A new class of nonstationary spatial models

Montserrat Fuentes[1] and Richard L. Smith[2]

ABSTRACT

Spatial processes are an important modeling tool for many problems of environmental monitoring. Classical geostatistics is based on processes which are stationary and isotropic, but it is widely recognized that real environmental processes are rarely stationary and isotropic. In this paper, a new class of nonstationary processes is proposed, based on a convolution of local stationary processes. This model has the advantage that the model is simultaneously defined everywhere, unlike "moving window" approaches, but it retains the attractive property that locally in small regions, it behaves like a stationary spatial processes. We discuss model fitting through exact and approximate likelihood maximization, and propose a hierarchical Bayes approach to allow predictive inference when the parameters of the model are unknown. Applications include obtaining the total loading of sulfur dioxide concentrations over different geo-political boundaries.

[1]Montserrat Fuentes is an assistant professor at the Statistics Department, North Carolina State University (NCSU), Raleigh, NC 27695-8203, and a visiting scientist at the US Environmental Protection Agency (EPA). Tel.:(919) 515-1921, Fax: (919) 515-1169, E-mail: fuentes@stat.ncsu.edu. This research was sponsored by a National Science Foundation grant DMS 0002790, by a US EPA award R-8287801 and a contract with the US EPA Acid Rain Division.

[2]Richard L. Smith is a professor at the Statistics Department, University of North Carolina, Chapel Hill, NC. Tel: (919) 962-2660, Fax: (919) 962-1279, rls@email.unc.edu. Supported in part by NSF grants DMS-997190 and DMS-0084375 and EPA Cooperative Agreement CR-827737-01-0.

Key words: air pollution, Bayesian statistics, change of support, kriging, Matérn covariance, nonstationary process, spatial statistics.

1 Introduction

A major focus of the Clean Air Act Amendments of 1990 has been the effect of atmospherically-transported pollutants on terrestrial and aquatic ecosystems. The 1990 amendments include new requirements that will appreciably reduce sulfur dioxide (SO_2) emissions. Monitoring data for SO_2 are analyzed as part of the process for assessing compliance with the Clean Air Act Amendments of 1990. The ground-level concentration of SO_2 will depend on the proximity to the source, the prevailing meteorology, and the nature and extent of atmospheric chemical reactions between the source and receptor. The interaction of these chemical and physical atmospheric processes and the source locations tend to produce data patterns that show large spatial variability. Because spatial patterns of SO_2 fluxes and concentrations are nonstationary processes, in the sense that the spatial structure changes with location, standard methods of spatial interpolation are inadequate. We present a new statistical methodology for prediction of nonstationary processes, with the objective of obtaining in an efficient way the total loading of SO_2 concentrations and fluxes over different geo-political boundaries.

In recent years, probably the most extensively studied method for nonstationary spatial processes is the deformation approach due to Sampson and Guttorp (1992); see also Guttorp and Sampson (1994), Guttorp, Meiring and Sampson (1994). Maximum likelihood versions of the method were developed by Mardia and Goodall (1993) and Smith (1996). In a series of papers best represented by Haas (1995, 1998), T. Haas has proposed an approach to nonstationary spatial kriging based on moving windows. Higdon, Swall and Kern (1999) give a model for accounting for heterogeneity in the spatial covariance function of a spatial process, using a moving average specification of a Gaussian process. Another approach has been developed by Nychka and Saltzman (1998) and Holland *et al.* (1999), that extends the "empirical orthogonal functions" (EOF) approach that is popular among atmospheric scientists.

In this paper we give a new methodology for spatial interpolation of nonstationary processes. More specifically, we represent the process locally as a stationary isotropic random field with some parameters that describe the local spatial structure. These parameters are allowed to vary across space and reflect the lack of stationarity of the process. We present a nonstationary process Z observed on a region D as a convolution of local stationary processes:

$$Z(\mathbf{x}) = \int_D K(\mathbf{x} - \mathbf{s}) Z_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x}) d\mathbf{s}. \tag{1}$$

where K is a kernel function and $Z_{\boldsymbol{\theta}}(\mathbf{x}),\ \mathbf{x} \in D$ is a family of (independent) stationary Gaussian processes indexed by $\boldsymbol{\theta}$. If K is a sharply peaked kernel function and $\boldsymbol{\theta}(\mathbf{s})$ varies slowly with $\mathbf{s}$, this has the property that for $\mathbf{x}$ near $\mathbf{s}$, the process "looks like" a stationary process with parameter $\boldsymbol{\theta}(\mathbf{s})$. On the other hand, since $\boldsymbol{\theta}(\mathbf{s})$ may vary substantially over the whole space, it also allows significant nonstationarity.

This paper is organized as follows. Section 2 describes the current kernel approaches to nonstationary spatial processes. In section 3, we propose a new model for nonstationarity and we introduce some fitting algorithms to estimate the spatial structure. Section 4 discusses prediction from a Bayesian point of view, to take into consideration in the prediction the uncertainty in the covariance parameters. In Section 5, we discuss the change-of-support problem that occurs when the modeled data and the required predictions are defined on different spatial scales. Section 6 is an application of the methodology presented in this paper to the SO_2 data. Finally, in Section 7 we present some conclusions and final remarks.

2 Models defined by kernel smoothing

A broad class of stationary Gaussian processes may be represented in the form

$$Z(\mathbf{s}) = \int K(\mathbf{s} - \mathbf{u}) X(\mathbf{u}) d\mathbf{u},$$

with $K(\cdot)$ some kernel function and $X(\cdot)$ a constant-variance Gaussian white noise process. The motivation for defining a spatial process as an integral of white noise can be said to go back to Whittle (1954), who gave a similar representation for discrete spatial processes. Matérn (1986) used this representation to derive a wide class of stationary spatial processes. Higdon, Swall and Kern (1999) considered extensions of the form

$$Z(\mathbf{s}) = \int K_{\mathbf{s}}(\mathbf{u})X(\mathbf{u})d\mathbf{u}, \tag{2}$$

where the kernel $K_{\mathbf{s}}$ depends on position $\mathbf{s}$. The idea of Higdon *et al.* was to model $K_{\mathbf{s}}(\mathbf{u})$ as an unknown function in terms of specific parameters which can then be estimated in a hierarchical Bayes framework. In the case where $K_{\mathbf{s}}$ is a Gaussian kernel for each $\mathbf{s}$, this leads to tractable expressions for the covariance function and hence the likelihood function for the process.

This approach is promising, and a quite different idea from earlier approaches for nonstationary processes, but it has the disadvantage of not being easily related to traditional spatial models. The development of Higdon *et al.* relies heavily on the Gaussian form of kernel function and it is not clear how restrictive this is. Our own approach also uses kernel representations, but has a quite different motivation.

3 A new model for nonstationary spatial processes

Fuentes, (2001a) and (2001b) presents some new spectral approaches to study the spatial structure of a nonstationary process, using a Fourier-Stieltjes representation of the process and assuming the process behaves locally as a stationary isotropic random field. The model used in Fuentes (2001a) and (2001b) represents a nonstationary process Z observed on a region D as a weighted average of orthogonal local stationary processes, Z_i for $i = 1, \ldots, k$, with $\mathrm{cov}(Z_i(\mathbf{x}), Z_i(\mathbf{y})) = 0$ for $i \neq j$,

$$Z(\mathbf{x}) = \sum_{i=1}^{k} Z_i(\mathbf{x})w_i(\mathbf{x}) \tag{3}$$

where $S_1, \ldots, S_k$ are well-defined subregions that cover D, and Z_i is a local stationary process in the subregion S_i, $w_i(\mathbf{x})$ is a positive kernel function centered at the centroid of S_i.

In this article, we propose to write model (3) as an integral to obtain a continuous representation of Z. Thus, we represent a Gaussian spatial process Z as a convolution of stationary processes:

$$Z(\mathbf{x}) = \int_D K(\mathbf{x} - \mathbf{s}) Z_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x}) d\mathbf{s}. \tag{4}$$

where K is a kernel function and $Z_{\boldsymbol{\theta}}(\mathbf{x}),\ \mathbf{x} \in D$ is a family of (independent) stationary Gaussian processes indexed by $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is the parameter associated with the covariance function. The stochastic integral (4) is defined as a limit (in mean square) of approximating sums (e.g., Cressie, 1993, p. 107, Yaglom, 1962, p. 23). The parameter function $\boldsymbol{\theta}$ is allowed to vary across space to reflect the lack of stationarity of the process.

The processes $Z_{\boldsymbol{\theta}(\mathbf{s})}$ are stationary with autocovariances $C_{\boldsymbol{\theta}(\mathbf{s})}$, then they can be represented in the form:

$$Z_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x}) = \int_{\mathbb{R}^2} K_{\mathbf{s}}(\mathbf{u} - \mathbf{x}) X_{\mathbf{s}}(\mathbf{u}) \tag{5}$$

where $K_{\mathbf{s}}$ is a kernel for each $\mathbf{s}$, and $X_{\mathbf{s}}$ are independent white noise process for each $\mathbf{s}$. A direct attempt to rewrite (4) in the form (2) does not work because the process $X_{\mathbf{s}}$ is an independent white noise process for each $\mathbf{s}$, and we conjecture that there is no way of reducing the present model to one of the structure of (2), with a common underlying white noise process X.

There is a possible alternative interpretation of (5), in which the process $X_{\mathbf{s}}$, instead of being an independent white noise process for each $\mathbf{s}$, is a common process for all $\mathbf{s}$, so the processes $Z_{\boldsymbol{\theta}(\mathbf{s})}$ are correlated. In that case it is straightforward to rewrite (4) in the form of (2), though the kernel $K_{\mathbf{s}}(\mathbf{u})$ would be difficult to evaluate in practice. Because of its additional computational complexity, we have not pursued that approach in the present paper.

In the proposed nonstationary model (4), if K is a sharply peaked kernel function and $\boldsymbol{\theta}(\mathbf{s})$

varies slowly with $\mathbf{s}$, this has the property that for $\mathbf{x}$ near $\mathbf{s}$, the process "looks like" a stationary process with parameter $\boldsymbol{\theta}(\mathbf{s})$. On the other hand, since $\boldsymbol{\theta}(\mathbf{s})$ may vary substantially over the whole space, it also allows significant nonstationarity. The method has features in common with Haas's approach, but in view of the representation (4), there is no problem about it being a well-defined process with a positive definite covariance function (this is also an attraction of (2), of course). We could even use a variant of Haas's approach to estimate the model, for instance, by estimating $\boldsymbol{\theta}(\mathbf{s})$ for a finite set of values of $\mathbf{s}$ assuming stationarity within some window, and then smoothing the function $\boldsymbol{\theta}(\mathbf{s})$ by kernels or splines. In the discussion to follow, however, we shall outline a number of more sophisticated estimation procedures. Another attraction of model (4) is that it allows all model parameters $\boldsymbol{\theta}$ to vary over D, unlike the two-dimensional version of the Guttorp and Sampson (1994) model.

We treat Z as a zero-mean process, and we have a separate smooth surface for the mean function. The covariance of $Z_{\boldsymbol{\theta}(\mathbf{s})}$ is stationary with parameter $\boldsymbol{\theta}(\mathbf{s})$,

$$\mathrm{cov}\{Z_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x_1}), Z_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x_2})\} = C_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x_1} - \mathbf{x_2}).$$

The covariance $C(\mathbf{x_1}, \mathbf{x_2}; \theta)$ of Z is a convolution of the local covariances $C_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x_1} - \mathbf{x_2})$,

$$C(\mathbf{x_1}, \mathbf{x_2}; \boldsymbol{\theta}) = \int_D K(\mathbf{x_1} - \mathbf{s}) K(\mathbf{x_2} - \mathbf{s}) C_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x_1} - \mathbf{x_2}) d\mathbf{s}. \tag{6}$$

We assume that $\boldsymbol{\theta}(\mathbf{s})$ is a continuous function of $\mathbf{s}$. As an example, the process $Z_{\boldsymbol{\theta}(\mathbf{s})}$ could have a Matérn stationary covariance:

$$C_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x}) = \frac{\sigma_s}{2^{\nu_s - 1}\Gamma(\nu_s)} (2\nu_s^{1/2}|\mathbf{x}|/\rho_s)^{\nu_s} \mathcal{K}_{\nu_s}(2\nu_s^{1/2}|\mathbf{x}|/\rho_s), \tag{7}$$

where $\mathcal{K}_{\nu_s}$ is a modified Bessel function and we have $\boldsymbol{\theta}(\mathbf{s}) = (\nu_s, \sigma_s, \rho_s)$. The parameter ρ_s measures how the correlation decays with distance, generally this parameter is called the *range*; σ_s is the variance of the random field, i.e. $\sigma_s = \mathrm{var}(Z_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x}))$, the parameter σ_s is usually refereed to as

the *sill*; and the parameter ν_s measures the degree of smoothness of the process $Z_{\boldsymbol{\theta}(\mathbf{s})}$, the higher the value of ν_s the smoother $Z_{\boldsymbol{\theta}(\mathbf{s})}$ would be, e.g. when $\nu_s = \frac{1}{2}$, we get the exponential covariance function. If we consider the limit as $\nu_s \to \infty$ we get the Gaussian covariance

$$C_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x}) = \sigma_s e^{-|\mathbf{x}|^2/\rho_s^2}.$$

In (6) every entry requires an integration. Since each such integration is actually an expectation with respect to a uniform distribution, we propose to evaluate (6) by Monte Carlo integration. We draw an independent set of locations $\mathbf{s}_m$, $m = 1, 2, ..., M$ over D, $C(\mathbf{x_1}, \mathbf{x_2}; \boldsymbol{\theta})$ by

$$\hat{C}(\mathbf{x_1}, \mathbf{x_2}; \boldsymbol{\theta}) = M^{-1} \sum_{m=1}^{M} K(\mathbf{x}_1 - \mathbf{s}_m) K(\mathbf{x}_2 - \mathbf{s}_m) C_{\boldsymbol{\theta}(\mathbf{s}_m)}(\mathbf{x_1} - \mathbf{x_2}) \tag{8}$$

In this notation, the "hat" denotes a Monte Carlo integration which can be made arbitrarily accurate and has nothing to do with the data Z. The size of the sample, M, could be selected using the same kinds of criteria as are used for assessing convergence of simulation procedures in Bayesian inference.

3.1 Choice of weight function and bandwidth

In most settings where kernel methods are employed, e.g. density estimation or nonparametric regression, it is critical to choose the bandwidth parameter in a way that achieves reasonable balance between bias and variance. In particular, too small bandwidth would typically create an estimator that is too rough (too large variance). In contrast, analytic form of the kernel is generally considered to be of much less importance.

In the present setting, the smoothness of the spatial process is largely dependent on how smoothly $\boldsymbol{\theta}(\mathbf{u})$ varies as a function of $\mathbf{u}$. The bandwidth of the kernel is more important. However, the proposed computational method requires that, for each point of prediction $\mathbf{x}$, at least one of the sampling locations $\mathbf{s}_m$ $(1 \leq m \leq M)$ satisfies $K(\mathbf{x} - \mathbf{s}_m) > 0$. This defines a lower bound on

the permissible bandwidth.

Therefore, in practice we use a kernel of the form $K(\mathbf{u}) = h^{-2}K_0(\mathbf{u}/\mathbf{h})$, where for K_0 we have used the two-dimensional version of the Epanechnikov (1969) kernel,

$$K_0(\mathbf{u}) = \frac{2}{\pi}(1 - \|\mathbf{u}\|^2)_+. \tag{9}$$

and the bandwidth h is chosen near the minimum value for the previously mentioned condition to be satisfied. In the case that $\mathbf{s}_1, \ldots, \mathbf{s}_M$ lie on a regular grid with distance l between neighboring grid points, it suffices to take $h = l/\sqrt{2}$. This is the value used later in Section 7.

Our convolution model could be also considered a generalization to spatial problems of the delta sequences used for probability density estimation (e.g.. Walter and Blum (1979)). With delta sequences, small values of h are ideal to preserve the "shape".

4 Hierarchical Bayesian model

The parameter function $\boldsymbol{\theta}(\mathbf{s})$ for $\mathbf{s} \in \mathbf{D}$, measures the lack of stationarity of the process Z. It would be natural to treat $\boldsymbol{\theta}(\mathbf{s})$ as a stochastic process, with correlated errors. Alternatively, we could use a spline approach to model $\boldsymbol{\theta}(\mathbf{s})$, though the approach presented here allows splines as a special case. We consider a hierarchical Bayesian approach to model and take into the account the spatial structure of the parameter $\boldsymbol{\theta}(\mathbf{s})$ in the prediction of Z. The essence of the hierarchy that we suggest is the specification of Z as a parameterized process model:

[process, parameters]=[process | parameters] [parameters].

Bayes' Theorem provides [process, parameters | data] by combining with an observational data model [data | process, parameters].

<u>Stage 1:</u>

The process Z is written as a convolution of local stationary processes:

$$Z(\mathbf{x}) = \int_D K(\mathbf{x} - \mathbf{s}) Z_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{x}) d\mathbf{s}. \tag{10}$$

where K is a kernel function, and $Z_{\boldsymbol{\theta}}(\mathbf{x}),\ \mathbf{x} \in D$ is a family of (independent) stationary Gaussian processes indexed by $\boldsymbol{\theta}$. Thus, the distribution of Z given $\boldsymbol{\theta}$ is Gaussian.

Stage 2:

We present an approach to model $\boldsymbol{\theta}$. The parameter function $\boldsymbol{\theta}$ is modeled using a Bayesian version of the median polish approach (e.g. Cressie, 1993). Assume the process Z is observed in a grid $n_1 \times n_2$, then we propose the following model for the parameter function $\boldsymbol{\theta}$

$$\boldsymbol{\theta}(\mathbf{s}) = a + r_i + c_j + \epsilon_{\boldsymbol{\theta}}(\mathbf{s}) \tag{11}$$

where $\mathbf{s} = (s_i, s_j)$, for $i = 1, \ldots, n_1$ and $j = 1, \ldots, n_2$. The process $\epsilon_{\boldsymbol{\theta}}(\mathbf{s})$ represents some spatially correlated zero-mean noise. We assume $\epsilon_{\boldsymbol{\theta}}(\mathbf{s})$ is Gaussian with mean zero and a Matérn stationary covariance, $\mathrm{cov}(\epsilon_{\boldsymbol{\theta}}(\mathbf{x}+\mathbf{y}), \epsilon_{\boldsymbol{\theta}}(\mathbf{x})) = C_{\boldsymbol{\tau}_0}(\mathbf{y})$, with parameters $\boldsymbol{\tau}_0$. We use $\boldsymbol{\beta}_0$ to denote the hyper-parameters a, r_i for $i = 1, \ldots, n_1$, and c_j for $j = 1, \ldots, n_2$,. The vector parameter $\boldsymbol{\beta}_0$ is unknown. If σ does not change very rapidly with location, then only few values of r and c (i.e. small n_1 and n_2) should be enough to characterize the large scale structure of σ.

The kernel function K in the representation of Z is a kernel smoother with bandwidth h. The function K is not representing the lack of stationarity. The crucial parameter here to explain the lack of stationarity of Z is $\boldsymbol{\theta}$. The bandwidth parameter h is chosen using the algorithm discussed in Section 3.1. Another alternative that has not been used in this paper, it is to include the bandwidth parameter h in this hierarchy. Thus, in stage 1, we would have $[Z|\boldsymbol{\theta}, h]$, and the Bayesian model would be completely specified once we give a prior distribution to h and $\boldsymbol{\theta}$.

If the goal is to predict Z at a location $\mathbf{x}_0$, given observations $\mathbf{Z} = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_N))$, then the

Bayesian solution is the predictive distribution of $Z(\mathbf{x}_0)$ given $\mathbf{Z}$,

$$p(Z(\mathbf{x}_0)|\mathbf{Z}) \propto \int p(Z(\mathbf{x}_0)|\mathbf{Z}, \boldsymbol{\theta}(\mathbf{s}) \text{ for } \mathbf{s} \in D, h, \sigma^2)\, p(\boldsymbol{\theta}(\mathbf{s}) \text{ for } \mathbf{s} \in D, \sigma^2, h|\mathbf{Z})\, dh\, d\sigma^2\, d\boldsymbol{\theta}, \tag{12}$$

and a MCMC approach was used to simulate m values from the posterior of the parameters $\boldsymbol{\theta}$, h, and σ^2, and the predictive distribution was approximated by the *Rao-Blackwellized estimator*:

$$p(Z(\mathbf{x}_0)|\mathbf{Z}) = \frac{1}{m}\sum_{i=1}^{m} p(Z(\mathbf{x}_0)|\mathbf{Z}, \boldsymbol{\theta}(\mathbf{s})^{(i)} \text{ for } \mathbf{s} \in D, h^{(i)}, \sigma^{2(i)})$$

where $\boldsymbol{\theta}(\mathbf{s})^{(i)}$ for $\mathbf{s} \in D, h^{(i)}, \sigma^{2(i)}$ constitute the i-th draw from the posterior distribution. Notice that to sample from the posterior of $\boldsymbol{\theta}$ we need to specify first the priors of the hyperparameters $\boldsymbol{\beta}_0$ and $\boldsymbol{\tau}_0$:

$$p(\boldsymbol{\theta}(\mathbf{s}) \text{ for } \mathbf{s} \in D|\mathbf{Z}) \propto \int p(\mathbf{Z}|\boldsymbol{\theta}(\mathbf{s}) \text{ for } \mathbf{s} \in D) p(\boldsymbol{\theta}(\mathbf{s}) \text{ for } \mathbf{s} \in D|\boldsymbol{\beta}_0, \boldsymbol{\tau}_0) p(\boldsymbol{\beta}_0, \boldsymbol{\tau}_0) d\boldsymbol{\tau}_0\, d\boldsymbol{\beta}_0.$$

Thus, once we determine the priors for the parameters h and σ, and for the hyperparameters $\boldsymbol{\beta}_0$ and $\boldsymbol{\tau}_0$, the Bayesian model is completely specified. The joint posterior for $\boldsymbol{\theta}, h$, and σ^2 is defined as follows:

$$p(\boldsymbol{\theta}(\mathbf{s}) \text{ for } \mathbf{s} \in D, h, \sigma^2|\mathbf{Z}) \propto p(\mathbf{Z}|\boldsymbol{\theta}(\mathbf{s}) \text{ for } \mathbf{s} \in D, h, \sigma^2) p(h) p(\sigma^2) \int p(\boldsymbol{\theta}(\mathbf{s}) \text{ for } \mathbf{s} \in D|\boldsymbol{\beta}_0, \boldsymbol{\tau}_0) p(\boldsymbol{\beta}_0) p(\boldsymbol{\tau}_0) d\boldsymbol{\tau}_0\, d\boldsymbol{\beta}_0,$$

where $p(h), p(\sigma^2), p(\boldsymbol{\beta}_0)$, and $p(\boldsymbol{\tau}_0)$ are the prior distributions for h, σ, $\boldsymbol{\beta}_0$ and $\boldsymbol{\tau}_0$.

5 Change of Support

The change-of-support problem occurs when the supports of the predictand and the data are not the same. In the present instance we are interested in making predictions about the random process at a point location $Z(\mathbf{x}_0)$ from data on block averages, $Z(B_1), \ldots, Z(B_N)$.

Here, we observe Z, the output of a physical model, averaged over regions, $B_1, \ldots, B_N$ of dimensions 36km $\times$ 36km and we want to predict Z at a location of interest $\mathbf{x}_0$, e.g. at the location where we have a monitoring site.

The covariance for the block averages is defined as follows

$$\mathrm{cov}(Z(B_i), Z(B_j)) = \int_{B_i}\int_{B_j} C(\mathbf{u}, \mathbf{v}) d\mathbf{u} d\mathbf{v} / |B_i||B_j|, \tag{13}$$

where

$$C(\mathbf{u}, \mathbf{v}) = \mathrm{cov}(Z(\mathbf{u}), Z(\mathbf{v}))$$

C is a nonstationary spatial function. In practice, for each pixel B_i we draw an independent set of locations $\mathbf{u}_{ir}$, $r = 1, 2, \ldots, L_i$ uniformly over B_i, and we approximate the integral in (13) with the following expression

$$\widehat{\mathrm{cov}}(Z(B_i), Z(B_j)) = L_i^{-1} L_j^{-1} \sum_r \sum_{r'} C(\mathbf{u}_{ir}, \mathbf{u}_{jr'}). \tag{14}$$

Gelfand *et al* (2000) have independently studied this problem using approximation (14) to the block covariances. We approximate the point prediction $E(Z(\mathbf{x}_0)|\mathbf{Z})$, where $\mathbf{Z} = (Z(B_1), \ldots, Z(B_N))$ with $\hat{E}(Z(\mathbf{x}_0)|\mathbf{Z})$, which is obtained using (14) as an approximation of (13). It is useful to note that, if we define $\hat{Z}(B_i) = L_i^{-1} \sum_r Z(u_{ir})$ for $i = 1, \ldots, N$, then the solution $\hat{E}(Z(\mathbf{s}_0)|\mathbf{Z})$ is actually $E(Z(\mathbf{s}_0)|\hat{\mathbf{Z}})$. In the Bayesian approach if we use (14) as an approximation of (13), the Bayesian predictive distribution we obtain to sample $Z(\mathbf{x}_0)$ is actually $\hat{f}(Z(\mathbf{x}_0)|\mathbf{Z})$. Note that $\hat{f}(Z(\mathbf{x}_0)|\mathbf{Z}) = f(Z(\mathbf{x}_0)|\hat{\mathbf{Z}})$. Hence, we need $\hat{Z}(B_i) \xrightarrow{P} Z(B_i)$. If $\mathrm{cov}(Z(\mathbf{u}), Z(\mathbf{v}))$ is a continuous function on $\mathbf{u}$ and $\mathbf{v}$, then Z is *mean square continuous* (see e.g. Stein, 1999, p. 20), which implies $\hat{Z}(B_i) \xrightarrow{P} Z(B_i)$. By the definition of the nonstationary covariance $\mathrm{cov}(Z(\mathbf{u}), Z(\mathbf{v}))$ in (6), if the local stationary covariances are Matérn, then $\mathrm{cov}(Z(\mathbf{u}), Z(\mathbf{v}))$ is continuous on $\mathbf{u}$ and $\mathbf{v}$.

In this paper, we define the process Z in terms of a pointwise covariance $C(\mathbf{u}, \mathbf{v})$, but then we use (13), approximated by (14), to derive the covariances of the block averages $Z(B_i)$, $i = 1, \ldots, N$, in terms of the pointwise covariance $C(\mathbf{u}, \mathbf{v})$. This is then used to define a likelihood function for the parameters of the covariance function for the process Z in terms of the observed block averages $Z(B_1), Z(B_2), \ldots, Z(B_N)$.

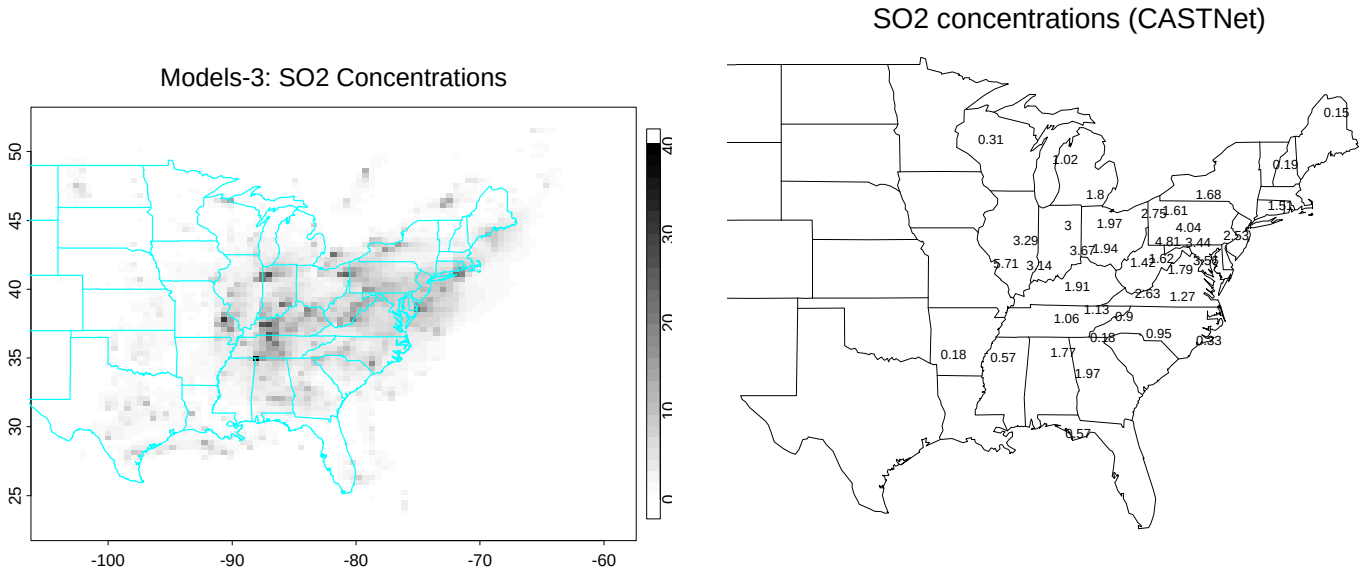
6 Application

We have two sources of data for dry deposition fluxes and concentrations of SO_2:

1. The first sources of information are the regional scale air quality models. These models, e.g. Models-3, are run by EPA and the U.S. States and provide SO_2 areal concentrations and fluxes in regular grids in parts of the US (see Figure 1 (a)). The current resolution of Models-3 is 36 km $\times$ 36 km. The primary objective of Models-3 is to improve the environmental management community's ability to evaluate the impact of air quality management practices for multiple pollutants at multiple scales, as part of the regulation process of the air pollutants standards.

2. EPA provides point measurements at 50 irregularly spaced sites in the eastern U.S. known as the Clean Air Status and Trends network (CASTNet) (see Figure 1 (b)). At each site, EPA measures dry deposition fluxes and concentrations of different atmospheric pollutants.

Models-3 is used to examine the response of the air pollution network to different control strategies under various high-pollution scenarios. To establish its credibility, however, it is essential that it should accurately reproduce observed measurements when applied to ground data. Models-3 uses as inputs metereological data, emissions data and boundary values of air pollution. The available emissions data are combined with numerical models of local weather (the Mesoscale Model version 5 (MM5)), the emissions process (the Sparse Matrix Operator Kernel Emissions (SMOKE)), as well as information about land use and cover, to estimate pollution levels in space and time (the Community Multiscale Air Quality (CMAQ) output) and produce maps (Dennis *et al*, 1996). Models-3. are not statistical models but numerical deterministic simulation models based on systems of differential equations that attempt to represent the underlying physics, and take the form of huge blocks of computer code.



(a) (b)

Figure 1: (a): SO_2 concentrations (ppb) from Models-3 for the week of July 11, 1995. The resolution is $36km \times 36km$. (b): SO_2 concentrations (ppb) at the CASTNet sites for the same week.

One of the main objectives of the work presented here is to evaluate Models-3, and therefore we need measures of how well Models-3 output and real data agree. We present a methodology that uses Models-3 output to come up with probabilistic predictions of SO_2 concentrations at the point measurement sites. In our approach we take into account the change of support problem between Models-3 (areal measurements) and CASTNet (point measurements). This application requires, as a first step, fitting some random field model to Models-3, but it is unrealistic to assume that a stationary isotropic model applies over such a large geographical area. Therefore, we need nonstationary spatial models.

The first step of our analysis is to understand and to quantify the spatial structure of air pollutants using the output of the regional scale air quality models (Models-3). Models-3 estimates hourly concentrations and fluxes of different pollutants. As an example we examine sulfur dioxide. The spatial domain, D, is a regular grid (81×87), and the dimensions of each pixel on the grid are 36km × 36km. Models-3 provides the estimated average concentration for each pixel. Figure

1 (a) shows the weekly average concentrations of SO_2 for the week starting July 11, 1995. We implemented here the nonstationary model (4) with a kernel function with compact support, the kernel used being the quadratic weight-function (9), presented in Section 3.1. We applied the algorithm discussed in Section 3.1 to choose the bandwidth, trying to preserve the general "shape" of the data. The value of h in this application is 229 km ($h = l/\sqrt{2}$), because $M = 81 = 9 \times 9$ (sample size) and l (distance between sampling points) is 324 km ($9 \cdot 36$km). Note that the value of h changes with M. We obtained the value of M using Bayesian criteria for convergence of the posterior distributions in the Bayesian hierarchical model. For the purpose of illustrating the need of the technique presented in this paper, Figures 2 (b) and 3 (a) show the posterior distributions of covariance parameters at the selected sites plotted in Figure 2 (a). We used vague gamma priors for all the covariance parameters, except for the sill parameter that we used

$$p(\sigma) \propto \sigma^{-1},$$

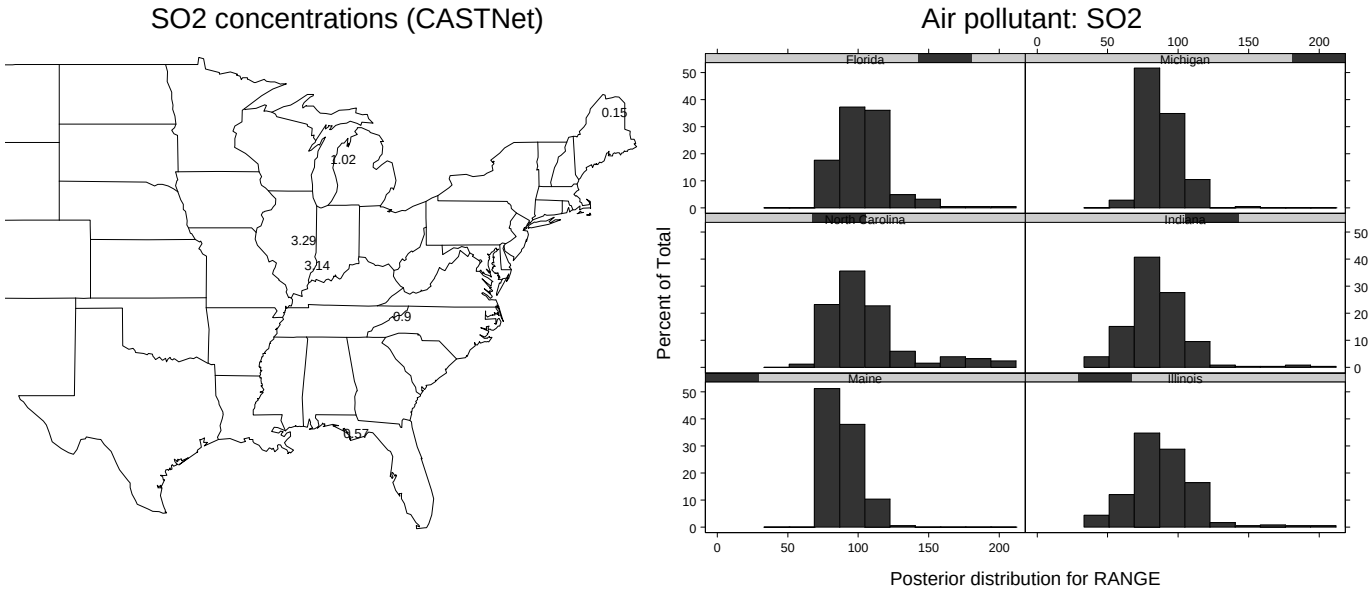
which is a uniform prior for $\log(\sigma)$. The sill parameter is changing with location as illustrated by the variation in the distributions in Figure 3 (a). This indicates lack of stationarity. We can clearly appreciate this fact in Figure 3 (b). Figure 3 (b) shows a map of the modes for the posterior distributions of the sill parameter of the Matérn covariance for the Models-3 SO_2 concentrations, for the week starting July 11, 1995. Figure 3 (b) indicates clearly that there is a deviation from stationarity. We modeled the covariance parameters ($\boldsymbol{\theta}$) using the model proposed in (11). Thus, for the sill parameter (σ) we have

$$\sigma(\mathbf{s}) = a + r_i + c_j + \epsilon_{\boldsymbol{\sigma}}(\mathbf{s})$$

where $\mathbf{s} = (s_i, s_j)$, for $i = 1, \ldots, n_1$ and $j = 1, \ldots, n_2$, the r_i's values explain the longitude effect and the c_j's values explain the latitude effect. The process $\epsilon_{\boldsymbol{\sigma}}(\mathbf{s})$ is Gaussian with mean zero and a Matérn stationary covariance, $\mathrm{cov}(\epsilon_{\boldsymbol{\sigma}}(\mathbf{x}+\mathbf{y}), \epsilon_{\boldsymbol{\sigma}}(\mathbf{x})) = C_{\boldsymbol{\tau}_0}(\mathbf{y})$, with parameters $\boldsymbol{\tau}_0$. We have

$n_1 = 9$ and $n_2 = 9$ (the number of sampling points is $M = 81$). The mode of the hyperparameters r_i for $i = 1, \ldots, n_1$, and the hyperparameters c_j for $j = 1, \ldots, n_2$, are shown in Figure 4 (a). In fact, Figure 4 (a) shows a smoothed version using a cubic spline of the longitude effect (r_i's) and the latitude effect (c_j's), and also the semivariogram of the error term ($\epsilon_{\boldsymbol{\sigma}}$) for the sill parameter, σ. The longitude effect shows higher sill values in a region between 80°W and 90°W, and the latitude effect suggests an increase of variability (sill) from 35°N to 40°N, followed by an almost linear decrease at higher latitudes. This corresponds to what we observed in Figure 3 (b), where we obtained higher values for the sill parameter in the Midwest. In Figure 4 (a) (bottom graph) we assumed a Matérn model for the semivariogram of the error term ($\epsilon_{\boldsymbol{\sigma}}$), the covariance parameters for this Matérn model are the modes of the posterior for the hyperparameter $\boldsymbol{\tau}_0$. The posterior mode of the smoothing parameter for this Matérn covariance estimates the spatial local behavior of the sill, in this case the mode of the smoothing parameter is only .09 which suggest that the sill parameter is not a very smooth process. The range and the smoothing parameters do not change much with location, the mode of the range is approximately 75 km almost everywhere (see Figure 2 (b)), and the mode of the smoothness parameter is approximately .5. The $S0_2$ varies rapidly with location, so it is not surprising to obtain these short ranges of autocorrelation. Regarding the smoothness of the process, we should keep on mind that we are working with preprocessed data, the output of a numerical model. Therefore, the smoothness parameter was not expected to vary much with location, because the data have been already smoothed and processed.

The evaluation of the regional air quality models is crucial. We use the ground measurements from the Clean Air Status and Trends Network (CASTNet) (see Figure 1 (b)), to evaluate weakly average concentrations of SO_2 estimated by Models-3. The coordinates for Models-3 do not match the location of the CASTNet sites. Thus, we spatially interpolate the output of the models at the location of the CASTNet sites. We should do the interpolation by taking into consideration the

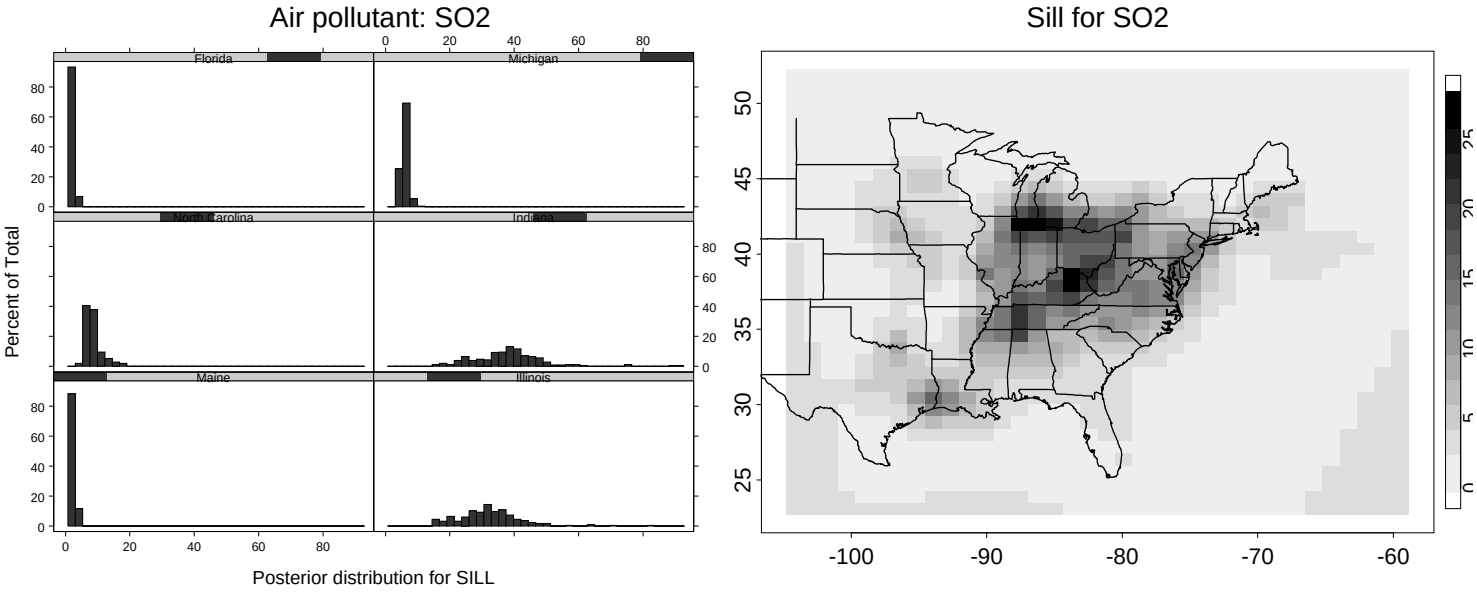


(a) (b)

Figure 2: (a): Selected CASTNet sites. (b): Posterior distributions for the range parameter (km) of the Matérn covariance for Models-3 SO_2 values, at the 6 selected sites.

spatial structure of the pollutant concentrations. Furthermore, Models-3 concentrations are block averages over the 36km $\times$ 36km. We use here a Bayesian approach to the interpolation technique for nonstationary fields taking also into account the change-of-support problem (we calculated the block averages covariances drawing a set of 4 locations in each pixel). We use posterior predictive checks (PPC) as suggested by Rubin (1984) for validation of the air quality numerical models. Thus, we compare the posterior predictive distributions of the physical models (Figure 4 (b)) at different locations to the observed data (Figure 2 (a)), to determine if the numerical models generate data that are similar to the CASTNet data. Figure 4 (b) shows the predictive values of the Models-3 SO_2 weekly average concentrations at the CASTNet selected sites. As expected, we get very high variability at the Indiana site, this site is very close to several coal power plants, and therefore the SO_2 levels can be very high or very low depending on wind speed, wind direction, and on the atmospheric stability. The sites in Maine and Florida have the lowest SO_2 levels and variability. The agricultural site in Illinois and the site in North Carolina have similar behavior regarding SO_2

levels. The site in NC is not far from the Tennessee power plants, and the site in Illinois is also relatively close to some Midwestern power plants. The site in Michigan, which is very close to the lake Michigan and relatively far from power plants has also low SO_2 levels.



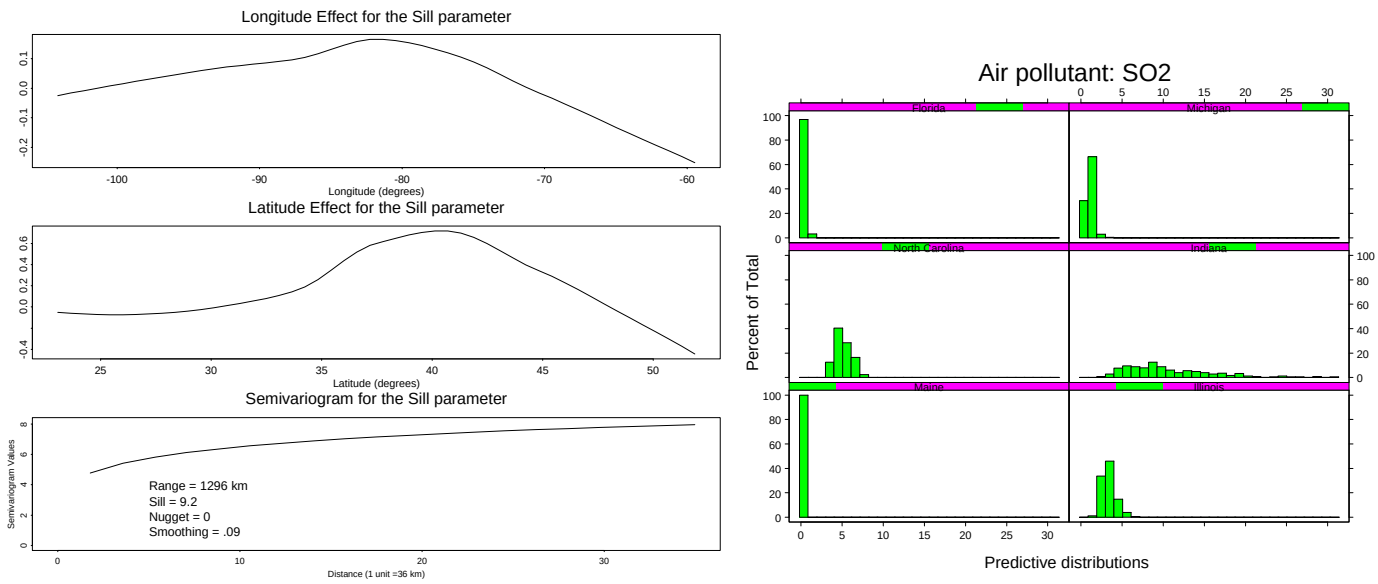
(a) (b)

Figure 3: (a):Posterior distributions for the sill parameter of the Matérn covariance for Models-3 SO_2 concentrations, for the week starting July 11, 1995. At the 6 selected locations showed in Figure 2 (a). (b):Map of the modes of the posterior distributions for the sill parameter of the Matérn covariance for Models-3 SO_2 concentrations, for the week starting July 11, 1995.

We do not consider CASTNet measurements to be the "ground truth", because there is measurement error. Thus, we assume there is an underlying (unobserved) field $Z(\mathbf{s})$, where $Z(\mathbf{s})$ measures the "true" concentration/flux of the pollutant at location $\mathbf{s}$. At station $\mathbf{s}$ we make an observation $\hat{Z}(\mathbf{s})$, corresponding to the CASTNet observation at this station, and we assume that

$$\hat{Z}(\mathbf{s}) = Z(\mathbf{s}) + e(\mathbf{s}), \tag{15}$$

where $e(\mathbf{s}) \sim N(0, \sigma_e^2)$ represents the measurement error (nugget) at location $\mathbf{s}$. The process $e(\mathbf{s})$ is independent of $Z(\mathbf{s})$. With the purpose of estimating σ_e, EPA has two collocated monitors at the



(a) (b)

Figure 4: (a):Longitude effect (r_i's), latitude effect (c_j's) and semivariogram of the error term ($\epsilon_{\boldsymbol{\sigma}}$) for the sill parameter (σ), where $\sigma(s_i, s_j) = a + r_i + c_j + \epsilon_{\boldsymbol{\sigma}}(s_i, s_j)$ for $i = 1, \ldots, n_1$ and $j = 1, \ldots, n_2$, and the process $\epsilon_{\boldsymbol{\sigma}}$ is Gaussian with mean zero and a Matérn stationary covariance with parameters $\boldsymbol{\tau}_0$. In the bottom graph we plot the Matérn covariance model of the the error term ($\epsilon_{\boldsymbol{\sigma}}$), the covariance parameters for this Matérn model are the modes of the posterior for the hyperparameters $\boldsymbol{\tau}_0$. (b):Predictive distributions for the Models-3 SO_2 concentrations, at the 6 selected locations showed in Figure 2 (a), for the week starting July 11, 1995.

Mackville site in Kentucky. The collocated monitors in Kentucky have been collecting weekly data since 1993. We estimate σ_e using the mean squared differences of all the available CASNet values at the two collocated sites, the estimated value is .3 ppb.

The graph on the left in Figure 5 presents a naive approach for evaluation of Models-3. This graph shows Models-3 versus CASTNet, without doing any spatial interpolation of Models-3. In this graph we simply have the values of Models-3 for the pixels that are the closest to each CASTNet site, without considering the change of support. In some areas the atmospheric pollutants vary significantly at scales smaller than the grid size of the model, therefore comparing the value of the grid cell with a point measurement in the ground would lead to erroneous conclusions. In Figure

5 the dotted lines indicate a 90% confidence region for CASTNet (CASTNet values $\pm 1.64 \cdot \sigma_e$). The graph on the right in Figure 5 shows the modes and 90% credible intervals of the predictive Bayesian distributions derived from Models-3 at the CASTNet locations versus the CASTNet data. The latter plot is much more informative about the fit of Models-3 to the real data, since we compare values that have the same spatial support instead of comparing grid cells with point measurements. The uncertainty in the estimated Models-3 values in this figure depends on location and is represented by the variance of the distributions in Figure 4 (b). The 90% credible intervals in Figure 5 show that at some locations, the bias in Models-3 is not significant. This Bayesian approach gives more reliable prediction errors, by taking into account the uncertainty in the covariance parameters, and the change of support. In Figure 6 we plot all CASTNet values versus the means of the predictive distributions of Models-3 at those locations. There are 3 sites where Models-3 overestimates the SO_2 values considerably, a site in Indiana, a site in Maryland, and a site in West Virginia. These three sites are close to power plants. In Figure 7 (a) we show these predictive values on a map, and in Figure 8 (a) we present the corresponding standard errors also on a map. Figure 6 is just an illustration of the powerful application of the technique presented in this paper, though it does arises some important questions that are currently being discussed with the Models-3 group regarding the formulation and improvement of the physical models. Some of the main sources of uncertainty that affect the performance of the models are the following; the photo-chemistry model parameterizations (which is the treatment of photo-chemistry phenomena varying at scales smaller than the grid size of the model), the boundary conditions, the treatment of the land use/land cover at smaller scales than the grid size, the quality of the emissions input that goes into the air quality models, and the air dispersion modeling of pollution plumes (at smaller scales than the grid size). The fact that the models perform worse in areas closer to power plants suggests that the dispersion modeling of pollution plumes in that areas needs to be improved. The Models-3 output used for

the analysis in this paper assumes that the SO_2 diffuses uniformly within each grid cell. New dispersion models are currently being added to Models-3. For more information about dispersion modeling of pollution plumes, see e.g. Beychok (1995).

We compare now our proposed modeling approach with a geostatistical kriging prediction approach (see, e.g. Cressie, 1993, for more information about kriging). In a geostatistical kriging approach for spatial prediction the covariance structure is estimated first, and then the estimated covariance is used for interpolation. The properties of the interpolants based on an estimated covariance structure are not well understood, and it is common practice to ignore the effect of the uncertainty in the covariance structure on subsequent predictions. Our Bayesian approach to interpolation of spatial nonstationary processes provides a general methodology for taking into account the uncertainty about parameters on subsequent predictions. We used for both approaches the same model for nonstationarity. For the kriging prediction the covariance parameters were the modes of the corresponding posterior distributions. In Figure 7 (a) we show the Bayesian predictive values (mean of predictive posterior distribution) at the CASTNet locations, and in Figure 7 (b) the kriging predictive values. Figure 8 shows the Bayesian standard errors (standard error of the predictive posterior distribution), and Figure 8 (b) the standard errors from kriging. Figure 9 plots the Bayesian standard errors at the CASTNet locations versus the Kriging standard error. The points in Figure 9 where the Bayesian standard errors are much larger than kriging, correspond to locations close to power plants.

As one would expect the Bayesian approach tends to originate larger standard errors, since it also takes into account the uncertainty about the covariance parameters. The Bayesian standard errors are particularly larger than kriging in areas with high emissions, mainly because the proximity of power plants and the uncertainty associated with that. In these areas the SO_2 values are high or low depending on the atmospheric stability, wind direction, and many other atmospheric factors.

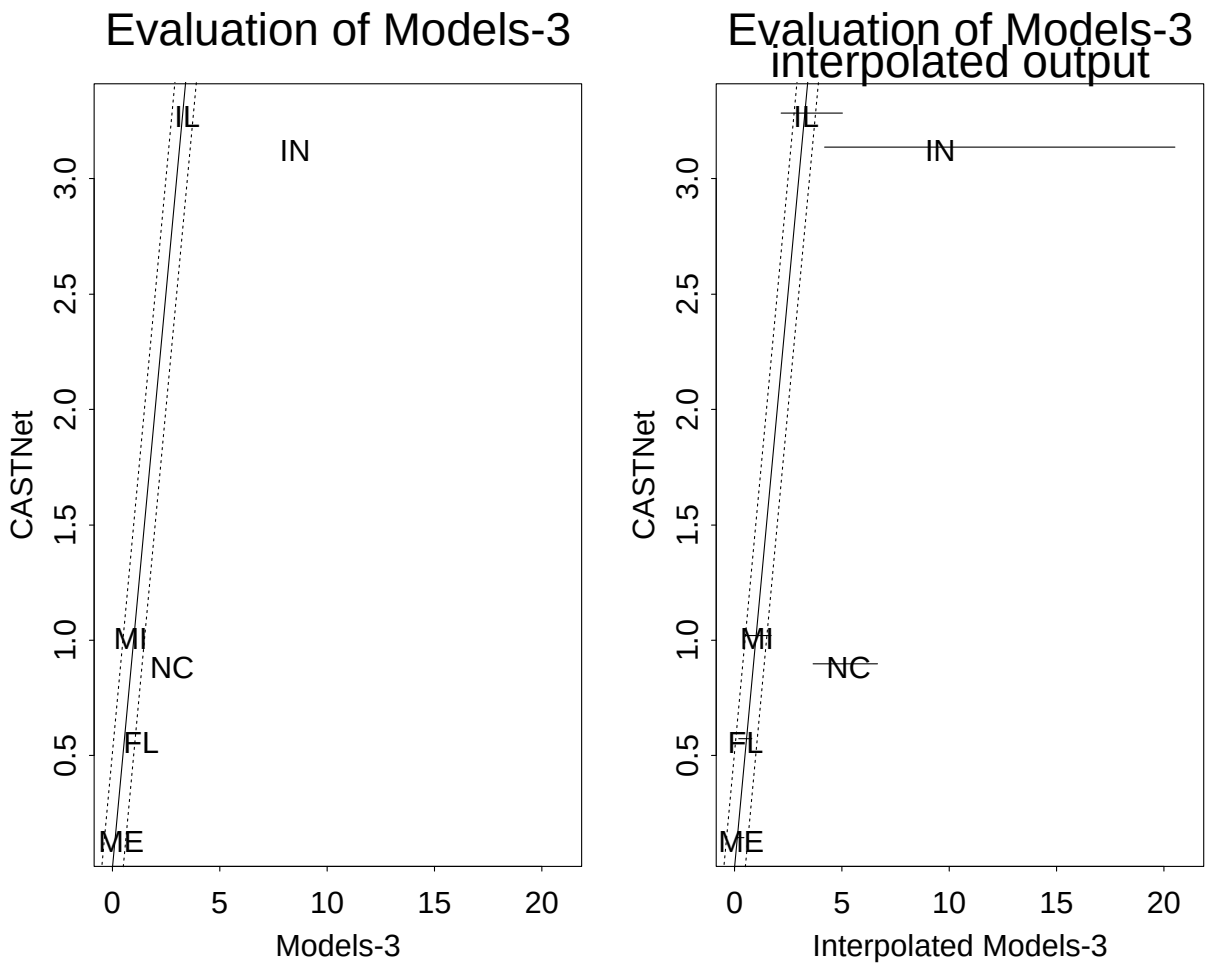


Figure 5: The graph on the left shows CASTNet measurements for the week starting July 11, 1995, versus the values of Models-3 for the pixels that are the closest to each CASTNet site, without considering the change of support. The graph on the right shows the CASTNet measurements versus the modes and 90% credible intervals of the predictive Bayesian distributions derived from Models-3 at the CASTNet locations.The dotted lines indicate a 90% confidence region for the CASTNet values.

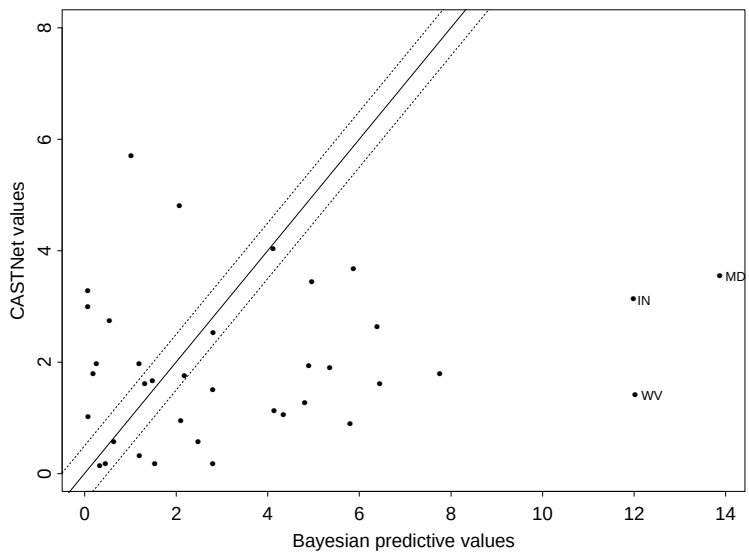


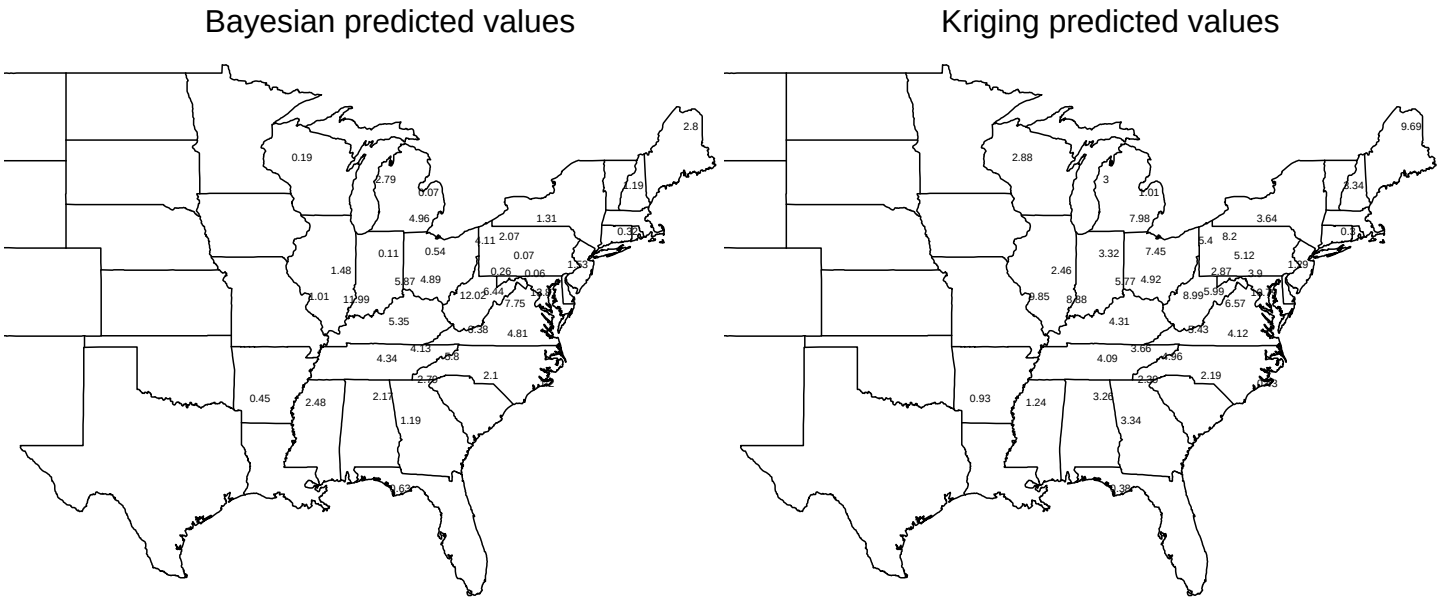
Figure 6: CASTNet values versus the mean of the predictive posterior distribution at each site.

The kriging standard errors are quite stable everywhere. This kriging method fails to capture the relatively larger uncertainty in areas close to power plants.

7 Conclusions and final remarks

We introduce in this paper a new statistical methodology for nonstationary models, the spatial field is represented locally as a stationary isotropic random field, but the parameters of the stationary random field are allowed to vary continuously across space. Kernel functions are used to ensure that the field is well-defined but also continuous. New fitting algorithms are developed. The methods are extended to prediction/interpolation questions using Bayesian approaches to account for parameter uncertainty. In this paper we also take into account the change of support problem that occurs when data and predictant have different spatial resolution.

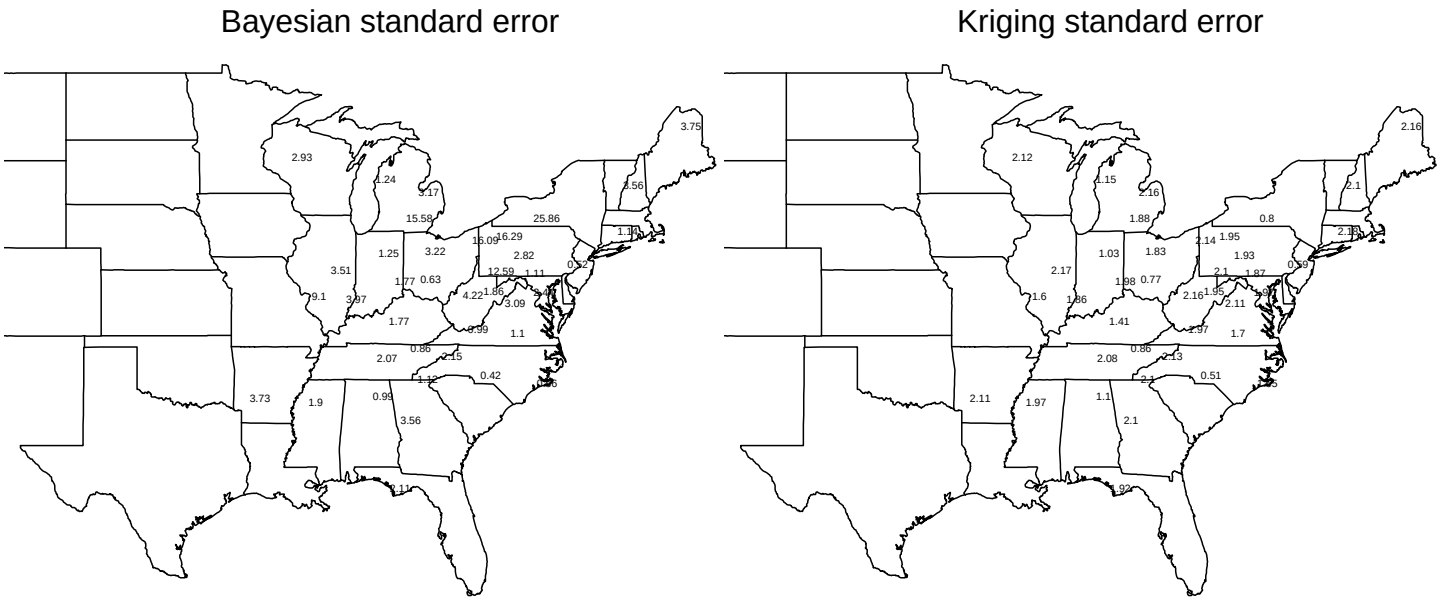
One of the main objectives of the application presented in this paper is to evaluate the physical air quality models (Models-3). Therefore, we need measures of how well Models-3 output and real



(a) (b)

Figure 7: (a): Mean of the predictive posterior distribution at each CASTNet site. (b): Kriging predictor at each CASNet location, with a nonstationary covariance.

data (CASTNet) agreed. We compared Models-3 with CASTNet at all the 50 CASTNet locations, taking into account the measurement error in CASTNet, and the change of support problem. Previous attempts to evaluate dry deposition values from Models-3, did not take into account the fact that Models-3 estimates grid-cell averages and not point values, leading to misleading conclusions about the performance of Models-3. For instance, in Figure 5 the CASNet site in NC does not seem to disagree with the models (if we ignore the change of support). But, a more careful analysis, using point predictions at the CASTNet sites suggests that Models-3 does not perform there as well as expected. Models-3 group is analyzing this site and other suggested locations, mainly in the areas close to power plants, where the models do not perform very well when compared to the ground measurements. It is very computationally expensive to run these physical models. Therefore, it is crucial for the modelers to understand where the numerical models do not perform very well. Because, they can then run the models for these areas of interest under



(a) (b)

Figure 8: (a): Standard error of the predictive posterior distribution at each CASTNet site. (b): Standard error of the Kriging predictor at each CASNet location.

different emissions scenarios, and under different model parameterizations and plume dispersion models, with the goal of improving the air quality models and understand better the reason for the disagreement between models output and ground observations.

We also compared our proposed modeling approach with a kriging prediction method, that ignores the uncertainty in the covariance parameters. Both approaches were implemented using the same model for nonstationarity. Our approach gave larger standard errors than kriging for the Models-3 predicted values, specially in areas close to power plants. This reflects the difficulty in estimating the covariance parameters in these areas. The kriging approach reported rather uniform standard errors everywhere. For the evaluation of the physical models, we studied where CASTNet values lay with respect to a confidence interval for the SO_2 predicted values from Models-3. Using kriging for the prediction we underestimated considerably the length of the confidence intervals leading to wrong conclusions about the performance of Models-3.

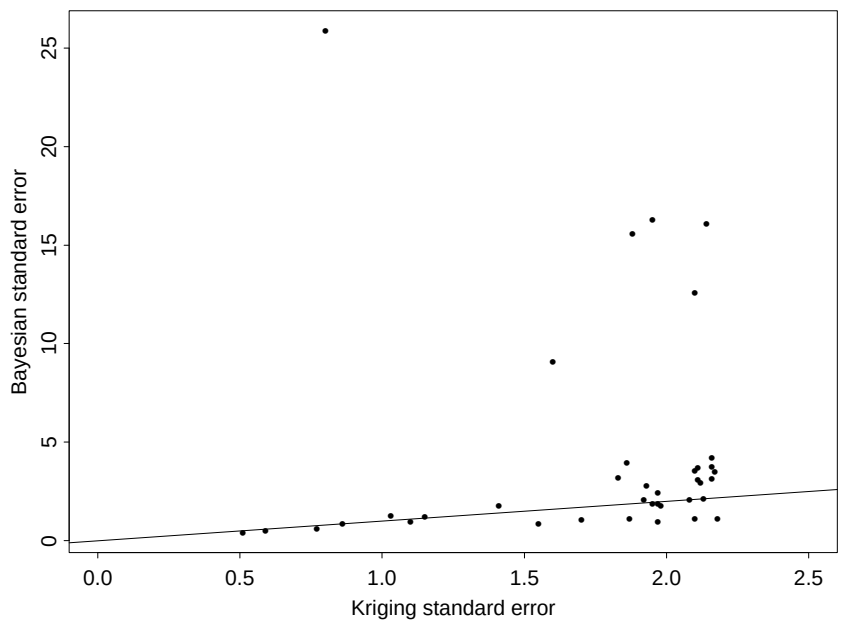


Figure 9: The standard error of the predictive posterior distribution at each CASTNet site, versus the standard error of the Kriging predictor.

The spatial model proposed in this paper could be also applied to spatial temporal data. Right now Models-3 is being run for short periods of time, but in a near future we should get access to longer periods of time than one week to conduct a spatial temporal analysis. Then, we would model Z as a spatio-temporal process,

$$Z(\mathbf{x}, t) = \mu(\mathbf{x}, t; \boldsymbol{\beta}) + \epsilon(\mathbf{x}, t)$$

where $E\left(Z(\mathbf{x}, t)\right) = \boldsymbol{\mu}(\mathbf{x}, t; \boldsymbol{\beta})$, and the process ϵ represents some spatial-temporally correlated zero-mean noise. We could represent $Z(\mathbf{x})$ using the model in Section 3 where $\mathbf{x}$ now varies over a domain D contained in a 3-dimensional Euclidean space $\mathbb{R}^3$. However, if we assume that the local stationary processes have a separable covariance, then we model Z as follows:

$$Z(\mathbf{x}, t) = \int_D K(\mathbf{x} - \mathbf{s}) Z_{\boldsymbol{\phi}(\mathbf{s})}(\mathbf{x}, t) d\mathbf{s}. \tag{16}$$

where $\boldsymbol{\phi}(\mathbf{s}) = (\boldsymbol{\theta}(\mathbf{s}), \boldsymbol{\rho}(\mathbf{s}))$ and the local processes $Z_{\boldsymbol{\phi}(\mathbf{s})}$ have a separable spatial-temporal covariance,

$$\text{cov}\left(Z_{\boldsymbol{\phi}(\mathbf{s})}(\mathbf{s}_1, t_1), Z_{\boldsymbol{\phi}(\mathbf{s})}(\mathbf{s}_2, t_2)\right) = C^{(1)}_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{s}_1 - \mathbf{s}_2) \cdot C^{(2)}_{\boldsymbol{\rho}(\mathbf{s})}(|t_1 - t_2|) \tag{17}$$

where $C^{(1)}_{\boldsymbol{\theta}(\mathbf{s})}$ is a stationary spatial covariance with parameter $\boldsymbol{\theta}(\mathbf{s})$, for instance a Matérn (7), and $C^{(2)}_{\boldsymbol{\rho}(\mathbf{s})}$ is a stationary temporal covariance with parameter $\boldsymbol{\rho}(\mathbf{s})$. For instance we could model $C^{(2)}_{\boldsymbol{\rho}(\mathbf{s})}$ as the covariance of an autoregressive AR(1) temporal model. Forms such as (17) have a history in spatial-temporal modeling; see e.g. Mardia and Goodall (1993) and references therein.

Then, the covariance of Z can be written as follows:

$$\text{cov}\left(Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2)\right) = \int_D K(\mathbf{s}_1 - \mathbf{s}) K(\mathbf{s}_2 - \mathbf{s}) C^{(1)}_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{s}_1 - \mathbf{s}_2) \cdot C^{(2)}_{\boldsymbol{\rho}(\mathbf{s})}(|t_1 - t_2|) d\mathbf{s} \tag{18}$$

this is a nonstationary spatial-temporal covariance. Note that the covariances for the processes $Z_{\boldsymbol{\phi}(\mathbf{s})}$ are separable but the covariance of Z is not separable. We should acknowledge that a separable model for Z would be no longer considered acceptable by many users. For each local process $Z_{\boldsymbol{\phi}(\mathbf{s})}$ spatial association at a fixed time point is captured through $C^{(1)}$; decay in such association over time is captured by $C^{(2)}$.

8 Acknowledgments

The authors would like to acknowledge the helpful insight and assistance provided by the US EPA Office of Research Development, RTP, NC, in particular by Dr. Peter Finkelstein and Dr. Robin Dennis.

References

Beychok, M. R. (1995). *Fundamentals of Stack Gas Dispersion*, 3rd Edition. Published by author, Irvine, California.

Cressie, N. A. (1993). *Statistics for spatial data.* Revised Edition. Wiley, New York.

Epanechnikov, V. A. (1969). Nonparametric estimates of a multivariate probability density. *Theor. Prob. Appl,* **14**, 153-158.

Fuentes, M. (2001a). A high frequency kriging approach for nonstationary environmental processes. *Envirometrics,* **12**, 469-483.

Fuentes, M. (2001b). Spectral methods for nonstationary spatial processes. *Biometrika,* to appear.

Gelfand, A., Zhu, L, and Carlin, B. P. (2000). On the change of support problem for spatio-temporal Data. Research Report 2000-011, Division of Biostatistics, University of Minnesota. To appear in *Biostatistics.*

Guttorp, P., Meiring, W. and Sampson, P. (1994), A space-time analysis of ground-level ozone data. *Environmetrics* **5**, 241–254.

Guttorp, P. and Sampson, P. (1994), Methods for estimating heterogeneous spatial covariance functions with environmental applications. In *Handbook of Statistics 12*, eds. G.P. Patil and C.R. Rao, Elsevier Science B.V., 661–689.

Haas, T.C. (1995), Local prediction of a spatio-temporal process with an application to wet sulfate deposition. *J. Amer. Statist. Assoc.* **90**, 1189–1199.

Haas, T.C. (1998), Statistical assessment of spatio-temporal pollutant trends and meteorological transport models. *Atmospheric Environment* **32**, 1865-1879.

Higdon, D., Swall, J. and Kern, J. (1999), Non-stationary spatial modeling. In *Bayesian Statistics 6*, eds. J.M. Bernardo *et al.*, Oxford University Press, pp. 761–768.

Holland, D., Saltzman, N., Cox, L.H. and Nychka, D. (1999), Spatial prediction of sulfur dioxide in the eastern United States. In *geoENV II — Geostatistics for Environmental Applications*, eds. Gómez-Hernández, J., Soares, A. and Froidevaux, R., Kluwer, Dordrecht, 65–76.

Mardia, K. V. and Goodall, C. R. (1993), Spatial-temporal analysis of mulitvariate environmental monitoring data. In *Multivariate Environmental Statistics*, eds. G.P. Patil and C.R. Rao, Elsevier Science Publishers, pp. 347–386.

Matérn, B. (1986). *Spatial Variation.* Lecture Notes in Statistics, Number 36. Springer Varlag, New York. (Second edition: originally published in 1960).

Nychka, D. and Saltzman, N. (1998), Design of air quality networks. In *Case Studies in Environmental Statistics*, eds. D. Nychka, W. Piegorsch and L.H. Cox, Lecture Notes in Statistics number 132, Springer Verlag, New York, pp. 51–76.

Rubin, D. B. (1984). Bayesian justifiable and relevant frequency calculations for the applied statistician. *Annals of Statistics,* **12**, 1151-1172.

Sampson, P.D. and Guttorp, P. (1992), Nonparametric estimation of nonstationary spatial covariance structure. *J. Amer. Statist. Assoc.* **87**, 108-119.

Smith, R.L. (1996), Estimating nonstationary spatial correlations. Preprint, University of North Carolina.

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

Walter, G. and Blum. H. (1979). Probability density estimation using delta sequences. *Annals of Statisticis,* **7**, 328-340.

Whittle, P. (1954). On stationary processes in the plane, *Biometrika,***41**, 434-449.

Yaglom, A. M. (1962). *An introduction to the theory of stationary random functions.* Prentice-Hall, NJ.