

# A Parametric Spatial Bootstrap

Liansheng Tang<sup>a</sup>; William R. Schucany<sup>\*b</sup>; Wayne A. Woodward<sup>b</sup>; and Richard F. Gunst<sup>b</sup>

July 17, 2006

<sup>a</sup> *Department of Biostatistics, University of Washington, P.O. Box 357232, Seattle, WA 98195, USA*

*This work was part of the first author's PhD dissertation at Southern Methodist University.*

<sup>b</sup> *Department of Statistical Science, Southern Methodist University, P.O. Box 750332, Dallas, TX 75275, USA*

<sup>\*</sup> *Corresponding author. Department of Statistical Science, Southern Methodist University, P.O. Box 750332, Dallas, TX 75275, USA*

*E-mail address: schucany@mail.smu.edu.*

**Abstract:** The classic bootstrap uses valid resamples whenever the observations are independent and identically distributed. Data from a spatial region usually have a correlated structure. If the bootstrap is naively applied to spatially dependent data, these correlations will be lost. This paper proposes a new parametric spatial bootstrap. The proposed method is motivated by an analysis of brain images in Spence (2004). The new method combines spatial modeling and the parametric bootstrap to produce valid resamples of spatially correlated normal data. The coverage of confidence intervals from the proposed method is estimated and compared to an existing spatial bootstrap method proposed by Solow (1985). The two procedures are simulated for several two-dimensional images. The new approach is based on fitting a Gaussian semivariogram, uses the Cholesky decomposition and generates standard normals. The method guarantees positive definiteness of covariance matrices and has significantly better coverage than the existing spatial bootstrap approach.

*Keywords:* Cholesky; dependent data; Gaussian; normal; resampling; semivariogram

# 1 Introduction

In spatial data any pair of observations may be correlated due to their positions in space. Such examples can be found in many areas, including brain imaging, geology, and environmental monitoring. In these cases data are collected from various spatial locations and statistical models are often used to represent the dependence between measurements at different locations. The spatial domain may be one, two, or three dimensional. Spatial images can be analyzed regardless of whether the spatial domain is a continuous region or a regular grid. For ease of presentation, we only consider regular grids in this article.

If spatial data have very weak correlations, then a simple basic bootstrap method can yield approximately valid resamples of the observed data. However, in strongly correlated spatial data, naively ignoring the spatial correlation inherent in the data and directly resampling the observations will inevitably destroy the correlation structure among the original observations. In addition, a point estimator  $\hat{T}$  may not be difficult to derive for a parameter of interest from observed spatial data. However, a confidence interval usually involves complicated calculations of the sampling distribution for  $\hat{T}$ , so fidelity to the correlation structure is critically important.

To improve the bootstrap method for dependent data, several different block bootstrap methods (Carlstein, 1986; Kunsch, 1989; Liu and Singh, 1992; Politis and Romano, 1992, 1994) have been proposed to nonparametrically replicate the dependence structure of the observations in the resamples. The block bootstrap preserves the dependence structure within the blocks, but distorts the correlation between blocks. Hall et al. (1995) pointed out that the bias and the variance of a block bootstrap estimator are heavily influenced by the block length. Solow (1985) introduces and Cressie (1993, Section 7.3.2) briefly describes a particular bootstrap procedure for spatially correlated data. Sjöstedt-de Luna and Young (2003) calibrated kriging prediction intervals from bootstrap resamples. Their resamples assume a Gaussian process with estimated covariance matrices. The present paper introduces a modification of a spatial bootstrap (SB) method proposed in Solow (1985). We adapt the parametric approach in Sjöstedt-de Luna and Young (2003) to obtain valid resamples from a Gaussian process. The proposed method offers a correct way to obtain confidence limits based on the statistic  $\hat{T}$ . The proposed parametric spatial bootstrap accounts for the spatial correlation in the data by estimating this structure and then imposing it in the resamples. Specifically, the new method is based upon the normality of semivariograms within deep brain regions found in SPECT studies (Spence, 2004).

This paper is organized as follows. Section 2 reviews basic terms and issues in spatial statistics. Section 3 introduces the parametric spatial bootstrap algorithm. Section 4 gives the algorithms for the spatial processes used in our simulation study. Different spatial data simulation techniques are discussed with respect to their computational efficiency. The simulation results of the proposed method and an existing method are compared with respect to the coverage of their confidence intervals on two-dimensional mean levels. Section 5 contains some conclusions and discussion of this parametric version of the spatial bootstrap. Technical issues on semivariograms and computational speeds are relegated to the appendix.

## 2 Spatial Statistics

### 2.1 Semivariogram

Suppose a location  $\mathbf{s} = (s_1, s_2, \dots, s_n)'$  is a vector in  $n$ -dimensional Euclidian space  $\mathfrak{R}^n$ , and  $Z(\mathbf{s})$  is a random magnitude at the location  $\mathbf{s}$ . A random field is

$$\{Z(\mathbf{s}) : \mathbf{s} \in D \subset \mathfrak{R}^n\},$$

where  $\mathbf{s}$  varies over the index set  $D$ , which is a subset of  $\mathfrak{R}^n$ . The random field  $Z(\mathbf{s})$  is decomposed into a deterministic mean function  $\mu(\mathbf{s})$  and a correlated error process  $\delta(\mathbf{s})$  as

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + \delta(\mathbf{s}), \quad \mathbf{s} \in D,$$

(see Cressie, 1993, Section 2.4). The error process  $\delta(\cdot)$  is assumed to be a zero-mean intrinsically stationary spatial process

There are a large number of conditionally negative definite semivariogram models in the literature on spatial modeling; e.g., Journel and Huijbregts (1978) and Cressie (1993, Section 2.3.1). Two quite popular isotropic semivariogram models are considered in this paper. The first is the Gaussian model

$$\gamma^G(d; \underline{\theta}) = \begin{cases} 0 & \text{if } d = 0 \\ \theta_1 + \theta_2 \{1 - \exp(-(d/\theta_3)^2)\} & \text{if } d > 0, \end{cases}$$

where  $\underline{\theta} = (\theta_1, \theta_2, \theta_3)$ , and  $\theta_1, \theta_2, \theta_3 \geq 0$ . The shape of the Gaussian model is displayed in Figure 1 using the nugget parameter  $\theta_1 = 0$ , the sill (because  $\theta_1 = 0$ )  $\theta_2 = 1$  and the range parameter  $\theta_3 = 3$ . The vertical dashed line shows the effective range,  $\sqrt{3}\theta_3$ , and the horizontal dashed line at unit variance is the sill  $\theta_2$ , where  $\gamma^G$  approaches its limit. The zero nugget is shown as the horizontal dotted line. It is clear from the plot that this semivariogram strictly increases with distance and becomes flat when the distance  $d$  gets close to the range.

The other semivariogram model investigated here is the exponential semivariogram model

$$\gamma^E(d; \underline{\theta}) = \begin{cases} 0 & \text{if } d = 0 \\ \theta_1 + \theta_2 \{1 - \exp(-d/\theta_3)\} & \text{if } d > 0, \end{cases}$$

where  $\underline{\theta} = (\theta_1, \theta_2, \theta_3)$ , and  $\theta_1, \theta_2, \theta_3 \geq 0$ .

Both parametric and nonparametric methods have been introduced for estimating the semivariogram. The parametric approach is usually based on weighted-least-squares (WLS) semivariogram fitting (Cressie, 1985). Appendix 1 gives detailed description of parametric estimation of the semivariogram. Nonparametric estimators of the semivariogram are introduced by several authors (Shapiro and Botha, 1991; Hall et al., 1994; Cherry et al., 1996; Genton and Gorsch, 2002). Even though the nonparametric approach does not require any parametric model assumption for the semivariogram, the resulting estimates are not conditionally negative-definite in many cases in practice due to numerical instability.

Initially, it seemed that the nonparametric approach would be preferable in our applications. However, realistic spatial data were simulated and the classic semivariogram estimates obtained. Shapiro-Botha semivariogram fits were obtained using R programs from Statlib

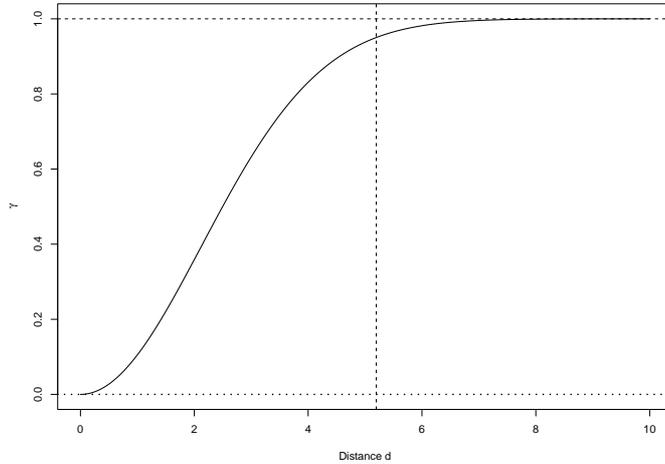


Figure 1. Gaussian semivariogram model with zero nugget, unit sill, and a range of 3.

nonparametric estimation algorithms contributed by Steve Cherry and Jeff Banfield. The corresponding covariance matrix estimates were calculated and then Cholesky decompositions were attempted on these covariance matrices. It turned out that none of the matrices were positive-definite so that the algorithm failed to give the Cholesky matrices. Even though the selection of a conditionally negative-definite parametric semivariogram family can be subjective, once the parameters of the semivariogram model are obtained, the corresponding fitted semivariogram model provides valid variance estimates. This property, coupled with results (e.g., Stein 1988) that show the critical feature is not the model selected but good estimation for small lags, led to the decision to use parametric semivariograms for further study.

## 2.2 Estimation of the Covariance Matrix

There is a close relationship between the semivariogram  $\gamma(\cdot)$  and the covariogram  $C(\cdot)$ . If  $\delta(\cdot)$  is second-order stationary, then

$$\gamma(\mathbf{s}_1 - \mathbf{s}_2) = C(0) - C(\mathbf{s}_1 - \mathbf{s}_2). \quad (1)$$

Furthermore if  $\gamma(\cdot)$  and  $C(\cdot)$  are functions only of Euclidean distance  $d$ , then the above equation implies that  $\delta(\cdot)$  is intrinsically stationary with  $\gamma(d) = C(0) - C(d)$ .

The method-of-moments covariogram estimator,  $\hat{C}(\cdot)$  in Appendix 1 yields an estimator of the covariance matrix by evaluating it for each of the  $n \times n$  elements. However, the resulting covariance matrix is not necessarily positive definite. Therefore it cannot be directly used for the parametric spatial bootstrap method described in Section 3.

The parametric semivariogram method gives a valid estimator  $\gamma(d; \hat{\theta})$ . Equation (1) suggests a relationship between the estimator,  $\hat{C}(d)$  and  $\gamma(d; \hat{\theta})$ , namely

$$\hat{C}(d) \simeq \hat{C}(0) - \gamma(d; \hat{\theta}),$$

where  $\hat{C}(0) = \hat{\sigma}_\delta^2 = \text{sill}$ . An estimate,  $\hat{\Sigma}$ , of the  $n \times n$  covariance matrix can then be obtained from

$$\hat{\Sigma}(i, j) = \hat{C}(d_{ij}),$$

where  $d_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$  for  $i, j = 1, \dots, n$ .

### 3 Parametric Spatial Bootstrap

The naive nonparametric bootstrap method fails to provide valid resamples whenever there is correlation in either time series or spatial data. When this bootstrap is applied to correlated data, it randomizes the residuals or the observations and destroys the correlation pattern inherent in the joint distribution.

For realizations of the spatial model in Section 2.1, the estimates for the deterministic components  $\hat{\boldsymbol{\mu}}$  are obtained using various methods (e.g., Cressie 1993, Section 3.1). Next the estimated spatial error process can be calculated as

$$\begin{aligned} \hat{\boldsymbol{\delta}} &= \{\hat{\delta}(\mathbf{s}_1), \dots, \hat{\delta}(\mathbf{s}_n)\} \\ &= \{Z(\mathbf{s}_1) - \hat{\mu}(\mathbf{s}_1), \dots, Z(\mathbf{s}_n) - \hat{\mu}(\mathbf{s}_n)\} \\ &= \mathbf{Z} - \hat{\boldsymbol{\mu}}. \end{aligned}$$

Appendix 1 discusses practical aspects of estimating the semivariograms and associated covariance matrix. The resulting positive definite covariance matrix is then decomposed using the Cholesky decomposition

$$\hat{\Sigma} = \hat{L}\hat{L}^T,$$

where  $\hat{L}$  is a lower triangular  $n \times n$  matrix.

Solow (1985) uses the Cholesky decomposition matrix inverse,  $\hat{L}^{-1}$ , to decorrelate the spatial error sequence

$$(\hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots, \hat{\varepsilon}_n) \equiv \hat{\boldsymbol{\varepsilon}} \equiv \hat{L}^{-1}\hat{\boldsymbol{\delta}}$$

and then centers the  $\hat{\boldsymbol{\varepsilon}}$  to obtain

$$\tilde{\varepsilon}_i = \hat{\varepsilon}_i - \frac{1}{n} \sum_{j=1}^n \hat{\varepsilon}_j$$

for  $i = 1, \dots, n$ . The decorrelated and centered residuals  $\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \dots, \tilde{\varepsilon}_n$  are then bootstrapped to provide the resampled residuals  $\boldsymbol{\varepsilon}_{SB}^* = (\varepsilon_1^*, \dots, \varepsilon_n^*)$ . The SB resample is obtained by transforming to recorrelate the bootstrapped residuals

$$\mathbf{Z}^* = \hat{\boldsymbol{\mu}} + \hat{L}\boldsymbol{\varepsilon}_{SB}^*.$$

In this paper Solow's method is modified for better coverage of the confidence intervals, as discussed in the simulations in Section 4. The new parametric bootstrap method generates spatially correlated residuals similar to the ones used in Sjöstedt-de Luna and Young (2003) to calibrate a kriging prediction interval. More importantly, the parametric spatial bootstrap (PSB) algorithm introduced in this paper is motivated by the Gaussian semivariograms fit in deep brain regions found in certain SPECT studies (Spence, 2004). The procedure follows

the same steps as SB up to the Cholesky decomposition of  $\hat{M}$ . At this point the parametric bootstrap residuals  $\boldsymbol{\varepsilon}_{PSB}^*$  are generated from a  $N(0, 1)$ ; i.e.,

$$\boldsymbol{\varepsilon}_{PSB}^* = (\varepsilon_1^*, \varepsilon_2^*, \dots, \varepsilon_n^*), \quad \text{where } \varepsilon_j^* \sim N(0, 1) \quad \text{for } j = 1, \dots, n.$$

Next, the spatial resamples are  $\mathbf{Z}^* = \hat{\boldsymbol{\mu}} + \hat{L}\boldsymbol{\varepsilon}_{PSB}^*$ . Finally, we calculate the statistic of interest,  $\hat{T}^*$ , from  $\mathbf{Z}^*$ . The above procedures are repeated  $B$  times to estimate the sampling distribution of  $\hat{T}$ .

PSB does not obtain the residuals  $\boldsymbol{\varepsilon}^*$  by decorrelating the spatial error process as in the SB algorithm. Instead, the residuals are independently generated from a standard normal distribution. This faithfully mimics the spatial model introduced in Section 4.1. The theoretical foundation is in Sjöstedt-de Luna and Young (2003).

## 4 SIMULATIONS

### 4.1 Simulation algorithm

Cressie (1993, Section 3.6) summarizes several spatial data simulation procedures, which are based on the Cholesky decomposition (Cressie and Laslett, 1987) or eigenvalue decomposition. The Cholesky decomposition procedure in Section 3 allows the covariance matrix  $\Sigma$  to be factored into the product of two matrices

$$\Sigma = LL^T.$$

The eigenvalue decomposition calculates the square root of  $\Sigma$  based on

$$\Sigma = Q \text{diag}\{\lambda_1, \dots, \lambda_n\}Q^T,$$

where  $\lambda_1, \dots, \lambda_n$  are the eigenvalues of  $\Sigma$  and the matrix  $Q$  consists of the corresponding eigenvectors of  $\Sigma$ . A matrix  $\Sigma^{1/2}$  is obtained by

$$\Sigma^{1/2} = Q \text{diag}\{\lambda_1^{1/2}, \dots, \lambda_n^{1/2}\}Q^T.$$

The Cholesky decomposition is preferred not only because it is widely available from most computer packages with matrix operations, but also because it is more computationally efficient than the eigenvalue decomposition. We report a detailed comparison in the Appendix justifying our use of the Cholesky decomposition.

The algorithm for simulating spatial data in this study is:

1. Choose a constant mean,  $\mu$ , for the entire random field ( $\mu=1, 2$  or  $4$ ).
2. Choose one of the two covariance matrices,  $\Sigma^P$  and  $\Sigma^G$  (see Section 4.2).
3. Obtain the Cholesky matrix  $L_P$  such that  $\Sigma^P = L_P L_P^T$  and similarly for  $L_G$ .
4. Generate independent random variables  $\boldsymbol{\varepsilon}$  from the standard normal distribution,

$$\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n), \quad \text{where } \varepsilon_j \sim N(0, 1) \quad \text{for } j = 1, \dots, n.$$

5. Simulate the spatial data  $\mathbf{Z}$  using the relationship

$$\mathbf{Z} = \mu + L_P \boldsymbol{\varepsilon}.$$

## 4.2 Data simulation

Two-dimensional spatial data  $\mathbf{Z}$  are considered here. For simplicity  $\mathbf{Z}$  has a constant mean  $\mu$  over the entire field. Thus the spatial errors  $\hat{\boldsymbol{\delta}}$  are reasonably estimated by subtracting the sample mean from the realization of the random field,

$$\hat{\boldsymbol{\delta}} = (\hat{\delta}(\mathbf{s}_1), \dots, \hat{\delta}(\mathbf{s}_n)) = \mathbf{Z} - \frac{1}{n} \sum_{j=1}^n Z(\mathbf{s}_j).$$

Two stationary covariance structures are studied here. The first,  $\Sigma^P$ , has the simple power form in which each location is correlated with all other locations according to

$$C(d) = \sigma^2 \rho^d,$$

where  $d$  is the Euclidean distance between the locations and  $\sigma^2$  is the variance of the spatial process. This power semivariogram is included in the simulation to investigate how the semivariogram models, Gaussian and exponential, perform when the model is incorrect.

Without loss of generality,  $\sigma^2 = 1$  in these simulations. The power covariance matrix may be written as  $\Sigma^P(i, j) = \rho^{d_{ij}}$ , where  $d_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$  for  $i, j = 1, \dots, n$ . For  $\rho = 0.7$  the covariance matrix is illustrated in a grayscale map, motivated by Stein (1988), in Figure 2(a) for an  $8 \times 8$  spatial image by ordering the pixels from top to bottom and from left to right, and calculating the correlation between each pair of pixels. The bands in the  $64 \times 64$  correlation image in Figure 2(a) indicate that the correlation is strong among local pixels and weaker with increased distance between pixels. The correlation pattern in this map has box-like shapes, because the ordered pixels from a two-dimensional image do not have strictly decreasing correlation with location in the correlation matrix. For example, the 1st and the 9th ordered pixels from an  $8 \times 8$  two-dimensional image do not have the correlation associated with pixels eight units apart. To the contrary, they are neighbors in the image and have the strongest correlation of only one unit apart. The theoretical power semivariogram is given by  $\gamma(d) = 1 - \rho^d$ .

The second covariance structure for data simulation is based on the Gaussian semivariogram model whose covariance matrix is

$$\begin{aligned} \Sigma^G(i, j) &= C(0) - \gamma^G(d_{ij}; \underline{\theta}) \\ &= \sigma^2 - (\theta_1 + \theta_2 \{1 - \exp(-(d_{ij}/\theta_3)^2)\}), \end{aligned}$$

where  $d_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$  for  $i, j = 1, \dots, n$ . In the simulation study the nugget parameter is set to  $\theta_1 = 0$  for continuity of the Gaussian semivariogram at the origin, the sill  $\theta_2 = 1$ , and the range parameter  $\theta_3 = 3$ , which means that the locations about 5 units apart are approximately uncorrelated. This range has been shown to be relevant within deep brain regions in SPECT studies (Spence, 2004). The covariance matrix for this Gaussian semivariogram is shown in a grayscale map in Figure 2(b) for an  $8 \times 8$  spatial image. The correlation here decreases with distance somewhat slower than the one in Figure 2(a) for the power covariance matrix. Figures 3(a) and 3(b) illustrate simulated 2D spatial processes and their perspective plots for both the power semivariogram with  $\rho = 0.7$  and the Gaussian semivariogram. The top two panels are the maps of  $16 \times 16$  2D images. From the perspective

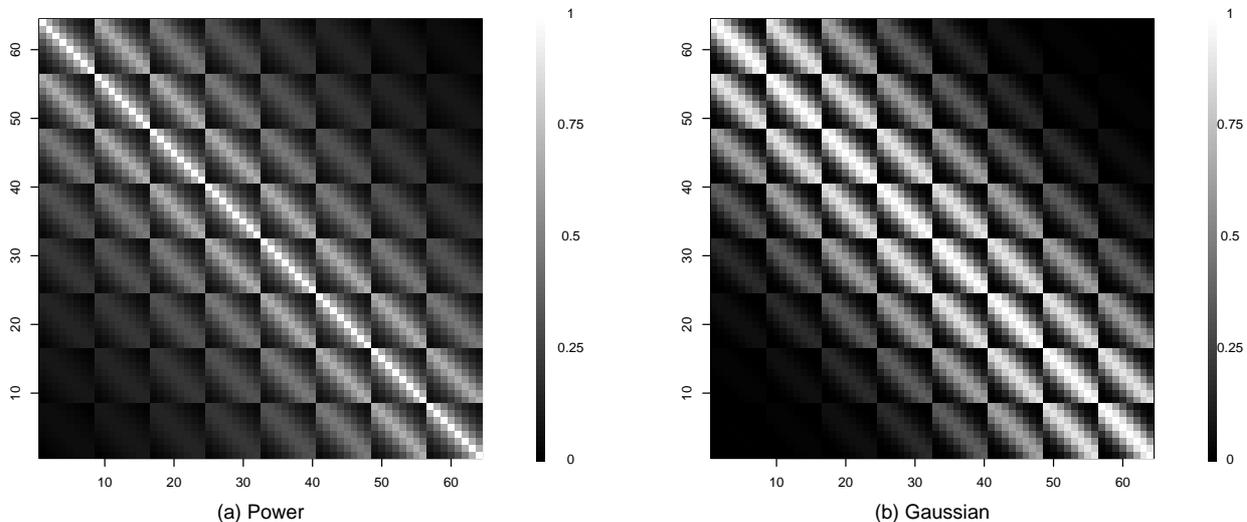


Figure 2. Correlation maps for Power and Gaussian covariance structures. The grayscale bar on the right is the magnitude of the correlations for each pair of pixels. (a) A Power covariance matrix :  $C(d) = \rho^d$  with  $\rho = 0.7$ ; (b) A Gaussian covariance matrix.

plots, it can be seen that the spatial process in Figure 3(c) has much more abrupt changes than the smoother Gaussian process in Figure 3(d).

For each simulated two-dimensional data set the PSB and SB methods are repeated 499 times to obtain approximate 90% confidence limits for the mean  $\mu$ . Sometimes a simulated data set may not yield a convergent semivariogram model fit within 30 iterations. When this occurs the simulation is considered invalid. This occurred about one out of ten times for the Gaussian semivariogram fit to Gaussian processes in this study. One thousand valid simulations are produced and the coverage of the 1000 resulting confidence intervals are obtained for each bootstrap method.

Both Gaussian and exponential semivariogram models are fit using WLS estimators. The experimental design accounts for three different constant means,  $\mu = 1, 2, 4$ , and two image sizes,  $8 \times 8$  and  $16 \times 16$ . All computations are performed in the R language. Figures are produced using R scripts available from the authors.

### 4.3 Simulation results

Table 1 gives simulation results at the nominal confidence level of 90% ( $SE \approx 0.9\%$ ) for  $\mu = 1, 2$  and 4, respectively. The SB method in the two columns on the right has coverage of about 80%, which is far below the nominal level ( $> 10 SE$ ) while the coverage for PSB is much better. The reason for the undercoverage using SB is that the estimated covariance matrix does not decorrelate the spatial error process. For example consider a  $16 \times 16$  realization generated using the Gaussian semivariogram. This corresponds to the smallest coverage (73.4%) in Table 1(b). The Gaussian semivariogram model was fit to the semivariogram values and then the covariance matrix was calculated. The spatial error process was decorrelated using the estimated inverse covariance matrix. Figure 4 shows one such

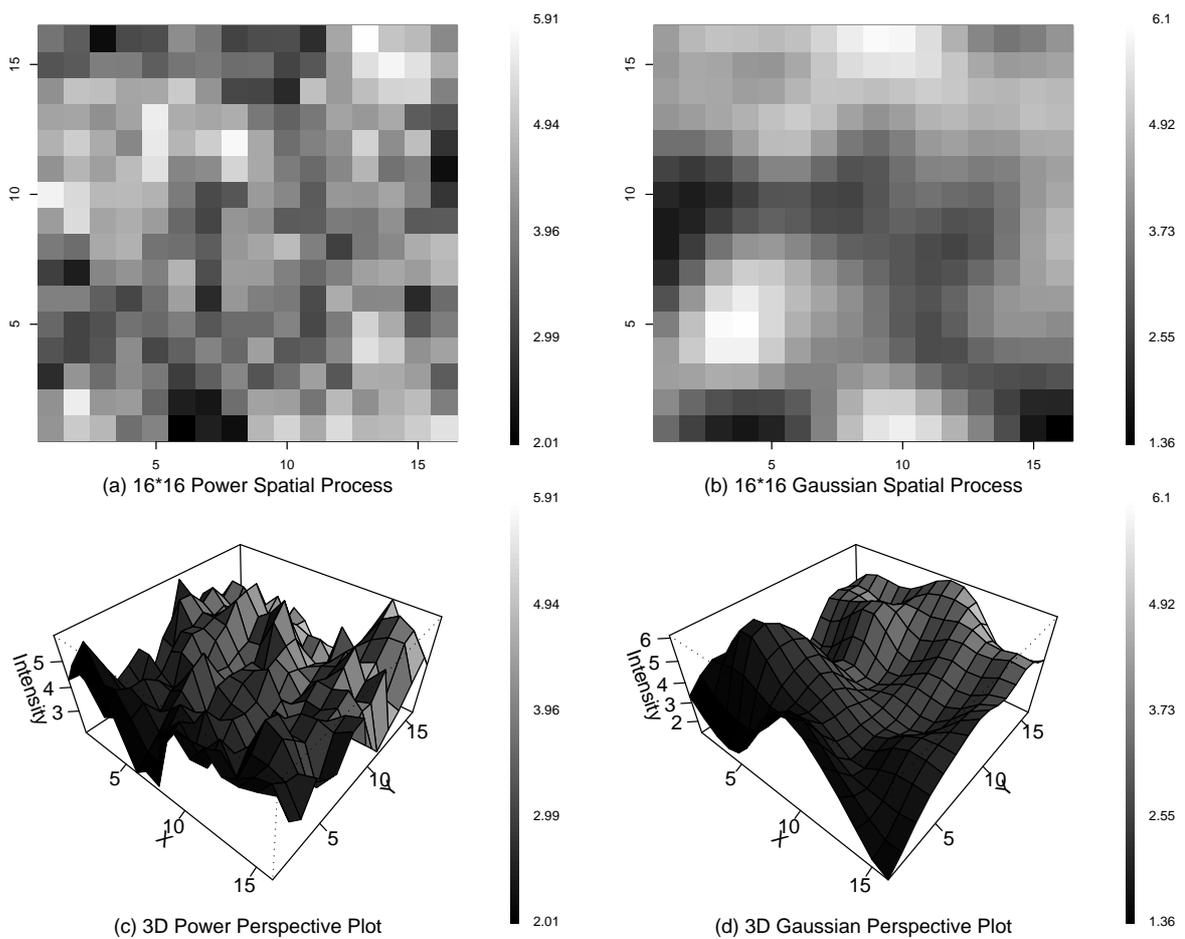


Figure 3. The simulated 2D spatial processes and their corresponding perspective plots. The constant mean for both processes is 4 and the image size is  $16 \times 16$ . Panels (a) and (c) give a power process image in a grayscale map and in a 3D setting, respectively; Panels (b) and (d) show a Gaussian process in a grayscale map and in a 3D setting, respectively.

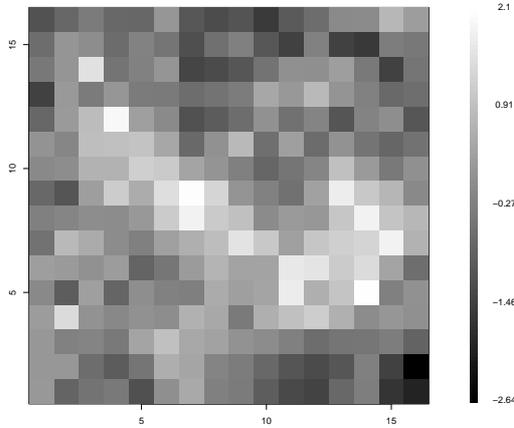


Figure 4. The “decorrelated” error. A  $16 \times 16$  spatial image is generated using the Gaussian semivariogram  $1 - \exp(-(d/3)^2)$  with a mean 4. The image shows the result of the decorrelated error process used in the SB procedure. Clear patterns of spatial correlation remain.

Table 1: Estimated confidence interval coverages in % using parametric spatial bootstrap and spatial bootstrap. The Power and Gaussian columns under bootstrap types indicate the semivariogram models generated. The rows labeled Gaussian and Exponential for each image size indicate the semivariogram type fit to the data. Simulations are from the Power semivariogram model with  $\rho = 0.7$  and a Gaussian model with  $\theta_1 = 0, \theta_2 = 1$  and  $\theta_3 = 3$ . The nominal confidence levels are 90% with standard errors of 0.9%.

(a) True mean $\mu = 1$					
		Parametric spatial bootstrap		Spatial bootstrap	
		Power	Gaussian	Power	Gaussian
Image size	Semivariogram				
$8 \times 8$	Gaussian	91.4	90.2	82.1	74.5
	Exponential	92.2	94.8	84.5	79.0
$16 \times 16$	Gaussian	91.2	93.6	79.7	75.1
	Exponential	94.6	95.8	77.2	79.2
(b) True mean $\mu = 2$					
$8 \times 8$	Gaussian	91.2	90.8	82.0	73.4
	Exponential	89.0	89.0	84.5	75.7
$16 \times 16$	Gaussian	93.4	94.2	74.5	74.2
	Exponential	92.5	96.0	73.6	73.6
(c) True mean $\mu = 4$					
$8 \times 8$	Gaussian	90.4	90.6	83.0	76.0
	Exponential	91.4	92.4	82.3	78.2
$16 \times 16$	Gaussian	93.0	94.1	77.5	74.0
	Exponential	91.2	95.1	76.8	76.4

approximately decorrelated error process for a typical realization. It is clear that correlation remains. It follows that the resamples obtained by bootstrapping these “decorrelated” error processes lead to the undercoverage of the resulting confidence intervals. On the other hand the coverage of the PSB intervals are not significantly less than the nominal confidence in all of the cases in Table 1.

## 5 Conclusions and Discussion

The PSB performed better than the spatial bootstrap under all conditions simulated. For smaller images the PSB using both the Gaussian and exponential semivariograms provide excellent coverage for the means, even when the fitted semivariogram is not the same form as the theoretical one used to generate the data. Figure 4(a) shows the model fits when the true semivariogram is of the power form. Figure 4(b) shows the semivariogram fits of the empirical semivariogram estimates as well as the true underlying Gaussian semivariogram curve for an  $8 \times 8$  spatial image. Note that the Gaussian fit approximates the empirical values well but the exponential fit is poor for small distances. Neither semivariogram fit approximates the empirical semivariogram very well. Figure 4(a) and Figure 4(b) also illustrate some departures of the empirical semivariogram estimates from the true underlying models. More severe departures occur in the simulated  $16 \times 16$  images. This appears to be the main reason that for  $16 \times 16$  images the PSB tends to have some overcoverage, as much as 5 *SE* above the nominal level in Table 1.

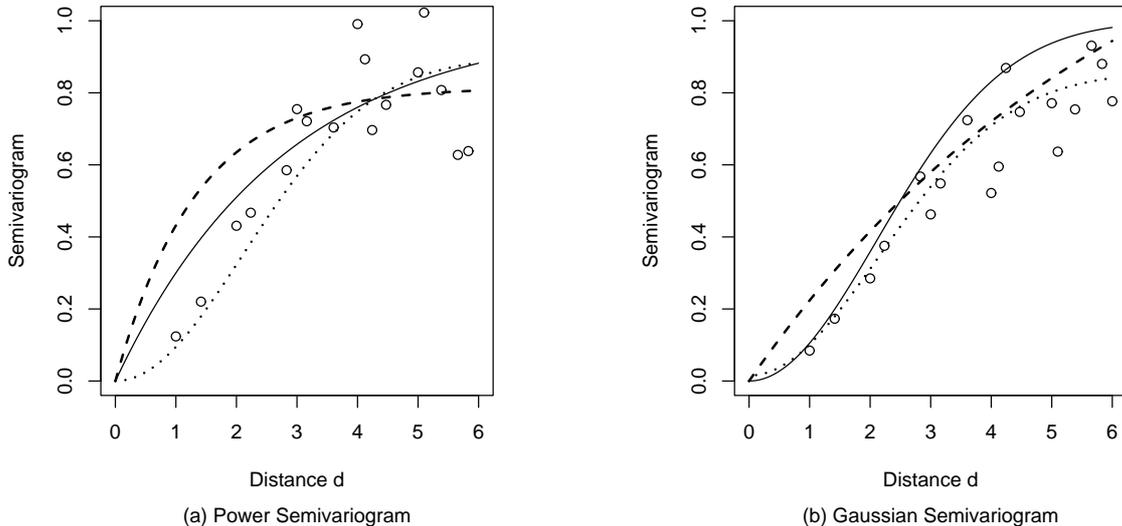


Figure 5. Semivariogram estimates for an  $8 \times 8$  spatial image. The solid lines are the true semivariograms. The dashed line shows the results of exponential model fitting and the dotted one is the result of Gaussian model fitting. (a) A Power semivariogram  $1 - 0.7^d$ ; (b) A Gaussian semivariogram  $1 - \exp(-(d/3)^2)$ .

The table also reveals that the Gaussian model fits tend to be closer to 90% coverage

than the exponential model for smaller image size, while for larger images these two model fits have similar overcoverage amounts. A one-factor, three-level ANOVA on the effect of the true means has a  $p$ -value of .56 for effect of different mean levels, which indicates that different levels of constant mean have little effect on the confidence interval coverages for PSB using both semivariogram models.

One of the main advantages of the PSB method over block bootstrap methods used in spatially correlated data is that it does not require one to determine block sizes and partition spatial data into blocks. The new method also guarantees positive definite covariance matrices and offers better coverage than the spatial bootstrap method proposed in Solow (1985). The tradeoff is that the proposed PSB method requires estimation of the correlation structure of a Gaussian process. Furthermore, if care is not taken to select a semivariogram model that reasonably approximates the empirical semivariogram, the parameter estimates might not converge. Extensions are being investigated to improve the stability of the proposed method for all forms of spatial data.

## APPENDIX 1: Estimation of the semivariogram

For a constant mean  $\mu(\mathbf{s}) = \mu$ , the method-of-moments estimator for the variogram was introduced in Matheron (1962). It is referred to as the classical variogram estimator by Cressie (1993, Section 2.4) and is given by

$$2\hat{\gamma}(d) = \frac{1}{|N(d)|} \sum_{N(d)} (Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2,$$

where  $N(d) = \{(\mathbf{s}_i, \mathbf{s}_j) : \|\mathbf{s}_i - \mathbf{s}_j\| = d \text{ for } i, j = 1, \dots, n\}$  including all distinct pairs of locations,  $\mathbf{s}_i, \mathbf{s}_j \in D \subset \mathfrak{R}^n$ , which are  $d$  units apart and  $|N(\mathbf{d})|$  is the number of these pairs. The semivariogram estimator is  $\hat{\gamma}(d)$ .

Several parametric methods for fitting a semivariogram model are summarized in Cressie (1993, Section 2.4.3). After the classic semivariogram estimator  $\hat{\gamma}(\cdot)$  is obtained as a vector

$$\hat{\boldsymbol{\gamma}} = (\hat{\gamma}(d_1), \hat{\gamma}(d_2), \dots, \hat{\gamma}(d_T))'$$

at  $T$  lags, the ordinary least squares (OLS) method obtains the value  $\underline{\boldsymbol{\theta}}$  that minimizes

$$(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}(\underline{\boldsymbol{\theta}}))'(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}(\underline{\boldsymbol{\theta}})).$$

If the correlation among estimator values  $\hat{\gamma}(d)$  at different lags is taken into account, the generalized least squares (GLS) criterion minimizes  $(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}(\underline{\boldsymbol{\theta}}))'V_G^{-1}(\underline{\boldsymbol{\theta}})(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}(\underline{\boldsymbol{\theta}}))$ , where  $V_G(\boldsymbol{\theta}) = \text{var}(\boldsymbol{\gamma})$ . However, calculation of all elements in  $V_G(\boldsymbol{\theta})$  is not a trivial task. Thus, a simpler approach, weighted least square (WLS), is widely adopted and only requires the estimation and inversion of a  $T \times T$  diagonal matrix  $V_W(\boldsymbol{\theta})$ , where

$$V_W(\boldsymbol{\theta}) = \text{diag}\{\text{var}(\gamma(d_1)), \text{var}(\gamma(d_2)), \dots, \text{var}(\gamma(d_T))\}.$$

Although GLS can be superior, Cressie (1984) and Grondona and Cressie (1995) show that there is negligible loss in efficiency using WLS instead of GLS. Thus, WLS is a more appropriate estimator due to its efficiency and ease of calculation. The weights in WLS are

given by the reciprocals of the variances of the semivariogram estimator  $\gamma$ . The diagonal terms of  $V_W$  can be approximated by  $\text{var}(\hat{\gamma}(d_i)) \simeq 2(\hat{\gamma}(d_i; \underline{\theta}))^2/|N(d_i)|$ . Thus, WLS obtains the estimate of  $\underline{\theta}$  such that

$$\hat{\underline{\theta}} = \underset{\underline{\theta}}{\text{argmin}} \left\{ \sum_{i=1}^n w(d_i) \{ \hat{\gamma}(d_i) - \gamma(d_i; \underline{\theta}) \}^2 \right\},$$

where

$$w(d_i) = 1/\text{var}(\hat{\gamma}(d_i)) = |N(d_i)|/2(\hat{\gamma}(d_i; \underline{\theta}))^2 \quad \text{for } i = 1, \dots, n.$$

The above equations illustrate that WLS gives more weight to squared deviations that have more pairs of observations or which are at smaller lags  $d_i$ . In particular, the latter condition means that the deviations are weighted more if their lags are closer to the origins, improving the fit of the semivariogram near the origin. Stein (1988) argues the critical importance of estimating semivariogram models well at small lags.

An initial value of the vector  $\underline{\theta}^{(0)}$  can be obtained from the sample semivariogram values  $\hat{\gamma}(d_i)$  using OLS to obtain the initial estimate  $\hat{\underline{\theta}}^{(0)}$ . Then  $\hat{\underline{\theta}}^{(0)}$  is substituted into the matrix  $V_W$  to calculate  $V_W(\hat{\underline{\theta}}^{(0)})$ , which is used in the  $w(d_i)$  to produce  $\hat{\underline{\theta}}^{(1)}$  and the resulting  $V_W(\hat{\underline{\theta}}^{(1)})$  from WLS. The above procedures are iterated until the parameter estimates converge within a specified tolerance. The final estimate of the parameters  $\hat{\underline{\theta}}$  are then substituted in the semivariogram model  $\gamma(d; \underline{\theta})$  yielding a valid semivariogram estimator,  $\gamma(d; \hat{\underline{\theta}})$ , for all further analyses.

## APPENDIX 2: Computation speeds of two decompositions

Table 2 gives a comparison of the two decompositions on different sizes of Gaussian covariance matrices. The calculations are performed using the R package under the Microsoft Windows XP operation system on a laptop computer with Pentium M 1.6GHz processor and 512MB SDRAM memory. The Cholesky decomposition runs about 10 times faster than the eigenvalue decomposition on the higher dimensions. The covariance matrices for spatial data usually have large dimensions. If the size of a two dimensional image is  $n \times n$ , the corresponding covariance matrix will be of size  $n^2 \times n^2$ . As  $n$  gets large, the covariance matrix quickly grows in size. A computationally efficient algorithm is desired to handle such a large number of matrix operations, and hence we use the Cholesky decomposition in this paper.

## ACKNOWLEDGEMENTS

This study was supported by the U.S. Army Medical Research and Materiel Command numbers DAMD17-97-2-7025 and DAMD17-01-1-0741 through a consortium agreement with the University of Texas Southwestern Medical Center at Dallas, U.S. Public Health Service grant M01-RR00633, and by a grant from the Perot Foundation, Dallas, Texas. The content

	$64 \times 64$	$256 \times 256$	$2304 \times 2304$	$4096 \times 4096$
Cholesky	0.01	0.2	68.08	438.58
Eigenvalue	0.03	0.72	626.17	4917.47

Table 2: Computation time in seconds for the Cholesky and eigenvalue decompositions. The first row lists the covariance matrix sizes. The decomposition methods are given on subsequent rows.

of this paper does not necessarily reflect the position or the policy of the U.S. government, and no official endorsement should be inferred.

## REFERENCES

- Carlstein, E. (1986). “The Use of Subseries Methods for Estimating the Variance of a General Statistic from a Stationary Time Series.” *Annals of Statistics*, **14**, 1171-1179.
- Cherry, S., Banfield, J., and Quimby, W. (1996). “An Evaluation of a Nonparametric Method of Estimating Semivariograms of Isotropic Spatial Processes.” *Journal of Applied Statistics*, **23**, 435-449.
- Cressie, N. (1985). “Fitting Variogram Models by Weighted Least Squares.” *Mathematical Geology*, **17**, 563-585.
- Cressie, N. (1993). *Statistics for Spatial Data* (revised ed.), New York: John Wiley and Sons.
- Cressie, N. and GM Laslett (1987). “G. M. Random Set Theory and Problems of Modeling.” *SIAM Review*, **29**, 557-574.
- Cressie, N. and Grondona, M. (1992). “A Comparison of Variogram Estimation with Covariogram Estimation.” *The Art of Statistical Science*, K.V. Mardia (ed.), Chinchester: Wiley, 191-208.
- Genton, M. and Gorsich, D. (2002). “Nonparametric Variogram and Covariogram Estimation with Fourier-Bessel Matrices.” *Computational Statistics & Data Analysis*, **41**, 47-57.
- Grondona, M. O. and Cressie, N. (1995). “Residuals Based Estimators of the Covariogram.” *Statistics*, **26**, 209-218.
- Hall, P. (1985). “Resampling A Coverage Pattern.” *Stochastic Processes and their Applications*, **20**, 231-246.
- Hall, P., Fisher, N., and Hoffmann, B. (1994). “On the Nonparametric Estimation of Covariance Functions.” *Annals of Statistics*, **22**, 2115-2134.
- Hall, P., Horowitz, J. L., and Jing, B. (1995). “On Blocking Rules for the Block Bootstrap with Dependent Data.” *Biometrika*, **82**, 561-574.
- Kunsch, H. R. (1989). “The Jackknife and the Bootstrap for General Stationary Observations.” *Annals of Statistics*, **17**, 1217-1261.
- Liu, R. Y. and Singh, K. (1992). “Moving Blocks Jackknife and Bootstrap Capture Weak Dependence.” *In Exploring the Limits of Bootstrap* (R. Lepage and L. Billard, eds.) 225-248. John Wiley, New York.

- Matheron, G. (1962). "Traite de Geostatistique Appliquee, Tome I." *Memoires du Bureau de Recherches Geologiques et Minieres*, No. 14 Editions Technip, Paris.
- Matheron, G. (1971). *The theory of regionalized variables and its applications*. Les Cahiers du Centre de Morphologie Mathématique, Fasc. 5, Centre de Géostatistique, Fontainebleau.
- Politis, D. and Romano, J. (1992). "A Circular Block Resampling Procedure for Stationary Data." In *Exploring the Limits of Bootstrap* (R. Lepage and L. Billard, eds.) 263-270. Wiley, New York.
- Politis, D. and Romano, J. (1994). "Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions." *Annals of Statistics*, **22**, 2031-2050.
- Sjöstedt-de Luna, S. and Young, A. (2003). "The Bootstrap and Kriging Prediction Intervals." *Scandinavian Journal of Statistics*, **30**, 175-192.
- Solow, A. (1985). "Bootstrapping Correlated Data." *Journal of the International Association for Mathematical Geology*, **17**, 769-775.
- Shapiro, A. and Botha, J. (1991). "Variogram Fitting with a General Class of Conditionally Nonnegative Definite Functions." *Computational Statistics & Data Analysis*, **11**, 87-96.
- Spence, J. (2004). "A Spatial Analysis of SPECT Brain Imaging Data: Optimal Predictions in Regions of Interest." PhD Dissertation, Southern Methodist University.
- Stein, M. (1988). "Asymptotically Efficient Prediction of a Random Field with a Misspecified Covariance Function." *Annals of Statistics*, **16**, 55-63.