



A comparative study of Gaussian geostatistical models and Gaussian Markov random field models[☆]

Hae-Ryoung Song^a, Montserrat Fuentes^{b,*}, Sujit Ghosh^b

^a *Biostatistics Department, University of South Carolina, Columbia, SC, United States*

^b *Department of Statistics, North Carolina State University, United States*

Received 12 September 2006
Available online 7 February 2008

Abstract

Gaussian geostatistical models (GGMs) and Gaussian Markov random fields (GMRFs) are two distinct approaches commonly used in spatial models for modeling point-referenced and areal data, respectively. In this paper, the relations between GGMs and GMRFs are explored based on approximations of GMRFs by GGMs, and approximations of GGMs by GMRFs. Two new metrics of approximation are proposed: (i) the Kullback–Leibler discrepancy of spectral densities and (ii) the chi-squared distance between spectral densities. The distances between the spectral density functions of GGMs and GMRFs measured by these metrics are minimized to obtain the approximations of GGMs and GMRFs. The proposed methodologies are validated through several empirical studies. We compare the performance of our approach to other methods based on covariance functions, in terms of the average mean squared prediction error and also the computational time. A spatial analysis of a dataset on PM_{2.5} collected in California is presented to illustrate the proposed method.

© 2008 Elsevier Inc. All rights reserved.

AMS 2000 subject classifications: 91B76; 86A32; 62H11; 91D72; 60J20

1. Introduction

In many applications in spatial and environmental epidemiology, data concerning a spatial process of interest are often observed at different spatial resolutions, and the overall problem

[☆] The research conducted by Fuentes has been partly supported by a National Science Foundation grants DMS 0353029 and DMS 0706731, the National Institutes of Health award R01 ES01 14884301A2, and the Environmental Protection Agency award R833863. Song has been supported by a cooperative agreement (CT829562) with the Office of Air Quality Planning and Standards/US Environmental Protection Agency.

* Corresponding author.

E-mail addresses: hrsong@gwm.sc.edu (H.-R. Song), fuentes@stat.ncsu.edu (M. Fuentes).

of incompatible spatial data has been encountered very commonly when relating two spatial variables with different supports. For example, in studies of the association between air pollution exposure and adverse health effects, relevant health outcomes are usually available as areal data due to confidentiality while the pollution data are available as a point level [6,19]. To investigate the relationships between two variables with different spatial resolutions, the mismatch problem in the support of the two variables needs to be resolved. One common solution to this spatial misalignment problem is to aggregate the point-referenced data to the area level, and create a common support for both variables. Once the point-referenced data are aggregated to the relevant level, the process representing the aggregated data is modeled using integrals of spatial continuous process [14,15]. Consider a continuous Gaussian process $Y(s)$ with mean function $\mu(s)$ and covariance function $c(s_i, s_j)$ for $s_i, s_j \subset D \in \mathbb{R}^d$, where s_i is a location in a fixed domain D . The aggregated process over a region B , $Y(B) = \int_B Y(s) ds$ has a multivariate normal distribution with mean function $\mu(B)$ and covariance function $\Sigma(B_1, B_2)$,

$$\mu(B) = E(Y(B)) = |B|^{-1} \int_B \mu(s) ds$$

$$\Sigma(B_1, B_2) = \text{cov}(Y(B_1), Y(B_2)) = |B_1|^{-1} |B_2|^{-1} \int_{B_1} \int_{B_2} c(s_1, s_2) ds_1 ds_2,$$

where $|B_i|$ denotes the area of a region B_i for $i = 1, 2$. Modeling aggregated data using these spatial integrals requires lots of computation. Hence, instead of using the aggregated models directly in modeling aggregated point-referenced data, it is becoming very common to use Gaussian Markov random fields.

The modeling of aggregated point-referenced data using GMRFs serves as one of our motivations to investigate the relations between GMRFs and GGMs. GGMs are used in modeling a process over a domain based upon a set of measurements taken at a finite number of sites in the domain. On the other hand, GMRFs are widely used for modeling areal data measured as averages of a quantity over subregions of a larger study region. There have been several attempts in the literature to explore the relationship between GGMs and GMRFs. Besag [3] showed that covariance function of GMRFs could be approximately represented in terms of a modified Bessel function that decreases monotonically with distance, which suggested the possibility of approximations of GMRFs by GGMs. Griffith and Csillag [12] investigated the approximation of GMRFs using several geostatistical models by minimizing the squared differences between the covariances of GMRFs and GGMs. Hrafnkelsson and Cressie [13] explored a relationship between a specific class of GMRFs and a Matérn covariance model through an empirical parametrization of the relations between parameters of two models. On the other hand, Rue and Tjelmeland [16] investigated approximations of GGMs by GMRFs using two approximation methods of GGMs by GMRFs, one using the Kullback–Leibler (KL) discrepancy between probability densities and the other one by matching the correlation functions of the models, showing that the matching correlation approach performed better than the KL method.

From these earlier studies, it appears that one of the key elements of this comparative study is the choice of the metric to measure the discrepancy between GGMs and GMRFs. Although various metrics are suggested in previous research, they are mainly based on covariance functions of models and, no evaluations have been conducted on suggested metrics. In this paper, we introduce two new metrics using spectral density functions to explore the relations between GGMs and GMRFs, and compare our metrics and covariance metrics based on the average mean squared prediction error and computation time. The aim of this study is to explore

the relations between GGMs and GMRFs through a variety of empirical studies based on our suggested metrics. In Section 2, we present general overviews of GGMs and GMRFs. In Section 3, we propose new approximation metrics using spectral density functions. In Section 4, we compare the performance of our spectral method to the other covariance-based type of approaches, using as criteria the averaged mean squared prediction error and also taking into account the computational time. Section 5 presents an illustration of our methods using PM_{2.5} data. Section 6 summarizes our findings and suggests possible extensions of this research.

2. Spatial models

In this section, we briefly review two general classes of spatial models, GGMs and GMRFs. The choice of models usually depends on the data types whether the data are collected as points in space or are observed as averages over regions. GGMs (e.g. [10]) are usually used in modeling geostatistical data such as measurements on several attributes at point-referenced spatial locations, s_1, \dots, s_n in a fixed region D where the measurement points vary continuously, while GMRFs (e.g. [2]) are used for modeling areal data which are collected over a certain region. These models specify the process in a different way. GGMs model spatial associations directly through the parametric covariance models while GMRFs specify spatial associations through the conditional specification of the precision matrix.

2.1. Gaussian geostatistical models (GGMs)

GGMs are used to model continuous spatial Gaussian processes using a spatial covariance that is often just a function of distance and direction between locations. GGMs are generally based on two common assumptions which are *second order stationarity* and *isotropy*. Second order stationarity implies that the mean of a process is constant and the covariance function depends on the spatial vector distance between two locations. When the covariance function only depends on the Euclidean distance (no direction) between two locations, the process is called isotropic. The spatial covariance of a stationary and isotropic spatial process could be modeled using parametric functions of Euclidean distances. The Matérn covariance is one of the commonly used parametric covariance functions, which is defined as,

$$C(h) = \begin{cases} \sigma^2 \frac{1}{2^{v-1} \Gamma(v)} \left(\frac{2\sqrt{v}h}{\phi} \right)^v 2K_v \left(\frac{2\sqrt{v}h}{\phi} \right) & \text{if } h > 0 \\ \tau^2 + \sigma^2 & \text{if } h = 0 \end{cases}$$

where h denotes the Euclidean distance between two points, K_v is a modified Bessel function of the third kind [1] of order v , τ^2 , σ^2 and ϕ represent the nugget effect, partial sill and effective range of the covariance, respectively. The nugget effect is a discontinuity at the origin due to microscale variations and measurement errors. The sill parameter is the variance of the process, and the partial sill is the sill minus the nugget effect. The effective range is the distance where the correlation drops below 0.05. The order $v > 0$ of the K_v function controls the smoothness of the function. In particular, when $v = 1/2$, the Matérn covariance reduces to an exponential model and when $v \rightarrow \infty$, it approaches a Gaussian covariance model (e.g. [17]). Due to its flexibility, the Matérn covariance is widely used in a variety of spatial applications.

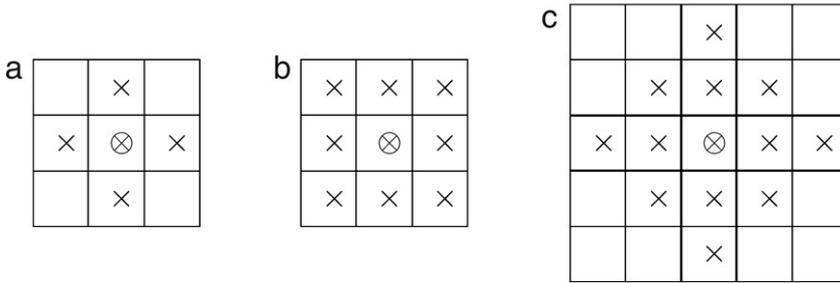


Fig. 1. (a) First order neighbors (b) Second order neighbors (c) Third order neighbors.

2.2. Gaussian Markov random models (GMRFs)

A GMRF is specified through full conditional based on the Markov property in space where the conditional distribution of data at a certain location given all of the other data depends only on the values in the neighborhood. Let Z_i denote a random variable observed at the site i . A GMRF is characterized by the following conditional distribution

$$Z_i | \{Z_j : j \neq i\} \sim N \left(\mu_i + \alpha \sum_{j=1}^n b_{ij}(z_i - \mu_j), \tau_i \right),$$

where μ_i is the mean of Z_i , and τ_i is the conditional variance of Z_i given $\{Z_j : j \neq i\}$. By Brook’s Lemma and the Hammersley–Clifford Theorem [2], the joint distribution of a GMRF $\mathbf{Z} = (Z_1, \dots, Z_n)^T$ is uniquely determined as,

$$\mathbf{Z} \sim N(\boldsymbol{\mu}, (\mathbf{I} - \alpha\mathbf{B})^{-1}\mathbf{M})$$

where $\boldsymbol{\mu}$ is an $n \times 1$ vector of means with elements μ_i , and \mathbf{M} is an $n \times n$ diagonal matrix with elements τ_i . The matrix \mathbf{B} is an $n \times n$ neighborhood matrix with elements b_{ij} , which satisfy $b_{ii} = 0, \sum_j b_{ij} = 1, b_{ij}\tau_j = b_{ji}\tau_i$ and $b_{ij} = 0$, unless i and j are neighbors. The parameter α is known as the spatial dependency parameter which somehow controls spatial dependence in the covariance. Specific choices of α lead to the covariance matrix being nonsingular. When $\alpha = 0$, the model becomes independent, and when $\alpha = 1$, the covariance matrix becomes singular and the joint distribution is improper. In particular, the model with $\alpha = 1$ is called an intrinsic GMRF, [5,4] which has been extensively used in spatial statistics as a prior for random effects in Bayesian hierarchical models [18,8].

A GMRF introduces spatial associations in the model through the specification of neighborhoods based on the arrangement of the regions in the map. We impose a neighborhood structure on a site i , and assign weights to the neighboring values, which are prespecified according to some criterion. In the case of a regular lattice, some popular neighborhood structures are defined as orders depending on the touching borders between grid cells. Some examples of neighborhoods with different orders are illustrated by Fig. 1. For irregular lattices, two sites can be considered a neighbor if they are within some specified distance of one another or they share a common boundary. The weights assigned to each neighborhood are determined in several ways. Common weight functions are binary functions with value 1 if two sites are neighbors and 0 otherwise, and scaled weights which are standardized by the row sum.

3. Spectral methods to measure discrepancy

Spectral densities and covariance functions are related through Fourier transforms, both of which are used to explore the second order properties of a spatial process. Although analyzing the variability of a process via the covariance function and the spectral density can be regarded as equivalent, they provide different ways of analyzing the process, and spectral analysis might offer some advantages compared to the analysis based on the covariance function. For instance, since spectral densities of stationary processes evaluated at different frequencies are approximately independent, the inference made based on the spectral density is usually easier to derive than by using the covariance function. In addition, the spectral density function can highlight subtle differences in the second order structure more easily than the covariance function [17]. Considering these advantages of the spectral approaches, we introduce approximation methods using spectral densities. One limitation of this method is that spectral analysis generally requires gridded data.

3.1. Spectral densities of spatial processes

Consider $Z(\mathbf{s})$ a stationary spatial process at location \mathbf{s} , defined on D where $D \subset R^2$. The process $Z(\mathbf{s})$ can be represented in terms of a spectral process X using a Fourier–Stieltjes integral,

$$Z(\mathbf{s}) = \int_{R^2} \exp(i\mathbf{w}^T \mathbf{s}) dX(\mathbf{w}),$$

where $\{X(\mathbf{w}), \mathbf{w} = (w_1, w_2) \in R^2\}$ is a process with zero mean and orthogonal increments, i.e.

$$E(X(\mathbf{w})) = 0$$

$$\text{Cov}(X(\mathbf{w}_3) - X(\mathbf{w}_2), X(\mathbf{w}_1) - X(\mathbf{w}_0)) = 0,$$

for disjoint rectangles $(\mathbf{w}_3, \mathbf{w}_2)$ and $(\mathbf{w}_1, \mathbf{w}_0)$. The process $X(\cdot)$ is called the spectral process associated to the stationary process $Z(\cdot)$.

Using the spectral process X , we define a monotonically increasing function $F(\mathbf{w})$ as

$$E(|dX(\mathbf{w})|^2) = dF(\mathbf{w}),$$

where $dX(\mathbf{w}) = \lim_{\mathbf{h} \rightarrow 0} \frac{(X(\mathbf{w}+\mathbf{h}) - X(\mathbf{w}))}{\mathbf{h}}$ and $dF(\mathbf{w}) = \lim_{\mathbf{h} \rightarrow 0} \frac{(F(\mathbf{w}+\mathbf{h}) - F(\mathbf{w}))}{\mathbf{h}}$. The function $F(\mathbf{w})$ represents the contribution to the total variance of the process by frequencies in the range $(\mathbf{0}, \mathbf{w})$. The derivative of $F(\mathbf{w})$ denoted by $f(\mathbf{w})$ (if it exists) is called the spectral density function of Z .

The covariance function $C(\mathbf{h})$ of a weakly stationary process and the spectral density function $f(\mathbf{w})$ form a Fourier transform pair,

$$C(\mathbf{h}) = \int_{R^2} \exp(i\mathbf{w}^T \mathbf{h}) f(\mathbf{w}) d\mathbf{w}$$

$$f(\mathbf{w}) = \frac{1}{(2\pi)^2} \int_{R^2} C(\mathbf{h}) \exp(-i\mathbf{w}^T \mathbf{h}) d\mathbf{h},$$

where $\{\mathbf{h} = \|\mathbf{s}_1 - \mathbf{s}_2\|, \mathbf{s}_1, \mathbf{s}_2 \in D \subset R^2\}$ [7,9,20]. If a stochastic process is defined on a lattice A , the integral in the expression of $f(\mathbf{w})$ is replaced by a sum, and the frequency domain is restricted to the open square $A = (-\pi, \pi]^2$,

$$C(\mathbf{h}) = \int_A \exp(i\mathbf{w}^T \mathbf{h}) f(\mathbf{w}) d\mathbf{w}$$

$$f(\mathbf{w}) = \frac{1}{(2\pi)^2} \sum_{\mathbf{h}} C(\mathbf{h}) \exp(-i\mathbf{w}^T \mathbf{h}).$$

When the underlying process is continuous, but observed only on a integer lattice Λ with an interval between neighboring observations δ , we cannot distinguish between the frequency components \mathbf{w} and $\mathbf{w} + \frac{2\pi\mathbf{l}}{\delta}$, for $\mathbf{l} \subset \{(l_1, l_2); l_1, l_2 = \pm 1, \pm 2, \dots\}$. This phenomenon is known as the aliasing effect, and the frequencies $\mathbf{w} + \frac{2\pi\mathbf{l}}{\delta}$ are called the aliases of \mathbf{w} for any \mathbf{w} in $(-\frac{\pi}{\delta}, \frac{\pi}{\delta}]^2$. Due to the aliasing effects, the power in the sampled process from the set of frequencies $\mathbf{w} \in (-\frac{\pi}{\delta}, \frac{\pi}{\delta})^2$ is the accumulation of power in the original process from all the sets of frequencies $\mathbf{w} + \frac{2\pi\mathbf{l}}{\delta}$. The spectral density of the sampled process is defined as follows,

$$f_{\delta}(\mathbf{w}) = \sum_{\mathbf{l} \in \mathbb{Z}^2} f\left(\mathbf{w} + \frac{2\pi\mathbf{l}}{\delta}\right),$$

where $\mathbf{w} \in (-\pi, \pi]^2$ and $\delta > 0$.

While the spectral density of a stationary process is represented by orthogonal increments, the corresponding spectral process of a nonstationary process is no longer orthogonal, but correlated. The Fourier transform pairs of a process defined on a lattice Λ are,

$$C(\mathbf{z}_1, \mathbf{z}_2) = \int_{R^4} \exp(i(\mathbf{z}_1 \mathbf{w}_1 - \mathbf{z}_2 \mathbf{w}_2)) f(\mathbf{w}_1, \mathbf{w}_2) d\mathbf{w}_1 d\mathbf{w}_2$$

$$f(\mathbf{w}_1, \mathbf{w}_2) = \frac{1}{(2\pi)^4} \sum_{\mathbf{z}_1, \mathbf{z}_2} C(\mathbf{z}_1, \mathbf{z}_2) \exp(-i(\mathbf{z}_1 \mathbf{w}_1 - \mathbf{z}_2 \mathbf{w}_2)),$$

where $\mathbf{z}_1, \mathbf{z}_2 \in \Lambda$ and $\mathbf{w}_1, \mathbf{w}_2 \in (-\pi, \pi]^2$. The nonstationary spectral density can be represented as the stationary spectral density for the special case in which the spectral density is zero for all the elements except for the ones along the diagonal ($\mathbf{w}_1 = \mathbf{w}_2$),

3.2. Approximation methods using spectral densities

We consider a continuous stationary and isotropic Gaussian spatial process on a $r \times c$ finite lattice Λ , with an interval δ between neighboring points, and define a GGM and a GMRF on this lattice. The GGM represents a weakly stationary and isotropic process, and the corresponding GMRF would be a nonstationary process due to the finiteness of the lattice and the edge effect. Since we define a GGM on a discrete lattice Λ , we observe only a sampled GGM on a lattice Λ . The spectral density of the sampled GGM g_l is defined as,

$$g_l(\mathbf{w}) = \sum_{\mathbf{Q} \in \mathbb{Z}^2} g\left(\mathbf{w} + \frac{2\pi\mathbf{Q}}{l}\right)$$

where $g(\mathbf{w}) = \frac{1}{(2\pi)^2} \int_{R^2} C(\mathbf{h}) \exp(-i\mathbf{w}^T \mathbf{h}) d\mathbf{h}$

for $\mathbf{w} \in (-\pi, \pi]^2$ and $Z = \{0, \pm 1, \pm 2, \dots\}$. Nonstationary processes corresponding to GMRFs defined on a lattice cannot be represented as a function with orthogonal increments but require correlated increments. The spectral density of a nonstationary process is

$$f(\mathbf{w}_1, \mathbf{w}_2) = \frac{1}{(2\pi)^4} \sum_{\mathbf{z}_1, \mathbf{z}_2 \subset \Lambda} C(\mathbf{z}_1, \mathbf{z}_2) \exp(-i(\mathbf{z}_1 \mathbf{w}_1 - \mathbf{z}_2 \mathbf{w}_2)),$$

where $\mathbf{w}_1, \mathbf{w}_2 \in (-\pi, \pi]^2$.

Using spectral densities, we suggest two approximation criteria; the KL discrepancy between two spectral densities and the chi-squared distance (CSD) between two spectral densities. We use these criteria to approximate the spectral density of a geostatistical process with the spectral density of a GMRF or vice versa by minimizing the differences of the spectral densities of two processes. The KL discrepancy between two spectral densities is defined as,

$$KL(f, \tilde{f}; \theta) = \int_{-\pi}^{\pi} f(\mathbf{w}) \log \frac{f(\mathbf{w})}{\tilde{f}(\mathbf{w}; \theta)} d\mathbf{w},$$

where $f(\mathbf{w})$ and $\tilde{f}(\mathbf{w}; \theta)$ are the spectral densities of the true and approximated model, respectively. Through the numerical minimization of the criterion, we estimate the parameters of the approximated model,

$$\hat{\theta} = \arg \min_{\theta} KL(f, \tilde{f}; \theta).$$

The CSD between two spectral densities is expressed as,

$$G(f, \tilde{f}; \theta) = \int_{-\pi}^{\pi} \frac{(f(\mathbf{w}) - \tilde{f}(\mathbf{w}; \theta))^2}{f(\mathbf{w})} d\mathbf{w},$$

where $f(\mathbf{w})$ and $\tilde{f}(\mathbf{w}; \theta)$ are the spectral densities of the true and approximated model, respectively. The best fitting model is obtained using the optimal parameters estimated by minimizing the CSD between the two spectral densities,

$$\hat{\theta} = \arg \min_{\theta} G(f, \tilde{f}; \theta).$$

3.3. Comparisons of spectral density approximation methods with covariance function approximation methods

We compare our approximation approaches using spectral densities in the frequency domain to the squared relative differences (SRD) between covariance functions in the spatial domain. The SRD approach in the spatial domain is to minimize the squared relative differences between the covariance functions of the two models,

$$\hat{\theta} = \arg \min_{\theta} \int_{R^2} \left(\frac{C(\mathbf{s}_1, \mathbf{s}_2) - \tilde{C}(\mathbf{s}_1, \mathbf{s}_2; \theta)}{C(\mathbf{s}_1, \mathbf{s}_2)} \right)^2 ds_1 ds_2,$$

where $C(\mathbf{s}_1, \mathbf{s}_2)$ and $\tilde{C}(\mathbf{s}_1, \mathbf{s}_2; \theta)$ are the true and approximated covariances at $\mathbf{s}_1, \mathbf{s}_2 \in R^2$. We compare the three approximation methods based on the mean squared prediction error (MSPE). The MSPE is the average squared difference between the actual and predicted values at different locations, defined as,

$$MSPE = \frac{1}{N} \sum_i (Z(\mathbf{s}_i) - \hat{Z}(\mathbf{s}_i))^2,$$

where $Z(\mathbf{s}_i)$ and $\hat{Z}(\mathbf{s}_i)$ are the observed and predicted values at site i , and N is the number of sites.

Instead of conducting simulations by generating data from a spatial process with known covariance models to calculate MSPE, we obtain the averaged mean squared prediction error

(AMSPE), which can be expressed as a function of the known covariances of GGMs or GMRFs to compare various metrics. We consider a spatial process with mean zero and a known covariance which follows either a GGM or a GMRF. In the approximations of GGMs by GMRFs, the predicted value at site \mathbf{s}_i given other values $\hat{Z}_{\text{GGM}}(\mathbf{s}_i)$ is expressed as

$$\hat{Z}_{\text{GGM}}(\mathbf{s}_i) = - \sum_{j=1, j \neq i}^N Q_{ij} Q_{ii}^{-1} Z(\mathbf{s}_j),$$

where Q_{ij} is the (i, j) th element of the precision matrix of the approximated GMRF. Then the AMSPE of a GGM is represented as

$$\text{AMSPE}_{\text{GGM}} = E \left(Z(\mathbf{s}_i) - \left(- \sum_{j=1, j \neq i}^N Q_{ij} Q_{ii}^{-1} Z(\mathbf{s}_j) \right) \right)^2 = E(\mathbf{C}'\mathbf{Z})^2 = \mathbf{C}' \Sigma_{\text{GGM}} \mathbf{C},$$

where \mathbf{C} is an $N \times 1$ vector with $c_i = 1$ and $c_j = Q_{ij} Q_{ii}^{-1}$, $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_N))'$ and Σ_{GGM} is the covariance matrix of a GGM. The expectation in the AMSPE of a GGM is taken with respect to the true underlying GGM. In the approximation of GMRFs by GGMs, the predicted value $\hat{Z}_{\text{GMRF}}(\mathbf{s}_i)$ is expressed as

$$\hat{Z}_{\text{GMRF}}(\mathbf{s}_i) = \sigma_{iJ} \sigma_{JJ}^{-1} \mathbf{Z}_{-i},$$

where σ_{iJ} is an $1 \times (N - 1)$ matrix with the j th element, $\sigma_{iJ(j)} = \text{cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j))$, σ_{JJ} is a $(N - 1) \times (N - 1)$ matrix with the (j, k) th element, $\sigma_{JJ(jk)} = \text{cov}(Z(\mathbf{s}_j), Z(\mathbf{s}_k))$ where $j, k = 1, \dots, i - 1, i + 1, \dots, N$, and $\mathbf{Z}_{-i} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_{i-1}), Z(\mathbf{s}_{i+1}), \dots, Z(\mathbf{s}_N))'$. The AMSPE of a GMRF is

$$\text{AMSPE}_{\text{GMRF}} = E(Z(\mathbf{s}_i) - \sigma_{iJ} \sigma_{JJ}^{-1} \mathbf{Z}_{-i})^2 = E(\mathbf{D}'\mathbf{Z})^2 = \mathbf{D}' \Sigma_{\text{GMRF}} \mathbf{D},$$

where \mathbf{D} is an $N \times 1$ vector with $d_i = 1$, $d_j = -(\sigma_{iJ} \sigma_{JJ}^{-1})_j$ for $j < i$, $d_j = -(\sigma_{iJ} \sigma_{JJ}^{-1})_{j-1}$ for $j > i$ and Σ_{GMRF} is the covariance matrix of a GMRF. The expectation in the AMSPE of a GMRF is taken with respect to the true underlying GMRF. We also compare the AMSPE of the approximated model with the AMSPE of the original model (AOM) to check the performance of the criteria. The AMSPE of the original GGM is defined as,

$$\text{AMSPE} = E(Z(\mathbf{s}_i) - Z^{\text{pred}}(\mathbf{s}_i))^2 = E(Z(\mathbf{s}_i) - \sigma_{iJ} \sigma_{JJ}^{-1} \mathbf{Z}_{-i})^2 = E(\mathbf{D}'\mathbf{Z})^2 = \mathbf{D}' \Sigma_{\text{GGM}} \mathbf{D},$$

and the AMSPE of the original GMRF is,

$$\begin{aligned} \text{AMSPE} &= E(Z(\mathbf{s}_i) - Z_{\text{GMRF}}^{\text{pred}}(\mathbf{s}_i))^2 \\ &= E \left(Z(\mathbf{s}_i) - \left(- \sum_{j=1, j \neq i}^N Q_{ij} Q_{ii}^{-1} Z(\mathbf{s}_j) \right) \right)^2 = E(\mathbf{C}'\mathbf{Z})^2 = \mathbf{C}' \Sigma_{\text{GMRF}} \mathbf{C}, \end{aligned}$$

where $Z_{\text{GGM}}^{\text{pred}}(\mathbf{s}_i)$ and $Z_{\text{GMRF}}^{\text{pred}}(\mathbf{s}_i)$ are the predicted values of the GGM and the GMRF at \mathbf{s}_i , respectively.

4. An empirical study

In this section, an empirical study is conducted to compare the spectral density approximation methods to the covariance approximation method in the approximations of GGMs and GMRFs.

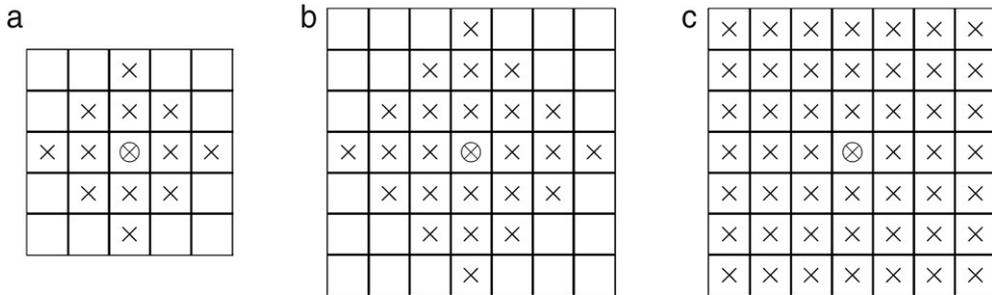


Fig. 2. (a) Neighborhood I, (b) neighborhood II, (c) neighborhood III.

Let us define a Gaussian Geostatistical process (GGP) and a GMRF on a 30×30 lattice. To reduce the nonstationarity problem of the GMRF which arises due to the edge effect, we consider a buffer zone by extending the lattice to a 32×32 grid to cover a larger domain, and we use only the covariance of the original lattice within the covariance of the extended lattice for approximations. We consider a Matérn model with various smoothness parameters for the GGP and different neighborhood structures that are shown in Fig. 2 with scaled weights for the GMRF.

The empirical study is composed of two steps, estimating parameters using each metric, and comparing them in terms of the AMSPE. In the estimation step, we estimate the parameters of the approximated models by minimizing the criteria suggested in the previous section: CSDS (chi-squared differences between spectral densities), KLS (the Kullback–Leibler discrepancy between spectral densities) and SRDC (squared relative differences between covariance functions). In the approximations of GMRFs by GGMs, we estimate the range (ϕ) and sill (σ^2) parameters of the Matérn model,

$$C(h) = \frac{2\sigma^2}{\Gamma(v)} \left(\frac{h}{2\phi}\right)^v K_v\left(\frac{h}{\phi}\right),$$

where h denotes the Euclidean distance and v is a smoothness parameter. For the approximations of GGMs by GMRFs, we estimate the spatial dependency (α), scale (τ) parameters in the GMRF covariance function,

$$(\mathbf{I} - \alpha\mathbf{B})^{-1}\mathbf{M}(\tau),$$

where \mathbf{B} is the neighborhood matrix and $\mathbf{M}(\tau)$ is the diagonal matrix with elements τ . These parameters are estimated using the `nlm` function in the R software (<http://www.r-project.org/>). We perform various empirical studies to explore the relation between GGMs and GMRFs. Table 1 shows the estimated parameters of GGMs when approximating GMRFs with different values for α and neighborhood III using GGMs. We obtain different estimated parameters depending on the matching criteria. When we use the covariance as an approximation criterion, the estimated sill and range vary severely over different values of α , while the estimated parameters are stable over different values of α in the case of using spectral density criteria. We compare our methods by the average mean squared prediction error (AMSPE). Table 2 shows that the AMSPEs obtained using the CSDS and the KLS criteria are smaller than those obtained using the SRDC, and they are close to the AOM. This suggest that the CSDS and KLS are good criteria for the approximations of GMRFs by GGMs.

In Table 3, we present the estimation results for GGMs in the approximation of GMRFs with $\alpha = 0.95$ and various neighborhood structures using GGMs. For the CSDS and the KLS, the

Table 1

Estimated parameters of GGMs in approximations of GMRFs with neighborhood III and various values of α by GGMs with $\nu = 0.5$

Method	$\alpha = 0.9$		$\alpha = 0.95$		$\alpha = 0.98$	
	σ^2	ϕ	σ^2	ϕ	σ^2	ϕ
CSDS	0.43	0.14	0.43	0.15	0.43	0.15
KLS	0.43	0.13	0.43	0.14	0.43	0.15
SRDC	0.43	0.14	0.001	10.40	0.002	19.37

Table 2

AMSPE in approximations of GMRFs with neighborhood III and various values of α by GGMs with $\nu = 0.5$

Method	AMSPE		
	$\alpha = 0.9$	$\alpha = 0.95$	$\alpha = 0.98$
CSDS	0.019	0.019	0.020
KLS	0.019	0.019	0.020
SRDC	0.024	0.023	0.023
AOM	0.018	0.018	0.018

Table 3

Estimated parameters of GGMs in approximations of GMRFs with $\alpha = 0.95$ and various neighborhood structures by GGMs with $\nu = 0.5$

Method	I		II		III	
	σ^2	ϕ	σ^2	ϕ	σ^2	ϕ
CSDS	2.35	0.42	0.90	0.21	0.43	0.15
KLS	2.31	0.28	0.90	0.18	0.43	0.14
SRDC	0.02	4.50	0.005	6.71	0.001	10.45

Table 4

AMSPE in approximations of GMRFs with $\alpha = 0.95$ and various neighborhood structures by GGMs with $\nu = 0.5$

Method	AMSPE		
	I	II	III
CSDS	0.103	0.043	0.019
KLS	0.109	0.043	0.019
SRDC	0.101	0.046	0.023
AOM	0.08	0.037	0.018

estimated sill and range parameters tend to decrease as the neighborhood increases. However, the SRDC gives increasing estimated range parameters along with the increasing neighborhood. Table 4 shows the prediction results using the estimated parameters. As the neighborhood increases, the AMSPE decreases, and the SRDC gives the largest AMSPE for the II and III neighborhoods. Comparing with the AOM, the CSDS and KLS yield closer values of the AMSPE than the SRDC.

We approximate GMRFs using Matérn models with different smoothness parameters (ν) to study the effect of the smoothness parameter on the approximations of GMRFs by Matérn models. Table 5 presents the estimated parameters of GGMs in the approximations of GMRFs

Table 5

Estimated parameters of GGMs in approximations of GMRFs with $\alpha = 0.95$ and neighborhood III using Matérn models with various values of ν

Method	$\nu = 0.1$		$\nu = 0.5$		$\nu = 1$	
	σ^2	ϕ	σ^2	ϕ	σ^2	ϕ
CSDS	0.66	0.15	0.43	0.15	0.42	0.15
KLS	0.68	0.12	0.43	0.14	0.42	0.15
SRDC	0.005	5.74	0.001	10.40	0.001	12.34

Table 6

AMSPE in approximations of GMRFs with $\alpha = 0.95$ and neighborhood III using Matérn models with various values of ν

Method	AMSPE		
	$\nu = 0.1$	$\nu = 0.5$	$\nu = 1$
CSDS	0.019	0.019	0.019
KLS	0.020	0.019	0.019
SRDC	0.020	0.023	0.023
AOM	0.018	0.018	0.018

Table 7

Estimated parameters of GMRFs in approximations of GGMs with $\sigma^2 = 1$, $\nu = 0.5$ and various values of ϕ by GMRFs with neighborhood II

Method	$\phi = 2 (\phi < \text{MDN})$		$\phi = 3 (\phi = \text{MDN})$		$\phi = 5 (\phi > \text{MDN})$	
	α	τ	α	τ	α	τ
CSDS	0.99	0.32	0.99	0.22	0.99	0.13
KLS	0.99	0.22	0.99	0.13	0.99	0.08
SRDC	0.63	101.52	0.79	166.42	0.91	140.61

with $\alpha = 0.95$ and neighborhood III using Matérn models with various values of ν . The parameters estimated by the SRDC vary over the various values of ν while those estimated by the CSDS and the KLS show a little change by the different values of ν . Table 6 displays the AMSPE results, which show that AMSPEs obtained by the SRDC are larger than the other methods.

Table 7 shows the parameter estimates of GMRFs corresponding to GGMs with $\sigma^2 = 1$ and various values of ϕ . We consider the three different cases, (1) when ϕ is smaller than the maximum distance within the neighborhood (MDN), (2) when ϕ is the same as the MDN and (3) when ϕ is larger than the MDN, and assign different values of ϕ for each case. The estimated α is close to one when ϕ is larger than the MDN. The AMSPE becomes smaller as ϕ increases, and we obtain values closer to the AOM by using the CSDS and the KLS criteria (Table 8).

To study the effects of σ^2 values on the approximations of GGMs by GMRFs, we fit GGMs with various values of σ^2 to GMRFs. Table 9 presents the estimated parameters of GMRFs which approximate GGMs with $\phi = 3$ and various values of σ^2 . As σ^2 increases, the estimated conditional variance increases for all criteria, which suggests a positive relation between the sill parameter in the GGM and the conditional variance in the GMRF. Table 10 presents the AMSPE results, showing that the CSDS and the KLS give closer values of the AMSPE to the AOM than the SRDC.

Table 8

AMSPE in approximations of GGMs with $\sigma^2 = 1$, $v = 0.5$, and various values of ϕ by GMRFs with neighborhood II

Method	AMSPE		
	$\phi = 2$	$\phi = 3$	$\phi = 5$
CSDS	0.70	0.54	0.36
KLS	0.70	0.54	0.36
SRDC	0.76	0.58	0.38
AOM	0.51	0.36	0.21

Table 9

Estimated parameters of GMRFs in approximations of GGMs with $\phi = 3$, $v = 0.5$ and various values of σ^2 by GMRFs with neighborhood II

Method	$\sigma^2 = 0.5$		$\sigma^2 = 1$		$\sigma^2 = 2$	
	α	τ	α	τ	α	τ
CSDS	0.99	0.11	0.99	0.22	0.99	0.44
KLS	0.99	0.06	0.99	0.13	0.99	0.26
SRDC	0.79	83	0.79	166.42	0.82	332.53

Table 10

AMSPE in approximations of GGMs with $\phi = 3$, $v = 0.5$ and various values of σ^2 by GMRFs with neighborhood II

Method	AMSPE		
	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 2$
CSDS	0.27	0.54	1.09
KLS	0.27	0.54	1.09
SRDC	0.29	0.58	1.17
AOM	0.17	0.36	0.71

Table 11

Estimated parameters of GMRFs in approximations of GGMs with $\phi = 3$, $\sigma^2 = 1$ and various values of v by GMRFs with neighborhood II

Method	$v = 0.1$		$v = 0.5$		$v = 1$	
	α	τ	α	τ	α	τ
CSDS	0.99	0.64	0.99	0.41	0.99	0.03
KLS	0.99	0.45	0.99	0.33	0.99	0.02
SRDC	0.93	37.23	0.85	90.06	0.67	250.92

We investigate the effects of the Matern smoothness parameter (v) on approximations of Matern models by GMRFs. In the CSDS and the KLS, α is constant and τ decreases, while in the SRDC, α decreases and τ increases as v increases (Table 11). Comparing our methods using the AMSPE, the AMSPE decreases as v increases, and the CSDS and the KLS yield smaller AMSPEs than the SRDC (Table 12). Comparing our methods using the AOM, AMSPEs obtained using the CSDS and the KLS are closer to the AOM than those obtained using the SRDC criterion.

Table 12

AMSPE in approximations of GGMs with $\phi = 3, \sigma^2 = 1$ and various values of v by GMRFs with neighborhood II

Method	AMSPE		
	$v = 0.1$	$v = 0.5$	$v = 1$
CSDS	0.80	0.49	0.44
KLS	0.80	0.49	0.44
SRDC	0.80	0.52	0.54
AOM	0.77	0.36	0.16

Table 13

Estimated parameters of GMRFs with parametrized weights in approximations of GGMs with $\phi = 3, \sigma^2 = 1$ and various values of v by GMRFs with neighborhood II

Method	$v = 0.1$			$v = 0.5$			$v = 1$		
	α	τ	θ	α	τ	θ	α	τ	θ
CSDS	0.99	1.65	2.59	0.99	0.018	0.35	0.99	0.0005	0.21
KLS	0.99	0.15	0.72	0.99	0.0006	0.20	0.99	0.0005	0.56
SRDC	0.97	0.02	0.24	0.94	0.015	0.17	0.92	0.002	0.13

Table 14

AMSPE in approximations of GGMs with $\phi = 3, \sigma^2 = 1$ and various values of v by GMRFs with parametrized weights and neighborhood II

Method	AMSPE		
	$v = 0.1$	$v = 0.5$	$v = 1$
CSDS	0.79	0.38	0.19
KLS	0.77	0.37	0.19
SRDC	0.81	0.37	0.21
AOM	0.77	0.36	0.16

To study the potential impact of the weight function of GMRFs on the approximations of GGMs by GMRFs, we use a distance-based weight parametrized by γ , and estimate γ in addition to the α and τ ,

$$b_{ij} = \frac{\exp\left(-\frac{d_{ij}}{\gamma}\right)}{\sum_{j=1}^N \exp\left(-\frac{d_{ij}}{\gamma}\right)},$$

where b_{ij} is the weight for sites i and j , γ is a scale parameter, and d_{ij} is the distance between sites i and j . We performed this study on 16 by 16 lattices. The estimated parameters are slightly different from those based on the scaled binary weights in GMRFs (Table 13). However, we still obtain smaller AMSPEs using the CSDS and the KLS (Table 14).

We also compare the suggested metrics taking into account the computational time. Tables 15 and 16 display the computational time in approximations of GGMs by GMRFs, and approximations of GMRFs by GGMs. The unit of time is seconds, and these results are obtained using a Pentium PC with a 3.20 GHz CPU, showing that the approximation methods based on the spectral density are faster than those based on covariance functions.

Table 15

Computation times in approximations of GGMs with $\sigma^2 = 1$, $\phi = 3$ and $v = 0.5$ by GMRFs with neighborhood II

Methods	Computation times
CSDS	26 698
KLS	226 585
SRDC	347 815

Table 16

Computational times in seconds in approximations of GMRFs with $\alpha = 0.95$ and neighborhood III by GGMs with $v = 0.5$

Methods	Computation times
CSDS	25 238
KLS	31 111
SRDC	54 991

5. Applications

GGMs and GMRFs are two different models commonly used in modeling spatial data, and they have different characteristics in terms of modeling the spatial structure of an underlying process. GGMs are usually used for modeling point level data, and explain the spatial relation based on the distance between two points, while GMRFs have been developed for modeling areal level data, and the spatial structure is considered in the model based on the conditional location between two areas which is specified in the neighborhood structure. GMRFs have some advantages over GGMs in that they are more convenient to use in Bayesian hierarchical models to capture the spatial structure in the data due to the conditional form of GMRFs, and the computational benefits. Because of this, point level data are often aggregated to a certain areal level, and GMRFs are used to model aggregated point level data. The main objective of this paper is to investigate the possibility of the interchangeable use of GGMs and GMRFs by suggesting approximations of GGMs by GMRFs, and approximations of GMRFs by GGMs. For the approximations, we use the theoretical covariance of each model and approximate the models based on the different proposed criteria. One widely used criterion to approximate one covariance model with another model is minimizing the squared differences between the two covariance models. We suggest some new criteria based on the spectral density and our simulation study shows that our suggested criteria are more accurate in terms of the AMSPE and easier to compute than the previously suggested criterion.

Particulate Matter (PM) is the general term used for a complex mixture of solid particles and liquid droplets suspended in air, which includes a broad class of chemically diverse particles whose range in size is from $0.005 \mu\text{m}$ to $100 \mu\text{m}$ in diameter. PM is usually sorted by its size because size is an important factor that determines the health effects of PM. In particular, $\text{PM}_{2.5}$, also known as fine PM, which includes the fine particles that are less than or equal to $2.5 \mu\text{m}$ in diameter, causes a variety of serious health disease such as respiratory and cardiovascular problems. $\text{PM}_{2.5}$ observations are typical point-referenced data collected at point monitoring stations. This type of data can be aggregated to a level of interest to associate spatial variables with different spatial resolutions. For example, in the study of associations between $\text{PM}_{2.5}$ and adverse health effects, since health outcomes are generally collected over geographic regions, point-referenced $\text{PM}_{2.5}$ values are commonly aggregated to the level of the health outcomes

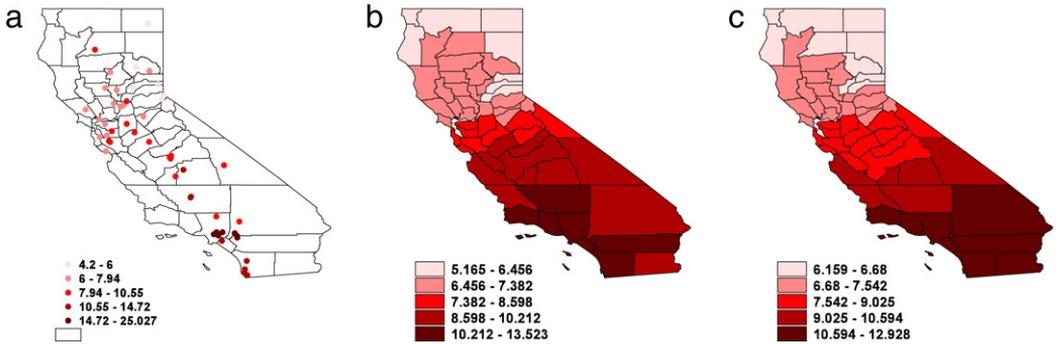


Fig. 3. (a) Point level $PM_{2.5}$ values, (b) county level averaged $PM_{2.5}$ values and (c) predicted values using GMRFs estimated by the CSDS.

Table 17

MSPE in the approximations of aggregated $PM_{2.5}$ using GMRFs with first and second order neighborhood structures

Method	MSPE	
	First order neighborhood	Second order neighborhood
CSDS	0.68	1.12
KLS	0.68	1.12
SRDC	0.68	1.19

to solve the mismatch support problem [11]. In this application, we apply the approximation methods described in the previous sections to approximate the aggregated $PM_{2.5}$ data at the county level in California by GMRFs. The aggregated point data at each county are obtained by averaging the predicted values of $PM_{2.5}$ at 20 locations within each county based on the $PM_{2.5}$ observations. Since the underlying process is a point process, the aggregated $PM_{2.5}$ process is also based on a point process. However, we can also consider the aggregated $PM_{2.5}$ as areal level data and model them using a GMRF. To approximate the aggregated PM process by a GMRF, we obtain the empirical covariance of GGMs based on the aggregated PM process and approximate the empirical covariance of the aggregated PM process by a GMRF using the several criteria we suggested in the previous section. We also investigate the effects of the neighborhood of GMRFs in the approximations by considering the first and second order neighborhoods in GMRFs. After estimating the parameters of the GMRF, the performance of the different methods are evaluated through the MSPE. Fig. 3 displays the point and aggregated county level $PM_{2.5}$ values and the GMRF with the first order neighborhood estimated by the CSDS with the first order neighborhood. In the point level map, we see that the Southern California areas are highly polluted with $PM_{2.5}$, these patterns are also observed in the county level map and the predicted map using approximated GGMs. We can also notice that the high magnitude of $PM_{2.5}$ found in the Southern California areas is reduced after the aggregation.

Table 17 is the MSPE obtained from approximations of the aggregated $PM_{2.5}$ county values using a GMRF. In our example, the MSPE values are the same for all the methods with the first order neighborhood and the SRDC gives a little larger MSPE than the other methods for the second order neighborhood. Also we notice that the GMRF with the first order neighborhood performs better than the GMRF with the second order neighborhood in that the first order neighborhood yields smaller MSPE than the second order neighborhood.

6. Conclusions

The major objective of this paper is to study relations between GMRFs and GGMs through approximations of GMRFs by GGMs, and vice versa. We approximate GMRFs by GGMs and GGMs by GMRFs using three approximation methods based on spectral density functions and covariance functions, which are evaluated in terms of the AMSPE and computation times. The approach presented here using spectral densities is easy and fast to compute compared to covariance-based criteria. Our approach reduces considerably the total computation time by the use of Fast Fourier transform algorithm and also yields smaller AMSPEs than the covariance criterion. One noticeable result is that there are positive relationships between range parameter and spatial dependency parameter, and sill and conditional variance which might be expected since these parameters control the strength of the correlation of the data and variance in GMRFs and GGMs. Since we use different criteria in approximations of GGMs and GMRFs, the parameter estimates are different depending on the criteria used. However, the results in terms of prediction were not very different, and show some patterns as we change parameters of GGMs and GMRFs. In the approximations of GMRFs by GGMs, as the neighborhood order increases, the AMSPE becomes smaller, while the AMSPE decreases as the range increases and the conditional variance decreases in the approximations of GGMs by GMRFs. In our application, we approximated county level averaged $PM_{2.5}$ using GMRFs. The smallest MSPE was obtained using the CSDS and the KLS criteria.

Further investigation is being conducted to study the effect of the level of aggregation on the approximations of averaged process using GMRFs. Also we plan to study the bias that might occur in estimated parameters when GMRFs are used in modeling aggregated data instead of averaged continuous spatial process.

References

- [1] G. Arfken, H.J. Weber, *Mathematical Methods for Physicists*, 4th edn, Academic Press, San Diego, CA, 1995.
- [2] J. Besag, Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society* 36 (1974) 192–236.
- [3] J. Besag, On a system of two-dimensional recurrence equations, *Journal of the Royal Statistical Society* 43 (1981) 302–309.
- [4] J. Besag, C. Kooperberg, On conditional and intrinsic autoregressions, *Biometrika* 82 (1995) 733–746.
- [5] J. Besag, J.C. York, A. Mollie, Bayesian image restoration, with two applications in spatial statistics (with discussion), *Annals of the Institute of Statistical Mathematics* 43 (1991) 1–59.
- [6] N.G. Best, K. Ickstadt, R.L. Wolpert, D.J. Briggs, Combining models of health and exposure data: The SAVIAH study, in: P. Elliott, J.C. Wakefield, N.G. Best, D.J. Briggs (Eds.), *Spatial Epidemiology: Methods and Applications*, Oxford, University Press, Oxford, New York, 2000, pp. 393–414.
- [7] D.R. Brillinger, *Time Series: Data Analysis and Theory*, Holden-Day, San Francisco, 1981.
- [8] B.P. Carlin, S. Banerjee, Hierarchical multivariate CAR models for spatio-temporally correlated survival data (with discussion), in: J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith, M. West (Eds.), *Bayesian Statistics 7*, Oxford University Press, Oxford, 2003, pp. 45–63.
- [9] H. Cramer, M.R. Leadbetter, *Stationary and Related Stochastic Processes. Sample Function Properties and their Applications*, Wiley, New York, 1967.
- [10] N. Cressie, *Statistics for Spatial Data*, Wiley, New York, 1993.
- [11] M. Fuentes, H.R. Song, S. Ghosh, D. Holland, J. Davis, Spatial association between speciated fine particles and mortality, *Biometrics* 62 (2006) 855–863.
- [12] D. Griffith, F. Csillag, Exploring relationships between semivariogram and spatial autoregressive models, *Papers in Regional Science* 72 (1993) 283–295.
- [13] B. Hrafnkelsson, N. Cressie, Hierarchical modeling of count data with application to nuclear fall-out, *Environmental and Ecological Statistics* 10 (2003) 179–200.

- [14] A.G. Journel, Ch.J. Huijbregts, *Mining Geostatistics*, Academic Press, London, 1978.
- [15] R.J. Martin, J.N. Dwyer, Approximations to the covariance properties of processes averaged over irregular spatial regions, *Communications in Statistics-Theory and Methodologies* 23 (3) (1994) 913–945.
- [16] H. Rue, H. Tjelmeland, Fitting gaussian markov random fields to Gaussian fields, *Scandinavian Journal of Statistics* 29 (2002) 31–49.
- [17] M.L. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*, Springer-Verlag, New York, 1999.
- [18] D. Sun, R.K. Tsutakawa, P.L. Speckman, Posterior distribution of hierarchical models using CAR(1) distributions, *Biometrika* 86 (1999) 341–350.
- [19] P. Tolbert, J. Mulholland, D. MacIntosh, F. Xu, D. Daniels, O. Devine, B.P. Carlin, M. Klein, J. Dorley, A. Butler, D. Nordenberg, H. Frumkin, P.B. Ryan, M. White, Air pollution and pediatric emergency room visits for asthma in Atlanta, *American Journal of Epidemiology* 151 (8) (2000) 798–810.
- [20] A.M. Yaglom, *Correlation Theory of Stationary and Related Random Functions*, Springer-Verlag, New York, 1987.