NIH Public Access
Author Manuscript
Stat Methodol. Author manuscript; available in PMC 2013 August 14.

Published in final edited form as:
Stat Methodol. 2012 ; 9(1-2): 265–274. doi:10.1016/j.stamet.2011.01.007.

Nonparametric Bayesian models for a spatial covariance

Brian J Reich[1] and **Montserrat Fuentes**
Department of Statistics, North Carolina State University

Abstract

A crucial step in the analysis of spatial data is to estimate the spatial correlation function that determines the relationship between a spatial process at two locations. The standard approach to selecting the appropriate correlation function is to use prior knowledge or exploratory analysis, such as a variogram analysis, to select the correct parametric correlation function. Rather that selecting a particular parametric correlation function, we treat the covariance function as an unknown function to be estimated from the data. We propose a flexible prior for the correlation function to provide robustness to the choice of correlation function. We specify the prior for the correlation function using spectral methods and the Dirichlet process prior, which is a common prior for an unknown distribution function. Our model does not require Gaussian data or spatial locations on a regular grid. The approach is demonstrated using a simulation study as well as an analysis of California air pollution data.

Keywords

Covariance estimation; Dirichlet process prior; Particulate matter; Spectral density

1 Introduction

An essential element of spatial data analysis is to estimate the spatial correlation function that determines the relationship between a spatial process at two locations. Let $y(s) \in \mathbb{R}$ be the response at spatial location $s = (\mathbf{s}_1, s_2)' \in \mathbb{R}^2$. A common spatial model is $y(s) = x(\mathbf{s})'\boldsymbol{\beta} + \mu(\mathbf{s}) + \varepsilon(\mathbf{s})$, where $x(\mathbf{s})$ is the p-vector of known covariates with regression parameters $\boldsymbol{\beta}$, $\mu(\mathbf{s})$ is a spatial process, and $\varepsilon(s) \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$ is error. We assume that the spatial process $\mu(\mathbf{s})$ is a mean-zero Gaussian process with stationary covariance function $\mathrm{Cov}(\mu(\mathbf{s}),\mu(\mathbf{s} + \mathbf{h})) = \tau^2 C(\mathbf{h})$. The standard approach to selecting the appropriate spatial correlation function $C(\mathbf{h})$ is to use prior knowledge or exploratory analysis of the residuals, such as a variogram analysis, to select the correct parametric correlation function, e.g., exponential, spherical, or Matérn (Cressie, 1993; Banerjee, Carlin, and Gelfand, 2004). However, this can be challenging if the number of covariates is large, since covariate estimates can change dramatically with different spatial models (Reich, Hodges, and Zadnik, 2006), or if the responses $y(\mathbf{s})$ are non-Gaussian, for example in a spatial analysis of binary data.

Rather that selecting a particular parametric correlation function, we treat the covariance function as an unknown function to be estimated from the data. Directly modeling the spatial

© 2011 Elsevier B.V. All rights reserved

[1]Corresponding author: brian_reich@ncsu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

correlation function is challenging because it must be nonnegative definite. That is, $\sum_{i=1}^{n}\sum_{j=1}^{n} C(s_i - s_j)a(s_i)a(s_j) \geq 0$ for any vector $(a(\mathbf{s}_1),\ldots, a(\mathbf{s}_n))'$. This requirement is difficult to check when specifying a spatial correlation. A more convenient way to specify a flexible model for the spatial covariance is through the corresponding spectral density, which is directly related to the correlation function but has fewer restrictions. When the spatial locations are on a grid, the periodogram is an asymptotically unbiased estimate of the spectral density, and many methods have been proposed to estimate the spectral density by smoothing the periodogram (e.g., Ripley, 1981; Schabenberger and Gotway, 2005). Recently, Zheng, Zhu, and Roy (2010) propose a nonparametric Bayesian prior for the spectral density of a Gaussian random field on a lattice.

As in Zheng, Zhu, and Roy (2010), we specify a nonparametric Bayesian prior for the spectral density. Our model does not require Gaussian responses or that the spatial locations be on a grid. We model the spectral density as in infinite mixture of simple spectral densities. By modeling the covariance function as unknown in a fully-Bayesian model for the spatial process, our uncertainty regarding the form of the correlation function is correctly propagated throughout the analysis, including the posterior of the regression coefficients and the posterior predictive distribution.

The paper proceeds as follows. In Section 2 we review spectral methods, and propose two models for the spectral density using the Dirichlet process prior and the Dirichlet process mixture priors in Section 2.2 and 2.3, respectively. The computational algorithm is described in Section 3. We compare methods using a simulation study in Section 4 and a spatial analysis of air pollution data in Section 5. Section 6 concludes.

2 Nonparametric priors for a spectral density

2.1 Review of spectral methods

In this section we provide a brief review of spectral methods for spatial statistics; a more comprehensive review can be found in Fuentes and Reich (2009). A key result in spectral analysis is Bochner's theorem, which states that a stationary covariance C(**h**) is non-negative definite if and only if there exists a positive, finite measure F so that

$$C(h)=\int_{R^2}\cos(\omega' h)F(d\omega), \quad (1)$$

where $\omega=(\omega_1,\omega_2)' \in \mathscr{R}^2$ is a frequency. Bochner's theorem directly relates the covariance function to the spectrum, $F(\boldsymbol{\omega})$. We assume the spectrum permits a density with respect to Lebesgue measure, so that $F(d\omega) = \tau^2 f(\boldsymbol{\omega})$, where $\tau > 0$ and f is the spectral density with $\int f(\omega)d\omega=1$. Sections 2.2 and 2.3 give nonparametric priors for the spectral density, and thus nonparametric priors for the spatial covariance function.

The spatial process is real if and only if the spectral density is even, i.e., $f(\omega)=f(-\omega)$. The spectral representation theorem states that the real-valued process $\mu(s)$ can be written

$$\mu(s)=\int_{\mathscr{R}^2}\cos(\omega' s)dU(\omega)+\int_{\mathscr{R}^2}\sin(\omega' s)dW(\omega), \quad (2)$$

where U and W are independent Gaussian processes with mean zero, orthogonal increments, and $E(|dU(\omega)|^2)=E\left(|dW(\omega)|^2\right)=F(\omega)<\infty$. The spectral representation formulates the spatial process as a convolution of trigonometric basis functions and stochastic processes in the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

frequency domain with independent increments. From this model, clearly $\mu(\mathbf{s})$ is a Gaussian process with mean zero and covariance C(**h**).

2.2 Dirichlet process prior for the spectral density

The spectral density f is a density that controls the distribution of mass in the spectral domain. Therefore, we model f using the Dirichlet process (DP) prior (Ferguson, 1973; Ferguson, 1974) which is often used in Bayesian nonparametrics as a prior for an unknown distribution. Using the stick-breaking representation (Sethuraman, 1994), the Dirichlet process prior for f can be written as the infinite mixture

$$f(\omega)=\sum_{j=1}^{\infty} p_j \delta(\alpha_j), \quad (3)$$

where the mixture probabilities p_j satisfy $\sum_{j=1}^{\infty} pj=1$ almost surely and $\delta(\boldsymbol{\alpha}_j)$ is the point mass distribution at $\boldsymbol{\alpha}_j \in \mathbb{R}^2$. The mixture probabilities have priors $p_1 = V_1$ and $p_j = V_j \Pi_{k<j}(1 - V_k)$ for $j > 1$, where $V_j \overset{\text{iid}}{\sim} \text{Beta}(1, D)$. The V_j "break the stick" into pieces that sum to one, with $\prod_{k<j}(1 - Vk) = 1 - \sum_{k=1}^{j-1} pj$ the proportion of the stick not attributed to the first $j - 1$ terms and V_j the proportion of the remaining mass attributed to term j. The frequencies $\alpha_j \overset{\text{iid}}{\sim} f_\theta$, where f_θ is a parametric centering distribution with hyperparameters θ.

To ensure that the spectral density is even, and thus a real-valued process, the prior could be modified as

$$f(\omega)=\frac{1}{2}\sum_{j=1}^{\infty} p_j\{\delta(\alpha_j)+\delta(-\alpha_j)\}, \quad (4)$$

where p_j and $\boldsymbol{\alpha}_j$ are modeled as before. Due to the DP's discreteness, the stochastic integral (2) becomes the countable sum

$$\mu(s)=\sum_{j=1}^{\infty}\cos(\alpha'_j s)U_{1j}+\cos(-\alpha'_j s)U_{2j}+\sin(\alpha'_j s)W_{1j}+\sin(-\alpha'_j s)W_{2j}, \quad (5)$$

where $U_{1j}, U_{2j}, W_{1j}, W_{2j} \overset{\text{iij}}{\sim} N(0, \tau^2 p_j/2)$, independent over j. This is equivalent to

$$\mu(s)=\sum_{j=1}^{\infty}\cos(\alpha'_j s)U_j+\sum_{j=1}^{\infty}(\alpha'_j s)W_j, \quad (6)$$

where $\alpha_j \overset{\text{iid}}{\sim} f_\theta$ and $U_j, W_j \overset{iid}{\sim} N(0, \tau^2 p_j)$, independent over j. We use this parameterization since it has fewer terms. The spatial process μ is a linear combination of sine and cosine basis functions. The sine and cosine functions are random functions of the unknown frequencies $\alpha_j \overset{iid}{\sim} f_\theta$. Note that without the restriction that the spectral density is even, the second sum in (6) would be the imaginary part of the complex process $\mu(\mathbf{s})$.

The spatial covariance is a random function of $\boldsymbol{\alpha}_j$ and V_j. Given $\boldsymbol{\alpha}_j$ and V_j (through p_j),

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$\mathrm{Cov}(\mu(s),\mu(t))=\tau^2\sum_{j=1}^{\infty}p_j\left[\cos(\alpha'_j s)\cos(\alpha'_j t)+\sin(\alpha'_j s)\sin(\alpha'_j t)\right]=\tau^2\sum_{j=1}^{\infty}p_j\cos(\alpha'_j h), \quad (7)$$

where $\mathbf{h} = \mathbf{s} - \mathbf{t}$. Therefore the conditional covariance is stationary and $V(\mu(\mathbf{s})) = \tau^2$ for all $\mathbf{s}$. Also, we note that although the spectral density is discrete, the resulting covariance is a continuous function of $\mathbf{h}$.

The prior mean and variance (with respect to ω_j and V_j) of the covariance function are

$$\begin{aligned} E[\mathrm{Cov}(\mu(s)),(\mu(t))]&=\tau^2 C_\theta(h)\\ V[\mathrm{Cov}(\mu(s)),(\mu(t))]&=\tau^4 V_\theta(h)/(D+1) \end{aligned} \quad (8)$$

where

$$\begin{aligned} C_\theta(h)&=E(\cos(\alpha' h))=\int_{R^2}\cos(\alpha' h)f_\theta(\alpha)d\alpha \text{ and}\\ V_\theta(h)&=\mathrm{Var}(\cos(\alpha' h))=\int_{R^2}\cos(\alpha' h)f_\theta(\alpha)d\alpha - C_\theta(h)^2 \end{aligned} \quad (9)$$

are the first two spectral moments. The prior for the spatial covariance function is centered on the covariance corresponding to the centering distribution f_θ. The variance of the spatial covariance is a function of both the variability in the centering distribution though V_θ and the stick-breaking parameter D; large D implies small prior variance and thus strong centering on the parametric covariance C_θ.

The infinite sum in (6) cannot be fit in practice. Although there are infinitely-many terms, they are stochastically decreasing, e.g., the prior mean of p_j is $[1/(D+1)][D/(D+1)]^{j-1}$. Therefore, for computational purposes we truncated the mixture to M terms by setting $\upsilon_M = 1$. This finite mixture approximates the full model arbitrarily well, and the posterior distribution of the mass on the final term, p_M, can be used to determine the number of terms needed to provide a sufficient approximation to the full model. We increase M until the posterior mean of p_M is small, say 0.001. For large data sets with strong spatial correlation, $M \ll n$ may provide a sufficient approximation. This gives a natural dimension reduction which is useful for large spatial data sets. A disadvantage of this truncation is that the process μ becomes periodic, which is unrealistic in most situations. However, when interest is restricted to a particular finite spatial domain, e.g., California as in Section 5, periodicity is not problematic unless there is interest in projecting outside of domain of interest.

For the centering distribution we use the bivariate-t centered at zero with 2 × 2 scale matrix $\frac{\rho^2}{2\nu+1}I_2$ and degrees of freedom $2\nu + 1$, i.e., $\theta=(\rho, \nu)$ and

$$f_\theta(\alpha)\propto\left[1+\frac{\|\alpha\|^2}{\rho^2}\right]^{-(\nu+1)}. \quad (10)$$

This corresponds to the Matérn (Matérn, 1960) correlation. The Matérn has two parameters, $\rho > 0$ controls the range of the spatial covariance, and $\nu > 0$ controls the smoothness. The Matérn has several interesting spatial cases, including the Gaussian correlation $C(\mathbf{h}) = \exp(-4\|\mathbf{h}\|^2/\rho^2)$ with $\nu = \infty$ and exponential correlation $C(\mathbf{h}) = \exp(-\|\mathbf{h}\|/\rho)$ with $\nu = 1/2$.

2.3 Dirichlet process mixture prior for the spectral density

The Dirichlet process prior for the spectral density is discrete. In some settings a continuous spectral density is desirable, e.g., to avoid periodicity as discussed in Section 2.2. Therefore

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

we also consider a Dirichlet process mixture model (Antoniak, 1974) for the spectral density. The Dirichlet process mixture model replaces the discrete point mass distribution $\delta(\boldsymbol{\alpha}_j)$ with a continuous parametric distribution $g\gamma_j(\omega|\alpha_j)$ with location $\boldsymbol{\alpha}_j$ and parameters $\boldsymbol{\gamma}_j$. As in (4), we restrict the spectral density to be even, and thus spatial process to be real, by restricting each component of the mixture to be even. To do this, we select $g\gamma(\omega|\alpha)$ to be a location family with location $\boldsymbol{\alpha}$ that is even in $\boldsymbol{\alpha}$, i.e., $g\gamma(\omega|\alpha)=g\gamma(-\omega|-\alpha)$. Then

$$f(\omega)=\frac{1}{2}\sum_{j=1}^{\infty}p_j\left[g\gamma_j(\omega|\alpha_j)+g\gamma_j(\omega|-\alpha_j)\right], \quad (11)$$

where p_j and $\boldsymbol{\alpha}_j$ are modeled as in Section 2.2.

A natural choice for g is be a bivariate non-central t with location $\boldsymbol{\alpha}_j$, scale matrix $\frac{\phi_j^2}{2\xi_j+1}I_2$ and degrees of freedom $2\xi_j+1$, so that $\gamma_j=(\phi_j,\xi_j)$. Clearly, this is a location family satisfying $g\gamma(\omega|\alpha)=g\gamma(-\omega|-\alpha)$. This mixing distribution has the advantage of reducing to the Matérn spectral density if $p_1=1$ and $\boldsymbol{\alpha}_1=0$. Also, restricting $\boldsymbol{\alpha}_j$ so that $\alpha_j=(\alpha_j,\alpha_j)'$ gives an isotropic correlation.

To study the correlation induced by this spectral density model, let $K\boldsymbol{\gamma}(\mathbf{h})=\int\cos(\mathbf{h}'\boldsymbol{\omega})g\boldsymbol{\gamma}(\boldsymbol{\omega}|\mathbf{0})d\boldsymbol{\omega}$ be the correlation function corresponding to the centered mixing distribution with $\boldsymbol{\alpha}=\mathbf{0}$. For example, if g is bivariate t as above, then $K\boldsymbol{\gamma}_j$ is the Matérn correlation with range $\boldsymbol{\phi}_j$ and smoothness ξ_j The Appendix shows that

$$C(h)=\sum_{j=1}^{\infty}p_j\cos(\alpha'_j h)K\gamma_j(h). \quad (12)$$

The correlation is a convex combination of correlations of the form $\cos(\boldsymbol{\alpha}'_j\mathbf{h})K\boldsymbol{\gamma}_j(\mathbf{h})$. Assuming a non-central bivariate t for g, each correlation is the product of a Matérn correlation $K\boldsymbol{\gamma}_j(\mathbf{h})$ and the trigonometric function $\cos(\boldsymbol{\alpha}'_j\mathbf{h})$. Taking $\xi_j=0.5$ gives the exponentially-dampened cosine function of Zastavnyi (1993). The central Matérn correlation is always positive and decreasing. Thus a mixture of central Matérn correlations is also positive and decreasing. In contrast, using a non-central t spectral density results in the Matérn correlation being multiplied by $\cos(\boldsymbol{\alpha}'_j\mathbf{h})$, and thus the non-central Matérn correlation $\cos(\boldsymbol{\alpha}'_j\mathbf{h})K\boldsymbol{\gamma}_j(\mathbf{h})$ is oscillating and can be negative. This provides greater flexibility than the central Matérn.

Modeling the spectral density as a mixture of non-central t densities can approximation a wide class of spectral densities, and thus a mixture of non-central Matérn correlations can approximate a wide class of correlation functions. The even discrete DP model in (4) spans the entire class of even distributions. Therefore, convolving the discrete DP with the mixture distribution gives non-zero mass to any even spectral density (Hoff, 2003), and therefore spans the full class of covariance functions for a real process.

Figure 1 plots an example covariance of the form (12) to illustrate the flexibility of this covariance model. The mixture is truncated to a three-component mixture and the correlations are isotropic with $\boldsymbol{\alpha}_j=(\alpha_j,\alpha_j)'$, and are therefore plotted against $h\in\mathbb{R}^1$. The first correlation is the usual Matérn with smoothness parameter 10. The second and third are non-central Matérn with $\alpha_2=50$ and $\alpha_3=25$. The resulting mixture decays rapidly at the origin, plateaus between $h=0.10$ and $h=0.25$, and then descends to zero at $h=0.5$.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

As in Section 2.2, we must truncate the mixture at M terms, where M is chosen so that $p_M \approx 0$. The DP mixture model can typically be approximated by far fewer terms than the discrete DP model, since the covariances for each mixture term are more flexible for the mixture model than the discrete DP (a non-central Matérn compared to a single-point spectral density).

Finally, we note that although we assume the spatial process μ is Gaussian, the prior for the spatial covariance can easily be applied to non-Gaussian data. For example, if the observations are binary, we may use the spatial probit model $P[y(\mathbf{s}) = 1] = \Phi[\mathbf{x}(\mathbf{s})\boldsymbol{\beta} + \mu(\mathbf{s})]$, where Φ is the standard normal distribution function. With a slight modification, the computational algorithm in Section 3 would apply to this model.

3 Computational details

The MCMC updates for all parameters are done using either Gibbs or random walk Metropolis sampling. The most problematic parameters are the frequencies $\boldsymbol{\omega}_j$ in Section 2.2's discrete DP model. Updating these parameters conditioned of the coefficients (U_j, W_j) lead to poor MCMC mixing. Therefore, we update $\boldsymbol{\omega}_j$ and (U_j, W_j) simultaneously using blocked Metropolis-Hastings sampling. We first generate the candidate frequency as $\omega_j^* \sim N(\omega_j, c^2 I_2)$ where $\boldsymbol{\omega}_j$ is the current value in the chain. Then the candidate for (U_j, W_j) is generated from the full conditional posterior of (U_j, W_j) given ω_j^*. We tune c to give acceptance rate around 40%. This dramatically improved MCMC convergence. Convergence is monitored using trace plots of the deviance and several representative parameters. For the simulation study in Section 4 we generate 5,000 iterations and discard the first 1,000 as burn-in; for the data analysis in Section 5 we generate 10,000 samples and discard the first 5,000 as burn-in. Code is available from the first author by request.

4 Simulation study

We conduct a brief simulation study to compare covariance models. We generate n_t independent realizations of a spatial process at n_s locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ with $s_i \overset{\text{iid}}{\sim} \text{Uniform}[0, 1[^2$. Let

$$y_t(s_i) = x(s_i)'\beta + \mu_t(s_i) + \varepsilon_t(s_i) \quad (13)$$

be the observation at location $\mathbf{s}_i$ for replication t, where $\mathbf{x}(\mathbf{s}_i)$ are spatial covariates and $\boldsymbol{\beta}$ are the corresponding regression coefficients, $\mu_t(\mathbf{s})$ are independent over t and correlated across space with mean zero and variance $\tau^2 = 1$, and $\varepsilon_t(s_i) \overset{\text{iid}}{\sim} N(0, 0.1^2)$ is error. We include only the intercept, i.e., $\mathbf{x}(\mathbf{s}_i) = 1$, and set $\boldsymbol{\beta} = 0$.

We generate $S = 25$ data sets with $n_s = 100$ sites and $n_t = 100$ replications with five isotropic spatial correlation models for $\mu_t(\mathbf{s})$, $C(h)$, where $h \in \mathscr{R}^1$ is spatial distance. The correlations are:

1. Exponential: Matérn(0.2,0.5)
2. Matérn: Matérn(0.1,2)
3. Spherical: $C(h) = \left[1 - \frac{3}{2}\frac{h}{\rho} + \frac{1}{2}\left(\frac{h}{\rho}\right)^3\right] I(h<\rho)$ with $\rho = 0.3$
4. Wave: $C(h) = I(h=0) + I(h>0)\frac{1}{h/\rho}\sin(h/\rho)$ with $\rho = 0.1$
5. Mixture: the three-component mixture plotted in Figure 1

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

These correlations are plotted in Figure 2. For each data set, half of the observations are used as training data to fit the model (selected randomly over space and replication) and the remaining locations are used as the test set.

Since the primary objective of covariance estimation is often to estimate the underlying spatial process, we compare models using coverage probabilities of 90% intervals and the mean square error of the posterior mean for $\mu_t(\mathbf{s})$, separately for training and testing sets. We fit two models. The first is Dirichlet process mixture (DPM) model described in Section 2.3. Although much of the intuition regarding the prior for the correlation function is derived from the discrete DP model in Section 2.2, we choose to focus on the continuous DPM spectral density in simulation study and data analysis in Section 5. For the DPM model we assume isotropy by fixing $\alpha_j=(\alpha_j, \alpha_j)'$ and truncate the infinite sum at $M = 5$ terms. Inspecting the posterior of p_M suggests that this provides a reasonable approximation to the full model. As a benchmark, we also include the usual isotropic Mat′ern. For all models we select the following uninformative priors: N(0, 10^2) for the mean parameters (i.e., the intercept) in $\boldsymbol{\beta}$, InvGamma(0.1,0.1) for all variances, Unif(0,2) for all spatial ranges, and Unif(0.5,10) for all Matérn smoothness parameters.

Table 1 presents the results. The DPM model is comparable to the Matérn correlation for the first two simulations where the true correlation is Matérn. The Matérn correlation also performs well under the spherical correlation, which decays like the Matérn near the origin, but is exactly zero for large distances. The DPM model gives considerably smaller mean squared error for the non-decreasing wave and mixture correlations. The relative mean square error of the DPM compared to the Matérn is especially small for the training sites for the mixture correlation and the testing sites for the wave correlation. The DPM's coverage probabilities are slightly less than 90% for the training sites, but near 90% for all correlations for the testing sites.

5 Analysis of fine particular matter

We illustrate the proposed models for the spatial covariance using fine particulate matter data. Fine particular matter, $PM_{2.5}$, are particles less than 2.5 micrometers in aerodynamic diameter. $PM_{2.5}$ concentration has been linked to several adverse health outcomes (Bates, Baker-Anderson, and Sizto, 1990; Ostro et al., 1991; Dockery, Schwartz, and Spengler, 1992; Schwartz, 1994; Pope, Dockery, and Schwartz, 1995) and is currently one of six air pollutants regulated by the US EPA. Spatial modeling of $PM_{2.5}$ is important because monitors are typically operational only every two, three, or six days, so spatial modeling to interpolate $PM_{2.5}$ at inactive monitoring locations is needed for health studies.

The daily data are from California from June–August, 2008, obtained from the California EPA at http://www.arb.ca.gov/homepage.htm. To apply a stationary spatial model we first project the spatial locations to a two-dimensional surface using the Mercator projection, and then scale them to the unit square. We analyze $y_t(\mathbf{s}) = \log[Z_t(\mathbf{s}) + 1]$, where $Z_t(\mathbf{s})$ is the observed concentration on day t at site $\mathbf{s}$, to justify the normality assumption. The data are modeled as in (13), except that we include in x(s) the spatial covariates latitude, longitude, and elevation, as well and the squares of latitude and longitude. There at n_s =84 spatial locations and n_t= 92 days. Of the 84 * 92 = 7728 potential observations, 2431 (31%) are observed. The data are plotted in Figure 3.

We fit the Matérn and DPM models described in Section 4, using the same priors as in Section 4 except that we use $M = 20$ mixture terms in the DPM model. Due to the regular sampling scheme at most stations, there are few consecutive observations to study temporal correlation. Therefore, we model the spatial effects $\mu_t(\mathbf{s})$ as independent over time, and

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

having the same spatial correlation each day. We compare correlation models by withholding 20% of the observations as a test set, selected randomly across space and time. We compute prediction mean squared error (MSE), mean absolute deviation (MAD), and coverage probabilities of 90% intervals. The DPM model has slightly smaller MSE and MAD (0.122 and 0.211) than the Matérn (0.126 and 0.214). Both models have coverage probability 0.87.

Figure 4 plots the posterior mean of the spatial correlation for the two correlation models. The posterior mean (95% interval) for the Matérn correlation parameters are 0.37 (0.31, 0.44) for the range and 0.52 (0.50, 0.54) for the smoothness parameter. Compared to the Matérn, the DPM model has higher posterior mean correlation for middle distances with $0.2 < h < 0.6$ and lower correlation at long distances with $h > 0.8$. The posterior standard deviation of the correlations is large relative to these differences, but the added flexibility of the DPM spatial correlation results in a modest improvement in prediction MSE.

6 Discussion

In this paper, we present flexible priors for a spatial correlation by applying the Dirichlet process prior and the Dirichlet process mixture prior to the spectral density. Using simulated data we demonstrate that adequately modeling the correlation function can improve estimation of the underlying spatial process. The methods are illustrated using a spatial analysis of fine particular matter.

In our analysis of $PM_{2.5}$ data we did not account for autocorrelation. It would be straightforward to model the spatiotemporal correlation function via the three-dimensional spectral density. However, this would lead to a challenging computational problem. It may be possible to use a reduced-dimensional Dirichlet Process model with $M \ll n$ to fit this model.

Another area of future work is allowing the spectral density to be nonstationary and nonparametric. Several methods have been proposed to model a spatially-varying density function (Gelfand, Kottas, and MacEachern, 2005; Griffin and Steel, 2006; Reich and Fuentes, 2007; Dunson and Park, 2008). These priors could be used to specify a spatially-varying spectral density, which would allow the spectral density to vary by location, but be smoothed towards the spectral densities at nearby sites.

Appendix – Covariance of the DPM spectral density

First we consider a single mixture component, i.e., we derive the correlation function corresponding to the spectral density $0.5g(\boldsymbol{\omega}|\boldsymbol{\alpha}) + 0.5g(\boldsymbol{\omega}|-\boldsymbol{\alpha})$. Using the sum-difference formula $cos(u \pm v) = cos(u)\cos(v) \mp \sin(u)\sin(v)$, the induced covariance is

$$\begin{aligned} C(h) = E_\omega[\cos(\omega' h)] &= \tfrac{1}{2}E_{\omega_{(+)}}\left[\cos(\omega'_{(+)}h)\right] + \tfrac{1}{2}E_{\omega_{(-)}}\left[\cos(\omega'_{(-)}h\right] \\ &= \tfrac{1}{2}E_\delta[\cos([\delta+\alpha]' h] + \tfrac{1}{2}E_\delta[\cos([\delta-\alpha]' h)] \\ &= E_\delta[\cos([\delta-\alpha]' h)] \\ &= E_\delta[\cos(\delta' h)\cos(\alpha' h) + \sin(\delta' h)\sin(\alpha' h)] \end{aligned} \tag{14}$$

where $\boldsymbol{\omega}_{(+)}$ follows g with location $\boldsymbol{\alpha}$, $\boldsymbol{\omega}_{(-)}$ follows g with location $-\boldsymbol{\alpha}$, $\boldsymbol{\delta}$ follows g with $\alpha=0$, $E_\delta[\cos([\delta+\alpha]' h)] = E_\delta[\cos([\delta-\alpha]' h)]$ since both $\cos(\bullet)$ and $g(\bullet|\mathbf{0})$ are even, $E_{\boldsymbol{\delta}}[\sin(\boldsymbol{\delta}'\mathbf{h})] = 0$ since $g(\bullet|\mathbf{0})$ is even, and $C_{\boldsymbol{\delta}}$ is the correlation induced by $g(\bullet|\mathbf{0})$. The covariance for the mixture is then $\sum_{j=1}^{M} p_j \cos(\alpha'_j h) C_{\delta_j}(h)$.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

References

Antoniak CE. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. The Annals of Statistics. 1974; 2:1152–1174.

Banerjee, S.; Carlin, BP.; Gelfand, AE. Hierarchical modeling and analysis for spatial data. Chapman and Hall; New York: 2004.

Bates DV, Baker-Anderson M, Sizto R. Asthma Attack Periodicity: A Study of Hospital Emergency Visits in Vancouver. Environmental Research. 1990; 51:51–70. [PubMed: 2298182]

Cressie, N. Statistics for spatial data. Wiley-Interscience; New York: 1993.

Dockery DW, Schwartz J, Spengler JD. Air pollution and daily mortality: associations with particulates and acid aerosols. Environmental Research. 1992; 59:362–373. [PubMed: 1464289]

Dunson DB, Park JH. Kernel stick-breaking processes. Biometrika. 2008; 95:307–323. [PubMed: 18800173]

Ferguson TS. A Bayesian analysis of some nonparametric problems. The Annals of Statistics. 1973; 1:209–230.

Ferguson TS. Prior distribution on spaces of probability measures. The Annals of Statistics. 1974; 2:615–629.

Fuentes, M.; Reich, BJ. Handbook of Spatial Statistics. CRC Press; 2009. Spectral Analysis for Spatial Data.

Gelfand AE, Kottas A, MacEachern SN. Bayesian nonparametric spatial modeling with Dirichlet process mixing. Journal of the American Statistical Association. 2005; 100:1021–1035.

Griffin JE, Steel MFJ. Order-based dependent Dirichlet processes. Journal of the American Statistical Association. 2006; 101:179–194.

Hoff PD. Nonparametric Estimation of Convex Models via Mixtures. Annals of Statistics. 2003; 31:174–200.

Matérn, B. Meddelanden fr an Statens Skogsforskningsinstitut. 2. Vol. 49. Springer-Verlag; Berlin: 1960. Spatial Variation. Almaenna Foerlaget, Stockholm(1986)

Ostro BD, Lipsett MJ, Wiener MB, Selner JC. Asthmatic responses to airborne acid aerosols. American Journal of Public Health. 1991; 81:694–702. [PubMed: 1851397]

Pope CA, Dockery D, Schwartz J. Review of epidemiological evidence of health effects of particulate air pollution. Inhalation Toxicology. 1995; 47:1–18.

Reich BJ, Hodges JS, Zadnik V. Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. Biometrics. 2006; 62:1197–1206. [PubMed: 17156295]

Reich BJ, Fuentes M. A multivariate semiparametric Bayesian spatial modeling framework for hurricane surface wind fields. The Annals of Applied Statistics. 2007; 1:249–264.

Ripley, BD. Spatial Statistics. New York: Wiley; 1981.

Schabenberger, O.; Gotway, CA. Statistical Methods for Spatial Data Analysis. Boca Raton: Chapman & Hall; 2005.

Schwartz J. Air pollution and daily mortality: a review and meta analysis. Environmental Research. 1994; 64:36–52. [PubMed: 8287841]

Sethuraman J. A constructive definition of Dirichlet priors. Statistica Sinica. 1994; 4:639–650.

Zastavnyi VP. Positive definite functions depending on a norm. Russian Academy of Sciences Doklady Mathematics. 1993; 46:112–114.

Zheng Y, Zhu J, Roy A. Nonparametric Bayesian inference for the spectral density function of a random field. Biometrika. 2010; 97:238–245.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

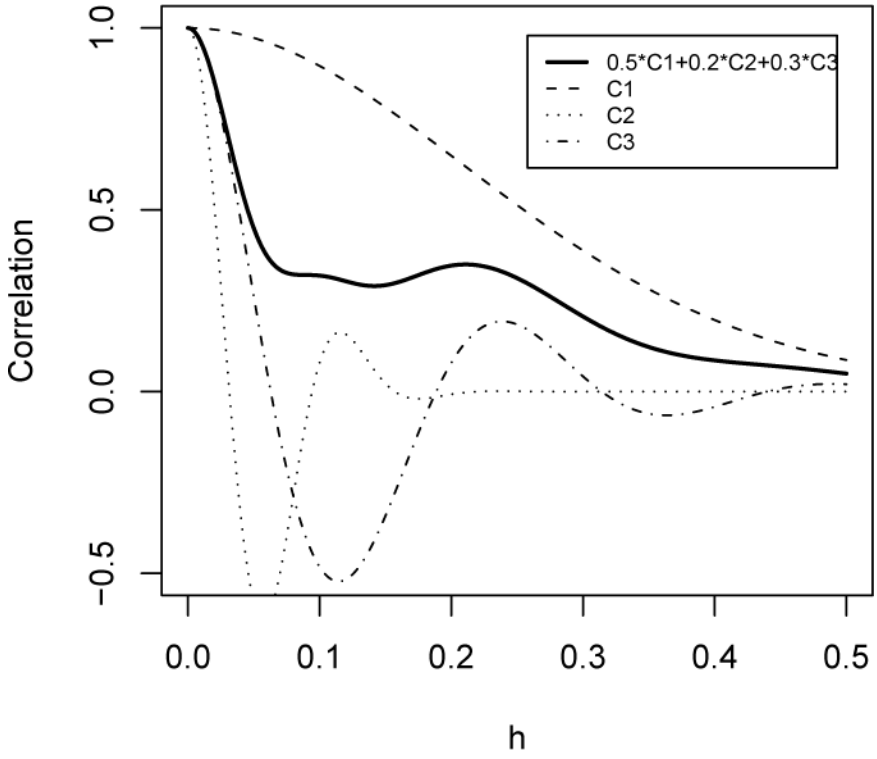

Figure 1.

Example of a mixture of non-central Matérn correlations. The three correlations (denoted "C1", "C2", and "C3") have location, scale, and smoothness (a_j, ϕ,ξ) equal to (0,0.05,10), (50,0.01,20), and (25,0.10,1), respectively.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

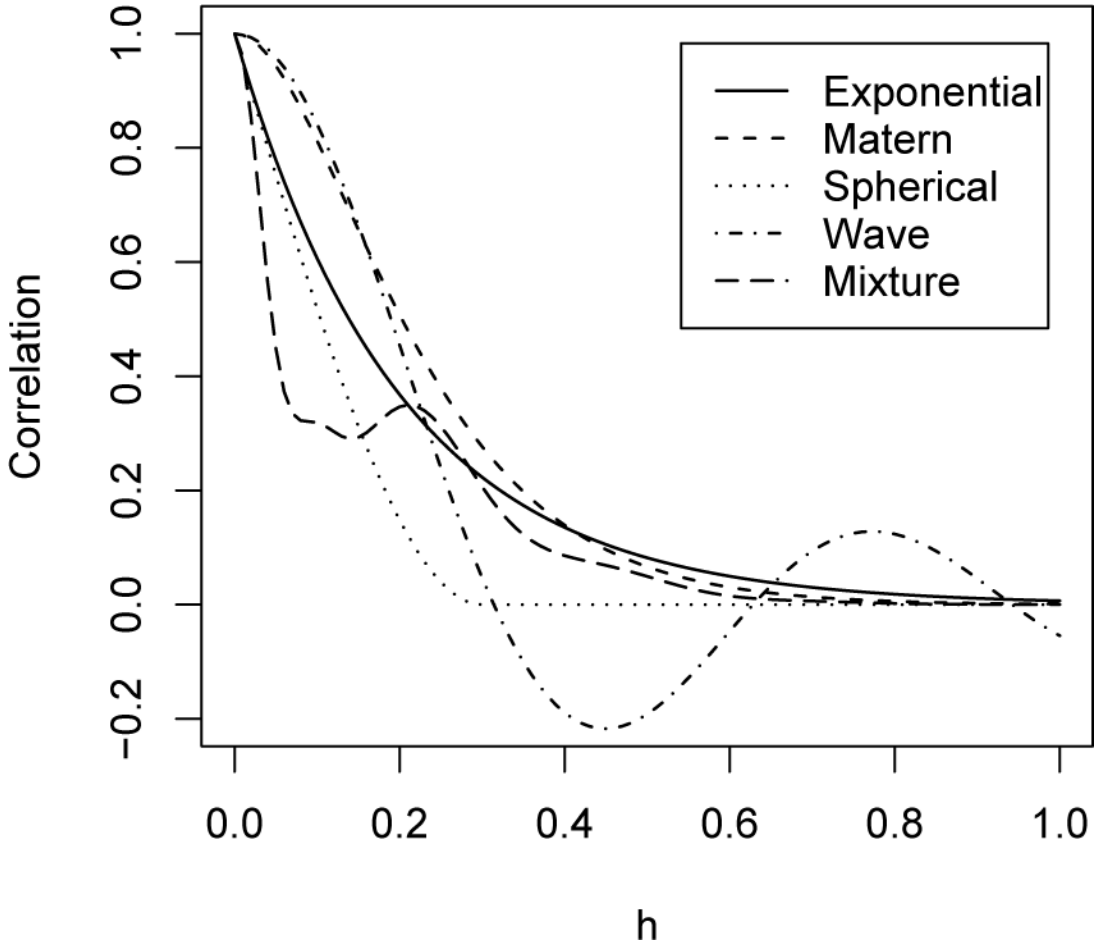

Figure 2.
Correlation functions used in the simulation study.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

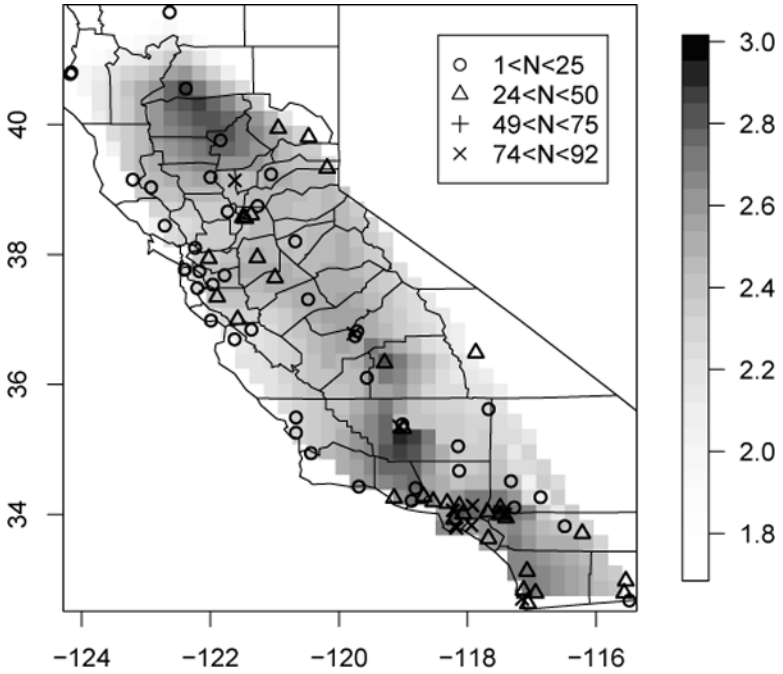

Figure 3.

Fine particulate matter in California, June–August, 2008. The shading indicates the average of the log(concentration+1), and the symbols indicate the number of observations.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

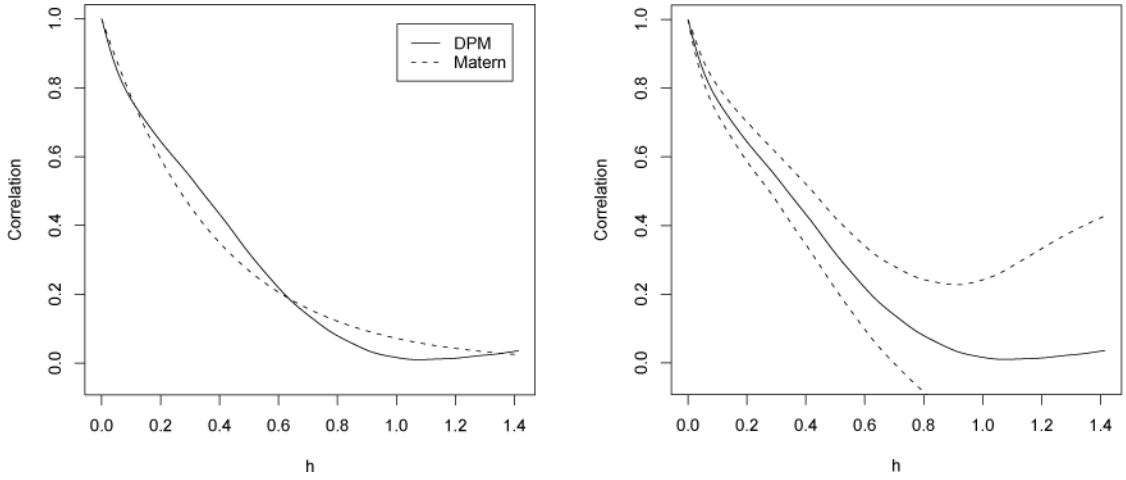

Figure 4.

Spatial correlation estimates for the California $PM_{2.5}$ data. The left panel gives the posterior mean correlation for the Matérn and Dirichet process mixture (DPM) model, and right panel plots the posterior mean plus/minus two times the standard deviation for the DPM model. The x-axis is projected distance using the Mercator projection to the unit square.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Simulation study results, reported as the mean (standard error) across data sets. All values are multiplied by 100 except relative mean squared error ("Rel MSE")

(a) Results for the training sites

	MSE			Coverage	
True covariance	**Matérn**	**DPM**	**Rel MSE**	**Matérn**	**DPM**
Exponential	1.16 (0.03)	1.20 (0.02)	1.05 (0.03)	84.3 (0.23)	83.8 (0.24)
Matérn	0.86 (0.01)	0.85 (0.01)	0.99 (0.01)	91.3 (0.12)	90.0 (0.13)
Spherical	1.20 (0.02)	1.15 (0.02)	0.96 (0.02)	84.0 (0.18)	84.5 (0.26)
Wave	0.63 (0.01)	0.49 (0.01)	0.77 (0.01)	90.9 (0.08)	90.0 (0.10)
Mixture	34.4 (0.54)	4.04 (1.68)	0.11 (0.04)	93.2 (0.09)	83.4 (0.59)

(b) Results for the testing sites

	MSE			Coverage	
True covariance	**Matérn**	**DPM**	**Rel MSE**	**Matérn**	**DPM**
Exponential	43.8 (0.24)	43.8 (0.23)	1.00 (0.01)	90.6 (0.09)	90.6 (0.08)
Matérn	14.7 (0.20)	14.3 (0.13)	0.97 (0.01)	90.1 (0.10)	90.4 (0.12)
Spherical	46.1 (0.34)	45.9 (0.39)	0.99 (0.01)	90.6 (0.10)	90.7 (0.09)
Wave	4.08 (0.11)	1.94 (0.07)	0.48 (0.01)	88.1 (0.13)	89.0 (0.12)
Mixture	71.0 (0.33)	65.2 (0.41)	0.92 (0.01)	94.4 (0.09)	90.9 (0.21)