Applied Spatial Statistics: Lecture 2A Random curves and surfaces

Douglas Nychka,
National Center for Atmospheric Research

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

Outline

- The additive model
- Expectation (review of lecture1)
- Covariance function $\rightarrow$ Random fields
- Spatial data $\rightarrow$ covariance function
  using a variogram

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

“Review” from last lecture

Mean of a random variable is also called the expectation

$E[U]$ another way of writing "mean of random variable" U.

$VAR[U]$ can be $E[(U - E[U])^2]$.

Linear properties of expectations

For any random variables U and V and any constant a.

$E[U + V] = E[U] + E[V]$

$E[aU] = aE[U]$

From these rules

$$E[\sum a_j U_j] = \sum a_j E[U_j]$$

and all the matrix/ random vector rules from Lecture 1.

Work variogram on the board.

$E[\frac{(U_1 - U_2)^2}{2}] = ?$

Covariance functions

$g(\boldsymbol{x})$ is a random surface with $E[g(\boldsymbol{x})] = \mu(\boldsymbol{x})$

Covariance function k has two spatial arguments

$$k(\boldsymbol{x}_1, \boldsymbol{x}_2) = E[(g(\boldsymbol{x}_1) - \mu(\boldsymbol{x}_1)(g(\boldsymbol{x}_2) - \mu(\boldsymbol{x}_2))]$$

Usually $\mu(\boldsymbol{x})$ is constant or even zero.

Some examples:

Exponential

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

Usual (Euclidean) distance $= \sqrt{(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T(\boldsymbol{x}_1 - \boldsymbol{x}_2)}$

Isotropic

$$k(\boldsymbol{x}_1, \boldsymbol{x}_2) = \rho\Phi(\text{distance}(\boldsymbol{x}_1, \boldsymbol{x}_2)/\theta)$$

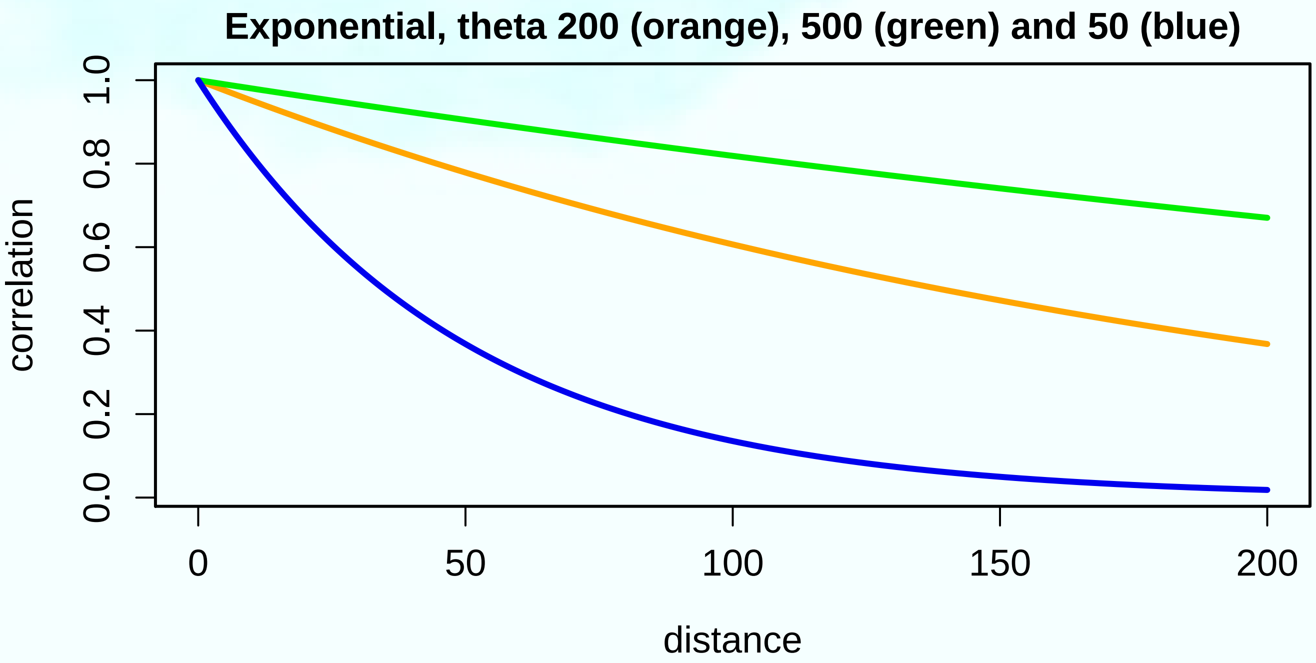
Exponential $\Phi(d) = e^{-d}$

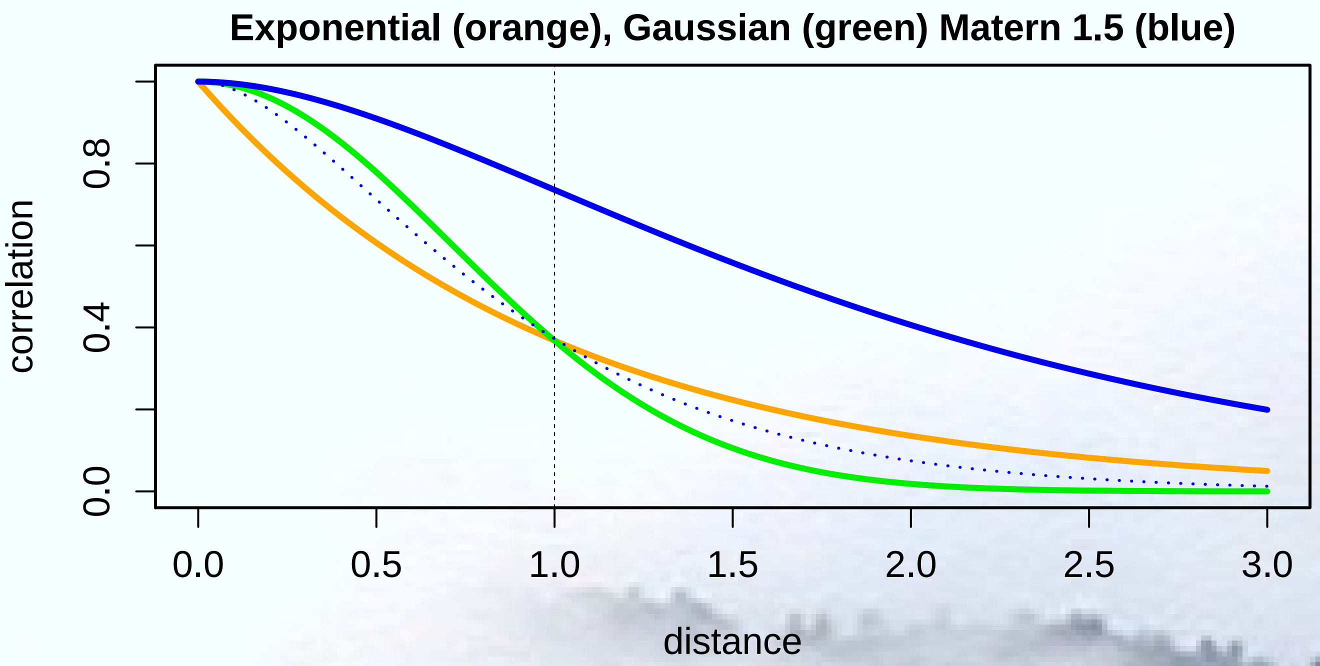
Gaussian $\Phi(d) = e^{-d^2}$

Member of Matern family $\Phi(d) = (1 + d)e^{-d}$
(smoothness$=$ 1.5)

Not every choice of k or Φ will be a covariance function

Examples using R





Simulating a process

- Need to assume that is Gaussian
- Assume that g has a mean (expectation) of zero

$\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ a set of locations
e.g. a grid or the data locations.

Simulate $\{g(\boldsymbol{x}_1), g(\boldsymbol{x}_2), \ldots, g(\boldsymbol{x}_n)\}$

Σ the covariance matrix for g at these locations:
$\Sigma_{i,j} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$

For example $\Sigma_{i,j} = e^{-\text{distance}(\boldsymbol{x}_i, \boldsymbol{x}_j)}$

```
M <- 30
N <- 30 # don't set these too large!
grid.list<- list( x= seq( 0,3,,M), y=seq(0,3,,N) )
x.grid<- make.surface.grid( grid.list)
Sigma<- exp( - rdist( x.grid, x.grid) )
```

Simulating a process (continued)

Decompose Σ as

$$\Sigma = B^T B$$

Always possible using the Cholesky (`chol`) decomposition.

If U is a random vector that has mean zero and identity covariance $U \sim (0, I)$

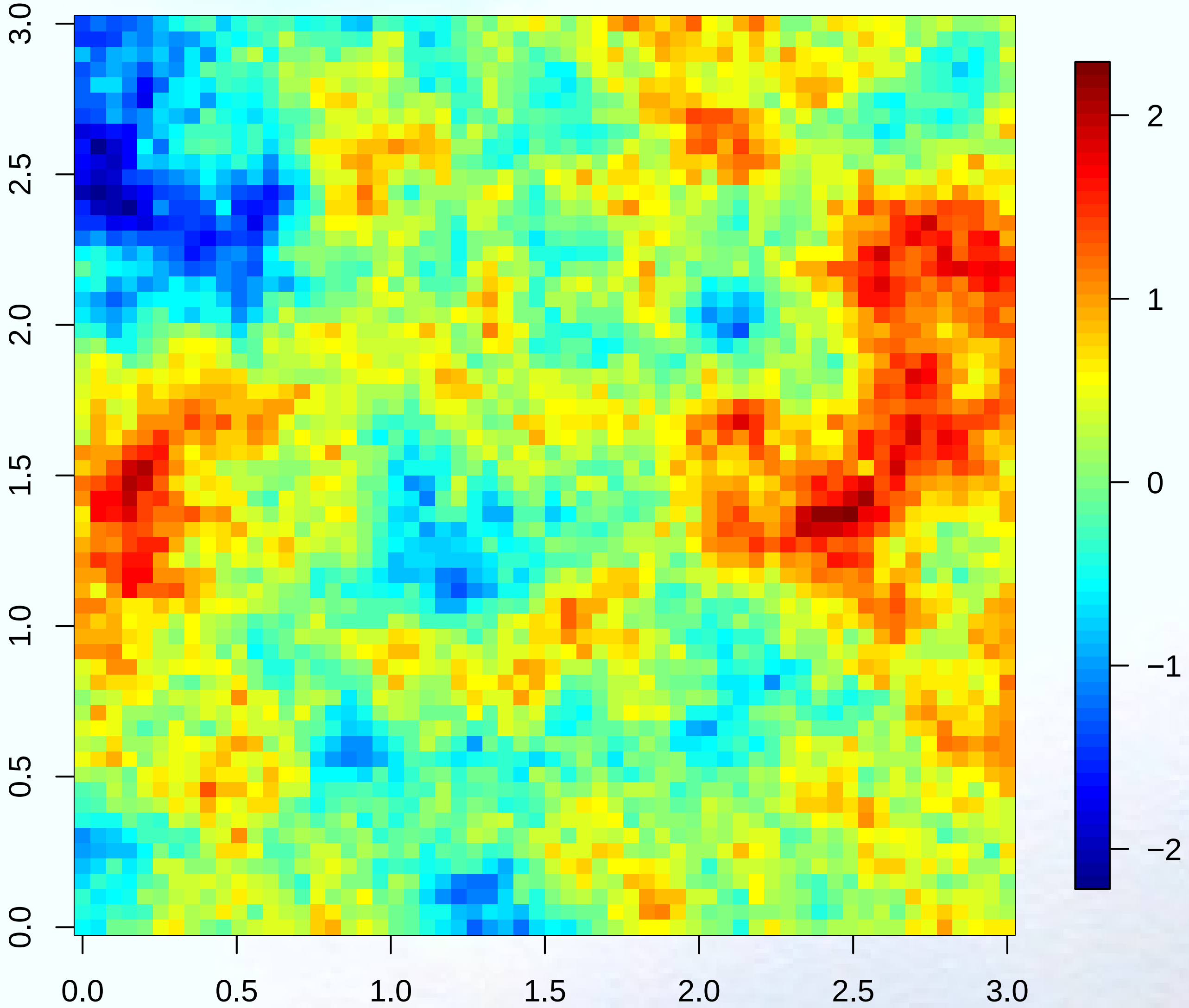
$$Z = B^T U$$

Z will have mean zero and covariance Σ
i.e. $Z \sim (0, \Sigma)$

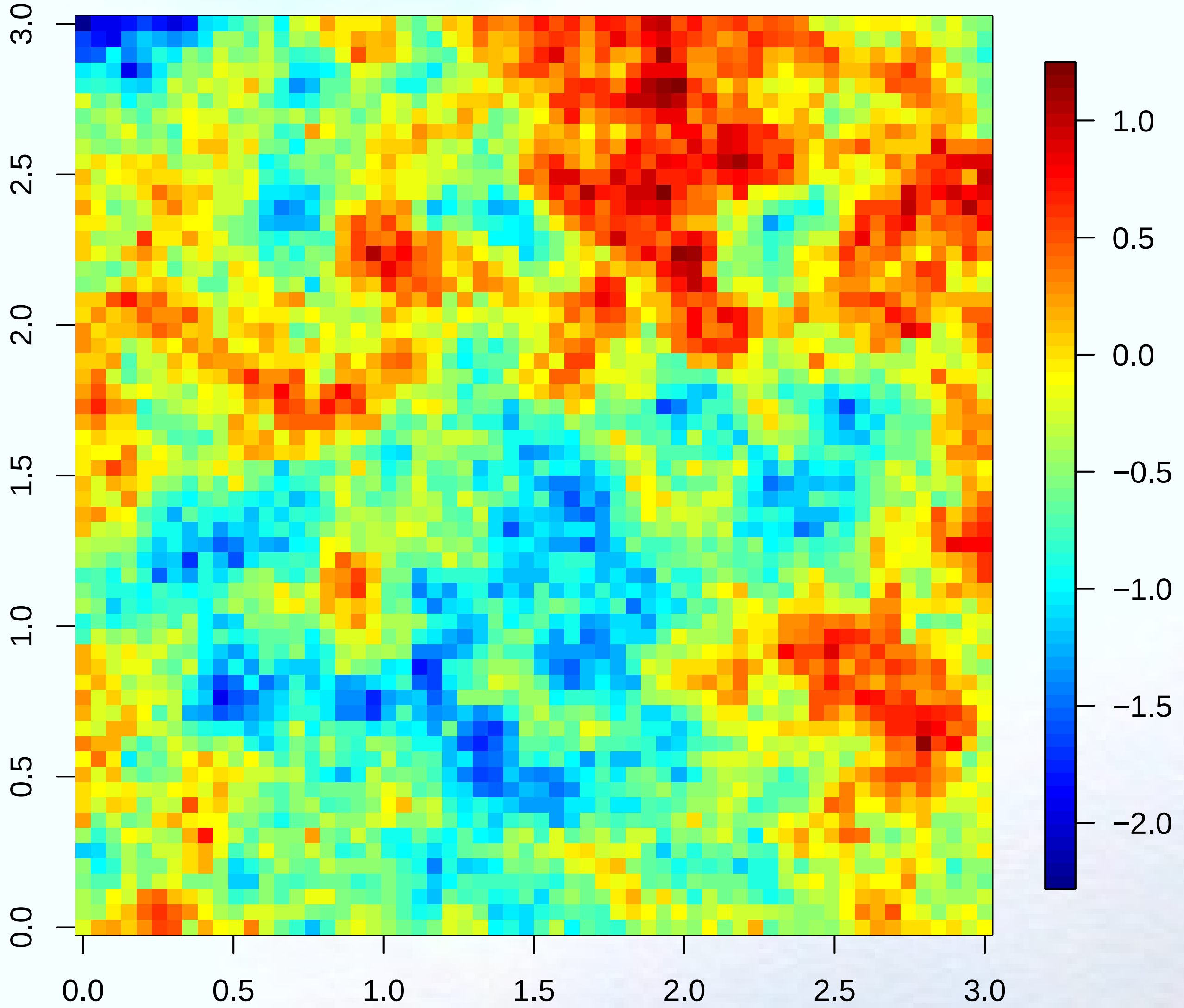
```
B<- chol( Sigma)
# check this by comparing t(B)%*%B to Sigma
U<- rnorm(M*N)
Z<- t(B)%*% U
image.plot( as.surface( x.grid, Z))
```

- *What are the dimensions of* `B`*?*
- *Why is the* `as.surface` *needed?*

One realization

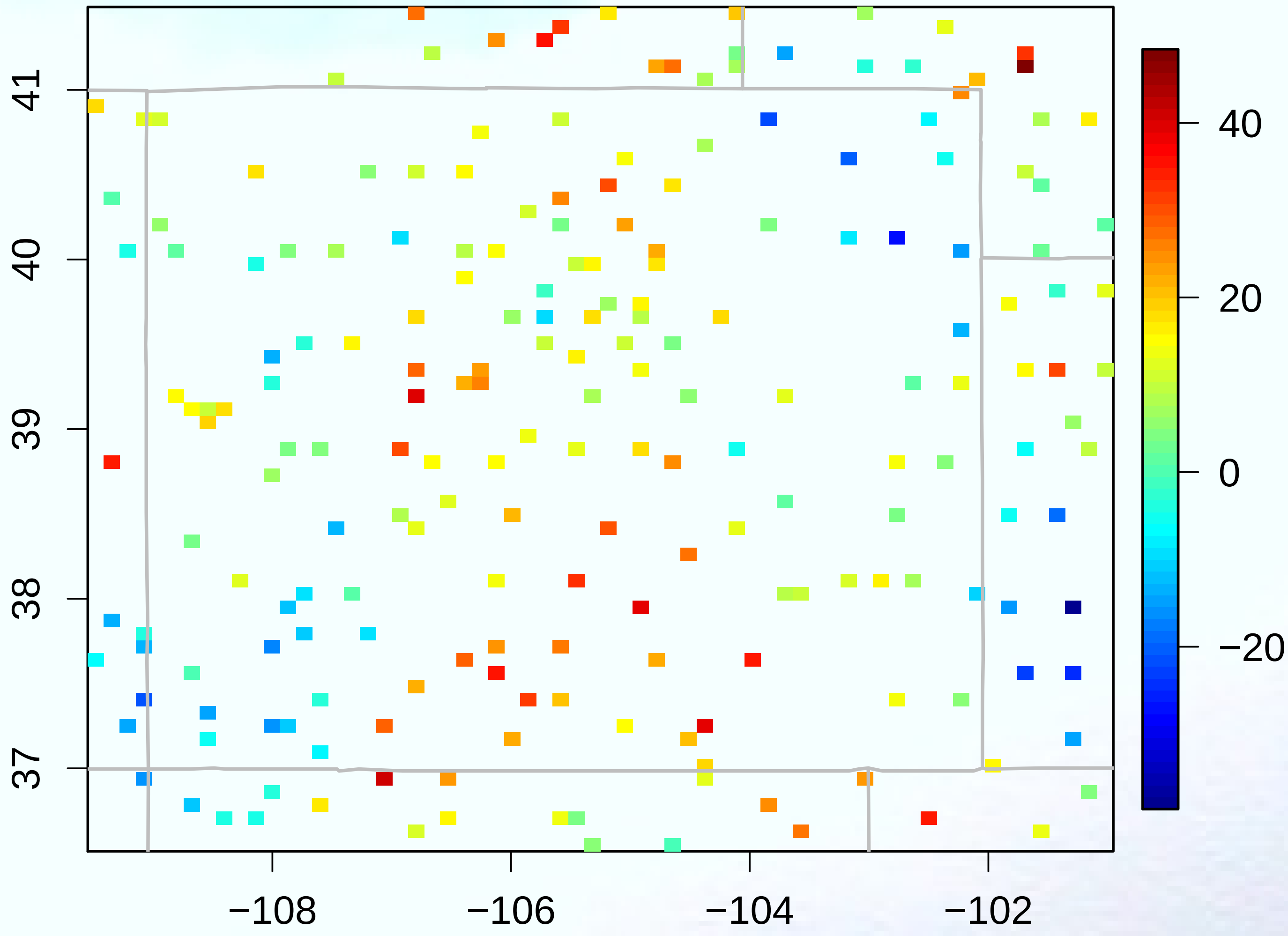


Another ...



From Homework 1

Method works for arbitrary locations and covariance functions



Realization using `Sigma<-825.1*exp(-rdist.earth(xHW1,xHW1)/187.8)`

BTW This is not a very good model for the data!