Outline

- Gaussian processes
- Some common covariance functions
- Processes from constructions
- Kriging estimate
- Basis functions → Kriging
- Kriging → basis functions
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.
To describe $g(x)$ as a Gaussian process, start with a covariance function:

$$\rho_k(x_1, x_2) = \text{COV}(g(x_1), g(x_2))$$

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

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

Specifying the covariance function (and the mean) is the recipe for describing the Gaussian process.

$\rho$ will be a key parameter in the spatial model.
The exponential covariance

$$\rho k(x_1, x_2) = \rho e^{-D/\theta}$$

D is the distance between the two locations $x_1$ and $x_2$

This depends on two parameters: $\rho$ (sill) and $\theta$ (range).

- $\text{VAR}(g(x)) = \rho$

- $\text{COV}(g(x_1), g(x_2))$ falls to $\approx .36$ of $\rho$ when the distance is equal to $\theta$.

- Correlations are just $e^{-D/\theta}$
Realizations of process

Covariance matrix

Three realizations of the process.
Families of covariances

**Matern:**
\[ \phi(d) = \rho \psi_\nu(d/\theta) \] with \( \psi_\nu \) a Bessel function.

**Wendland:**
Polynomial that is exactly zero outside given range.

- \( \theta \) a range parameter
- \( \nu \) smoothness at 0.
- \( \psi_\nu \) is an exponential for \( \nu = 1/2 \) as \( \nu \to \infty \) Gaussian.
- As \( \nu \) increases the process is smoother.

Compactly supported Wendland covariance
These are computed in fields with the functions Matern and Wendland.
What do these processes look like?

Varying the smoothness:
Matern (.5) Matern(1.0)

Matern (2.0) and Wendland (2.0)

Simulations found by circulant embedding

D. Nychka Spatial Processes
A constructivist approach

If you build it ...  

Marginal variance

- Start with a simple spatial process $U(x)$ $\text{VAR}(U(x)) = 1$
- Specify a positive function $\rho(x)$, the marginal variance
- Scale $U$: $g(x) = \sqrt{\rho(x)}U(x)$

Convolution:

- Start with a simple spatial process $U(x)$.
- Smooth it: $g(x) = \int \Omega(x, u)U(u)du$
  or approximately $g(x) = \sum_{j=1}^{M} \Omega(x, u_j)U(u_j)$
More constructions

**Basis functions**

- Start with a basis \( \phi_j(x) \).
- Generate coefficients \( c \sim MN(0, P) \).
- Sum: \( g(x) = \sum_{j=1}^{M} \phi_j(x)c_j \)

\[
k(x, x') = \sum_{j,k=1}^{M} \phi_j(x)P_{j,k}\phi_k(x')
\]

Also, \( M \) can be infinite if it makes sense e.g. \( k(x, x) \) is finite.

**Deformation/transformation of spatial coordinates**

- Start with your favorite homogenous process \( U(u) \).
- Specify an (nonlinear) invertible transformation of coordinates \( x = \Gamma(u) \) with \( x \) being the physical domain and \( u = \Gamma^{-1}(x) \).
- Evaluate \( g(x) = U(\Gamma^{-1}(x)) \).
Stochastic Partial Differential Equations

- Start with a partial differential operator $\mathcal{L}$
  e.g. $\mathcal{L}(g) = \Delta g + \alpha g$ with $\Delta = \partial^2/\partial x_1^2 + \partial^2/\partial x_2^2$
- $U(x)$ your favorite homogenous process.
- Solve the SPDE for $g \mathcal{L}(g) = U$

This is the same as convolution by choosing the right $\Omega$ (the Green’s function).
Basis functions

For any reasonable covariance we have the Karhunen-Loeve expansion

$$
cov(g(x), g(x')) = k(x, x') = \sum_{j=1}^{\infty} c_j \phi_j$$

where $$\{c_j\}$$ are uncorrelated!

So in some sense all of the strategies might be approximated by basis functions and random coefficients.
An example

A basis

Marginal variance

Coefficient covariance matrix

Covariance of process
Some preliminaries

Suppose we know $Y_1$ what is the conditional distribution of $Y_2$?

Some useful bracket/bar notation:

- $[Y_1]$ the density for just $Y_1$
- $[Y_1, Y_2]$ the joint density for just $Y_1$ and $Y_2$
- $[Y_2|Y_1]$ the conditional density for $Y_2$ given $Y_1$
  | say "given"

Algebraically this is given by: $f(y_1, y_2)/g(y_1)$

with $g$ the $N(\mu_1, \sigma_1^2)$ density for just $Y_1$. 

The Multivariate Normal

\[ z \sim N(\mu, \Sigma) \]

**Density function**

\[ f(y) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(y-\mu)^T \Sigma^{-1} (y-\mu)\right) \]

\( Q = \Sigma^{-1} \) is the *precision* matrix.

**Proving things**

The easiest way to manipulate this random vector is by the representation:

\[ z = \Sigma^{1/2} u + \mu \]

\( u \) is a vector of independent \( N(0,1) \)'s.

This also tells us how to simulate the random vector!
Partitioning and conditioning

Divide these into two parts: what we observe and what we want to predict.

\[ z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \]

\( \Sigma \) is divided up so that

- e.g. \( \Sigma_{11} \) is the covariance matrix for \( z_1 \)

\[ [z_2|z_1] = N(\Sigma_{2,1} \Sigma_{1,1}^{-1}(z_1 - \mu), \Sigma_{2,2} - \Sigma_{2,1} \Sigma_{1,1}^{-1} \Sigma_{1,2}) \]

(distribution of \( z_2 \) given \( z_1 \))
Our application is

\[ z_1 = y \] (the Data)

and

\[ z_2 = \{ g(x_1), \ldots g(x_N) \} \]

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

*Sorting through pieces of \( \Sigma \)*

\[ \Sigma_{1,1} = \text{cov}(y, y) \quad \Sigma_{2,1} = \text{cov}(y, g) \quad \Sigma_{2,2} = \text{cov}(g, g) \]
The Kriging Estimator

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

**Conditional mean**

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

rows of $A$ are the Kriging weights.

**Conditional covariance**

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

These two pieces characterize the entire conditional distribution

Diagonal elements of the conditional are the prediction errors.
Posterior mode and ridge regression

Basis function model:

\[ g(x) = \sum_{j=1}^{M} \phi_j(x)c_j \]

Observations:

\[ y = \Phi c + \epsilon \]

\[ [y|c, \sigma, \rho] \sim MN(\Phi c, \sigma^2 I) \text{ with } \Phi_{i,j} = \phi_j(x_i) \]

Prior:

\[ [c] = MN(0, \rho P) \]
finding the posterior

The log posterior:

$$-\left(\frac{(y - \Phi c)^T(y - \Phi c)}{2\sigma^2} + \frac{c^T P^{-1} c}{2\rho}\right) - n \log(\sigma) - M \log(\rho)/2 + C$$

Posterior maximum at:

$$\hat{c} = (\Phi^T \Phi + \lambda P^{-1})^{-1} \Phi^T y \quad \text{with} \quad \lambda = \sigma^2/\rho$$

Posterior is MN

$$[c|y, \sigma, \rho] = MN(\hat{c}, \left[COV(c, c) - COV(c, y)(COV(y)^{-1}COV(y, c))\right])$$
Connections between estimates
From the posterior mode to Kriging

Posterior mode = Kriging

- Assuming everything is multivariate normal (MN).
- Kriging estimate = $E[g(x)|y] = \sum \phi_j(x)E[c_j|y] = \sum \phi_j(x)\hat{c}_j$

A more direct way is to work out the matrix algebra and show the estimates are equal.

$$\text{cov}(y, y) = (\rho \Phi^T P \Phi + \sigma^2 I)$$
$$\text{cov}(g(x), y) = \phi(x)^T(\rho P)\Phi$$

Need to use the Sherman-Morrison-Woodbury matrix identity.
From a covariance to basis functions

From a covariance function $k(.,.)$ define a particular basis and coefficient covariance.

$$\phi_j(x) = k(x, x_j)$$

A basis function for each location.

$$Q_{i,j} = \lambda k(x_i, x_j) \text{ and } P^{-1} = Q$$

Using this basis and this penalty matrix, $\hat{g}(x)$ will be identical to the Kriging estimator.

*But the conditional covariance will not be correct.*
Estimating covariance parameters

Integrating out $c$ gives

$$y \sim MN(0, \rho \Phi P \Phi^T + \sigma^2 I)$$

with log likelihood

$$-(1/2)y^T(\rho \Phi P \Phi^T + \sigma^2 I)^{-1}y - (n/2) \log |\rho \Phi P \Phi^T + \sigma^2 I| + C$$

Reparametrize with $\lambda = \sigma^2 / \rho$ and $\rho$.

The likelihood can be maximized over $\rho$ leaving a concentrated likelihood over $\lambda$ and any other covariance parameters in $P$. 
An example for climate

Data: Mean spring (MAM) maximum temperature for Colorado compiled from COOP observing network.

Spatial Process model:

- Fixed part: (lon, lat) (lon,lat, elevation)
- Covariance: (sill, range, smoothness) great circle distance using in covariance function
- Observation model: Measurement variance $\sigma^2$

Goal is to predict onto the 4km PRISM grid.
Data with elevations (1000, 2000, 3000 m)
Profile over smoothness and range

Elevation as linear fixed effect

Contours at 50%, 95%, 99% approximate confidence

Range is in miles.
Profiles for a few models

\[(\text{Range, Smoothness}) = (200, 0.5) (300, 1.0) (\infty, 2.0), \ Tps \ m = 2\]

From fields

\[
\text{obj3}\leftarrow \text{Krig}(\ x, \ y, \ \text{Covariance="Matern"}, \ \text{smoothness}=.5, \\
\quad \text{theta}=200, \ \text{Distance="rdist.earth"}, \\
\quad Z=\text{COelevation}, \ \text{method="REML"} )
\]
Fitted surfaces

Exponential, full surface, smooth

Thin plate spline, full surface, smooth

Reference Elevations
Summary

- Several different ways to build spatial process models
- Posterior distribution easy to find if covariance is known
- Posterior mode = posterior mean = Kriging

- Ridge regression model =
  Kriging with a specific (perhaps funky) spatial process.

- Kriging =
  Ridge regression with specific basis functions and penalty.