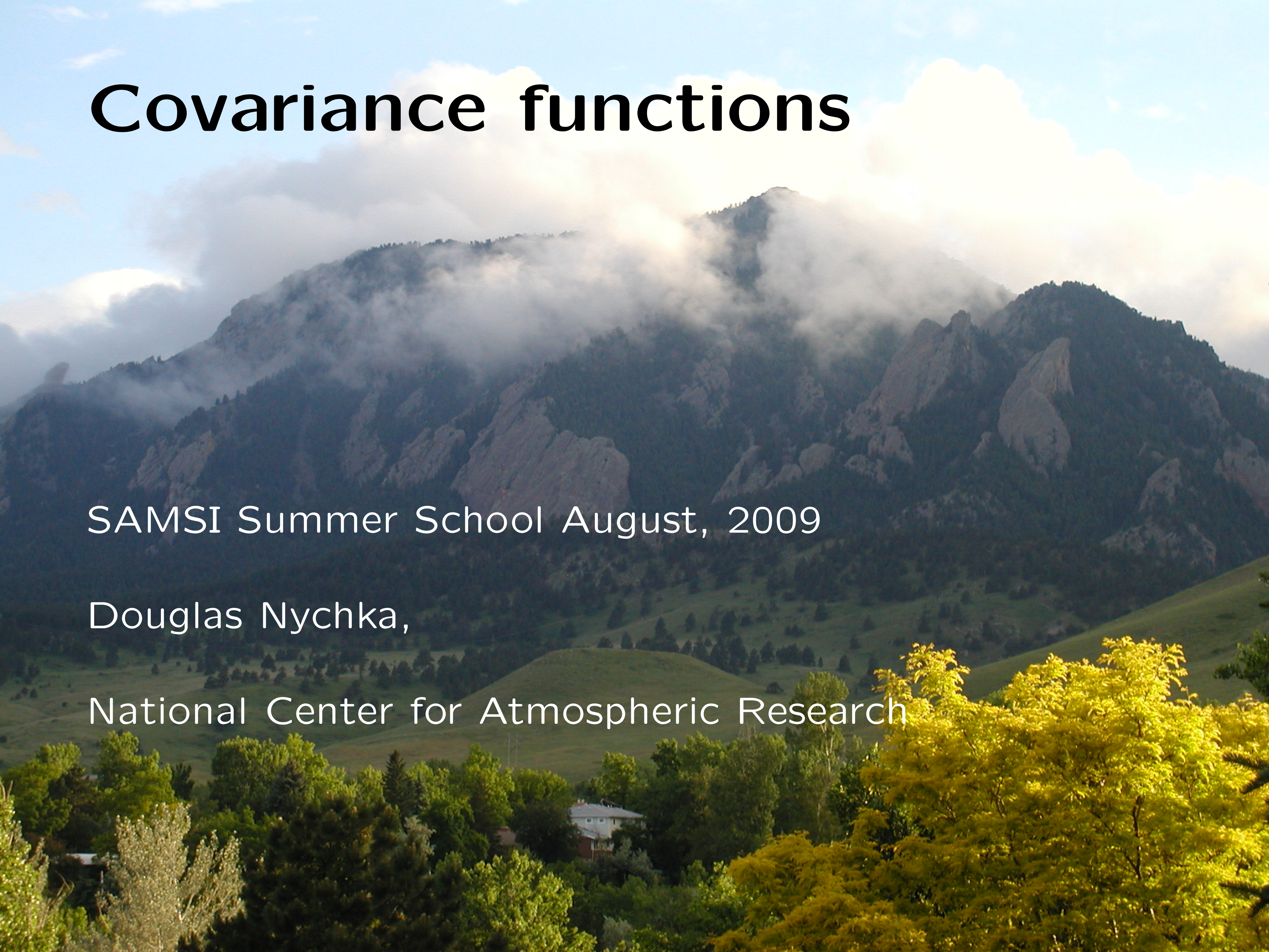
Covariance functions



SAMSI Summer School August, 2009

Douglas Nychka,

National Center for Atmospheric Research

Outline

- Restricted maximum likelihood

- Tapering a covariance

Supported by the National Science Foundation *SAMSI July 2009*

The spatial model

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

$$y_i = [T\boldsymbol{d}]_i + h(x_i) + \epsilon_i$$

ϵ_i's are random errors.

$\boldsymbol{\epsilon}$ are $MN(0, \sigma^2 I)$

$h(x)$ is a realization of a Gaussian process
$E(h(x)) = 0$ and $COV(h(x), h(x')) = \rho k_\theta(x, x')$

In vector notation:

$$\boldsymbol{y} = T\boldsymbol{d} + \boldsymbol{h} + \boldsymbol{\epsilon}$$

$\rho K(\theta)$ is the covariance matrix for $\boldsymbol{h}$.

The likelihood

- 2 log likelihood:

$$(\boldsymbol{y} - T\boldsymbol{d})^T (M)^{-1} (\boldsymbol{y} - T\boldsymbol{d}) + log(|M|) + c$$

$M = \rho K(\theta) + \sigma^2 I$ and c is a constant.

Eliminate fixed part

Choose U an orthogonal matrix so that $U^T T = 0$

$$U^T \boldsymbol{y} = U^T \boldsymbol{h} + U^T \boldsymbol{\epsilon}$$

Restricted likelihood

$$(\boldsymbol{y}^*)^T (M^*)^{-1} (\boldsymbol{y}^*) + log(|M^*|) + c$$

$\boldsymbol{y}^* = U^T \boldsymbol{y}$ and $M^* = U^T M U$

Linear algebra for the likelihood

Cholesky decomposition of M^*:

$M^* = C^T C$ where C is an upper triangular matrix.

Quadratic Form:

$$(\boldsymbol{y}^*)^T (M^*)^{-1} (\boldsymbol{y}^*) = (\boldsymbol{y}^*)^T (C^T C)^{-1} (\boldsymbol{y}^*)$$
$$= (\boldsymbol{y}^*)^T C^{-1} C^{-T} (\boldsymbol{y}^*) = ||C^{-T} \boldsymbol{y}^*||^2$$

solve the linear system $C^T \boldsymbol{w} = \boldsymbol{y}^*$.
so $(\boldsymbol{y}^*)^T (M^*)^{-1} (\boldsymbol{y}^*) = ||\boldsymbol{w}||^2$

For the determinant: $|M^*| = |C^T C| = |C|^2$

Determinant of an upper triangular matrix is the product of the diagonal elements.

Maximizing the likelihood

- All these computations need to be done for a single evaluation. and are order n^3. Although the Cholesky decomposition and solution are faster that a full matrix inversion.

- Difficult to find the derivatives of the likelihood although it can be concentrated about σ or ρ

Maximum likelihood estimates are usually not feasible beyond about 1000 observations.

Alternatives:

- If replicates are available can fit a correlogram using nonlinear regression. This tends to have more variance.

- Use a compact covariance or tapering to make M a sparse matrix.

Tapering

- 2 log likelihood:

$$(\boldsymbol{y} - T\boldsymbol{d})^T (M)^{-1} (\boldsymbol{y} - T\boldsymbol{d}) + log(|M|) + c$$

$$tr\left((M)^{-1}(\boldsymbol{y} - T\boldsymbol{d})(\boldsymbol{y} - T\boldsymbol{d})^T\right) + log(|M|) + c$$

$$M_{i,j} = \rho k(\boldsymbol{x}_i, \boldsymbol{x}_j) + \sigma^2$$

The idea is to introduce many zeroes into this matrix.

$w(u, v)$ a covariance that is zero if $||u - v|| > \Delta$

Tapered version of covariance

Replace: $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$

with $k(\boldsymbol{x}_i, \boldsymbol{x}_j)w(\boldsymbol{x}_i, \boldsymbol{x}_j)$

Covariances with locations with distance larger than Δ will zero!

The one taper -2 log likelihood:

$$(\boldsymbol{y} - T\boldsymbol{d})^T (M \cdot W)^{-1} (\boldsymbol{y} - T\boldsymbol{d}) + log(|M \cdot W|) + c$$

$[M \cdot W]_{i,j} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)w(\boldsymbol{x}_i, \boldsymbol{x}_j)$

The two tapered -2 log likelihood:

Also taper the matrix $(\boldsymbol{y} - T\boldsymbol{d})(\boldsymbol{y} - T\boldsymbol{d})^T$ in the same way

$$tr([M \cdot W]^{-1}[(\boldsymbol{y} - T\boldsymbol{d})(\boldsymbol{y} - T\boldsymbol{d})^T \cdot W]) + log(|M \cdot W|) + c$$

This will tend to give less biased estimates than the one taper.

Some closing remarks

- The likelihood can be simplified by eliminating the fixed part.
- Cholesky decompositions lead to efficient methods for evaluating the likelihood.
- Tapering can substantially reduce the amount of computation by introducing sparse covariance matrices.