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 $(x_i, y_i)$, $i = 1, \ldots, n$

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

$\epsilon_i$’s are random errors.

Assume that $g$ is a realization of a Gaussian process. and $\epsilon$ 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 $g(x)$ is a Gaussian process,

$$\rho_k(x, x') = \text{COV}(g(x), g(x'))$$

For the moment assume that $E(g(x)) = 0$.

(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 $x$!

$$\hat{g}(x) = E[g(x)|\text{data}]$$

i.e. Just use the conditional multivariate normal distribution.
A review of the conditional normal

\[ u \sim N(0, \Sigma) \]

and

\[ u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{11}, \Sigma_{12} \\ \Sigma_{21}, \Sigma_{22} \end{pmatrix} \]

\[ [u_2|u_1] = N(\Sigma_{2,1} \Sigma_{1,1}^{-1} z_1, \quad \Sigma_{2,2} - \Sigma_{2,1} \Sigma_{1,1}^{-1} \Sigma_{1,2}) \]

Our application is

\[ u_1 = y \quad \text{(the Data)} \]

and

\[ u_2 = (g(x_1, \ldots g(x_N))) \quad \text{a vector of function values where we would like to predict.} \]
The Kriging weights

Conditional distribution of $g$ given the data $y$ is Gaussian.

**Conditional mean**

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

rows of $S$ are the Kriging weights.

**Conditional variance**

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

With these two pieces we characterize the entire conditional distribution.
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

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

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.
The connection to penalized least squares, splines and the smoothing parameter

Let $COV(g) = \rho K$

(recall $\rho$ is a just scale factor)

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

and so

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

where $\lambda = \sigma^2 / \rho$

Have we seen this before?
The main results

Recall

\[ \hat{g}(x) = \sum_{l=1}^{n} \hat{\theta}_k \psi_k(x) \]

\[ \min_{\theta} \sum_{i=1}^{n} (y - \left[ W \theta \right]_i)^2 + \lambda \theta^T B \theta \]

The Kriging estimator is identical to a penalized least squares estimator with basis functions \( k(., x_i) \), penalty matrix \( B = K \) and \( \lambda = \sigma^2 / \rho \).

The Kriging estimator is a spline with reproducing kernel \( k \)!

\( \lambda \) 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 $z$.

$[y|g]$ the likelihood for the data

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

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

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

Bayes Theorem: the posterior

$$[g|y] = \frac{[y|g][g]}{[y]} \sim [y|g][g]$$
The Posterior mode: where $[g|y]$ has a maximum.

Maximizing $[g|y]$ is the same as

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

or

$$\min_g [-2\ln([y|g]) - 2\ln([g])]$$

or plugging in the densities and some hand waving

$$\min_{\theta} \sum_{i=1}^{n} \frac{(y - [K\theta]_i)^2}{\sigma^2} + \rho\theta^TK\theta$$

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

This is true even we let "$g$" 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:

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

The Matern class of covariances:

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

$\theta$ a range parameter, $\nu$ smoothness at 0. $\psi_{\nu}$ is an exponential for $\nu = 1/2$ as $\nu \to \infty$ Gaussian.
Matern family: the shape $\nu$

\[ (m^{th} \text{ order thin plate spline in } \mathbb{R}^d \nu = 2m - d!) \]
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

seasonal mean (PPB)

seasonal sd (PPB)

Covariance model:
\[ k(x, x') = \rho \sigma(x) \sigma(x') \exp\left(-\frac{|x - x'|}{\theta}\right) \]

Mean model: \[ E(z(x)) = \mu(x) \]

where \( \mu \) is also a Gaussian spatial process.
Spatial estimate and uncertainty

Posterior 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