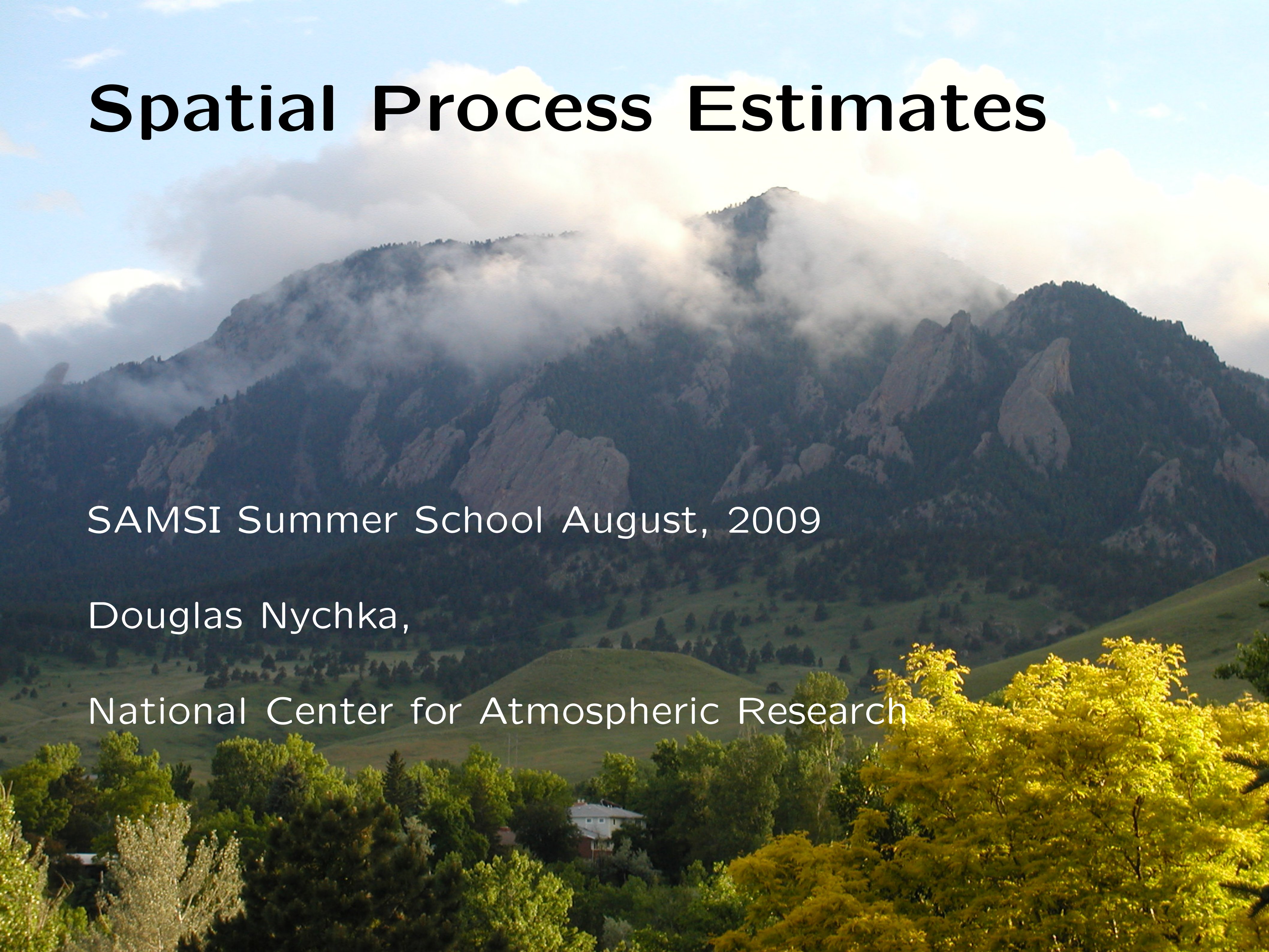
Spatial Process Estimates



SAMSI Summer School August, 2009

Douglas Nychka,

National Center for Atmospheric Research

Outline

- A spatial model and Kriging
- Kriging = Penalized least squares = splines
- Robust Kriging.
- Identifying a covariance function
- Inference

IMAGe
Institute for Mathematics Applied to Geosciences

Supported by the National Science Foundation

The additive model

Given n pairs of observations (x_i, y_i), $i = 1, \ldots, n$

$$y_i = g(x_i) + \epsilon_i$$

ϵ_i's are random errors.

Assume that g is a realization of a Gaussian process. and $\boldsymbol{\epsilon}$ are $MN(0, \sigma^2 I)$

Formulating a statistical model for g
makes a very big difference in how we solve the problem.

A Normal World

We assume that $g(\boldsymbol{x})$ is a Gaussian process,

$$\rho k(\boldsymbol{x}, \boldsymbol{x}') = COV(g(\boldsymbol{x}), g(\boldsymbol{x}'))$$

For the moment assume that $E(g(\boldsymbol{x})) = 0$.

(A Gaussian process $\equiv$ any subset of the field locations has a multivariate normal distribution.)

We know what we need to do!

If we know k we know how to make a prediction at $\boldsymbol{x}$!

$$\widehat{g}(\boldsymbol{x}) = E[g(\boldsymbol{x})|data]$$

i.e. Just use the conditional multivariate normal distribution.

A review of the conditional normal

$$\boldsymbol{u} \sim N(0, \Sigma)$$

and

$$\boldsymbol{u} = \begin{pmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{11}, \Sigma_{12} \\ \Sigma_{21}, \Sigma_{22} \end{pmatrix}$$

$$[\boldsymbol{u}_2 | \boldsymbol{u}_1] = N(\Sigma_{2,1}\Sigma_{1,1}^{-1}\boldsymbol{u}_1, \quad \Sigma_{2,2} - \Sigma_{2,1}\Sigma_{1,1}^{-1}\Sigma_{1,2})$$

(distribution of u_2 *given* u_1)

Our application is

$\boldsymbol{u}_1 = \boldsymbol{y}$ (the Data)

and

$$\boldsymbol{u}_2 = \{g(\boldsymbol{x}_1), ...g(\boldsymbol{x}_N)\}$$

a vector of function values where we would like to predict.

The Kriging weights

Conditional distribution of $\boldsymbol{g}$ given the data $\boldsymbol{y}$ is Gaussian.

Conditional mean

$$\hat{\boldsymbol{g}} = COV(\boldsymbol{g}, \boldsymbol{y})\,[COV(\boldsymbol{y})]^{-1}\,\boldsymbol{y} = A\boldsymbol{y}$$

rows of A are the Kriging weights.

Conditional variance

$$COV(\boldsymbol{g}, \boldsymbol{g}) - COV(\boldsymbol{g}, \boldsymbol{y})\,[COV(\boldsymbol{y})]^{-1}\,COV(\boldsymbol{y}, \boldsymbol{g})$$

These two pieces characterize the entire conditional distribution

Kriging as a smoother

Suppose the errors are uncorrelated Normals with variance σ^2.

$\rho K = COV(\boldsymbol{g}, \boldsymbol{y}) = COV(\boldsymbol{g}, \boldsymbol{g})$

and

$COV(\boldsymbol{y}) = (\rho K + \sigma^2 I)$

$$\hat{g} = \rho K(\rho K + \sigma^2 I)^{-1}\boldsymbol{y}$$

$$= K(K + \lambda I)^{-1}\boldsymbol{y} = A(\lambda)\boldsymbol{y}$$

My geostatistics/BLUE overhead

For any covariance and any smoothing matrix (not just A above) we can easily derive the prediction variance.

Question: Find the minimum of

$$E\left[(g(\boldsymbol{x}) - \widehat{g}(\boldsymbol{x}))^2\right]$$

over all choices of A.

The answer: The Kriging weights ... or what we would do if we used the Gaussian process and the conditional distribution.

Folklore and intuition: The spatial estimates are not very sensitive if one uses suboptimal weights, especially if the observations contain some measurement error.
It does matter for measures of uncertainty.

Kriging with a fixed part

Adding a fixed component

$$g(x) = \sum_i \phi_i(x) d_i + h(x)$$

$\boldsymbol{d}$ *is fixed*

h *is a mean zero process with covariance,* k*.*

BLUE/Universal Kriging

Find $\boldsymbol{d}$ by Generalized least squares

$$\hat{\boldsymbol{d}} = \left(T^T M^{-1} T\right)^{-1} T^T M^{-1} \boldsymbol{y}$$

$M = K + \lambda I$

"Krig" the residuals

$$\hat{h} = K(K + \lambda I)^{-1}(\boldsymbol{y} - T\hat{\boldsymbol{d}})$$

In general:

$$\widehat{g}(x) = \sum_i \phi_i(x)\widehat{d}_i + \sum_j k(x, x_j)\widehat{c}_j$$

with

$$\widehat{\boldsymbol{c}} = (K + \lambda I)^{-1}(\boldsymbol{y} - T\widehat{\boldsymbol{d}})$$

The spline connection

Basis functions: determined by the covariance function

Penalty function: $\Omega = K$ is based on the covariance.

The minimization criteria:

$$\min_{\boldsymbol{d},\boldsymbol{c}} \sum_{i=1}^{n} (\boldsymbol{y} - (T\boldsymbol{d} + K\boldsymbol{c})_i)^2 + \lambda \boldsymbol{c}^T K \boldsymbol{c}$$

Kriging estimator is a spline with reproducing kernel k.

λ is proportional to the measurement (nugget) variance

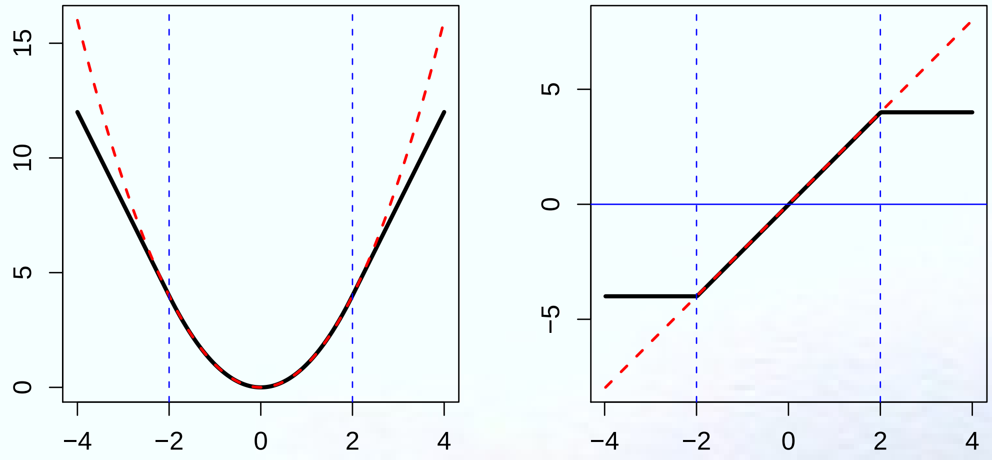
Robust Kriging

$\rho(u)$ a robust function that grows slower than u^2

The robust minimization criteria:

$$\min_{\boldsymbol{d},\boldsymbol{c}} \sum_{i=1}^{n} \rho(\boldsymbol{y} - (T\boldsymbol{d} + K\boldsymbol{c})_i) + \lambda \boldsymbol{c}^T K \boldsymbol{c}$$

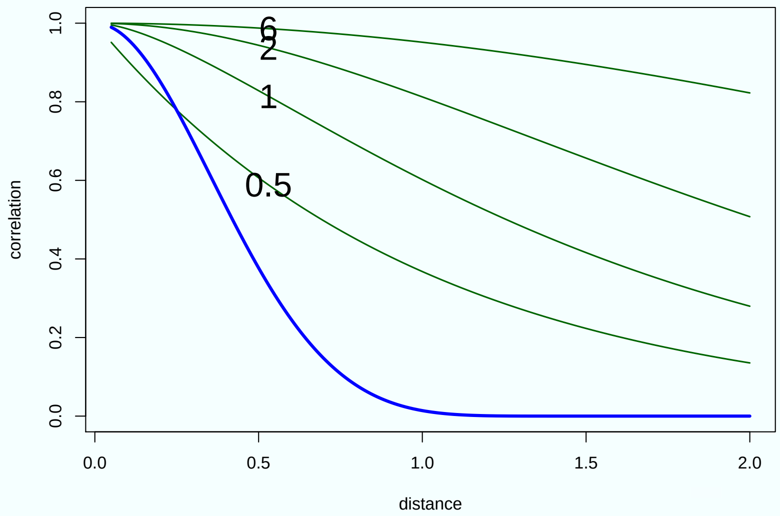
An example of ρ and its derviative



Identifying a covariance function

The Matern class of covariances:

$\phi(d) = \rho\psi_\nu(d/\theta))$ with ψ_ν a Bessel function.



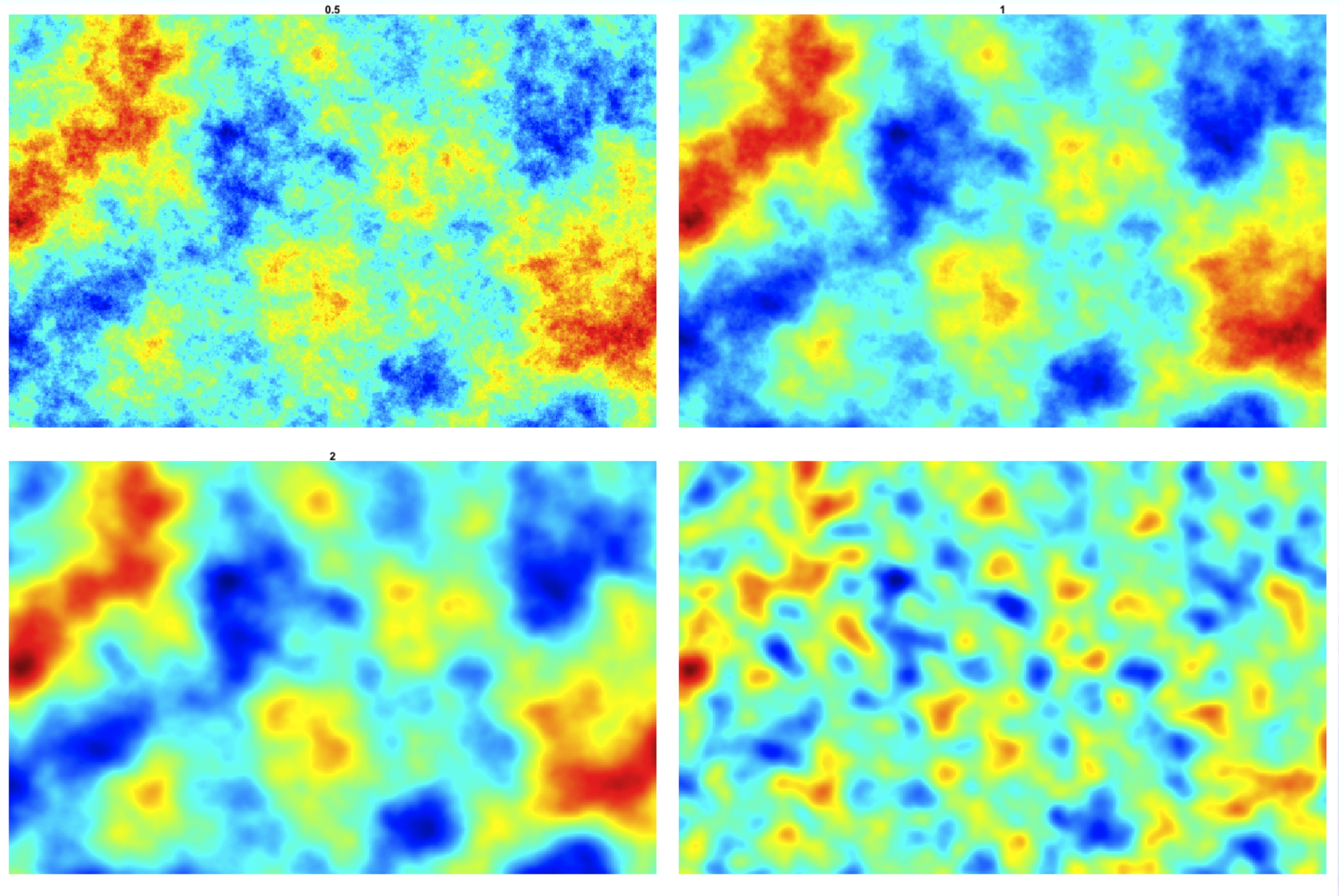
θ a range parameter, ν smoothness at 0.

- ψ_ν is an exponential for $\nu = 1/2$ as $\nu \to \infty$ Gaussian.
- mth order thin plate spline in $\Re^d$ $\nu = 2m - d$.)

Compactly support Wendland covariance

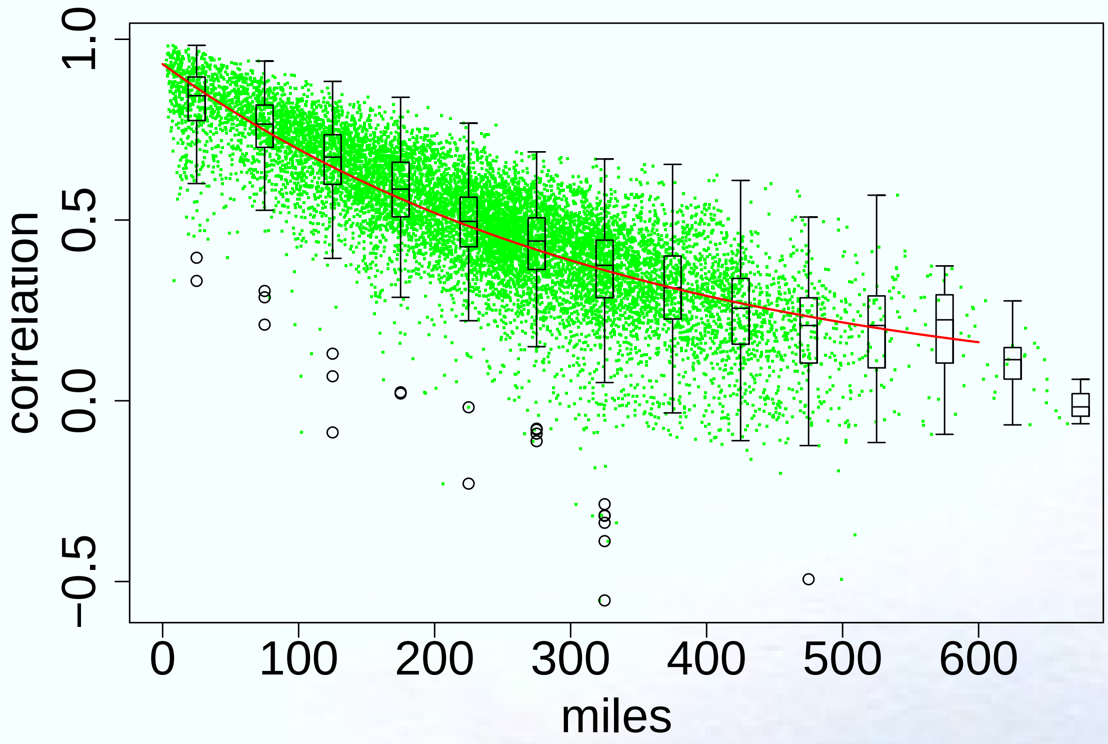
What kind of processes are these?

Matern (.5,1.0,2.0) and Wendland (2.0)



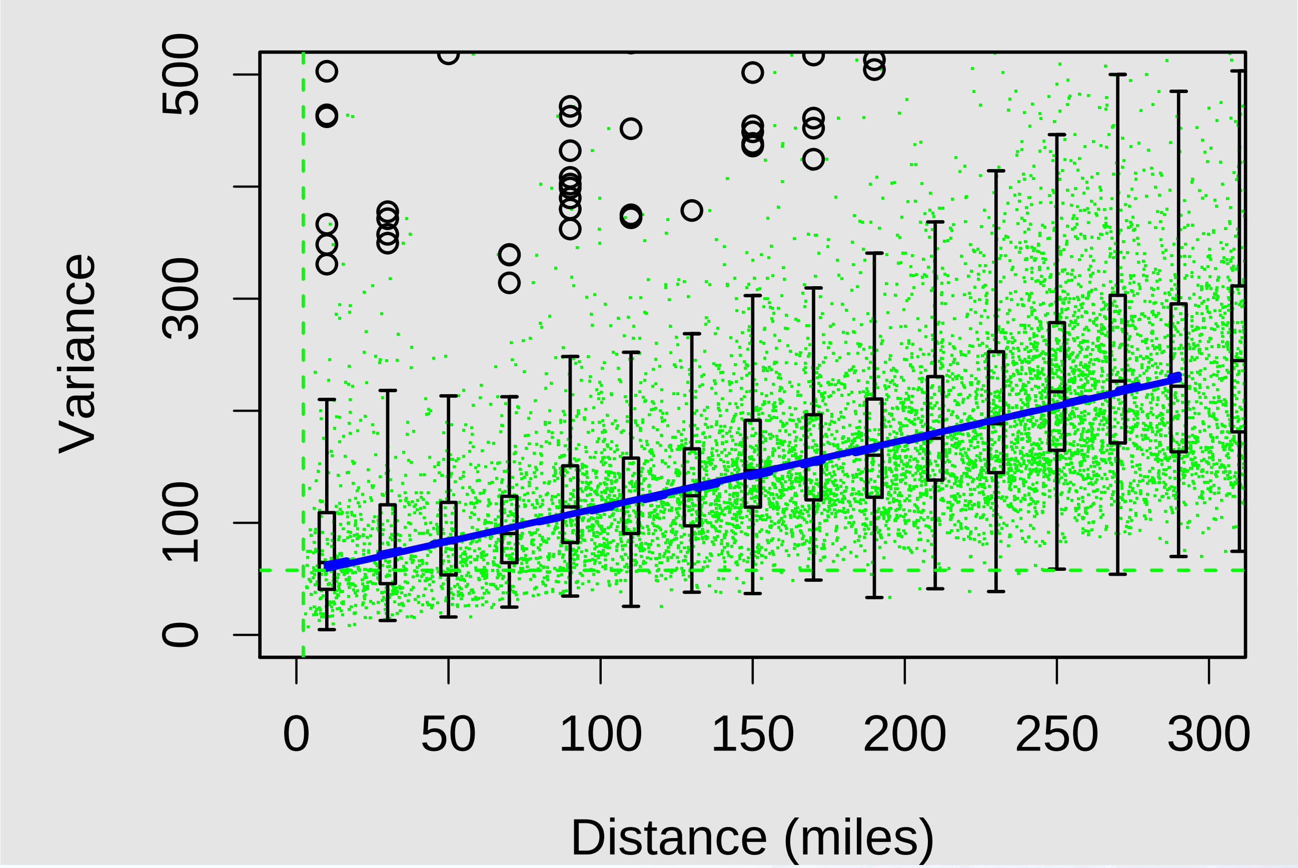
Correlations among ozone

In many cases spatial processes also have a temporal component. Here we take the 89 days over the "ozone season" and just find sample correlations among stations.



Comparison to the variogram

Sample variogram for 89 days during summer 1987 :

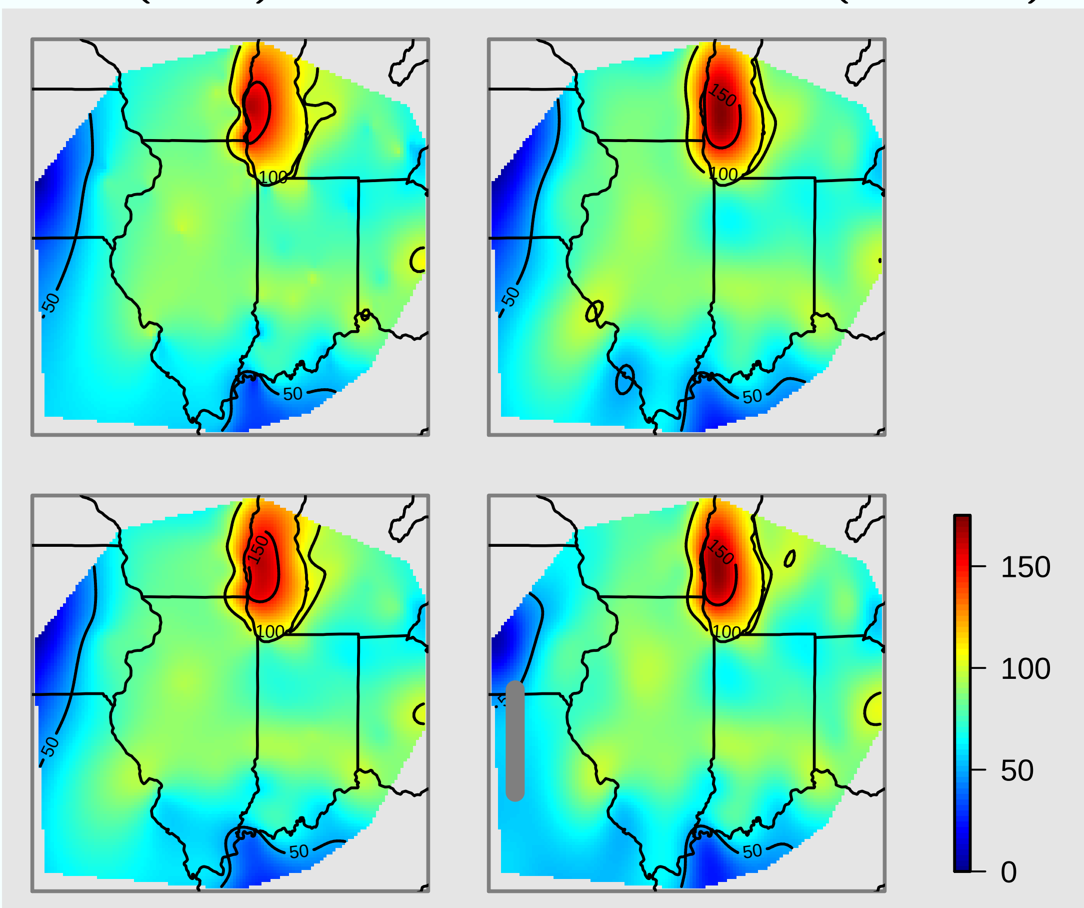


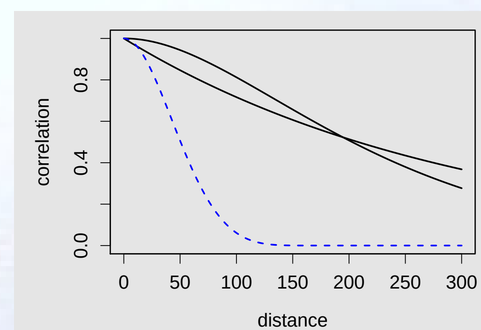
"." Sample variogram values

"–" fitted linear function Variogram ~ distance/θ

Sensitivity to the covariance

Exp (200) Matern (2, 200)

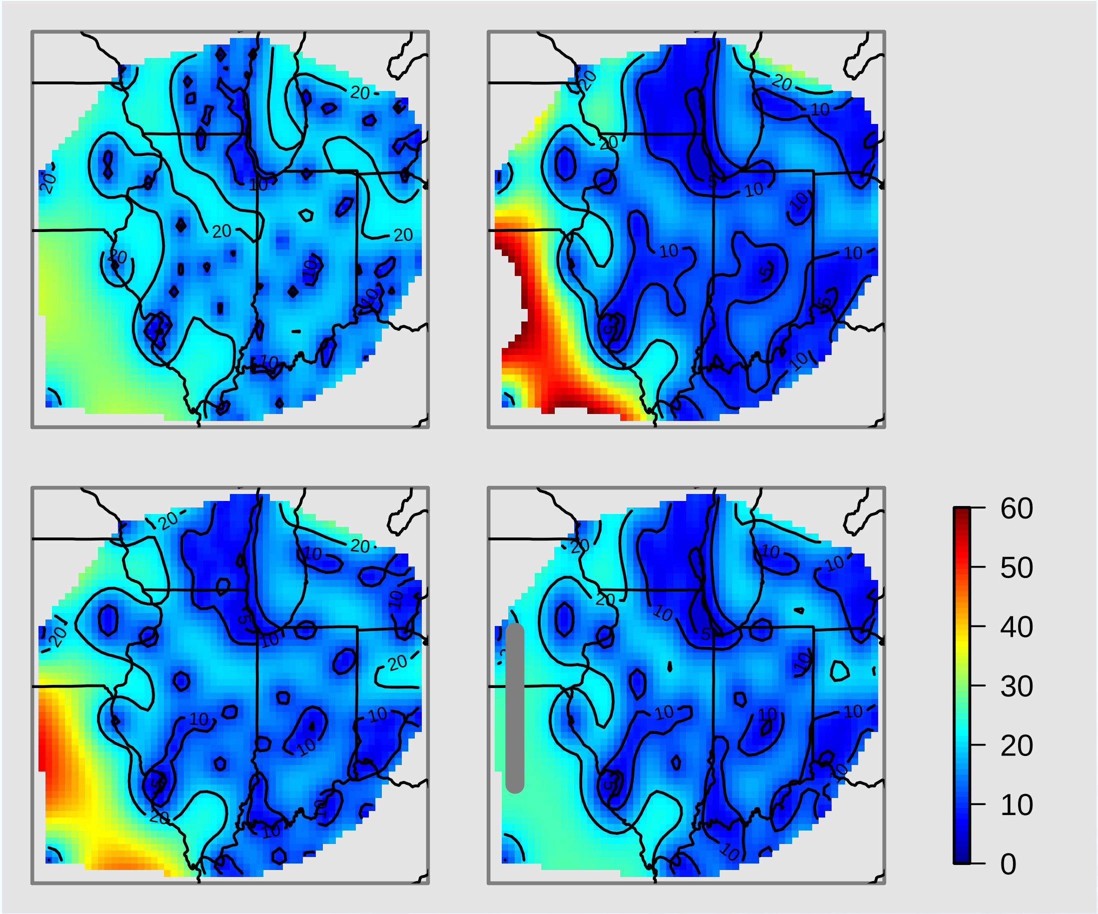




Thin plate spline Wendland (2, 180)

Sensitivity to the covariance (SE)

Exp (200) Matern (2, 200)

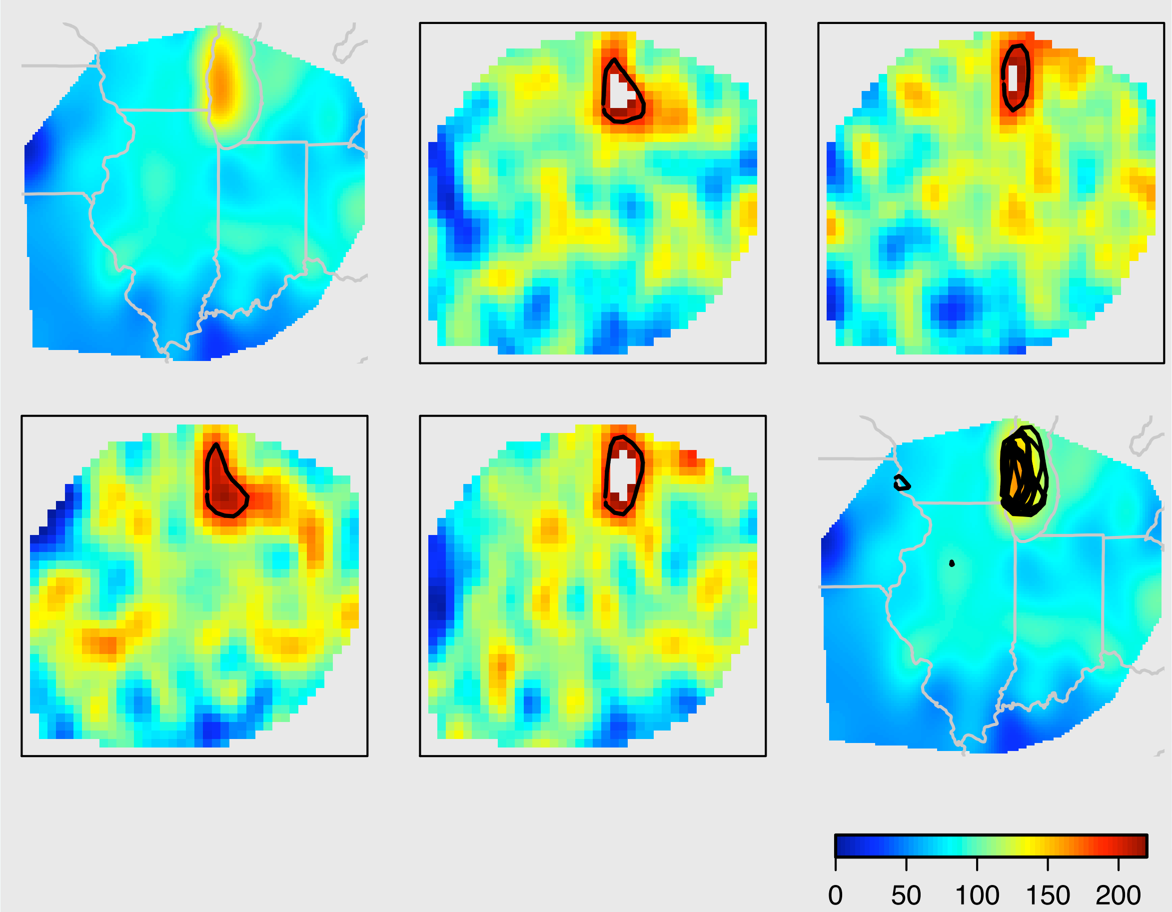


Thin plate spline Wendland (2, 180)

Uncertainty in exceeding 140 PPB

Mean field, four realizations of the conditional distribution

Contours from 10 cases



Summary

A spatial process model leads to a penalized least squares estimate

A spline = Kriging estimate= Bayesian posterior mode

For spatial estimators the basis functions are related to the covariance functions and can be identified from data