# A comparative study of Gaussian geostatistical and Gaussian Markov random field models [1]

**Hae-Ryoung Song, Montserrat Fuentes, and Sujit Ghosh**

Institute of Statistics Mimeo Series# 2577

**Abstract**

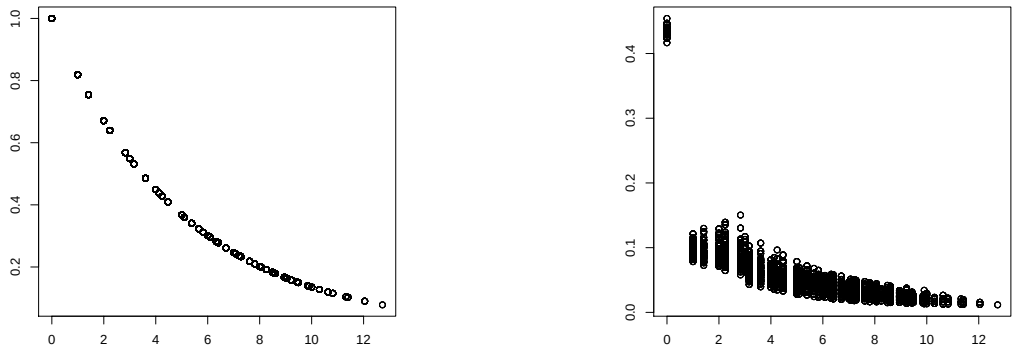
Gaussian geostatistical models (GGMs) and Gaussian Markov random fields (GMRFs) are two distinct approaches commonly used in modeling point referenced and areal data, respectively. In this work the relations between GMRFs and GGMs are explored based on approximations of GMRFs by GGMs, and vice versa. The proposed framework for the comparison of GGMS and GMRFs is based on minimizing the distance between the corresponding spectral density functions. In particular, the Kullback-Leibler discrepancy of spectral densities and the chi-squared distance between spectral densities are used as the metrics for the approximation. The proposed methodology is illustrated using simulation studies. We also apply the methods to a air pollution dataset in California to study the relation between GMRFs and GGMs.

[1]H.R. Song is a Ph.D. student at the Statistics Department, North Carolina State University (NCSU), Raleigh, N.C. M. Fuentes is an Associate Professor at the Statistics Department, NCSU. S. Ghosh is an Associate Professor at the Statistics Department, NCSU. *Contact author's email address:* fuentes@stat.ncsu.edu.
The research conducted by Fuentes has been partly supported by a National Science Foundation grant DMS 0353029. Song has been supported by a cooperative agreement (CT829562) with the Office of Air Quality Planning and Standards/U.S. Environmental Protection Agency.
*Keywords:* covariance, spatial geostatistical model, spatial random field, spectral density.

# 1 Introduction

In many applications in environmental sciences and in epidemiology, data concerning a spatial process of interest are often observed at different spatial resolutions. For example, in studies of the association between air pollution exposure and adverse health effects, the pollution data and the health outcomes may be available as point-referenced data and as areal data, respectively (Best *et al.* 2000; Tolbert *et al.* 2000). One common solution to this spatial misaligment 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 (Journel & Huijbregts, 1978; Martin and Dwyer, 1994). Modeling aggregated data using these spatial integrals is computationally very expensive. Hence, because of the computational benefits it is becoming very common to use Gaussian Markov random fields (GMRFs) in modeling aggregated point-referenced data.

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 point-referenced data. On the other hand, GMRFs are widely used for areal data, such as, counts or spatial 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 (1981) showed that covariance function of GMRFs could be approximately represented in terms of a modified Bessel function that decreases monotonically with distance. Griffith and Csillag (1993) investigated the approximation of GMRFs using several geostatistical models by minimizing the squared differences between the covariances of GMRFs and GGMs. Hrafnkelsson and Cressie (2003) defined a specific class of GMRFs that was approximated in an empirical setting by a GGM using a Matérn covariance model. On the other hand, Rue and Tjelmeland (2002) investigated approximations of GGMs by GMRFs, because of the computational benefits that GMRFs offer. Rue and Tjelmeland (2002) suggested two approximation methods of GGMs by GMRFs, one using the Kulback-Leibler (KL) discrepancy and the other one by matching the correlation functions of the models.




(a) (b)

Figure 1: a) covariance function of a Matérn model with nugget=0, range=5, smoothness=.5, and partial sill=1, versus distance b) estimated covariance function of a GMRF with a 1st order neighborhood structure versus distance. Note the different scale in the vertical axis.

They showed 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. One common metric is the likelihood function, in which using samples generated from the original model, the parameters of the corresponding model can be estimated by maximizing the likelihood function. However, based on simulations studies we conducted, it seems that the maximum likelihood approach tends to underestimate the variance. Figure 1 displays the covariance of the original GGM and the estimated GMRF with a 1st order neighborhood versus distance, showing the underestimation of the variance of the original process. In this paper, we introduce two new metrics using spectral density functions. In Section 2, we present general overviews of GGMs and GMRFs. In Section 3, we propose new approaches to approximate GGMs with GMRFs and vice versa, using spectral density functions. In Section 4, we compare the performance of our spectral approach to other methods using covariance functions. 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 classes of spatial models, GGMs and GMRFs. The choice of model is usually determined by whether the data are collected as points in space or are observed as averages over subregions. GGMs (e.g. Cressie, 1993) are used in modeling point-referenced data of continuous process and GMRFs (e.g. Besag, 1974) are used for modeling areal data whose underlying process is discrete.

## 2.1 Gaussian geostatistical models (GGMs)

GGMs are used to represent continous spatial Gaussian processes with a spatial covariance that is often just a function of distance and direction between locations. Two common assumptions in GGMs 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,

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

where $h$ denotes the Euclidean distance between two points, $K_v$ is a modified Bessel function of the third kind (Arfken and Weber, 1995). $\tau^2$, $\sigma^2$ and $\phi$ 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 function $K_v$ is the modified Bessel function of order $v > 0$ which controls the smoothness of the function.

In particular, when $v = 1/2$, the Matérn covariance reduces to an exponential model and when $v \to \infty$, it approaches to a Gaussian covariance model (e.g. Stein, 1999). Due to its flexibility, the Matérn covariance is widely used in a variety of spatial applications.

## 2.2 Gaussian Markov random models (GMRFs)

Let $Z_i$ denote a random variable observed at the site $i$. A GMRF is characterized by the following conditional distribution

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

where $\mu_i$ is the mean of $Z_i$, and $\tau_i$ is the conditional variance of $Z_i$ given $\{Z_j : j \neq i\}$. By Brook's Lemma and Hammersley-Clifford Theorem (Besag 1974), the joint distribution of a GMRF $\mathbf{Z} = (Z_1, \cdots, Z_n)^T$ is uniquely determined as,

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

where $\boldsymbol{\mu}$ is an n×1 vector of means with elements $\mu_i$, and $\mathbf{M}$ is an n×n diagonal matrix with elements $\tau_i$. The matrix $\mathbf{B}$ is an n×n neighborhood matrix with elements $b_{ij}$, which satisfy $b_{ii} = 0$, $b_{ij}\tau_j = b_{ji}\tau_i$ and $b_{ij} = 0$, unless $i$ and $j$ are neighbors.

The parameter $\alpha$ is known as the spatial dependency parameter which somehow controls spatial dependence in the covariance. Specific choices of $\alpha$ 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, (Besag et al. 1991; Besag and Kooperberg, 1995) which has been extensively used in spatial statistics as a prior for random effects in Bayesian hierarchical models (Sun et al. 1999; Carlin and Banerjee, 2003).

A GMRF involves the specification of neighborhoods. We impose a neighborhood structure on a site $i$ and then we assign weights to the neighboring values, which are prespecified

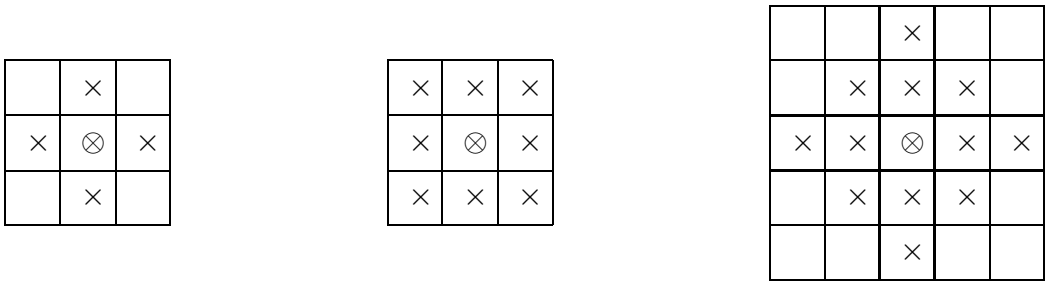


Figure 2: (a) First order neighbors (b) Second order neighbors (c) Third order neighbors

according to some criterion. In the case of a regular lattice, some popular neighborhoods structures are defined as orders depending on the touching borders between grid cells. Figure 2 illustrates different order of neighbors. 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. Scale weights are usually restricted to have each row of weights sum to one.

# 3 Spectral methods to measure discrepancy

Spectral densities and covariance functions are related through Fourier transforms and they are both 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. Spectral analysis might offer some advantages. 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 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 (Stein, 1999). Considering these advantages of the spectral approaches, we introduce approximation methods using spectral densities. There are also some limitations, since spectral analysis generally require 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.

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

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

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

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

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

where $dX(\mathbf{w}) = \lim_{h\to 0} \frac{(X(\mathbf{w}+\mathbf{h}) - X(\mathbf{w}))}{h}$ and $dF(\mathbf{w}) = \lim_{h\to 0} \frac{(F(\mathbf{w}+\mathbf{h}) - F(\mathbf{w}))}{h}$. 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 $\{\mathbf{h} = ||\mathbf{s}_1 - \mathbf{s}_2||, \mathbf{s}_1, \mathbf{s}_2 \in D \subset R^2\}$. If a stochastic process is defined on a lattice $\Lambda$. 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 $\Lambda$ with an interval between neighboring observations $\delta$, we can not distinguish between the frequency components $\mathbf{w}$ and $\mathbf{w} + \frac{2\pi\mathbf{l}}{\delta}$, for $\mathbf{l}$ in $L$, where $L = \{(l_1, l_2); l_1, l_2 = \pm 1, \pm 2, \cdots\}$. The power in the sampled process from the set of frequencies $\mathbf{w} \in (-\frac{\boldsymbol{\pi}}{\delta}, \frac{\boldsymbol{\pi}}{\delta})$ is the accumulation of power in the original process from all the sets of frequencies $\mathbf{w} + \frac{2\pi\mathbf{l}}{\delta}$ . For any $\mathbf{w}$ in $(-\frac{\boldsymbol{\pi}}{\delta}, \frac{\boldsymbol{\pi}}{\delta}]$, the frequencies $\mathbf{w} + \frac{2\pi\mathbf{l}}{\delta}$ are called the aliases of $\mathbf{w}$. The spectral density of the sampled process is defined as follows,

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

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

A nonstationary process can also be represented by sinusoidal functions, but the corresponding spectral process does no longer have orthogonal increments but correlated. If a process is defined on a lattice $\Lambda$, then the Fourier transform pairs 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 $\mathbf{w_1}, \mathbf{w_2} \in (-\pi, \pi]^2$. For the special case in which the spectral density is zero for all the elements, except for the ones along the diagonal ($\mathbf{w_1} = \mathbf{w_2}$), then the corresponding process is stationary.

## 3.2 Approximation methods using spectral densities

We assume that we observe a continuous stationary and isotropic Gaussian spatial process on a $r \times c$ finite lattice $\Lambda$, with an interval $\delta$ between neighboring points. Then, we 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. For the GGM, the underlying process is continuous, but observed only on a discrete lattice $\Lambda$. Hence, the spectral density of the sampled GGMs $g_\delta$ is defined as,

$$g_\delta(\mathbf{w}) = \sum_{\mathbf{l} \in \mathcal{Z}^2} g\left(\mathbf{w} + \frac{2\pi\mathbf{l}}{\delta}\right)$$

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

for $\mathbf{w} \in (-\pi, \pi]^2$ and Z=$\{0, \pm1, \pm2, \cdots\}$.

Nonstationary processes corresponding to GMRFs defined on a lattice can not 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 stationary geostatistical

process with the spectral density of a nonstationary GMRF or vice versa by minimizing the differences of the spectral densities of stationary and nonstationary processes when $\mathbf{w}_1 = \mathbf{w}_2 = \mathbf{w}$. This corresponds to the stationary component of the GMRF, we basically ignore the nonstationary contribution, due to the edge effect.

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,

$$\hat{\theta} = \text{argmin}_\theta \, KL(f, \tilde{f}; \theta).$$

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,

$$\hat{\theta} = \text{argmin}_\theta \, G(f, \tilde{f}; \theta).$$

## 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 those using covariance functions in the spatial domain. The KL approach in the spatial domain consists in estimating the parameters of the approximated model by minimizing the

KL discrepancy between the distribution functions of the two models,

$$\hat{\theta} = \text{argmin}_\theta KL(g, \tilde{g}; \theta) = \text{argmin}_\theta \int_{R^2} g(\mathbf{x}) \log \frac{g(\mathbf{x})}{\tilde{g}(\mathbf{x}; \theta)} d\mathbf{x},$$

where $g(\mathbf{x})$ and $\tilde{g}(\mathbf{x}; \theta)$ are probability densities of the true and approximation model, respectively. The squared relative differences (SRD) approach in the spatial domain consits in minimizing the squared relative differences between the covariance functions of the two models,

$$\hat{\theta} = \text{argmin}_\theta \int_{R^2} \left( \frac{C(\mathbf{s}_1, \mathbf{s}_2) - \tilde{C}(\mathbf{s}_1, \mathbf{s}_2; \theta)}{C(\mathbf{s}_1, \mathbf{s}_2)} \right)^2 d\mathbf{s}_1 d\mathbf{s}_2,$$

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 four 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, and it is 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 true and predicted value at site $i$, and $N$ is the number of sites. Using the MSPE, we predict the values at some locations where we have data, and then we compare them to the actual values at these locations.

# 4 A simulation study

In this Section, a simulation study is conducted to compare the spectral density approximation approaches to the covariance approximation methods. Let us define a Gaussian Geostatistical process (GGP) and a GMRF on a 10×10 lattice. To reduce the nonstationarity problem of the GMRF which arises due to the edge effect, we extend the lattice to a 12×12 grid to cover a larger domain, and we consider only the covariance of the original lattice within the covariance of the extended lattice. We consider a Matérn model with

smoothness parameter 0.5 and nugget 0 for the GGP and a binary scaled weight neighborhood structure for the GMRF. The simulation is composed of two steps, the estimation of parameters and the prediction of data. In the estimation step, the parameters of the approximated models are estimated by minimizing the criteria suggested in the previous section: CSDS (chi-squared differences between spectral densities), KLS (Kullback-Leibler discrepancy between spectral densities), SRDC (squared relative differences between covariance functions), and KLP (Kullback-Leibler discrepancy between probability density functions). We estimate the range ($\phi$) and sill ($\sigma^2$) parameters of the Matérn model,

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

where $h$ denotes the Euclidean distance. For the GMRF, we estimate the spatial dependency ($\alpha$) and scale ($\tau$) parameters in the 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 $\tau$. These parameters are estimated using the optim function in the R software (http://www.r-project.org/).

We repeat the experiment 50 times to obtain an average MSPE and the standard deviation (SD) of the average MSPE across the 50 simulations.

Table 1 shows the estimated parameters of a GGM, when we approximate a GMRF (with different values for $\alpha$ and a 6th order neighborhood) using a GGM. As $\alpha$ increases, the estimated range parameter increases.

Using the estimated parameters of the GGM, we predict the observations generated from the GMRF, and then we obtain the average mean squared prediction error (AMSPE) (see Table 2). As $\alpha$ increases, the AMSPE increases using all criteria. The smallest AMSPE is obtained with the KLP, and the SRDC gives the largest AMSPE when $\alpha = 0.9$ and $\alpha = 0.95$. However, when $\alpha$ is 0.98, the smallest AMSPE is obtained by the SRDC. However,

Table 1: Approximations of GMRFs with a 6th neighborhood order and various values of $\alpha$ using GGMs. This table presents the estimated parameters of the GGM

| | $\alpha$=0.9 | | $\alpha$=0.95 | | $\alpha$=0.98 | |
|---|---|---|---|---|---|---|
| method | sill (SE) | range (SE) | sill (SE) | range (SE) | sill (SE) | range (SE) |
| CSDS | 2.56 ( 0.68) | 0.55 (0.31) | 2.60 ( 0.71) | 0.72 (0.45) | 1.99 (0.60) | 2.60 ( 0.53) |
| KLS | 2.57 (0.43) | 0.46 (0.17) | 2.65 (0.38) | 0.50 (0.42) | 2.72 (0.47) | 0.54 (0.19) |
| SRDC | 0.22 (0.004) | 3.68 (0.043) | 0.32 (0.0058) | 5.82 (0.10) | 0.60 (0.01) | 10.69 (0.28) |
| KLP | 1.02 (0.14) | 0.38 (0.17) | 1.03 (0.14) | 0.40 (0.17) | 1.03 (0.15) | 0.41 (0.17) |

Table 2: Approximations of GMRFs with a 6th order neighborhood and various values of $\alpha$ using GGMs. Predicting GMRFs using GGMs

| | AMSPE | | |
|---|---|---|---|
| method | $\alpha$=0.9 | $\alpha$=0.95 | $\alpha$=0.98 |
| CSDS | 1.225 (0.210) | 1.304 (0.244) | 1.519 (0.362) |
| KLS | 1.198 (0.200) | 1.265 (0.240) | 1.470 (0.410) |
| SRDC | 1.332 (0.241) | 1.364 (0.244) | 1.513 (0.349) |
| KLP | 1.172 (0.191) | 1.238 (0.238) | 1.443 (0.427) |

the differences are not statistically significant.

In Table 3 we present the estimation results for GGMs, in the approximation of a GMRF with $\alpha$=0.95 and various orders for the neighborhood structure using a GGM. For the CSDS, KLS and KLP, the estimated sill and range parameters tend to decrease as the neighborhood order increases. However, the SRDC gives an estimated range parameter that increases with the neighborhood order. Table 4 shows the prediction results using the estimated parameters. As the neighborhood order increases, the AMSPE decreases and the SRDC gives the largest AMSPE for all neighborhood orders.

Table 5 shows the parameter estimates of a GMRF corresponding to a GGM with a sill=1 and with various values for the range parameter. The estimated $\alpha$ is always close to one. This might be a justification for the popular intrinsic GMRF models. The AMSPE becomes smaller as the range parameter increases (Table 6).

Table 3: Approximations of GMRFs with $\alpha$=0.95 and various orders of neighborhood using GGMs. This table shows the estimated parameters of a GGM.

| | 2nd order | | 3rd order | | 6th order | |
|---|---|---|---|---|---|---|
| method | sill (SE) | range (SE) | sill (SE) | range (SE) | sill (SE) | range (SE) |
| CSDS | 3.59 ( 0.76) | 1.13 (0.33) | 3.07 (0.76) | 1.09 (0.38) | 2.60 (0.71) | 0.72 (0.45) |
| KLS | 3.63 (0.76) | 0.92 (0.27) | 3.14(0.60) | 0.72(0.24) | 2.65 (0.38) | 0.50 (0.42) |
| SRDC | 0.65 (0.013) | 2.67 (0.025) | 0.48 (0.0092) | 3.56 (0.043) | 0.32 (0.005) | 5.82 (0.10) |
| KLP | 1.42 (0.28) | 0.94 (0.30) | 1.17(0.19) | 0.67 (0.22) | 1.03 (0.14) | 0.40 (0.17) |

Table 4: Approximations of GMRFs that have $\alpha$=0.95 and various orders of neighborhood structures using GGMs. This table shows the performance in terms of prediction of a GMRF using a GGM.

| | AMSPE | | |
|---|---|---|---|
| method | 2nd order | 3rd order | 6th order |
| CSDS | 1.756 (0.508) | 1.458 (0.330) | 1.304 (0.244) |
| KLS | 1.737 (0.500) | 1.419 (0.314) | 1.265 (0.240) |
| SRDC | 1.789 (0.522) | 1.493 (0.344) | 1.364 (0.244) |
| KLP | 1.740 (0.501) | 1.408 (0.310) | 1.238 (0.238) |

# 5 Applications

Particulate Matter (PM) is the general term used for a complex mixture of solid particles and liquid droplets suspended in the air. It is an air pollutant that represents a broad class of chemically diverse particles whose range in size is from 0.005 $\mu m$ to 100 $\mu m$ in diameter. $PM_{2.5}$, also known as fine PM, includes the fine particles that are less than or equal to 2.5 $\mu m$ in diameter. Our $PM_{2.5}$ observations are point-referenced data collected at monitoring stations. This type of data can be aggregated to a level of interest. For example, in the study of associations between $PM_{2.5}$ and adverse health effects, health outcomes are generally collected over geographic regions and point-referenced $PM_{2.5}$ values are aggregated to the level of the health outcomes (Fuentes *et al.*, 2005). In this application, we apply the approximation methods described in the previous sections to aggregated $PM_{2.5}$ data at the county level in California (see Figure 3). Let us define $Y(s)$ be a continuous Gaussian

Table 5: Approximations of a GGM with sill=1 and with various values of the range parameter using a GMRF with a 6th order neighborhood. Parameter estimates of the GMRF.

| | range=2 (range<MDN) | | range=3 (range=MDN) | | range=5 (range>MDN) | |
|---|---|---|---|---|---|---|
| method | $\alpha$ (SE) | $\tau$ (SE) | $\alpha$ (SE) | $\tau$ (SE) | $\alpha$ (SE) | $\tau$ (SE) |
| CSDS | 0.99(0.092) | 0.175 (1.054) | 0.99(0.016) | 0.140 (0.634) | 0.99(0.0003) | 0.162 (0.331) |
| KLS | 0.99(0.009) | 0.153(0.025) | 0.99( 0.002) | 0.106(0.017) | 0.99(0.0009) | 0.065 (0.010) |
| SRDC | 0.81(0.001) | 2.818(0.043) | 0.90 (0.001) | 2.413(0.042) | 0.96 (0.0007) | 1.798 (0.041) |
| KLP | 0.98 (0.019) | 0.391(0.055) | 0.99(0.009) | 0.270 (0.038) | 0.99( 0.021) | 0.166 (0.023) |

Table 6: Approximations of a GGM with sill=1 and with various values of the range parameter using a GMRF with a 6th order neighborhood. This table shows the performance in terms of prediction of the GMRF.

| | AMSPE | | |
|---|---|---|---|
| method | range=2 | range=3 | range=5 |
| CSDS | 0.602 (0.103) | 0.475 (0.103) | 0.404 (0.125) |
| KLS | 0.601 (0.103) | 0.475 (0.103) | 0.404 (0.125) |
| SRDC | 0.643 (0.103) | 0.492 (0.112) | 0.410 (0.130) |
| KLC | 0.602 (0.121) | 0.475 (0.103) | 0.404 (0.125) |

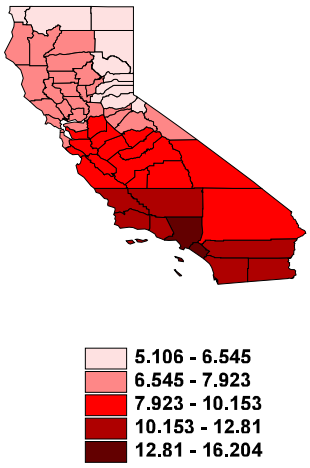
process with the $PM_{2.5}$ values, with mean function $\mu(s)$ and covariance function $c(s_i, s_j)$ for $s_i, s_j \subset D \in R^d$ and let $|B| = \int_B ds$ denote the area of county $B$. The aggregated process $Y(B) = \int_B Y(s)ds$ has a multivariate normal distribution with mean function $\mu(B)$ and covariance function $\Sigma(B_i, B_j)$,

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

$$\Sigma(B_i, B_j) = cov(Y(B_i), Y(B_j) = |B_i|^{-1}|B_j|^{-1} \int_{B_i} \int_{B_j} c(||s_i - s_j||)ds_i ds_j.$$

To obtain the aggregated $PM_{2.5}$ data, $Y(B)$, we predict $PM_{2.5}$ at 20 locations within each county and then we get the average of the predicted values. Figure 3 displays the county level $PM_{2.5}$ values, showing high concentrations in the Southern California area.

For county aggregated $PM_{2.5}$ values the first and second order neighborhoods are used.




(a)

Figure 3: a) county averaged $PM_{2.5}$ values

Table 7: MSPE in the approximations of aggregated $PM_{2.5}$ using GMRFs with a first and second order neighborhood structures.

| | MSPE | |
|---|---|---|
| method | first order neighborhood | second order neighborhood |
| CSDS | 1.29 | 1.42 |
| KLS | 0.69 | 1.35 |
| SRDC | 0.58 | 1.22 |
| KLP | 1.04 | 1.23 |

After estimating the parameters of the GMRF, the performance of the different methods are evaluated through the MSPE. Table 7 is the MSPE obtained from approximations of the aggregated $PM_{2.5}$ county values using a GMRF. In terms of the prediction of point-referenced $PM_{2.5}$ values, the smallest MSPE is obtained by the SRDC method. We study the effect of the neighborhood selection on the MSPE of the aggregated $PM_{2.5}$. In our example, the first order neighborhood gives smaller MSPE than the second order neighborhood, and the SRDC gives the smallest MSPE.

# 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 four approximation methods based on spectral density functions and covariance functions, which are evaluated in terms of an average MSPE. The simulation results show that approximations of GGMs by GMRFs give smaller average MSPE than those of GMRFs by GGMs. In the approximations of GMRFs by GGMs, the estimated range parameter increases as $\alpha$ increases. This result might be expected since $\alpha$ and the range parameter control the strength of the correlation of the data in GMRFs and GGMs, respectively. We also study the impact of the neighborhood selection on the estimated range parameter in the approximations of GMRFs by GGMs. Although the parameter estimates are different depending on the criteria used, the results in terms of prediction were not very different. In most cases, the SRDC method gave the largest AMSPE. In the approximations of GMRFs by GGMs, as the neighborhood order increases, the AMSPE becomes smaller, while the AMSPE becomes larger as $\alpha$ increases. The approximations of GGMs with various values of the range parameter using GMRFs gave similar results with the four approximation methods. The smallest AMSPE was obtained for the larger range value. In our application, we approximated county-level averaged $PM_{2.5}$ using GMRFs. The smallest MSPE was obtained using the SRDC criterion.

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.

# Acknowledgement

# References


Arfken, G., and Weber, H.J. (1995) *Mathematical Methods for Physicists,* 4th ed. San Diego, California: Academic Press.

Besag, J.(1974). Spatial interaction and the statistical analysis of lattice systems, *journal of the royal statistical society*, **36**, 192-236.

Besag, J.(1981). On a system of two-dimensional recurrence equations, *journal of the royal statistical society*, **43**, 302-309.

Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, **82**, 733-746.

Besag, J., York, J.C., and Mollie, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics*, **43**, 1-59.

Best, N.G., Ickstadt, K., Wolpert, R.L., and Briggs, D.J., Combining models of health and exposure data: the SAVIAH study, in *Spatial epidemiology: methods and applications,* eds. P Elliott, JC Wakefield, NG Best, and DJ Briggs, Oxford; New York: Oxford University Press, pp. 393-414, 2000 .

Carlin, B.P. and Banerjee, S., Hierarchical multivariate CAR models for spatio-temporally correlated survival data (with discussion), in *Bayesian Statistics 7*, eds. J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith, and M. West, Oxford: Oxford University Press, pp. 45–63, 2003.

Cressie, N., *Statistics for spatial data.* New York: Wiley, 1991.

Fuentes, M., Song, H.R., Ghosh, S., Holland, D. and Davis, J. (2005). Spatial association between speciated fine particles and mortality. *Biometrics*, to appear.

Griffith, D. and Csillag, F. (1993). Exploring relationships between semivariogram and spatial autoregressive models, *papers in regional science*, **72**, 283-295.

Hrafnkelsson, B. and Cressie, N. (2003). Hierarchical modeling of count data with application to nuclear fall-out, *Environmental and ecological statistics*, **10**, 179-200.

Kaiser, M. S. (2002). Ananlysis of particulate matter air pollution using Markov random field models of spatial dependence. *Environmetrics,* **13**, 615-628.

Journel, A.G. and Huijbregts, Ch.J. (1978). *Mining Geostatistics,* Academic Press: London.

Martin, R.J. and Dwyer, J.N. (1994). Approximations to the covariance properties of processes averaged over irregular spatial regions, *Communications in statistics-theory and methodologies,* **23:3**, 913-945.

Rue, H. and Tjelmeland, H.(2002). Fitting gaussian markov random fields to gaussian fields , *Scandinavian journal of statistics*, **29**, 31-49.

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

Tolbert, P., Mulholland, J., MacIntosh, D., Xu, F., Daniels, D., Devine, O., Carlin, B.P., Klein, M., Dorley, J., Butler, A., Nordenberg, D., Frumkin, H., Ryan, P.B., and White, M. (2000). Air pollution and pediatric emergency room visits for asthma in Atlanta. *American Journal of Epidemiology,* **151:8**, 798-810.

Werner, L. (2004) Spatial inference for non-lattice data using Markov Random fields, PhD thesis, Lund University.