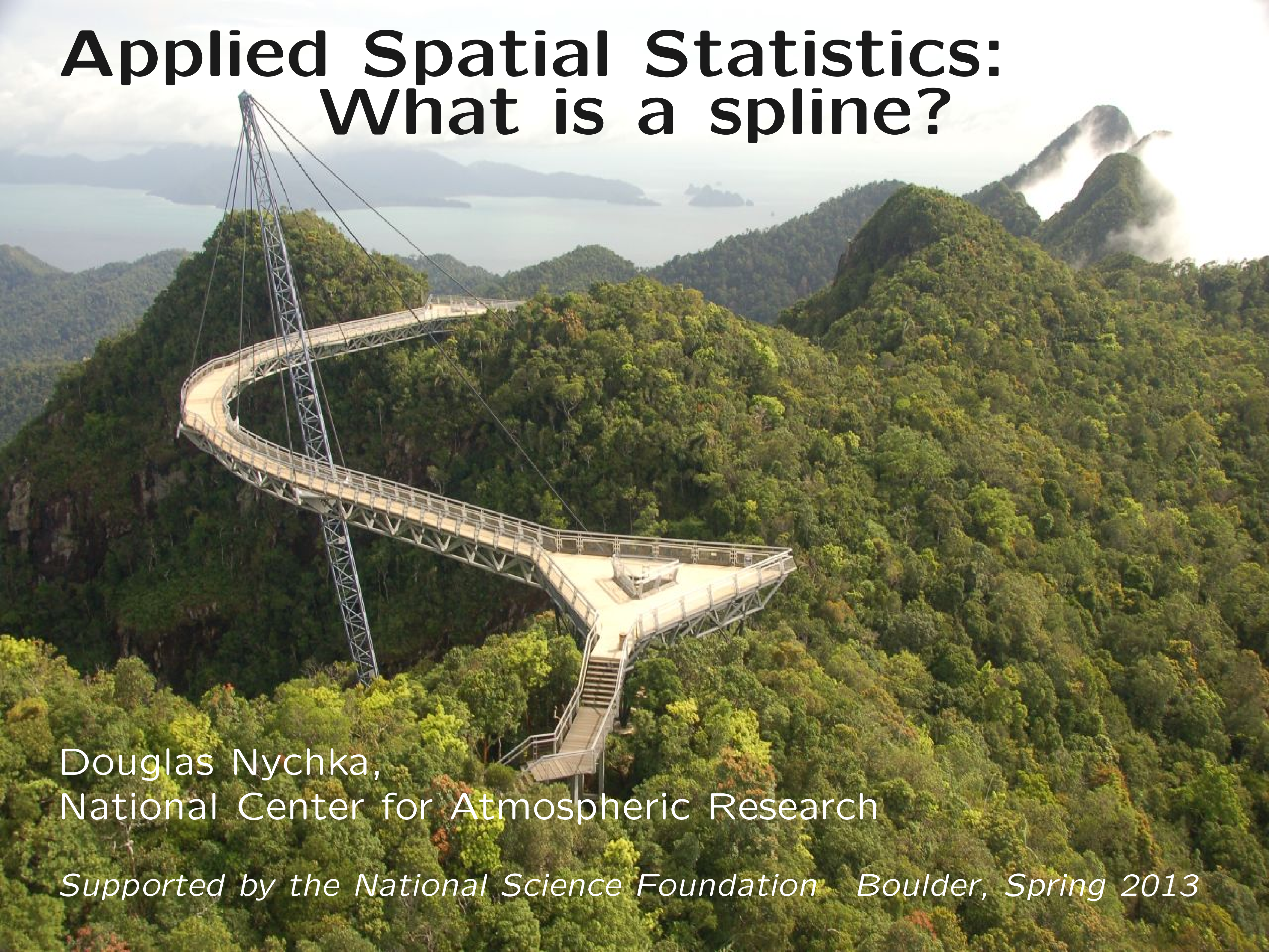


Applied Spatial Statistics: What is a spline?

Douglas Nychka,
National Center for Atmospheric Research

Supported by the National Science Foundation   *Boulder, Spring 2013*

Outline

- Cubic splines
- Reproducing kernels
- Thin plate splines
- Kriging is a type of spline.

The spline roughness penalty is equivalent to assuming a specific covariance function.

The Kriging covariance can be identified with a specific roughness penalty.

Estimating a curve or surface.

The 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 – usually $\boldsymbol{e} \sim (0, \sigma^2 I)$
and g is an unknown smooth function.

The goal is to estimate a function g based on the observations

Develop a statistical model for g that suggests an estimate and also its uncertainty

The classic cubic smoothing spline

Splines are the solutions to variational problems.

For curve smoothing in one dimension,

$$\min_f \sum_{i=1}^{n} (y_i - f(\boldsymbol{x}_i))^2 + \lambda \int (f''(x))^2 dx$$

The second derivative measures the roughness of the fitted curve.

In general:

$$\min_f \sum_{i=1}^{n} (y_i - f(\boldsymbol{x}_i))^2 + \lambda \mathcal{R}(f)$$

$\mathcal{R}(f) = \langle f, f \rangle \geq 0$.

Usually $\mathcal{R}$ depends on derivatives and integrals.

Easy to use in `fields`*:*

with λ found by GCV

```
obj<- sreg( x,y)
plot( x,y)
lines( obj$predicted)
```

or equivalently

```
obj<- Tps( x,y)
#
plot( x,y)
xg<- seq(min(x), max(x),,200)
fhat<- predict( obj,xg)
lines( xg, that)
```

In both cases default is to find λ by Generalized cross-validation (GCV).

Form of the cubic spline

$$\mathcal{R}(f) = \int (f''(x))^2 dx$$

$\hat{g}$ is continuous and with continuous first and second derivatives

It is a piecewise, cubic polynomial in between the observation points.

What does this have to do with spatial statistics?

Spline estimator

$$\hat{f}(\boldsymbol{x}) = \sum_j \phi_j(\boldsymbol{x}) d_j + \sum_i k(\boldsymbol{x}, \boldsymbol{x}_i) c_i$$

Null space:

$\{\phi_j\}$ basis functions where $\mathcal{R}(\phi_j) = 0$.
For 1-d cubic spline constant a linear functions.

Reproducing Kernel:

$k(.,.)$ a special covariance function that is derived from the roughness penalty.

Has the property that

$$\mathcal{R}(\sum_i k(\boldsymbol{x}, \boldsymbol{x}_i) c_i) = \boldsymbol{c}^T K \boldsymbol{c}$$

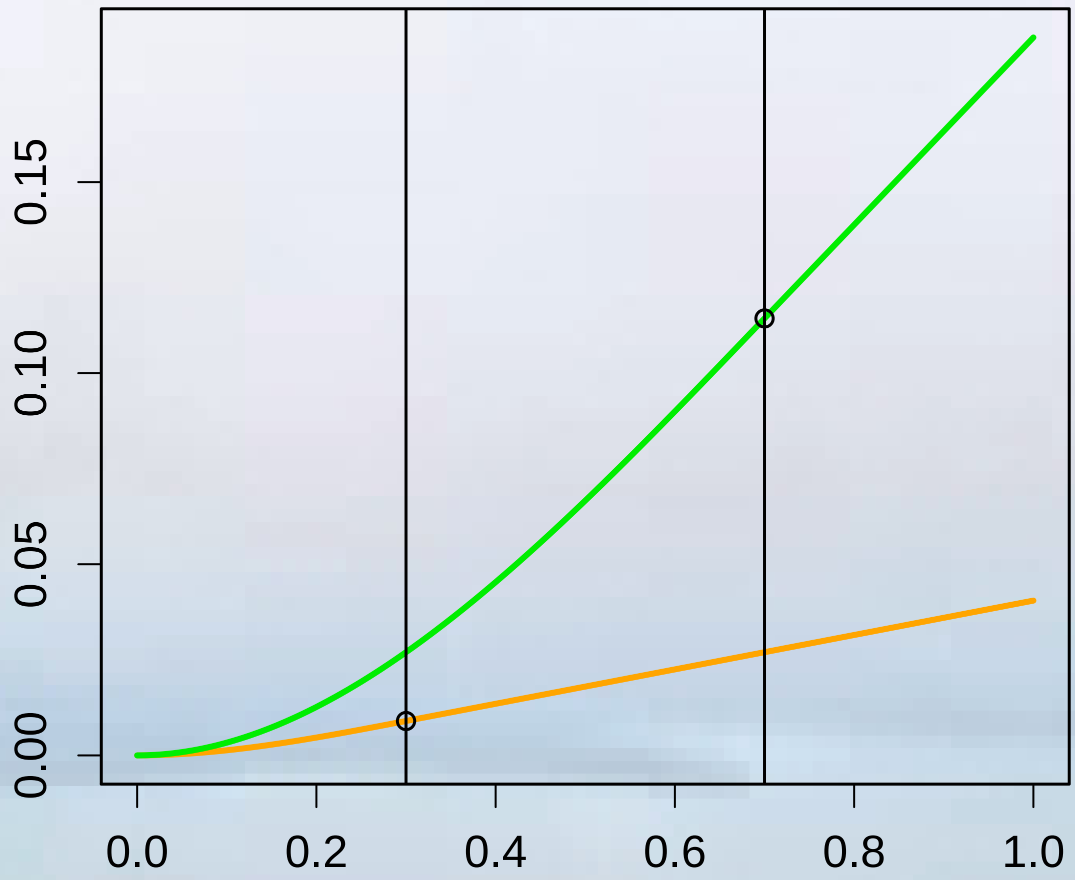
where $K_{i,j} = k(\boldsymbol{x}_i \boldsymbol{x}_j)$ for all $\boldsymbol{c}$.

RK for the cubic spline

$$k(u,v) = \begin{cases} u^2v/2 - u^3/6 & u < v \\ uv^2/2 - v^3/6 & u \geq v \end{cases}$$

Verify this using integration by parts and the boundary conditions at 0.

$k(u, .3)$ and $k(u, .7)$



Original variational problem:

$$\min_{f} \sum_{i=1}^{n} (y_i - f(\boldsymbol{x}_i))^2 + \lambda \mathcal{R}(f)$$

plugging in the form of the solution:

$$\min_{\boldsymbol{d},\boldsymbol{c}} \sum_{i=1}^{n} \left(y_i - \sum_k \phi(\boldsymbol{x}_i) \boldsymbol{d}_k + h_{\boldsymbol{c}}(\boldsymbol{x}_i) \right)^2 + \lambda \boldsymbol{c}^T K \boldsymbol{c}$$

In matrix/vector form:

$$\min_{\boldsymbol{d},\boldsymbol{c}} (\boldsymbol{y} - X\boldsymbol{d} - K\boldsymbol{c})^T (\boldsymbol{y} - X\boldsymbol{d} - K\boldsymbol{c}) + \lambda \boldsymbol{c}^T K \boldsymbol{c}$$

$X_{i,k} = p_k(x_i)$ regression matrix for polynomials.
$K_{i,j} = k(x_i, x_j)$ and $h_{\boldsymbol{c}}(\boldsymbol{x}) = \sum_i k(\boldsymbol{x}, \boldsymbol{x}_i) \boldsymbol{c}_i$

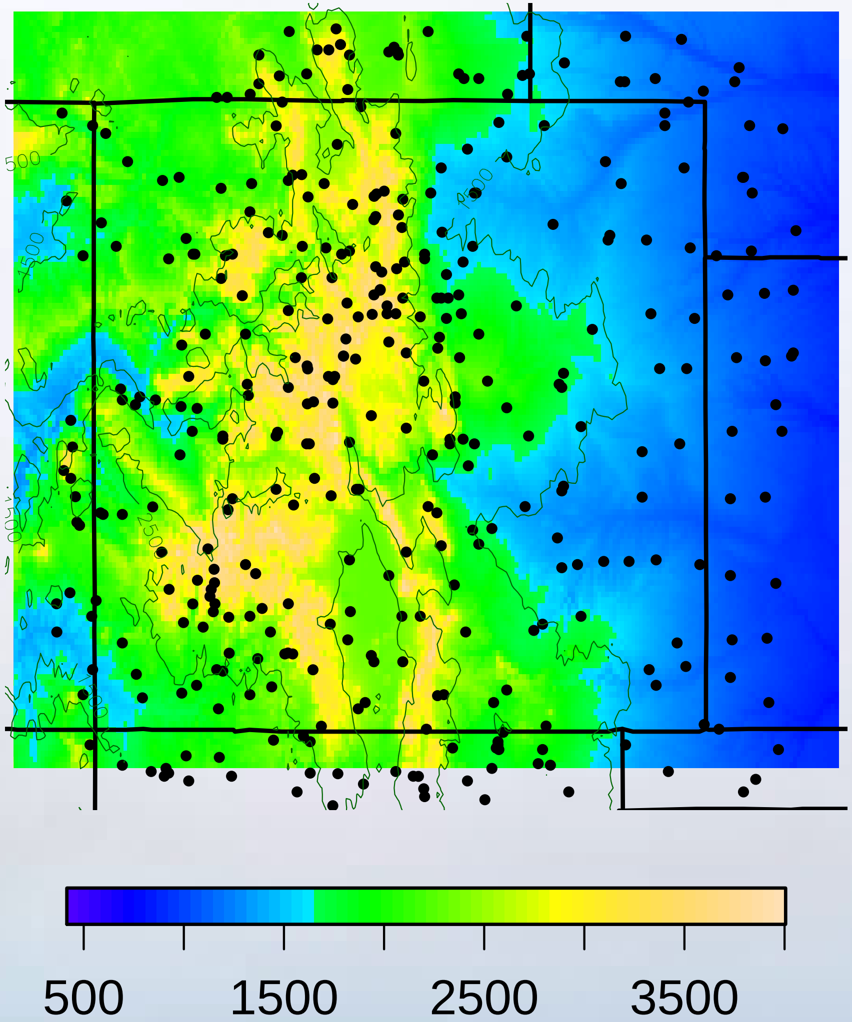
- Note that the polynomial part does not figure into $\mathcal{R}$.

- From HW3 we know that the minimization over $\boldsymbol{d}$ and $\boldsymbol{c}$ gives the Kriging estimate if k is the covariance function.

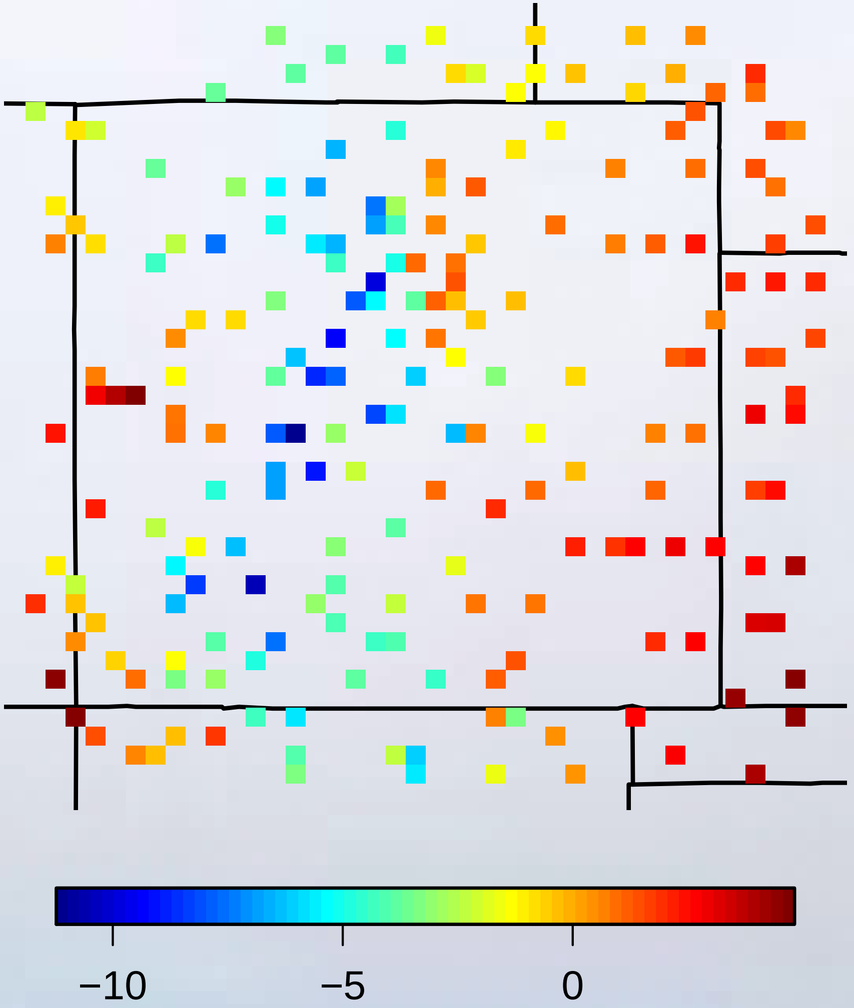
- λ is interpreted as σ^2/ρ from the Kriging perspective.

Climate for Colorado

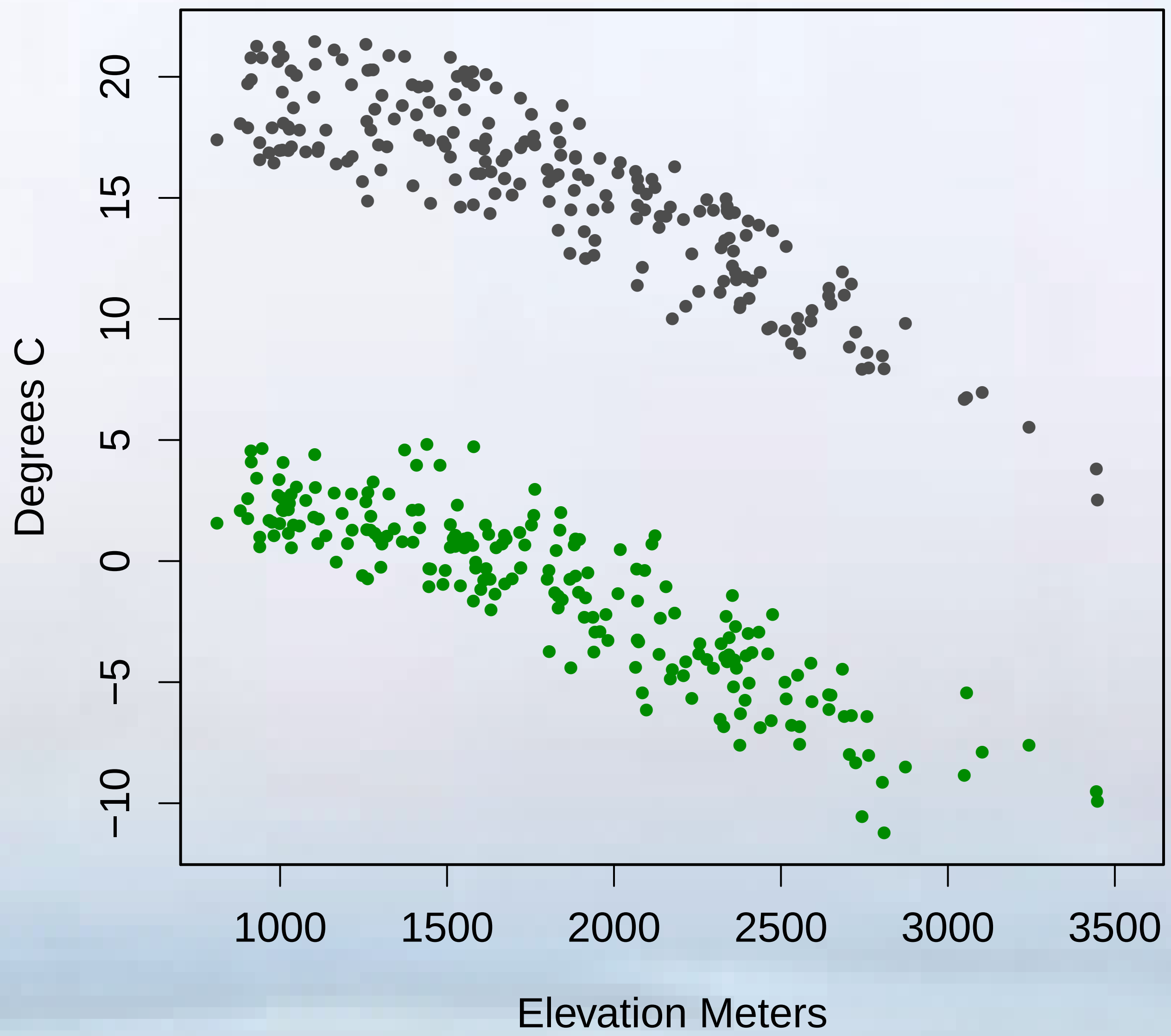
Elevations



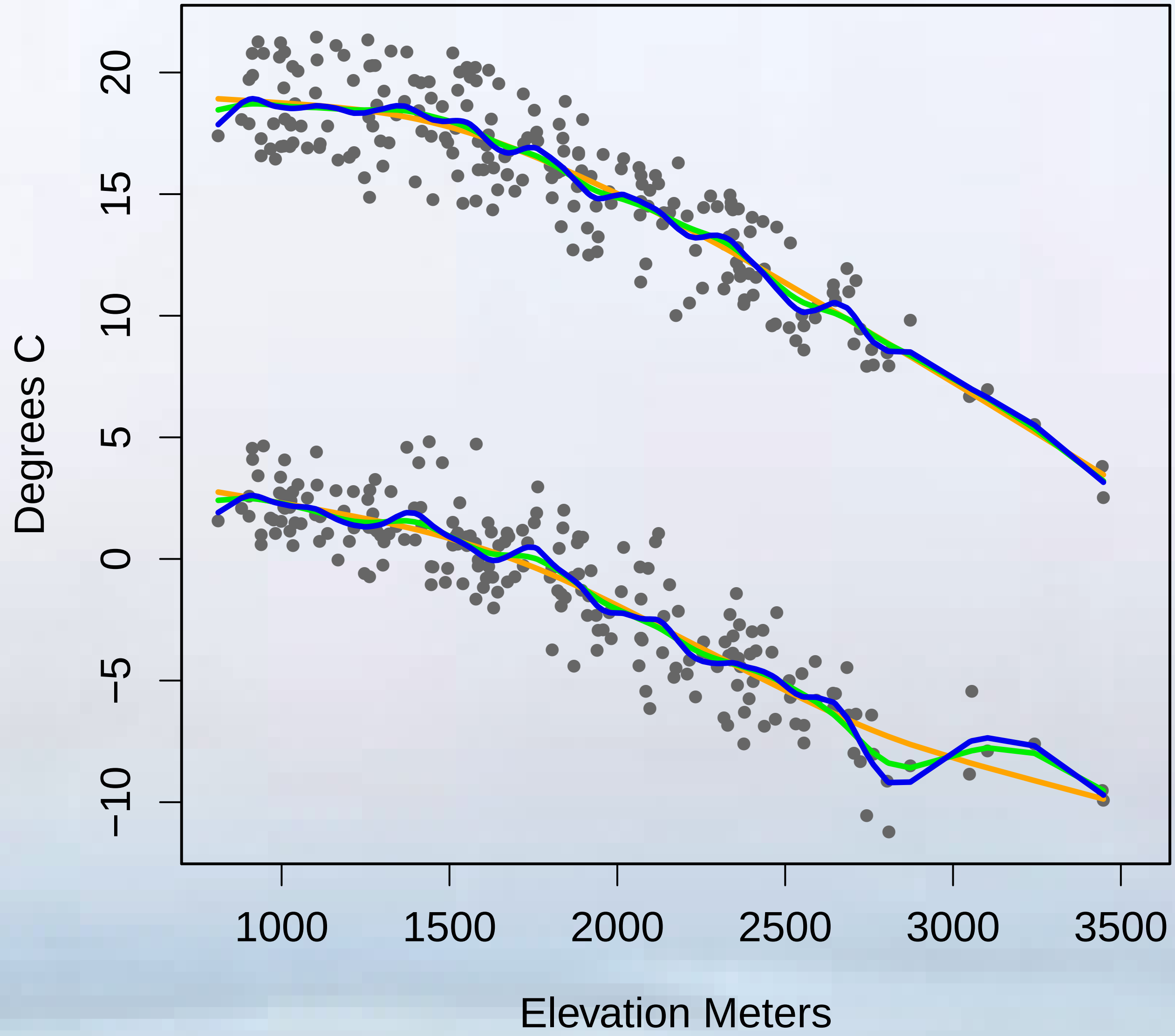
Spring average daily max temperatures



Max/Min spring temperatures



Cubic splines with different λ



Cross-validation

The magic formula

Recall for the full data set there is an "A" matrix

$$\begin{bmatrix} \hat{f}(\boldsymbol{x}_1) \\ \hat{f}(\boldsymbol{x}_2) \\ \vdots \\ \hat{f}(\boldsymbol{x}_n) \end{bmatrix} = A(\lambda)\boldsymbol{y}$$

Leave-one-out CV

$\hat{f}_{-i}$ the prediction at $\boldsymbol{x}_i$ having omitted $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ from the data set.

CV residual

residual for $\hat{f}(x_i)$ having omitted y_i is

$$(y_i - \hat{f}_{-i}) = (y_i - \hat{f}_i)/(1 - A(\lambda))_{i,i}$$

This has a simple form because adding the data pair $(\boldsymbol{x}_i, \hat{f}_{-1})$ to the data does not change the sums of squares.

CV and Generalized CV criterion

$CV(\lambda)$

$$(1/n)\sum_{i=1}^{n}(y_i - \widehat{g}_{-i})^2 = (1/n)\sum_{i=1}^{n}\frac{(y_i - \widehat{g}_i)^2}{(1 - A(\lambda)_{i,i})^2}$$

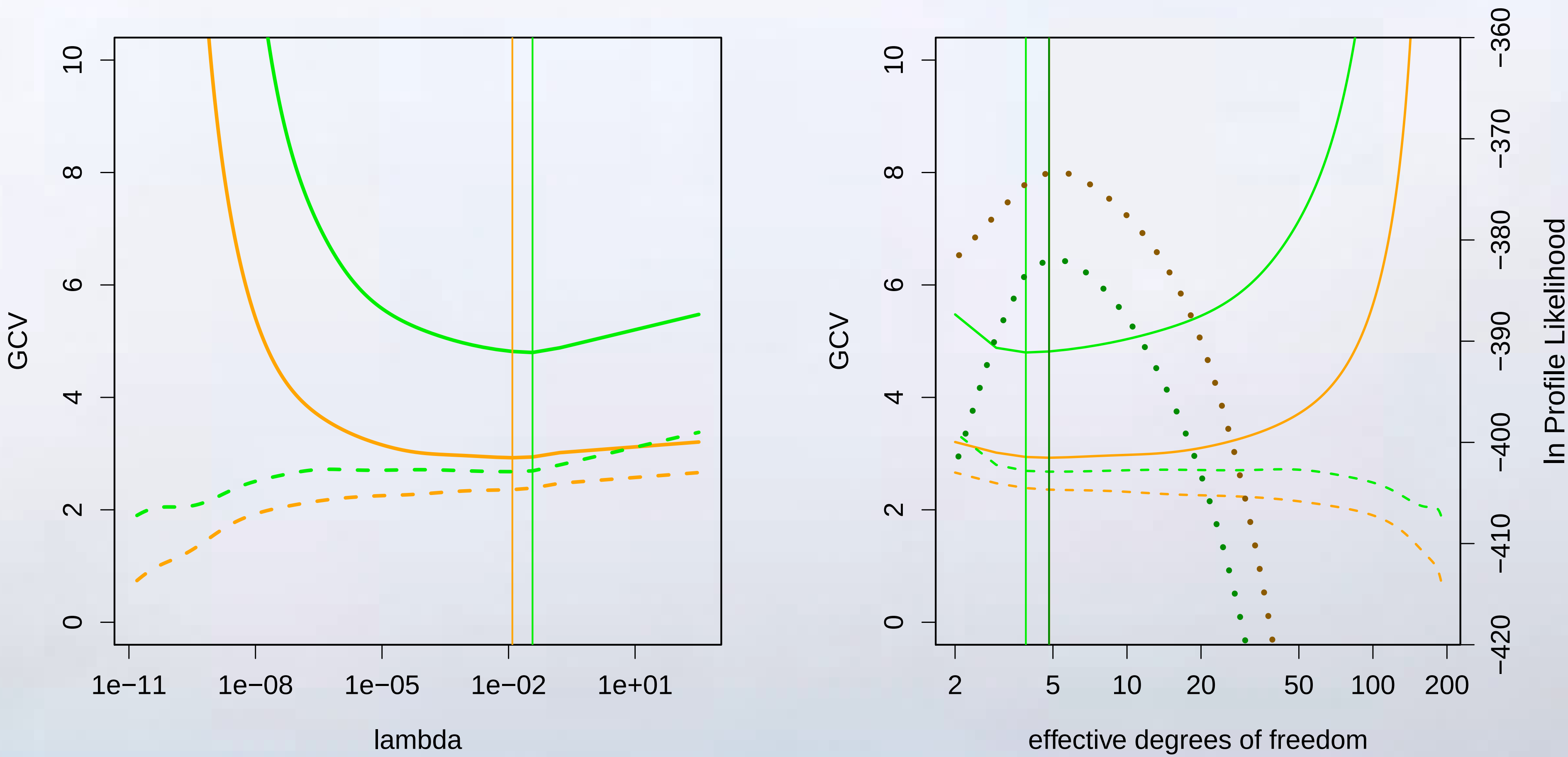
$GCV(\lambda)$ $A(\lambda)_{i,i} \approx \mathrm{tr}A(\lambda)/n$

$$(1/n)\frac{\sum_{i=1}^{n}(y_i - \widehat{g}_i)^2}{(1 - \mathrm{tr}A(\lambda)/n)^2}$$

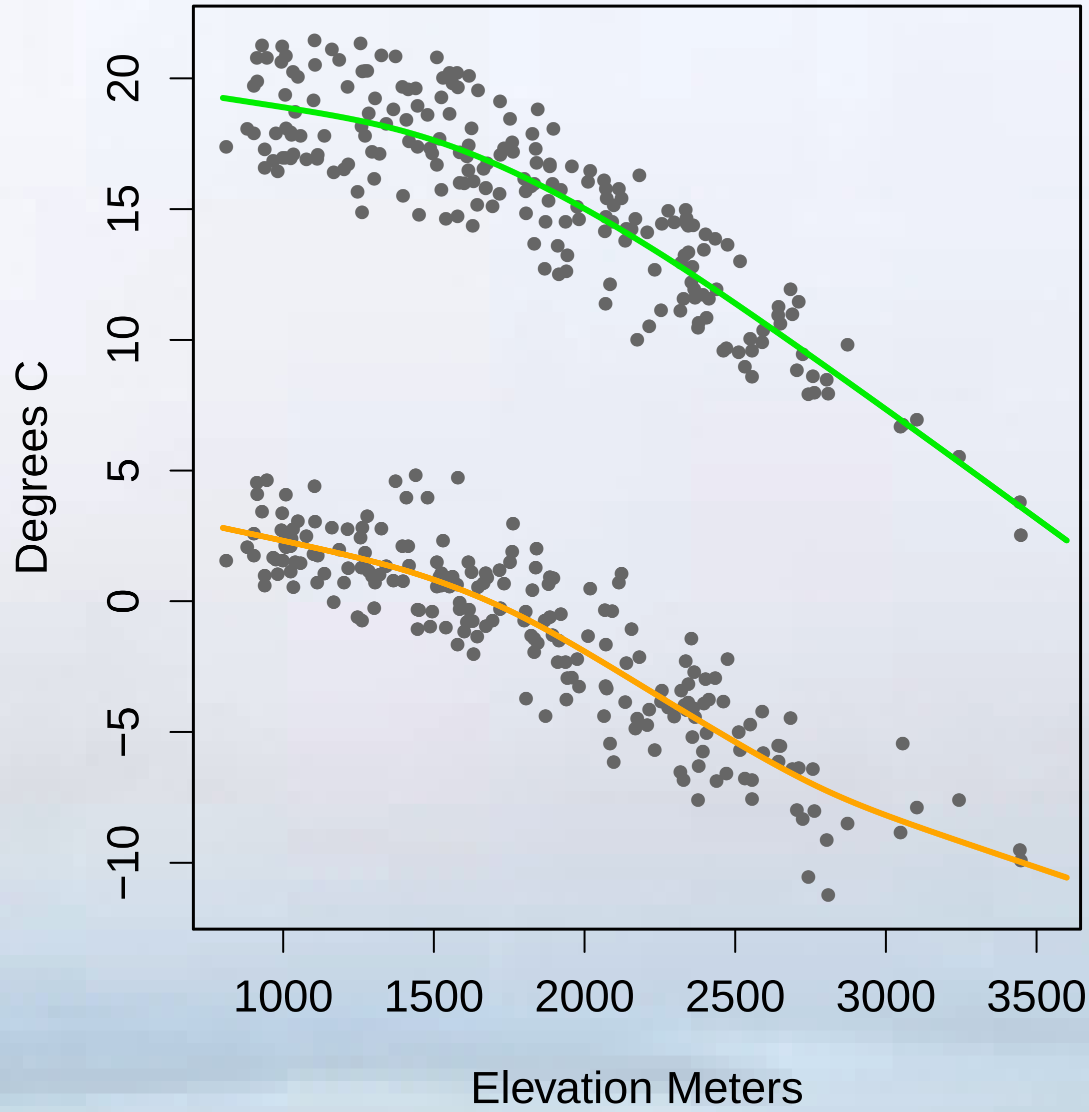
Minimize CV or GCV over λ to determine a good value

GCV for the temperature data

min max dashed – estimate $\hat{\sigma}^2$



GCV estimated curves



A 2-d thin plate smoothing spline

$$\min_f \sum_{i=1}^{n} (y_i - f_i)^2 + \lambda \int_{\Re^2} \left(\frac{\partial^2 f}{\partial^2 u}\right)^2 + 2\left(\frac{\partial^2 f}{\partial u \partial v}\right)^2 + \left(\frac{\partial^2 f}{\partial^2 v}\right)^2 du dv$$

Collection of second partials is invariant to a rotation.

The linear part of f is in the null space

$$\min_f \sum_{i=1}^{n} (y_i - f(\boldsymbol{x}_i))^2 + \lambda \mathcal{R}(f)$$

Thin plate reproducing kernel:

$$k(\boldsymbol{x}, \boldsymbol{x}') = ||\boldsymbol{x} - \boldsymbol{x}'||^2 log(||\boldsymbol{x} - \boldsymbol{x}'||) + \text{linear terms}$$

Harder to show this works – but it does

Null space:

span of $1, x_1, x_2$

In General

The thin plate spline in d dimensions and derivative order m.

Null space:

All polynomials up to degree $m-1$

Reproducing kernel, $k(x, x')$:

d - even $||\boldsymbol{x} - \boldsymbol{x}'||^{2m-d} log(||\boldsymbol{x} - \boldsymbol{x}'||) +$ polynomial terms

d - odd $||\boldsymbol{x} - \boldsymbol{x}'||^{2m-d} +$ polynomial terms

Once the basis and penalty matrix are found this is just a ridge regression!

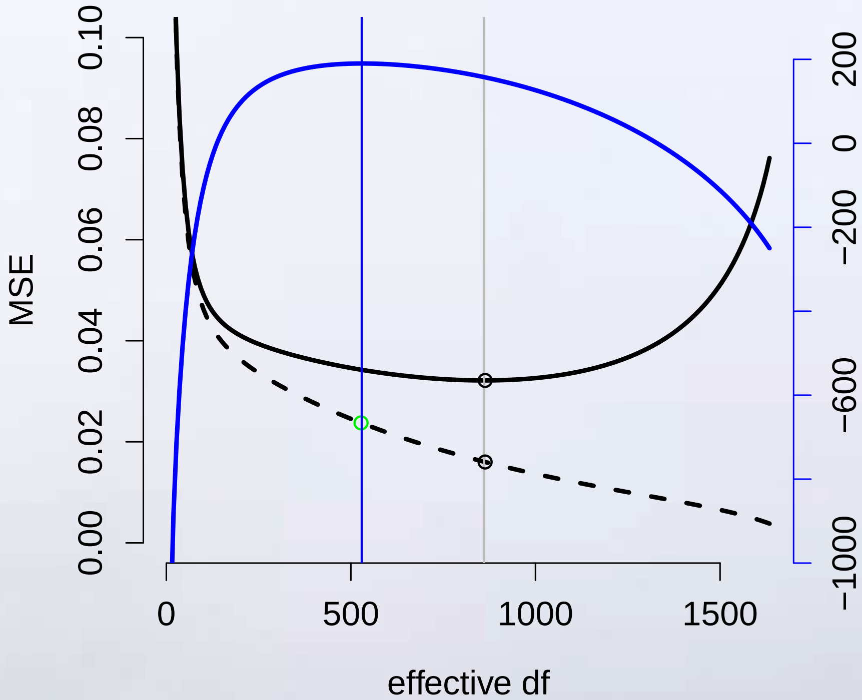
defining a spline estimate

Inner product $\rightarrow$ Reproducing kernel $\rightarrow$ funky ridge regression.

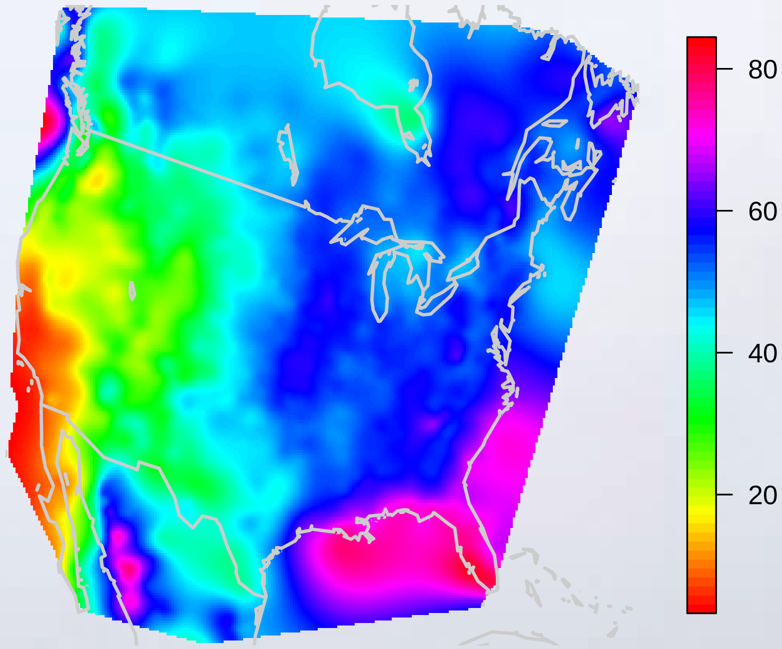
The spline implicitly assumes a spatial process model with the reproducing kernel as the covariance function.

Tps in action

GCV function and estimate of σ^2
and profile log likelihood



Fitted surface using GCV



In `fields` (no elevation fixed effect).

```
obj<- Tps( x,y)
surface( obj)
```

What about Kriging?

There is an equivalence between

roughness penalties $\Longleftrightarrow$ reproducing kernels

reproducing kernels $\Longleftrightarrow$ covariance functions

Kriging $\rightarrow$ spline with the roughness penalty determined by the covariance function.

but the penalty may be strange ...

Spline $\rightarrow$ Kriging estimate with the covariance determined by the roughness penalty.

but the covariance may be strange ...