Inference for spatial fields

Douglas Nychka,
`www.image.ucar.edu/~nychka`

- **A spatial model and Kriging**
- **Kriging = Penalized least squares**
- **The Bayes connection**
- **Identifying a covariance function**

Supported by the National Science Foundation *L' Ecole D'Ete, Ovronnaz Sep 2005*

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 ϵ 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 $g(x)$ 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 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

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

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

Our application is

$\boldsymbol{u}_1 = y$ **(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 g given the data y is Gaussian.

Conditional mean

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

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

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

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

My geostatistics/BLUE overhead

For any covariance and any smoothing matrix (not just S 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 S. *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

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

$\mathbf{d}$ *is fixed*

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

The BLUE/Universal Kriging estimate is:

Find d by Generalized least squares

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

"Krig" the residuals

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

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 connection to penalized least squares, splines and the smoothing parameter

Basis functions:

determined by the covariance function

Penalty function

K is based on the covariance.

The minimization criteria:

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

The Kriging estimator is a spline with reproducing kernel k.

λ is proportional to the measurement (nugget) variance

The Bayes connection

Bracket notation is very useful:

$[Z]$ **the density function for the random variable** Z

$[Y|z]$ **the conditional density function for the random variable** Y **given** z.

$[\boldsymbol{y}|g]$ **the likelihood for the data**

$$[\boldsymbol{y}|g] \sim MN(\boldsymbol{g}, \sigma^2 I)$$

$[g]$ **the prior for** g.

$$[g] \sim MN(0, \rho K)$$

Bayes Theorem: the posterior

$$[g|y] = \frac{[y|g][g]}{[y]} \sim [\boldsymbol{y}|g][g]$$

The Posterior mode: where $[g|y]$ has a maximum.

Maximizing:

$[g|y]$ is the same as

Minimizing:

$$-2ln[g|y] = -2ln([y|g]) - 2ln([g]) + 2ln([y])$$

The posterior mode is the penalized least squares estimate where the penalty is equivalent to a prior!

This is true even we let ''$\boldsymbol{g}$'' be the entire field, not just its values at the observations.

A causal example of identifying a covariance function

A useful form for k are isotropic correlations:

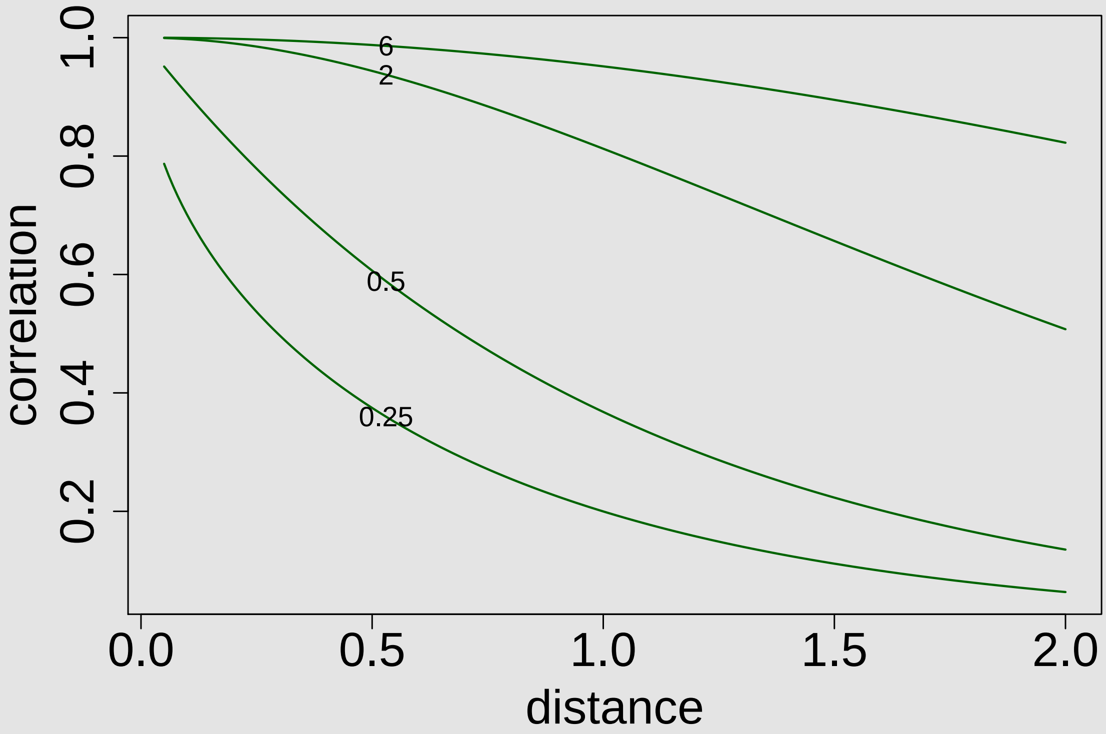
$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma(\boldsymbol{x})\sigma(\boldsymbol{x}')\phi(\|\boldsymbol{x} - \boldsymbol{x}'\|)$$

The Matern class of covariances:

$$\phi(d) = \rho\psi_\nu(d/\theta))$$

θ a range parameter, ν smoothness at 0.
ψ_ν is an exponential for $\nu = 1/2$ as $\nu \to \infty$ Gaussian.

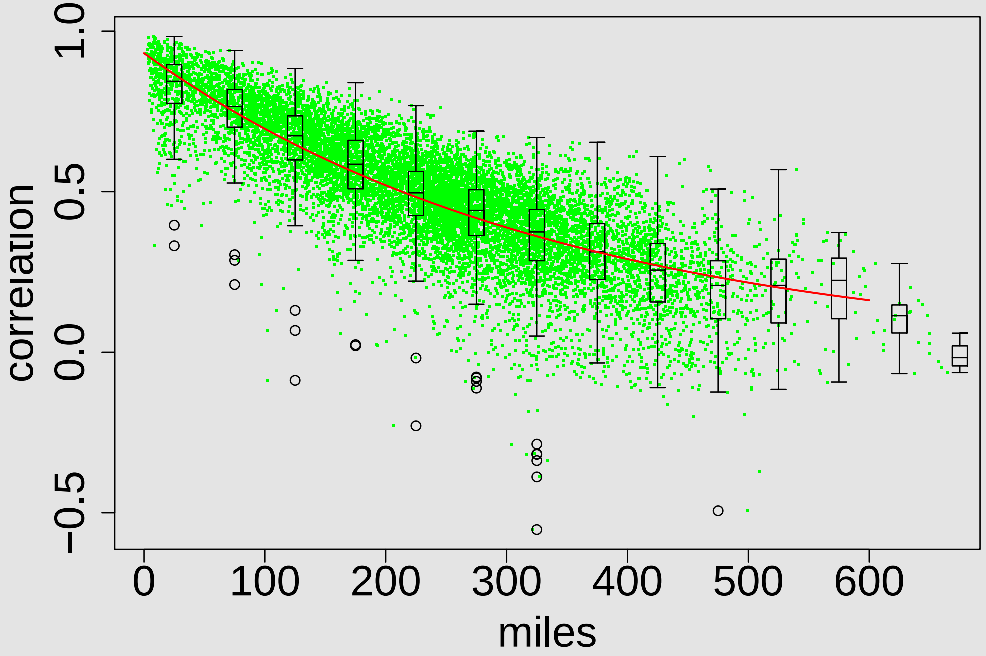
Matern family: the shape ν



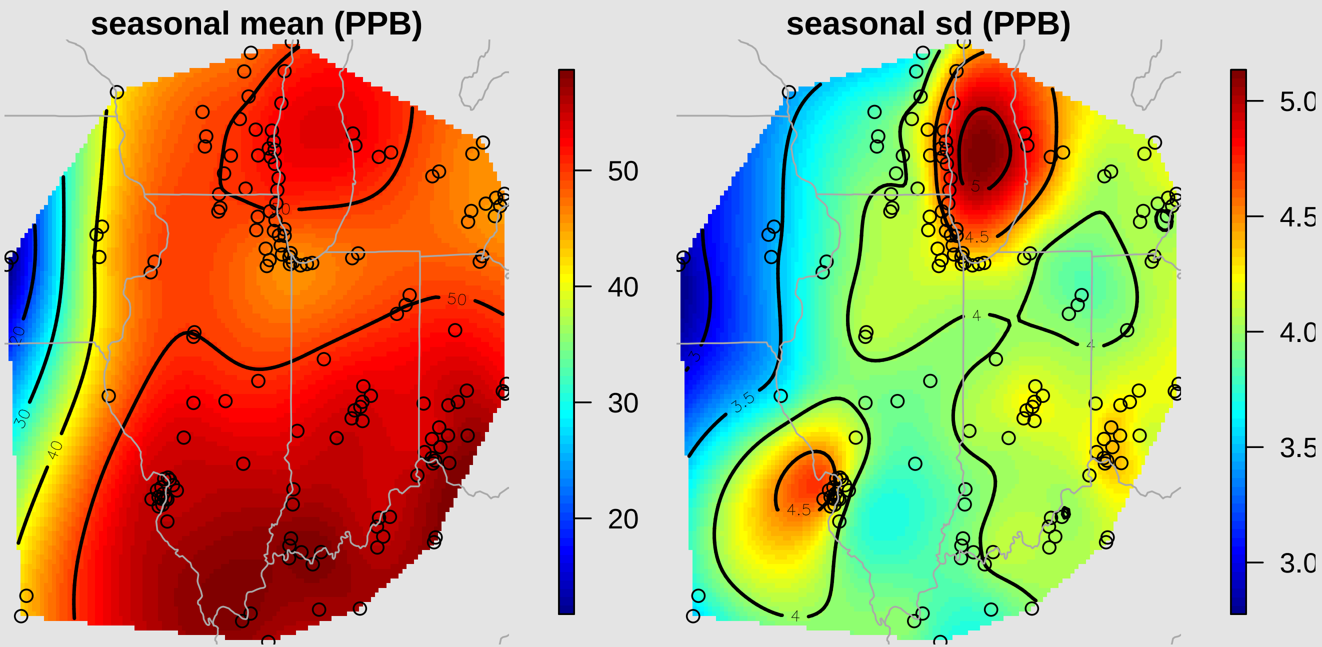
(mth order thin plate spline in $\Re^d$ $\nu = 2m - d$!)

Using the temporal information

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.



Mean and SD surfaces for 1987 ozone



Covariance model:

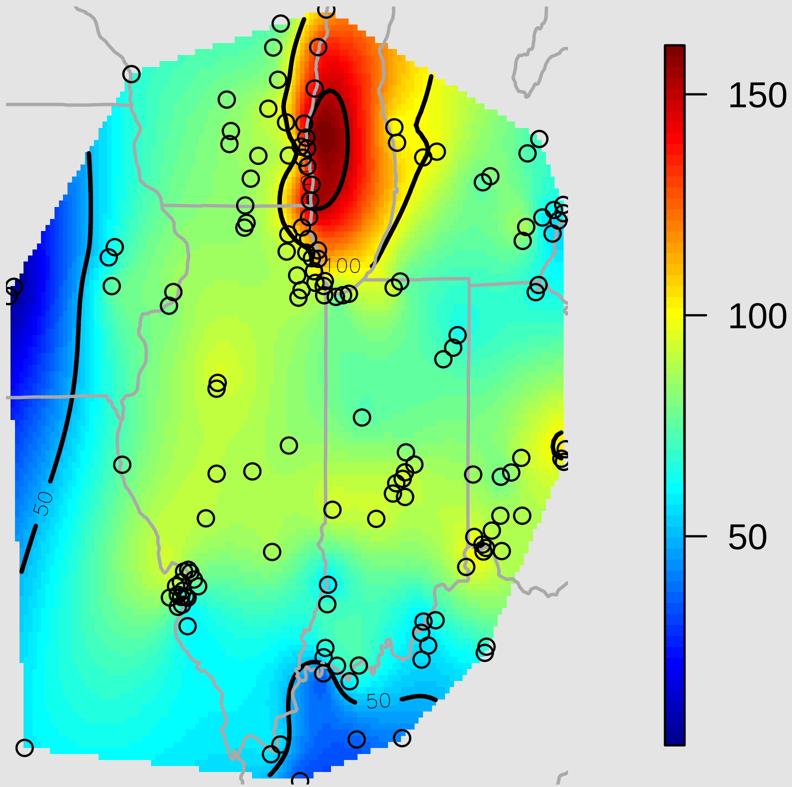
$$k(\boldsymbol{x}, \boldsymbol{x}') = \rho\sigma(\boldsymbol{x})\sigma(\boldsymbol{x}')exp(-||\boldsymbol{x} - \boldsymbol{x}'||/\theta)$$

Mean model: $E(z(\boldsymbol{x})) = \mu(\boldsymbol{x})$

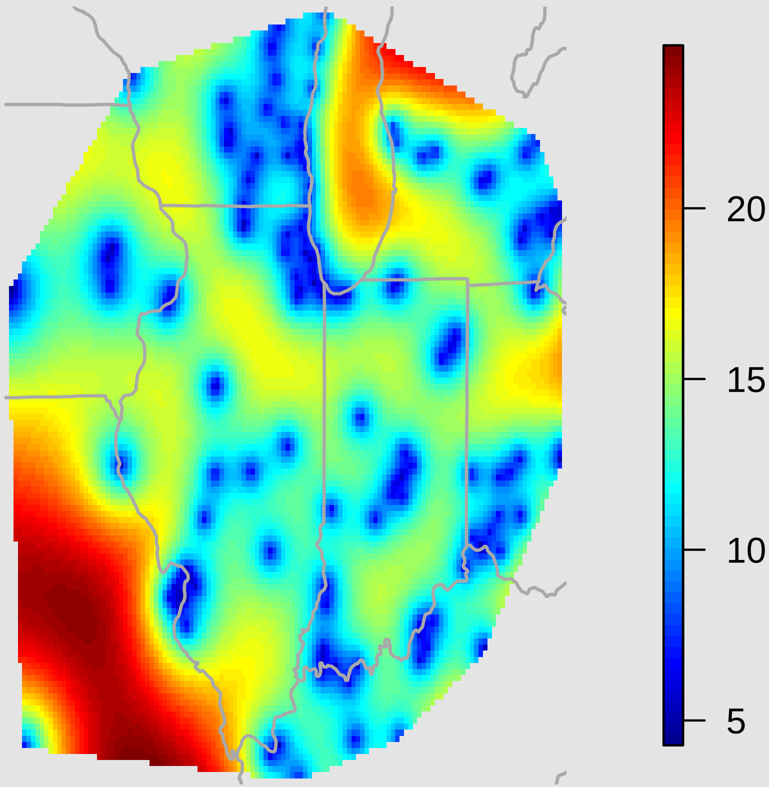
where μ is also a Gaussian spatial process.

Spatial estimate and uncertainty

Posteror mean



Posterior standard deviation.



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