

*University of Warwick. May 2007*

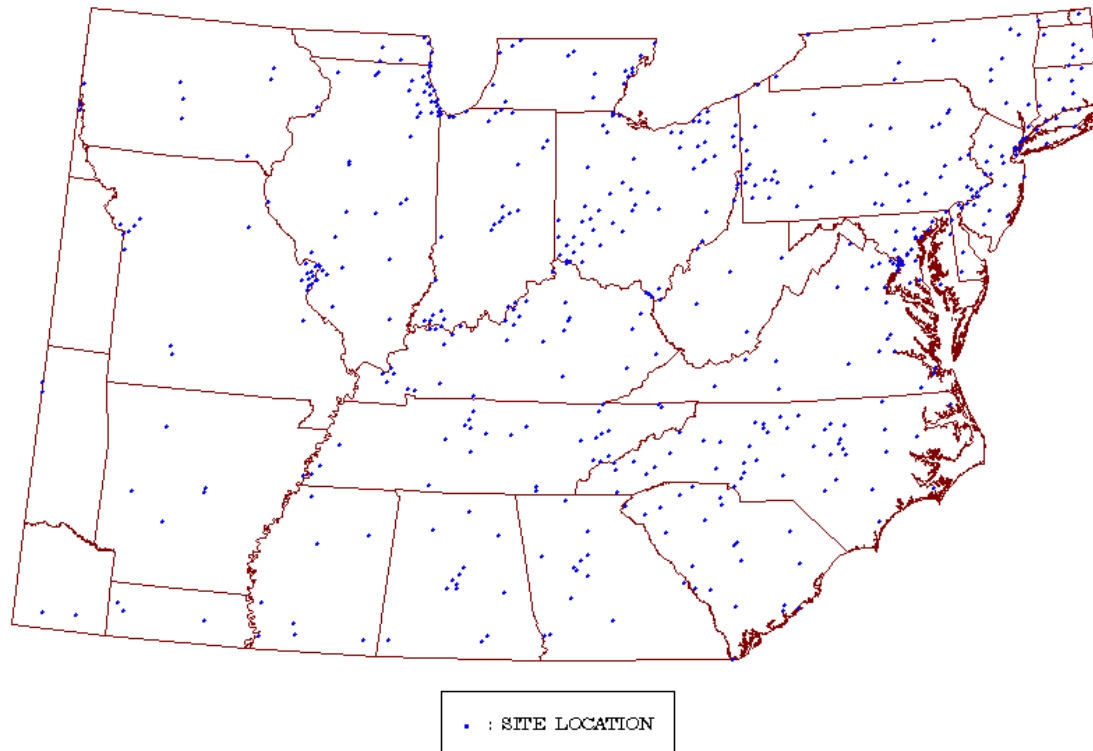
# Nonstationary covariance models

**Montse Fuentes**

North Carolina State University

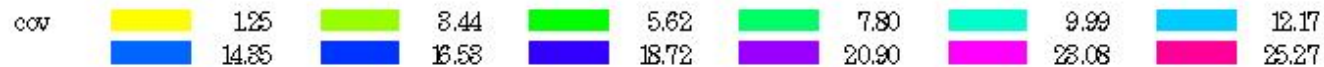
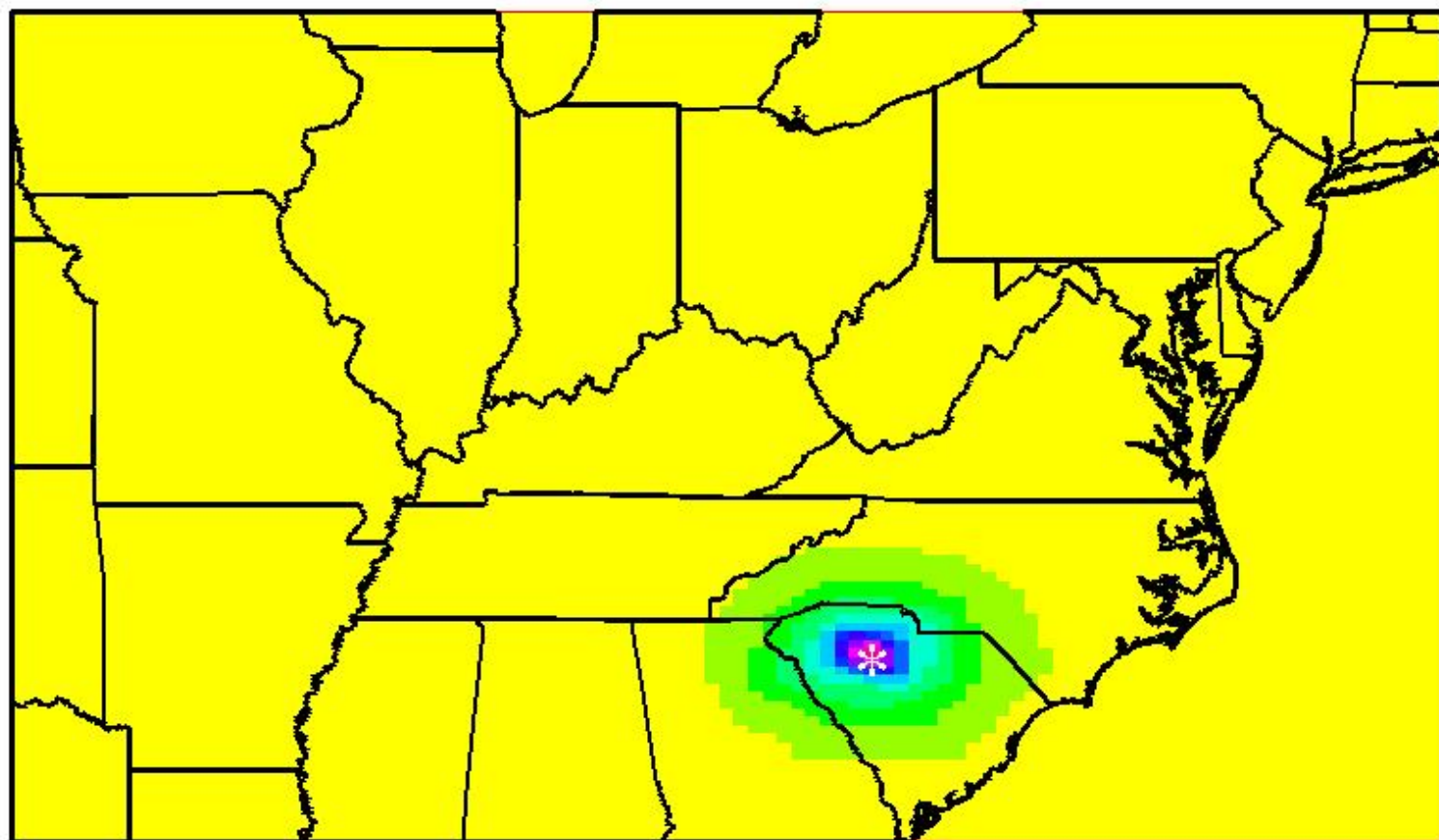
## Example of nonstationary kriging (Fuentes, 2002b)

### LOCATION OF THE SITES

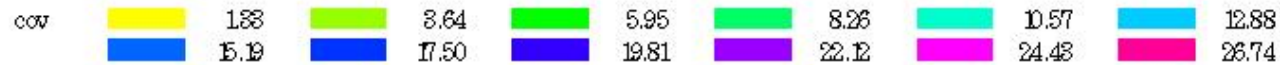
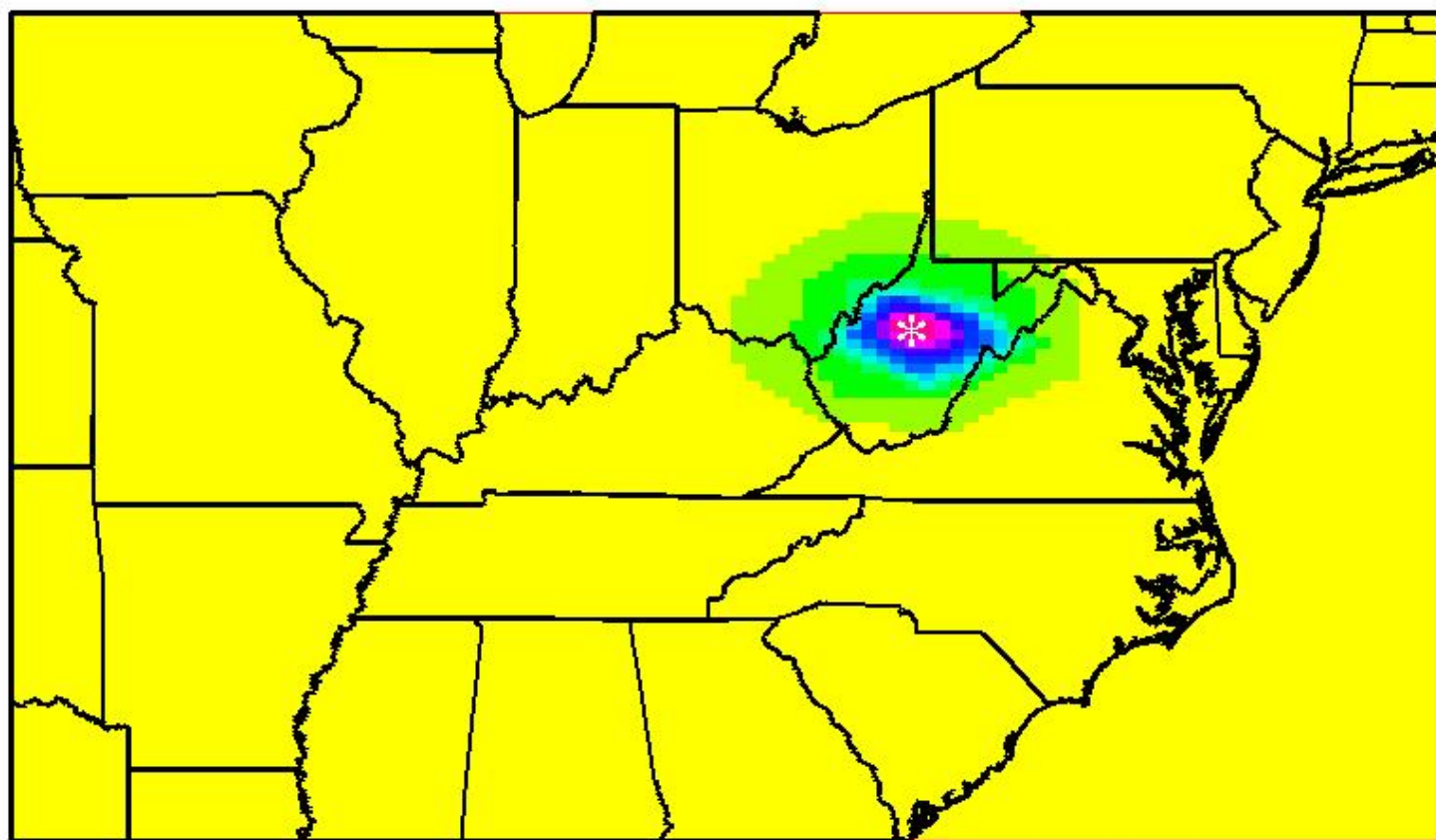


The graph shows the locations of the 513 sites where the ozone ambient concentrations are measured hourly.

# Covariance plot



# Covariance plot



## Background in spatial methods

$Z$  is *second-order stationary* process if the mean is constant  $\mu(\mathbf{s}) \equiv \mu$ , where  $E(Z(\mathbf{s})) = \mu(\mathbf{s})$ , and

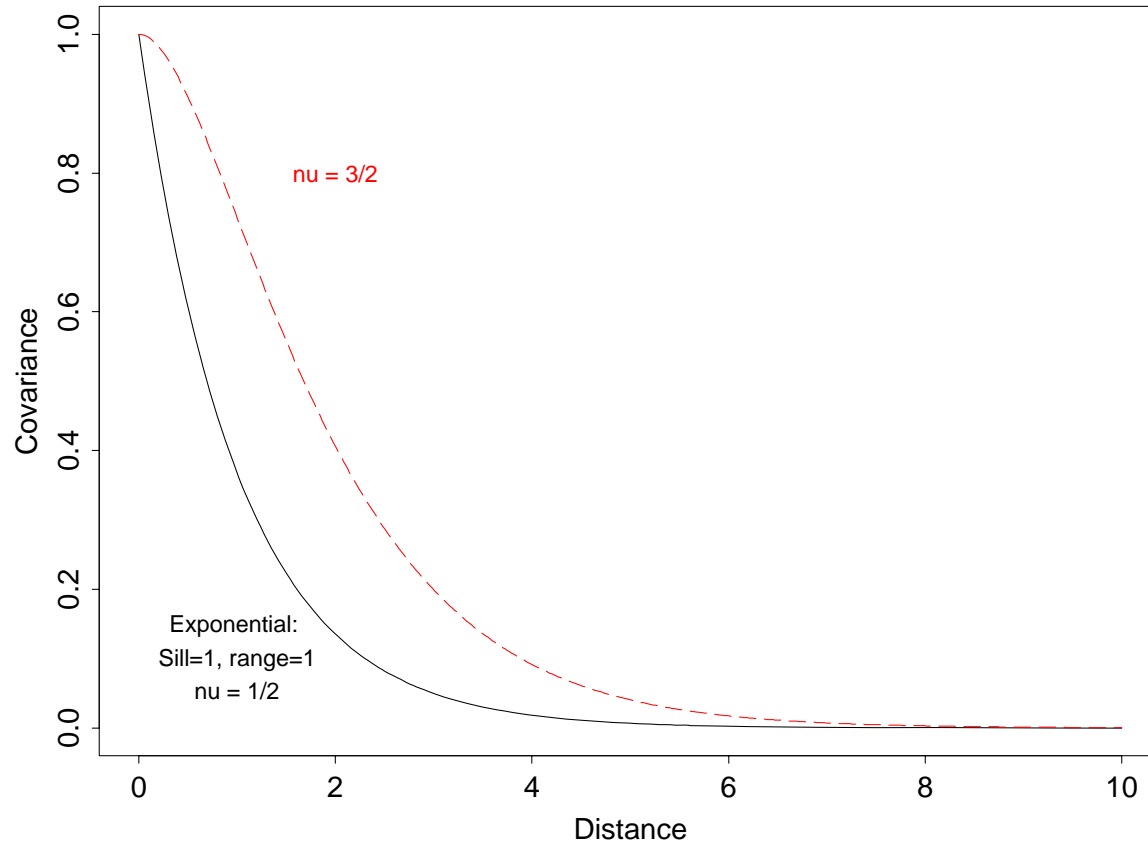
$$\text{cov}\{Z(\mathbf{s}_1), Z(\mathbf{s}_2)\} = C(\mathbf{s}_1 - \mathbf{s}_2)$$

where  $C(\mathbf{s})$  is the covariance function.

When  $C$  is only a function of distance (not direction), the process is called *isotropic*:

$$\text{cov}\{Z(\mathbf{s}_1), Z(\mathbf{s}_2)\} = C(\|\mathbf{s}_1 - \mathbf{s}_2\|)$$

## Matern Covariances



Covariances. Matérn Class for  $\nu = 1/2$  (exponential covariance) and  $\nu = 3/2$ .

## Background in spatial statistics

$C_{\boldsymbol{\theta}}(\mathbf{x})$  is a Matérn stationary (and isotropic) covariance:

$$C_{\boldsymbol{\theta}}(\mathbf{x}) = \frac{\sigma}{2^{\nu-1}\Gamma(\nu)} (2\nu^{1/2}|\mathbf{x}|/\rho)^{\nu} \mathcal{K}_{\nu}(2\nu^{1/2}|\mathbf{x}|/\rho),$$

where  $\mathcal{K}_{\nu}$  is a modified Bessel function and we have  $\boldsymbol{\theta} = (\nu, \sigma, \rho)$ . The parameter  $\rho$  measures how the correlation decays with distance, generally this parameter is called the range;  $\sigma$  is the variance of the random field, i.e.  $\sigma = \text{var}(Z_{\boldsymbol{\theta}}(\mathbf{s}))$ , the covariance parameter  $\sigma$  is usually referred to as the sill; and the parameter  $\nu$  measures the degree of smoothness of the process  $Z$ , the higher the value of  $\nu$  the smoother  $Z$  would be.

## Background in spatial statistics

### Separable models

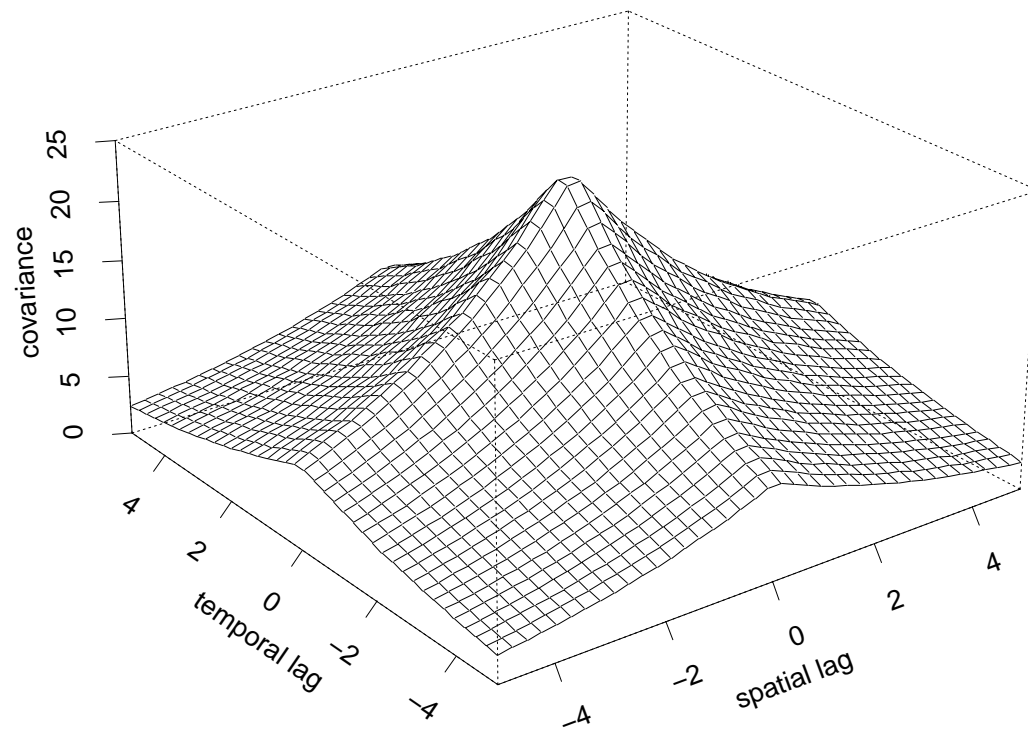
Many of the problems of space and time modeling can be overcome by using separable processes.

A spatial-temporal field  $Z(\mathbf{s}, t)$ , where  $\mathbf{s}$  represent space and  $t$  time, is separable if

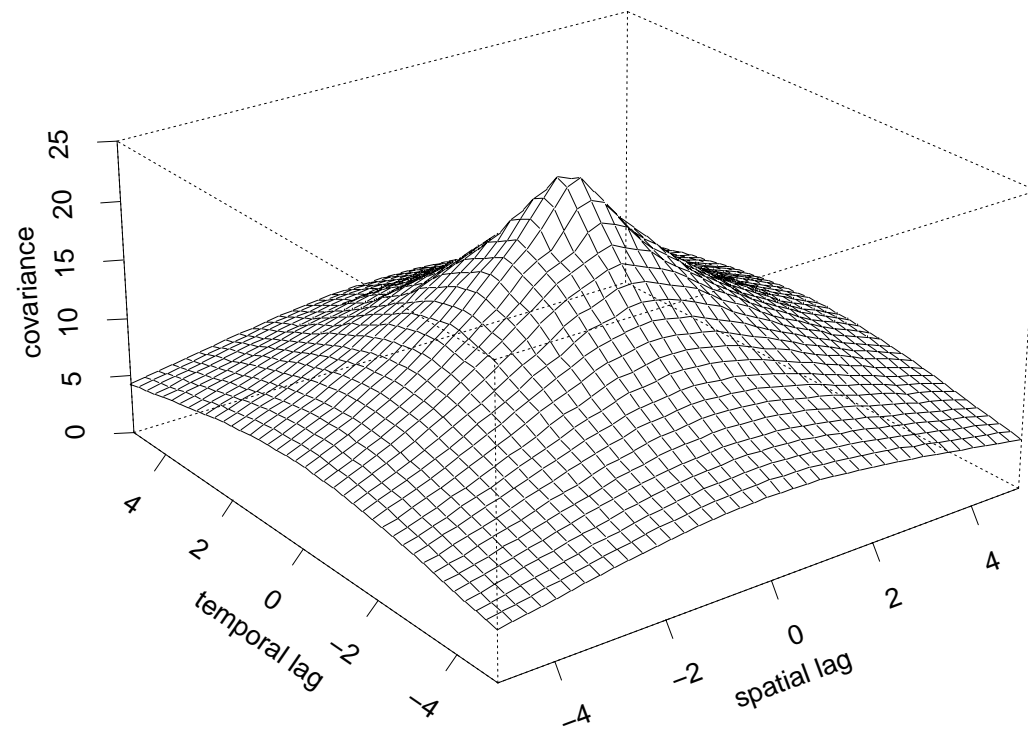
$$\text{Cov}\{Z(\mathbf{s}, t), Z(\mathbf{s}', t')\} = C_1(\mathbf{s}, \mathbf{s}')C_2(t, t')$$

for some spatial autocorrelation  $C_1$  and temporal autocorrelation  $C_2$ .

A separable covariance



A non-separable covariance

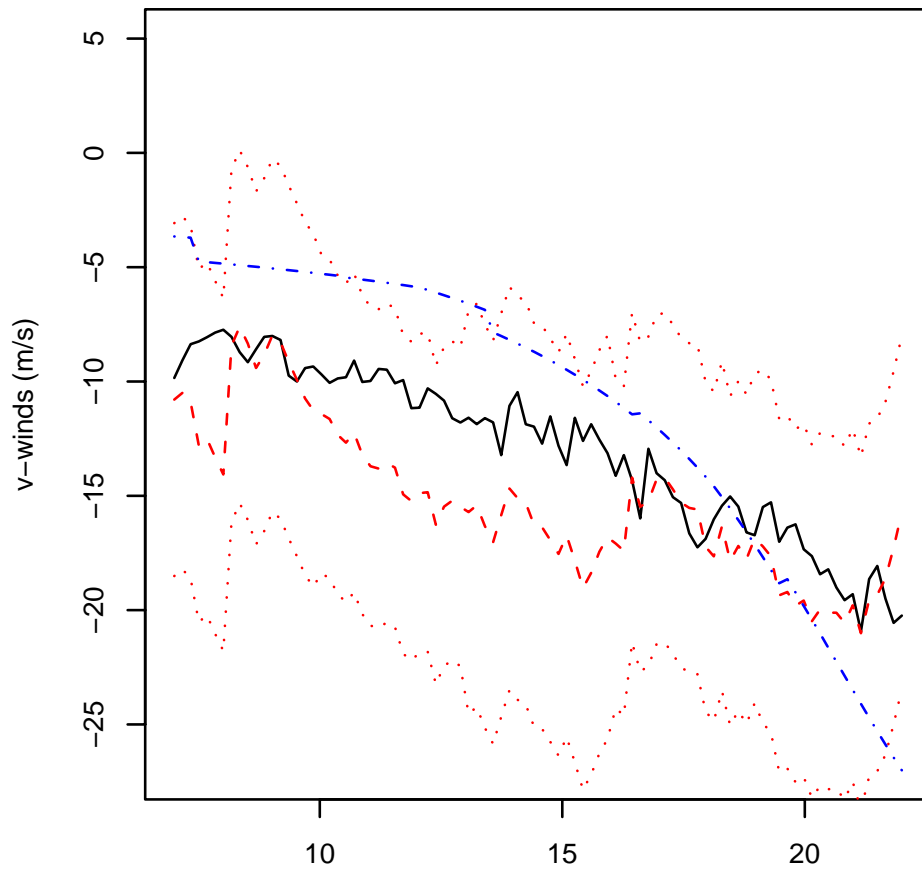


### Need for nonstationary models

Next example with hurricane Floyd, illustrates the need for nonstationary models, and the impact of assuming stationarity for wind fields.

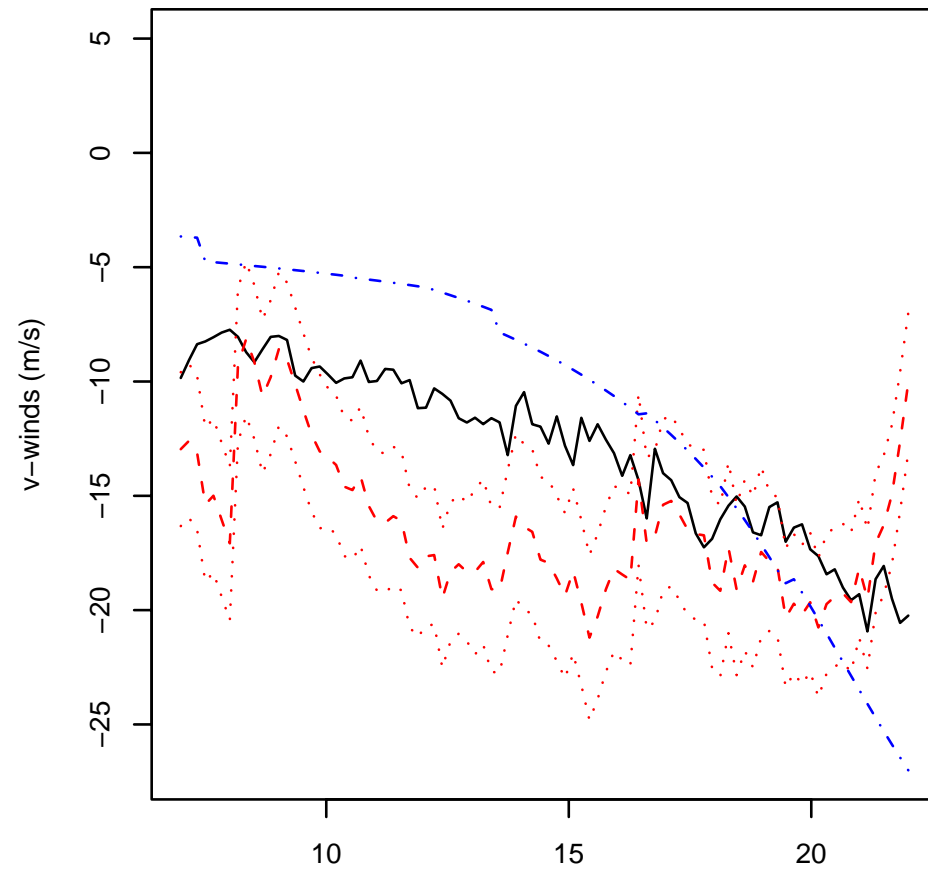
**Wind Prediction at 2 buoys for Hurricane Floyd** September 10-17 1999. Observed  $v$  winds in m/s at each buoy in solid black for September 15th 730 UTC to 2230 UTC. The dashed and dotted line is the Holland output based on posterior median values for  $B$  and  $R_{max}$ . The dashed lines are the predicted values and the dotted lines are 95% posterior predictive intervals using nonstationary and stationary LMC models.

Buoy FBIS1: Nonstationary Model



Sept. 15 1999, UTC hour

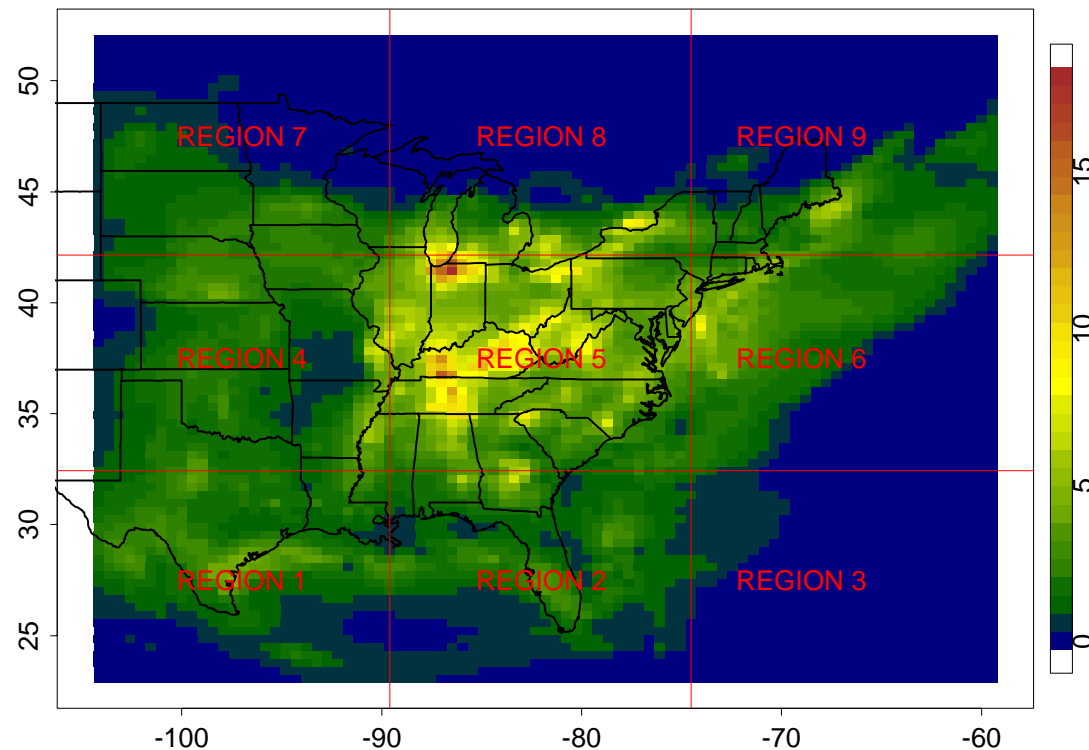
Buoy FBIS1: Stationary Model



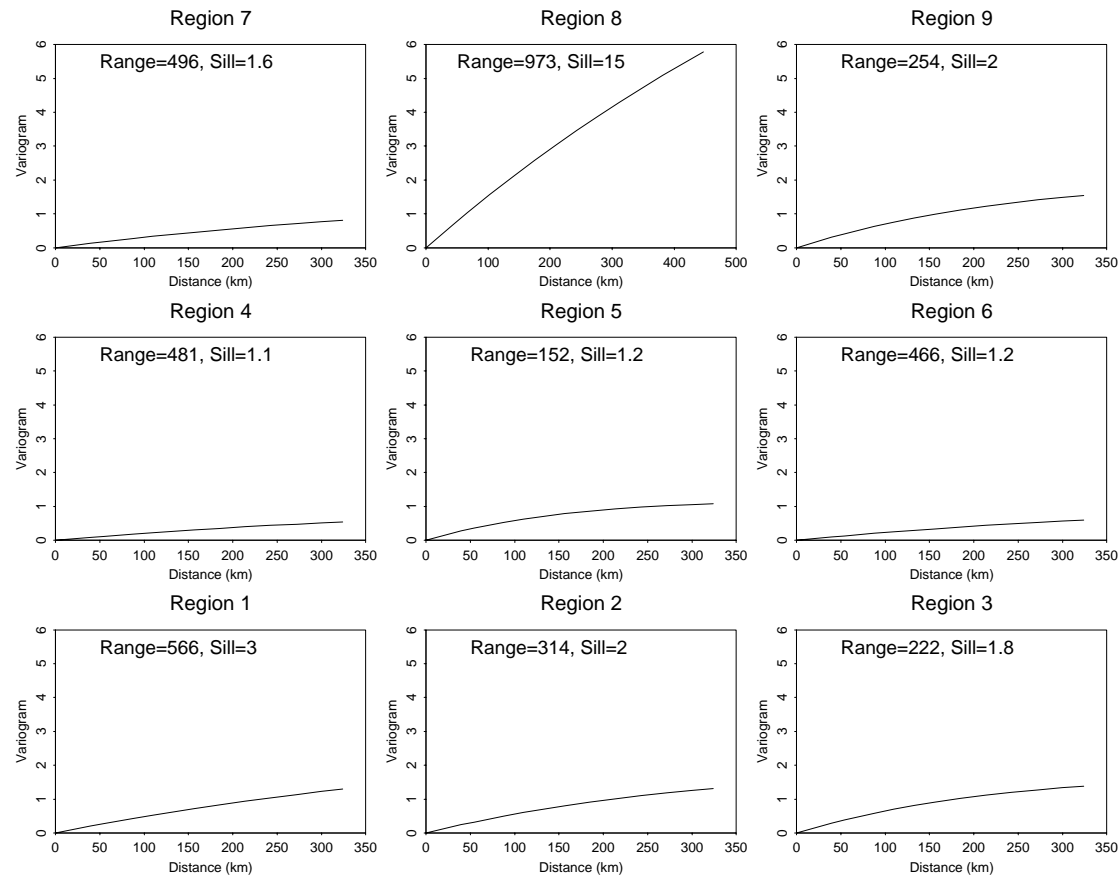
Sept. 15 1999, UTC hour

## The need for nonstationary models

Nitric Acid concentrations



This figure shows the output of Models-3 for the week starting July 11, 1995. We divide the domain in 9 subregions.



This figure shows the Matérn variograms  $\gamma_i$ , corresponding to the processes  $Z_i$ . **Note:**  $\gamma_i(\mathbf{x}) = C_i(\mathbf{0}) - C_i(\mathbf{x})$ , where  $C_i$  is the corresponding covariance.

In this chapter we consider a number of different approaches to processes which are not spatially stationary.

(a) **Deformation methods**, in which it is assumed that the process is stationary and isotropic only after some nonlinear deformation of the sampling space.

(b) **Moving-window methods**, in which the predictor or interpolator at a particular location, is based on a “window” of observations centered at that location.

(c) **EOF** and methods based on an eigenfunctions expansion of the covariance function.

(d) **Kernel-based methods**.

## Deformation methods

We consider a spatial process  $\{Z(s), s \in D\}$ , where  $D \subseteq \mathbb{R}^d$  is a domain of spatial locations. Usually  $d = 2$  or  $3$  (here  $d = 2$ ). Spatial dependence is usually characterized in terms of either the *covariance function*

$$C(s_1, s_2) = \text{Cov} \{Z(s_1), Z(s_2)\}, \quad s_1, s_2 \in D,$$

or the *dispersion*

$$D(s_1, s_2) = \text{Var} \{Z(s_1) - Z(s_2)\}, \quad s_1, s_2 \in D.$$

Much of the literature is concerned with processes which satisfy some or all of

(i) *intrinsic stationarity*:  $D(s_1, s_2)$  depends on  $s_1$  and  $s_2$  only through the (vector) difference  $s_1 - s_2$ ,

(ii) *stationarity*:  $C(s_1, s_2)$  depends on  $s_1$  and  $s_2$  only through  $s_1 - s_2$  (this implies intrinsic stationarity, but not conversely),

(iii) *isotropy*:  $D(s_1, s_2)$  or  $C(s_1, s_2)$  depends only on  $\|s_1 - s_2\|$  (the Euclidean norm of  $s_1 - s_2$  or, equivalently, the Euclidean distance between the locations  $s_1$  and  $s_2$ ). In this case we often write  $D(s_1, s_2) = 2\gamma_0(\|s_1 - s_2\|)$ , where  $\gamma_0$  is an isotropic *semivariogram* function.

When all of (i)–(iii) hold we shall call the process *homogeneous*.

Classical geostatistics is concerned primarily with homogeneous processes for which, by now, a very extensive literature exists, e.g. Cressie (1993). Until recently, however, not much was known about the modeling of inhomogeneous processes. One old approach (Journel and Huijbregts 1978) for stationary, non-isotropic processes is to write the dispersion in the form

$$D(s_1, s_2) = 2\gamma_0(\|A_0(s_1 - s_2)\|)$$

or by extension

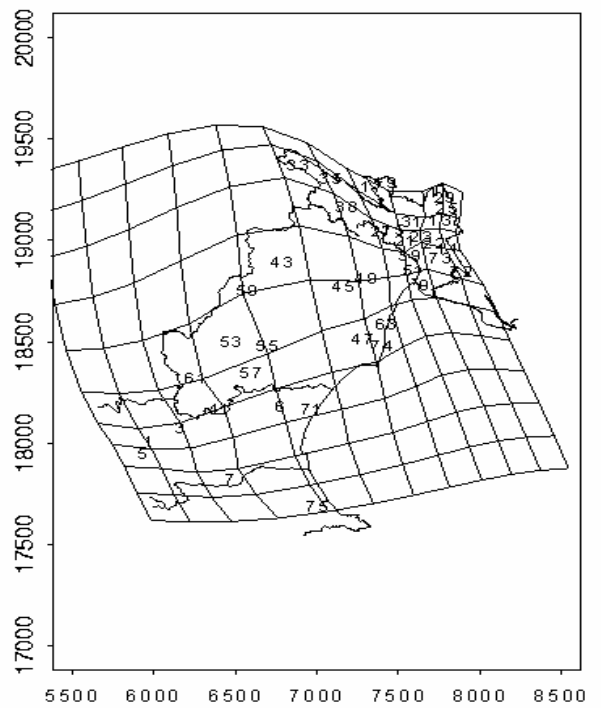
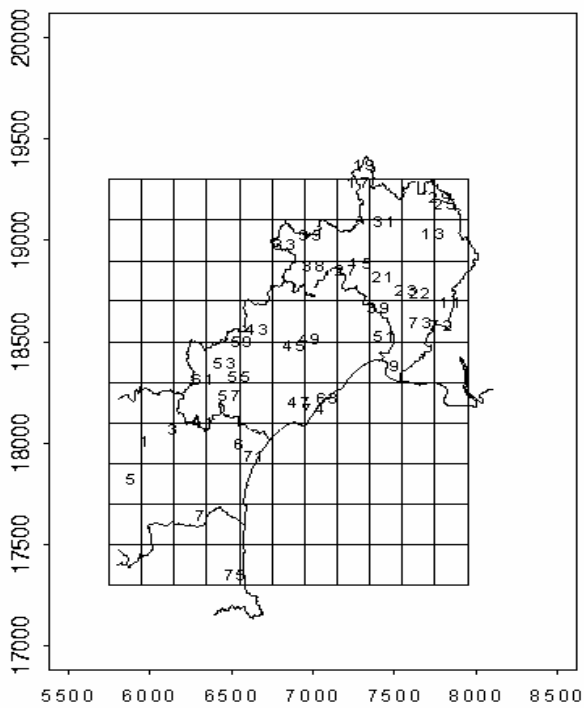
$$D(s_1, s_2) = 2 \sum_{j=0}^{J-1} \gamma_j(\|A_j(s_1 - s_2)\|).$$

Here  $A_0, A_1, \dots$ , are arbitrary matrices and  $\gamma_0, \gamma_1, \dots$ , isotropic semivariogram functions. However, this is still quite a restrictive class of models.

## Application to Languedoc-Roussillon Precipitation Data

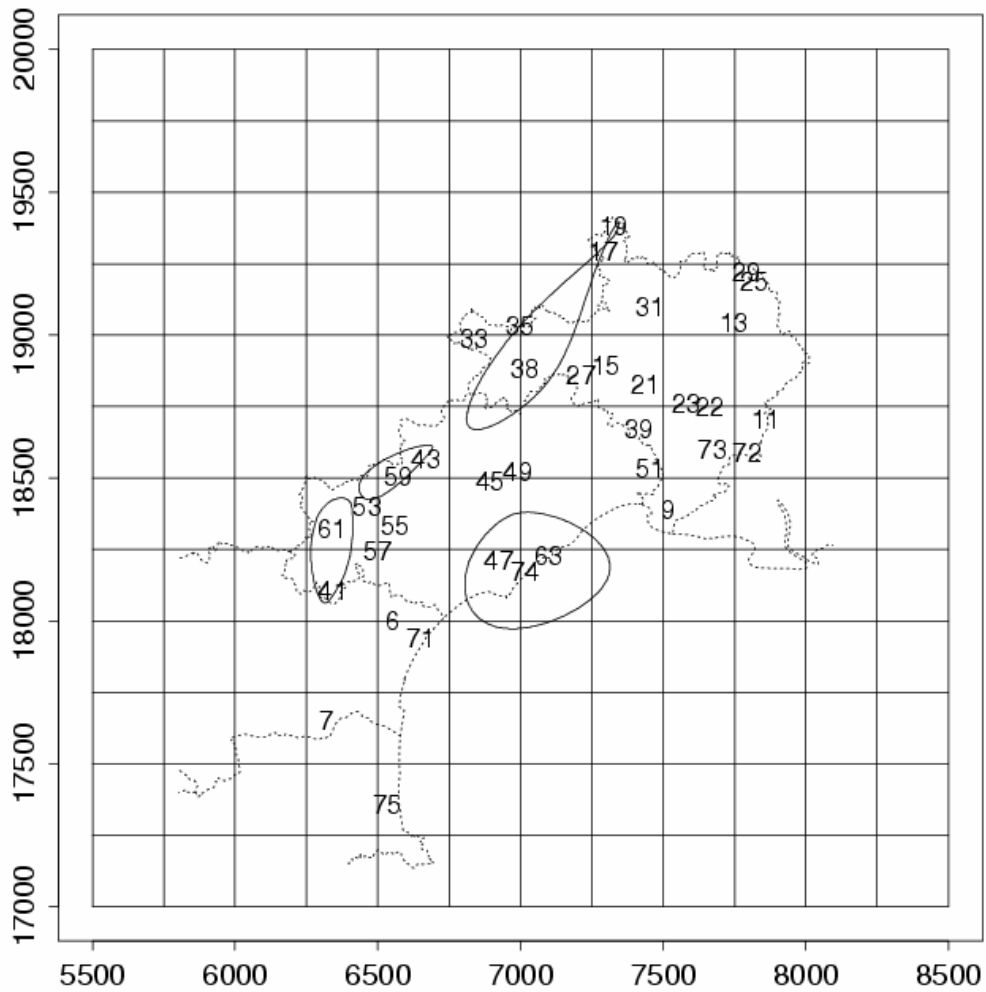
- 108 altitude-adjusted, 10-day aggregate precip records at 39 sites (Nov-Dec, 1975-1992).
- Data log-transformed and site-specific means removed (for this analysis).
- Estimated deformation is non-linear: correlation stronger in the NE region, weaker in the SW.

## Precipitation in Southern France - an example of a non-linear deformation



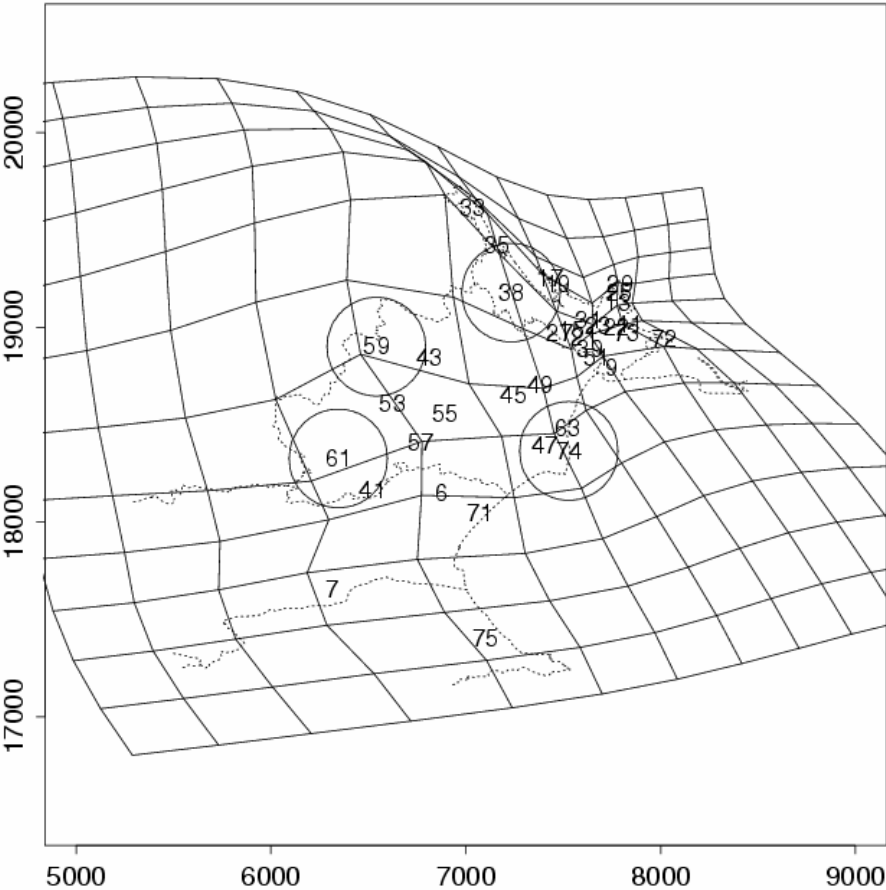
# G-plane Equicorrelation Contours

Equi-Correlation (0.9) Contours around 4 points (G-Plane)



# D-plane Equicorrelation Contours

Equi-Correlation (0.9) Contours around 4 points (D-Plane)



A much more radical extension has been proposed in a series of papers by Sampson and Guttorp — see, in particular, Sampson and Guttorp (1992). They considered models of the form

$$D(s_1, s_2) = 2\gamma_0(f(s_1), f(s_2))$$

with  $\gamma_0$  again an isotropic semivariogram and  $f$  a smooth nonlinear map from  $\mathbb{R}^d$  to  $\mathbb{R}^{d'}$ . In principle one may permit  $d' \neq d$  though in most of the Sampson-Guttorp work it is assumed that  $d' = d$  and we shall continue to assume that here.

The idea is that the map  $f$  takes the coordinates from the real, geographical or “G” space, into an alternative dispersion or “D” space in which the process is **homogeneous**. This approach may not be universally applicable to inhomogeneous processes.

The precise original methodology used by Sampson and Guttorp contained a number of rather *ad hoc* features. Briefly, it consists of three stages:

(a) A mapping of the  $n$  sampling points from the G space into the D space is found to minimize a stress criterion

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

where  $d_{ij}$  is the observed dispersion between sites  $i$  and  $j$ ,  $h_{ij}$  is the distance between sites  $i$  and  $j$  in D space and the minimization is taken over all monotonically increasing functions  $\delta$ . This formulation of the problem permits it to be solved by a [multidimensional scaling](#) (MDS) algorithm.

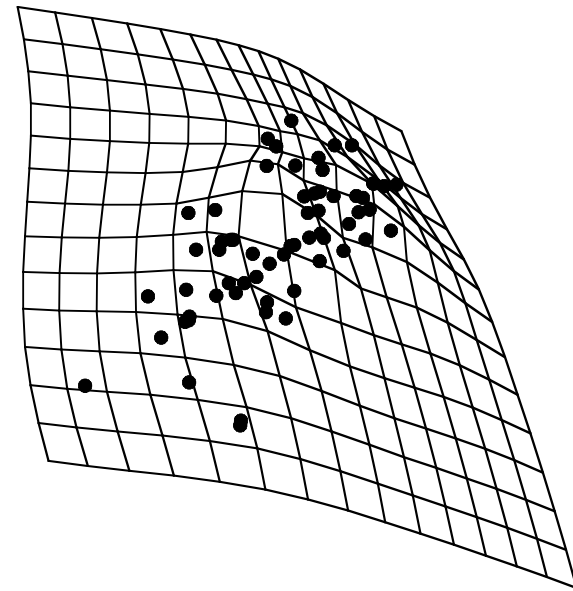
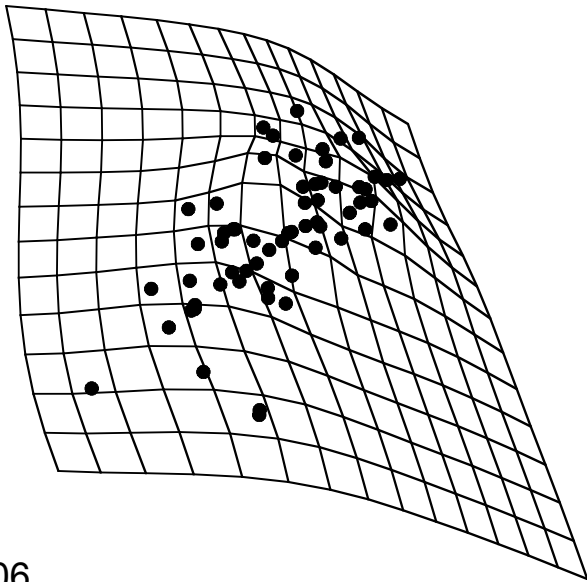
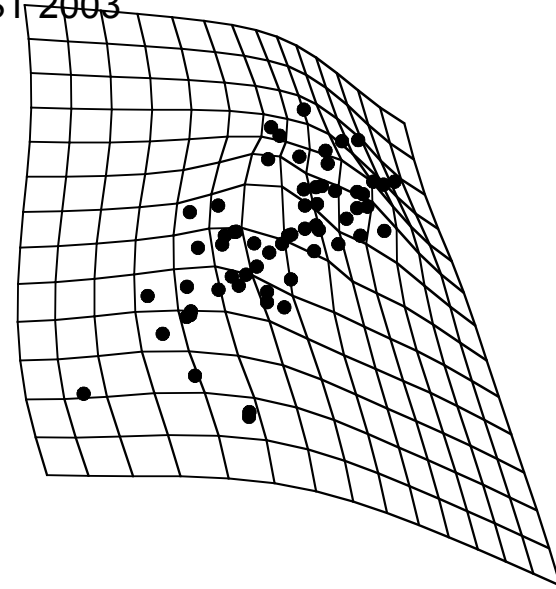
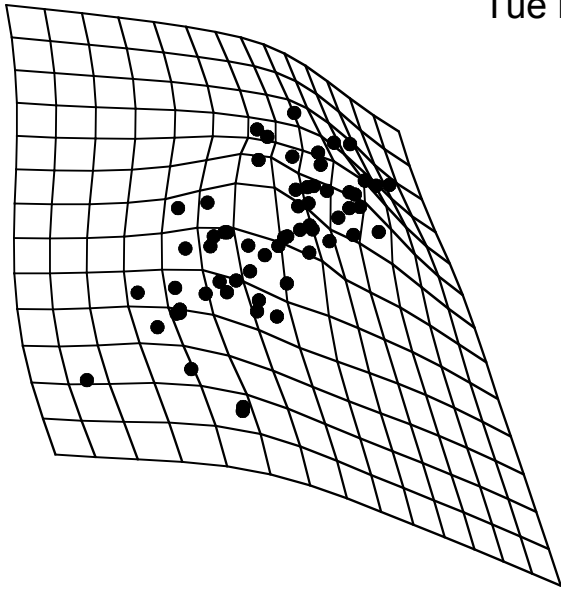
(b) The mapping of the  $N$  sampling points is then extended to a *smooth* function from the entire  $G$  space into the  $D$  space, using a representation based on [thin plate splines](#).

(c) The function  $\delta$  is replaced by a smooth function  $g$  (so  $d_{ij} \approx g(h_{ij})$ ), which satisfies the [positive definiteness condition](#) required for  $g$  to be the variogram of a homogeneous process. For this purpose, Sampson and Guttorp used a very general representation of  $g$  as a mixture of Gaussian-type variograms.

The Sampson-Guttorp approach implies some [restrictions](#) on the models considered. In particular, by using MDS to model the locations so that increasing distances correspond to increasing dispersions, the possibility that  $g$  may be non-monotone is excluded.

[Maximum likelihood](#) versions of the method were developed by Mardia and Goodall (1993) and Smith (1996). There are also two recent [Bayesian approaches](#) due to Damian *et al.* (2001) and Schmidt and O'Hagan (2000).

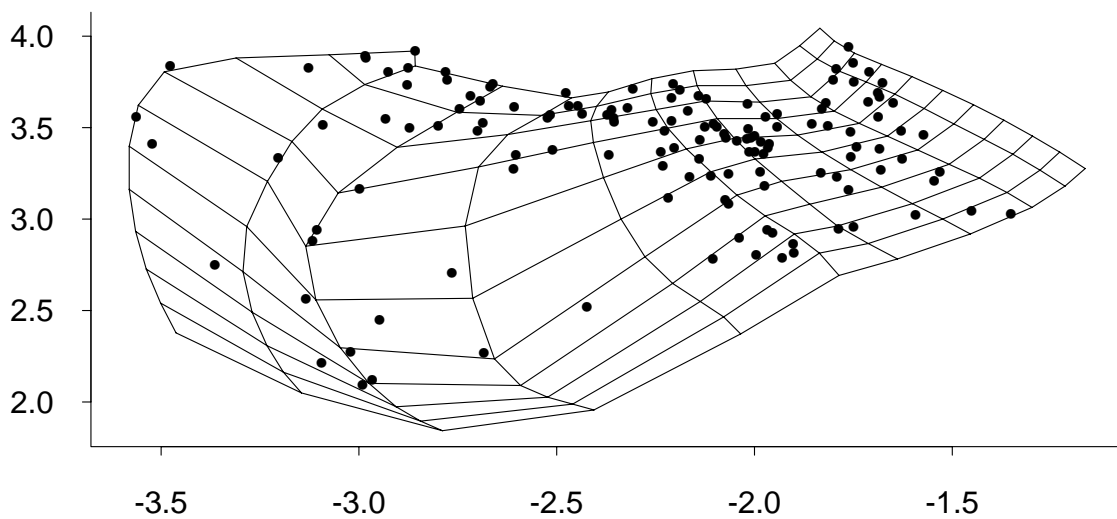
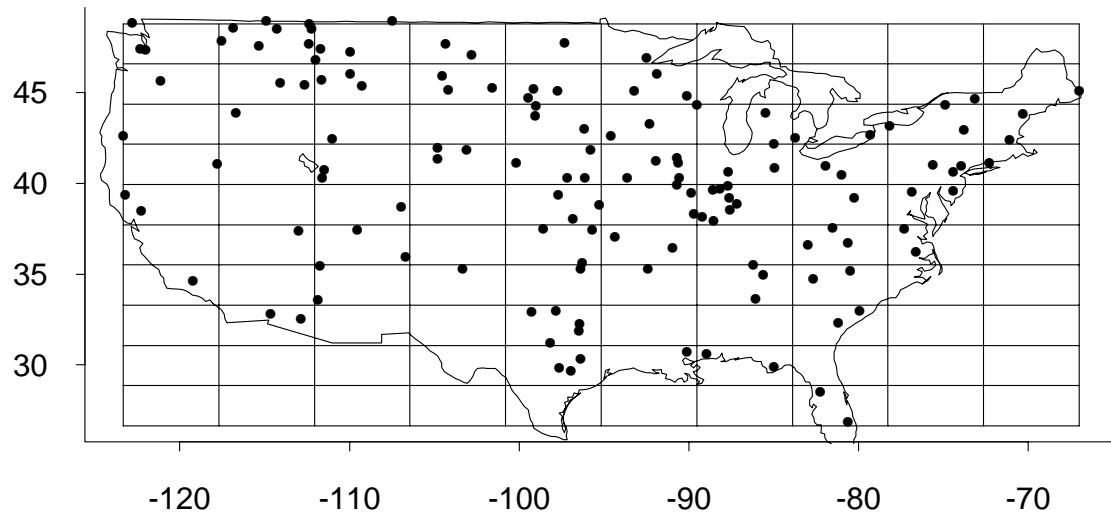
N=62, Northeast: 4 samples from the posterior distribution of deformations reflecting spatial covariance  
Tue Feb 04 18:34:34 PST 2003



1/31/06

Annual average temperature measurements from 138 HCN (40 years)

## Original and Transformed Maps



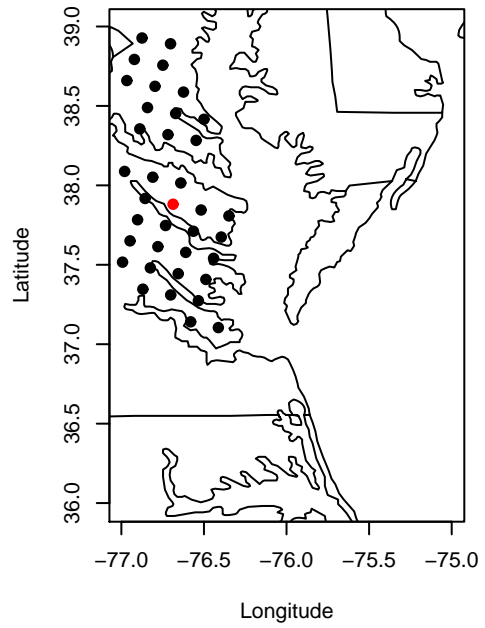
Richard Smith (1996)

## Moving-Window Approaches

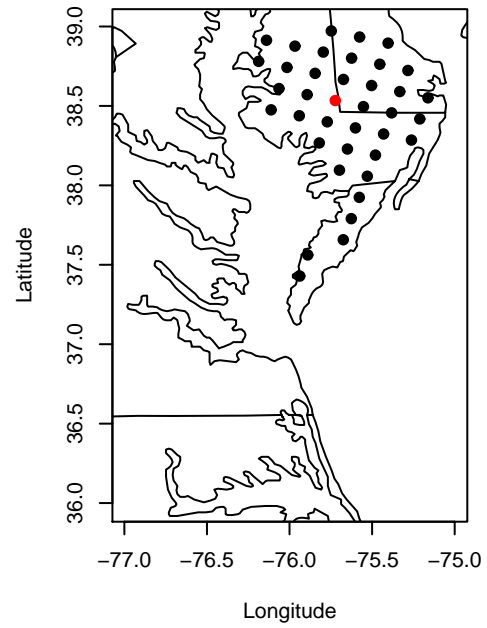
The idea of a moving-window approach is that to fit a spatial model and to perform kriging at a sampling location  $s$ , we should restrict ourselves to a “**window**” of sampling stations close to  $s$ , within which it is reasonable to assume a **homogeneous** model. Thus the method retains all the mathematical techniques of homogeneous processes, while not assuming that homogeneity applies across the whole sampling region.

For the present description we will follow Haas (1995), who develops the method in the context of spatio-temporal processes.

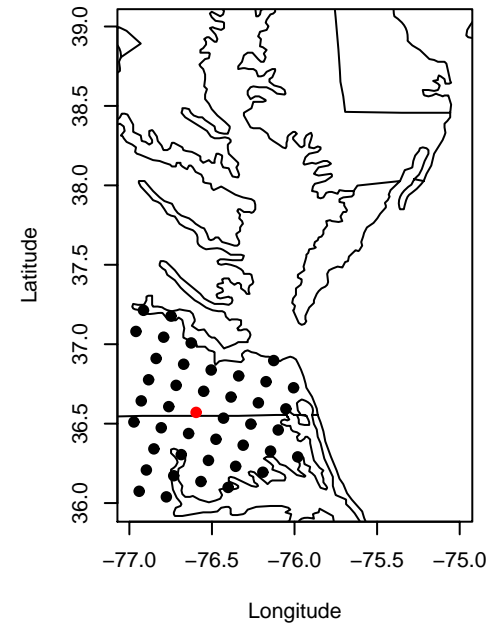
**subregion 1**



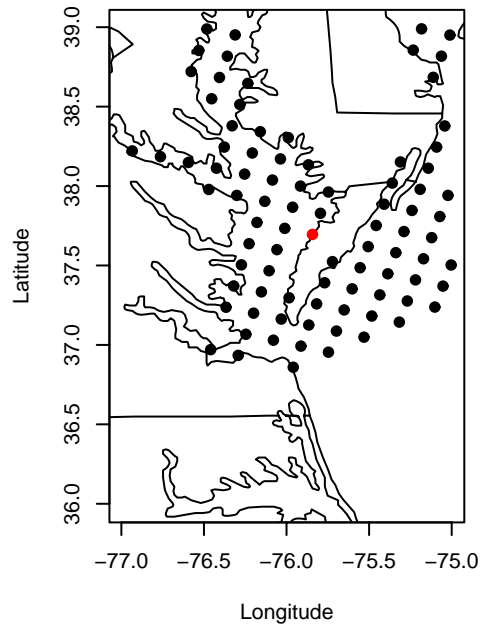
**subregion 2**



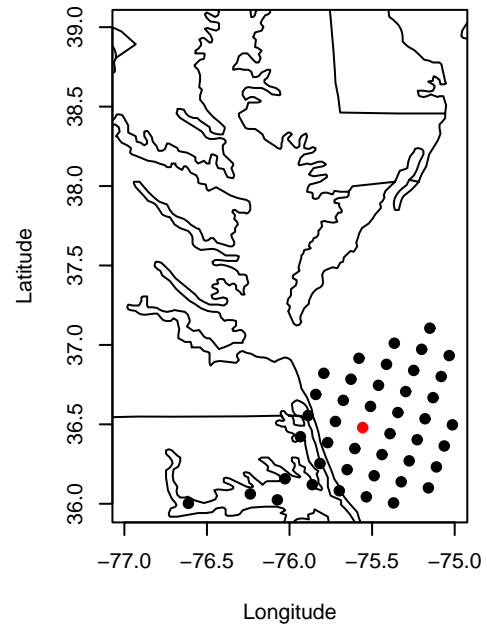
**subregion 3**



**subregion 4**



**subregion 5**



Suppose we have spatio-temporal data  $Z(t, s)$  where  $t$  denotes time and  $s$  denotes space. Specifically, we have a sample  $\{Z(t_i, s_i)\}$  at  $n$  time-space points  $\{(t_i, s_i), 1 \leq i \leq n\}$ . In most environmental applications this will consist of a fixed time series of observations at each measuring station  $s_i$ , but this format is not required for the methodology.

- The method requires the specification of two parameters, the *time window*  $m_T$  and the *sampling fraction*  $f_c$ .
- Once these parameters are specified, the window is defined as follows. Suppose we want to predict or interpolate at a specific time  $t_0$  and location  $s_0$ . Restrict the observations to those which lie within the time window  $(t_0 - \frac{m_T}{2}, t_0 + \frac{m_T}{2})$ . Within that window, pick out observations in order of space, i.e. first select all the observations at the spatial location closest to  $s_0$ , then those at the location second closest to  $s_0$ , and so on until a fixed number  $n_c = n f_c$  of observations has been selected.

**Prediction** at  $(t_0, s_0)$  will be based entirely on this group of  $n_c$  observations.

- The next step is to consider the form of regression model suitable for both the mean and standard deviation of  $Z$ . Haas considered a general model of the form

$$Z(t, s) = \mu(t, s, \beta) + \psi(\mu(t, s, \beta))R(t, s)$$

in terms of additional functions  $\mu$  and  $\psi$ , where  $\mu$  is typically a regression function of covariates such as meteorology, in terms of additional parameters  $\beta$ , which are also estimated separately within each window.

- For model fitting and kriging, it is necessary to specify a suitable spatio-temporal **corelation structure** for the residual process  $R(s, t)$ , restricted to the given window. The basic covariance model assumed by Haas is

$$C\{R(t_1, s_1), R(t_2, s_2)\} = C_T(t_2 - t_1)C_S(s_2 - s_1) \quad (1)$$

where  $C_T$  denotes the temporal covariance function and  $C_S$  the spatial covariance function, each of which has been assumed stationary within the window. For the functions  $C_T$  and  $C_S$ , he assumed the “spherical” form of isotropic covariance structure. The same form of covariance is assumed for both the spatial and temporal scale, though of course the parameters may be quite different for the two functions.

The product form in which the spatio-temporal covariance function is written as a product of a function of space and a function of time, is known as the *separability* assumption and is widely discussed in the context of time-space processes. It is an assumption which is very widely used because of its convenience, though it is often criticized as unrealistic when applied to actual time-space data.

Once the model functions for  $\mu$ ,  $\psi$ ,  $C_T$  and  $C_S$  are parametrically specified, under an assumption of joint normality, we could estimate the model by [maximum likelihood](#). Haas avoided this but instead described an algorithm including first OLS and later GLS regression to estimate the parameters of  $\mu$  and  $\psi$ , along with the approximate WLS procedure to estimate the parameters of  $C_T$  and  $C_S$ .

- Finally, once the model is fitted, [kriging](#) is used to calculate an optimal predictor at  $(t_0, s_0)$ , say  $Z^*(t_0, s_0)$ , and its prediction standard error.

*Reconstructing the full covariance matrix.*

One disadvantage of the woving-window approach is that it does not lead to a single [model](#) to describe the whole data set. For example, different covariance functions are fitted to different portions of the data set, and if we simply combine these together to form an overall estimated covariance matrix, the result may not be positive definite.

Though Haas (1996) suggested one way to get a [positive definite estimate](#) of the full spatial covariance by finding the nearest (in some metric) positive definite matrix to the spatial covariance matrix estimated from the moving window approach, this is not very intuitive and does not correspond to a model for the continuous underlying field.

## The EOF method

**EOF:** Empirical orthogonal functions.

A very classical approach to spatial statistics, dating back at least as far as Cohen and Jones (1969), is to represent a spatial field in terms of the Karhunen-Loève expansion of its covariance function. This leads to representations of the form:

$$Z(\mathbf{s}) = \sum_{\nu=1} a_{\nu} \lambda_{\nu}^{1/2} \psi_{\nu}(\mathbf{s})$$

where  $\{\psi_{\nu}\}$  are a fixed basis of orthogonal functions,  $\{\lambda_{\nu}\}$  are coefficients to be estimated, and  $\{a_{\nu}\}$  are independent standard normal RVs. Models of this form have become very widely used in geophysical sciences (see, e.g., Creutin and Obled (1982)).

The covariance of  $Z$  would be of the form:

$$C(\mathbf{s}, \mathbf{y}) = \sum_{\nu=1} \lambda_{\nu} \psi_{\nu}(\mathbf{s}) \psi_{\nu}(\mathbf{y}).$$

The corresponding approximation to the covariance  $C$  of  $Z$  corresponds to an empirical orthogonal function (EOF) decomposition ([principal component analysis](#)) of the covariance matrix of the observations (we truncate the sum at  $\nu = M$ ). This covariance (or variogram) is calculated with observations assumed to be realizations of the spatial process replicated over time.

The estimated coefficient  $\lambda_{\nu}$ , is also called the  $\nu$ -th (spatial) principal component, is the orthogonal projection of the original process at location  $\mathbf{s}$  onto the  $\nu$ -th spatial EOF vector  $\hat{\psi}_{\nu}(\mathbf{s})$ .

We need replications over time to use this EOF method.

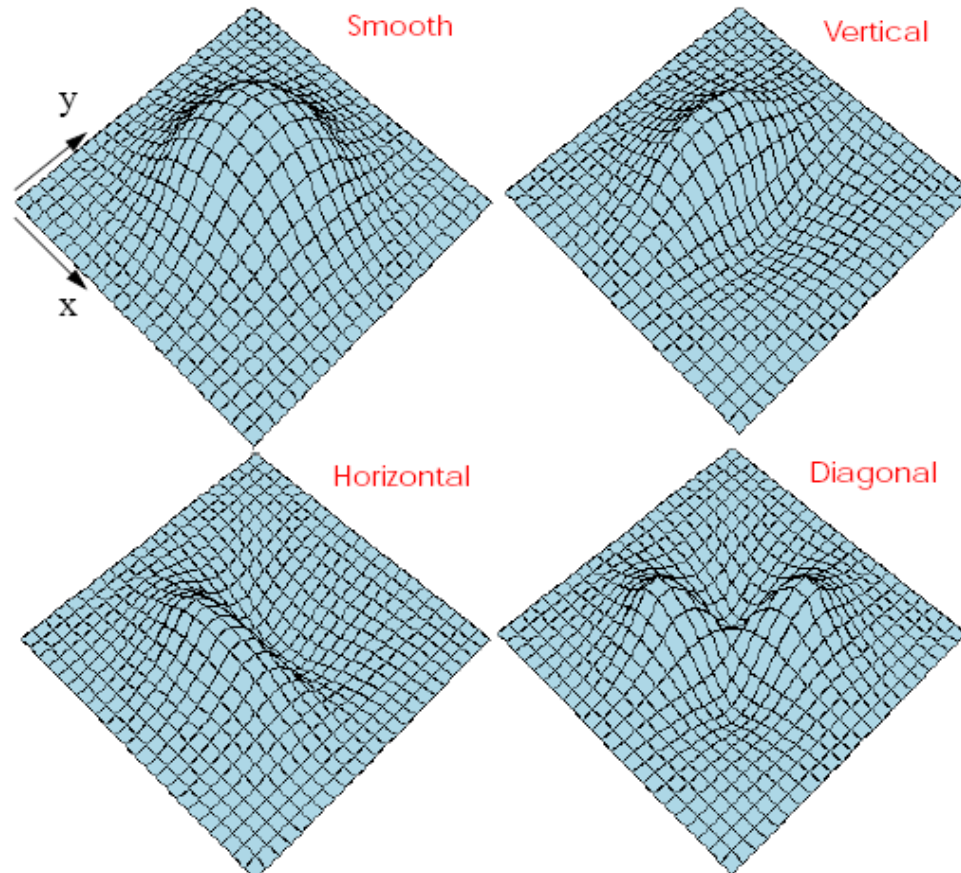
Nychka *et al.* (2002) have recently proposed models of the EOF-form in which  $\{\psi_\nu\}$  are replaced by [wavelet basis](#) functions. The wavelet representation is motivated by nonstationarity and they also emphasize the computational applicability of the approach in very large systems.

There is also the possibility of a mixture of the two kinds of models (Nychka and Saltzman (1998), Holland *et al.* (1999)) based on representations of the form

$$Z(\mathbf{s}) = \sigma(\mathbf{s}) \left\{ \rho^{1/2} Z_0(\mathbf{s}) + \sum_{\nu=1}^M a_\nu \lambda_\nu^{1/2} \psi_\nu(\mathbf{s}) \right\}$$

in which  $Z_0(\mathbf{s})$  is a stationary isotropic process,  $\rho$  is a positive constant and  $\sigma(\mathbf{s})$  is a scaling function.

A 2-d wavelet basis is obtained by repeated translations and scalings of four 2-D functions: **Smooth**, **Horizontal**, **Vertical** and **Diagonal** (presentation by Nychka).



## 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.

Then, the covariance of the process is  $C(h) = \int K(u - h)K(u)du$ .

For a Gaussian kernel  $K(u) \propto \exp(-\frac{1}{2}\|u\|^2)$ . Then, the covariance of the process is  $C(u) \propto \exp(-\frac{1}{4}\|u\|^2)$ .

## Extension to nonstationary processes

Higdon, Swall and Kern (1999) considered extensions of the form

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

where the kernel  $K_{\mathbf{s}}$  depends on position  $\mathbf{s}$ . The idea is to model  $K_{\mathbf{s}}(\mathbf{u})$  as an unknown function in terms of specific parameters which can then be estimated in a Bayes framework.

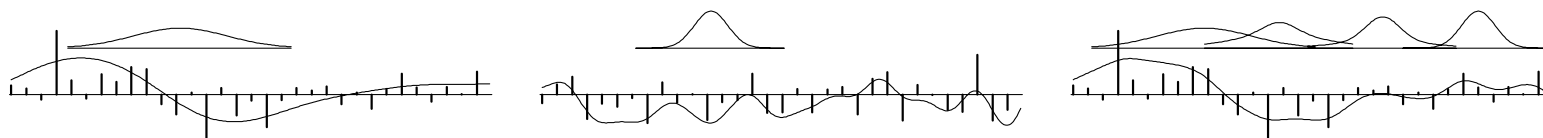


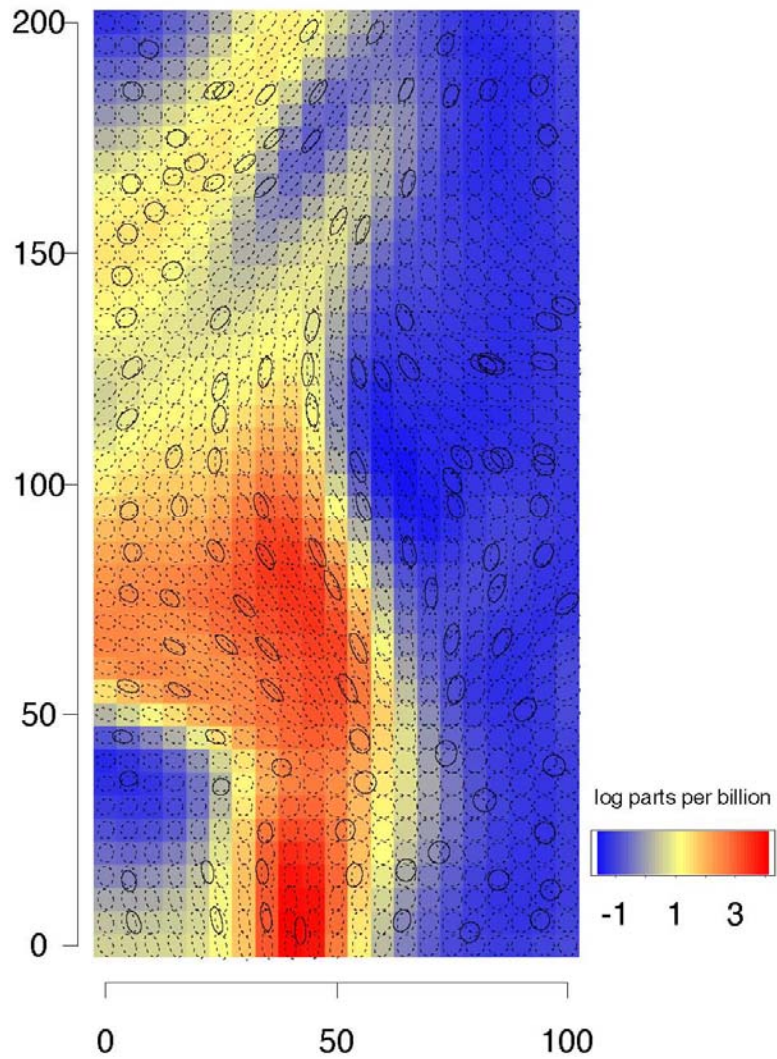
Figure 3: Convolution approach (Higdon) with different kernel functions.

Then, the covariance of the process is  $C(u, v) = \int K_s(u)K_s(v)ds$ .

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.

Gaussian kernel  $K_{\mathbf{s}}(\mathbf{u}) \propto \exp\left(-\frac{1}{2}\mathbf{u}^T \sigma_{\mathbf{s}}^{-1} \mathbf{u}\right)$ .

**Swall & Higdon. Process convolution approach,  
Posterior mean and covariance kernel ellipses.**



## A new model for nonstationarity

Consider a Gaussian spatial process  $Z(x)$ . We represent  $Z$  as a convolution of local stationary processes (Fuentes (2001), Fuentes (2002), and Fuentes and Smith (2001)):

$$Z(x) = \int_D K(x - s) Z_{\boldsymbol{\theta}(s)}(x) ds.$$

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

The covariance  $C(\mathbf{s}_1, \mathbf{s}_2; \boldsymbol{\theta})$  of  $Z$  is a convolution of the local covariances  $C_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{s}_1 - \mathbf{s}_2)$ ,

$$C(\mathbf{s}_1, \mathbf{s}_2; \boldsymbol{\theta}) = \int_D K(\mathbf{s}_1 - \mathbf{s})K(\mathbf{s}_2 - \mathbf{s})C_{\boldsymbol{\theta}(\mathbf{s})}(\mathbf{s}_1 - \mathbf{s}_2)ds.$$

If  $K$  is a sharply peaked kernel function and  $\boldsymbol{\theta}(\mathbf{s})$  varies slowly with  $s$ , 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 there is no problem about it being a well-defined process with a positive definite covariance function.

## DISCRETE VERSION:

Suppose  $\theta(\mathbf{s})$  is constant within subregions of stationarity  $S_i$ . Then, the nonstationary process  $Z$  observed on a region  $D$  is a **MIXTURE** of orthogonal local stationary processes.

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

where  $S_1, \dots, S_k$  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  $S_i$ .

The value of  $k$  (number of subregions) is chosen using a BIC or AIC criteria.

The nonstationary covariance of  $Z$  is defined in terms of the local stationary covariances of the processes  $Z_i$  for  $i = 1, \dots, k$ ,

$$\text{cov}(Z(\mathbf{x}), Z(\mathbf{y})) = \sum_{i=1}^k w_i(\mathbf{x})w_i(\mathbf{y})\text{cov}(Z_i(\mathbf{x}), Z_i(\mathbf{y}))$$

this is a valid nonstationary covariance.

The covariance parameters can be estimated with the MLEs or using a Bayesian approach. We use the estimated covariance for prediction (kriging).

Example of nonstationary kriging (Fuentes, 2002b)

LOCATION OF THE SITES

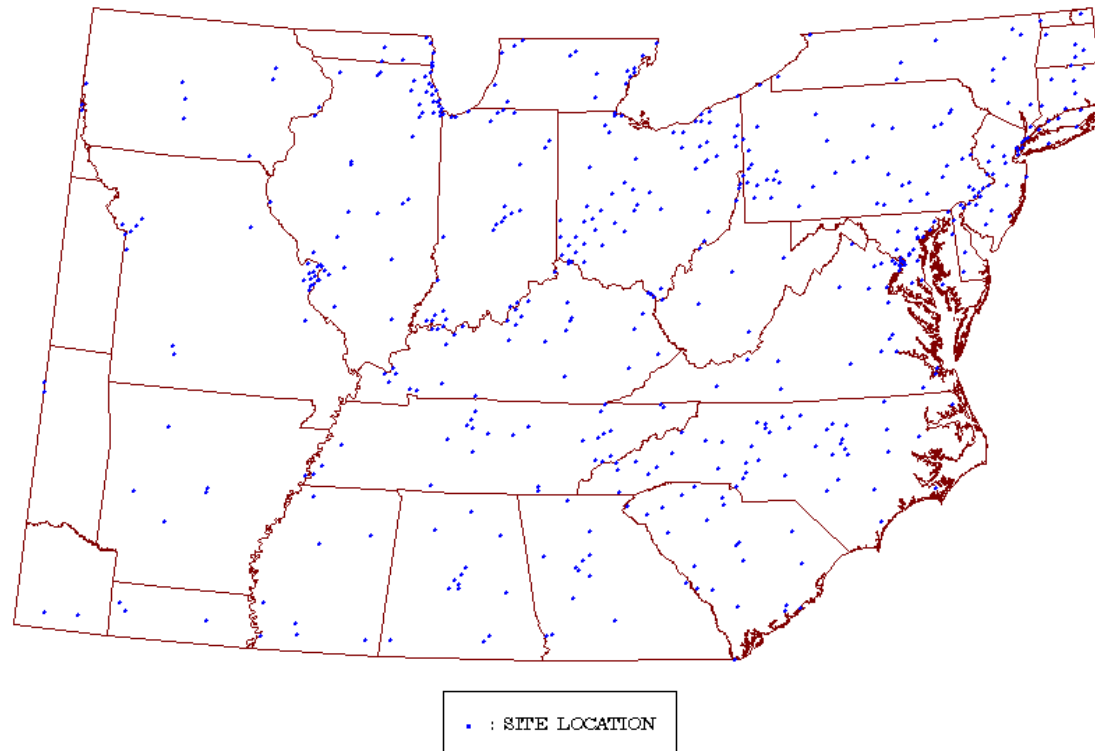


Figure 4: The graph shows the locations of the 513 sites where the ozone ambient concentrations are measured hourly.

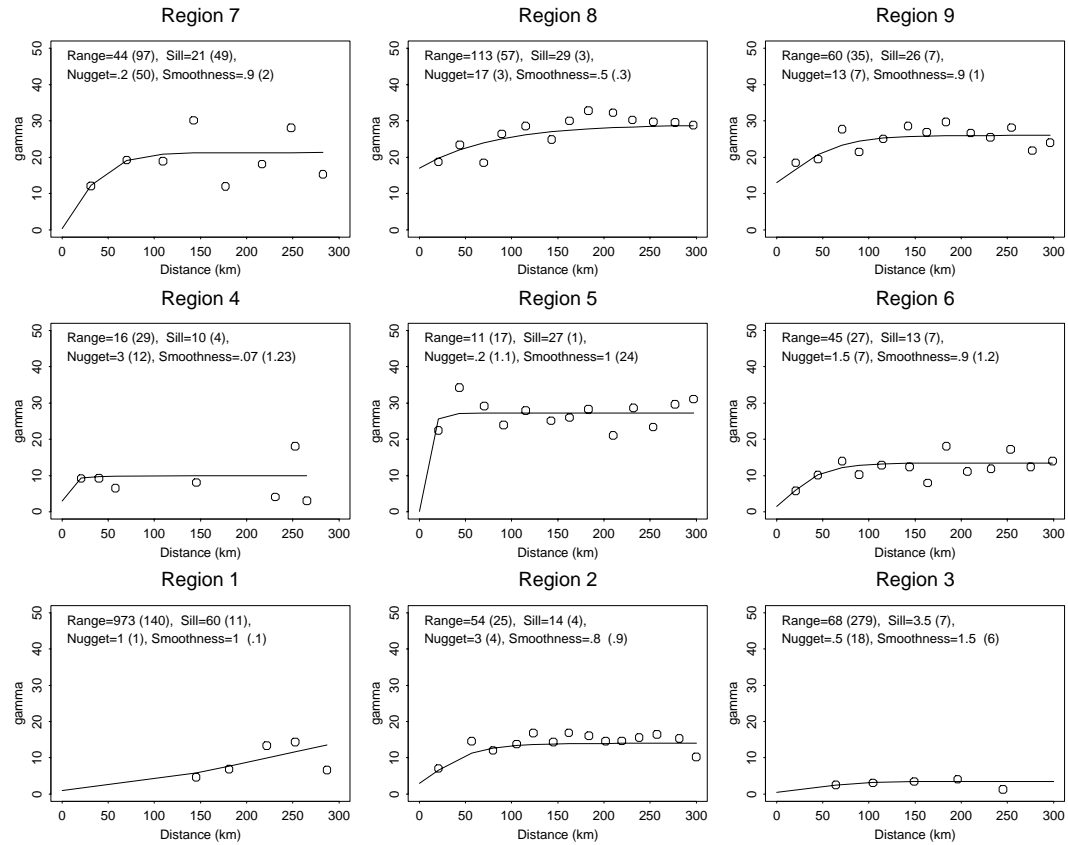


Figure 5: Each graph shows for each subregion, the empirical semivariogram (circles) and the likelihood estimate (solid line) based on the covariance model,  $\text{cov}(Z(\mathbf{x}), Z(\mathbf{y})) = \sum_{i=1}^9 K(\mathbf{x} - \mathbf{s}_i)K(\mathbf{y} - \mathbf{s}_i)C_{\theta(\mathbf{s}_i)}(\mathbf{x} - \mathbf{y})$ .

### Ozone Standard (years 95-97)

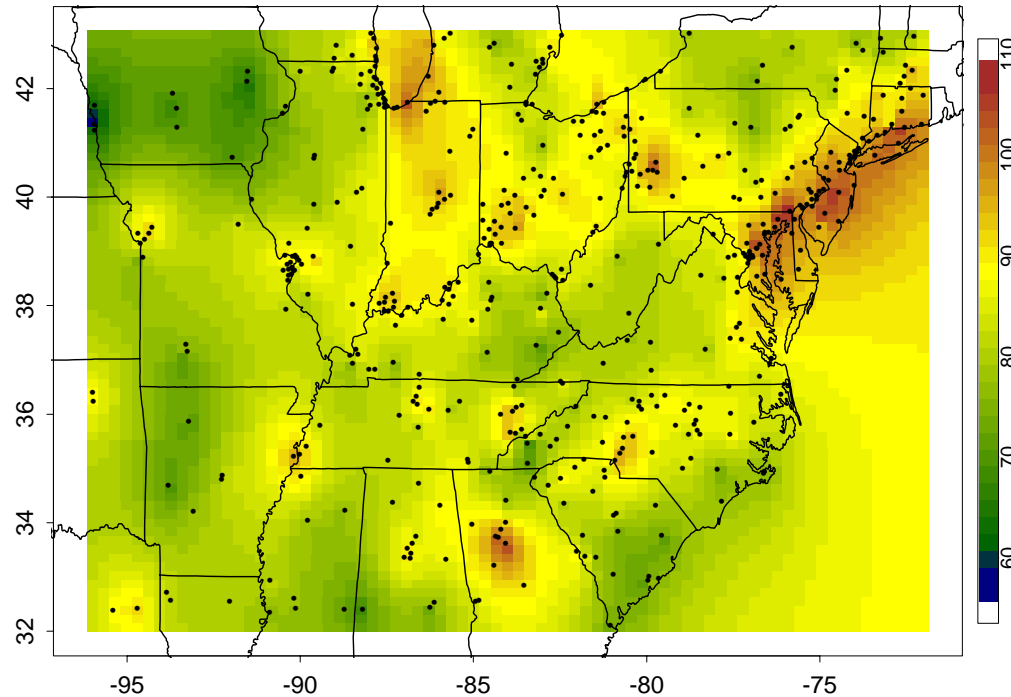


Figure 6: The graph shows the interpolated (kriging) values of the ozone air quality design values (ppb). The dots show the location of the monitoring sites, the design values are calculated as the 3-year average (95-97) of the annual fourth-highest daily maximum 8-hour average ozone concentration.

### Standard Error in the Prediction

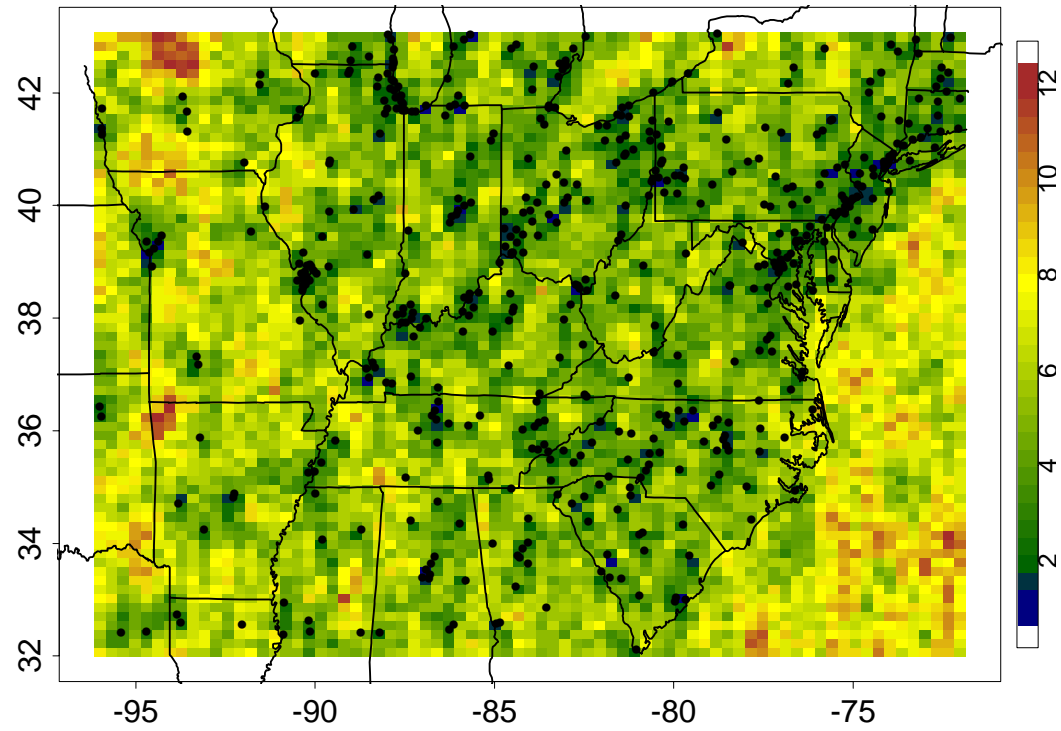
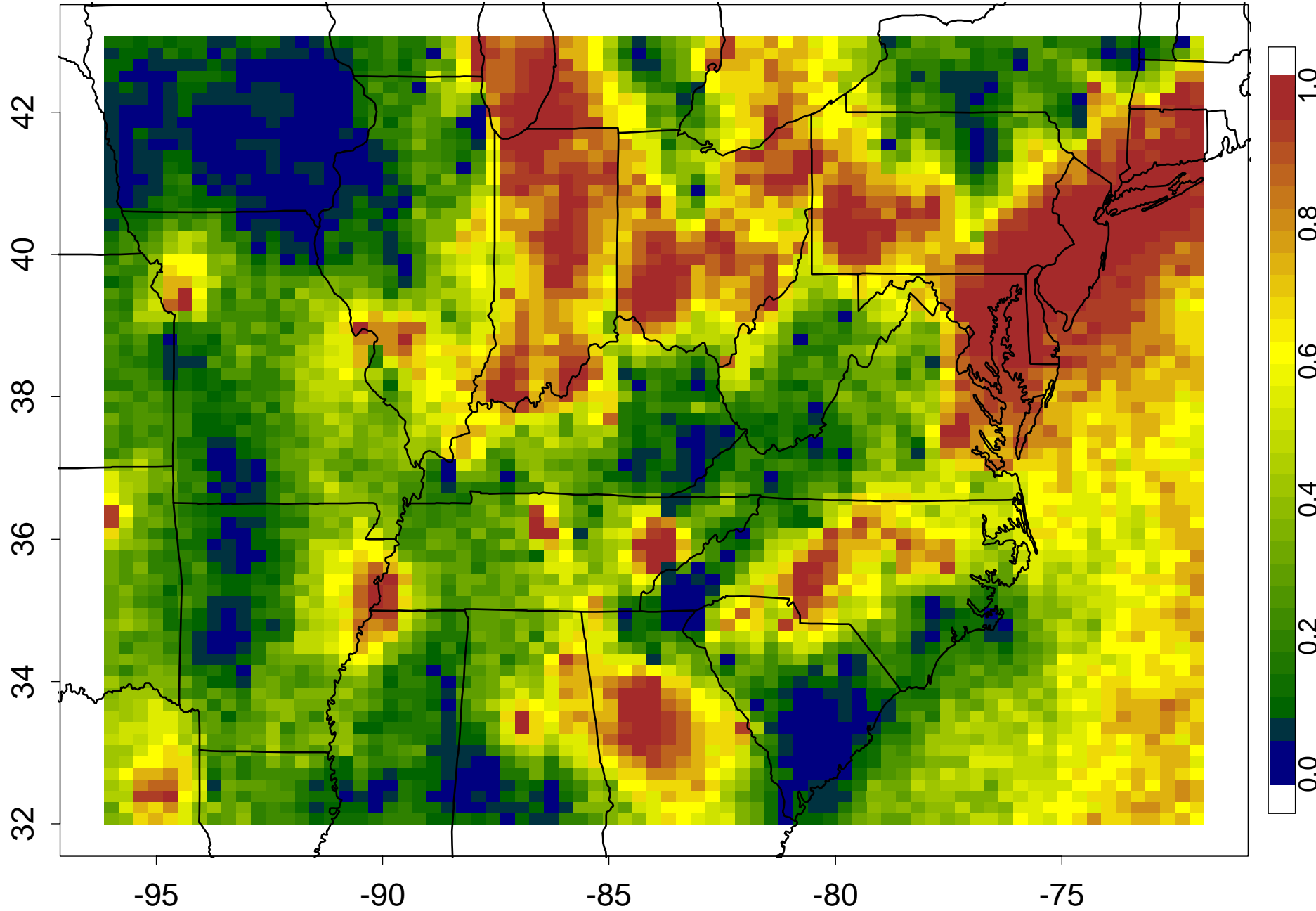


Figure 7: Standard error for the posterior predictive distribution of the ozone ambient air quality design values (ppb) using data from 1995 to 1997. The dots show the locations of the monitoring sites for the ozone.

Probability of non-attainment for the 8-h ozone standard for the 3 year average: 1995,1996,1997.



Here, we extend the stick-breaking prior to spatially smooth the potentially non-Gaussian random effects  $\tilde{V}(s)$ .

### Our model

Assuming finite  $m$ , the spatial stick-breaking model can be written as a mixture model where  $g(s) \in \{1, \dots, m\}$  indicates site  $s$ 's group, i.e,

$$\begin{aligned}
 \tilde{V}(s) &= \theta_{g(s)} & (5) \\
 \theta_j &\stackrel{iid}{\sim} N(0, \tau^2), j = 1, \dots, m \\
 g(s) &\sim \text{Categorical}(p_1(s), \dots, p_m(s)) \\
 p_j(s) &= w_j(s) V_j \prod_{k < j} [1 - w_k(s) V_k], \text{ where } V_j \stackrel{iid}{\sim} \text{Beta}(a, b)
 \end{aligned}$$

The spatial effects are each assigned a different prior distribution, i.e.,  $\tilde{V}(s) \sim F(s)$ .

- The spatial stick-breaking prior is  $F(s) \stackrel{d}{=} \sum_{i=1}^m p_i(s) \delta(\theta_i)$ .
- The probabilities  $p_i(s) = w_i(s) V_i \prod_{j=1}^{i-1} (1 - w_j(s) V_j)$  vary spatially according to the kernel functions  $w_i(s)$ .
- For example, the kernel may be  $w_i(s) = I(|s - \psi_i| < \epsilon_i)$ .
- The knots  $\psi_i$  and bandwidths  $\epsilon_i$  are unknown and given priors.
- The locations  $\theta_i \sim N(0, \tau^2)$  are shared across space.

## REFERENCES

Cohen, A. and Jones, R.H. (1969), Regression on a random field. *J. Amer. Statist. Assoc* **64**, 1172–1182.

Creutin, J.D. and Obled, C. (1982), Objective analysis and mapping techniques for rainfall fields: an objective comparison. *Water Resources Research* **18**, 413–431.

Cressie, N. (1993), *Statistics for Spatial Data*. Second edition, John Wiley, New York.

Damian, D., Sampson, P. D., and Guttorp, P. (2000). Bayesian estimation of semi-parametric non-stationary spatial covariance structure. *Environmetrics* **12**,161-176.

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

Fuentes, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika*, **89** 197-210.

Fuentes, M. (2002b). Interpolation of nonstationary air pollution processes: a spatial spectral approach. *Statistical Modeling* **2**, 281-298.

Fuentes, M. (2003). Statistical assessment of geographic areas of compliance with air quality standards. *Journal of Geophysical Research - Atmospheres*, to appear.

Fuentes, M. and Smith, R. (2001). A new class of nonstationary models. Tech. report at North Carolina State University, Institute of Statistics Mimeo Series #2534.

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. (1996), A method for statistically assessing spatio-temporal pollutant trends and meteorological transport models. Preprint, University of Wisconsin-Milwaukee.

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.

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

Mardia, K.V. and Goodall, C.R. (1993), Spatial-temporal analysis of multivariate 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 Verlag, 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.

Nychka, D., Wikle, C. and Royle, A. (2002). Multiresolution models for nonstationary spatial covariance functions. *Statistical Modeling* **2**, 299-314.

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

Schmidt, A. M. and O'Hagan, A. (2003). Bayesian inference for nonstationary spatial covariance structure via spatial deformations. *Journal of the Royal Statistical Society, Series B*, to appear. Research Report No. 498/00 Department of Probability and Statistics, University of Sheffield.

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