Inconsistent Estimation and Asymptotically Equal Interpolations in Model-Based Geostatistics

Hao ZHANG

It is shown that in model-based geostatistics, not all parameters in the Matérn class can be estimated consistently if data are observed in an increasing density in a fixed domain, regardless of the estimation methods used. Nevertheless, one quantity can be estimated consistently by the maximum likelihood method, and this quantity is more important to spatial interpolation. The results are established by using the properties of equivalence and orthogonality of probability measures. Some sufficient conditions are provided for both Gaussian and non-Gaussian equivalent measures, and necessary conditions are provided for Gaussian equivalent measures. Two simulation studies are presented that show that the fixed-domain asymptotic properties can explain some finite-sample behavior of both interpolation and estimation when the sample size is moderately large.

KEY WORDS: Equivalent measures; Generalized linear mixed model; Kriging; Matérn class; Minimum mean squared error; Model-based geostatistics; Prediction.

1. INTRODUCTION

Geostatistics is a field of statistics concerned with spatial variation in a continuous spatial region. It has its origins in problems connected with estimation of ore reserves in mining (Krige 1951) and has found applications in many other areas, including hydrology, agriculture, natural resource evaluation, and environmental sciences. (See Cressie 1993 and Chilés and Delfiner 1999 for an introduction to geostatistics.) In many geostatistical problems, interpolation is the ultimate objective. A class of linear interpolation methods commonly called "kriging" has been developed. Stein (1999) has provided a rigorous account of the mathematical theory underlying linear kriging. However, in many applications, interpolations are made when spatial counts are observed. Gotway and Stroup (1997), Diggle, Tawn, and Moyeed (1998), and Zhang (2002) provided real examples of interpolation given spatial counts. These spatial counts are generally related to binomial sample sizes or lengths of time during which the counts are collected. Although this information should be incorporated into prediction, linear prediction generally cannot do this. Diggle et al. (1998) considered model-based geostatistics that use explicit parametric stochastic models and likelihood-based inferences. This approach effectively incorporates sample sizes into the binomial models, for example, and allows for calculation of minimum mean squared error (MMSE) prediction.

In model-based geostatistics, spatial generalized linear mixed models (GLMM's) are used to model both Gaussian and non-Gaussian variables, such as spatial counts. Although distributional assumptions are not needed for linear interpolation, it becomes possible to study asymptotic properties of estimation under distributional assumptions, and these asymptotic properties are useful for explaining finite-sample behaviors of estimators and interpolators. For example, for a one-dimensional Gaussian process with an exponential covariogram $\sigma^2\exp(-\alpha h)$, Ying (1991) pointed out that neither of the two parameters σ^2 or α can be estimated consistently given that the process is observed in the unit interval, but showed that the maximum likelihood estimator (MLE) of the product $\sigma^2\alpha$ is strongly consistent under the infill asymptotics. Using equivalence of Gaussian measures as a tool, Stein (1990, thm. 3.1) showed that an incorrect covariogram that is compatible with the correct covariogram yields asymptotically optimal interpolation relative to the predictions based on the correct covariogram. Because two exponential covariograms are compatible if they have the same product $\sigma^2\alpha$, this product matters more to interpolation than do the individual parameters.

There are two distinct asymptotics in spatial statistics: increasing domain asymptotics, where more data are collected by increasing the domain, and fixed-domain or infill asymptotics, where more data are collected by sampling more densely in a fixed domain. Asymptotic properties of estimators are quite different under the two asymptotics. For example, both the variance σ^2 and the scale parameter α in the exponential covariogram can be estimated consistently under the increasing domain asymptotics (Mardia and Marshall 1984), whereas such consistent estimators do not exist under infill asymptotics. Which asymptotics to use, or even whether any asymptotics is valuable to a given problem may be disputable, because only a finite number of spatial locations are encountered. Here we adopt Stein's position that we use asymptotics not because we actually plan to take more and more observations by increasing the domain or sampling more densely in a fixed-domain, but rather because we hope that the asymptotic results obtained will be useful for the specific problem at hand (Stein 1999, sec. 3.3, p. 62). Simulation studies can reveal how appropriate the asymptotic results are in a specific finite-sample setting. All asymptotic statements in this article are restricted to the fixed-domain asymptotics.

We consider a wide class of covariance functions, the Matérn class, that has received more attention in recent years because of its capacity to model the variogram's behavior near the origin. It consists of exponential variograms as a special case. Unlike other popular covariograms, such as exponential, powered-exponential, or spherical covariograms, the Matérn class has a parameter that controls the smoothness of the process. For this reason, Stein (1999) strongly recommended using the Matérn class to model spatial correlations. The Matérn class has also been used by Handcock and Stein (1993), Handcock and Wallis (1994), Williams, Santner, and Notz (2000), and Diggle, Ribeiro, and Christensen (2002).

Hao Zhang is Associate Professor, Department of Statistics, Washington State University, Pullman, WA 99164-3144 (E-mail: *zhanghao@wsu.edu*). The author thanks the editor, the associate editor, four referees, and Jave Pascual for helpful comments, and Michael Stein for communication on Theorem 2.

Journal of the American Statistical Association
March 2004, Vol. 99, No. 465, Theory and Methods
DOI 10.1198/016214504000000241

I show in this article that in model-based geostatistics with Gaussian or non-Gaussian observations, one cannot correctly distinguish between two Matérn covariograms with probability 1 no matter how many sample data are observed in a fixed region. Consequently, not all covariogram parameters are consistently estimable. This might suggest that the covariogram may very well be incorrectly estimated, and explains why estimates of covariograms usually have large variations. However, as I show later, an incorrect covariogram may (but does not always) yield asymptotically equal predictions in model-based geostatistics. I also study the quantity that is more important to interplolation than any individual parameters, and establish the strong consistency of the MLE of this quantity. My results may also partially explain the difficulties in likelihood estimation of covariogram parameters reported in the literature (e.g., Warnes and Ripley 1987; Mardia and Watkins 1989; Diggle et al. 1998; Zhang 2002), and the ineffectiveness of cross-validating a variogram in model-based geostatistics (Zhang 2003).

The rest of the article is organized as follows. Section 2 reviews the Matérn class and the stochastic models in model-based geostatistics. Section 3 contains main theoretical results, showing that two Matérn variograms may define two equivalent probability measures. It also provides a new result about orthogonal Gaussian measures, from which I establish the strong consistency of the MLE of a quantity that is important to interpolation. Section 4 provides two simulation studies that show how well the fixed-domain asymptotic results apply to finite-sample cases. The final section provides a discussion and open problems for future research.

2. MODEL–BASED GEOSTATISTICS AND THE MATÉRN CLASS

In model-based geostatistics, spatial GLMM's are used to provide a unified approach to modeling Gaussian and non-Gaussian data. For example, the following spatial GLMM has been used to model spatial counts (see, e.g., Diggle et al. 1998; Heagerty and Lele 1998; Zhang 2002, 2003; Christensen and Waagepetersen 2002; Diggle et al. 2002; Zhang and Wang 2002):

1. Let $\{b(\mathbf{s}),\ \mathbf{s}\in\mathbb{R}^d\}$ be a second-order stationary Gaussian process with mean 0 such that $b(\mathbf{s})$ represents the local variation at site $\mathbf{s}$.
2. Conditional on $\{b(\mathbf{s}),\ \mathbf{s}\in\mathbb{R}^d\}$, the random variables $\{Y(\mathbf{s}), \mathbf{s}\in\mathbb{R}^d\}$ are mutually independent, and for any $\mathbf{s}$, $Y(\mathbf{s})$ follows a generalized linear model with a distribution specified by the value of the conditional mean $\mu(\mathbf{s}) = E(Y(\mathbf{s})|b(\mathbf{s}))$. For some link function g, $g(\mu(\mathbf{s})) = \beta + b(\mathbf{s}) + \sum_{i=1}^{p} x_i(\mathbf{s})\beta_i$, where $x_i(\mathbf{s})$ is the value of the ith explanatory variable at location $\mathbf{s}$, $i = 1,\dots,p$.

Note that this model excludes Gaussian models by requiring that the distribution of $Y(\mathbf{s})$ given $b(\mathbf{s})$ depends only on $E(Y(\mathbf{s})|b(\mathbf{s}))$. It can be extended to include the following Gaussian model with measurement error:

$$Y(\mathbf{s}) = \mu + \epsilon(\mathbf{s}) + b(\mathbf{s}),$$

where $\epsilon(\mathbf{s})$ is an iid Gaussian process with mean 0, $b(\mathbf{s})$ is a stationary Gaussian process with mean 0, and the two processes are mutually independent.

For simplification, I consider the spatial GLMM with no explanatory variables. This is a particularly interesting case for interpolations because no covariates need to be observed for interpolation. The distribution of the Gaussian process $b(\mathbf{s})$ is determined solely by its covariance function, or covariogram, that is often assumed to have a parametric form and depends on some vector of parameters $\boldsymbol{\theta}$. The model parameters are then β and $\boldsymbol{\theta}$. Given observations of $Y(\mathbf{s})$ at sampling locations $\mathbf{s}_1,\dots,\mathbf{s}_n$, model parameters can be estimated using maximum likelihood (ML) techniques (Zhang 2002) or a Bayesian approach (Diggle et al. 1998). Using the estimates as the true values, the plug-in MMSE prediction of $Y(\mathbf{s})$ at an unsampled location is $E\{Y(\mathbf{s})|Y(\mathbf{s}_i), i = 1,\dots,n\}$, where the expectation is evaluated under the estimates of parameters.

The covariogram of the Gaussian process $b(\mathbf{s})$ and the parameter β completely determine the probability distribution of the process $Y(\mathbf{s})$ in the spatial GLMM. One of the important classes of isotropic covariograms is the Matérn class, defined as

$$K(x;\ \sigma^2,\alpha,\nu) = \frac{\sigma^2(\alpha x)^\nu}{\Gamma(\nu)2^{\nu-1}}K_\nu(\alpha x), \qquad x \ge 0, \tag{1}$$

where σ^2, $\alpha > 0$, and $\nu > 0$ are parameters and K_ν is the modified Bessel function of order ν. (See Abramowitz and Stegun 1967, pp. 375–376, for the definition and properties of the modified Bessel function.) Because $K_\nu(x)x^\nu \to 2^{\nu-1}\Gamma(\nu)$ as $x \to 0$, $K(0) = \sigma^2$ is the variance of the process. When $\nu = 1/2$, the Matérn covariogram becomes the exponential one, $K(x) = \sigma^2\exp(-\alpha x)$. Hereafter, I call ν the smoothness parameter and call α the scale parameter.

A process having the Matérn covariogram (1) is $[\nu]-1$ times mean square differentiable, where $[\nu]$ is the largest integer less than or equal to ν. Other classes of covariograms do not have such a parameter to yield a preferred mean square differentiability.

When a stationary process is isotropic, the isotropic spectral density is often used instead of the second-order spectral density. Recall that the second-order spectral density $f^*(\lambda)$ of an isotropic process depends only on the module of λ, and the function $f(|\lambda|) = f^*(\lambda)$ is called the isotropic spectral density. For the Matérn covariogram (1) in $\mathbb{R}^d$, the corresponding isotropic spectral density is (see, e.g., Stein 1990, pp. 48–49)

$$f(u) = \frac{\sigma^2\alpha^{2\nu}}{\pi^{d/2}(\alpha^2+u^2)^{\nu+d/2}}, \qquad u \ge 0. \tag{2}$$

This functional form of the spectral density is used in the proof of Theorem 2 in the next section.

3. EQUIVALENCE OF PROBABILITY MEASURES AND MAIN RESULTS

I first review the concept of equivalence of probability measures and its applications to statistical references. Recall that for two probability measures P_i, $i = 1, 2$, defined on the same measurable space $(\Omega,\mathcal{F})$, P_1 is said to be absolutely continuous with respect to P_2, denoted by $P_1 \ll P_2$, if $P_1(A) = 0$ for any $A\in\mathcal{F}$ such that $P_2(A) = 0$. P_1 and P_2 are equivalent, denoted by $P_1 \equiv P_2$, if $P_1 \ll P_2$ and $P_2 \ll P_1$. If $P_1 \equiv P_2$ on $\mathcal{F}$ and $\mathcal{F}$ is the σ-algebra generated by a stochastic process $Y(\mathbf{s})$, $\mathbf{s}\in T$ for any set T, then P_i, $i = 1, 2$, are said to be equivalent on the paths of $Y(\mathbf{s})$, $\mathbf{s} \in T$. Obviously, if two measures are equivalent on $\mathcal{F}$, then they must be also equivalent on any σ-algebra $\mathcal{F}_0 \subset \mathcal{F}$.

The equivalence of probability measures has two major applications to statistical references. First, if $P_1 \equiv P_2$, then P_1 cannot be correctly distinguished from P_2 with P_1-probability 1 regardless of what is observed. Moreover, if $\{P_\theta, \theta \in \Theta\}$ is a family of equivalent measures and $\hat{\theta}_n, n \geq 1$ is a sequence of estimators, then, irrespective of what is observed, $\hat{\theta}_n$ cannot be weakly consistent estimators of θ for all $\theta \in \Theta$. Otherwise, for any fixed $\theta \in \Theta$ there exists a strongly consistent subsequence $\{\theta_{n_k},\ k \geq 1\}$—that is, $P_\theta(\hat{\theta}_{n_k} \to \theta, k \to \infty) = 1$ (see, e.g., Dudley 1989, thm. 9.2.1, p. 226). For any $\theta' \in \Theta$ such that $\theta' \neq \theta$, it follows from the equivalence of the two measures P_θ and $P_{\theta'}$ that $P_{\theta'}(\hat{\theta}_{n_k} \to \theta, k \to \infty) = 1$. On the other hand, the weak consistency of the subsequence $\{\theta_{n_k},\ k \geq 1\}$ under the probability measure $P_{\theta'}$ implies the existence of a sub-subsequence that converges to θ' with $P_{\theta'}$-probability 1. This sub-subsequence converges to two different values under the same measure $P_{\theta'}$. This apparent contradiction shows that $\hat{\theta}_n$ cannot be weakly consistent.

The second application is on prediction. Its theoretical foundation is the theorem of Blackwell and Dubins (1962). I now rephrase the theorem to make it directly applicable to model-based geostatistics:

Let Y_i, $i \geq 1$, be random variables on a measurable space $(\Omega, \mathcal{F})$ and P_i, $i = 1, 2$, be two probability measures on $\mathcal{F}$ such that $P_1 \ll P_2$ constrained on $\sigma(Y_i, i \geq 1)$, the σ-algebra generated by Y_i, $i \geq 1$. Then with P_1-probability 1,

$$\sup \left| P_1(A|Y_1, \ldots, Y_n) - P_2(A|Y_1, \ldots, Y_n) \right| \to 0 \quad \text{as } n \to \infty,$$

where the supremum is taken over all $A \in \sigma(Y_i, i > n)$. In particular,

$$\sup_{i>n, B} \left| P_1(Y_i \in B|Y_1, \ldots, Y_n) - P_2(Y_i \in B|Y_1, \ldots, Y_n) \right| \to 0 \quad \text{as } n \to \infty. \tag{3}$$

This says that given $Y_1, \ldots, Y_n$, predictions for Y_i, $i > n$, under both measures tend to agree as $n \to \infty$. Note that constrained on $\sigma(Y_i, i \geq 1)$, the probability measures are predictive, as defined by Blackwell and Dubins (1962), and therefore their main theorem applies.

The concept of equivalence of measures is more complex than the definition might suggest, particularly when an infinite stochastic sequence is involved. Apparently, any two nonsingular Gaussian measures in a finite-dimensional Euclidian space are equivalent. However, they may be orthogonal in an infinite space. For example, if $Y_1, Y_2, \ldots$ are iid $\mathrm{N}(0, \sigma_i^2)$ under P_i, $i = 1, 2$, with $\sigma_1^2 \neq \sigma_2^2$, then on the paths of the infinite sequence $Y_i, n \geq 1$, the two measures are orthogonal, because if

$$A = \left\{ (1/n) \sum_{i=1}^{n} Y_i^2 \to \sigma_1^2 \text{ as } n \to \infty \right\},$$

then $P_1(A) = 1$ and $P_2(A) = 0$ by the law of large numbers. For a correlated process, comprehending the equivalence of probability measures becomes less intuitive. Consider, for example, a stationary isotropic Gaussian random process $Y(\mathbf{s})$, $\mathbf{s} \in \mathbb{R}^d$, with mean 0 and an isotropic covariogram $K(h) = \sigma_i^2 \exp(-h/\theta_i)$, $h > 0$, under measures P_i, $i = 1, 2$. Then $P_1 \equiv P_2$ on the paths of $\{Y(\mathbf{s}), \mathbf{s} \in T\}$ for any bounded subset T of $\mathbb{R}^d$ if $\sigma_1^2/\theta_1 = \sigma_2^2/\theta_2$ (see, e.g., Stein 1999, p. 120, for $d = 1$ and Stein 2004, thm. A.1, for $d > 1$). If T is finite and bounded, then $\sigma_1^2/\theta_1 \neq \sigma_2^2/\theta_2$ implies that the two measures are orthogonal, as implied by Theorem 2. Hence this ratio can be well estimated given sufficient data from a bounded region, as seen from Theorem 3.

I now state the main theorem on the equivalence of probability measures defined through the spatial GLMM. For convenience, I assume that both $\{b(\mathbf{s}), \mathbf{s} \in T\}$ and $\{Y(\mathbf{s}), \mathbf{s} \in T\}$ are defined on some probability space $(\Omega, \mathcal{F})$ and that $P_{\beta,\boldsymbol{\theta}}$ is a probability measure indexed by the parameters β and $\boldsymbol{\theta}$, where $\boldsymbol{\theta} = (\sigma^2, \alpha, \nu)$ consists of the covariogram parameters, such that under each $P_{\beta,\boldsymbol{\theta}}$, $\{b(\mathbf{s}), \mathbf{s} \in T\}$ is a mean-0 Gaussian process with a Matérn covariogram (1) with the parameter $\boldsymbol{\theta}$, and $\{Y(\mathbf{s}), \mathbf{s} \in T\}$ are independent conditional on $\{b(\mathbf{s}), \mathbf{s} \in T\}$, and the conditional distribution of $Y(\mathbf{s})$ depends only on the parameter β and not on $\boldsymbol{\theta}$. Note that the construction of the probability measures includes the spatial GLMM and the Gaussian model with measurement error.

Theorem 1. Let T be a bounded subset of $\mathbb{R}^d$ for some integer $d > 0$, and the processes $\{Y(\mathbf{s}), \mathbf{s} \in T\}$, $\{b(\mathbf{s}), \mathbf{s} \in T\}$ and the measure $P_{\beta,\boldsymbol{\theta}}$ be the same as previously defined. For any β, $\boldsymbol{\theta}_1$, and $\boldsymbol{\theta}_2$, $P_{\beta,\boldsymbol{\theta}_1} \equiv P_{\beta,\boldsymbol{\theta}_2}$ on the paths of $Y(\mathbf{s})$, $\mathbf{s} \in T$, if $P_{\beta,\boldsymbol{\theta}_1} \equiv P_{\beta,\boldsymbol{\theta}_2}$ on the paths of $b(\mathbf{s})$, $\mathbf{s} \in T$.

Proofs of this theorem and other theorems in this section are given in the Appendix.

Because $P_{\beta,\boldsymbol{\theta}}$ depends only on $\boldsymbol{\theta}$ when restricted to $\sigma(b(\mathbf{s}), \mathbf{s} \in T)$, we see from the theorem that if two covariograms define two equivalent Gaussian measures on $b(\mathbf{s})$, $\mathbf{s} \in T$, then the induced measures on $Y(\mathbf{s})$, $\mathbf{s} \in T$, are equivalent for any fixed β. Sufficient conditions exist for equivalent Gaussian measures that are expressed in terms of the second-order spectral densities (see Gihman and Skorohod 1974, thm. 3, p. 509, and Ibragimov and Rozanov 1978, thm. 17, chap. III, for $d = 1$ and Yadrenko 1983, p. 156, and Stein 1999, p. 120, for $d > 1$). Stein (2004, thm. A1) provided the following sufficient conditions for equivalence of two Gaussian measures, which are easy to verify for the Matérn class:

Let $P_i, i = 1, 2$, be two probability measures such that under P_i, the process $X(\mathbf{s})$, $\mathbf{s} \in \mathbb{R}^d$, is stationary Gaussian with mean 0 and a second-order spectral density $f_i(\mathbf{v})$, $\mathbf{v} \in \mathbb{R}^d$. If, for some $\alpha > 0$, $f_1^*(\mathbf{v})|\mathbf{v}|^\alpha$ is bounded away from 0 and ∞ as $|\mathbf{v}| \to \infty$, and for some finite c,

$$\int_{|\mathbf{v}|>c} \left\{ \frac{f_2^*(\mathbf{v}) - f_1^*(\mathbf{v})}{f_1^*(\mathbf{v})} \right\}^2 d\mathbf{v} < \infty, \tag{4}$$

then $P_1 \equiv P_2$ on the paths of $X(\mathbf{s})$, $\mathbf{s} \in T$, for any bounded subset $T \subset \mathbb{R}^d$.

Condition (4) can be expressed in terms of the isotropic spectral densities $f_i(u)$, $i = 1, 2$:

$$\int_c^\infty u^{d-1} \left\{ \frac{f_2(u) - f_1(u)}{f_1(u)} \right\}^2 du < \infty. \tag{5}$$

Theorem 2. Let P_i, $i = 1, 2$, be two probability measures such that under P_i, the process $X(\mathbf{s})$, $\mathbf{s} \in \mathbb{R}^d$, is stationary Gaussian with mean 0 and an isotropic Matérn covariogram in $\mathbb{R}^d$ with a variance σ_i^2, a scale parameter α_i, $i = 1, 2$, and the same smoothness parameter ν, where $d = 1, 2$ or 3. For any bounded infinite set $T \subset \mathbb{R}^d$, $P_1 \equiv P_2$ on the paths of $X(\mathbf{s})$, $\mathbf{s} \in T$ if and only if $\sigma_1^2 \alpha_1^{2\nu} = \sigma_2^2 \alpha_2^{2\nu}$.

An immediate corollary is that the following exponential covariograms are equivalent: $K_i(x) = \varphi \alpha_i^{-1} \exp(-\alpha_i x)$, $i = 1, 2$, where $\varphi > 0$ is a constant. Theorem 2 has several applications. I first state the following obvious corollaries about parameter estimation and prediction.

Corollary 1. Let $Y(\mathbf{s})$, $\mathbf{s} \in T$, follow the spatial GLMM with the random effects having a Matérn covariogram, where T is a bounded subset of $\mathbb{R}^d$. Also let D_n, $n \geq 1$, be an increasing sequence of subsets of T. Given observations of $Y(\mathbf{s})$ for $\mathbf{s} \in D_n$, there do not exist estimators σ_n^2 and α_n that are weakly consistent—that is, for any β and $\boldsymbol{\theta} = (\sigma^2, \alpha, \nu)$, $P_{\beta,\boldsymbol{\theta}}(|\sigma_n^2 - \sigma^2| > \epsilon) \to 0$ or $P_{\beta,\boldsymbol{\theta}}(|\alpha_n - \alpha| > \epsilon) \to 0$, $n \to \infty$, for any $\epsilon > 0$.

Handcock and Wallis (1994) recommended an alternative reparameterization of the Matérn covariogram (1) by letting $\rho = 2\nu^{1/2}/\alpha$. Clearly, ρ cannot be estimated consistently. In fact, any parameterization cannot make all parameters consistently estimable, although it is possible that reparameterizing could enable consistent estimation of one of the new parameters. For example, write $c = \sigma^2\alpha^{2\nu}$, and reparameterize by using c and α. This c can be estimated consistently by Theorem 3. However, α still cannot be estimated consistently. Otherwise, if both c and α can be estimated consistently, then $\sigma^2 = c\alpha^{-2\nu}$ can be estimated consistently.

Corollary 2. Let $Y(\mathbf{s})$, $\mathbf{s} \in T$, follow the spatial GLMM with the random effects having a Matérn covariogram. Write $\boldsymbol{\theta}_i = (\sigma_i^2, \alpha_i, \nu)$ for some $\nu > 0, \sigma_i^2 > 0$, and $\alpha_i > 0, i = 1, 2$, such that $\sigma_1^2\alpha_1^{2\nu} = \sigma_2^2\alpha_2^{2\nu}$. Let $\mathbf{s}_i$, $i = 1, 2, \ldots$, be locations in a bounded domain T. Then for any β,

$$\sup\big|P_{\beta,\boldsymbol{\theta}_1}\big(A|Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)\big) - P_{\beta,\boldsymbol{\theta}_2}(A|Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n))\big| \to 0 \quad \text{as } n \to \infty, \tag{6}$$

where the supremum is taken over all $A \in \sigma(Y(\mathbf{s}_i), i > n)$.

This corollary implies that given $Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)$, the distributions of $Y(\mathbf{s}_{n+k})$ for any $k > 0$ are asymptotically equal under equivalent measures. When $Y(\mathbf{s}_{n+k})$ takes a finite number of values like the binomial variables, then, for any function ϕ,

$$\sup_k\big|E_{\beta,\boldsymbol{\theta}_1}\{\phi(Y(\mathbf{s}_{n+k}))|\mathbf{Y}\} - E_{\beta,\boldsymbol{\theta}_2}\{\phi(Y(\mathbf{s}_{n+k}))|\mathbf{Y}\}\big| \to 0 \quad \text{as } n \to \infty. \tag{7}$$

It is often of interest to predict a function of $b(\mathbf{s})$ at a site $\mathbf{s}$ such as $p(\mathbf{s}) = \exp(\beta + b(\mathbf{s}))/(1 + \exp(\beta + b(\mathbf{s})))$ for the logistic model. It has been shown that if $\psi(b(\mathbf{s})) = E\{\phi(Y(\mathbf{s}))|b(\mathbf{s})\}$ for some function ϕ, then $E\{\phi(Y(\mathbf{s}))|\mathbf{Y}\} = E\{\psi(b(\mathbf{s}))|\mathbf{Y}\}$ (Zhang 2003). Therefore, predictions of such a function $\psi(b(\mathbf{s}))$ will be asymptotically equal under two equivalent measures. $p(\mathbf{s})$ is clearly such a function, because $p(\mathbf{s}) = E\{Y(\mathbf{s})/n(\mathbf{s})|b(\mathbf{s})\}$ if $Y(\mathbf{s})$ follows the spatial GLMM with the logit link function and $n(\mathbf{s})$ is the binomial sample size. For a logistic model, in many situations predicting $p(\mathbf{s})$ is more interesting than predicting other functions of $b(\mathbf{s})$. In the second simulation study in the next section, I argue heuristically that prediction variances of $p(\mathbf{s})$ are also asymptotically equal under two equivalent measures.

Several authors have commented on the usefulness of cross-validating a fitted variogram (Davis 1987; Cressie 1993, p. 104; Stein 1999, sec. 6.9). In general, it is considered a method of model checking to prevent blunders and to highlight potentially troublesome prediction points; it is not a foolproof method for detecting problems with the fitted spatial model. From Corollary 2, cross-validation clearly cannot effectively detect an incorrect covariogram if the incorrect covariogram defines a measure equivalent to the one defined by the correct covariogram.

Corollary 2 states that an incorrect covariogram may yield similar interpolation results as the correct covariogram, provided that a sufficiently large number of locations are observed in a fixed domain. This is true only when the two covariograms define equivalent probability measures, however. For the Matérn class, this equivalence translates into the property that the two covariograms have the same quantity $\sigma^2\alpha^{2\nu}$. Hence it is this product and not the individual parameters that matters more to interpolation. Next, I show that for a Gaussian process with a Matérn covariogram (1) with a known ν, the quantity $\sigma^2\alpha^{2\nu}$ can be estimated consistently. For an exponential covariogram (i.e., $\alpha = 1/2$), Ying (1991) considered strong consistency of this quantity in one-dimensional space.

Theorem 3. Let the underlying process $\{X(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d\}$, $d = 1, 2$, or 3, be second order stationary Gaussian with mean 0 and possess an isotropic Matérn covariogram (1) with the unknown parameter values σ_0^2, α_0 and a known ν. Let D_n, $n = 1, 2, \ldots$, be an increasing sequence of finite subsets of $\mathbb{R}^d$ such that $\bigcup_{n=1}^{\infty} D_n$ is bounded and infinite, and $L_n(\sigma^2, \alpha)$ be the likelihood function when the process is observed at locations in D_n. For any fixed $\alpha_1 > 0$, let $\hat{\sigma}_n^2$ maximize $L_n(\sigma^2, \alpha_1)$. Then $\hat{\sigma}_n^2\alpha_1^{2\nu} \to \sigma_0^2\alpha_0^{2\nu}$, with P_0 probability 1, where P_0 is the Gaussian measure defined by the Matérn covariogram corresponding to parameter values σ_0^2, α_0, and ν.

4. NUMERICAL RESULTS

Asymptotic results are meant to help for inferences from finite samples. The applicability of asymptotic results to a finite-sample case in spatial statistics is complicated by the fact that there are two distinct asymptotics and the results are quite different under the two asymptotics, as mentioned earlier. Hence it is interesting to see which asymptotics, if any, is helpful in finite-sample cases. The simulation studies given in this section are done for this purpose, with an emphasis on examining the fixed-domain asymptotics. In particular, I intend to discover the practical implications of the consistent and inconsistent estimation discussed in the previous section. I use ML estimation in the simulation because asymptotic properties of ML estimation are available under both asymptotics.

Example 1. I simulate a Gaussian process on some sampling locations with mean 0 and an exponential covariogram

$$K_0(x) = \sigma_0^2 \exp(-x/\theta_0), \qquad x \geq 0,$$

where $\sigma_0^2 > 0$ and $\theta_0 > 0$ are known values. In the simulations, σ_0 is fixed at 1 and θ_0 takes values .1, .2, and .3. I show later that to use different values for σ_0^2 is not necessary. For each set of the parameters, I simulate 1,000 independent realizations of the Gaussian process with mean 0 and the exponential covariogram at each of the three sets of locations. Set 1 comprises $(i/10, j/10), i, j = 0, 1, \ldots, 10$, and four more locations $(x, y), x, y = .05, .15$; set 2 has 221 locations and is the union of set 1 and $\{(.05 + .1i, .05 + .1j), i, j = 0, \ldots, 9\}$; set 3 contains 289 locations, as shown in Figure 1, including all of the locations of set 2 and 68 additional locations. I let the true

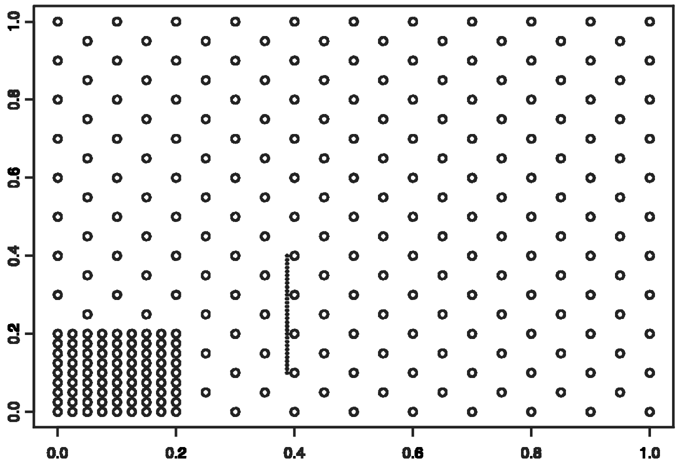

Figure 1. Sampling Locations in the Simulations (○) and Predicted Locations (· · · ·).

θ value and sample size vary, so that it can be seen how the estimates of different parameters change accordingly. It is a general belief that including some closely spaced locations leads to more efficient estimation of covariogram parameters (Stein 1999, p. 197). This is the reason why I included some closely spaced locations in sets 1 and 3.

For each dataset, I fit the following exponential covariogram by the ML method,

$$K(h;\sigma^2,\theta)=\sigma^2\exp(-h/\theta), \qquad h\geq 0. \tag{8}$$

The loglikelihood is, apart from an additive constant,

$$L(\sigma^2,\theta)=-(1/2)\log\{\det(V(\sigma^2,\theta))\}-(1/2)\mathbf{X}'V^{-1}(\sigma^2,\theta)\mathbf{X},$$

where $\mathbf{X}$ is the vector of simulated normal variables and $V^{-1}(\sigma^2,\theta)$ is the inverse of $V(\sigma^2,\theta)$, the covariance matrix of $\mathbf{X}$ corresponding to parameters σ^2 and θ.

I have mentioned that two orthogonal Gaussian measures can be distinguished correctly with probability 1 given an infinite sample, whereas two equivalent Gaussian measures cannot be. This property should be reflected in the behavior of the likelihood function for a large finite sample. For this purpose, Figure 2 plots the loglikelihood function $L(\sigma^2,\theta)$ along $\sigma^2/\theta=c$, where $c=5$ and 2 and θ ranges from .05 to 1. It also plots $L(\sigma^2,\theta)$ for σ^2 fixed at 1 and θ ranging from .05 to 1. The data are the first five simulations corresponding to sample size 289 and $\theta=.2$. When σ^2 is fixed, the log-likelihood $L(1,\theta)$ has a unique maximum around the true value, and decreases or increases sharply on either side of the maximum. Different behavior of $L(\sigma^2,\theta)$ is observed along the the curve $\sigma^2/\theta=c$, where it is quite flat on the right side of the maximum. This difference can be attributed to the difference between equivalence and orthogonality of probability measures, because different θ values define orthogonal Gaussian measures when σ^2 is fixed, whereas these different values define equivalent Gaussian measures along the curve $\sigma^2/\theta=c$. This does help explain some numerical results observed by others. For example, Warnes and Ripley (1987) described long and very flat ridges of the likelihood function, but did not relate them to the equivalence of probability measures. Other authors also have pointed out problems with finding the global maximum of the likelihood of spatial data (e.g., Ripley 1988; and Mardia and Watkins 1989). However, none associated the difficulties with the equivalence of probability measures.

Next, I found the MLE's for σ^2, θ, and σ^2/θ. I first used the Fisher-scoring method as used by Mardia and Marshall (1984) and Zimmerman and Zimmerman (1991), but found that the algorithm converged very slowly and occasionally failed to converge. I then used the profile likelihood function, which for θ is defined as

$$\begin{aligned}PL(\theta)&=\sup_{\sigma}L(\sigma^2,\theta)\\&=-(n/2)\log\big(\mathbf{X}'\boldsymbol{\Gamma}(\theta)^{-1}\mathbf{X}/n\big)\\&\quad-(1/2)\log(|\boldsymbol{\Gamma}(\theta)|)-n/2,\end{aligned}$$

where $\boldsymbol{\Gamma}(\theta)$ is the correlation matrix of $\mathbf{X}$ corresponding to θ and $\boldsymbol{\Gamma}^{-1}(\theta)$ is the inverse of $\boldsymbol{\Gamma}(\theta)$. (The correlation matrix

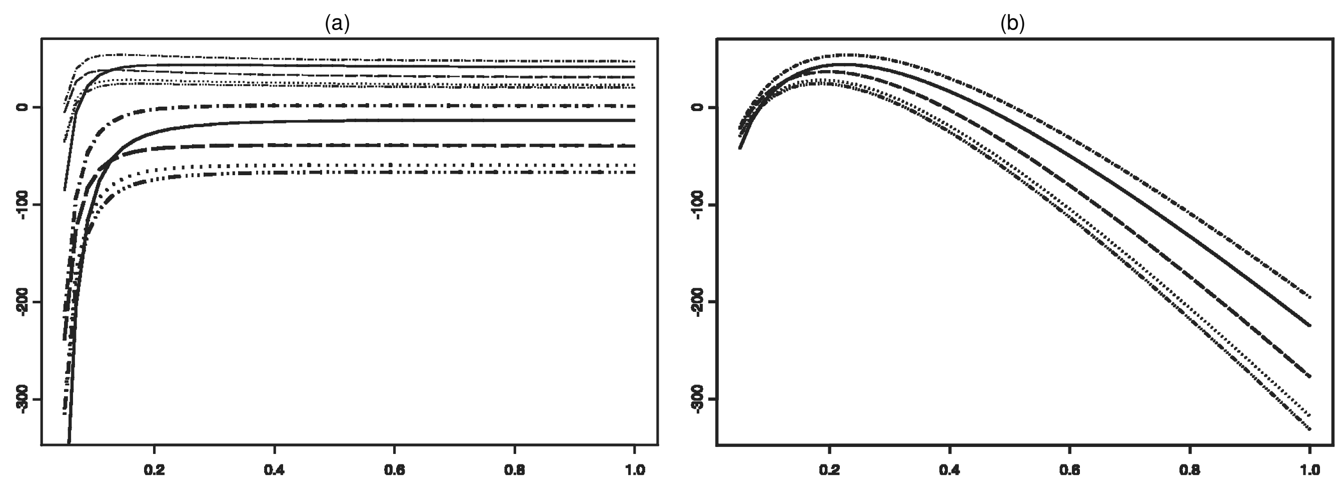

Figure 2. Log-Likelihood Function $L(\sigma^2,\theta)$ (a) on the Set $\sigma^2/\theta=c$ for $c=5$ (—— dataset 1; ······· dataset 2; ·-·-·-·- dataset 3; – – – dataset 4; ·-··-··- dataset 5) and $c=2$ (—— dataset 1; ······ dataset 2; ·-·-·- dataset 3; – – – dataset 4; -·- dataset 5) and (b) When σ^2 Is Fixed at 1 (—— dataset 1; ········ dataset 2; ·-·-·-·- dataset 3; – – – dataset 4; ·-··-·· dataset 5).

depends only on θ.) Maximizing PL(θ) through the Newton–Raphson algorithm yields the MLE $\hat{\theta}_n$. The MLE for σ^2 is $\hat{\sigma}_n^2 = (1/n)\mathbf{X}'\mathbf{\Gamma}^{-1}(\hat{\theta}_n)\mathbf{X}$. Nonconvergence never occurred for this algorithm.

We note that if $\mathbf{Y} = c\mathbf{X}$ for some constant $c > 0$ (so that the two correlation matrices are the same but the variances differ), the two log profile likelihood functions for θ differ only by an additive constant. The estimators for θ are the same, and the estimator of variance of $\mathbf{Y}$ is c^2 times the estimate of variance of $\mathbf{X}$. For this reason, I fixed σ_0 at 1 in the simulations.

Histograms of the estimates for θ and σ^2 and the ratio σ^2/θ are shown in Figures 3, 4, and 5. Each figure comprises nine histograms of the estimates corresponding to nine different combinations of the sample size n and θ_0. Figures 3 and 4 show that increasing the sample size from $n = 125$ does not result in a significant decrease in the variance of the estimates of θ or σ^2 and/or improvement of symmetry of the distributions of these estimators, especially when the spatial correlation is stronger. In contrast, Figure 5 shows that the distribution of the estimator for the ratio σ^2/θ becomes more symmetric with a smaller variance as the sample size increases, particularly when the spatial correlation is stronger. This difference in a sense supports the fixed-domain asymptotic results; the MLE's for θ and σ^2 are not consistent and hence cannot be asymptotically normal, whereas the MLE for the ratio is consistent. This consistency likely indicates that the variance of the estimator will vanish as the sample size increases, and that the estimator may be asymptotically normal, although the asymptotic distribution is not given in this article.

Tables 1–3 summarize, for each sample size and θ_0 value, the estimates of θ, σ^2, and the ratio σ^2/θ by listing the percentiles, biases, and sample standard deviations. These tables provide a better way to show how the variances are influenced by sample size. Overall, the MLE's for all parameters have negligible biases. Zimmerman and Zimmerman (1991) noted some negative biases of the estimates for σ^2, but they used sample sizes of 16 and 36, much smaller than the ones in this work.

A larger value of θ corresponds to a stronger spatial correlation of data. When $\theta = .1$, the correlation coefficient decreases to about .05 at the lag distance .3, and therefore this case presents a very weak spatial correlation. Estimators in this case have more symmetric distributions than the corresponding ones in the cases of stronger spatial correlations. However, the sample size still does not influence the variances of the estimators for θ and σ^2 as much as it does those for the ratio.

The practical implication of these estimation results is that sampling more data in a fixed domain may not improve estimates of the parameters θ and σ^2 as much as the estimates of the ratio σ^2/θ. Indeed, a sample size of 125 seems large enough to yield reasonably good estimates for θ and σ^2, and a larger sample may result in only minor improvements to the estimation of these two parameters. Sampling more from a fixed domain seem to be always helpful for estimating the ratio, as evidenced in the biases and standard deviations in Table 3.

I now obtain interpolations using three different exponential covariograms that correspond to (σ^2, θ) =(1, .2), (2, .4), and (1.8, .4). The first set represents true parameter values, and the second set defines an equivalent Gaussian measure to

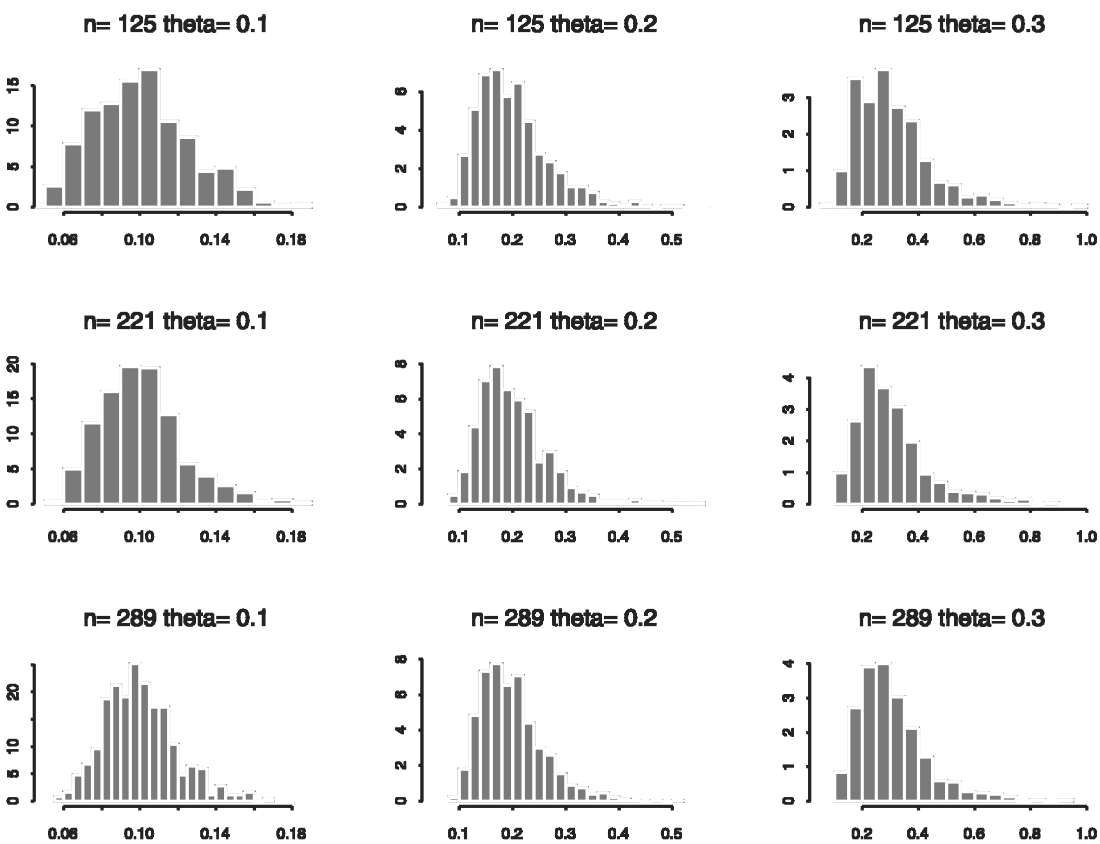

Figure 3. Histograms of Estimates of θ for Different Sample Sizes, 125 (top row), 221 (center row), and 289 (bottom row), and Different True θ Values, .1 (left column), .2 (center column), and .3 (right column).

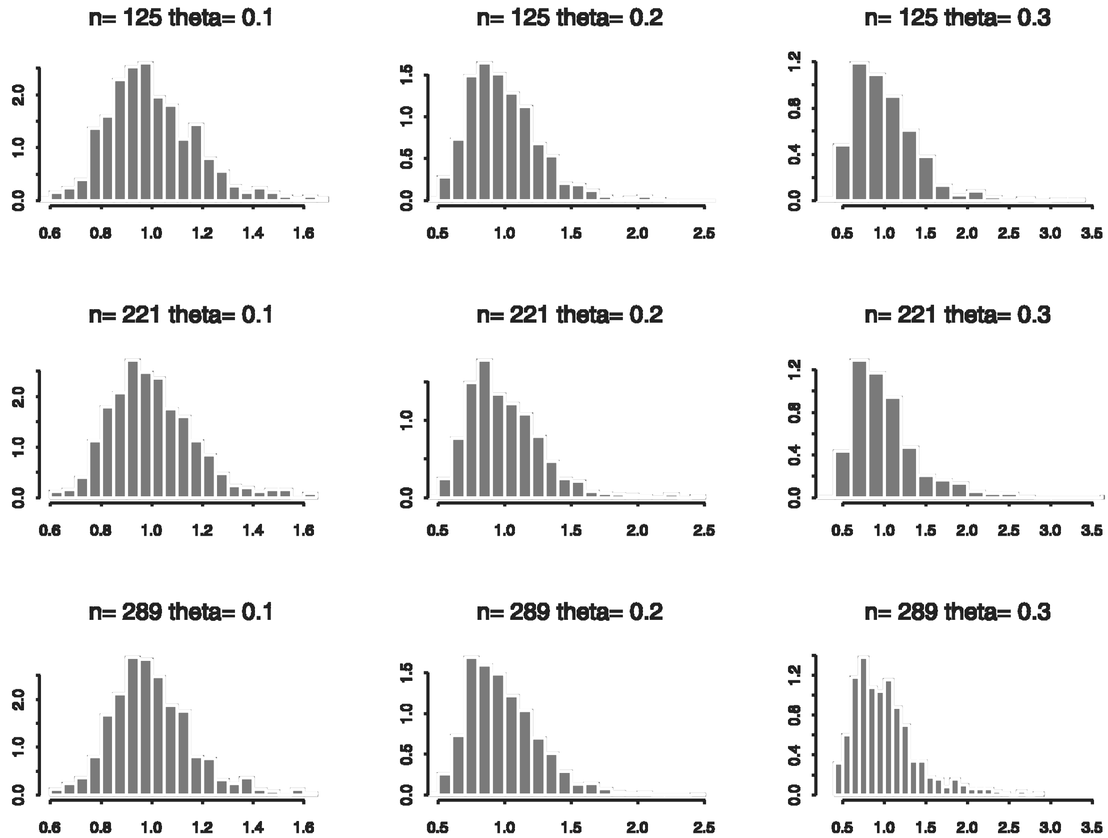

Figure 4. Histograms of Estimates of σ^2 for Different Sample Sizes, 125 (top row), 221 (center row), and 289 (bottom row), and Different True θ Values: .1 (left column), .2 (center column), and .3 (right column).

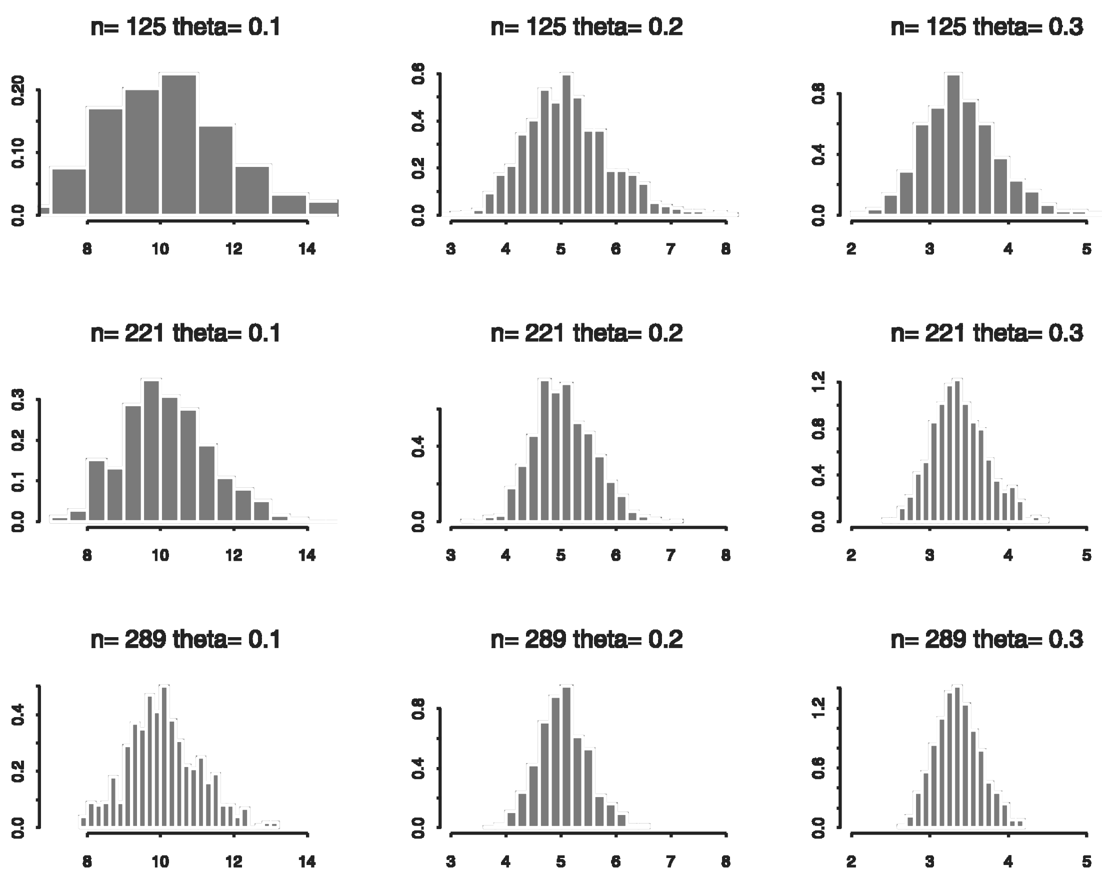

Figure 5. Histograms of Estimates of σ^2/θ for Different Sample Sizes, 125 (top row), 221 (center row), and 289 (bottom row), and Different True θ Values, .1 (left column), .2 (center column), and .3 (right column).

Table 1. Summary of Estimates of θ: Percentiles, Means, and Sample Standard Deviations (SD)

θ_0	n	5%	25%	50%	75%	95%	BIAS	SD
.1	125	.06298	.08156	.09877	.11548	.14671	.00062	.02507
	221	.06977	.08524	.09825	.11129	.14146	.00024	.02139
	289	.07272	.08712	.09870	.11113	.13253	.00001	.01868
.2	125	.11474	.15234	.18856	.23343	.32772	−.00029	.06629
	221	.12134	.15464	.18908	.23315	.31632	.00138	.06459
	289	.12293	.15651	.18881	.22819	.31050	−.00132	.05848
.3	125	.1465	.2077	.2820	.3744	.5651	.0072	.1334
	221	.1500	.2170	.2707	.3612	.5859	.0048	.1335
	289	.1569	.2167	.2815	.3589	.5371	.0041	.1212

Table 3. Summary of Estimates of σ^2/θ: Percentiles, Means, and Sample Standard Deviations (SD)

θ_0	n	5%	25%	50%	75%	95%	BIAS	SD
.1	125	7.443	8.931	10.078	11.357	13.887	.274	1.953
	221	8.259	9.277	10.107	10.900	12.304	.151	1.231
	289	8.5363	9.430	10.011	10.659	11.739	.077	.9747
.2	125	3.9332	4.5871	5.0616	5.5932	6.5001	.1204	.7849
	221	4.2139	4.6723	5.0092	5.4116	5.9900	.0481	.5421
	289	4.3105	4.7247	5.0197	5.3242	5.8517	.0364	.4554
.3	125	2.6568	3.0724	3.3434	3.6836	4.2379	.0581	.4922
	221	2.8443	3.1330	3.3562	3.6024	4.0092	.0430	.3456
	289	2.9028	3.1601	3.3525	3.5434	3.8600	.0233	.2875

the first set on the paths of $X(\mathbf{s}), \mathbf{s} \in [0,1]^2$. The third set defines an orthogonal Gaussian measure to the first two. The data are the first simulation corresponding to $n = 289$ and $\theta = .2$. Figure 6 plots the empirical semivariogram, as well as the three semivariograms used for interpolation. Figure 7 shows the interpolated values and prediction variances for 31 locations $(.387, .1 + .01n)$, $n = 0, \ldots, 30$, under the three distinct covariograms $(\sigma^2, \theta) = (1, .2)$, $(\sigma^2, \theta) = (2, .4)$ and $(\sigma^2, \theta) = (1.8, .4)$. The first two covariograms yielded very similar predicted values and prediction variances, but the third covariogram yielded different prediction variances, although it also produced similar predicted values. It is striking that the third covariogram graphically does not deviate from the first covariogram as much as the second covariogram (see Fig. 6), and yet it yields much more different interpolation results. Therefore, when interpolation is the objective of study, the ratio σ^2/θ matters more than each individual parameter.

Figure 7 can be explained using the fixed-domain asymptotic properties of interpolation discussed in the previous section, though the sample size is finite. To further check whether the asymptotic results are applicable to a less denser lattice, I used the sample data on a subset of the 289 locations, $(i/11, j/11)$, $i, j = 0, 1, \ldots, 10$, to predict for the same 31 locations. Figure 8 plots the predicted values and prediction variances. With this smaller sample, the same conclusions are reached. I also used data from another subset of 221 locations, set 2, to predict for the sample locations, and again reached similar conclusions.

I repeated the interpolation for 14 other datasets and reached the same conclusions each time. Although 15 samples is not a large number, I believe that fixed-domain asymptotics is appropriate in geostatistics when interpolation is concerned. Moreover, this is the only theory that can explain the interpolation results seen repeatedly in the simulation study.

Table 2. Summary of Estimates of σ^2: Percentiles, Means, and Sample Standard Deviations (SD)

θ_0	n	5%	25%	50%	75%	95%	BIAS	SD
.1	125	.7577	.8742	.9736	1.1006	1.3041	−.0003	.1765
	221	.7682	.8881	.9827	1.0999	1.2939	.0018	.1684
	289	.7702	.8930	.9812	1.0857	1.2730	−.0015	.1594
.2	125	.6369	.7951	.9514	1.1365	1.5021	−.0040	.2812
	221	.6371	.7994	.9525	1.1552	1.5092	.0017	.2865
	289	.6474	.7891	.9583	1.1364	1.4773	−.0093	.2687
.3	125	.5307	.7217	.9364	1.2178	1.7158	.0138	.3990
	221	.5400	.7263	.9326	1.1676	1.8371	.0134	.4138
	289	.5379	.7379	.9334	1.1762	1.8380	.0120	.3930

Example 2. Let $\{b(\mathbf{s}), \mathbf{s} \in \mathbb{R}^2\}$ be a mean-0 Gaussian stationary process with an isotropic covariogram $K_0(h) = \exp(-h/.2)$. Conditional on $\{b(\mathbf{s}), \mathbf{s} \in \mathbb{R}^2\}$, $\{Y(\mathbf{s}), \mathbf{s} \in \mathbb{R}^2\}$ is a set of binomial variables so that $Y(\mathbf{s})$ has a binomial probability $p(\mathbf{s}) = \exp(-2 + b(\mathbf{s}))/(1 + \exp(-2 + b(\mathbf{s})))$ and size $n(\mathbf{s})$. I simulate $b(\mathbf{s})$ and $Y(\mathbf{s})$ on the same 289 locations as in set 3 in the previous example. The sample size $n(\mathbf{s})$ at each of these 289 locations is fixed at 10. I use these data to predict $p(\mathbf{s})$ for the same 31 locations used in the previous example. I calculate the predicted values and prediction variances by fixing $\beta = -2$ and assuming three different "fitted" exponential covariograms, $K(h; \sigma^2, \theta) = \sigma^2 \exp(-h/\theta)$, $h \geq 0$, for $(\sigma^2, \theta) = (1, .2), (2, .4)$, and $(1.8, .4)$, to show explicitly how an incorrect covariogram affects interpolation.

The MMSE prediction and prediction variance can be computed using a Markov chain Monte Carlo (MCMC) approach, as done by in Diggle et al. (1998) and Zhang (2003). In particular, Zhang (2003) showed that combining partial analytic results with the MCMC approach can significantly reduce the necessary run length for a satisfactory convergence. Here I follow the approach of Zhang (2003). For any function ψ of $b(\mathbf{s})$, by theorem 1 of Zhang (2003),

$$E\{\psi(b(\mathbf{s}))|\mathbf{Y}\} = E\big[E\{\psi(b(\mathbf{s}))|\mathbf{b}\}|\mathbf{Y}\big],$$

where $\mathbf{b} = (b_1, \ldots, b_{289})$ and $\mathbf{Y} = (Y_1, \ldots, Y_{289})$ denote the random effects and the observed binomial variables at the sampling locations. Because the process $\{b(\mathbf{s})\}$ is Gaussian, the

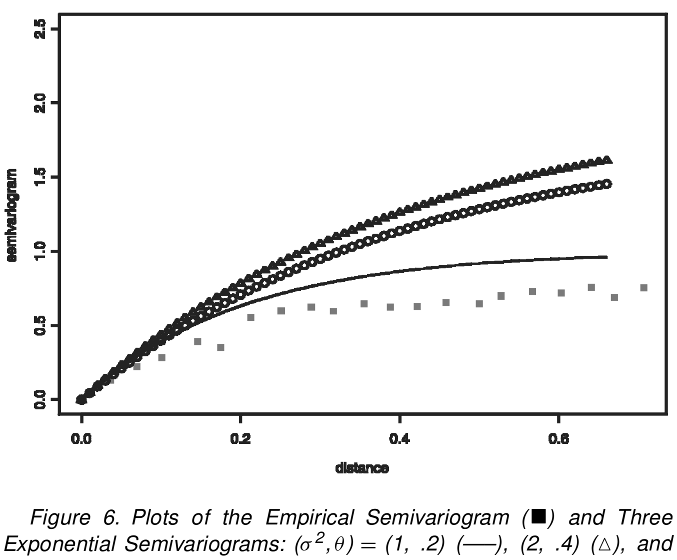

Figure 6. Plots of the Empirical Semivariogram (■) and Three Exponential Semivariograms: $(\sigma^2, \theta) = (1, .2)$ (——), (2, .4) (△), and (1.8, .4) (○).

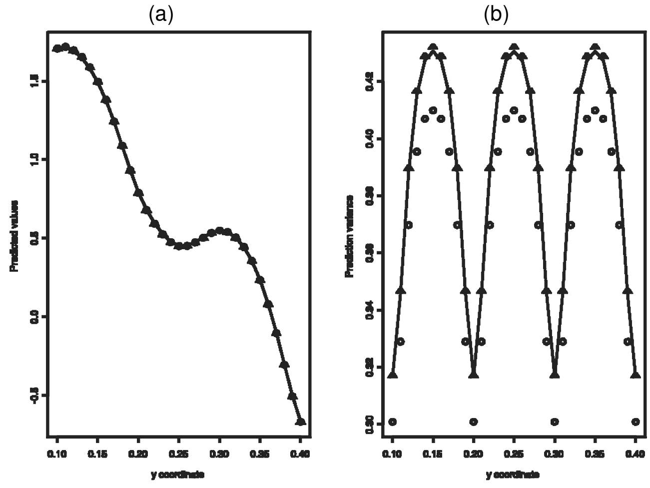

Figure 7. Comparison of Interpolation Results [(a) predicted values; (b) prediction variance] Under Three Exponential Covariograms Using Data on 289 Locations [——, $(\sigma^2, \theta) = (1, .2)$; △,(2, .4); ○,(1.8, .4)].

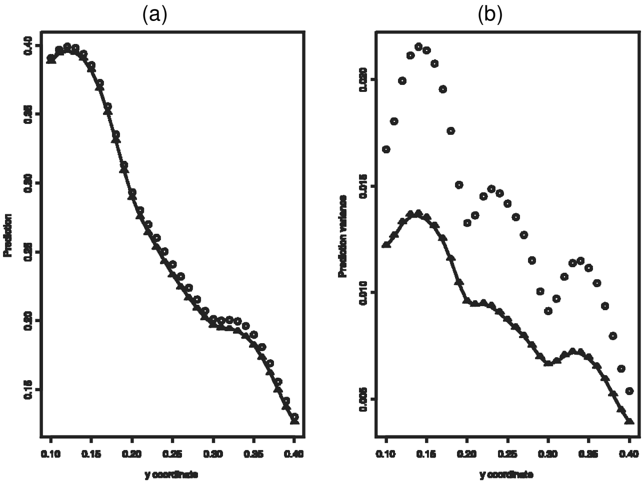

Figure 9. Comparisons of (a) Prediction Values and (b) Prediction Variances at 31 Locations Using Three Different Exponential Covariograms $K(x) = \sigma^2 \exp(-x/\theta)$ [——, $(\sigma^2, \theta) = (1, .2)$; △, (2, .4); ○, (1.8, .4)].

conditional expectation $E\{\psi(b(\mathbf{s}))|\mathbf{b}\}$ for any function ψ is of the form $\int f(t) \exp(-t^2)\, dt$, which can be fairly easily approximated to any given precision (Crouch and Spiegelman 1990) if it cannot be computed in closed form. The Metropolis–Hastings algorithm can be easily implemented to generate a Markov chain $\mathbf{b}^{(m)}, m \geq 1$, with the stationary distribution being the conditional distribution of $\mathbf{b}$ given $\mathbf{Y}$. Therefore, $E[E\{\psi(b(\mathbf{s}))|\mathbf{b}\}\,|\mathbf{Y}]$ can be approximated by the average of $E\{\psi(b(\mathbf{s}))|\mathbf{b}^{(m)}\}$, $m = 1, \ldots, M$. Here I adopt the Metropolis–Hastings algorithm of Zhang (2002) and choose $M = 2{,}000$. I graphically checked the convergence by plotting predicted values versus the run length M, and $M = 2{,}000$ showed a satisfactory convergence.

Figure 9 plots the predicted values and prediction variances for each of the 31 locations under the three sets of parameters. The plots show that predictions corresponding to the first two covariograms are nearly identical. Although the predicted values under the third covariogram are close to those under the other two covariograms, prediction variances are quite different. These results are interpretable under the fixed-domain asymptotics.

It is known that at an unsampled location $\mathbf{s}$, $E\{p(\mathbf{s})|\mathbf{Y}\} = E\{Y(\mathbf{s})/n(\mathbf{s})|\mathbf{Y}\}$, and hence (7) implies that predictions of $p(\mathbf{s})$ under two equivalent measures are similar. It can be argued heuristically that the prediction variances for $p(\mathbf{s})$ under two equivalent measures are similar as well. The assumptions in the model imply that for any prediction site $\mathbf{s}$, $E\{p^2(\mathbf{s})|\mathbf{Y}\}$ does not depend on $n(\mathbf{s})$. Choosing $n(\mathbf{s}) = 2$ and predicting the probability that $Y(\mathbf{s}) = 1$, the best prediction is

$$\begin{aligned}\Pr\big(Y(\mathbf{s}) = 1|\mathbf{Y}\big) &= E\big(\Pr\{Y(\mathbf{s}) = 1|b(\mathbf{s})\}|\mathbf{Y}\big) \\ &= 2E\big(p(\mathbf{s})\big(1 - p(\mathbf{s})\big)|\mathbf{Y}\big).\end{aligned}$$

This probability will be asymptotically the same under two equivalent measures, and, because the same statement applies to $E(p(\mathbf{s})|\mathbf{Y})$, so will $E(p^2(\mathbf{s})|\mathbf{Y})$, and therefore $\mathrm{var}(p(\mathbf{s})|\mathbf{Y})$.

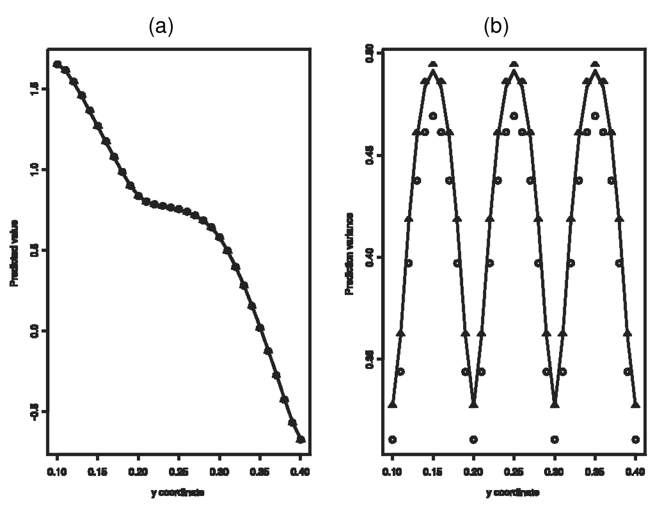

Figure 8. Comparison of Interpolation Results [(a) predicted values; (b) prediction variance] for 31 Locations Under Three Exponential Covariograms Using Data at 121 Locations: (i/10, j/10): $i, j = 0, \ldots, 10$ [——, $(\sigma^2, \theta) = (1, .2)$; △, (2, .4); ○, (1.8, .4)].

5. SUMMARY AND DISCUSSION

In this article I have used properties of equivalence of probability measures to show that not all parameters in a spatial GLMM are consistently estimable, but one quantity can be estimated consistently by the ML method under the fixed-domain asymptotics. This quantity is more important to interpolation than individual parameters. I also showed the impact of equivalent probability measures on interpolation under the fixed-domain asymptotics.

I ran simulation studies to discover the practical implications of the theoretical results. The simulation results show that the MLE for the ratio σ^2/θ in an exponential variogram has a more symmetric distribution with a smaller variance when more data are sampled in a fixed and bounded region. However, less noticeable is the improved estimation for the parameters θ and σ^2 achieved by sampling more data, particularly when the spatial correlation is not too weak. The MLE's have negligible biases for all parameters when the sample size is large. This does not contradict the inconsistency of the estimators for σ^2 and θ, however, because the variances of these estimators may not vanish when the sample size increases to infinity.

Note that the results in this article were established under specific statistical models—spatial GLMM's. For non-Gaussian data like counts data, some marginal models have been proposed that use only the first two moments for interpolation, and use generalized estimating equations for parameter estimation (see, e.g., Albert and McShane 1995; Gotway and Stroup 1997; McShane, Albert, and Palmatier 1997). Because the marginal models do not fully specify the probability distribution of the process, studing the equivalence of probability measures under these models is difficult. Although generalized estimating equation (GEE) estimators for marginal models have been shown to be consistent and asymptotically normal for longitudinal data under some regularity conditions, their properties under spatial dependence remain to be established. A key difference between spatial data and longitudinal data is that for longitudinal data, those from different subjects are independent. This independence is important in establishing the asymptotic properties of GEE estimation and is no longer true for spatial data.

Gotway and Stroup (1997) developed a generalized linear model approach to spatial prediction given spatial discrete or categorical data, in which only the first two moments are used to construct a linear prediction. This prediction differs from the classical kriging prediction, such as indicator kriging, in that the mean structure is estimated nonlinearly. Although the general results of Stein (1990, thm. 3.1; 1999, thm. 8, chap. 4) on asymptotic optimality of linear predictions under an incorrect covariogram are still applicable in principle, some specific details—such as verification of the equivalence of corresponding Gaussian measures in this case—remain to be worked out for the specific problem at hand.

Also note that this work is focused on the Matérn class. Other covariance functions are often used in practice, such as the spherical covariogram and the powered-exponential model. Unlike the Matérn class, spectral densities corresponding to those covariograms do not have a closed form, and there are no results on the equivalence of Gaussian measures induced by these covariograms. If Theorem 2 can be established, for example, for the spherical covariogram $K(h; \sigma^2, \theta) = \sigma^2(1 - 1.5(h/\theta) + .5(h/\theta)^3)\chi_{0\le h<\theta}$ such that two such covariograms define equivalent Gaussian measures of the same mean if and only if both have the same ratio σ^2/θ, then the results in this article, including Theorem 3, are all extendable to the spherical covariogram.

Some open problems remain for future research. It is known that Gaussian measures on a separable Hilbert space are either equivalent or orthogonal. It is not known, however, whether the similar statement applies to measures on the paths of $Y(\mathbf{s})$, $\mathbf{s} \in T$ for a bounded T, where $Y(\mathbf{s})$ follows a non-Gaussian spatial GLMM. By Theorem 1, the equivalence of Gaussian measures on $b(\mathbf{s})$ implies the equivalence of measures on $Y(\mathbf{s})$ in the spatial GLMM for any β. It is not yet known what the measures on $Y(\mathbf{s})$ would be (i.e., orthogonal, equivalent, or neither orthogonal nor equivalent) were the two measures on $b(\mathbf{s})$ orthogonal. If the orthogonality of measures on $b(\mathbf{s})$ implies the orthogonality of measures on $Y(\mathbf{s})$, then Theorem 3 may be extendable to the non-Gaussian case. The prediction results in Example 2 seem to suggest that probability measures on $Y(\mathbf{s})$ cannot be equivalent if the measures on $b(\mathbf{s})$ are orthogonal.

Theorem 3 was established under a known mean, although in reality the mean may need to be estimated. It will be a more difficult problem to establish Theorem 3 when the mean is unknown and must be estimated. A key technique in my proof is using the concavity of the log-likelihood function. When the mean is estimated, the likelihood function is a multivariate function, and the concavity may not be true or may be hard to establish. The restricted ML estimation avoids estimation of the mean, making it difficult to apply the martingale convergence theorem that is essential in my proof. Ying (1991, thm. 3) established strong consistency and asymptotic normality of the MLE for σ^2/θ for a one-dimensional Gaussian process with an unknown mean and an exponential covariogram $\exp(-h/\theta)$, $h > 0$. The proof explicitly uses the directional nature of one-dimensional space and is difficult to extend to high dimensions. Another interesting problem is to establish the asymptotic normality of the consistent estimator, as was done by Ying (1991) in high-dimensional space for the Matérn class. This problem remains open even for the exponential covariogram in high dimensions.

Corollary 2 implies that given observations of Y, the conditional distributions of $Y(\mathbf{s})$ under two equivalent measures are asymptotically equal. It will be interesting to learn wheather the conditional distributions of $b(\mathbf{s})$ given Y are also asymptotically equal under two equivalent measures, because it is of interest to predict a function of the random effects $b(\mathbf{s})$. It also will be interesting to see wheather (7) holds with a general function of $Y(\mathbf{s}_{n+k})$ or $b(\mathbf{s}_{n+k})$.

This article emphasizes the interpolation aspect of spatial GLMM. Although in many situations the ultimate goal is interpolation, the underlying problem in other situations may be estimating the linear coefficients to find significant explanatory variables. Indeed, an important application of spatial GLMM is on disease mapping, where a major objective is to find those significant variables that affect disease rate. The covariogram may impact estimation of the linear coefficients. This is clearly seen in a spatial linear model. However, Stein (1990) showed that using fixed but incorrect values of linear coefficients yields asymptotically optimal linear predictors (see also comments of Stein 1999, in the last paragraph of sec. 4.3). It will be interesting to investigate whether analogous results hold for nonlinear predictors in a non-Gaussian GLMM.

APPENDIX: PROOFS

This appendix provides some technical details and proofs of the main results in Section 2.

Proof of Theorem 1

Let $\mathcal{F}_y$ and $\mathcal{F}_b$ denote the σ-algebras generated by $Y(\mathbf{s}), \mathbf{s} \in T$, and $b(\mathbf{s}), \mathbf{s} \in T$, and let E_i denote the expectation with respect to $P_{\beta,\boldsymbol{\theta}_i}$, $i = 1, 2$. Then for any $A \in \mathcal{F}_y$, $P_{\beta,\boldsymbol{\theta}_2}(A) = E_2\{E_2(\mathbb{1}_A|\mathcal{F}_b)\}$, where $\mathbb{1}_A$ is the indicator function. Conditional on $\mathcal{F}_b$, $\{Y(\mathbf{s}), \mathbf{s} \in T\}$ has the same distribution under both measures. Hence $E_2(\mathbb{1}_A|\mathcal{F}_b) = E_1(\mathbb{1}_A|\mathcal{F}_b)$. Consequently,

$$P_{\beta,\boldsymbol{\theta}_2}(A) = E_2\{E_1(\mathbb{1}_A|\mathcal{F}_b)\}.$$

Constrained on $\mathcal{F}_b$, the two measures are equivalent. Let ρ denote the Radon–Nikodym derivative of $P_{\beta,\boldsymbol{\theta}_2}$ constrained on $\mathcal{F}_b$ with respect to $P_{\beta,\boldsymbol{\theta}_1}$ constrained on $\mathcal{F}_b$. Then ρ is necessarily $\mathcal{F}_b$ measurable, and for any $\mathcal{F}_b$-measurable function g,

$$E_2(g) = E_1(\rho g).$$

Taking $g = E_1(\mathbb{1}_A|\mathcal{F}_b)$,

$$E_2\{E_1(\mathbb{1}_A|\mathcal{F}_b)\} = E_1\{\rho E_1(\mathbb{1}_A|\mathcal{F}_b)\}$$
$$= E_1\{E_1(\rho\mathbb{1}_A|\mathcal{F}_b)\} = E_1(\rho\mathbb{1}_A).$$

I have shown that $E_2(\mathbb{1}_A) = E_1(\rho\mathbb{1}_A)$. Because ρ is integrable, it follows that $E_1(\mathbb{1}_A) = 0$ implies $E_2(\mathbb{1}_A) = 0$. Therefore, on $\mathcal{F}_y$, $P_{\beta,\boldsymbol{\theta}_2} \ll P_{\beta,\boldsymbol{\theta}_1}$. Similarly, it can be shown that $P_{\beta,\boldsymbol{\theta}_1} \ll P_{\beta,\boldsymbol{\theta}_2}$ on $\mathcal{F}_y$. The theorem is proved.

Proof of Theorem 2

First, assume that $\sigma_1^2\alpha_1^{2\nu} = \sigma_2^2\alpha_2^{2\nu}$. For $i = 1, 2$, the isotropic spectral density corresponding to $K(x; \sigma_i^2, \alpha_i, \nu)$ is, by (2), $f_i(u) = \sigma_i^2\alpha_i^{2\nu}\pi^{-d/2}(\alpha_i + u^2)^{-\nu-d/2}$. Obviously, $f_1(u)u^{2\nu+d}$ is bounded away from 0 and ∞ as $u \to \infty$. To prove the equivalence of the two measures, I need only show that (5) is satisfied.

If $\sigma_1^2\alpha_1^{2\nu} = \sigma_2^2\alpha_2^{2\nu}$, then by (2),

$$\left|\frac{f_2(u) - f_1(u)}{f_1(u)}\right| = \left|\frac{(\alpha_1^2 + u^2)^{\nu+d/2}}{(\alpha_2^2 + u^2)^{\nu+d/2}} - 1\right|$$
$$\le \left|(\alpha_1^2 + u^2)^{\nu+d/2} - (\alpha_2^2 + u^2)^{\nu+d/2}\right|/u^{2\nu+d}$$
$$\le \left|\big((\alpha_1/u)^2 + 1\big)^{\nu+d/2} - \big((\alpha_2/u)^2 + 1\big)^{\nu+d/2}\right|.$$

Note that

$$(x+1)^\alpha = 1 + \alpha x + O(x^2), \qquad x \to 0.$$

Then, as $u \to \infty$,

$$\left|\big((\alpha_1/u)^2 + 1\big)^{\nu+d/2} - \big((\alpha_2/u)^2 + 1\big)^{\nu+d/2}\right|$$
$$\le (\nu + d/2)|\alpha_1^2 - \alpha_2^2|u^{-2} + O(u^{-4}).$$

The integral in (5) is finite for $d = 1, 2, 3$. Therefore, the two measures are equivalent.

If $\sigma_1^2\alpha_1^{2\nu} \ne \sigma_2^2\alpha_2^{2\nu}$, let $\sigma_0^2 = \sigma_2^2(\alpha_2/\alpha_1)^{2\nu}$. Then $\sigma_0^2\alpha_1^{2\nu} = \sigma_2^2\alpha_2^{2\nu}$, and the two Matérn covariograms $K(x; \sigma_0^2, \alpha_1, \nu)$ and $K(x; \sigma_2^2, \alpha_2, \nu)$ define two equivalent measures. I just need to show that $K(x; \sigma_0^2, \alpha_1, \nu)$ and $K(x; \sigma_1^2, \alpha_1, \nu)$ define two orthogonal Gaussian measures. It is helpful to note that the two covariograms define the same correlogram and differ only in variance. I can show in general that any such covariograms define two orthogonal Gaussian measures. Let P_i be the Gaussian measure for $X(\mathbf{s})$, $\mathbf{s} \in T$, with mean 0 and covariance function $K(\cdot; \sigma_i^2, \alpha_1, \nu)$, $i = 0, 1$.

Let ψ_j, $j \ge 1$, be an orthonormal basis of the Hilbert space generated by $X(\mathbf{s})$, $\mathbf{s} \in T$, with the inner product

$$\langle \xi, \eta \rangle = \int \xi\eta\, dP_0.$$

Each ψ_j can be chosen to be a linear combination of $X(\mathbf{s}_{j,k})$, $k = 1, \dots, n_j$, for some $n_j > 0$ and $\mathbf{s}_{j,k} \in T$, $k = 1, \dots, n_j$. The existence of such a basis follows from the continuity of the covariance function. By lemma 1 of Ibragimov and Rozanov (1978, p. 72), the two measures P_i, $i = 0, 1$, are equivalent on $X(\mathbf{s})$, $\mathbf{s} \in T$, if and only if they are equivalent on ψ_j, $j \ge 1$.

Because $K(x; \sigma_1^2, \alpha_1, \nu) = (\sigma_1^2/\sigma_0^2)K(x; \sigma_0^2, \alpha_1, \nu)$ for any $\mathbf{s}$ and $\mathbf{t}$,

$$E_1\big(X(\mathbf{s})X(\mathbf{t})\big) = (\sigma_1^2/\sigma_0^2)E_0\big(X(\mathbf{s})X(\mathbf{t})\big).$$

This equation also holds for any linear combinations of $X(\mathbf{s})$, $\mathbf{s} \in T$. It follows that

$$E_1(\psi_j\psi_k) = (\sigma_1^2/\sigma_0^2)E_0(\psi_k\psi_k) = (\sigma_1^2/\sigma_0^2)\delta_{jk}.$$

Then

$$\sum_{i,k=1}^{\infty}\big(E_1(\psi_j\psi_k) - E_0(\psi_j\psi_k)\big)^2 = \infty.$$

It follows that the two measures are not equivalent on ψ_j, $j \ge 1$ (Stein 1990, thm. 7, p. 129), and hence must be orthogonal, because the two Gaussian measures are either equivalent or orthogonal. The proof is completed.

Proof of Corollary 1

If there exist weakly consistent estimators σ_k^2 such that for any β and $\boldsymbol{\theta} = (\sigma^2, \alpha, \nu)$, then σ_k^2 converges to σ^2 in probability under $P_{\beta,\boldsymbol{\theta}}$. Then, by a well-known fact (see, e.g., Dudley 1989, thm. 9.2.1, p. 226), there is an almost-surely convergent subsequence $\sigma_{k_j}^2$ such that

$$P_{\beta,\boldsymbol{\theta}}\Big(\lim_{j\to\infty}\sigma_{k_j}^2 = \sigma^2\Big) = 1. \tag{A.1}$$

Let $\boldsymbol{\theta}' = (2^{2\nu}\sigma^2, \alpha/2, \nu)$. Then the two measures $P_{\beta,\boldsymbol{\theta}}$ and $P_{\beta,\boldsymbol{\theta}'}$ are equivalent by Theorem 2, and, consequently, (A.1) implies that

$$P_{\beta,\theta'}\Big(\lim_{j\to\infty}\sigma_{k_j}^2 = \sigma^2\Big) = 1.$$

On the other hand, the weak consistency of σ_k^2 under $P_{\beta,\boldsymbol{\theta}'}$ implies that for any almost-surely convergent subsequence, the limit equals $2^{2\nu}\sigma^2$. This contradiction shows that consistent estimators for σ^2 do not exist. Similarly, consistent estimators for α do not exist.

Corollary 2 directly follows from Theorem 2 and the theorem of Blackwell and Dubins (1962).

Proof of Theorem 3

Write $\theta = \sigma^{-2}$ and denote by P_θ the Gaussian measure on the paths of $X(\mathbf{s})$, $\mathbf{s} \in D$, corresponding to Matérn covariogram $K(\cdot; \theta^{-1}, \alpha_1, \nu)$ and mean 0. Let $f_{n,\theta}$ be the probability density function of $X(\mathbf{s})$, $\mathbf{s} \in D_n$ under the probability measure P_θ and write $\theta^* = \sigma_0^{-2}(\alpha_1/\alpha_0)^{2\nu}$. It is well known that for any θ, the Radon–Nikodym derivative $\rho_n(\theta) = f_{n,\theta}/f_{n,\theta^*} = dP_\theta^{D_n}/dP_{\theta^*}^{D_n}$ converges with P_{θ^*}-probability 1, and the limit equals the density of the absolutely continuous component of measure P_θ with respect to measure P_{θ^*}, where $P_\theta^{D_n}$ denotes the measure of P_θ restricted on $\sigma(X(\mathbf{s}), \mathbf{s} \in D_n)$ (see, e.g., Gihman and Skorohod 1974, thm. 1, p. 442). In particular, if P_θ and P_{θ^*} are orthogonal, then $\rho_n(\theta) \to 0$. By Theorem 2, the two measures P_θ and P_{θ^*} are orthogonal if $\theta \ne \theta^*$. Hence, with P_{θ^*}-probability 1,

$$\lim_{n\to\infty}\log\rho_n(\theta) = \begin{cases} -\infty & \text{if } \theta \ne \theta^* \\ 1 & \text{if } \theta = \theta^*. \end{cases}$$

The theorem holds if $\sigma_n^{-2} \to \theta^*$ with P_0-probability 1. Because $P_{\theta^*} \equiv P_0$, where P_0 is defined in the theorem, I need only show $\sigma_n^{-2} \to \theta^*$ with P_{θ^*}-probability 1. To this end, it suffices to show that for any $\epsilon > 0$, with P_{θ^*}-probability 1 there exists an integer N such that for $n > N$ and $|\theta - \theta^*| > \epsilon$,

$$\log\rho_n(\theta) = \log f_{n,\theta} - \log f_{n,\theta^*} \le -1. \tag{A.2}$$

First note that for any n, the log-likelihood function $L_n(\theta) = \log f_{n,\theta}$ is concave. Indeed, the covariance function of the variables $X(\mathbf{s})$, $\mathbf{s} \in D_n$, can be written as $(1/\theta)\boldsymbol{\Gamma}_n$, where the matrix $\boldsymbol{\Gamma}_n$ does not depend on θ. It is clear that

$$L_n(\theta) = (1/2)(n\log\theta - \theta\mathbf{X}'\boldsymbol{\Gamma}_n^{-1}\mathbf{X}) + R_n,$$

where R_n does not depend on θ. Obviously, $\partial^2 L_n/\partial\theta^2 = -2n\theta^{-2}$, and the function L_n is strictly concave for each n.

For any $\epsilon > 0$, let $\theta_1 = \theta^* - \epsilon$ and $\theta_2 = \theta^* + \epsilon$. Because $\rho_n(\theta_i) \to -\infty$, $i = 1, 2$, there exists an integer N such that for all $n > N$, $\log(\rho_n(\theta_i)) \le -1$, $i = 1, 2$. In view of this and $\log(\rho_n(\theta^*)) = 0$ for any n, the concavity implies that (A.2) holds for all $n > N$. Indeed, if there exist an $n > N$ and a $\theta > \theta_2$ such that (A.2) is not true, then

$$\log(\rho_n(\theta)) > \log(\rho_n(\theta_2)) \quad\text{and}\quad \log(\rho_n(\theta^*)) = 0 > \log(\rho_n(\theta_2)),$$

which is impossible, because $\theta^* < \theta_2 < \theta$ and the concavity implies that $\log(\rho_n(\theta_2))$ cannot be smaller than both $\log(\rho_n(\theta))$ and $\log(\rho_n(\theta^*))$. This contradiction shows that (A.2) must be true for $\theta > \theta_2$. Similarly, it must be true for $\theta < \theta_1$. The proof of the theorem follows.

[Received July 2002. Revised September 2003.]

REFERENCES

Abramowitz, M., and Stegun, I. (eds.) (1967), *Handbook of Mathematical Functions*, Washington, DC: U.S. Government Printing Office.

Albert, P. S., and McShane, L. M. (1995), "A Generalized Estimating Equations Approach for Spatially Correlated Data: Applications to the Analysis of Neuroimaging Data," *Biometrics*, 51, 627–638.

Blackwell, D., and Dubins, L. (1962), "Merging of Opinions With Increasing Information," *The Annals of Mathematical Statistics*, 33, 882–886.

Chilés, J. P., and Delfiner, P. (1999), *Geostatistics: Modeling Spatial Uncertainty*, New York: Wiley.

Christensen, O. F., and Waagepetersen, R. P. (2002), "Bayesian Prediction of Spatial Count Data Using Generalized Linear Mixed Models," *Biometrics*, 58, 280–286.

Cressie, N. (1993), *Statistics for Spatial Data* (rev. ed.), New York: Wiley.

Crouch, E. A. C., and Spiegelman, D. (1990), "The Evaluation of Integrals of the Form $\int_{-\infty}^{\infty} f(t)\exp(-t^2)\,dt$: Application to Logistic-Normal Models," *Journal of the American Statistical Association*, 85, 464–469.

Davis, B. (1987), "Uses and Abuses of Cross-Validation in Geostatistics," *Mathematical Geology*, 19, 241–248.

Diggle, P. J., Ribeiro, P. J., and Christensen, O. F. (2002), "An Introduction to Model-Based Geostatistics," in *Spatial Statistics and Computational Methods*, ed. J. Møller, New York: Springer-Verlag, pp. 43–86.

Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998), "Model-Based Geostatistics" (with discussion), *Journal of the Royal Statistical Society*, Ser. C, 47, 299–350.

Dudley, R. M. (1989), *Real Analysis and Probability*, Pacific Grove, CA: Wadsworth.

Gihman, I., and Skorohod, A. V. (1974), *The Theory of Stochastic Processes I*, New York: Springer-Verlag.

Gotway, C. A., and Stroup, W. W. (1997), "A Generalized Linear Model Approach to Spatial Data Analysis and Prediction," *Journal of Agricultural, Biological and Environmental Statistics*, 2, 157–178.

Handcock, M., and Stein, M. (1993), "A Bayesian Analysis of Kriging," *Technometrics*, 35, 403–410.

Handcock, M., and Wallis, J. R. (1994), "An Approach to Statistical Spatial-Temporal Modeling of Meteorological Fields" (with discussion), *Journal of the American Statistical Association*, 89, 368–390.

Heagerty, P. J., and Lele, S. R. (1998), "A Composite Likelihood Approach to Spatial Binary Data," *Journal of the American Statistical Association*, 93, 1099–1111.

Ibragimov, I., and Rozanov, Y. (1978), *Gaussian Random Processes*, New York: Springer-Verlag.

Krige, D. G. (1951), "A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand," *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52, 119–139.

Mardia, K. V., and Marshall, R. J. (1984), "Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Statistics," *Biometrika*, 71, 135–146.

Mardia, K. V., and Watkins, A. J. (1989), "On Multimodality of the Likelihood in the Spatial Linear Model," *Biometrika*, 76, 289–295.

McShane, L. M., Albert, P. S., and Palmatier, M. A. (1997), "A Latent Process Regression Model for Spatially Correlated Count Data," *Biometrics*, 53, 698–706.

Ripley, B. D. (1988), *Statistical Inferences for Spatial Processes*, New York: Cambridge University Press.

Stein, M. L. (1990), "Uniform Asymptotic Optimality of Linear Predictions of a Random Field Using an Incorrect Second-Order Structure," *The Annals of Statistics*, 18, 850–872.

——— (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, New York: Springer.

——— (2004), "Equivalence of Gaussian Measures for Some Nonstationary Random Fields," *Journal of Statistical Planning and Inference*, in press.

Warnes, J., and Ripley, B. D. (1987), "Problems With Likelihood Estimation of Covariance Functions of Spatial Gaussian Processes," *Biometrika*, 74, 640–642.

Williams, B. J., Santner, T. J., and Notz, W. I. (2000), "Sequential Design of Computer Experiments to Minimize Integrated Response Functions," *Statistica Sinica*, 10, 1133–1152.

Yadrenko, M. (1983), *Spectral Theory of Random Fields*, New York: Optimization Software.

Ying, Z. (1991), "Asymptotic Properties of a Maximum Likelihood Estimator With Data From a Gaussian Process," *Journal of Multivariate Analysis*, 36, 280–296.

Zhang, H. (2002), "On Estimation and Prediction for Spatial Generalized Linear Mixed Models," *Biometrics*, 56, 129–136.

——— (2003), "Optimal Interpolation and the Appropriateness of Cross-Validating Variogram in Spatial Generalized Linear Mixed Models," *Journal of Computational and Graphical Statistics*, 12, 698–713.

Zhang, H., and Wang, H. H. (2002), "A Study on Prediction of Spatial Binomial Probabilities With an Application to Spatial Design," in *Computing Science and Statistics*, 34, eds. E. Wegman and A. Braverman, Fairfax Station, VA: Interface Foundation of North America, Inc., pp. 263–276.

Zimmerman, D. L., and Zimmerman, M. B. (1991), "A Comparison of Spatial Semivariogram Estimators and Corresponding Ordinary Kriging Predictors," *Technometrics*, 23, 77–91.