Kriging

Douglas Nychka,
National Center for
Atmospheric Research

Lecture 3 of Applied Spatial Statistics

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

Outline

- The additive model
- Prediction variance for linear estimator
- Kriging estimator and standard error
- Additional components in model

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

Find the estimator that minimizes prediction variance based on a covariance model for g.

Prediction error, some ingredients

Predict $g(\boldsymbol{x}^*)$, $\boldsymbol{x}^*$ an arbirtrary location

A linear estimator:

- $\widehat{g}(\boldsymbol{x}^*) = \sum_{i=1}^{n} \boldsymbol{w}_i \boldsymbol{y}_i = \boldsymbol{w}^T \boldsymbol{y}$

Covariances for data:

- $K_{i,j} = COV(g(\boldsymbol{x}_i), g(\boldsymbol{x}_j))$

Covariance matrix for observations: $COV(\boldsymbol{y}) = K + \sigma^2 I$

Covariance between observed locations and $g(\boldsymbol{x}^*)$:

- $\boldsymbol{k}_i^* = COV(\boldsymbol{g}(\boldsymbol{x}_i), g(\boldsymbol{x}^*)) = k(\boldsymbol{x}^*, \boldsymbol{x}_i)$,

  $\boldsymbol{k}^*$ a vector with length n.

For example if g has a standard exponential covariance

$$K_{i,j} = e^{-||\boldsymbol{x}_i - \boldsymbol{x}_j||} \qquad \boldsymbol{k}_j^* = e^{-||\boldsymbol{x}^* - \boldsymbol{x}_j||}$$

Adding random vectors

Adding together two vectors with different weights and sizes

$$U \sim (\mu_U, \Sigma_U) \text{ and } V \sim (\mu_V, \Sigma_V)$$

$COV(U,V) = \Sigma_{U,V}$

$$Z = A_U U + A_V V$$

$Z \sim ($

$A_U \mu_U + A_V \mu_V,$

$A_U \Sigma_U A_U^T + A_U \Sigma_{U,V} A_V^T + A_V \Sigma_{V,U} A_U^T + A_V \Sigma_V A_V^T$

$)$

Red terms are equal.

$$COV(Z) = A_U \Sigma_U A_U^T + 2 A_U \Sigma_{U,V} A_V^T + A_V \Sigma_V A_V^T$$

Variance of the error

$Z = \hat{g}(\boldsymbol{x}^*) - g(\boldsymbol{x}^*) = \boldsymbol{w}^T \boldsymbol{y} - g(\boldsymbol{x}^*)$

$A_U \equiv \boldsymbol{w}^T$, $U \equiv \boldsymbol{y}$, $A_V \equiv -1$ and $V = g(\boldsymbol{x}^*)$

Mean

Both g and $\boldsymbol{y}$ have expected values of 0 so Z has expected value of zero too.

Variance

$\Sigma_U = K + \sigma^2 I$, $\Sigma_V = VAR(g(\boldsymbol{x}^*))$, and $\Sigma_{UV} = \boldsymbol{k}^*$

Plugging in everything:

$$VAR(Z) = \boldsymbol{w}^T (K + \sigma^2 I) \boldsymbol{w} - 2\boldsymbol{w}^T \boldsymbol{k}^* + VAR(g(\boldsymbol{x}^*))$$

Minimizing the error variance

Take the derivative with respect to $\boldsymbol{w}$, set equal to zero and solve for minimum.

One obtains

$$(K + \sigma^2 I)\boldsymbol{w} = \boldsymbol{k}^*$$

so the minimum is at

$$\boldsymbol{w}^* = (K + \sigma^2 I)^{-1}\boldsymbol{k}^*$$

Kriging estimate

$$\widehat{g}(\boldsymbol{x}^*) = (\boldsymbol{k}^*)^T (K + \sigma^2 I)^{-1} \boldsymbol{y}$$

In terms of basis functions

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

$$\widehat{g}(\boldsymbol{x}^*) = \sum_{i=1}^{n} k(\boldsymbol{x}^*, \boldsymbol{x}_i) \widehat{\mathbf{c}}_i$$

Predictions at the observation locations

$$\begin{bmatrix} \widehat{g}(\boldsymbol{x}_1) \\ \widehat{g}(\boldsymbol{x}_2) \\ \vdots \\ \widehat{g}(\boldsymbol{x}_n) \end{bmatrix} = K (K + \sigma^2 I)^{-1} \boldsymbol{y} \tag{1}$$

Predicting at the observation locations

- K is a positive definite matrix so $K = QDQ^T$ where D is diagonal and positive and Q is orthogonal ($Q^TQ = I$).

- It is easy to show $K(K + \sigma^2 I)^{-1} = Q(D(D + I)^{-1})Q^T$ so this matrix has eigenvalues $\{D_i/(D_i + \sigma^2)\}$ all in the interval $(0, 1)$.

- $K(K + \sigma^2 I)^{-1}$ has a smoothing property in that

$$||K(K + \sigma^2 I)^{-1}\boldsymbol{y}||^2 < ||\boldsymbol{y}||^2$$

for all $\boldsymbol{y}$. The sum of squares of the predicted values is always less than the sum of squares of the observations. This is a property of the eigenvalues all being in $(0, 1)$.

Minimum prediction variance

Evaluate the variance with this best weight.

Recall:

$$VAR(Z) = \boldsymbol{w}^T(K + \sigma^2 I)\boldsymbol{w} - 2\boldsymbol{w}^T\boldsymbol{k}^* + VAR(g(\boldsymbol{x}^*))$$

$$\boldsymbol{w} = (K + \sigma^2 I)^{-1}\boldsymbol{k}^*$$

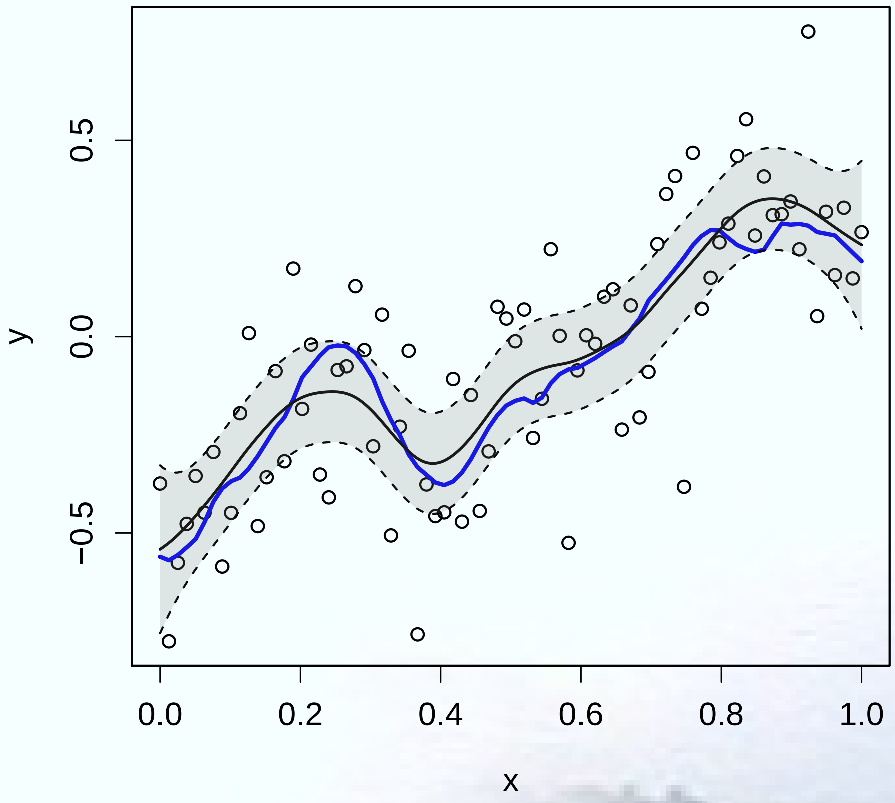
Kriging variance

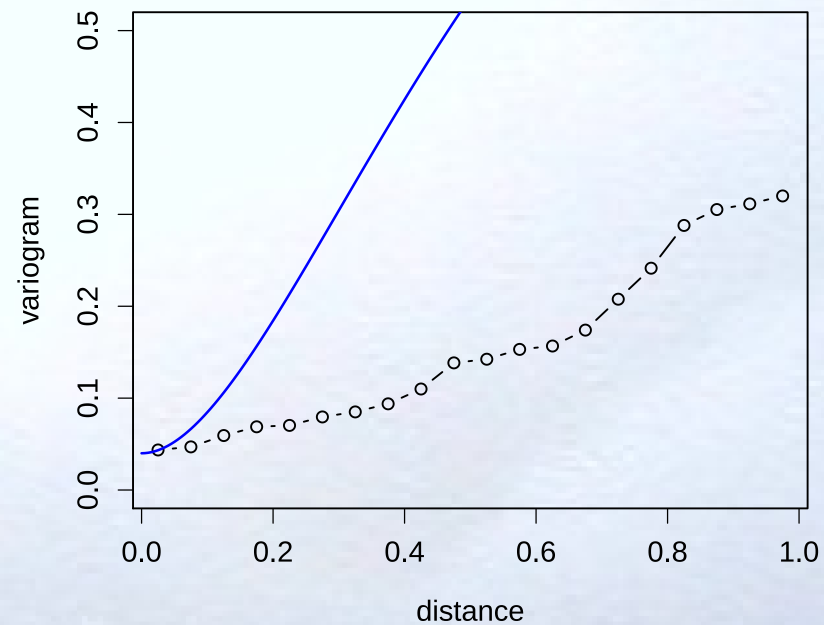
$$VAR(g(\boldsymbol{x}^*)) - (\boldsymbol{k}^*)^T(K + \sigma^2 I)^{-1}\boldsymbol{k}^*$$

A synthetic 1-d example

- Matern covariance smoothness of 1.5, scale is .3, sill is 1.0
- nugget variance is $.2^2$.

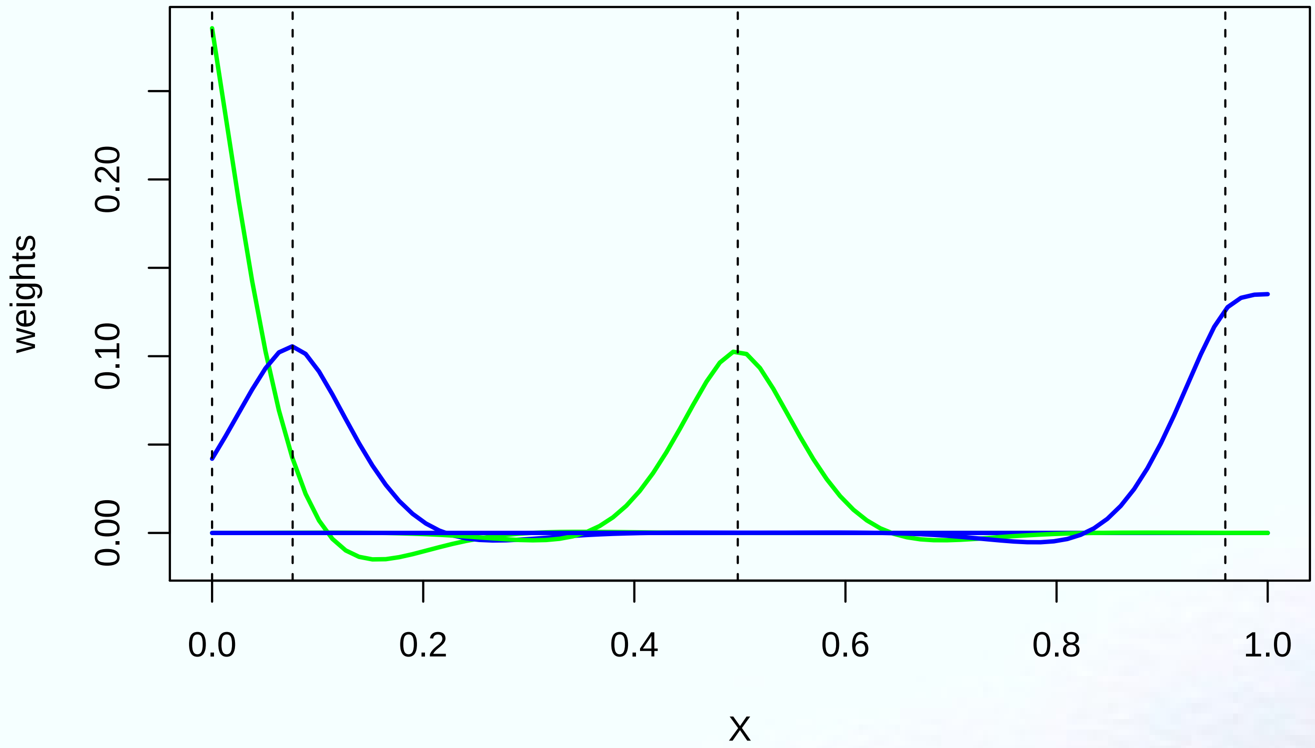
True function, Kriging estimate, and $\pm$ 2 **standard errors** Variogram





The weight function

Weight function applied to predict at different points



In R: how to do it.

Simulate the data x, y ...

```
K<- Matern( rdist( x,x)/.3 , smoothness=1.5)
xgrid<- seq( 0,1,,250)
Kstar<- Matern(rdist( xgrid, x)/.3, smoothness=1.5)
ghat<- Kstar %*% solve( K + diag( sigma^2, nobs)) %*% y
```

Note use of "vectorizing" the computations

Additive models

Spatial models usually have other components besides a mean zero spatial process.

Observation =
linear function of spatial coordinates

+ function of other variables

+ spatial process

+ error

CO climate data

Try linear regression using lon/lat and elevation (`HW1x` and use `CO.elev`)

```
elevHW1<- CO.elev[good]
X<- cbind( as.matrix(xHW1) , elevHW1)
out<- lm( yHW1 ~ X)
```

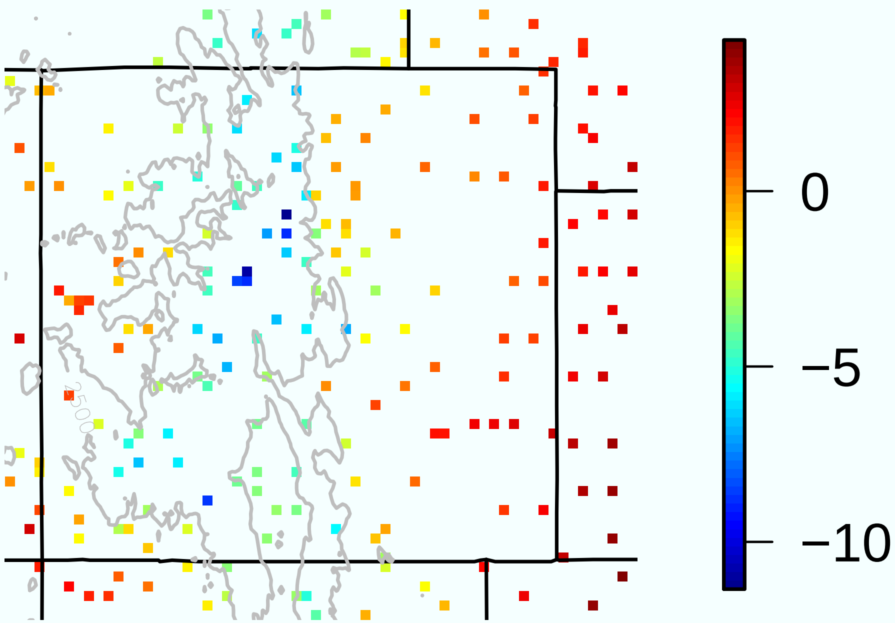
regression summary

```
summary( out)
  Call:
lm(formula = yHW1 ~ X)

Residuals:
    Min      1Q  Median      3Q     Max
-4.6706 -0.8541 -0.1308  0.9088  3.5183

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.9372343  5.9584509   1.332    0.184
Xlon        -0.1982295  0.0499149  -3.971 9.83e-05 ***
Xlat        -0.4951096  0.0691068  -7.164 1.32e-11 ***
XelevHW1    -0.0059590  0.0002029 -29.374  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.412 on 209 degrees of freedom
Multiple R-squared: 0.8415,Adjusted R-squared: 0.8393
F-statistic:   370 on 3 and 209 DF,  p-value: < 2.2e-16
```

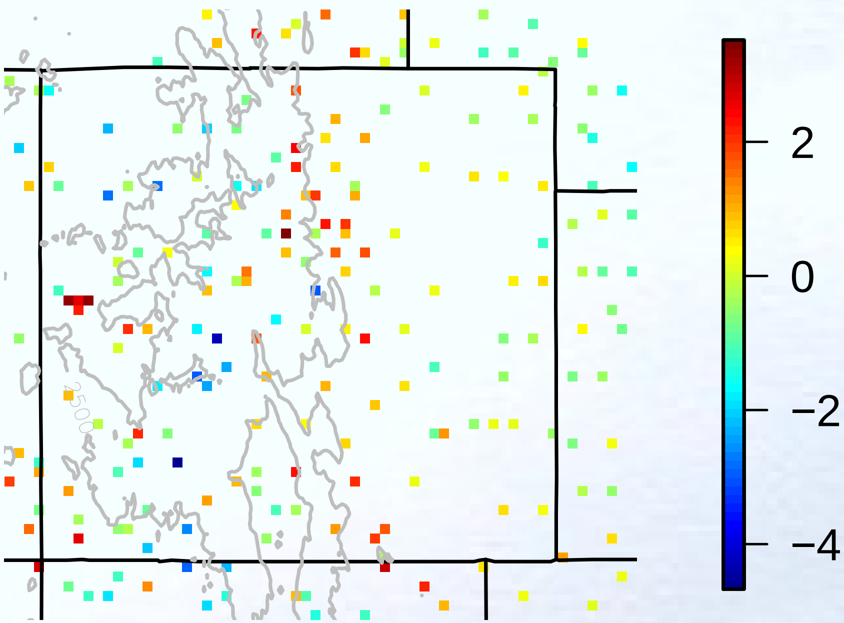
Checking residuals

Use `quilt.plot` to look at predicted values and residuals.
(`out$fitted.values` and `out$residuals`) contour at 2500m

Linear function lon/lat and elevation



Residuals



Some spatial signal left over.

How do we combine the linear regression with the spatial model?