Estimating curves and surfaces

Lecture One

Douglas Nychka

National Center for Atmospheric Research

National Science Foundation *KAUST, March 2014*

These lectures

Four parts:

- An introduction to spatial statistics
- Covariance choices for large data sets
- Multi-resolution spatial model
- Spatial statistics and inverse problems

Big ideas:

- Separate what you observe from what you want to see.
- Create a model for the unknown surface or curve
- Quantify uncertainty using a statistical model.

Outline

- Basis function
- Least squares smoothers
- Gaussian processes and covariance functions
- Kriging as penalized least squares

Estimating a curve or surface.

An additive statistical model:

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

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

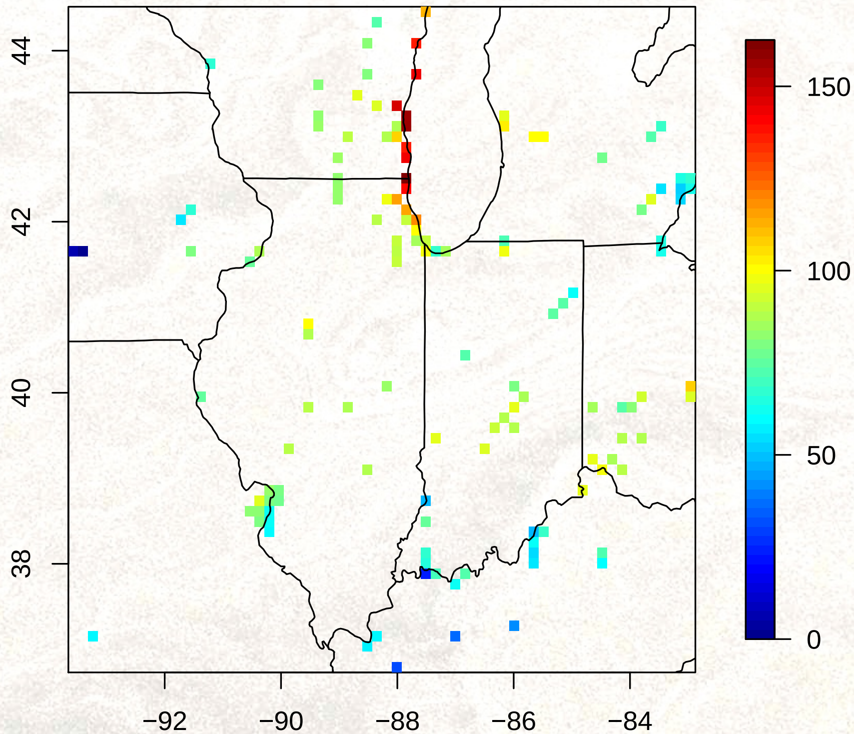
ϵ_i's are random errors and g is an unknown, smooth function.

The goal is to estimate g based on the observations

A two dimensional example

Predict surface ozone where it is not monitored.

Ambient daily ozone in PPB June 16, 1987, US Midwestern Region.



Representing a curve

Start with your favorite m basis functions $\{b_1(x), b_2(x), \ldots, b_m(x)\}$

The estimate has the form

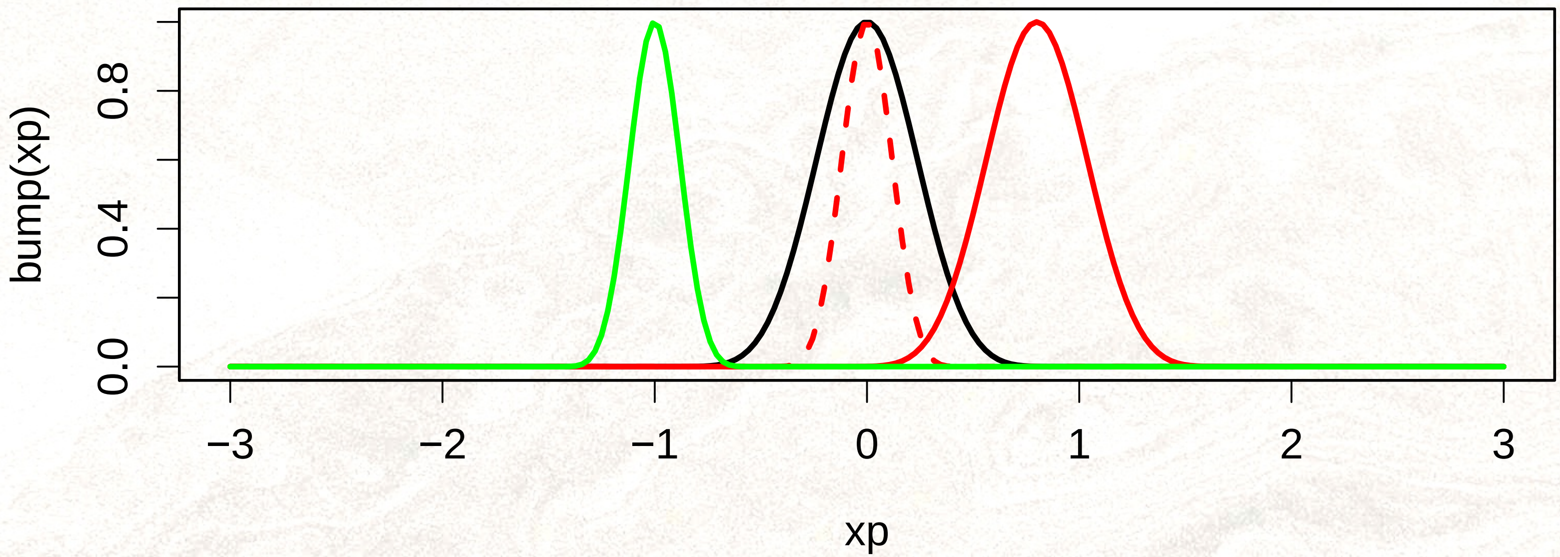
$$\hat{g}(x) = \sum_{k=1}^{m} \beta_k b_k(x)$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)$ are the coefficients.

The basis functions are fixed and so the problem is to just find the coefficients.

Many spatial statistics problems have this general form or can be approximated by it.

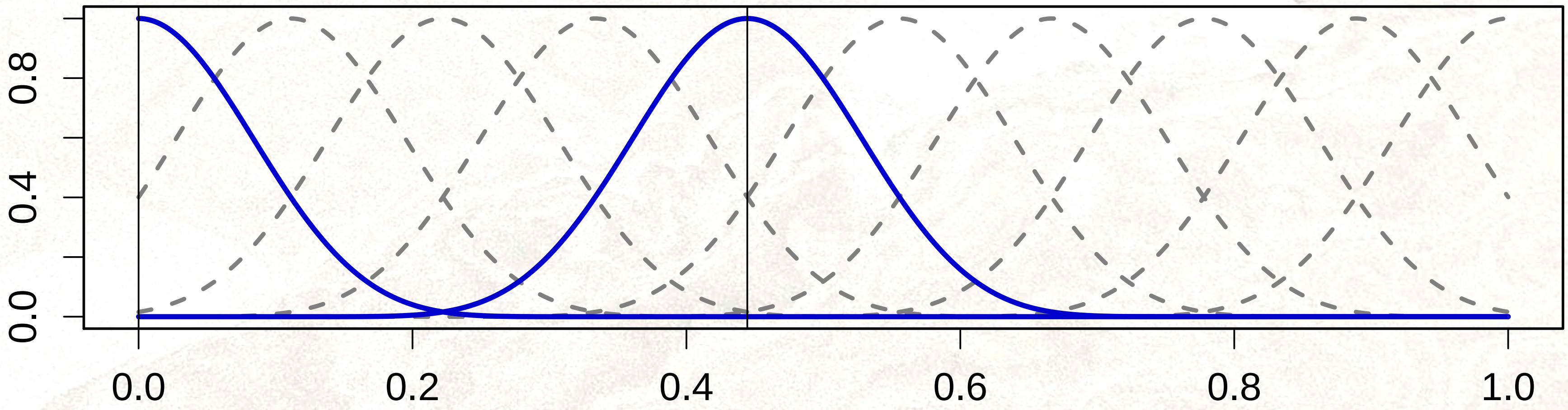
Example of basis functions



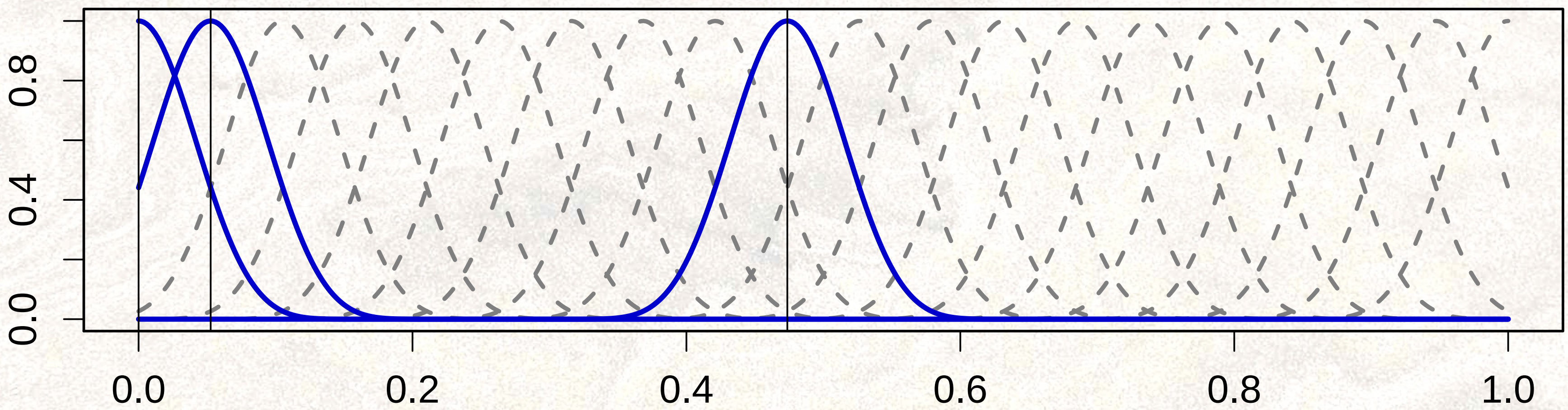
- Build a basis by translating and scaling a bump shaped curve
- Not your usual sine/cosine or polynomials!

Two Bases

10 Functions:



20 Functions:



Least squares.

$X_{i,j} = b_k(\boldsymbol{x}_i)$

$(\hat{g}(x_1), \hat{g}(x_2), \ldots, \hat{g}(x_n))^T = X\boldsymbol{\beta}$ and $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$

minimize over β:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} (\boldsymbol{y} - [X\boldsymbol{\beta}]_i)^2$$

Solution:

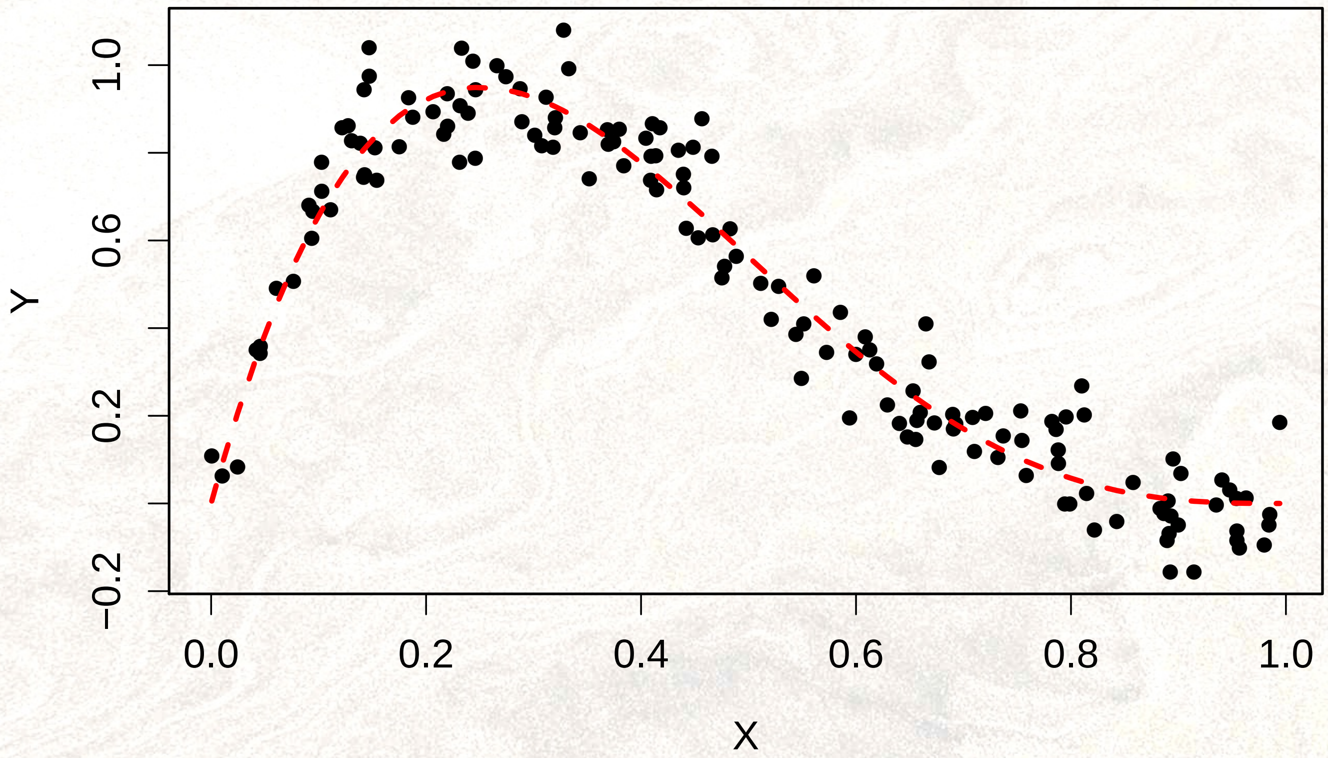
$$\hat{\beta} = (X^T X)^{-1} X^T \boldsymbol{y}$$

Spatial estimate:

$$\hat{g}(x) = \sum_{k=1}^{m} \hat{\beta}_k b_k(x)$$

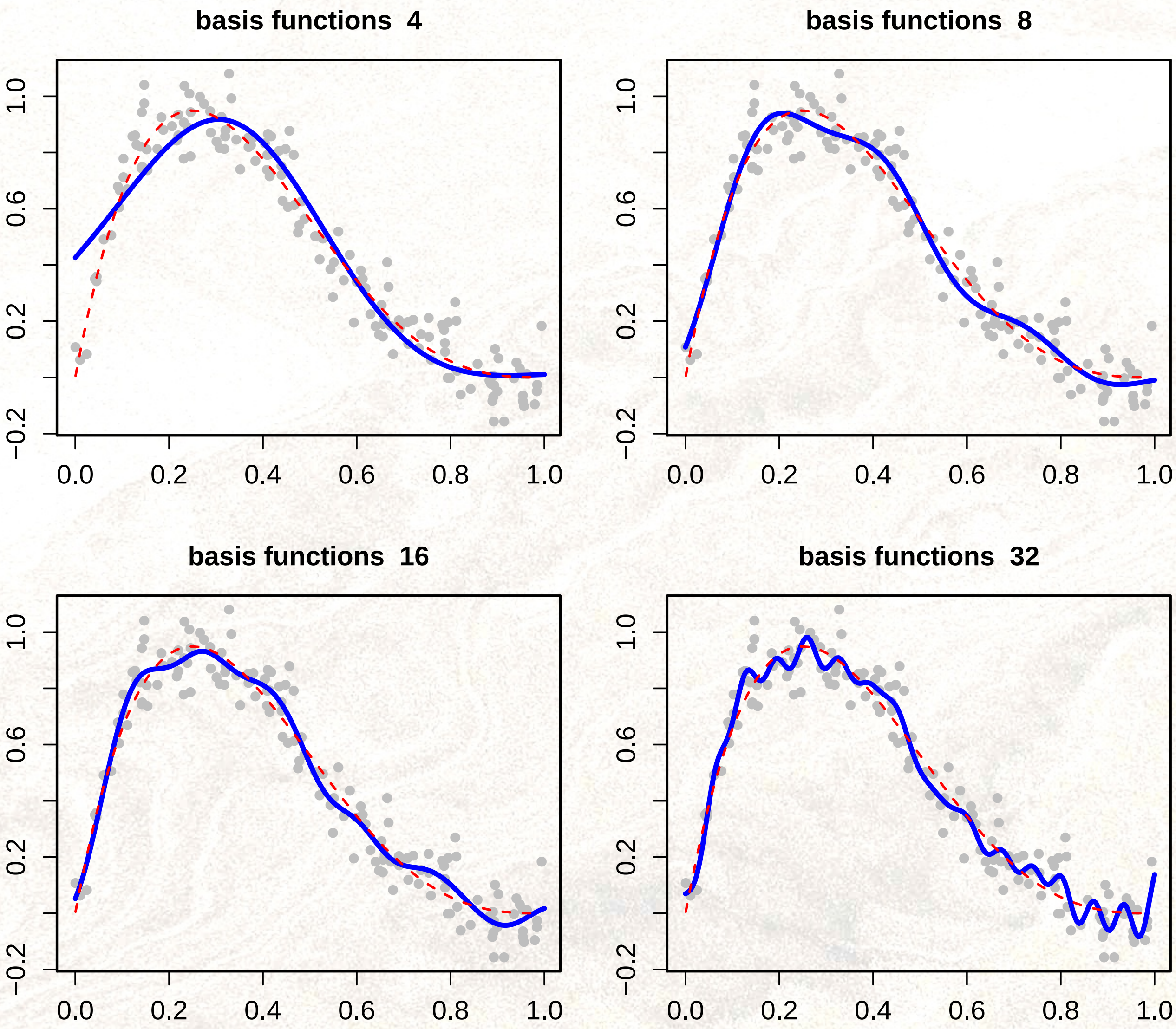
A specific example

Some synthetic data: $Y_k = h(x_k) + e_k$
$h(x) = 9x(1-x)^3$



x_k are 150 unequally spaced points in $[0, 1]$
$h(x) = 9x(1-x)^3$, $e_k \sim N(0, (.1)^2)$

Varying basis size



Problem: How to choose number of basis functions? Uncertainty of estimate?

Penalized least squares.

$g(x) = \sum_{k=1}^{m} \beta_k b_k(x)$

minimize over β:

Sum of squares($\boldsymbol{\beta}$) + penalty on $\boldsymbol{\beta}$

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} (\boldsymbol{y} - [X\boldsymbol{\beta}]_i)^2 + \lambda \boldsymbol{\beta}^T Q \boldsymbol{\beta}$$

Fit to the data + penalty for complexity/smoothness

- $\lambda > 0$ a smoothing parameter
- Q a nonnegative definite matrix.

More on Q

- Q is an identity $\boldsymbol{\beta}^T Q \boldsymbol{\beta} = \sum (\beta_l)^2$

- Choose Q so that $\boldsymbol{\beta}^T Q \boldsymbol{\beta} = \sum (\beta_l - \beta_{l-1})^2$

- Cubic smoothing spline type: $\boldsymbol{\beta}^T Q \boldsymbol{\beta} = \int (g''(x))^2 dx$

i.e. $Q_{kl} = \int b''_k(x) b''_l(x) dx$

- Spatial model: $Q^{-1}_{kl} = exp(-|x_k - x_l|/\theta)$

Solution to penalized problem

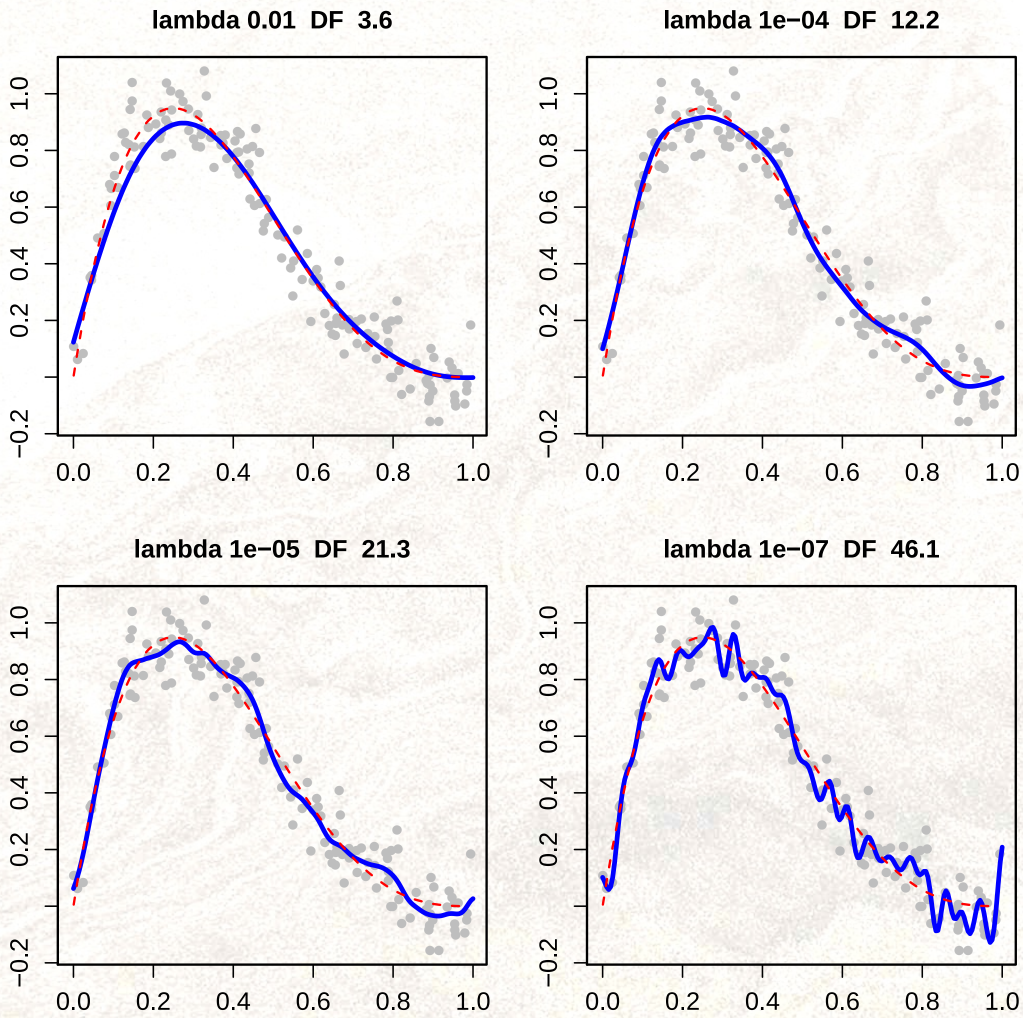
$$\hat{\beta} = (X^T X + \lambda Q)^{-1} X^T \boldsymbol{y}$$

Also known as ridge regression

Varying smoothing parameter

Cubic spline-type choice with 50 basis functions

Effective number of parameters (basis functions) depends on λ.



Problem: How to choose λ ? Uncertainty of estimate?

A Normal World

To describe $g(\boldsymbol{x})$ as a Gaussian process, start with a covariance function:

$$\rho k(\boldsymbol{x}_1, \boldsymbol{x}_2) = COV(g(\boldsymbol{x}_1), g(\boldsymbol{x}_2))$$

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.

Fill in the elements of the covariance matrix using the locations and the covariance function.

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

A simple covariance function: the exponential

$$\rho k(\boldsymbol{x}_1, \boldsymbol{x}_2) = \rho e^{-D/\theta}$$

D is the distance between the two locations $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$

In general:

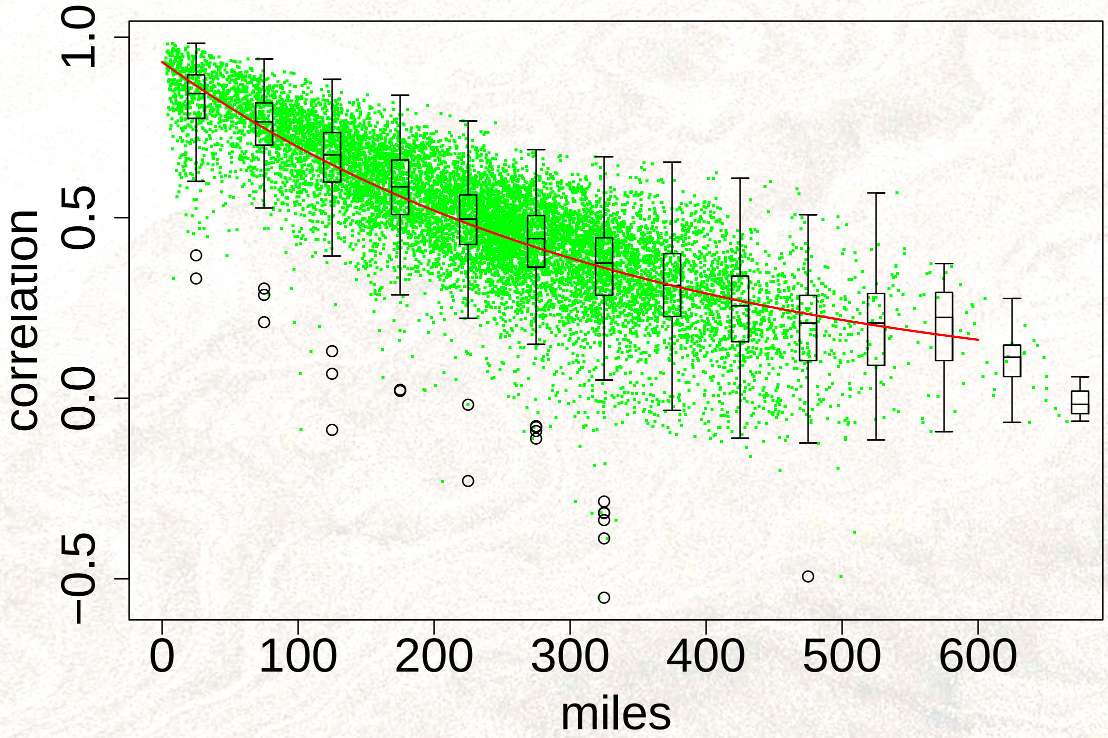
$$\rho k(\boldsymbol{x}_1, \boldsymbol{x}_2) = \rho e^{-||\boldsymbol{x}_1 - \boldsymbol{x}_2||/\theta}$$

This depends on two parameters: ρ and θ.

- VAR($g(\boldsymbol{x})) = \rho$
- COV($g(x_1)$,$g(x_2)$) falls to $\approx$ .36 of ρ when the distance is equal to θ.
- Correlations are just $e^{-||\boldsymbol{x}_1 - \boldsymbol{x}_2||/\theta}$

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.

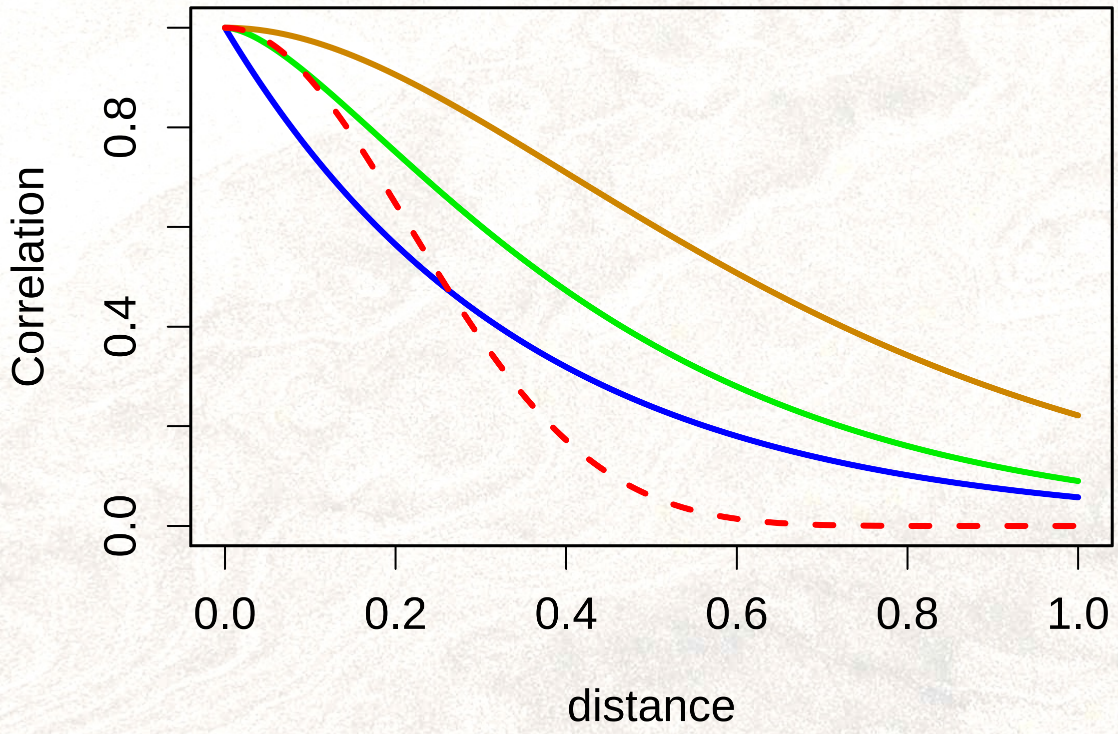


Families of correlation functions

Matern:

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

$\nu =$ .5, 1.0 , 2.0



- θ a range parameter
- ν smoothness at 0.
- ψ_ν is an exponential for $\nu = 1/2$ as $\nu \to \infty$ Gaussian.
- As ν increases the process is smoother.

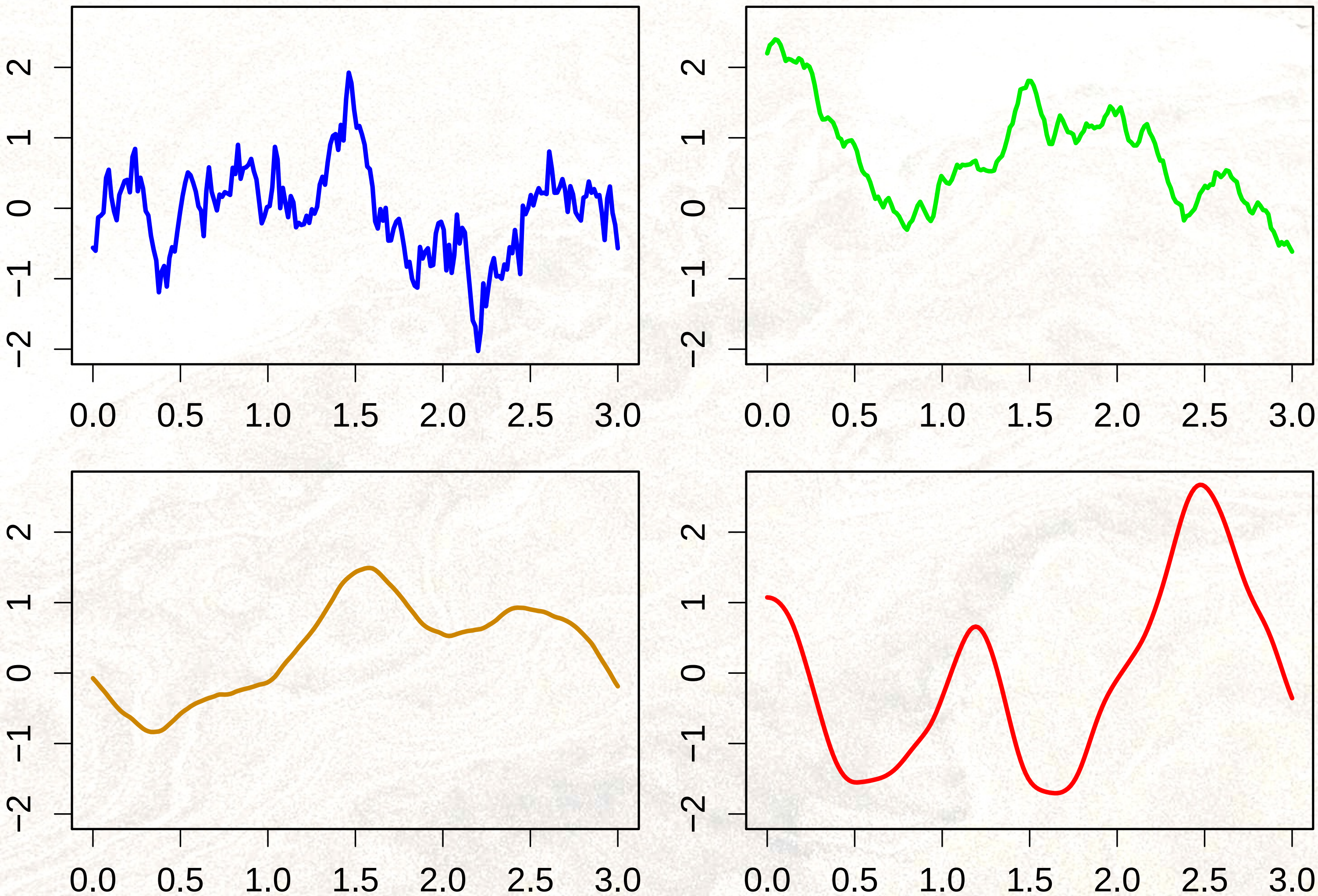
Wendland:

Polynomial that is exactly zero outside given range.

Compactly supported Wendland covariance (d=2, k=3)

What do these processes look like?

Matern (.5) Matern(1.0)

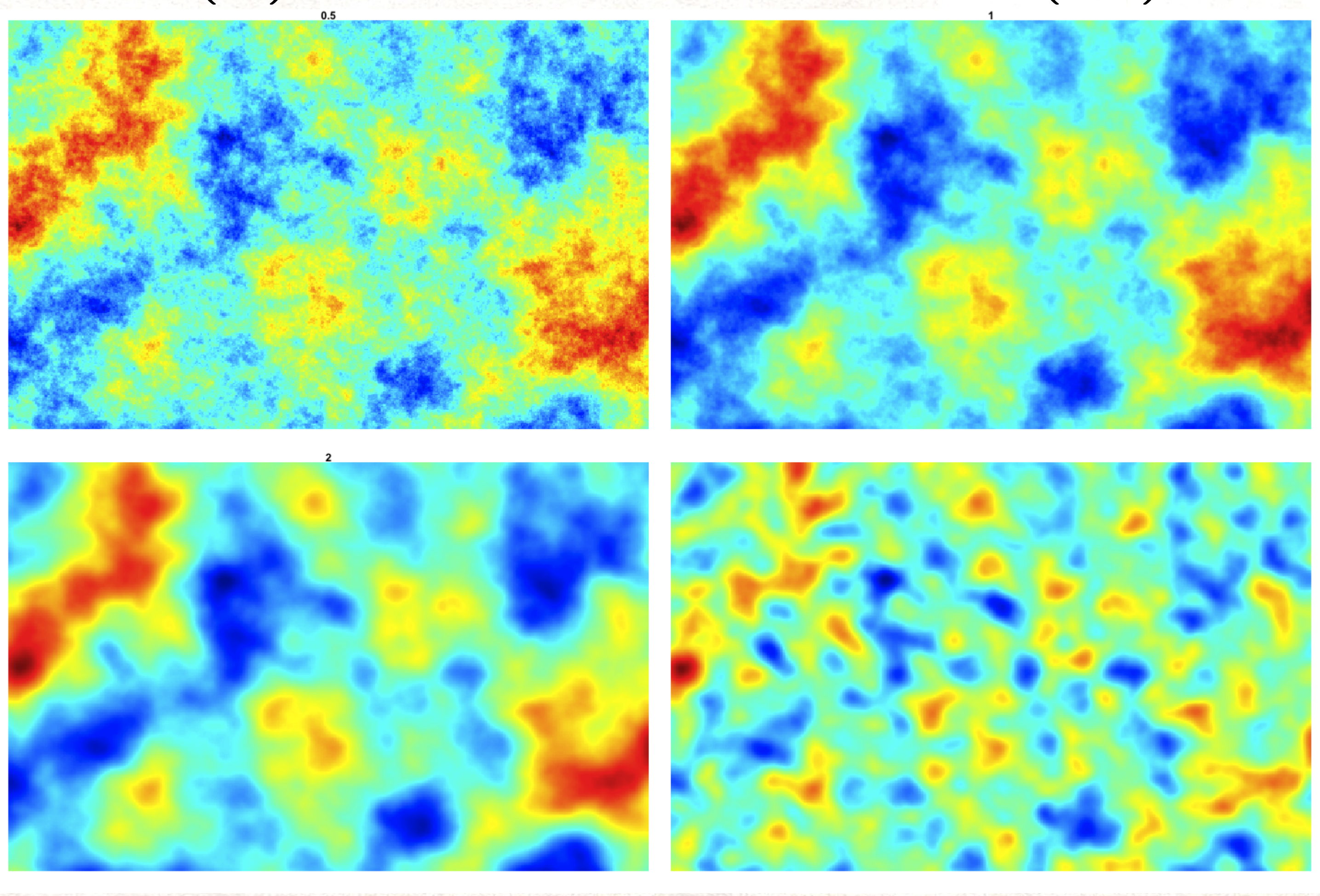


Matern (2.0) Wendland (2.0)

What do these processes look like?

Varying the smoothness:

Matern (.5) Matern(1.0)



Matern (2.0) Wendland (2.0)

A statistical model for spatial data

$$y_i = g(\boldsymbol{x}_i) + \epsilon_i$$

Observation = fixed component + spatial process + error

- ϵ_i's are uncorrelated $N(0, \sigma^2)$
- $g(x)$ mean zero process $VAR(g(\boldsymbol{x})) = \rho$
- Covariance function $\rho k_\theta(.,.)$ with θ some parameters.

Covariance matrix of g at observations: $K_{i,j} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$

Kriging

Assume that the covariance parameters are known.

Traditionally:

$\hat{g}$ is derived as the conditional expectation of g given $\boldsymbol{y}$

e. g. $\hat{g}(x) = E[g(x)|\boldsymbol{y}]$

Here is another way to view this estimator.

Kriging as penalized least squares:

Key insights are

- Basis functions come from covariance function $b_j(x) = k(x, \boldsymbol{x}_j)$
- penalty comes from inverse covariance matrix $Q = K^{-1}$

Find β by:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} (\boldsymbol{y}_i - [X\boldsymbol{\beta}]_i)^2 + \lambda \boldsymbol{\beta}^T Q \boldsymbol{\beta}$$

Fit to the data $+$ penalty for complexity/smoothness

$\lambda = \sigma^2/\rho$ and recall $K_{i,j} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$

Simplifying

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} (\boldsymbol{y}_i - [K\boldsymbol{\beta}]_i)^2 + \lambda \boldsymbol{\beta}^T K^{-1} \boldsymbol{\beta}$$

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

$\hat{g}(x) = \sum_j k(x, \boldsymbol{x}_j) \hat{\boldsymbol{\beta}}_j$

Uncertainty

Closed form formula for the standard error for $\widehat{g}(x)$ using

- the Gaussian process assumption for g
- the normal assumption for errors.

standard error for $\widehat{g}(x)$:

$$\rho(1 - \boldsymbol{u}(x)^T (K + \lambda I)^{-1} \boldsymbol{u}(x))$$

$$\boldsymbol{u}(x)^T = [k(x, x_1), k(x, x_2), \ldots, k(x, x_n)]$$

Likelihood for statistical parameters

$$\boldsymbol{y} \sim MN(0, (\rho K + \sigma^2 I))$$

Likelihood

$$\frac{1}{\sqrt{2\pi}^n} \quad e^{-\frac{1}{2}(\boldsymbol{y})^T(\rho K+\sigma^2 I)^{-1}(\boldsymbol{y})} \quad (\det \rho K + \sigma^2)^{-1/2}$$

log Likelihood

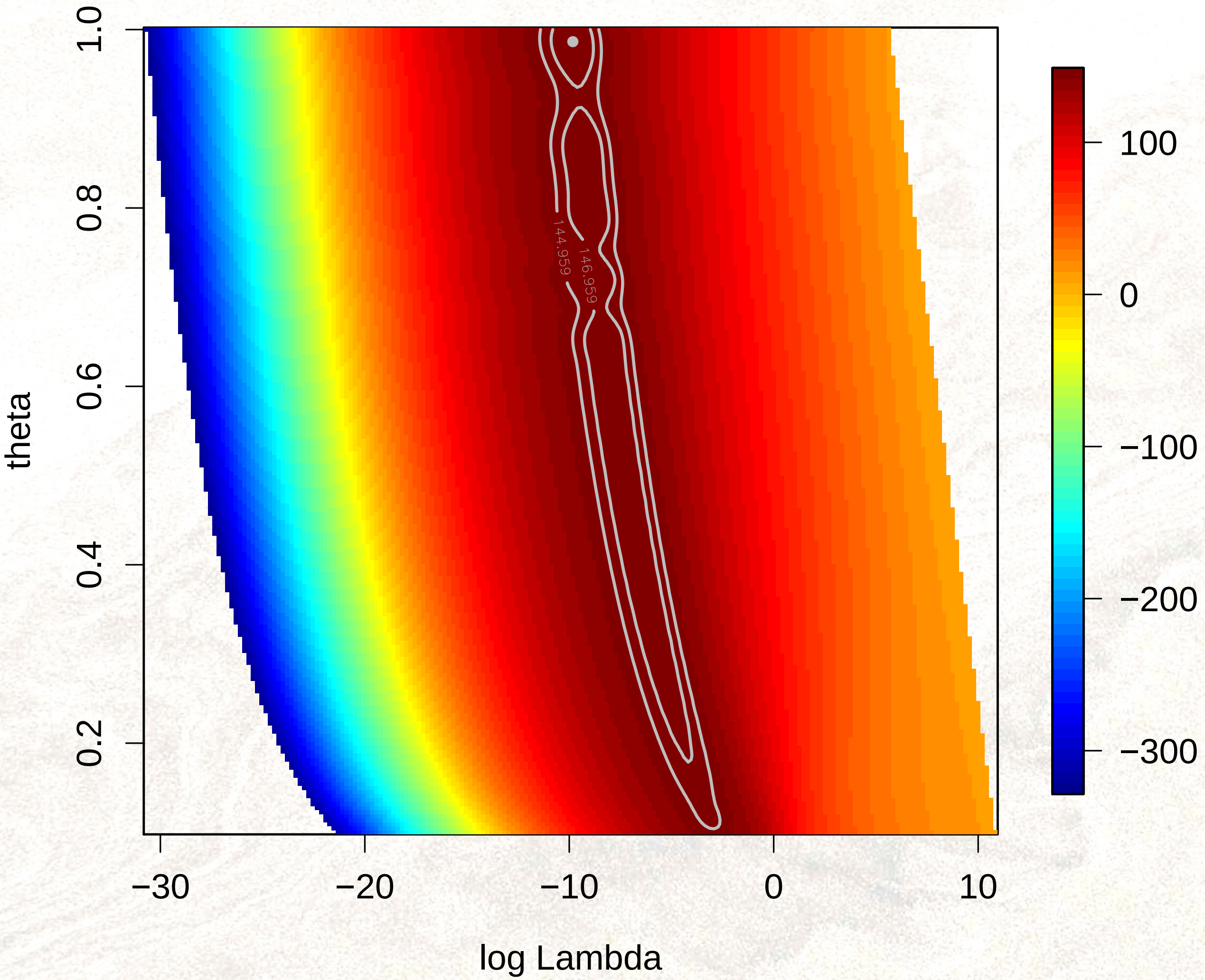
$$-\frac{1}{2}(\boldsymbol{y})^T(\rho K + \sigma^2 I)^{-1}(\boldsymbol{y}) - \frac{1}{2} log(\text{det}(\rho K + \sigma^2 I)) + \text{stuff}$$

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

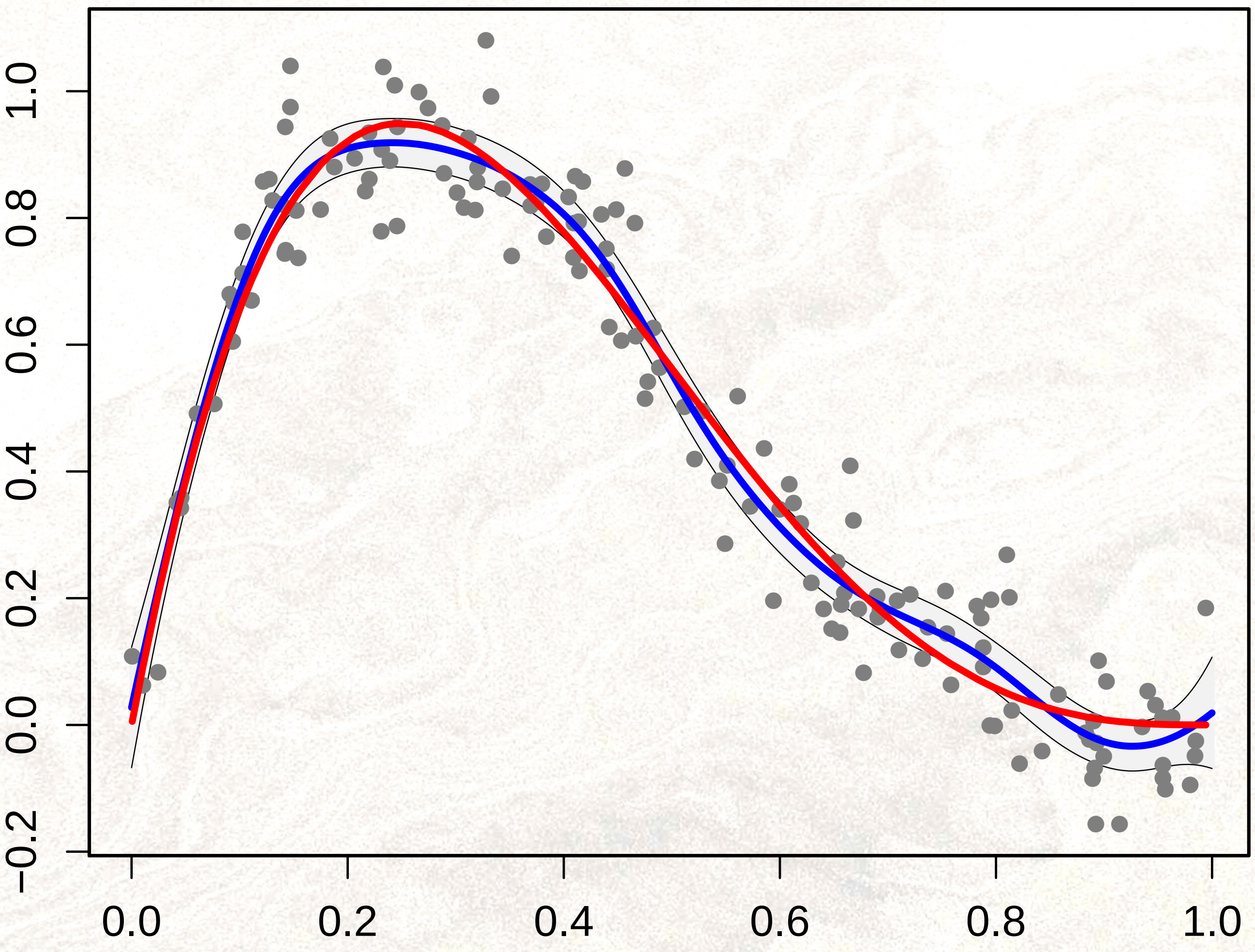
$$-\frac{1}{2\rho}(\boldsymbol{y}^T(K + \lambda I)^{-1}\boldsymbol{y}) - \frac{1}{2} log(\text{det}(K + \lambda I)) + -\frac{n}{2} log(\rho) + \text{stuff}$$

- Can maximize this analytically for ρ
- K may depend on other parameters e.g. scale, shape

log Likelihood for example



Curve estimate with uncertainty



Matern covariance
smoothness = 2, $\hat{\theta} = .98$, $\hat{\sigma} = .08$

Thank you!