Predicting nonstationary processes using Reversible MCMC

Arin Chaudhuri,Monsterrat Fuentes,David Holland

April 26, 2004 -- Mimeo Series 2553

Abstract

Spatial processes for air pollutants defined over large geographic areas rarely exhibit stationary behavior. We propose a non-stationary spatial covariance model that is a mixture of stationary processes. The number of stationary processes and their parameters are estimated using a Reversible Jump MCMC approach in a hierarchical Bayesian framework. We also introduce methods for an efficient implementation and fast computation. We apply this method to the modeling, prediction of ambient ozone in the Eastern US.

1 Introduction

Atmospheric pollutants have significant impact on health. The same chemical properties that allow high concentrations of ozone to react with organic material outside the body give it the ability to react with similar organic material that makes up the body, and potentially cause harmful health consequences. When inhaled, ozone can damage the lungs. Relatively low amounts can cause chest pain, coughing, shortness of breath, and, throat irritation. Ozone may also worsen chronic respiratory diseases such as asthma and compromise the ability of the body to fight respiratory infections. To study the the impact of various atmospheric pollutants health and terrestrial and aquatic aqua-systems the Clean Air Act Amendments (CAAA) established a monitoring network to to assess improvements in air quality throughout the United States. To meet the objective Clean Air status and trends Network (CASTNet) was established by the United States Environment Protection

Agency (EPA) to monitor rural areas. CASTNet is made up of 51 sites mostly in eastern United States. CASTNet monitors various pollutants including Ozone (O_3), sulfur dioxide (SO_2), Nitric Acid (HNO_3). The CAAA also established ambient air quality standards for carbon monoxide (CO), lead (Pb), nitrogen dioxide (NO_2), particulate matter and Ozone. State and Local Air Monitoring Network Stations(SLAMS) and National Air Monitoring Stations(NAMS) were set up to monitor compliance with the air quality standards, these monitors are predominantly in the urban areas. In this paper we study the distribution Ozone from the SLAMS/NAMS/CASTNet networks. The distribution of Ozone is known to exhibit non-stationary behavior, i.e., the spatial distribution of Ozone depends on where it is being measure, taking this into consideration we fit a hierarchical Bayesian model arising out of the convolution of stationary processes centered at various locations. We do not assume the number of local stationary process required to fit the data adequately to be known in advance and hence our model consists of a variable number of parameters and we use the methods of RJMCMC to determine the number of the processes and to estimate their parameters. The next section is a literature review of the work done in non-stationary modeling so far.

2 Literature Review

In recent years, probably the most extensively studied method for non-stationary spatial processes is the deformation approach due to Sampson and Guttorp [11], In a series of papers best represented by Haas [5] , T. Haas has proposed an approach to non-stationary spatial kriging based on moving windows. Higdon, Swall and Kern [6] 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 [8] and Holland *et al.* [7], that extends the "empirical orthogonal functions" (EOF) approach that is popular among atmospheric scientists. We describe below these approaches to modeling non-stationarity in the spatial context.

2.1 Deformation Approach of Sampson and Guttrop

The Sampson-Guttrop [11] approach does not assume the underlying process to be either stationary or isotropic. The approach involves the following steps:

- Using multidimensional-scaling a two dimensional coordinate representation of the sampling stations is obtained with the distances representing the *dispersion* [1] between sites, i.e., given locations $x_1, \ldots x_n \in R^2$ and a spatial process Z which has been observed at locations $x_1, \ldots x_n$ they construct the transformed coordinates $y_1, \ldots y_n \in R^2$ such that

  $$\|y_i - y_j\| = E(Z(x_i) - Z(x_j))^2$$

  The *deformed* space is called the D-space and the original space is called the G-space. The points are obtained by minimizing the *stress criterion*

  $$\min_{\delta} \frac{\sum_{i<j} \{\delta(d_{ij}) - h_{ij}\}^2}{\sum_{i<j} {h_{ij}}^2}$$

  Where δ varies over all positive monotonically increasing functions.

- Next using thin plate splines they estimate a function *s*mooth function f which satisfies $y_i = f(x_i)$ for i=1, ..., n over the whole domain.

- The *dispersion* between points at x_1 and x_2 is estimated as $\|f(x_1) - f(x_2)\|$

In the approach of Sampson and Guttrop [11] the function f might be a mapping from $\mathcal{R}^2$ to $\mathcal{R}^d$ where might be greater than two, but they assume d=2 in their applications. . Damian, Sampson and Guttrop [10] have also approached the problem in a bayesian framework in which they estimate the deformed coordinates and the parameters of the correlation function and the variance parameters. In the Bayesian case they parametrize the *dispersion* between two points, i.e, they consider the following form

$$\rho_\theta(|f(x_1) - f(x_2)|)$$

where ρ_θ is a known parametric family.

[1] Sampson and Guttrop define dispersion between the spatial variables $Z(x)$ and $Z(y)$ as $E(Z(x) - Z(y))^2$

2.2 Models based on EOF

The EOF (*e*mpirical orthogonal functions) approach is similar to the method of *Principal Components*. Here it is assumed that the variable of interest is obtained by smoothing a random field $X(s)$ and adding an uncorrelated error term to it. The model is given by

$$y_i = \int_D X_i(s)B(s)ds + \epsilon_i \; i = 1, \ldots, n$$

Where y_i is the variable of interest and ϵ_i's are orthogonal to the $X_i(.)$'s. Solving for $B()$ involves solving a functional system of equations

$$g_i(t) = \int_D C_i(s,t)B(s)ds \tag{1}$$

Where $g_i(t) = cov(y_i, X_i(t))$ and $C_i(s,t) = cov(X_i(s), X_i(t)$
Assuming that the *Karhunen-Lóeve expansion* holds, we are led to the following form for $C_i(s,t)$

$$C_i(s,t) = \sum_k \lambda_{ki}\psi_{ki}(s)\psi_{ki}(t)$$

Where λ's are the eigenvalues and ψ's are the eigenvectors of the functional h_i defined by :

$$h_i : f \to \int_D f(s)C_i(s,t)ds$$

The conditions for the number of eigenvectors being countable and of the eigenvectors being complete are usually satisfied. The spectral decomposition of the covariance also suggests the following representation of the process

$$Y_i = \sum_k Z_k \lambda_{ki}^{\frac{1}{2}} \psi_{ki}(s) \tag{2}$$

Where $\{Z_k\}$ is a collection of uncorrelated random variables. In fact we can start with a collection of orthogonal functions and construct a process by the formula given in (2) and obtain a non-stationary process.

Nychka and Saltzmann [8] extend this idea and suggest the following form of non-stationary covariance

$$C(s_1, s_2) = \sigma(s_1)\sigma(s_2)\left\{\rho \exp(-\frac{\|s_1 - s_2\|}{\theta}) + \sum_{k=1}^{M} \lambda_k \psi_k(s_1)\psi_k(s_2)\right\}$$

The above model departs in stationarity from the exponential family of covariances, the value of M determines the extent of the non-stationarity.

Taking the idea of orthogonal expansions further Nychka,Wilke and Royle [9] have proposed a wavelet basis expansion :

$$Z(s) = \sum_{k=1}^{MN} a_k \psi_k(s)$$

Where the wavelets are centered around a $M \times N$ grid , and that a_k's are distributed as MVN(0,Σ_a). Let $s_1, \ldots s_n$ be any arbitrary collections of points, and let $\lambda_{i,j} = \psi_i(s_j)$. Then the covariance of $Z = (Z(s_1), \ldots Z(s_n))$ is $\Lambda\Sigma_a\Lambda$ The computations get complicated since Σ_a need not be diagonal, Nychka *et al.* approximate it by a near diagonal matrix in applications.

2.3 Kernel based approaches

Higdon, Swall and Kern [6] have proposed an approach based on *spatially evolving kernels*. The model proposed by them is of the form :

$$Z(s) = \int_D K_s(u) X(u) du \tag{3}$$

Where $\{K_s() : s \in D\}$ is a family of of kernels and $X(u)$ is white noise. The correlation between points at x_1 and x_2 is proportional to $\int_D K_{x_1}(u) K_{x_2}(u) du$ which is not a function of the difference $x_1 - x_2$ between the locations. The kernels are chosen to vary smoothly with space, Hidgon *et al.* choose a family of normal kernels whose parameters are continuous functions of the location. The covariance model depends solely on the choice of the kernel functions.

Fuentes [2] considers non-stationary processes which arise as a weighted average of the stationary processes with the weights depending on the location. The model is :

$$Z(x) = \sum_{i=1}^{k} Z_i(x) w_i(x) \tag{4}$$

Where $Z_i()$'s are mutually orthogonal stationary processes with covariances $C_i()$. Under the assumptions it is easily seen that

$$cov(Z(x), Z(y)) = \sum_{i=1}^{k} w_i(x) w_i(y) C_i(x - y)$$

.

Some special cases :

- By choosing $w_i(x)$ to be constants for all i we obtain a stationary model.
- Let $S_1, \ldots S_M$ be a partition of the domain into disjoint subregions and let $w_i(x) = I(x \in S_i)$ for $i = 1, \ldots, M$ then the covariance structure given by (4) is :

$$C(x, y) = \begin{cases} C_i(x - y) & \text{if x,y} \in S_i \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

  A non-stationary model obtained by the restriction of different stationary models in disjoint regions.

Fuentes and Smith [3] have proposed a model obtained by assuming a family of mutually orthogonal stationary process and convoluting them to obtain the final non-stationary process. In the particular case where the number of stationary processes is finite the model is given by :

$$Z(x) = \sum_{i=1}^{k} K(x - s_i) Z_{s_i}(x) \tag{6}$$

Here $K(.)$ is the kernel function and $Z_{s_i}(.)$'s are mutually orthogonal stationary process with covariance $C_{s_i}(.,.)$. The covariance function of the above process in this case is easily seen to be

$$C(x_1, x_2) = \sum_{j=1}^{k} K(x_1 - s_j) K(x_2 - s_j) C_{s_i}(x_1 - x_2)$$

The natural extension of such a covariance to the continuous case leads to the Fuentes-Smith model

$$C(x_1, x_2) = \int_D K(x_1 - s) K(x_2 - s) C_s(\|x_1 - x_2\|) ds \tag{7}$$

Here $\{C_s(.,.) : s \in D\}$ is a family of stationary covariances. The above covariance is clearly non-stationary in general. In this approach, the family of covariances are smoothed by a kernel to obtain the final covariance structure, with the contribution of the covariance at s diminishing at distances far away from s.

Note that the above model is a special case of [2] with

$$w_i(x) = K(x - s_i)$$

We propose a model that is based on a generalization of (7) and has both continuous and discrete models as special cases.

3 Our Model

We will discuss about various forms of the covariance structure conditional on the number of *components* k and the spatial process $S = (s_1, \ldots s_k)$.

A problem with the model with covariance structure given by

$$c(x, y) = \sum_{i=1}^{k} K(x - s_i)K(y - s_i)C_{s_i}(x - y) \tag{8}$$

is that it does not give a stationary model unless the kernel K is a constant. Since we will not choose a trivial Kernel it makes sense to add a stationary component to the above model. So, we suggest the following model conditional on k and $s_1, \ldots s_k$

$$c(x, y) = C_0(x - y) + \alpha \sum_{i=1}^{k} K(x - s_i)K(y - s_i)C_{s_i}(x - y) \tag{9}$$

Here $C_0(\,)$ is a stationary covariance, and $\alpha \geq 0$ is a parameter that measures the deviation from stationarity. A covariance of the above form arises from processes of the following nature,

$$z(x) = Z_0(x) + \sqrt{\alpha} \sum_{i=1}^{k} K(x - s_i)Z_i(x) \tag{10}$$

Here, $Z_0(\,), Z_1(\,) \ldots$ are mutually orthogonal stationary processes with covariance functions $C_0(\,), C_1(\,) \ldots$ and $\alpha \geq 0$ and k are parameters that determine the extent of non-stationarity in the process. We also assume that $s_1, \ldots s_k$ are distributed as per a possion process with rate λ

The covariance in 10 conditional on the poisson process $s_1, \ldots s_k$ is given by

$$[C_(x, y)|k, s_1, \ldots s_k] = C_0(x, y) + \alpha \sum_{i=1}^{k} K(x - s_i)K(y - s_i)C_{s_i}(x - y)$$

Averaging out $s_1, \ldots s_k$ we get

$$C(x, y) = C_0(x, y) + \mathrm{E}([\alpha \sum_{i=1}^{k} K(x - s_i)K(y - s_i)C_{s_i}(x - y)|k, s_1, \ldots s_k]) \tag{11}$$

In the special case that the poissoin process is uniform with rate λ we have

$$C_(x, y) = C_0(x, y) + \lambda \frac{\int_D K(x - s)K(y - s)C_s(x - y)}{|D|}$$

4 Reverse Jump Markov Chain Monte Carlo

In a large class of statistical problems, the dimensionality of the parameters is not fixed but has to be determined from the data. One such example would be in the problem of mixture devconvolution with unknown number of components.

So, the problem is to choose from a family of countable models $\{\mathcal{M}_k : k \in \mathcal{K}\}$, here $\mathcal{M}_k$ has n_k parameters, say $\theta^{(k)}$ The joint distribution of $(k, \theta^{(k)}, y)$ can be modeled as

$$p(k, \theta^{(k)}, y) = p(k)p(\theta^{(k)}|k)p(y|k, \theta^{(k)})$$

Let $\mathcal{C}_k$ denote the subspace $\{k\} \times \theta^{(k)}$. Green [4] devises a MCMC scheme under which at each step transitions are made from the current subspace $\mathcal{C}_1 = \{n_1\} \times \mathcal{R}^{n_1}$ (say) to a possibly different subspace $\mathcal{C}_2 = \{n_2\} \times \theta^{(n_2)}$ say is proposed with probability $j(x)$ where $x = (n_1, \theta^{(n_1)})$ is our current state. The transition is made to a point in $\mathcal{C}_2$ as follows, generate a continuous random variable of length m_1 say $u^{(1)}$,and set $\theta^{(2)}$ to be deterministic function of $\theta^{(1)}$ and $u^{(1)}$, similarly to switch back, generate a continuous random variable of length m_2 say $u^{(2)}$,and set $\theta^{(1)}$ to be deterministic function of $\theta^{(2)}$ and $u^{(2)}$, m_1 and m_2 should be chosen such that

$$m_1 + n_1 = m_2 + n_2 \tag{12}$$

and the mapping

$$(\theta^{(1)}, u^{(1)}) \leftrightarrow (\theta^{(2)}, u^{(2)}) \tag{13}$$

is a diffeomorphism. Conditions (12) and (13) are called the *dimension balancing* conditions.

The transition probability is given by

$$\min\left\{1, \frac{p(2, \theta^{(2)}|y)j(2, \theta^{(2)})q_2(u^{(2)}|(2, \theta^2))}{p(1, \theta^{(1)}|y)j(1, \theta^{(1)})q_1(u^{(1)}|(1, \theta^1))} \left|\frac{\partial(\theta^{(2)}, u^{(2)})}{\partial(\theta^{(1)}, u^{(1)})}\right|\right\} \tag{14}$$

Here $q_1()$ and $q_2()$ are densities of $u^{(1)}$ and $u^{(2)}$ respectively and $\dfrac{\partial(\theta^{(2)}, u^{(2)})}{\partial(\theta^{(1)}, u^{(1)})}$ is the Jacobian of the transformation between $(\theta^{(1)}, u^{(1)})$ and $(\theta^{(2)}, u^{(2)})$

To illustrate this approach with an example let us consider the following problem: let our target distribution be triangular on $[-\frac{1}{2}, \frac{1}{2}]$ with probability p_1 and uniform on $[-\frac{1}{2}, \frac{1}{2}] \times [-\frac{1}{2}, \frac{1}{2}]$ with probability p_2. We propose only

one kind of move, that which switches dimensions. Consider the following diffeomorphism on $\mathcal{R}^3$

$$(x_1, x_2, x_3) \longrightarrow (2x_1 + x_2, x_1 + x_3, x_3)$$

Using this we set up the following transition rule:

If we have a point in $\mathcal{R}^1$ say x we generate two $N(0,1)$ random variables u_1, u_2 and jump to the point $(2x + u_1, x + u_2)$ with probability r_1. If we are currently at point in $\mathcal{R}^2$ say (x_1, x_2) we generate a $N(0,1)$ random variable u the inverse of the mapping above determines the point to jump back to i.e., $x_2 - u$ with probability r_2. Note, that the dimension matching condition is trivially satisfied (1+2=2+1) and that the jacobian of the transformation is 2. The transition probability for jumping from dimension 1 to 2, i.e., from $x \longrightarrow (2x + u_1, x + u_2)$ is given by

$$\min\left\{1, \frac{p_2 I(|2x+u_1|<0.5) I(|x+u_2|<0.5)\sqrt{2\pi}\exp(-\frac{u_2^2}{2})r_2}{4p_1(1-|2x|)\exp(-(\frac{u_1^2+u_2^2}{2}))r_1}\right\}$$

and probability of the reverse jump from (x_1, x_2) to $x_2 - u$ is given by

$$\min\left\{1, \frac{4p_1(1-|2(x-u)|) I(|2(x-u)|<1)\exp(-\frac{(2u-x_2)^2+u^2}{2})r_1}{p_2\sqrt{2\pi}\exp(-u^2/2)r_2}\right\}$$

5 Setting up the hierarchical Bayesian model

We consider a hierarchichal Bayesian approach with the following steps

- $(Z(x_1), Z(x_2) \ldots Z(x_n)) \sim \mathrm{N}(0, \Sigma(\theta))$ where $\Sigma(\theta)[i,j] = C_0(x-y) + \sum_{l=1}^{k} K(x_i - s_l)K(x_j - s_l)C_l(x_i - x_j)$
- $(s_1, s_2, \ldots s_k)$ is distributed as a poisson process with rate λ
- Conditional on $(s_1, s_2, \ldots s_k)$ the range, sill and nuggets of $C_i()$ are distributed as $U[0,\delta]$,$\Gamma(\alpha_1, 1)$ and $\Gamma(\alpha_2, 1)$ for all i

5.1 RJMCMC moves

We implement the RJMCMC algorithm by choosing one among a choice of transition functions $q_1(.,.), \ldots q_m(.,.)$. If our markov chain is currently at state t the $j_k(t)$ denotes the probability of choosing the transition method $q_k(.,.)$.

We propose the following kind of RJMCMC moves

- Insert a new *center* and update the value of θ

In this move we propose a parameter with an additional center compared to the current state of the markov chain. This leads to a jump in dimensionality. We also have to simulate additional parameters for the stationary process associated with the additional center. To implement this move we partition the who domain into zones and depending on which zone the center falls we choose the distribution of the parameters of the stationary process.

- Delete a *cluster point* and update the value of θ

This move leads to a drop in dimensionality. Suppose the the current state of the markov chain we has l centers, we choose one of the centers at random and delete the center and the associated stationary distribution from the model. This move and the and previous move have to be coordinated in a way so that the RJMCMC condition is met.

- No new cluster point added or deleted, value of θ updated

In this step there is no change in dimensionality and the form of the transition probability is same as that of an usual MCMC transition probability.

The parameterization of our process can be represented as $\theta = (k, \theta^{(k)})$, where $\theta^{(k)} = (l_1, \ldots l_k)$ and $l_k = (s_k, c^{(k)})$ where s_k denotes a position and $c^{(k)}$ represents the parameters of the Matern process centered around s_k.

Let $y = (z(x_1) \ldots z(x_n))$ be the locations where the process is observed, then the distribution of the process conditional on the parameters is given by

$$f(z|(k, \theta^{(k)}) = \frac{1}{\sqrt{(2\pi)|\Sigma(k, \theta^{(k)})|^k}} \exp(\frac{-y^T \Sigma(k, \theta^{(k)})^{-1} y}{2})$$

To implement the RJMCMC process we have to specify the probability of choosing each of the moves and method of generating the extra parameters when the dimension is increased and chopping off parameters when the dimension is decreased. We do this as follows

1. Each of the above moves is chosen with the same probability $\frac{1}{3}$

2. In the case we decide to increase the dimension by 1 we do as follows:

   We partition the domain D into $\{D_1, \ldots D_P\}$. When we increase the dimension from k to k+1 we have to generate an extra point $l_{k+1} = (s_{k+1}, c^{(k+1)})$. To do this we first generate s_{k+1} uniformly over D and then conditional on s_{k+1} we generate c_{k+1} as

   $$f(c^{(k+1)}|s_{k+1}) = \sum_{i=1}^{P} g_i(c^{(k+1)}) I_{(s_{k+1} \in D_i)}$$

   that is with each block of the partition D_i we have an density of the parameters of the matern process associated $g_i(\,)$ and depending on where the new location falls we generate the parameters of the process with the appropriate density.

3. When we decide to drop a dimension from k to k-1 we do as follows:

   We random choose one of $l_1, \ldots l_k$ and delete it to get the reduced dimension

4. When we decide not to change dimensions and we only update $\theta^{(k)}$ we first choose on of $l_1 \ldots l_k$ randomly say the point chosen is $l_{k'} = (s_{k'}, c^{(k')})$, we shift the $s_{k'}$ in an ϵ neighborhood and conditional on the new value of $s_{k'}$ we adjust the value $c^{(k')}$ as per step 2.

5.2 Estimation

We propose estimation of the parameters by first estimating the value of k and then conditional on this estimated value of k we estimate the value $\theta^{(n_k)}$. So, suppose, given a sample from RJMCMC, say $(k_1, \theta^{(n_{k_1})}) \ldots (k_N, \theta^{(n_{k_N})})$ We estimate k by $\hat{k} = \text{mode}\{k_1, \ldots k_N\}$

Now using the fact that $[k, \theta^{(k)}|y] = [k|y] \times [\theta^{(k)}|k, y]$ we can infer about the appropriate value of k. Within sample MCMC would help us infer about both k and $\theta^{(k)}|k$

5.3 Prediction

We want to predict the value of the process at a particular point, say at x_0. The prediction is done in following stages: we first construct a grid on

the real line, say $X = \{x_i \in \mathcal{R}\}$, next we calculate the posterior predictive distribution of $z(x_0)$ at each and every point of the grid and estimate the posterior predictive distribution of the process at x_0 at each point in X and we finally estimate the mean of posterior predictive distribution using the values of the posterior predictive density at points of X for an estimate of $z(x_0)$.

The implement the second step we simulate from the posterior distribution of parameters, say a sample is $(\theta^1, k^1), \ldots (\theta^m, k^m)$. Let $f(|k, \theta)$ denote the density of $z(x_0)$, then the posterior predictive density of $z(x_0)$ at $x_i \in X$ can be calculated as

$$\hat{f}_{pos}(x_i) = \frac{1}{m} \sum_{i=1}^{m} f(x_i|(k^i, \theta^i))$$

The expected value of the posterior predictive density, i.e., $\int x f_{pos}(x) dx$ can be estimated by among other ways

1. Estimating the density from its values in X using kernel smoothing and calculating the area under the curve of the estimate

2. Simply estimate it as

$$\sum_{i=2}^{m} x_i \hat{f}_{pos}(x_i)(x_i - x_{i-1})$$

   provided the x_i's lie sufficiently close to each other.

One can also consider using the mode, rather than the mean to get our prediction, that can be easily estimated by choosing the x_i that maximizes $\hat{f}_{pos}(x_1), \hat{f}_{pos}(x_2), \ldots$

6 Application

The data for the application consists of 236 sites from the SLAMS/NAMS/CastNET netowrk. The observations consist of Ozone standard values observed at these 236 sites. For the purposes of computation we restricted the number of sites to be at least 3 and to be 5 at the maximum. Conditional on k, all the ranges and the partial sills were given gamma priors $\Gamma(70,1)$ and $\Gamma(50,1)$ respectively, the nuggets were uniformly distributed in the interval [0,6]. The

prior on k was uniform between its maximum and minimum values. 50,000 iterations of the simulation were run and the first 10,000 were disregarded as burn-off values, because of the large magnitude of the numbers involved some calculations were done in extended precision arithmetic.

The number of parameters in different iterations is variable so to get an idea of the posterior distribution of certain parameters we restricted ourselves to those observations which had the number of centers finally selected. The predicted and standard error surfaces where obtained by breaking up the whole region into a grid of size 800×800. For ease of computation we used the mode of the posterior predictive distribution as our estimate.

The number of centers chosen is 3. The centers of the process seem to be well separated and lie in the different regions of the US, the nuggets take small values close to each other. The partial sills seem to vary significantly across the three centers with values and centers labeled 1 and 3 significantly higher than the others, the ranges in the three regions also vary sharply as can be seen in the figure. The credible intervals at the locations where estimated centers lie.

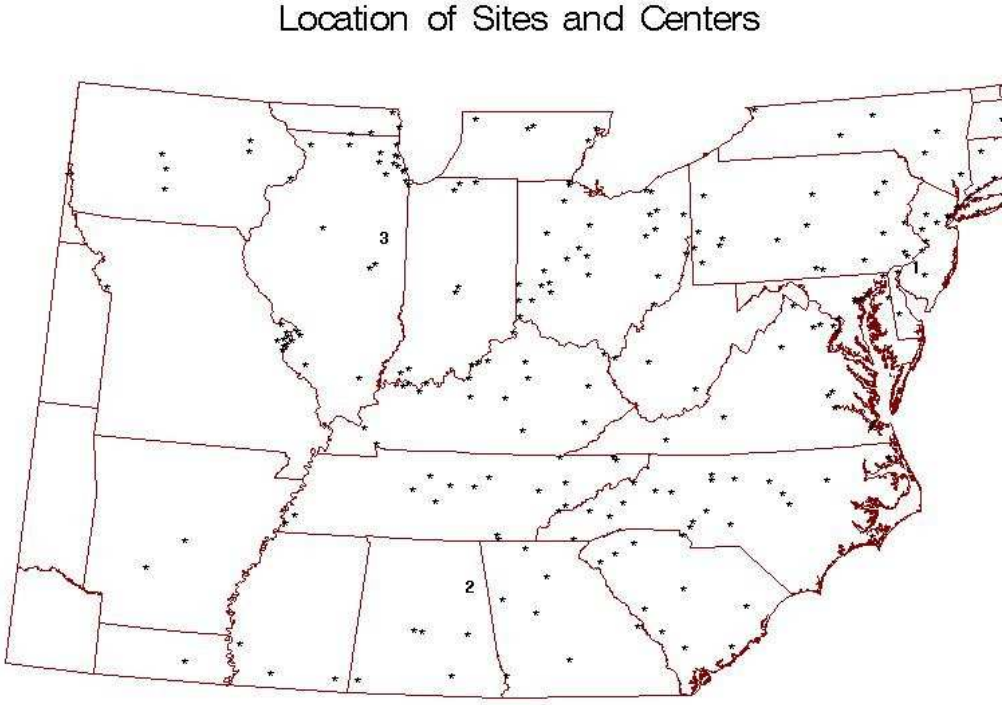

Figure 1: Location of the the sites, the centers are marked 1,2 and 3

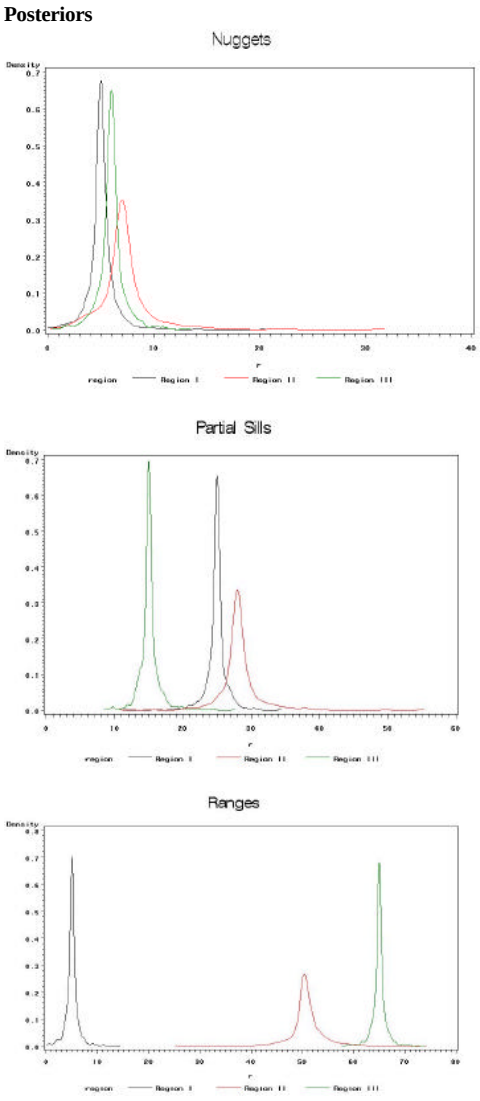

Figure 2: Posterior Distribution at the range sill and nugget parameters

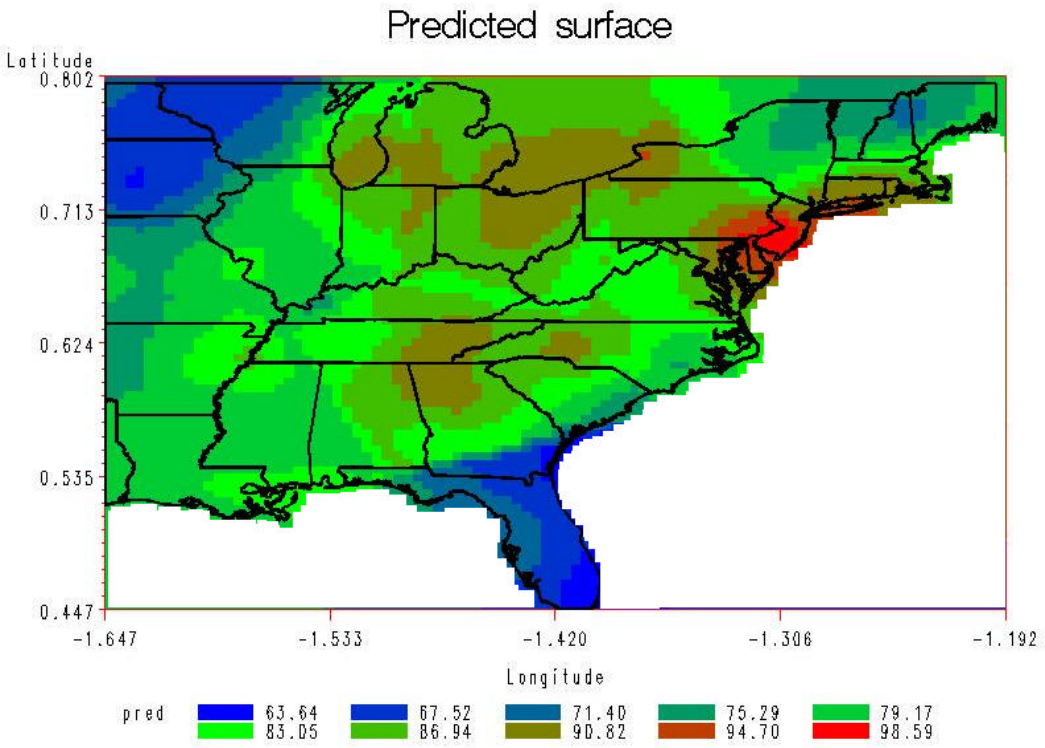

Figure 3: Predicted Surface

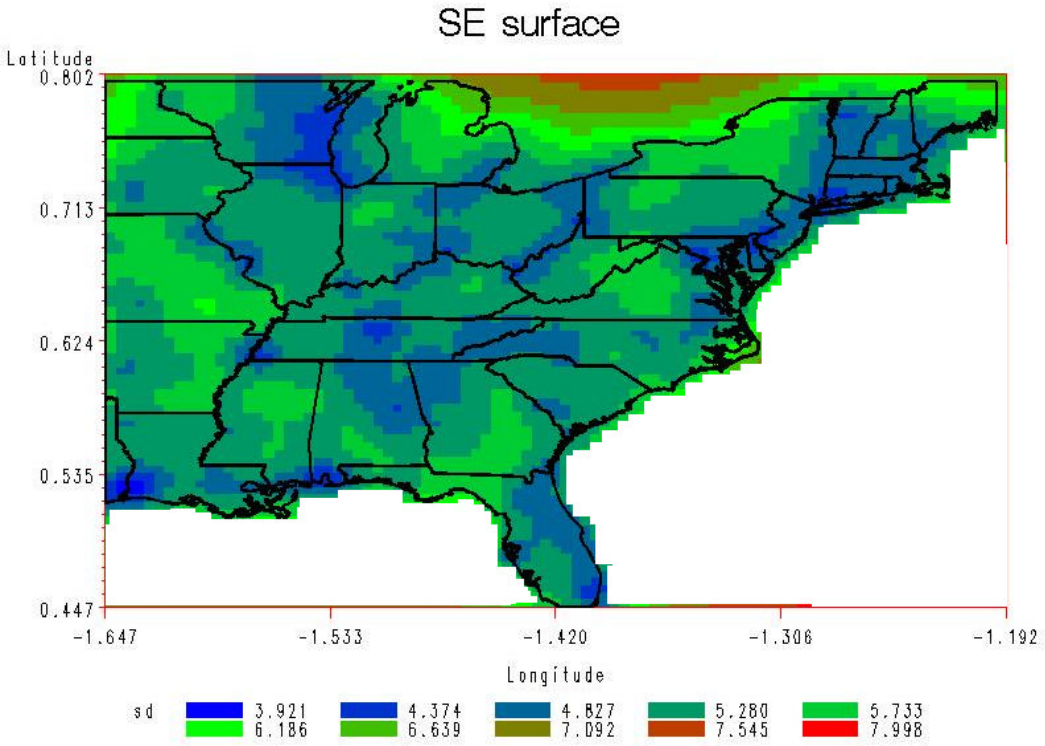

Figure 4: Standard Error Surface

References

[1] S. P. Brooks, P. Giudici, and G. O. Roberts. Efficient construction of reversible jump markov chain monte carlo proposal distributions.

[2] M. Fuentes. High frequency kriging appoach for nonstationary environmental processes. *Environmetrics*, 12:469–483, 2001.

[3] M. Fuentes and R. Smith. A new class of non-stationary spatial models. 2001.

[4] P. Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrica*, 82:711–732, 1995.

[5] Timothy C. Haas. Kriging and automated variogram modeling within a moving window. *Atmospheric Environment*, 24A:1759–1769, 1990.

[6] D. Higdon, J. Swall, and J. Kern. Non-stationary spatial modeling. pages 761–768.

[7] D. Holland, N. Saltzman, L. H. Cox, and D. Nychka. Spatial prediction of sulfur dioxide in eastern united states. In J. Gomez-Hernandez, A. Soares, and R. Froideveaux, editors, *geoENV-II - geostatistics for environmental applications*, pages 65–76. Kluwer, 1999.

[8] D. Nychka and N. Saltzman. Design of air quality networks. In D. Nychka, W. Piegorsch, and L. H. Cox, editors, *Case studies in enivoromental statistics*, pages 51–76. Springer Verlag, 1998.

[9] D. Nychka, C. Wilke, and J. Andrew Royle. Multiresolution models for nonstationary spatial covariance functions. *Statistical Modeling*, 2(4):315–331, 2002.

[10] D. Damian P. D. Sampson and P. Guttrop. Bayesian estimation of semi-parametric non-stationary spatial covariance structures. *Environmetrics*, 12:161–178, 2001.

[11] P. D. Sampson and P. Guttrop. Nonparametric estimation of nonstationary covariance structure. *Journal of the American Statistical Association*, 87(417):108–119, 1992.