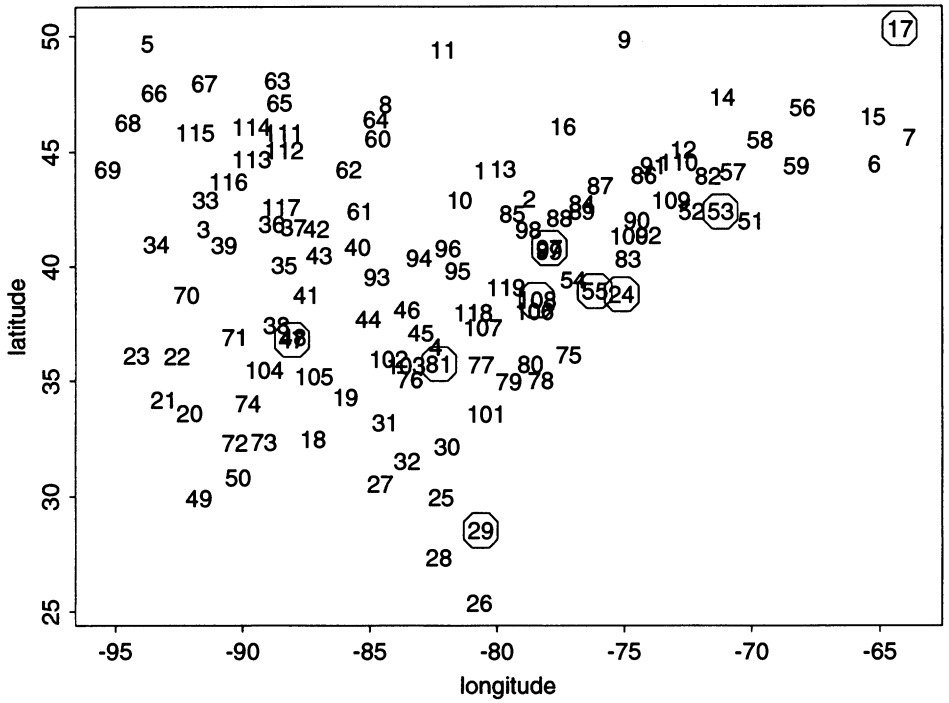

Figure 1. Geographic location of the United States National Atmospheric Deposition Network monitoring sites. The numbers refer to site labels.

for monitoring. Acid rain is evaluated by measuring the amounts of various chemicals deposited annually by precipitation (wet deposition). The major United States wet deposition network is the National Atmospheric Deposition Network (NADP), which began in 1978. The data considered here consist of summarized total ammonium deposition measurements collected by NADP for 1995 at 119 sites in eastern North America (Figure 1) and are available at http://www.biostat.umn.edu/~melanie/Data/.

Preliminary data analysis suggested a north-south trend in the log transformed ammonium deposition values with higher values tending to be in the north. In this analysis, kriging is performed on the detrended data using a variogram model for the spatial structure estimated from the same data. An empirical variogram of the detrended data appears to reasonably follow the shape of an exponential variogram (figure not shown) and has fitted range parameter 1.63, partial sill 0.96, and nugget 0.

These estimated values for the range, partial sill, and nugget are then considered as the true values for the spatial parameters and are plugged into the kriging equation and the associated mean squared prediction error for creating spatial prediction intervals. But when this is done, the sampling variability in the estimated spatial parameters is completely ignored and consequently the prediction intervals tend to be too short. These prediction intervals tend to have less than nominal coverage probability which may lead to problems in monitoring whether ammonium deposition is above or below a given thresholds at different spatial locations. Section 7 revisits this example and compares the usual ("naive") prediction

interval just described with the proposed bootstrap prediction intervals presented in this article.

3. REVIEW OF KRIGING WITH A FITTED PARAMETRIC VARIOGRAM

Kriging is a method of interpolation for a random process $Z(s)$ based on BLUP for the linear model, $Z(\mathbf{s}) = \mu + \epsilon(\mathbf{s})$, where s varies continuously throughout a region D, $E(\epsilon(\mathbf{s})) = 0$, and $\text{cov}(\epsilon(\mathbf{s}+\mathbf{h}), \epsilon(\mathbf{s})) = C(\mathbf{h})$. The function $C(\mathbf{h})$ is called the covariogram, where $\mathbf{h}$ represents the vector that separates two spatially located error terms.

The predictor $\hat{Z}(s_0)$ is the BLUP if $C(\mathbf{h})$ is known. Let $C(\mathbf{h}) \equiv C(\mathbf{h}; \boldsymbol{\alpha})$ denote a parametric covariogram completely specified by the parameters in vector $\boldsymbol{\alpha}$. Also let the notation $p(\mathbf{Z}; s_0; \boldsymbol{\alpha}) \equiv \hat{Z}(s_0)$ represent the BLUP of $Z(s_0)$ when $C(\mathbf{h}) \equiv C(\mathbf{h}; \boldsymbol{\alpha})$ and $m_1(\boldsymbol{\alpha}) \equiv \text{MSPE}(p(\mathbf{Z}; s_0; \boldsymbol{\alpha}))$ represent the mean squared prediction error (MSPE).

The variogram may also be used to describe the covariance structure and is more general than the covariogram since it does not require a second order stationary process assumption. It is defined as $2\gamma(\mathbf{h}) = \text{var}(Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s}))$ with the assumption $E(Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})) = 0$. It is common to assume the variogram has a parametric form, that is, $\gamma(\mathbf{h}) \equiv \gamma(\mathbf{h}; \boldsymbol{\alpha})$. Estimation of $\boldsymbol{\alpha}$ is then performed by minimizing a certain distance between the nonparametric empirical variogram estimator and the parametric form $\gamma(\mathbf{h}; \boldsymbol{\alpha})$. Lahiri, Lee, and Cressie (2002) developed the large sample distributions of $\hat{\boldsymbol{\alpha}}$ using various least squares approaches (ordinary least squares (OLS), weighted least squares (WLS), and generalized least squares (GLS)) under a pure increasing domain asymptotic framework and what they called a "mixed-increasing-domain" framework which still requires the spatial region to become unbounded but allows for infilling as well. For finite samples, several authors, notably Zimmerman and Zimmerman (1991) and Genton (1998), have used simulation to consider the bias and sampling variability for $\hat{\boldsymbol{\alpha}}$ obtained by using various nonparametric empirical variogram estimators followed by various minimization criterion. It is this sampling variability in $\hat{\boldsymbol{\alpha}}$ which is being ignored when forming a prediction interval using $m_1(\hat{\boldsymbol{\alpha}})$. The consequence of ignoring the parameter uncertainty is that the coverage probability will be less than nominal as shown in the next section.

4. COVERAGE PROBABILITY OF THE NAIVE METHOD

Because $\boldsymbol{\alpha}$ has to be estimated in practice, the naive MSPE $m_1(\hat{\boldsymbol{\alpha}})$, that is, treating $\hat{\boldsymbol{\alpha}}$ as if it were known without error, is usually smaller than the true MSPE of $p(\mathbf{Z}; s_0; \hat{\boldsymbol{\alpha}})$ (denoted here as $m_2(\boldsymbol{\alpha})$):

$$m_2(\boldsymbol{\alpha}) \equiv E(Z(s_0) - p(\mathbf{Z}; s_0; \hat{\boldsymbol{\alpha}}))^2, \tag{4.1}$$

where $Z(s_0)$ is the unobserved random variable to be predicted. Thus, the $100(1-c)\%$ prediction interval $p(\mathbf{Z}; s_0; \hat{\boldsymbol{\alpha}}) \pm z_{c/2}\sqrt{m_1(\hat{\alpha})}$ will have less than nominal coverage in general.

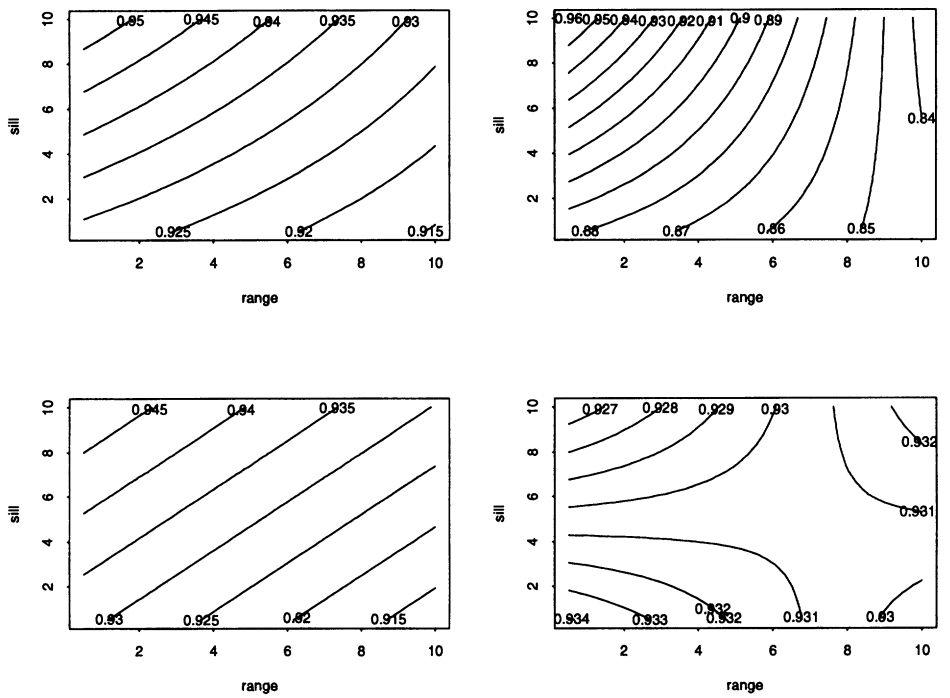

Figure 2. Contour plots of the coverage probability at four different sites using the naive method in which the estimated parameter vector is treated as known without error. Upper left: (3.5, 3.5), upper right: (3.9, 3.9), lower left: (5.5,1.5), lower right: (3.5, 7). Note the contour plots represent a smoothed surface through the simulated coverage probabilities for 400 different range and sill values.

Here we consider a simulation study to examine the performance of the naive method in terms of coverage probability. On a 6×6 grid, we sampled 5,000 spatial datasets of size 36 from a normal random field with an exponential semi-variogram with true parameter $\boldsymbol{\alpha} = (a, b, c)$, that is,

$$\gamma(\| \mathbf{h} \|) = \begin{cases} 0, & \| \mathbf{h} \| = 0 \\ a + b\{1 - \exp(- \| \mathbf{h} \| / c)\}, & \| \mathbf{h} \| > 0 \end{cases}, \tag{4.2}$$

where a is called the nugget effect, b the partial sill, and c the range. For each dataset we sampled four additional points at sites with coordinates (3.5, 3.5), (3.9, 3.9), (5.5, 1.5), and (3.5, 7). Note, the bottom left corner of the 6×6 grid was anchored at (0,0). We deleted the four additional points from each dataset for prediction. For each dataset, $\hat{\boldsymbol{\alpha}}$ was obtained using OLS to fit the exponential variogram to the classical empirical variogram (Matheron 1962). We chose this method because it is simple and the main goal of this article is to introduce a parametric bootstrap method which can be used to improve prediction intervals created using any estimation method for $\boldsymbol{\alpha}$. Then, for each dataset, we formed 95% prediction intervals at the four deleted sites using the naive method (i.e., $p(\mathbf{Z}; s_0; \hat{\boldsymbol{\alpha}}) \pm 1.96\sqrt{m_1(\hat{\boldsymbol{\alpha}})}$). We use the percentage of the covered cases as the estimated coverage probability. This procedure was repeated for 400 different combinations of true parameters to see how the performance of the naive method changes with the true variogram parameters.

In this simulation, we considered values for the true partial sill and range (i.e., b and c in (4.2)) from 1 to 10 with the nugget (i.e., a in (4.2)) fixed at 0.5.

From the contour plots, the coverage probability at each of the four points appears smaller than the nominal 95% coverage for most range and sill combinations (Figure 2). The different pattern observed in each plot shows that the coverage not only depends on what the true range and sill are, but also where the prediction point is located. Furthermore, the coverage probability is not only related to the point's position in the grid, but also related to the distance to its neighbors. Points (3.9, 3.9) and (3.5, 3.5) are both in the center area, but the former point has much lower coverage probability due to its proximity to an observed point. For the first three locations, coverage becomes worse as the true range of process increases. We repeated this simulation over a 12×12 grid and found the coverage probability was still smaller than nominal for most parameter combinations, but the overall coverage probability at each point appears better than that of 6×6 grid as we expect because of the larger sample size.

5. INCORPORATING THE VARIABILITY OF $\hat{\alpha}$ INTO PREDICTION INTERVALS

5.1 ANALYTICAL ADJUSTMENTS TO THE NAIVE MSPE

An approximation of $m_2(\boldsymbol{\alpha})$ using Taylor series expansion was proposed for use in mixed linear models with estimated variance components by Kackar and Harville (1984), for use in the general linear model with estimated covariance parameters by Harville and Jeske (1992), and again for use in spatial linear model with estimated covariance parameters by Zimmerman and Cressie (1992). The estimator given by Kackar and Harville (1984) is

$$m^*(\hat{\boldsymbol{\alpha}}) = m_1(\hat{\boldsymbol{\alpha}}) + tr[A(\hat{\boldsymbol{\alpha}})B(\hat{\boldsymbol{\alpha}})], \tag{5.3}$$

where $A(\boldsymbol{\alpha})$ is the covariance matrix of the partial derivatives of $p(\mathbf{Z}; s_0, \boldsymbol{\alpha})$ with respect to $\boldsymbol{\alpha}$ and $B(\hat{\boldsymbol{\alpha}})$ approximates $E[(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})']$. To adjust for the possible underestimation of $m^*(\hat{\boldsymbol{\alpha}})$ to $m_2(\boldsymbol{\alpha})$, Prasad and Rao (1986) and Harville and Jeske (1992) proposed the estimator

$$m^{**}(\hat{\boldsymbol{\alpha}}) = m_1(\hat{\boldsymbol{\alpha}}) + 2\text{tr}[A(\hat{\boldsymbol{\alpha}})B(\hat{\boldsymbol{\alpha}})]. \tag{5.4}$$

This is approximately unbiased for $m_2(\boldsymbol{\alpha})$ under several conditions, including that the covariance function $\mathbf{C}(\mathbf{h}; \boldsymbol{\alpha})$ is a linear function of the elements of $\boldsymbol{\alpha}$. This condition is not satisfied for most frequently used covariance functions for spatial data such as exponential, spherical, and Gaussian covariance functions.

The matrix $B(\boldsymbol{\alpha})$ can be obtained from the information matrix if we use maximum likelihood (ML) or restricted ML (REML). Analytically obtaining the partial derivatives of $p(\mathbf{Z}; s_0, \boldsymbol{\alpha})$, $A(\boldsymbol{\alpha})$, is not a trivial task. First of all, these partial derivatives must exist. Even

if they do, the prediction is a complicated function of the covariance matrix which itself is a complicated function of $\boldsymbol{\alpha}$. Hence, to get the derivative with respect to $\boldsymbol{\alpha}$ is not easy, and it is still necessary to calculate the covariance matrix of the derivative. Therefore, despite the simple form of $m^*(\hat{\boldsymbol{\alpha}})$, it is difficult to calculate in practice.

5.2 Parametric Bootstrap Prediction Error Estimate

To incorporate the variability of $\hat{\boldsymbol{\alpha}}$ into the prediction interval, we propose two methods using the parametric bootstrap to estimate the MSPE, that is, $m_2(\boldsymbol{\alpha})$. The first method uses the bootstrap to estimate the variability due to $\hat{\boldsymbol{\alpha}}$, and uses the bootstrap estimate to adjust the naive estimate of MSPE, and the second method uses the parametric bootstrap directly to estimate the MSPE.

5.2.1 Parametric Bootstrap Estimate of Variability in MSPE Due to $\hat{\boldsymbol{\alpha}}$

In (5.3), $\text{tr}[A(\hat{\boldsymbol{\alpha}})B(\hat{\boldsymbol{\alpha}})]$ is a two-step approximation to $E(p(\mathbf{Z}; s_0; \boldsymbol{\alpha}) - p(\mathbf{Z}; s_0; \hat{\boldsymbol{\alpha}}))^2$, the variability due to $\hat{\boldsymbol{\alpha}}$. We propose bootstrapping $E(p(\mathbf{Z}; s_0; \boldsymbol{\alpha}) - p(\mathbf{Z}; s_0; \hat{\boldsymbol{\alpha}}))^2$ directly rather than approximating it with $\text{tr}[A(\hat{\boldsymbol{\alpha}})B(\hat{\boldsymbol{\alpha}})]$.

Let $\hat{\mu}$, $\hat{\boldsymbol{\alpha}}$, and $p(\mathbf{Z}; s_0; \hat{\boldsymbol{\alpha}})$ be the parameter estimates and prediction at s_0 from the observed data $\mathbf{Z} \sim F(\mu, \boldsymbol{\alpha})$. To form our parametric bootstrap estimate of variability due to $\hat{\boldsymbol{\alpha}}$, we generate B bootstrap samples $\mathbf{Z}_1^*, \ldots, \mathbf{Z}_B^*$ from $F(\hat{\mu}, \hat{\boldsymbol{\alpha}})$. For each $\mathbf{Z}_i^*$, we calculate a new parameter estimate $\hat{\boldsymbol{\alpha}}_i^*$ and a new prediction $p(\mathbf{Z}; s_0; \hat{\boldsymbol{\alpha}}_i^*)$ at s_0. Note that the new prediction is based on the new parameter estimate $\hat{\boldsymbol{\alpha}}_i^*$ and the original data $\mathbf{Z}$ instead of $\mathbf{Z}_i^*$ in order to capture the variability only due to the $\hat{\boldsymbol{\alpha}}$. Furthermore, the estimator for μ used in $p(\mathbf{Z}; s_0; \hat{\boldsymbol{\alpha}}_i^*)$ is based on $\hat{\boldsymbol{\alpha}}^*$ and $\mathbf{Z}$, not $\mathbf{Z}_i^*$. The mean squared error due solely to variability of $\hat{\boldsymbol{\alpha}}$ is then approximated in the following way:

$$E[p(\mathbf{Z}; s_0; \hat{\boldsymbol{\alpha}}) - p(\mathbf{Z}; s_0; \boldsymbol{\alpha})]^2 \approx \frac{1}{B-1} \sum_{i=1}^{B} [p(\mathbf{Z}; s_0; \hat{\boldsymbol{\alpha}}_i^*) - p(\mathbf{Z}; s_0; \hat{\boldsymbol{\alpha}})]^2 \equiv s^2. \quad (5.5)$$

Therefore our proposed parametric bootstrap adjusted MSPE of $p(\mathbf{Z}; s_0; \hat{\boldsymbol{\alpha}})$ is given by

$$m_B^*(\hat{\boldsymbol{\alpha}}) \quad = \quad m_1(\hat{\boldsymbol{\alpha}}) + s^2. \quad (5.6)$$

Even though $m_B^*(\hat{\boldsymbol{\alpha}})$ is an improvement over the naive MSPE $m_1(\hat{\boldsymbol{\alpha}})$, since $m_1(\hat{\boldsymbol{\alpha}})$ usually underestimates $m_1(\boldsymbol{\alpha})$, the adjusted MSPE $m_B^*(\hat{\boldsymbol{\alpha}})$ could still underestimate the true MSPE $m_2(\alpha)$. In light of (5.4), we propose a similar parametric bootstrap adjusted MSPE of $p(\mathbf{Z}; s_0; \hat{\boldsymbol{\alpha}})$ as

$$m_B^{**}(\hat{\boldsymbol{\alpha}}) \quad = \quad m_1(\hat{\boldsymbol{\alpha}}) + 2s^2. \quad (5.7)$$

5.2.2 Direct Parametric Bootstrap Estimate of MSPE

In the previous bootstrap adjustment method, we bootstrap only the variability due to $\hat{\boldsymbol{\alpha}}$. Now we consider bootstrapping the whole MSPE of $p(\mathbf{Z}; s_0, \hat{\boldsymbol{\alpha}})$ or $m_2(\boldsymbol{\alpha})$ directly,

without using $m_1(\hat{\boldsymbol{\alpha}})$. This eliminates the concerns that $m_1(\hat{\boldsymbol{\alpha}})$ usually underestimates $m_1(\boldsymbol{\alpha})$. Using similar notation in Section 5.2.1, the direct bootstrap estimate of MSPE can be given as

$$E[p(\mathbf{Z}; s_0, \hat{\boldsymbol{\alpha}}) - Z(s_0)]^2 \approx m_B^d = \frac{1}{B-1}\sum_{i=1}^{B}[p(\mathbf{Z}_i^*; s_0, \hat{\boldsymbol{\alpha}}_i^*) - Z_i^*(s_0)]^2, \tag{5.8}$$

where $Z_i^*(s_0)$ is the ith bootstrap sample from $F(\hat{\boldsymbol{\alpha}})$ at unobserved location s_0 and $p(\mathbf{Z}_i^*; s_0, \hat{\boldsymbol{\alpha}}_i^*)$ is the usual prediction at s_0 based on bootstrap sample $\mathbf{Z}_i^*$ and $\hat{\boldsymbol{\alpha}}_i^*$.

In this method, we not only generate the parametric bootstrap sample $\mathbf{Z}_i^*$ at observed locations, but also generate $Z_i^*(s_0)$ at the sites for prediction. This leads to the advantage as well as the disadvantage of this method. The advantage is that we can generate instead of predict the "true" values, making the calculation of MSPE more accurate and so improving the coverage probability of the corresponding prediction intervals. The disadvantage is that the performance of this method depends on the assumed distribution $F(\mu, \boldsymbol{\alpha})$. That is, if we assumed the wrong distribution, we will lose the ability of generating "true" values at predicted sites and the calculation of MSPE is not correct, therefore, the corresponding prediction interval may not necessarily have proper coverage probability.

We evaluate the performances of the parametric bootstrap adjustment methods given in (5.6) and (5.7), and the direct parametric bootstrap method given in (5.8) by simulation study in the next section.

6. SIMULATION STUDY

To assess the coverage probability of the parametric bootstrap prediction intervals proposed in the previous section, we considered data generated on four different grids (5×5, 6×6, 9×9, and 12×12). The way we expand the grid is to fix the lower left corner of the grid at the (0,0) origin and increase the size of the grid only in first quadrant while keeping the internode distance fixed, that is, we increase the sampling domain instead of the sampling density. Additionally we generated four data points at the four locations with the same coordinates as used in Section 4. The four points were set aside as the "true" values to assess the different prediction intervals. The coordinates of the four data points do not change with the grid size. For each "observed" dataset of size n, we obtained the naive prediction interval and bootstrap prediction intervals as described in the previous section at the four additional locations. We then counted how many times the true values fell in the prediction intervals, and used the ratio as the coverage probability of the prediction intervals.

Recall that in Section 4 we considered the coverage probability of prediction intervals using the naive method for 200 different combinations of range and sill based on the 6×6 grid. Here on all the four different sized grids, we considered normal data with two sets of true parameters, $\boldsymbol{\alpha}_1 = (0.5, 1, 1)$ and $\boldsymbol{\alpha}_2 = (0.5, 1, 8)$ for the exponential variogram. From the contour plot in Figure 2, we know these parameters will give us both good and bad coverage probabilities at different points, so that we can see how well the bootstrap

methods perform when the performance of the naive method varies. Choosing the range parameter as either 1 or 8 is similar to simulations in Zimmerman and Zimmerman (1991) and Genton (1998) where only the range parameter varied. Because the weights of ordinary kriging are invariant by linear transformation of the variogram (Genton 1998, lemma 4) the range parameter is considered most influential for prediction.

For this simulation, we generated 1,000 "observed" datasets for each grid from a normal process with the given true parameters α_1 or α_2. For each "observed" dataset $\mathbf{Z}$, first we

Table 1. Simulated Coverage Probability for the 95% Prediction Intervals Based on Increasing Grid Size from 5 × 5 to 12 × 12. Exponential variogram parameters α_1 or α_2 (nugget, partial sill and range) are used for generating data. Location represents the four prediction sites. The mean squared prediction error (MSPE) represents the different methods used for creating the prediction intervals: the naive method $m_1(\hat{\alpha})$, the bootstrap adjustment method $m_B^*(\hat{\alpha})$, the bootstrap adjustment method $m_B^{**}(\hat{\alpha})$, and the direct bootstrap adjustment method $m_B^d(\hat{\alpha})$. The numbers in parenthesis are the average length of prediction intervals.

		$\alpha_1 = (0.5,1,1)$				$\alpha_2 = (0.5,1,8)$			
Location	*MSPE*	*5 × 5*	*6 × 6*	*9 × 9*	*12 × 12*	*5 × 5*	*6 × 6*	*9 × 9*	*12 × 12*
(3.5,3.5)	$m_1(\hat{\alpha})$	.911	.933	.942	.945	.898	.926	.935	.943
		(3.94)	(4.03)	(4.14)	(4.17)	(2.92)	(2.96)	(2.99)	(3.03)
	$m_B^*(\hat{\alpha})$	.923	.942	.950	.950	.909	.933	.939	.948
		(4.07)	(4.11)	(4.28)	(4.30)	(3.00)	(3.03)	(3.05)	(3.07)
	$m_B^{**}(\hat{\alpha})$	.927	.947	.958	.954	.917	.937	.945	.951
		(4.18)	(4.18)	(4.40)	(4.41)	(3.07)	(3.10)	(3.11)	(3.12)
	$m_B^d(\hat{\alpha})$	.922	.941	.945	.951	.906	.932	.938	.946
		(4.07)	(4.12)	(4.24)	(4.28)	(2.99)	(3.03)	(3.04)	(3.06)
(3.9,3.9)	$m_1(\hat{\alpha})$	.900	.899	.907	.910	.854	.859	.909	.930
		(3.52)	(3.55)	(3.70)	(3.71)	(2.69)	(2.70)	(2.83)	(2.93)
	$m_B^*(\hat{\alpha})$	.929	.921	.934	.936	.885	.884	.927	.938
		(3.88)	(3.78)	(4.06)	(4.04)	(2.93)	(2.92)	(2.99)	(3.03)
	$m_B^{**}(\hat{\alpha})$	.941	.939	.949	.944	.904	.908	.937	.944
		(4.19)	(3.97)	(4.35)	(4.30)	(3.13)	(3.09)	(3.14)	(3.13)
	$m_B^d(\hat{\alpha})$	.928	.928	.945	.941	.890	.890	.929	.939
		(3.91)	(3.91)	(4.03)	(4.05)	(2.91)	(2.93)	(2.98)	(3.03)
(5.5,1.5)	$m_1(\hat{\alpha})$	.915	.939	.945	.948	.909	.913	.929	.937
		(4.23)	(4.05)	(4.15)	(4.18)	(3.11)	(2.99)	(3.00)	(3.04)
	$m_B^*(\hat{\alpha})$	.926	.944	.950	.952	.928	.932	.938	.943
		(4.14)	(4.14)	(4.28)	(4.31)	(3.25)	(3.08)	(3.07)	(3.09)
	$m_B^{**}(\hat{\alpha})$	.937	.947	.953	.957	.938	.942	.942	.948
		(4.58)	(4.23)	(4.39)	(4.42)	(3.38)	(3.17)	(3.13)	(3.14)
	$m_B^d(\hat{\alpha})$	.922	.941	.949	.950	.923	.924	.938	.939
		(4.42)	(4.16)	(4.26)	(4.29)	(3.23)	(3.05)	(3.05)	(3.08)
(3.5,7)	$m_1(\hat{\alpha})$	.915	.928	.941	.940	.897	.931	.937	.933
		(4.98)	(4.54)	(4.06)	(4.09)	(3.64)	(3.36)	(2.96)	(3.01)
	$m_B^*(\hat{\alpha})$	.923	.933	.952	.950	.912	.939	.946	.936
		(5.13)	(4.62)	(4.22)	(4.25)	(3.77)	(3.46)	(3.03)	(3.06)
	$m_B^{**}(\hat{\alpha})$	.930	.937	.963	.954	.919	.946	.950	.940
		(5.28)	(4.69)	(4.36)	(4.38)	(3.90)	(3.56)	(3.11)	(3.11)
	$m_B^d(\hat{\alpha})$	.922	.933	.954	.946	.908	.936	.945	.938
		(5.17)	(4.64)	(4.20)	(4.23)	(3.76)	(3.43)	(3.03)	(3.05)

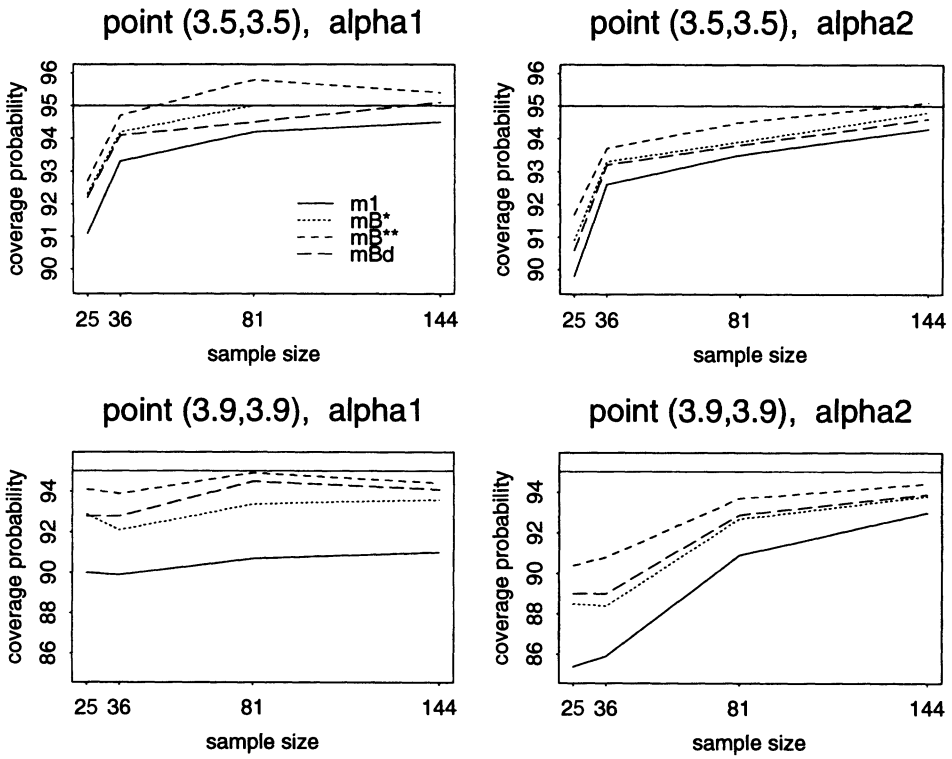

Figure 3. Simulated 95% coverage probability for prediction intervals at the points (3.5,3.5) and (3.9,3.9) by increasing regular grid size. The legends m_1, m_B^*, m_B^{**}, and m_B^d correspond to the naive method, the bootstrap adjustment method with s^2, the bootstrap adjustment method with $2s^2$, and the direct bootstrap method, respectively.

computed the moment-based empirical variogram (Matheron 1962). For the 5×5 grid, the number of lags used is 4, for 6×6, 9×9, and 12×12, the numbers are 7, 13, and 17, respectively. We then fit an exponential variogram using OLS to obtain the estimates of the exponential variogram parameters $\hat{\boldsymbol{\alpha}}$. The GLS estimator treating $\hat{\boldsymbol{\alpha}}$ as known is used to obtain $\hat{\mu}$, and then these are both used to generate the $B = 1{,}000$ bootstrap samples $\mathbf{Z}^*$ from the normal distribution $N(\hat{\mu}\mathbf{1}_n, C(\hat{\boldsymbol{\alpha}}))$. Based on the B bootstrap samples, we computed the prediction intervals for the 4 prediction points using the methods described in previous section.

In all cases (Table 1) the naive method has lower coverage probability than the other methods. The proposed bootstrap adjustment method using $m_B^*(\hat{\boldsymbol{\alpha}})$ shows some improvement, but cannot make the coverage probability as close to nominal as we expect when the naive method performs poorly. The adjustment method using $m_B^{**}(\hat{\boldsymbol{\alpha}})$ performs better in most cases because the underestimation of $m_1(\hat{\boldsymbol{\alpha}})$ to $m_1(\boldsymbol{\alpha})$ is partially balanced by the additional adjustment. We also see that the improvement from the adjustment methods is bigger when the naive method gives lower coverage probability (e.g., at point (3.9,3.9)) while the improvement is smaller when the naive method gives coverage probability closer to nominal (Table 1 and Figure 3). This phenomenon is also seen in the length of prediction

Table 2. Summary of Cross-Validations: Estimated Coverage Probability ($\hat{p}$), Mean and Variance of the Prediction Interval (PI) Lengths, and Labels of the Sites Which did not Fall Within Their Estimated Prediction Intervals

Method	$\hat{p}$	*Mean PI len*	*Var. PI len*	*Sites not covered*							
Naive(m_1)	0.9412	3.053	0.2373	17	24	29	53	55	81		108
Boot. adjust.(m_B^*)	0.9496	3.149	0.2310	17	24	29	53		81		108
Boot. adjust.(m_B^{**})	0.9496	3.236	0.2743	17	24	29	53		81		108
Direct boot. (m_B^d)	0.9412	3.148	0.1856	17	24	29	53	55	81		108
Bayesian	0.9580	3.183	0.0640		24	29	53	55		101	

intervals (Table 1). The direct bootstrap prediction interval (using $m_B^d(\hat{\boldsymbol{\alpha}})$) performs better than naive method and similarly to the adjustment method using $m_B^*(\hat{\boldsymbol{\alpha}})$. Overall the coverage probability for data generated with $\boldsymbol{\alpha}_1$ are better than the more spatially correlated data generated by $\boldsymbol{\alpha}_2$. And, as expected, the coverage probabilities are all closer to nominal as the sample size increases from 25 to 144 (Table 1 and Figure 3).

To see how well these methods perform when the data are nonnormal, first we generated normal data with an exponential variogram with the same $\boldsymbol{\alpha}_2$ as we used before. We then exponentiated the data to make it log normal and repeated the simulation for data on the 6×6 grid. In particular, the bootstrap samples are still drawn from a $N(\hat{\mu}\mathbf{1}_n, C(\hat{\boldsymbol{\alpha}}))$ distribution even though the data are clearly not normal. Note the variability of the "observed" data is larger because of the exponentiation. We found both the naive method and parametric bootstrap methods performed similarly to when the data were normal. It is not expected that the prediction intervals will perform well for all types of nonnormal data especially in small samples. Further investigation of the robustness for nonnormality is necessary.

7. APPLICATION

We applied the proposed bootstrap methods to compute the prediction intervals in the example introduced in Section 2. Bayesian methods also provide a convenient way of incorporating parameter uncertainty into predictive inferences in the form of posterior predictive distributions. The intractability of computation time precluded us from considering the Bayesian method in our simulation study, but we consider it here for comparison in our example.

We evaluated the coverage probability of each method against the others using cross validations. That is, we deleted one observation Z_i from the dataset and then used the remaining data $\mathbf{Z}_{(-i)}$ to predict at that location, and check to see whether the prediction interval covered the true value Z_i or not. This procedure was repeated by subsequently deleting each of the data points one at a time and calculating the overall coverage of the prediction intervals. We used WinBUGS/GeoBUGS (Spiegelhalter, Thomas, and Best 1999) software to do the Bayesian cross validations. Following an example in the lecture notes by Best, Marshall, and Thomas ("Spatial Modeling Using WinBUGS and GeoBUGS," short course, Brisbane, Nov. 30–Dec. 1, 2000), the vague priors (gamma$(0.01, 0.01)$ for nugget

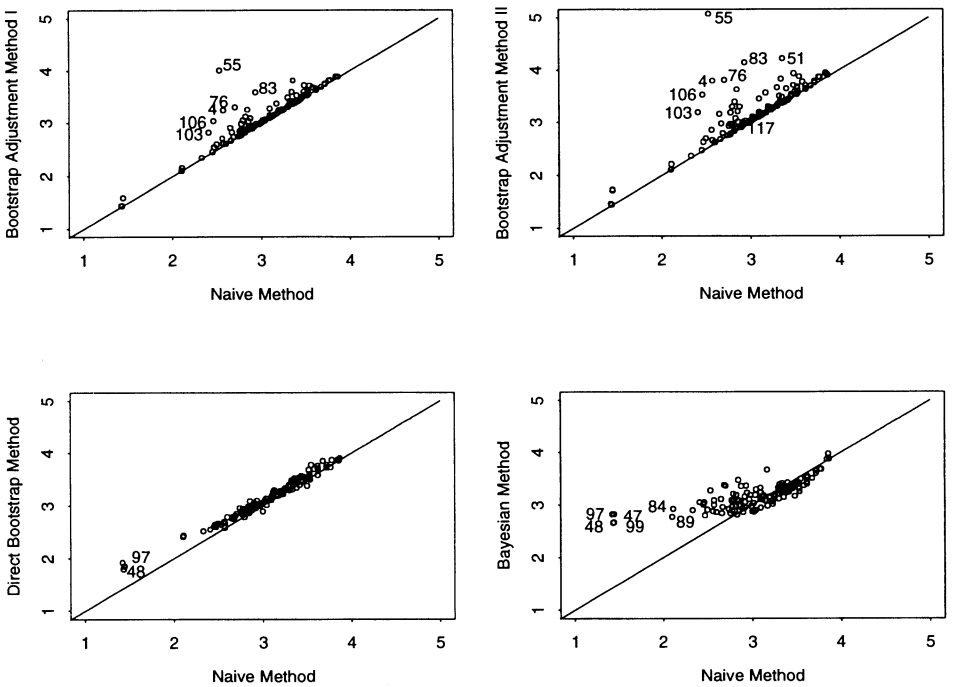

Figure 4. Scatterplots of the length of prediction intervals.

and sill parameters, uniform$(0.01, 4)$ for the range parameter) were used.

From the summary of the cross-validations (Table 2), we see that the estimated coverage probabilities for the five different methods are similar to each other. In this example with sample size of 119, we will focus on comparing the length of predictions intervals of these methods.

Note that site 55 is covered only by the bootstrap adjustment methods (Table 2). The reason for this can be seen from the scatterplot of the length of prediction intervals for naive method (using $m_1(\hat{\alpha})$) versus the other methods respectively (Figure 4). The length of the bootstrap adjustment method follows the length of the naive method in general. But at sites 4, 55, 76, 83, 103, and 106, the length differences between the two methods are much bigger than usual. It is this increase in length that makes the point 55 covered by bootstrap adjustment methods. From Figure 1, the geographic locations of observed sites, we find that these sites with larger interval length using the bootstrap adjustment methods are all very close to the uncovered sites or "outliers." That is, 4, 76, and 103 are near 81; 55 and 83 are near 24; 106 is near 108. From Table 2, we know 24, 81, and 108 are all outliers. We know that when a predicted site is very close to an observed site, the prediction will tend to be close to the observed site due to spatial correlations, hence if the observed site is an "outlier," the prediction will also be outlying. The spatial proximity does not play the same role in calculating the prediction error. The prediction error of the naive method does not depend on the data directly, that is, it only depends on the data through the estimated parameter $\hat{\alpha}$. Hence the prediction error from the naive method is not influenced by the

outliers in the data but only by the overall spatial correlations. On the other hand, from (5.5), we see that in the bootstrap adjustment method the adjustment term does depend on the data directly, so it is sensitive to outlying neighbors. Intuitively, if a predicted site is very close to an outlying neighbor, we would be less certain about the accuracy of the predicted value. To reflect this uncertainty, the prediction interval should be wider.

The Bayesian method is not very responsive to nearby observed points (Figure 4). This can be best seen at sites 47, 48, 97, 99 where the predicted site is very close to an observed site. At these sites, the length of intervals from naive and bootstrap methods are much smaller than the Bayesian prediction intervals. The Bayesian prediction interval seems not to be influenced much by its proximity to other neighbors. Therefore, the length of Bayesian prediction intervals stays relatively constant through the region. The variance of the length of Bayesian prediction intervals is 0.06, much smaller than the corresponding variances from the other methods (0.19 to 0.27) (Table 2).

This example suggests that the proposed parametric bootstrap methods allow us to incorporate the parameter uncertainty into prediction intervals to get better coverage probability. No complicated formula derivation is needed, so they are more easily implemented than the methods proposed by Kackar and Harville (1984) and Harville and Jeske (1992). They are more responsive to data than the Bayesian method and do not require priors.

8. CONCLUSION

In this article we demonstrated that the naive prediction interval achieves less-than-nominal coverage probability and varies as the true range and sill are varied. We proposed three parametric bootstrap adjustments to the prediction interval. Two are based on the analytic adjustment by Kackar and Harville (1984) and Harville and Jeske (1992), and the third directly estimates the MSPE by using the parametric bootstrap to generate realizations at the prediction sites.

From the simulation study, we find the parametric bootstrap adjustment methods show improvement over the naive method and, in particular, that m_B^{**} is closest to nominal coverage probability in most cases. The improvement is small when the naive method performs well, bigger when the naive method performs poorly. This desirable property is related to the fact that the adjustment methods not only depend on the overall spatial correlations but also uses the observed data directly. The Bayesian method incorporates the parameter uncertainty into the prediction inference more naturally, but it is less responsive to data than the bootstrap methods as we have seen in Section 7.

Although the simulation study shows that the adjustment methods outperforms the naive method, we note that the naive method performs reasonably well when the sample size is 144 under the conditions investigated here. This may be in part due to the fact that for prediction, Stein (1988) has shown that it is most important to capture the behavior of the variogram near the origin, and even though the estimator $\hat{\alpha}$ may be quite variable, the estimated variogram function near the origin might not be very variable. Nevertheless, the bootstrap adjustment methods are recommended for use with small sample sizes, where it

has been demonstrated that the naive method will tend to underestimate the true MSPE.

Although our proposed parametric bootstrap prediction interval methods can be used with any estimator for α, the results of our simulation study are limited to the case of OLS fit to the classical empirical variogram, for exponential variogram and increasing domain sample size. Further study is needed to investigate the performance of the prediction intervals under different methods for estimating α as well as for infill increasing sample size.

ACKNOWLEDGMENTS

The authors thank an associate editor and two referees for their detailed comments which improved the quality of this article. We also thank K. Cowles for providing us with a cleaned version of the NADP data used in our example. This work was supported by National Center for Health Statistics grant NCHS-UR6/CCU517477-01.

[Received February 2002. Revised February 2003.]

REFERENCES

Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998), "Model-Based Geostatistics" (with discussion), *Applied Statistics*, 47, 299–350.

Genton, M. G. (1998), "Variogram Fitting by Generalized Least Squares Using an Explicit Formula for the Covariance Structure," *Mathematical Geology*, 30, 323–345.

Harville, D. A., and Jeske, D. R. (1992), "Mean Squared Error of Estimation or Prediction under a General Linear Model," *Journal of the American Statistical Association*, 87, 724–731.

Kackar, R. N., and Harville, D. A. (1984), "Approximations for Standard Errors of Estimators of Fixed and Random Effect in Mixed Linear Models," *Journal of the American Statistical Association*, 79, 853–862.

Lahiri, S. N., Lee, Y., and Cressie, N. (2002), "On Asymptotic Distribution and Asymptotic Efficiency of Least Squares Estimation of Spatial Variogram Parameters," *Journal of Statistical Planning and Inference*, 103, 65–85.

Matheron, G. (1962), "Traite de Geostatistique Appliquee, Tome I," *Memoires du Bureau de Recherches Geologiques et Minieres*, 14, Paris: Editions Technip.

——— (1971), "The Theory of Regionalized Variables and Its Application," *Ecole Nationale Superiore des Mines de Paris*.

Prasad, N. G., and Rao, J. N. K. (1986), "On the Estimation of Mean Square Error of Small Area Predictors," Technical Report 97, Carleton University, Laboratory for Research in Statistics and Probability, Ottawa.

Putter, H., and Young, A. (2001), "On the Effect of Covariance Estimation on the Accuracy of Kriging Predictors," *Bernoulli*, 7, 421–438.

Spiegelhalter, D. J., Thomas, A., and Best, N. G. (1999), *WinBUGS Version 1.2 User Manual*, MRC Biostatistics Unit.

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

Zimmerman, D., and Cressie, N. (1992), "Mean Squared Prediction Error in the Spatial Linear Model with Estimated Covariance Parameters," *Annals of the Institute of Statistical Mathematics*, 44, 27–43.

Zimmerman, D. L., and Zimmerman, M. B, (1991), "A Comparison of Spatial Semivariogram Estimators and Corresponding Ordinary Kriging Predictors," *Technometrics*, 33, 77–91.