Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Analyzing Nonstationary Spatial Data Using Piecewise Gaussian Processes

Hyoung-Moon Kim, Bani K Mallick & C. C Holmes

To cite this article: Hyoung-Moon Kim, Bani K Mallick & C. C Holmes (2005) Analyzing Nonstationary Spatial Data Using Piecewise Gaussian Processes, Journal of the American Statistical Association, 100:470, 653-668, DOI: 10.1198/016214504000002014

To link to this article: https://doi.org/10.1198/016214504000002014

Analyzing Nonstationary Spatial Data Using Piecewise Gaussian Processes

Hyoung-Moon KIM, Bani K. MALLICK, and C. C. HOLMES

In many problems in geostatistics the response variable of interest is strongly related to the underlying geology of the spatial location. In these situations there is often little correlation in the responses found in different rock strata, so the underlying covariance structure shows sharp changes at the boundaries of the rock types. Conventional stationary and nonstationary spatial methods are inappropriate, because they typically assume that the covariance between points is a smooth function of distance. In this article we propose a generic method for the analysis of spatial data with sharp changes in the underlying covariance structure. Our method works by automatically decomposing the spatial domain into disjoint regions within which the process is assumed to be stationary, but the data are assumed independent across regions. Uncertainty in the number of disjoint regions, their shapes, and the model within regions is dealt with in a fully Bayesian fashion. We illustrate our approach on a previously unpublished dataset relating to soil permeability of the Schneider Buda oil field in Wood County, Texas.

KEY WORDS: Bayes factor; Kriging; Model averaging; Reversible-jump Markov chain Monte Carlo; Voronoi tessellation.

1. INTRODUCTION

Many naturally occurring spatial processes exhibit highly irregular (i.e., nonsmooth) behavior in their covariance structure. For instance, in geostatistical applications where the response variable of interest is highly dependent on rock type, we can expect to see sharp transitions in the covariance between points across strata (i.e., rock boundaries). In such cases conventional spatial models built on stationarity assumptions are inappropriate. Moreover, most nonstationary models, in which the covariance between points depends on location, also implicitly assume that the covariance dies off smoothly as a function of distance.

In this article we propose a model that can deal with sharp transitions in covariance structure. The method that we adopt is a random (Bayesian) partition model in which the spatial domain is partitioned using a Voronoi tiling such that within regions (tiles), the process is assumed to be stationary, whereas across regions, we assume independence. By treating the model structure within a Bayesian framework, we are able to assess uncertainty in the position of the boundaries as well as the form of the model within regions. This leads us to model averaging when making predictive statements, which tends to smooth the nonsmooth individual models in areas of greatest uncertainty.

Our motivating example is taken from the Schneider Buda oil field in the Wood County, Texas. The response variable of interest is soil permeability, and the task is to recover a two-dimensional surface over the field based on 92 measurement of permeability taken at distinct locations. The ability to accurately model the permeability of the oil field is a vital task for petroleum engineers, because the profitability of secondary stage oil extraction from the field is a function of the permeability.

Although we concentrate on this example from petroleum engineering, our approach is generic and applicable to any situation in which sharp changes in covariance are suspected. For instance, whenever a natural or man-made boundary exists in the spatial field, there is likely to be a possibility of highly irregular changes in covariance between points.

In this article we assume that the reader is familiar with conventional approaches to modeling Gaussian spatial fields using stationary kriging methods. Cressie (1993), Stein (1999), Handcock and Wallis (1994) have provided comprehensive overviews of this approach. The seminal articles by Le and Zidek (1992), Handcock and Stein (1993), and Brown, Le, and Zidek (1994) detail Bayesian alternatives to kriging. The rest of this article is organized as follows. Section 2 overviews current approaches to nonstationary spatial processes and highlights why they would be inappropriate for data where the covariance structure changes sharply across boundaries. Section 3 describes our procedure using a piecewise Gaussian process within a Bayesian hierarchical model and the computation strategies. Section 4 applies the method to several simulated datasets and shows that it results in good consequences. Section 5 analyzes permeability data from the Schneider Buda field in Wood County, Texas. Section 6 ends the article with a discussion.

2. EXISTING APPROACHES FOR NONSTATIONARY SPATIAL PROCESSES

In this section we briefly review some current approaches to model nonstationary spatial fields.

2.1 Moving-Window Approaches

Haas (1990, 1995) proposed the idea of a moving-window approach where to predict at a spatial location s, one first constructs a local spatial window around it and then assumes a stationary random field model within the window and performs kriging. This operation is reperformed at every point in which one is interested. The size of the window is chosen by cross-validation, although there is always a question as to whether using a particular cross-validation criterion is the best or most reasonable approach. The window is by design sym-

Hyoung-Moon Kim is Assistant Professor, Department of Applied Statistics, Konkuk University, Seoul, Korea (E-mail: *hmkim@konkuk.ac.kr*). Bani K. Mallick is Professor, Department of Statistics, Texas A&M University, College Station, TX 77843 (E-mail: *bmallick@stat.tamu.edu*). C. C. Holmes is Lecturer, Department of Statistics, University of Oxford, Oxford, OX1 3TG, U.K. (E-mail: *c.holmes@stats.ox.ac.uk*). The second author's research was supported in part by National Science Foundation grant DMS-03-27713, National Cancer Institute grant CA 57030, a Superfund basic research program grant from the National Institute of Environmental Health Sciences and a Texas A&M Interdisciplinary grant. The authors thank Akhil Datta-Gupta for the data and Michael Sherman, Sang Heon Lee, and Naijun Sha for helpful comments. The authors also thank the associate editor and two referees for their constructive suggestions that led to significant improvements in the presentation of the article.

Journal of the American Statistical Association
June 2005, Vol. 100, No. 470, Theory and Methods
DOI 10.1198/016214504000002014

metric around the point s and hence is unable to capture very sharp changes in covariance structure. In addition, by specifically defining a local model in this way, joint predictions at, say, two locations do not take correlation into account.

2.2 The Method of Empirical Orthogonal Functions

This method is based on Karhunen–Loeve expansions of the covariance function introduced by Cohen and Jones (1969) and have been used by geophysicists (Creutin and Obled 1982). Nychka, Wikle, and Royle (1999) have modified the model using wavelet basis functions. Although this is clearly a very general and flexible approach, it has not proved popular, due to a lack of connection with traditional approaches based on kriging and variograms. Our piecewise approach is more interpretable, because we use conventional kriging models within each region, and is more connected with traditional approaches.

2.3 Deformation Approaches

The very popular idea of the spatial deformation approach is to perform a change of coordinates that results in isotropic spatial correlation. The principal advocates of this methodology are Sampson and Guttorp (1992) and Guttorp and Sampson (1994). Usually the mapping is done from the original geographical space to an alternative dispersion space where stationarity is assumed, and usually the transformation functions are modeled by thin plate splines (Sampson and Guttorp 1992). This method works only with multiple observations. Moreover, the single smooth mapping will not be well suited to handling sharply changing covariance structures, because the thin-plate spline implicitly penalizes this. Sampson, Damian, and Guttorp (2001) have described more recent developments. Smith (1996) pointed out that spatial correlations of temperature data had different structures in the western and eastern United States and that the deformation approach does not handle this situation very well. Recently, Schmidt and O'Hagan (2003) proposed a Bayesian version of the deformation approach by developing a fully Bayesian model and using Markov chain Monte Carlo (MCMC) based computation. They modeled the transformation function using Gaussian processes rather than thin-plate splines.

2.4 Models Using Kernels

A stationary Gaussian process $Z(\mathbf{s})$ can be expressed as the convolution of a Gaussian white noise process $X(\mathbf{s})$ and convolution kernel $K(\mathbf{s})$ as $Z(\mathbf{s}) = \int_{\mathbb{R}^2} K(\mathbf{s}-\mathbf{u})X(\mathbf{u})\,d\mathbf{u}$. Higdon, Swall, and Kern (1999) extended this method with the novel idea of a spatially evolving kernel, where they considered the process $Z(\mathbf{s}) = \int_{\mathbb{R}^2} K_{\mathbf{s}}(\mathbf{u})X(\mathbf{u})\,d\mathbf{u}$, where the kernel $K_{\mathbf{s}}$ depends on position $\mathbf{s}$. They treated $K_{\mathbf{s}}$ as an unknown smooth function and used a Bayesian hierarchical model to estimate it with other model parameters. This method is ideally suited for smoothly varying spatial covariance structure and may not be that successful in capturing sharp changes.

Other recent developments for modeling nonstationary processes include those of Nott and Dunsmuir (2002) and Fuentes and Smith (2001). Nott and Dunsmuir (2002) extended the deformation approach of Sampson and Guttorp (1992), which is more useful for smoothly varying structures. The method suggested by Fuentes and Smith (2001) proposes a superposition of locally stationary processes, which again are assumed to evolve smoothly over the spatial domain.

2.5 Multiple Models

The work most closely related to our approach is that of Fuentes (2001), who used a mixture of locally stationary process defined within disjoint regions to model the globally nonstationary process. Fuentes (2001) assumed that the shape of the regions are known a priori and the number of regions is estimated using the Akaike information criterion or Bayes information criterion for model choice. Kernels are used to weight the influence of the local processes outside of their regions.

A key feature of our method is the treatment of model uncertainty through a prior distribution on the shape and the number of regions. Model uncertainty relates to the fact that many different spatial models may offer nearly equally plausible representations of the data. The uncertainty in the partition locations also allows us to achieve data-adaptive smooth transitions in the covariance structure, as well as sharp changes. Although our model is motivated mainly for data with sharp changes, the marginal predictive densities can capture smoothly changing processes through model averaging (Draper 1995).

In all of the aforementioned approaches, the methods implicitly assume smoothness in the covariance structure. In many situations this will be entirely appropriate. However, as expressed in Section 1, there are also situations in which a priori knowledge of the application area dictates that this smoothness assumption is likely to be violated. In the next section we discuss our approach to these situations.

3. PIECEWISE GAUSSIAN PROCESSES AND THE BAYESIAN MODEL

In this section we describe a piecewise Gaussian process for modeling a nonstationary process. The idea behind the model is that the field D can be partitioned into several regions such that the data are assumed to be homogeneous within regions and independent across regions. In particular, within each partition, we assume that the data follow a Gaussian process with an isotropic correlation function. Hence the overall field is a piecewise stationary Gaussian process. Once we have specified that within regions the data follow a conventional Gaussian process, the vital questions are (a) how to create these partitions (regions), (b) where to create these partitions, and (c) how many partitions that we need.

We choose to construct the regions by a Voronoi tessellation (Green and Sibson 1978). This tessellation is defined by a number of centers, $\mathbf{c} = (\mathbf{c}_1, \ldots, \mathbf{c}_M)$, that divide D into M disjoint regions, $R_1, \ldots, R_M$, such that points within R_i are closer to $\mathbf{c}_i$ than any other center $\mathbf{c}_j, j \neq i$, that is, $R_i = \{\mathbf{x} \in D : \|\mathbf{x} - \mathbf{c}_i\| < \|\mathbf{x} - \mathbf{c}_j\| \ \forall j \neq i\}$, where we take $\|(x_1, \ldots, x_n)\|^2 = \sum_{i=1}^n x_i^2$ (Euclidean distance). Figure 1 represents a Voronoi tessellation with several regions.

Our choice of a Voronoi partitioning scheme is made on grounds of tractability and computational simplicity. Our approach is not restricted to this choice; any partitioning structure for the spatial domain can be adopted. Within each region, we allow for a different correlation structure, which gives us flexibility, so in a sense we are performing local kriging within each region rather than the usual global kriging. By specifying hard or nonoverlapping regions, we are allowing for discontinuity in the correlation structure, which is natural in our applications.

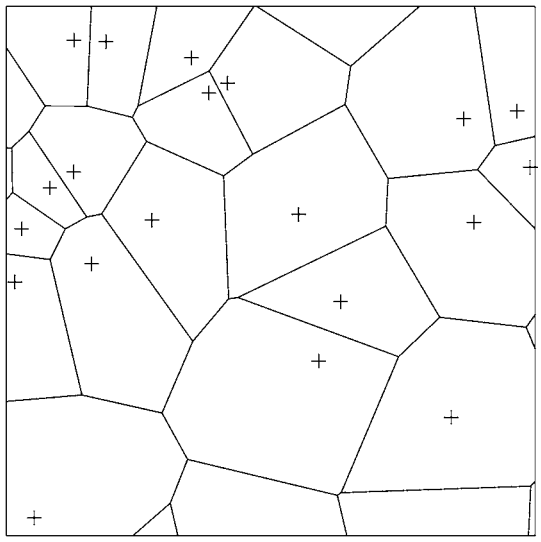

Figure 1. Voronoi Tessellation.

Now we assume that the number and center location of the regions are unknown components of the model that must be inferred alongside the other model parameters. This induces a model space of unknown dimensionality, giving added flexibility. In our Bayesian framework we adopt a prior distribution on all parameters in the model.

3.1 Piecewise Gaussian Processes

We now define some notations for our model. Let $\{Z(\mathbf{x}), \mathbf{x} \in D\}, D \subseteq \mathbb{R}^d$, be the random field of interest, and suppose that we have n observations $\mathbf{Z} = (Z(\mathbf{x}_1), \ldots, Z(\mathbf{x}_n))'$ from a single realization of this field, where $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are known distinct locations in D. We take M to be the number of partitions or disjoint regions and define $Z(x_{i,j})$ as jth observation of region R_i. Further, let n_i be the number of data points in R_i, so we can write $\mathbf{Z}_i = (Z(x_{i,1}), \ldots, Z(x_{i,n_i}))'$ with $\mathbf{X}_i = (x_{i,1}, \ldots, x_{i,n_i})'$ and $\mathbf{X} = \{\mathbf{X}_1', \ldots, \mathbf{X}_M'\}$. Let $\mathbf{t}$ denote the number and location of the region centers that define the tessellation, $\mathbf{t} = \{M, \mathbf{c}\}$.

The data across regions are assumed to be independent; hence we concentrate on the distribution within a region. We assume that within the ith region, the data follow a multivariate Gaussian distribution with an isotropic correlation function, so that

$$(\mathbf{Z}_i|\mathbf{x}_i, \boldsymbol{\beta}_i, \mathbf{t}) \sim \mathrm{N}_{n_i}(\mathbf{G}_i\boldsymbol{\beta}_i, \boldsymbol{\Sigma}_i), \tag{1}$$

where $\mathbf{G}_i = [g_i(x_{i,j})]_{n_i \times p_i}$ is a known design matrix used to capture trend within the ith region and $\boldsymbol{\Sigma}_i$ is a positive-definite covariance matrix that captures local spatial dependence, $(\boldsymbol{\Sigma}_i)_{jk} = \mathrm{cov}(Z(x_{i,j}), Z(x_{i,k}))$. All of our posterior and predictive distribution calculations in the next sections are based on this very general setup of the mean function, which allows a piecewise polynomial trend surface or any other basis function model by properly choosing the design matrix $\mathbf{G}$. But in our numerical examples, we consider only the simplest situation of piecewise constant-mean model, choosing $\mathbf{G}$ as column of 1's and $p_i = 1$.

In particular, we assume that

$$(\boldsymbol{\Sigma}_i)_{jk} = \sigma_i^2 \mathbf{K}_{\boldsymbol{\theta}_i}(x_{i,j}, x_{i,k}),$$

where σ_i controls the overall variance in the ith region and $\mathbf{K}_{\boldsymbol{\theta}_i}(\cdot)$ is a parametric correlation function with parameters $\boldsymbol{\theta}_i$. As our working family of isotropic correlation functions in each partition, we use the general (power) exponential correlation function (Yaglom 1987; De Oliveira, Kedem, and Short 1997)

$$\mathbf{K}_{\boldsymbol{\theta}_i}(x_{i,j}, x_{i,k}) = \mathbf{K}_{\boldsymbol{\theta}_i}(l_{jk}) = \exp\left(-\nu_i l_{jk}^{\theta_{2_i}}\right) = \theta_{1_i}^{l_{jk}^{\theta_{2_i}}}, \quad i = 1, \ldots, M, \tag{2}$$

where $l_{j,k}$ represents the Euclidean distance between $x_{i,j}$ and $x_{i,k}$, $\nu_i > 0$ and $\theta_{1_i} = e^{-\nu_i} \in (0, 1)$ and $\theta_{2_i} \in (0, 2]$ are unknown parameters.

Let $\boldsymbol{\Omega}$ denote the set of all model parameters (except $\mathbf{t}$) and let $\boldsymbol{\omega}$ be a realization of $\boldsymbol{\Omega}$, that is, $\{\boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{\theta}\}$, with $\boldsymbol{\beta} = \{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_M\}$, $\boldsymbol{\sigma}^2 = \{\sigma_1^2, \ldots, \sigma_M^2\}$, and $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M\}$. Then $\boldsymbol{\Omega}_i = \{\boldsymbol{\beta}_i, \sigma_i^2, \boldsymbol{\theta}_i\}$. Using the notation described so far, the likelihood of the data given the tessellation and model parameters is

$$L(\mathbf{Z}|\mathbf{x}, \boldsymbol{\omega}, \mathbf{t}) = \prod_{i=1}^{M} f(\mathbf{Z}_i|\mathbf{x}_i, \boldsymbol{\omega}_i, \mathbf{t}), \tag{3}$$

where $f(\mathbf{Z}_i|\mathbf{x}_i, \boldsymbol{\omega}_i, \mathbf{t})$ is a multivariate normal density of the response $\mathbf{Z}_i$ that has corresponding locations $\mathbf{x}_i$ and is given at (1).

In this article we note that any suitable model can be used to describe the data in each region.

3.2 Bayesian Analysis

We assign the following prior structure for the parameters:

$$\pi(\boldsymbol{\omega}|\mathbf{t}) = \prod_{i=1}^{M} \pi(\boldsymbol{\beta}_i|\mathbf{t}, \sigma_i^2)\pi(\sigma_i^2|\mathbf{t})\pi(\boldsymbol{\theta}_i|\mathbf{t}), \tag{4}$$

and we assign conjugate priors within the disjoint regions as

$$(\boldsymbol{\beta}_i|\mathbf{t}, \sigma_i^2) \sim \mathrm{N}_{p_i}(\mathbf{0}, \sigma_i^2\boldsymbol{\lambda}_i^{-1}), \qquad i = 1, \ldots, M, \tag{5}$$

$$(\sigma_i^2|\mathbf{t}) \sim \mathrm{IG}\left(\frac{\gamma_1}{2}, \frac{\gamma_2}{2}\right), \tag{6}$$

and

$$(\boldsymbol{\theta}_i|\mathbf{t}) \sim \mathrm{U}(0, 1) \cdot \mathrm{U}(0, 2]. \tag{7}$$

Here p_i is the dimension of the trend coefficients $\boldsymbol{\beta}_i$, $\boldsymbol{\lambda}_i/\sigma_i^2$ is the prior precision matrix of $\boldsymbol{\beta}_i$, and γ_1 and γ_2 are hyperparameters.

The parameter $\mathbf{t} = \{M, \mathbf{c}\}$ defines the partition or tessellation structure. We choose to assign the uniform priors

$$\pi(\mathbf{t}) = \pi(M)\pi(\mathbf{c}|M) \tag{8}$$

$$= \mathrm{DU}(M|1, \ldots, M_{\max}) \times \mathrm{DU}\left(\mathbf{c}\middle|1, \ldots, \binom{n}{M}\right) \tag{9}$$

$$= \frac{1}{M_{\max}} \times \frac{1}{\binom{n}{M}}, \tag{10}$$

where $\mathrm{DU}(x|1, \ldots, n)$ means discrete uniform on $1, \ldots, n$ and $M_{\max}$ is the maximum allowable number of centers (usually specified as n). Note that $\mathbf{c}|M$ is uniformly distributed over all possible combinations of M centers from known distinct locations in D. Bayesian methods, being fully probabilistic, contain a natural penalty against overly complex models; hence it is possible to adopt a discrete uniform prior over the positions of the region centers and the number of centers.

Under these prior distributions, the posterior distribution is proportional to $f(\mathbf{z}|\mathbf{x}, \boldsymbol{\omega}, \mathbf{t})\pi(\boldsymbol{\omega}|\mathbf{t})\pi(\mathbf{t})$. The marginal likelihood or the integrated likelihood of any proposed partition structure and the predictive distribution for a new observation are given by the following results. Using the conjugate priors as in (4)–(6), the marginal likelihood (after integrating out $\boldsymbol{\beta}$ and σ explicitly) is given by

$$\begin{aligned} f(\mathbf{z}|\mathbf{x}, \mathbf{t}) &= \int_{\boldsymbol{\Omega}} f(\mathbf{z}|\mathbf{x}, \boldsymbol{\omega}, \mathbf{t})\pi(\boldsymbol{\omega}|\mathbf{t})\, d\boldsymbol{\omega} \\ &= \int_{\boldsymbol{\theta}} f(\mathbf{z}|\mathbf{x}, \boldsymbol{\theta}, \mathbf{t})\pi(\boldsymbol{\theta}|\mathbf{t})\, d\boldsymbol{\theta} \qquad (11) \\ &= \prod_{i=1}^{M} \frac{\gamma_2^{\gamma_1/2}|\boldsymbol{\lambda}_i|^{1/2}\Gamma((\gamma_1+n_i)/2)}{\pi^{n_i/2}\Gamma(\gamma_1/2)} \\ &\quad \times \int_{\boldsymbol{\theta}_i} (\gamma_2+\alpha_i)^{-(\gamma_1+n_i)/2}\left|\mathbf{K}_{\boldsymbol{\theta}_i}\right|^{-1/2}|\mathbf{V}_i|^{1/2}\pi(\boldsymbol{\theta}_i|\mathbf{t})\, d\boldsymbol{\theta}_i, \end{aligned} \qquad (12)$$

where $\alpha_i = \mathbf{z}_i'(\mathbf{K}_{\boldsymbol{\theta}_i} + \mathbf{G}_i\boldsymbol{\lambda}_i^{-1}\mathbf{G}_i')^{-1}\mathbf{z}_i$, $\mathbf{V}_i = (\mathbf{G}_i'\mathbf{K}_{\boldsymbol{\theta}_i}^{-1}\mathbf{G}_i + \boldsymbol{\lambda}_i)^{-1}$, $|\cdot|$ denotes the determinant of a matrix, and $\Gamma(\cdot)$ is the usual gamma function.

Moreover, the predictive distribution for $\mathbf{z}(\mathbf{x}_0)$ to predict at a new location $\mathbf{x}_0$ is given by (13)–(16) as

$$f(\mathbf{z}(\mathbf{x}_0)|\mathbf{z}) = \int_{\mathbf{T}} f(\mathbf{z}(\mathbf{x}_0)|\mathbf{z}, \mathbf{t})\pi(\mathbf{t}|\mathbf{z})\, d\mathbf{t}, \qquad (13)$$

and if $\mathbf{x}_0 \in R_i$, then it is proportional to $f(\mathbf{z}(\mathbf{x}_0)|\mathbf{z}_i, \mathbf{t})$,

$$f(\mathbf{z}(\mathbf{x}_0)|\mathbf{z}_i, \mathbf{t}) \propto \int_{\boldsymbol{\theta}_i} f(z(\mathbf{x}_0)|\boldsymbol{\theta}_i, \mathbf{z}_i, \mathbf{t})\pi(\boldsymbol{\theta}_i|\mathbf{z}_i, \mathbf{t})\, d\boldsymbol{\theta}_i, \qquad (14)$$

where

$$(\mathbf{z}(\mathbf{x}_0)|\boldsymbol{\theta}_i, \mathbf{z}_i, \mathbf{t}) \sim Mt_1\left(\hat{\boldsymbol{\mu}}_{1.2_i}, \left(\frac{\gamma_2 + ss(\mathbf{z}_i)}{n_i+\gamma_1}\right)\boldsymbol{\Psi}_i, n_i+\gamma_1\right) \qquad (15)$$

and

$$\begin{aligned} (\boldsymbol{\theta}_i|\mathbf{z}_i, \mathbf{t}) &\propto \left|\mathbf{K}_{\boldsymbol{\theta}_i}\right|^{-1/2}\left|\mathbf{G}_i'\mathbf{K}_{\boldsymbol{\theta}_i}^{-1}\mathbf{G}_i\right|^{-1/2}\left|\mathbf{K}_{11.2_i}\right|^{-1/2} \\ &\quad \times \left|\mathbf{I} + \left(\mathbf{G}_i'\mathbf{K}_{\boldsymbol{\theta}_i}^{-1}\mathbf{G}_i\right)^{-1}\left\{\mathbf{P}_i'\mathbf{K}_{11.2_i}^{-1}\mathbf{P}_i + \boldsymbol{\lambda}_i\right\}\right|^{-1/2} \\ &\quad \times (\gamma_2 + ss(\mathbf{z}_i))^{-(n_i+\gamma_1)/2}\Psi_i^{1/2}\pi(\boldsymbol{\theta}_i|\mathbf{t}), \end{aligned} \qquad (16)$$

with

$$\begin{aligned} \hat{\boldsymbol{\mu}}_{1.2_i} &= \mathbf{g}_i'(\mathbf{x}_0)\hat{\boldsymbol{\beta}}_i + \mathbf{k}_{\boldsymbol{\theta}_i}'\mathbf{K}_{\boldsymbol{\theta}_i}^{-1}(\mathbf{z}_i - \mathbf{G}_i\hat{\boldsymbol{\beta}}_i), \\ \hat{\boldsymbol{\beta}}_i &= \left(\mathbf{G}_i'\mathbf{K}_{\boldsymbol{\theta}_i}^{-1}\mathbf{G}_i\right)^{-1}\mathbf{G}_i'\mathbf{K}_{\boldsymbol{\theta}_i}^{-1}\mathbf{z}_i, \\ ss(\mathbf{z}_i) &= (\mathbf{z}_i - \mathbf{G}_i\hat{\boldsymbol{\beta}}_i)'\mathbf{K}_{\boldsymbol{\theta}_i}^{-1}(\mathbf{z}_i - \mathbf{G}_i\hat{\boldsymbol{\beta}}_i), \\ \boldsymbol{\Psi}_i &= \mathbf{K}_{11.2_i} + \mathbf{P}_i\left(\mathbf{G}_i'\mathbf{K}_{\boldsymbol{\theta}_i}^{-1}\mathbf{G}_i\right)^{-1}\mathbf{P}_i', \\ \mathbf{K}_{11.2_i} &= \mathbf{k}_{\boldsymbol{\theta}_i}(\mathbf{x}_0, \mathbf{x}_0) - \mathbf{k}_{\boldsymbol{\theta}_i}'\mathbf{K}_{\boldsymbol{\theta}_i}^{-1}\mathbf{k}_{\boldsymbol{\theta}_i}, \end{aligned}$$

and

$$\mathbf{P}_i = \mathbf{g}_i'(\mathbf{x}_0) - \mathbf{k}_{\boldsymbol{\theta}_i}'\mathbf{K}_{\boldsymbol{\theta}_i}^{-1}\mathbf{G}_i.$$

We let $Mt_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, m)$ denote an n-dimensional multivariate t-distribution with density function

$$\frac{\Gamma((n+m)/2)}{(\pi m)^{n/2}\Gamma(m/2)}|\boldsymbol{\Sigma}|^{-1/2} \times \left(1 + \frac{1}{m}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)^{-(n+m)/2}. \qquad (17)$$

The derivations are given in the Appendix.

3.3 Computational Strategy

The posterior distribution, $\pi(\mathbf{t}|\mathbf{z}, \mathbf{x})$, is of complicated form owing to the varying dimensionality of our problem. Hence we make use of tailored MCMC to draw samples from the posterior model space. In particular, we must design a Markov chain to move between models of different dimension, relating to models of differing number of partitioning regions. The generating algorithm that we propose for this is as follows:

1. Initialize the tessellation structure, $\mathbf{t}$, with one randomly chosen generating position from known distinct locations in D.
2. When $1 < M < M_{\max}$, with probability 1/3, either add (birth step), delete (death step), or move (moving step) a generating position. Any new (birth) center is picked uniformly from the location values that do not already have generating point, and any deleting center is chosen uniformly from the present ones.
3. After a user-determined burn-in period, preserve every ith $\mathbf{t}$ generated.
4. Return to step 2 and repeat until a sufficient number of samples have been generated.

Using the notation of Green (1995), the acceptance probability for each of the move types in our problem is

$$\alpha = \min(1,\ \text{likelihood ratio} \times \text{prior ratio} \times \text{proposal ratio}).$$

For the moving step, the prior ratio and proposal ratio are both 1. The birth step prior ratio is equal to $(M+1)/(n-M)$, and the proposal ratio is $(n-M)/(M+1)$. Thus the proposed movement for the moving and birth steps at step 2 is accepted with probability

$$\alpha = \min\left(1, \frac{f(\mathbf{z}|\mathbf{x}, \mathbf{t}_p)}{f(\mathbf{z}|\mathbf{x}, \mathbf{t}_c)}\right), \qquad (18)$$

where the subscripts c and p refer to the current and proposed models, and the marginal likelihood, $f(\mathbf{z}|\mathbf{x}, \mathbf{t})$, is defined at (11)–(12). For the death step, the acceptance probability is the same, except that the fraction in (18) is inverted.

When M equals 1, we cannot jump to the death step, so when we propose adding a center in this state, we must multiply the fraction in (18) by 2/3 to adjust for this. Thus when $M = 2$ and we propose a death to maintain reversibility, we must multiply the fraction by 3/2. The reverse is applied when $M = M_{\max}$.

The fraction in (18) is the Bayes factor (Kass and Raftery 1995) for the proposed model, against the current model, which can be viewed as measuring the weight of evidence of the proposed model and the current model at predicting the data under the prior. In our case the integral, $f(\mathbf{z}|\mathbf{x}, \mathbf{t}) = \int_{\boldsymbol{\theta}} f(\mathbf{z}|\mathbf{x}, \boldsymbol{\theta}, \mathbf{t})\pi(\boldsymbol{\theta}|\mathbf{t})\, d\boldsymbol{\theta}$, is not explicitly available, and thus we

use a direct numerical integration (Berntsen, Espelid, and Genz 1991a,b) to calculate each marginal likelihood, which is given at (11)–(12).

Convergence of the Markov chain is monitored for each application. As a default, we preserve every 50th partition generated after 20,000 burn-ins. For predictions, we need to generate samples from the posterior distribution for the Gaussian process in a region R_i where a predicted point is located. We use a Metropolis–Hastings update for $\boldsymbol{\theta}$, where the acceptance probability is $\alpha(x, y) = \min(\frac{\varphi(y)}{\varphi(x)}, 1)$, where $\varphi(y) = |\mathbf{K}_{\boldsymbol{\theta}_y}|^{-1/2}|\mathbf{G}_i'\mathbf{K}_{\boldsymbol{\theta}_y}^{-1}\mathbf{G}_i|^{-1/2}|\mathbf{K}_{11.2_i}|^{-1/2}|\mathbf{I} + (\mathbf{G}_i'\mathbf{K}_{\boldsymbol{\theta}_y}^{-1}\mathbf{G}_i)^{-1}\{\mathbf{P}_i' \times \mathbf{K}_{11.2_i}^{-1}\mathbf{P}_i + \boldsymbol{\lambda}_i\}|^{-1/2}(\gamma_2 + ss(\mathbf{z}_i))^{-(n_i+\gamma_1)/2}\Psi_i^{1/2}$. We consider 1,000 generated values of $f(z(\mathbf{x}_0)|\boldsymbol{\theta}_i, \mathbf{z}_i, \mathbf{t})$ as samples from a predictive distribution by the composition method (Tanner 1996; Gelfand 1996) after a burn-in of 3,000 iterations.

4. SIMULATION STUDY

Here we report some simulation studies done to explore the behavior of our modeling approach. We generated several datasets using S+Spatial Stats (1997). Hereinafter, we use reasonably noninformative priors with $\boldsymbol{\lambda}_i = 10^{-4}\mathbf{I}_{p_i}$ and $\gamma_1 = \gamma_2 = .1$. We use a single intercept term, β_i, to model within-region trend, so that $p_i = 1$.

Simulation 1. The first simulation is a one-dimensional problem to show how the piecewise process can capture a smooth process even though the underlying model is nonsmooth. We adapted a simple one-dimensional example from Higdon (2001). In this example the sample size was 30, the spatial locations varied between 1 and 10, and the observations were created according to the formula

$$z(s_i) = \sin(2\pi[s_i/10]) + .2\sin(2\pi[s_i/2.5]) + \epsilon_i,$$

where the ϵ_i's are iid $N(0, .1^2)$. We used our piecewise Gaussian model. Figure 2 plots the true function and the estimated function. It is clear that due to mixing over the posterior model space, the estimated function is much smoother than any individual piecewise constant function. In this example the dependence is contained within the trend without spatial correlation. The following examples contain several types of spatial correlation.

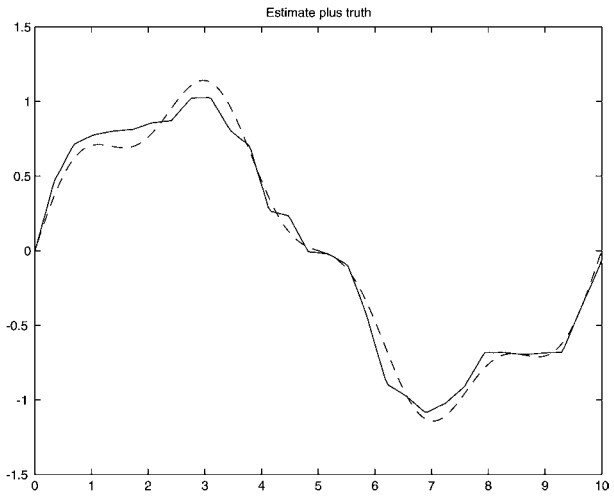

Figure 2. Plot of the Original (- - - - -) and Estimated (——) Functions.

Simulation 2. This simulation uses a single smooth spatial Gaussian process. To generate samples from a smooth Gaussian process, we used standard deviation (SD) 1, sill 1, and nugget 0 if not explicitly noted otherwise. This dataset with 64 points was generated using an exponential covariance function with range 2 and a Gaussian process with mean 0. Figure 3 plots the position and values of the simulated data and a histogram with a nonparametric density estimate of the generated z's. The contours were computed from the grid and plotted using S+ (using the command "contour"), which adds the contour values (heights) in bold.

We ran our model and retained samples every 50th partition after a burn-in of 20,000 iterations. Note that the number of partitions is equal to the (number of regions) − 1. We found that after 20,000 iterations, the marginal likelihoods of the model had settled down, so we chose this as our burn-in period. In the MCMC output, we obtained one region (1 center) in 95% of the samples. This suggests that the Bayesian model is not greedy and can recover the single Gaussian process for simple datasets. The probability distribution for number of regions is presented in Table 1.

Simulation 3. In this simulation we considered two Gaussian processes with a large difference in means. This is a fairly extreme example included to highlight the fact that we can recover the true structure in these circumstances, which can often occur in geologic data. We combined two generated datasets each with 36 data points into one dataset. The first combined dataset was generated using an exponential covariance with range 4 and a Gaussian process with mean 0 and SD 1. The second component of combined data used an exponential covariance function with range 1.5 and a Gaussian process with mean 10 and SD 1. Figure 4 plots the position and value of the simulated data and a histogram with a nonparametric density estimate of the observations z.

Again, more than 95% of the times in the posterior samples, we obtained two regions (one partition). The posterior modal sample of Voronoi tessellation with centers (marked with ◯) is plotted in Figure 5, which shows that our model searched out the location of the partition correctly as well.

To check model adequacy for prediction purposes, we used a cross-validation approach based on single-point-deletion predictive distributions as described by Gelfand, Dey, and Chang (1992). Let $Z(\mathbf{x}_i)$ be the random variable, let $z_{i,\text{obs}}$ be the observed value of $Z(\mathbf{x}_i)$, and let $\mathbf{z}_{(i)} = (z_{1,\text{obs}}, \ldots, z_{i-1,\text{obs}}, z_{i+1,\text{obs}}, \ldots, z_{n,\text{obs}})'$, the data vector with the ith observation deleted, $i = 1, \ldots, n$. The idea was then to quantify how well the model predicts each $Z(\mathbf{x}_i)$ based on $\mathbf{z}_{(i)}$. This predictive approach to model checking was the most natural in the present situation, because prediction is the model's intended use.

As a model choice criterion, we used mean squared prediction residuals (MSPRs) $= \sum_{i=1}^{n} r_i^2/n$, where $r_i = z_{i,\text{obs}} - \hat{z}_i$ and $\hat{z}_i =$ mean of $(Z(\mathbf{x}_i)|\mathbf{z}_{(i)})$, $i = 1, \ldots, n$. We compared a single Gaussian process (SGP) with our piecewise Gaussian process (PGP) using the mixture of partition models. Unsurprisingly, the PGP did much better, with an MSPR of 1.81, compared with 3.21 for SGP.

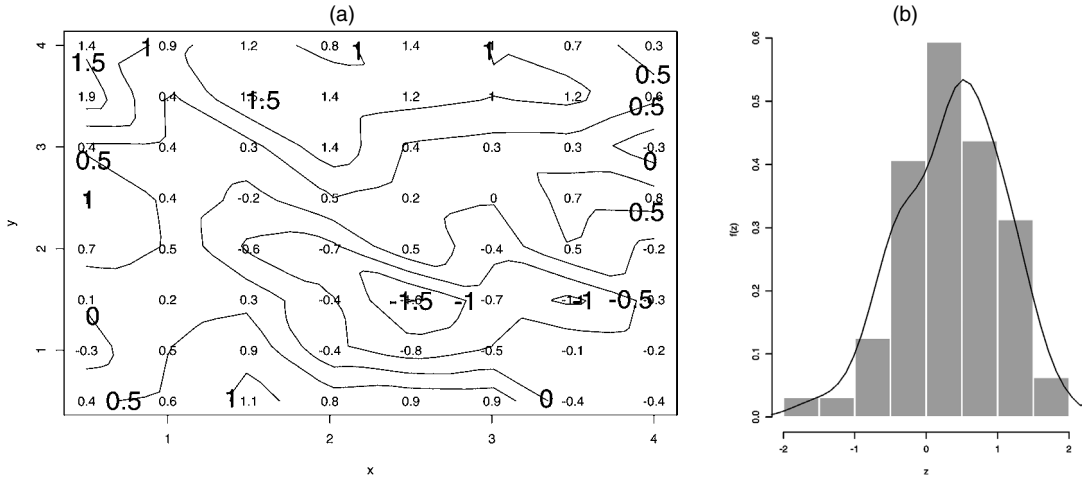

Figure 3. Position and Value of Simulated Dataset 2 With (a) a Contour Plot and (b) Histogram With a Nonparametric Density Estimate.

Table 1. Probability Distribution for Number of Regions for Simulation 2

No. of regions	*Probability (%)*
1	95.2
2	3.4
3	1.4

Simulation 4. Simulation dataset 3 is pretty extreme and seems dominated by the difference in mean level, so in the fourth simulation we combined two datasets each with 36 data points with the same mean and standard deviation but with different covariance structures. We generated three examples as follows. We adopted the same partition structure shown in Figure 5. In the first region we used the same data across all three examples, with mean 0, SD 1, sill 1, nugget 0, and range 9. In the second region we generated different datasets with the same value for mean, SD, nugget, and range but different values for the sill parameter, namely 2.5, 5, and 7. The position and histogram of the data are given in Figures 6–8. Table 2 summarizes our results of correct identification in these three situations. We also calculated the MSPR using the mixture models; these results are given in Table 3.

5. REAL DATA ILLUSTRATION: SPATIAL PERMEABILITY PREDICTION

Petroleum reservoirs are complex geologic formations that exhibit a wide range of physical and chemical heterogeneities that span multiple length scales and are impossible to describe in a deterministic fashion. Geostatistics and, more specifically, stochastic modeling of reservoir heterogeneities are being increasingly considered by reservoir analysts and petroleum en-

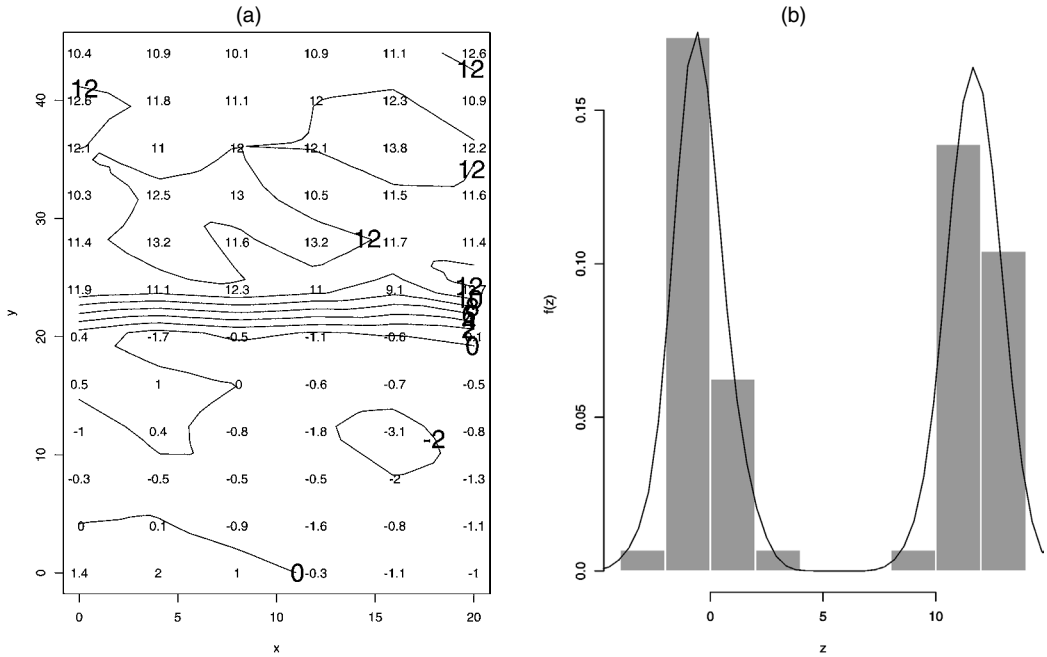

Figure 4. Position and Value of Simulated Dataset 3 With (a) Contour Plot and (b) Histogram With a Nonparametric Density Estimate.

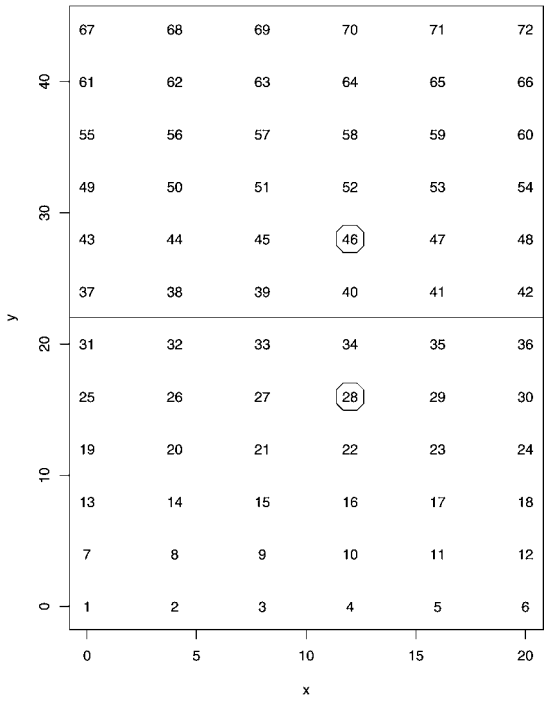

Figure 5. The Posterior Modal Sample of Voronoi Tessellation for Simulation Dataset 3.

gineers for their potential to generate more accurate reservoir models together with realistic measures of spatial uncertainty. The goal of reservoir characterization is to provide a numerical model of such reservoir attributes as hydraulic conductivities (permeability), storativities (porosity), and fluid saturation. These attributes are then used as inputs into complex transfer functions represented by various flow simulators to forecast future reservoir performance and oil recovery potential. In predicting future reservoir performance, it is imperative to have a geologic model that can be considered a "plausible" replica of the actual reservoir with acceptable uncertainty. Toward this objective, more flexible modeling approaches are needed to reproduce complex geologic/morphologic patterns and the wide variety of architectural heterogeneities observed in petroleum reservoirs.

In most flow situations, the single most influential input is permeability spatial distribution. Permeability is an important concept in porous media flow, such as the flow of underground oil. Physically, permeability arises both from the existence of pores and from the average structure of the connectivity of pores. Mathematically, fluid flow can typically be described by Darcy's law, which states that for steady-state flow in a porous medium, $v = -\rho \frac{\delta p}{L} \frac{1}{\mu}$, where ρ is the permeability, v is the volume flux per surface area of some region of length L, μ is the viscosity, and p is the pressure. The key role of permeability is evident from Darcy's law. In practice, therefore, dealing with the variability and uncertainties about permeability is critical for modeling porous flow.

Hence permeability predictions are a vital aspect of a reservoir description, but due to petrophysical variations rooted in digenesis, grain size variation, and cementation, highly and sharply heterogeneous behavior of the process are observed at different regions of the reservoir (Lee and Datta-Gupta 1999). Modeling these sharply changing heterogeneities is important, because they affect the amount of recovered oil. This has been confirmed by many theoretical examples and case studies; for instance, a postmortem study of North Sea fields developments (Dromgoole and Speers 1997) showed that such heterogeneities had a crucial impact on field behavior and that an inadequate, oversimplified smooth model can have direct impact on capital expenditures likely, in unfavorable oil price scenarios, to threaten the profitability of an oil field. The existing models fail

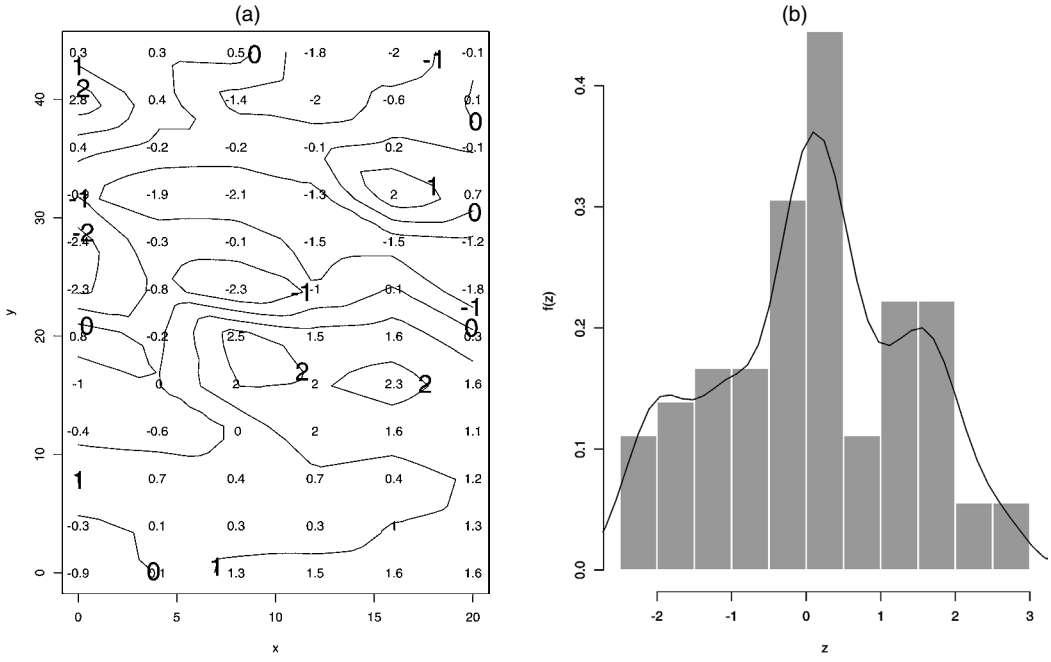

Figure 6. Position and Value of Simulated Dataset 4, With (a) Contour Plot and (b) a Histogram With a Nonparametric Density Estimate (with sill 2.5 for region 2).

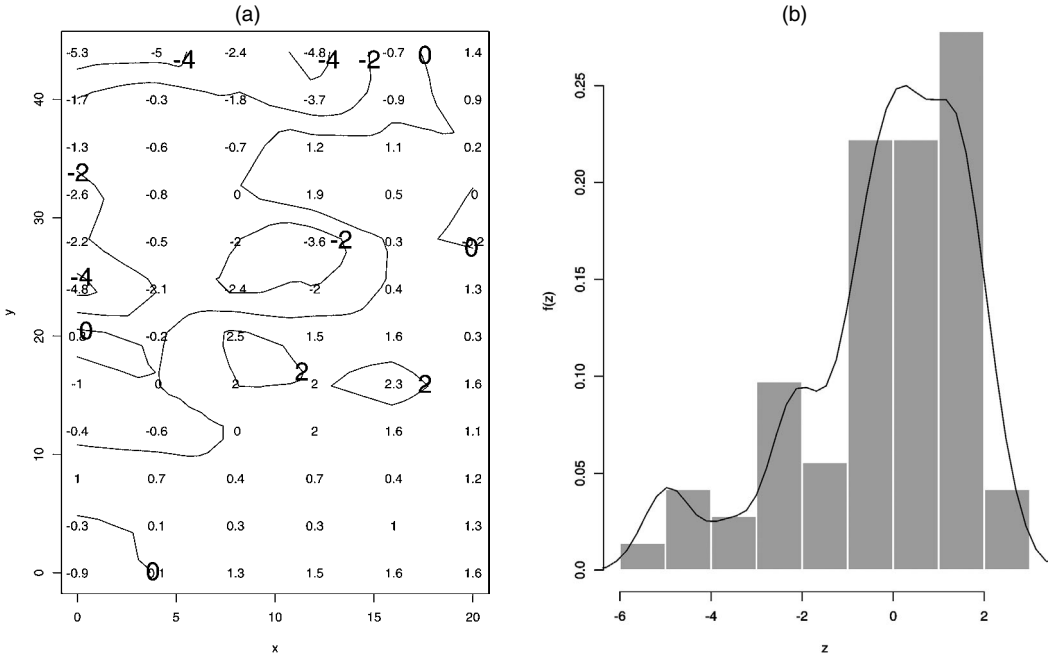

Figure 7. Position and Value of Simulated Dataset 4, With (a) a Contour Plot and (b) a Histogram With a Nonparametric Density Estimate (with sill 5 for region 2).

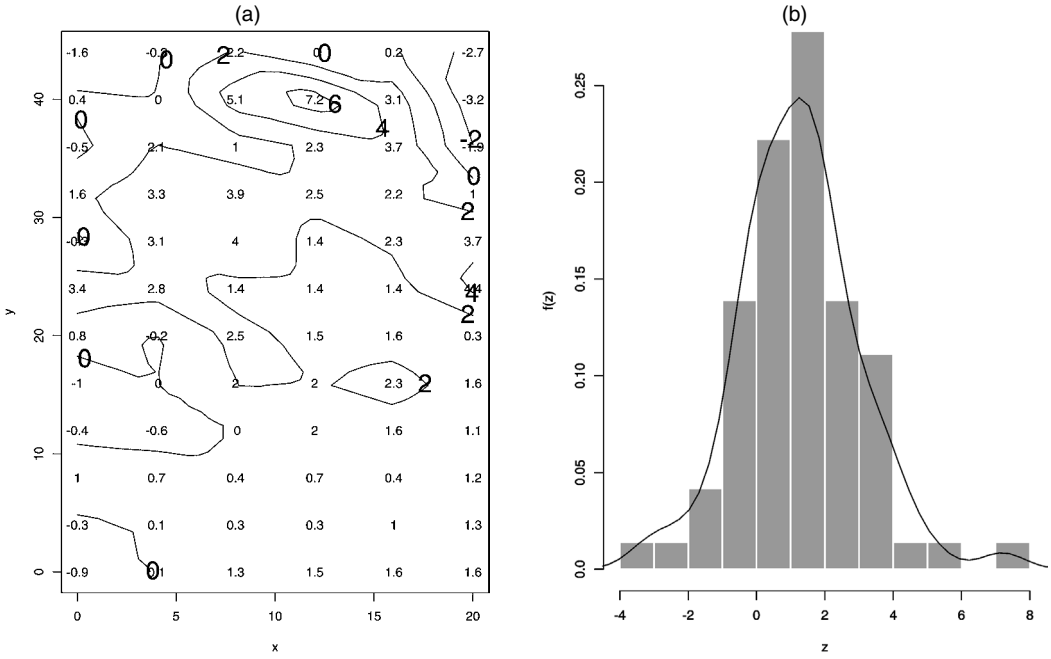

Figure 8. Position and Value of Simulated Dataset 4, With (a) a Contour Plot and (b) a Histogram With a Nonparametric Density Estimate (with sill 7.5 for region 2).

Table 2. Percentage of Correct Identification for Different Sill Values in Simulation 4

Sill parameter (region 1)	*Sill parameter (region 2)*	*Percentage of times with correct identification*
1	2.5	47.21
1	5	69.85
1	7.5	77.32

Table 3. Comparison of SGP and PGP for Simulation 4

Sill parameter (region 1)	*Sill parameter (region 2)*	*MSPR: SGP, PGP*
1	2.5	1.69, 1.38
1	5	2.91, 1.89
1	7.5	4.21, 2.57

to capture these sharp changes in the spatial structure (Barman, Sharma, Walker, and Datta-Gupta 1998).

The Schneider Buda field is located in Wood County, Texas. The reservoir rock type is Cretaceous reefal limestone. The field structure is an anticline, 10 km × 8 km, with the major axis north–south trending. The field produces from four zones: A, B, C, and D. The B zone is the main producing zone, yielding most of the cumulative oil production. The field has been divided into 12 areas for operations management. This analysis here focuses on area 05. Furthermore, we concentrate on permeability (Schon 1996; Barman et al. 1998; Lee and Datta-Gupta 1999), because this is a vital (and the hardest to determine) property for all reservoir problems, controling whether or not the rock can deliver or transmit fluids. More precise information on this has been given by Peddibhotla, Cubillos, Datta-Gupta, and Wu (1996). This type of reservoir can exhibit permeability variations of several orders of magnitude within a relatively short distance because of changes in the lithologic and diagenetic environment. In addition, reservoir properties can vary sharply and abruptly across faults caused by, for example, tectonic activities and also across stratigraphic zones caused by changes in the depositional environment or other geologic events.

The permeability is measured in different spatial locations of this field and is expressed in the unit mD, where 1 mD = 10^{-3} Darcy = 10^{-12} m^2. The high permeability values are associated with the center of the fluvial channel deposits, and the low permeability values correspond to a poor-quickly rock characterized as alluvial plain lithofacies (Canas 1992).

An initial exploratory analysis of the data suggested that the response had non-Gaussian form with heavy tails. We then considered a Box–Cox transformation of the data with two settings of the transformation parameter, λ, namely $\lambda = 1/2$ (square root transformation) and $\lambda = 0$ (log transformation). In particular, we amended our MCMC algorithm to include a jump proposal to sample this parameter value during the simulation (using a prior probability .5 on each of the two transforms). This allows the data to decide on the appropriate form. We found that the data slightly favored the square root transformation, with posterior probability .6. In the analysis and results that follow, we condition on the use of the square root transformation for ease of exposition. Clearly, we could use a mixture over the two transforms or allow for a continuous range of Box–Cox transformations; see Section 6.

Figure 9 presents a bubble plot of the square root–transformed that shows huge heterogeneities in permeabilities within local regions. Due to this heterogeneity and the inherent spatial dependence, we believe that a piecewise stationary process will be useful in modeling the permeability field acurately, which is the basis of the reservoir modeling.

As our practical usage, we adopt the general (power) exponential isotropic correlation function (Yaglom 1987; De Oliveira et al. 1997). In the following we use the second parameterization in (2), because it eases the assignment of noninformative priors and the MCMC methods. This is a flexible family containing the exponential ($\theta_2 = 1$) and the squared exponential ($\theta_2 = 2$) correlation functions as two of its members. It is easy to compute and is parameterized by physically interpretable quantities. The parameter θ_1 controls the range of correlation, viewed as the correlation between two observations one unit apart.

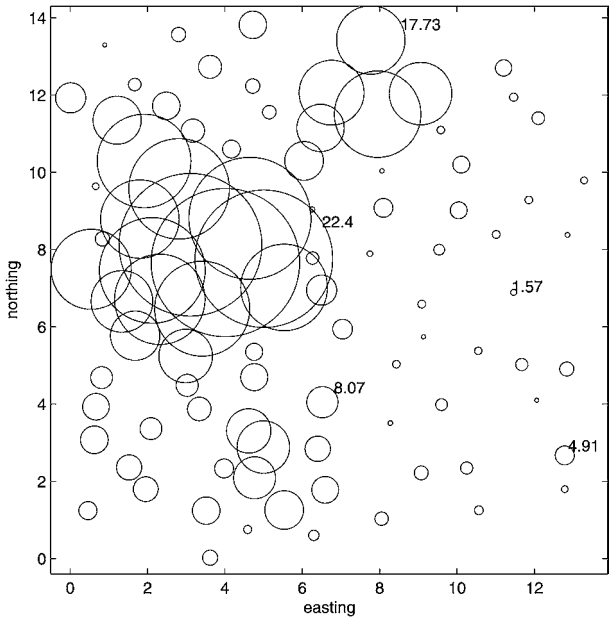

Figure 9. Bubble Plot of the Transformed Data.

Table 4 gives the probability distribution for the number of partitions, preserving every 50th partition after 20,000 burn-in iterations.

Figure 10 illustrates the posterior modal sample of the generated Voronoi tessellation with its centers (marked with •). We think that the result is physically meaningful, because petroleum reservoirs can be highly heterogeneous and exhibit nonstationary behavior of subsurface properties. The partitions divide the field into several high-permeability and low-permeability zones, which usually control the fluid flow in the reservoir. The formation of different permeability regions depends on depositional environment of rocks and subsequent processes like dolmalization. The deposition environment depends on the relative energy of the river that deposited; the flow of the river can change very quickly, which creates these sharp changes. The identification of the partitions (i.e., separation regions) are important for identifying the water barriers, and this information is used to design the oil recovery processes. We have performed some exploratory approaches using the method suggested by Handcock and Wallis (1994) to check the resulting Gaussian distribution assumption within each partition, using our square root transform. The method is based on a transformation of the spatially correlated variable to whitened residuals using the covariance matrix. These whitened residuals are independent, and we can use standard methods, like plots of marginal histograms, Q–Q plots, or the Kolmogorov–Smirnov (KS) test, to check the Gaussian assumption. We present the histograms and Q–Q plots in Figures 11, 12, and 13 for whitened residuals within each partition;

Table 4. Probability Distribution for Number of Regions

No. of regions	*Probability (%)*
1	2.1
2	11.2
3	65.3
4	21.4

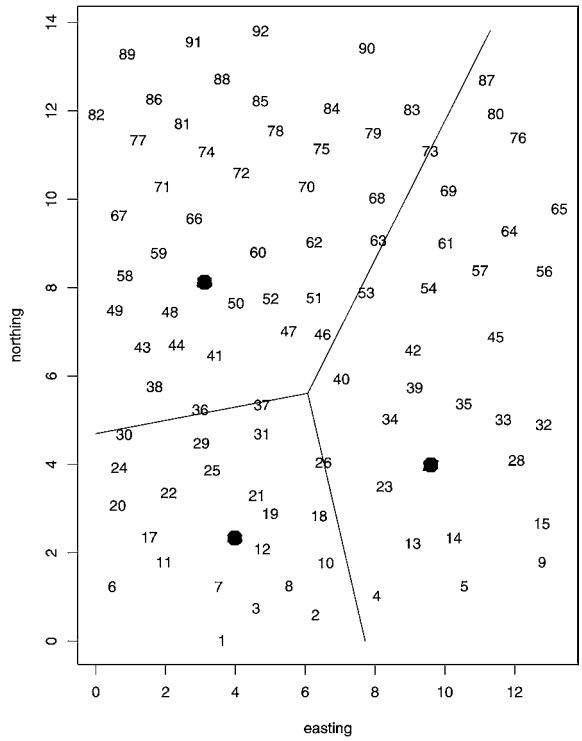

Figure 10. The Posterior Modal Sample of Voronoi Tessellation (with numbered locations).

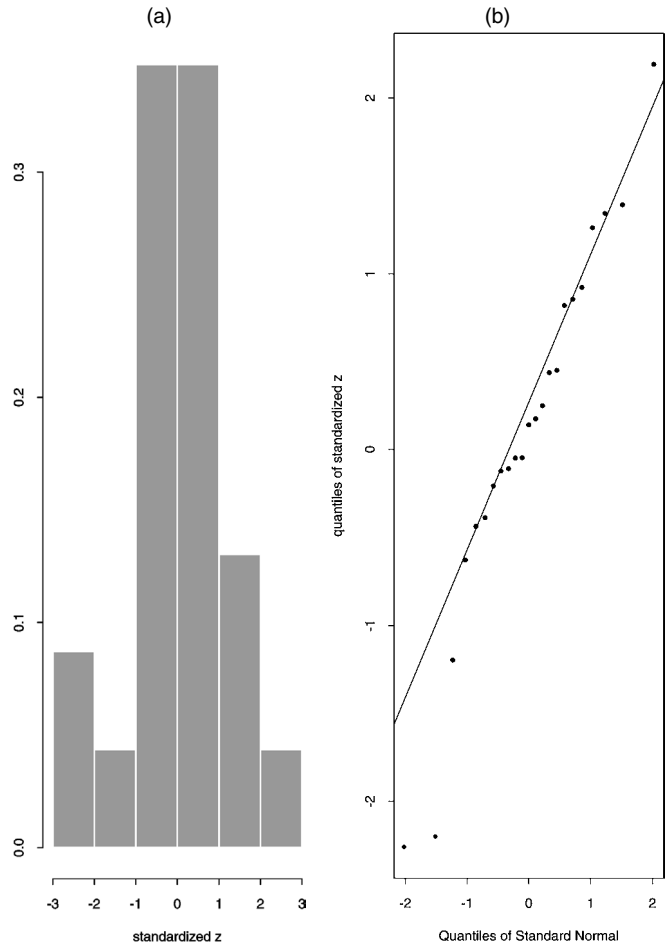

Figure 11. The (a) Histogram and (b) Q–Q Plot for the Whitened Residuals Within Partition 1 (southwest).

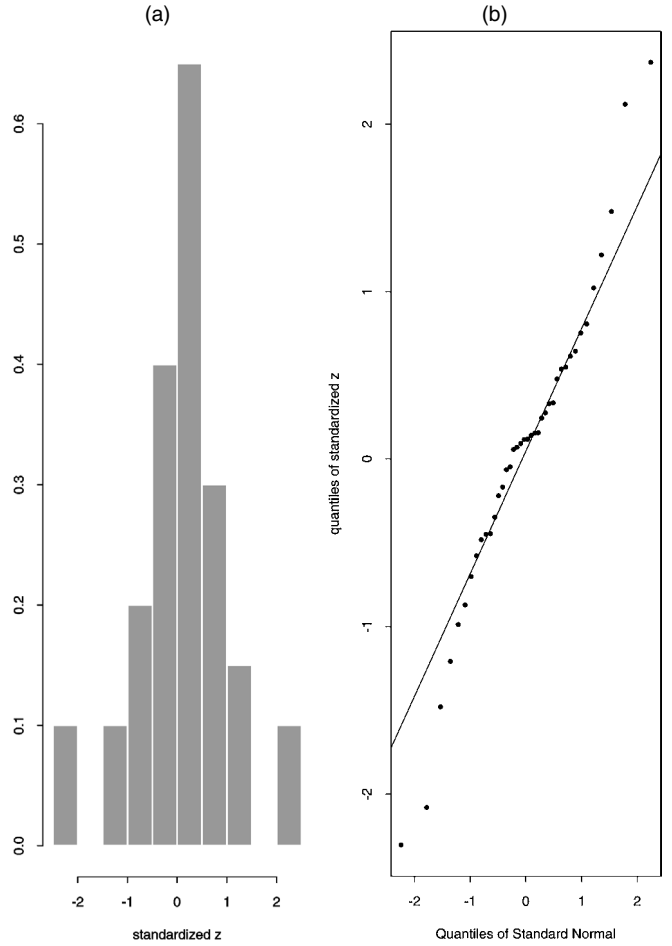

Figure 12. The (a) Histogram and (b) Q–Q Plot for the Whitened Residuals Within Partition 2 (northwest).

they look satisfactory, although there is still some evidence of kurtosis. We further performed KS tests for Gaussian distribution with KS test statistic values .13, .10, and .11 consecutively for the three partitions, and each time the assumption of Gaussian distribution failed to reject at the 5% level.

Due to nonstationarity, each location will have a different form of correlation structure. We obtained the posterior mean correlation surface for each point by averaging over correlation kernels for the different models in our MCMC output (many with different partition structures). We plotted the correlation surfaces for two location points, 33 and 46. Due to sharp nonstationarity, these surfaces are very different from each other. For the first plot in Figure 14, the point is well within the surface and looks like it can be modeled by a symmetric kernel. For the second location (a point near the partition) plotted in Figure 15, the correlation surface is highly asymmetric. The location itself is almost at the boundary of the surface, so there is a sharp change in the correlation structure in this situation. A symmetric kernel will fail to capture this situation and will tend to oversmooth the data. Our method is sufficiently flexible to model the kernels (or correlation surfaces) of different shapes automatically and data-adaptively. Prespecified regular shaped kernels may not be able to model these sorts of variation.

We present the following predictive results in square root scale, which is the working transformation of our model. Transforming them to the original scale can be done by simply

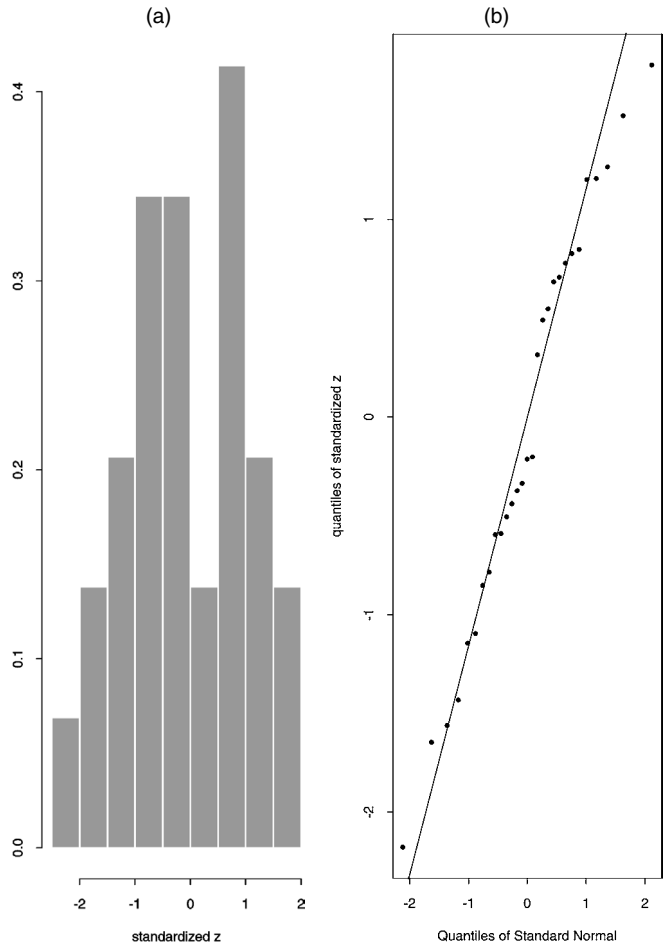

Figure 13. The (a) Histogram and (b) Q–Q Plot for the Whitened Residuals Within Partition 3 (east).

squaring the values obtained from the output of each MCMC sample. We use $\tilde{Z}$ to denote that we are working on $\sqrt{Z}$.

Mixing over different partition structures, we obtained the posterior predictive density function for $\tilde{Z}(\mathbf{x}_0)|\tilde{\mathbf{Z}}$ using the

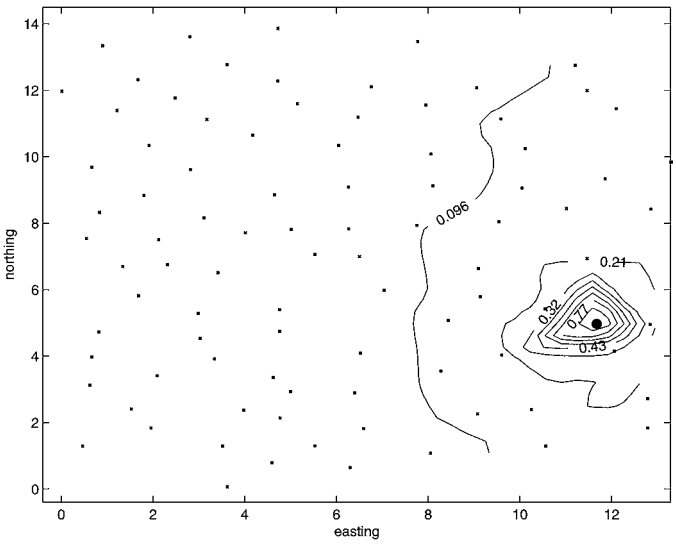

Figure 14. Marginal Correlation Surface for the 33rd Location (marked with a bullet).

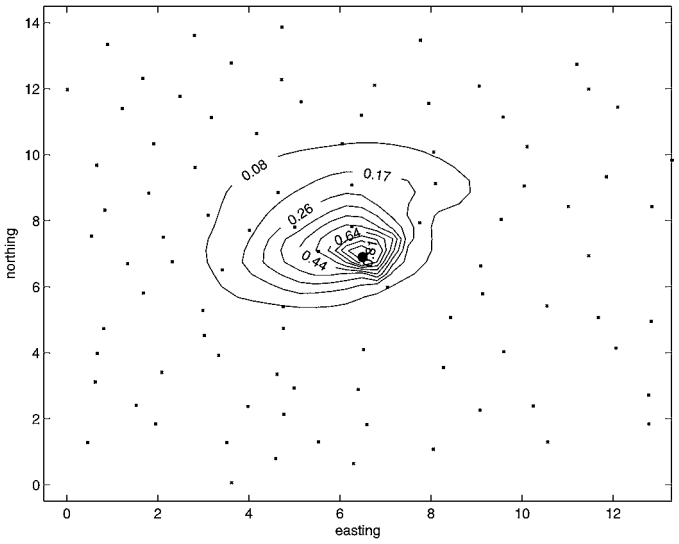

Figure 15. Marginal Correlation Surface for the 46th Location (marked with a bullet).

Metropolis–Hastings algorithm and the composition method (Tanner 1996; Gelfand 1996) for several locations covering different sections of the region of interest. The predicted positions (with numbering 92 locations) are plotted in Figure 16 (with positions shown by •). The posterior predictive densities are plotted in Figure 17. These posterior predictive densities have different shape and spread characteristics depending on the partition structures and the relative positions of $\mathbf{x}_0$, $\mathbf{x}_i$, $i = 1, \ldots, 92$. The 95% prediction intervals and means are summarized in Table 5.

Figure 18 shows the predicted map with a contour plot obtained by computing the mean of the posterior predictive density of $(\tilde{Z}(\mathbf{x}_0)|\tilde{\mathbf{Z}})$ on every $\mathbf{x}_0$ of a 40×40 grid. As a measure of

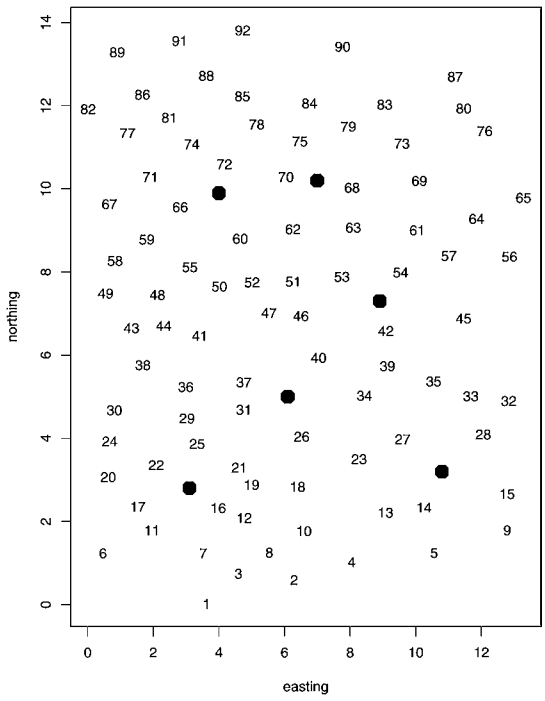

Figure 16. Predicted Positions With Numbering at 92 Locations.

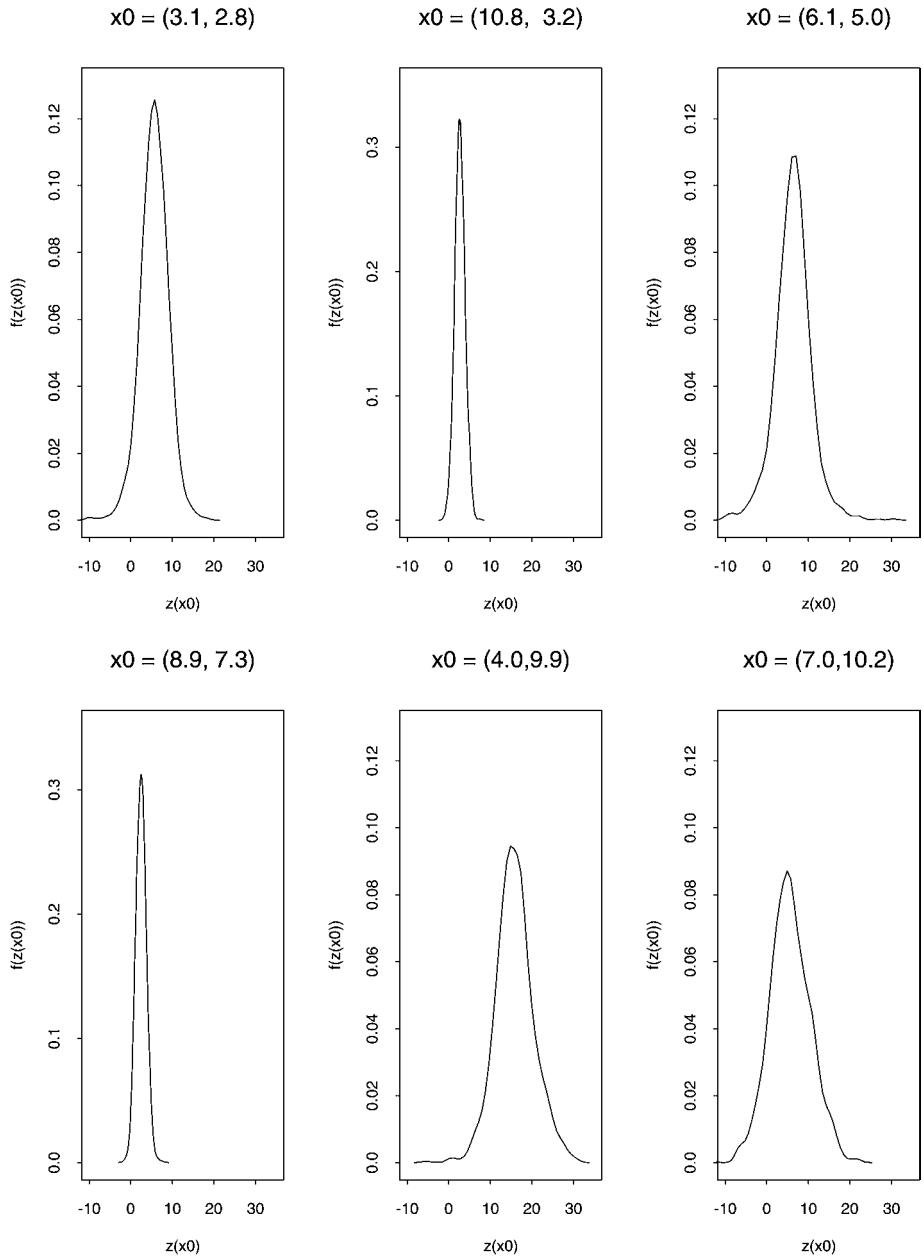

Figure 17. Predicted Densities for Different Locations for the Square Root of the Spatial Process.

predictive uncertainty at any location $\mathbf{s}_0$, we used the standard deviation. Figure 19 shows the map of this uncertainty measure with a contour plot.

We also plotted the predicted map generated from a single nonsmooth modal model in Figure 20. This is the mean of $(\tilde{Z}(\mathbf{x}_0)|\tilde{\mathbf{Z}}, \mathbf{t})$, fixing $\mathbf{t}$ at its modal value. It is clear that the predicted surface is smooth within each partition, but it is a step function between two different subregions.

Table 5. 95% Prediction Intervals and Mean of the Square Root of the Spatial Process

Location	*Prediction interval*	*Mean*
(3.1, 2.8)	(−1.31, 11.92)	5.67
(10.8, 3.2)	(.29, 5.12)	2.67
(6.1, 5.0)	(−2.75, 15.29)	6.29
(8.9, 7.3)	(.34, 4.79)	2.48
(4.0, 9.9)	(6.83, 25.42)	15.87
(7.0, 10.2)	(−3.85, 15.52)	5.58

We ran the analysis with two different prior settings with $\boldsymbol{\lambda}_i = 10^{-3}\mathbf{I}_{n_i}$, $\gamma_1 = \gamma_2 = .1$ and $\boldsymbol{\lambda}_i = 1 \cdot \mathbf{I}_{n_i}$, $\gamma_1 = 1$, $\gamma_2 = 7$. The results remain almost indistinguishable.

We can compute the evidence in favor of a partition model in terms of the Bayes factor. The Bayes factor is $B_{m0} = f(\tilde{\mathbf{z}}|\mathbf{x}, \mathbf{t}_m)/f(\tilde{\mathbf{z}}|\mathbf{x}, \mathbf{t}_0)$, where $f(\tilde{\mathbf{z}}|\mathbf{x}, \mathbf{t}_m)$ is the marginal likelihood of our chosen model and $f(\tilde{\mathbf{z}}|\mathbf{x}, \mathbf{t}_0)$ is the marginal likelihood of the model with no partition structure, which can be readily obtained from the acceptance probability (18). The calculated value of the Bayes factor is more than 100, which is interpretable as very strong evidence against no partition structure (Kass and Raftery 1995).

We also reanalyzed the data using the Matérn class of correlation functions used by Hancock and Stein (1993) and

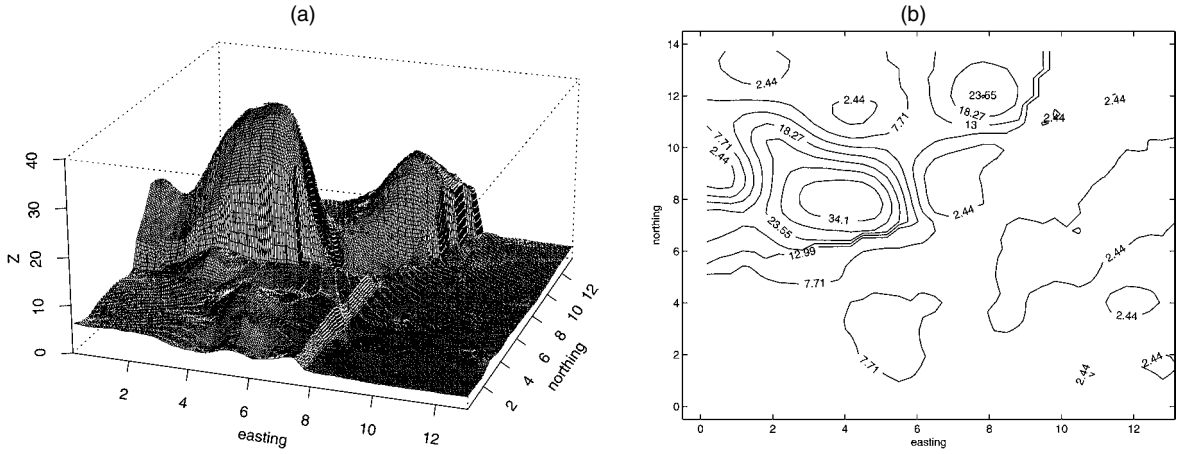

Figure 18. Predicted Map With a Contour Plot.

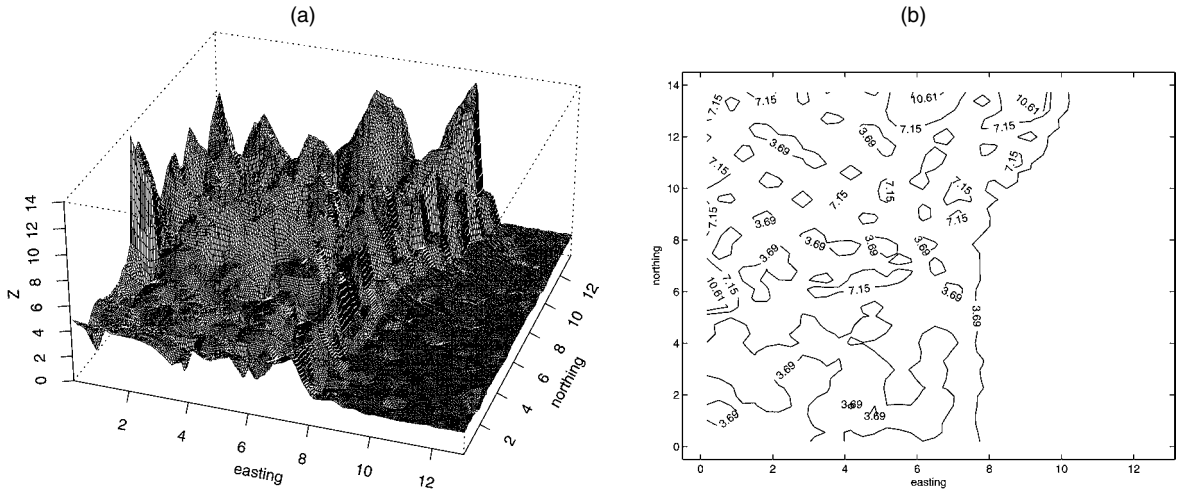

Figure 19. Uncertainty Map With a Contour Plot.

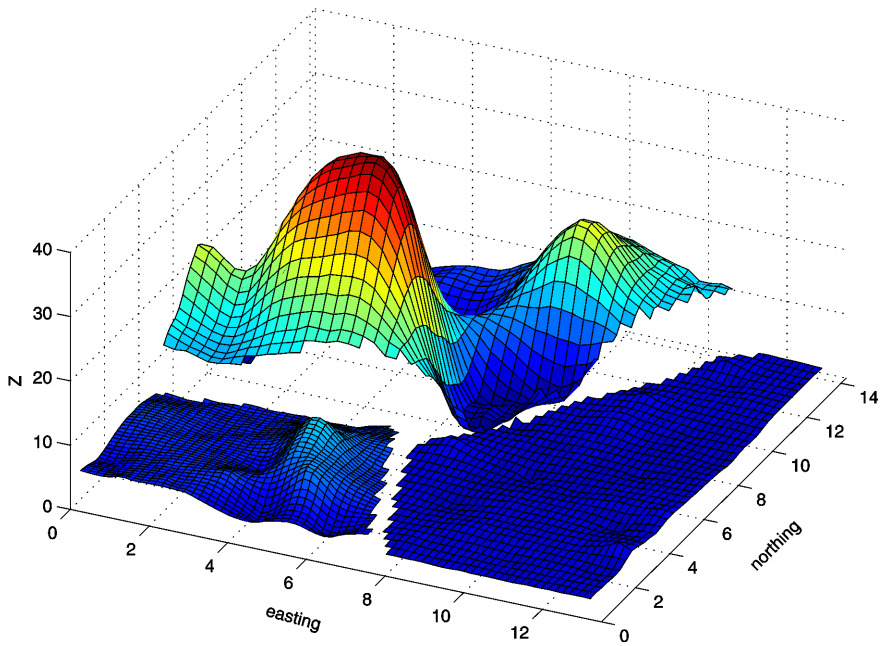

Figure 20. Single Nonsmooth Plot.

Handcock and Wallis (1994) and given by

$$K_{\boldsymbol{\theta}_i}(l) = \frac{1}{2^{\theta_2-1}\Gamma(\theta_2)}\left(\frac{l}{\theta_1}\right)^{\theta_2} B_{\theta_2}\left(\frac{l}{\theta_1}\right), \qquad \theta_1 > 0, \theta_2 > 0,$$

where $B_{\theta_2}(\cdot)$ is the modified Bessel function of the second type and of order θ_2. This model also supports the two-partitions (three-regions) model more than 65% of the time. For this particular dataset, the Matérn class model offers slightly improved MSPR at 27.73, compared with 31.89 for the exponential covariance model; however, it is computationally much more expensive due to the need for repeated calculations of the Bessel functions within each iteration.

6. DISCUSSION

We have extended the popular single Gaussian process model to piecewise Gaussian processes. The broad availability of fast, inexpensive computing helps us perform better predictions using this rich class of models.

If the partitions are curves rather than straight lines, then we can extend our model using additively weighted Voronoi models (Okabe, Boots, and Sugihara 1992). In these models we extend the definition of the region R_i as $R_i = \{\mathbf{x} \in D : \|\mathbf{x} - \mathbf{c}_i\| - w_i < \|\mathbf{x} - \mathbf{c}_j\| - w_j \ \forall j \neq i\}$, where the w_i's and w_j's are additional sets of parameters that will control the shape of the curve. We plot such a curved tessellation and compare it with an ordinary Voronoi tessellation in Figure 21. We can parameterize the model with $\alpha = |\mathbf{x}_i - \mathbf{x}_j|$ and $\beta = w_i - w_j$. The shape of the curvature at the boundary will depend on α and β. We can generate different curved boundaries depending on different values of α and β; in Figure 22 we show one example of this. In our Bayesian hierarchical model, we can treat these α's and β's as additional parameters and simulate them from the posterior distributions using the MCMC procedure.

A more flexible class of models is compoundly weighted Voronoi models, where the regions are defined as $R_i = \{\mathbf{x} \in D : w_{i1} \cdot \|\mathbf{x} - \mathbf{c}_i\| - w_{i2} < w_{j1} \cdot \|\mathbf{x} - \mathbf{c}_j\| - w_{j2} \ \forall j \neq i\}$. The additional set of parameters w_{i1}'s, and w_{i2}'s will have added flexibility to capture different type of curvature at the partitions.

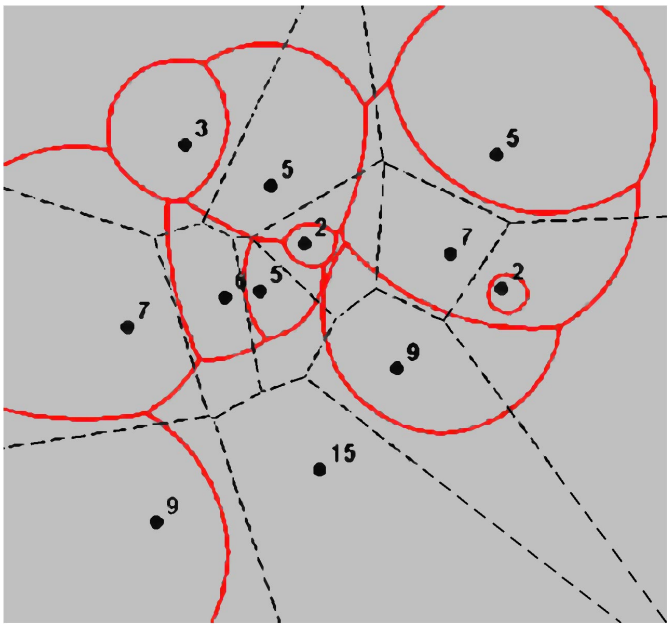

Figure 21. Plot of the Weighted Voronoi Tessellation (——) and an Ordinary Voronoi Tessellation (- - - - -).

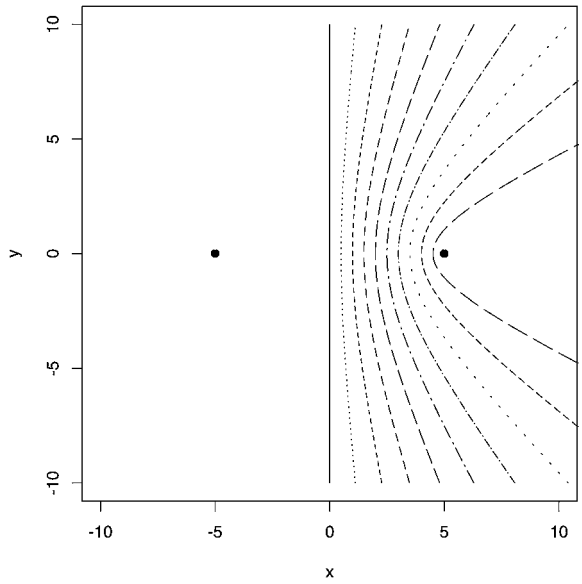

Figure 22. The Curved Boundary With the Additively Weighted Distance for Parameter Values $\alpha = 10$ and $\beta = 0$ (——); 1 (·····), 2 (- - - - -), 3 (– – –), 4 (— —), 5 (— · —), 6 (– · – · –), 7 (· · · ·), 8 (– – – –), and 9 (— – —).

One advantage of our procedure (as identified by one of the referees) is that we need only invert several covariance matrices of smaller dimension than the global one. This considerably aids computation on large datasets. We performed a simulation with sample size of 2,000 and 20 partitions with different combinations of means and covariance matrices. In the first simulation we kept all of the parameter values for the neighboring regions very distinct (similar to Simulation 3). Our method is successful in inferring the right number of partitions (the posterior probability of $M = 20$ is .87), and we avoid the problem of inverting a $2{,}000 \times 2{,}000$ matrix by inverting 20 100×100 covariance matrices. Following a suggestion from one of the referees, we then repeated our simulation with similar-looking parameter values for the neighboring regions (similar to Simulation 4). Still our method successfully identifies the right number of partitions as the posterior modal value of M is 20, but in this situation the posterior probability of $M = 20$ dropped to .22. We experienced slow movements of our MCMC sampler in this situation, and using adaptive parallel MCMC (Liang and Wong 2001) accelerated our algorithm considerably. A detailed discussion of piecewise Gaussian process models for large datasets with applications to data mining and corresponding efficient MCMC methods will appear in future work.

It may be reasonable to constrain our algorithm to contain some minimum number of points in each partition, so that we have adequate information for estimating the covariance function. Furthermore, the choice of a suitable correlation function is of clear interest. We can extend the general (power) exponential correlation function used in most of the examples to more complex ones. One possible choice of correlation function is the Matérn class of correlation functions as described in Section 5. Due to our flexible modeling structure through the number and location of partitions, our predictions may be less sensitive toward this choice.

The assumption of piecewise Gaussian distribution may not hold all of the time, and in those situations where it does not hold, we may need to use an excessive number of partitions to capture the process. To handle this situation, we may use a piecewise transformed Gaussian random field model. In this article we adopted a global transform, the square root. However, we could consider a continum of transforms seperate within each region. This would be an extension of the method of De Oliveria et al. (1997), using separate power transformations within each partition and estimating them using the data.

APPENDIX: DERIVATIONS OF MARGINAL LIKELIHOOD AND PREDICTIVE DENSITY

A.1 Marginal Likelihood for the Model

We have that

$$\begin{aligned} f(\mathbf{z}|\mathbf{x},\mathbf{t}) &= \int_{\boldsymbol{\Omega}} f(\mathbf{z}|\mathbf{x},\boldsymbol{\omega},\mathbf{t})\pi(\boldsymbol{\omega}|\mathbf{t})\,d\boldsymbol{\omega} \\ &= \int_{\boldsymbol{\Omega}} \prod_{i=1}^{M} (2\pi\sigma_i^2)^{-(n_i+p_i)/2} |\boldsymbol{\lambda}_i|^{1/2} \left|\mathbf{K}_{\boldsymbol{\theta}_i}\right|^{-1/2} \\ &\quad \times \exp\left(-\frac{1}{2\sigma_i^2}\left\{\boldsymbol{\beta}_i'\boldsymbol{\lambda}_i\boldsymbol{\beta}_i + (\mathbf{z}_i - \mathbf{G}_i\boldsymbol{\beta}_i)'\mathbf{K}_{\boldsymbol{\theta}_i}^{-1}(\mathbf{z}_i - \mathbf{G}_i\boldsymbol{\beta}_i)\right\}\right) \\ &\quad \times \pi(\sigma_i^2|\mathbf{t})\pi(\boldsymbol{\theta}_i|\mathbf{t})\,d\boldsymbol{\omega}, \end{aligned}$$

because the prior distribution of $\boldsymbol{\beta}_i|\mathbf{t}, \sigma_i^2$ is defined as $\mathrm{N}_{p_i}(\mathbf{0}, \sigma_i^2\boldsymbol{\lambda}_i^{-1})$, $i = 1, \ldots, M$, and $|\cdot|$ denotes the determinant of a matrix. After rearranging the terms containing $\boldsymbol{\beta}_i$, we find that the distribution of $\boldsymbol{\beta}_i$ in the integration follows $\mathrm{N}_{p_i}(\mathbf{V}_i\mathbf{G}_i'\mathbf{K}_{\boldsymbol{\theta}_i}^{-1}\mathbf{z}_i, \sigma_i^2\mathbf{V}_i)$, where $\mathbf{V}_i = (\mathbf{G}_i'\mathbf{K}_{\boldsymbol{\theta}_i}^{-1}\mathbf{G}_i + \boldsymbol{\lambda}_i)^{-1}$. Thus

$$\begin{aligned} f(\mathbf{z}|\mathbf{x},\mathbf{t}) = \prod_{i=1}^{M} \frac{|\boldsymbol{\lambda}_i|^{1/2}}{(2\pi)^{n_i/2}} \int_{\boldsymbol{\theta}_i}\int_{\sigma_i^2} \left|\mathbf{K}_{\boldsymbol{\theta}_i}\right|^{-1/2} |\mathbf{V}_i|^{1/2} (\sigma_i^2)^{-n_i/2} \\ \times \exp\left(-\frac{1}{2\sigma_i^2}\alpha_i\right)\pi(\sigma_i^2|\mathbf{t})\pi(\boldsymbol{\theta}_i|\mathbf{t})\,d\sigma_i^2\,d\boldsymbol{\theta}_i, \end{aligned}$$

where $\alpha_i = \mathbf{z}_i'(\mathbf{K}_{\boldsymbol{\theta}_i} + \mathbf{G}_i\boldsymbol{\lambda}_i^{-1}\mathbf{G}_i')^{-1}\mathbf{z}_i$. Furthermore, the distribution of σ_i^2 in the integration follows $\mathrm{IG}(\frac{\gamma_1+n_i}{2}, \frac{\gamma_2+\alpha_i}{2})$, and because the prior distribution of $\sigma_i^2|\mathbf{t}$ is assumed to be $\mathrm{IG}(\frac{\gamma_1}{2}, \frac{\gamma_2}{2})$, $i = 1, \ldots, M$, the result follows.

A.2 Predictive Density

After integrating out all other parameters within the model, we get

$$\begin{aligned} f(\mathbf{z}(\mathbf{x}_0)|\mathbf{z}_i,\mathbf{t}) \propto \int_{\boldsymbol{\theta}_i} [ss(\mathbf{z}_i^*) + \gamma_2]^{-(n_i+\gamma_1+1)/2} \\ \times \left|\mathbf{K}_{\boldsymbol{\theta}_i}^*\right|^{-1/2} \left|\mathbf{G}_i^{*\prime}\mathbf{K}_{\boldsymbol{\theta}_i}^{*-1}\mathbf{G}_i^* + \boldsymbol{\lambda}_i\right|^{-1/2} \pi(\boldsymbol{\theta}_i|\mathbf{t})\,d\boldsymbol{\theta}_i, \end{aligned}$$

where

$$\mathbf{G}_i^* = \begin{pmatrix} \mathbf{g}_i'(\mathbf{x}_0) \\ \mathbf{G}_i \end{pmatrix},$$

$$\mathbf{K}_{\boldsymbol{\theta}_i}^* = \begin{pmatrix} \mathbf{k}_{\boldsymbol{\theta}_i}(\mathbf{x}_0,\mathbf{x}_0) & \mathbf{k}_{\boldsymbol{\theta}_i}' \\ \mathbf{k}_{\boldsymbol{\theta}_i} & \mathbf{K}_{\boldsymbol{\theta}_i} \end{pmatrix},$$

$$ss(\mathbf{z}_i^*) = (\mathbf{z}_i^* - \mathbf{G}_i^*\hat{\boldsymbol{\beta}}_i)'\mathbf{K}_{\boldsymbol{\theta}_i}^{*-1}(\mathbf{z}_i^* - \mathbf{G}_i^*\hat{\boldsymbol{\beta}}_i), \qquad \text{and}$$

$$\mathbf{Z}_i^* = \begin{pmatrix} Z(\mathbf{x}_0) \\ \mathbf{Z}_i \end{pmatrix}.$$

Using the well-known matrix theorems (Henderson and Searle 1981), we can prove the following:

$$\begin{aligned} \left|\mathbf{K}_{\boldsymbol{\theta}_i}^*\right| &= \left|\mathbf{K}_{\boldsymbol{\theta}_i}\right| \left|\mathbf{k}_{\boldsymbol{\theta}_i}(\mathbf{x}_0,\mathbf{x}_0) - \mathbf{k}_{\boldsymbol{\theta}_i}'\mathbf{K}_{\boldsymbol{\theta}_i}^{-1}\mathbf{k}_{\boldsymbol{\theta}_i}\right| \\ &= \left|\mathbf{K}_{\boldsymbol{\theta}_i}\right| \left|\mathbf{K}_{11.2_i}\right|, \\ \left|\mathbf{G}_i^{*\prime}\mathbf{K}_{\boldsymbol{\theta}_i}^{*-1}\mathbf{G}_i^* + \boldsymbol{\lambda}_i\right| &= \left|\mathbf{G}_i'\mathbf{K}_{\boldsymbol{\theta}_i}^{-1}\mathbf{G}_i\right| \\ &\quad \times \left|\mathbf{I} + \left(\mathbf{G}_i'\mathbf{K}_{\boldsymbol{\theta}_i}^{-1}\mathbf{G}_i\right)^{-1}\left\{\mathbf{P}_i'\mathbf{K}_{11.2_i}^{-1}\mathbf{P}_i + \boldsymbol{\lambda}_i\right\}\right|, \\ ss(\mathbf{z}_i^*) &= ss(\mathbf{z}_i) \\ &\quad \times \left[1 + \left(\mathbf{z}(\mathbf{x}_0) - \hat{\boldsymbol{\mu}}_{1.2_i}\right)'\frac{\boldsymbol{\Psi}_i^{-1}}{ss(\mathbf{z}_i)}\left(\mathbf{z}(\mathbf{x}_0) - \hat{\boldsymbol{\mu}}_{1.2_i}\right)\right], \end{aligned}$$

and

$$\Psi_i^{-1/2} = |\boldsymbol{\Psi}_i|^{-1/2} = \left|\mathbf{K}_{11.2_i} + \mathbf{P}_i\left(\mathbf{G}_i'\mathbf{K}_{\boldsymbol{\theta}_i}^{-1}\mathbf{G}_i\right)^{-1}\mathbf{P}_i'\right|^{-1/2}.$$

Hence the result follows. This proof is similar to the proof of Le and Zidek (1992), who used a conjugate Wishart distribution for the covariance matrix.

[Received February 2001. Revised September 2004.]

REFERENCES

Barman, I., Sharma, A. K., Walker, R. F., and Datta-Gupta, A. (1998), "Permeability Predictions in Carbonate Reservoirs Using Optimal Non-Parametric Transformations: An Application at the Salt Creek Field Unit, Kent County, TX," SPE 39667, presented at the 1998 SPE/DOE improved Oil Recovery Symposium, Tulsa, Oklahoma, April, 1998.

Berntsen, J., Espelid, T. O., and Genz, A. (1991a), "An Adaptive Algorithm for the Approximate Calculation of Multiple Integrals," *ACM Transactions on Mathematical Software*, 17, 437–451.

——— (1991b), "An Adaptive Multidimensional Integration Routine for a Vector of Integrals," *ACM Transactions on Mathematical Software*, 17, 452–456.

Brown, P. J., Le, N. D., and Zidek, J. V. (1994), "Multivariate Spatial Interpolation and Exposure To Air Pollutants," *Canadian Journal of Statistics*, 22, 489–505.

Canas, J. A. (1992), "Use of a Hydraulic Interwell Connectivity Concept for Sandstone Reservoir Characterization," unpublished masters thesis, Texas A&M University.

Cohen, A., and Jones, R. H. (1969), "Regression on a Random Field," *Journal of the American Statistical Association*, 64, 1172–1182.

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

Creutin, J. D., and Obled, C. (1982), "Objective Analysis and Mapping Techniques for Rainfall Fields: An Objective Comparison," *Water Resources Research*, 18, 413–431.

De Oliveira, V., Kedem, B., and Short, D. S. (1997), "Bayesian Prediction of Transformed Gaussian Random Fields," *Journal of the American Statistical Association*, 92, 1422–1433.

Draper, D. (1995), "Assessment and Propagation of Model Uncertainty" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 57, 45–97.

Dromgoole, P., and Speers, R. (1997), "Geoscore: A Method for Quantifying Uncertainty in Field Reserve Estimates," *Petroleum Geoscience*, 3, 1–12.

Fuentes, M. (2001), "A New High-Frequency Kriging Approach for Nonstationary Environmental Processes," *Environmetrics*, 12, 469–483.

Fuentes, M., and Smith, R. (2001), "A New Class Of Nonstationary Spatial Models," technical report, North Carolina State University.

Gelfand, A. E. (1996), "Model Determination Using Sampling-Based Methods," in *Markov Chain Monte Carlo in Practice*, eds. W. R. Gilks, R. Richardson, and D. J. Spiegelhalter, Washington, DC: Chapman & Hall, pp. 145–161.

Gelfand, A. E., Dey, D. K., and Chang, H. (1992), "Model Determination Using Predictive Distributions With Implementation via Sampling-Based Methods," *Bayesian Statistics*, 4, 147–167.

Green, P. J. (1995), "Reversible-Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, 82, 711–732.

Green, P. J., and Sibson, R. (1978), "Computing Dirichlet Tessellations in the Plane," *The Computer Journal*, 21, 168–173.

Guttorp, P., and Sampson, P. (1994), "Methods for Estimating Heterogeneous Spatial Covariance Functions With Environmental Applications," in *Handbook of Statistics 12*, eds. G. P. Patil and C. R. Rao, New York: Elsevier Science, pp. 661–689.

Haas, T. C. (1990), "Kriging and Automated Variogram Modeling Within a Moving Window," *Atmospheric Environment*, 24A, 1759–1769.

——— (1995), "Local Prediction of a Spatio-Temporal Process With an Application to Wet Sulfate Deposition," *Journal of the American Statistical Association*, 90, 1189–1199.

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

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

Henderson, H. V., and Searle, S. R. (1981), "On Deriving the Inverse of a Sum of Matrices," *SIAM Review*, 23, 53–60.

Higdon, D. (2001), "Space and Space–Time Modeling Using Process Convolutions," technical report, Duke University.

Higdon, D., Swall, J., and Kern, J. (1999), "Nonstationary Spatial Modeling," in *Bayesian Statistics 6*, eds. J. M. Bernardo et al., Oxford, U.K.: Oxford University Press, pp. 761–768.

Kaluzny, S. P., Vega, S. C., Cardoso, T. P., and Shelly, A. A. (1997), *S+SpatialStats*, New York: Springer-Verlag.

Kass, R. E., and Raftery, A. E. (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–795.

Le, N. D., and Zidek, J. V. (1992), "Interpolation With Uncertain Spatial Covariances: A Bayesian Alternative to Kriging," *Journal of Multivariate Analysis*, 43, 351–374.

Lee, S. H., and Datta-Gupta, A. (1999), "Electrofacies Characterization and Permeability Predictions in Carbonate Reservoirs: Role of Multivariate Analysis and Nonparametric Regression," SPE 56658, presented at the 1999 SPE Annual Technical Conference and Exhibition, Houston, Texas, October 1999.

Liang, F., and Wong, W. H. (2001), "Real-Parameter Evolutionary Sampling With Application in Bayesian Mixture Models," *Journal of the American Statistical Association*, 96, 653–666.

Nott, D. J., and Dunsmuir, T. M. (2002), "Estimation of Nonstationary Spatial Covariance Structure," *Biometrika*, 89, 819–829.

Nychka, D., Wikle, C., and Royle, J. A. (1999), "Large Spatial Prediction Problems and Nonstationary Random Fields," preprint, Geophysical Statistical Program, National Center for Atmospheric Research.

Okabe, A., Boots, B., and Sugihara, K. (1992), *Spatial Tessellations*, New York: Wiley.

Peddibhotla, S., Cubillos, H., Datta-Gupta, A., and Wu, C. H. (1996), "Rapid Simulation of Multiphase Flow Through Fine-Scale Geostatistical Realizations Using a New 3-D Streamline Model: A Field Example," SPE 36008, presented at PCC, Dallas, June 1996.

Sampson, P. D., Damian, D., and Guttorp, P. (2001), "Advances in Modeling and Inference for Environmental Processes With Nonstationary Spatial Covariance," in *GeoENV 2000: Geostatistics for Environmental Applications*, eds. E. Monesteiz, D. Allard, and R. Froidvaux, Dordrecht: Kluwer Academic, pp. 17–32.

Sampson, P. D., and Guttorp, P. (1992), "Nonparametric Estimation of Nonstationary Spatial Covariance Structure," *Journal of the American Statistical Association*, 87, 108–119.

Schmidt, A. M., and O'Hagan, A. (2003), "Bayesian Inference for Nonstationary Spatial Covariance Structure via Spatial Deformations," *Journal of Royal Statistical Society*, Ser. B, 65, 743–758.

Schon, J. H. (1996), *Physical Properties of Rocks: Fundamentals and Principles of Petrophysics*, New York: Pergamon Press.

Smith, R. L. (1996), "Estimating Nonstationary Spatial Correlation," preprint, University of North Carolina.

Stein, M. L. (1999), *Statistical Interpolation of Spatial Data: Some Theory for Kriging*, New York: Springer-Verlag.

Tanner, M. A. (1996), *Tools for Statistical Inference*, New York: Springer-Verlag.

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