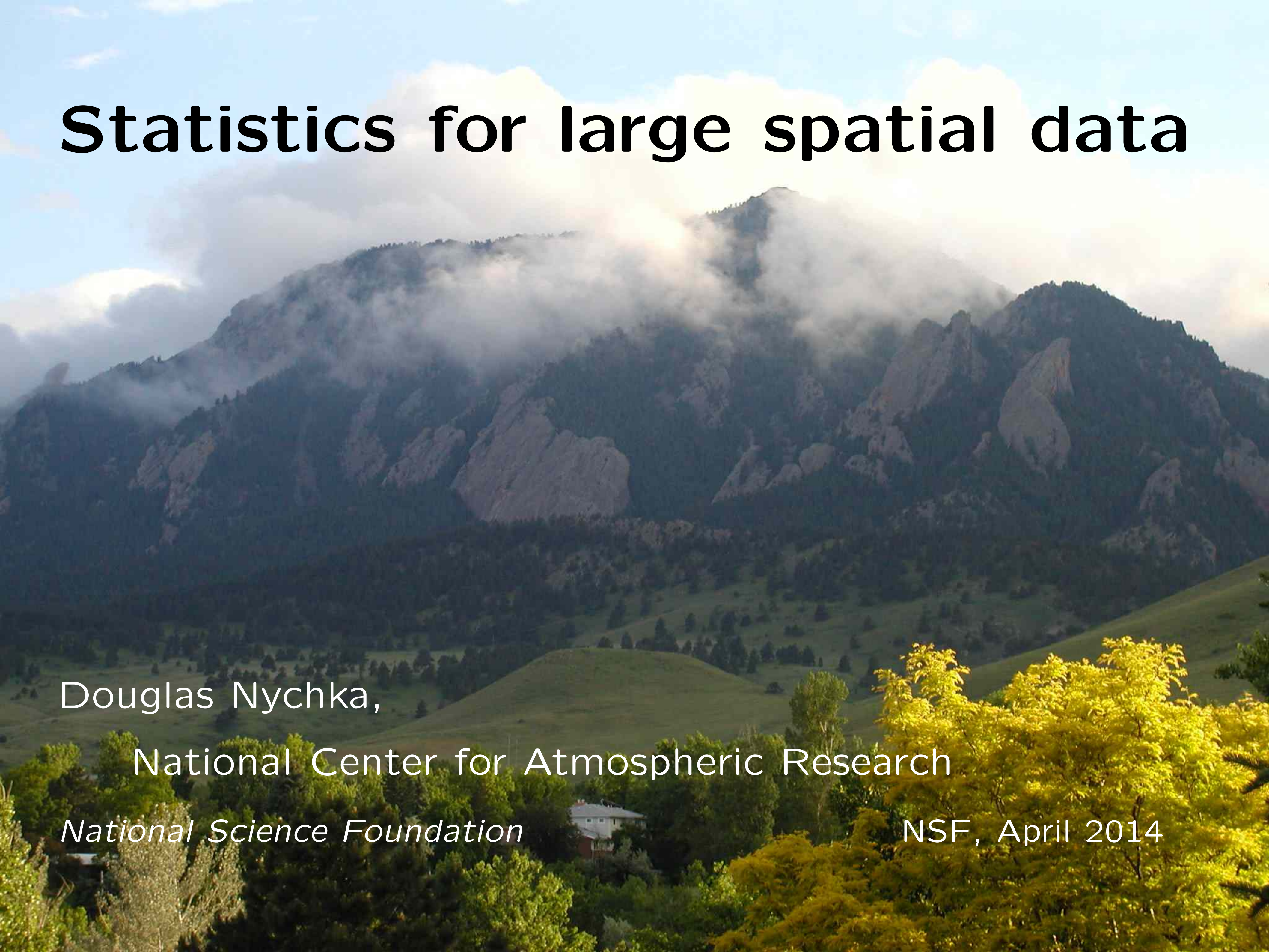


Statistics for large spatial data

Douglas Nychka,

National Center for Atmospheric Research

National Science Foundation

NSF, April 2014

Introductions

Doug Nychka,

North Carolina State Univ. 1983 – 1997, NCAR 1997 – present

Director and Scientist 4,

Institute for Mathematics Applied to Geosciences (IMAGe)

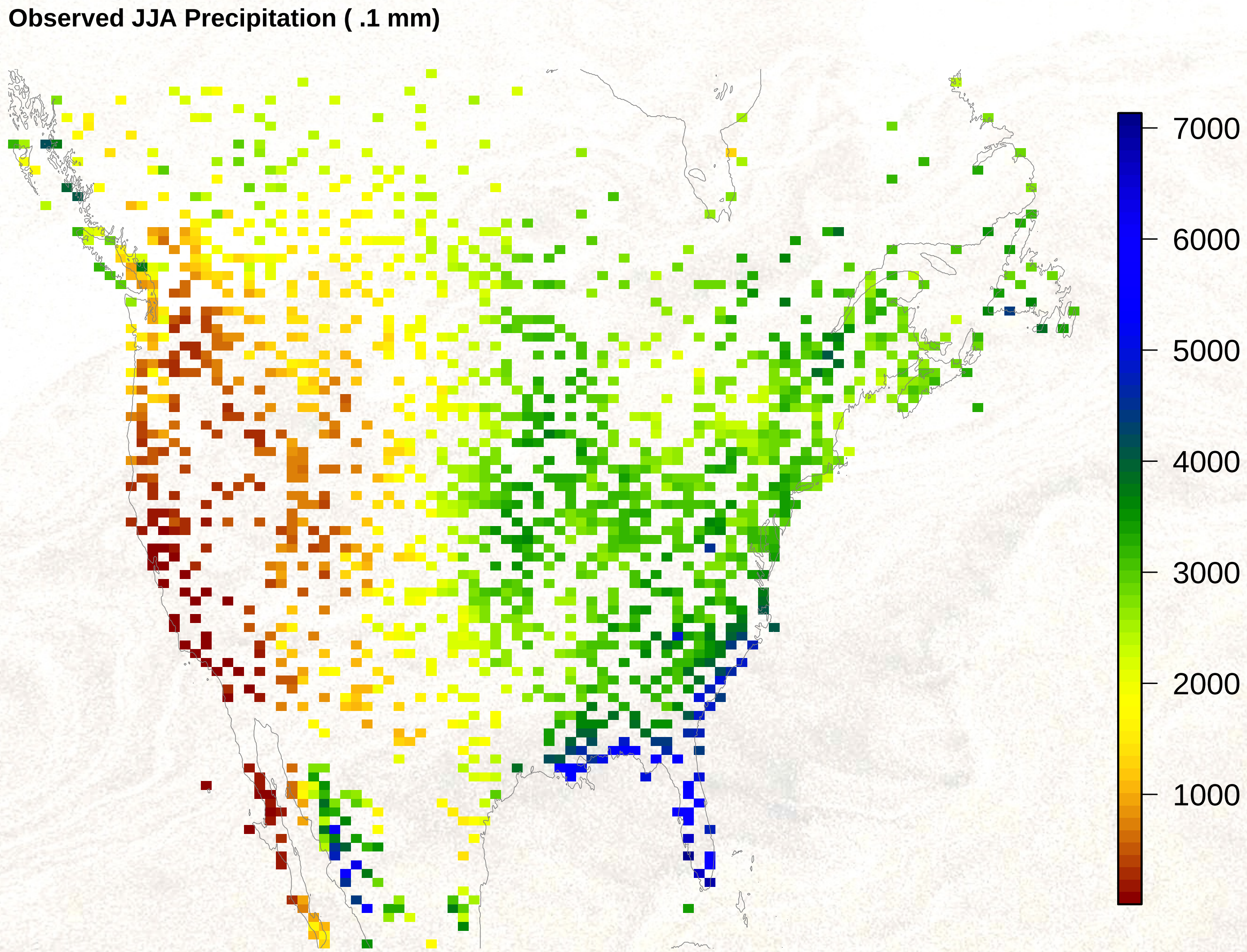
(26 staff, scientists, and post docs)

IMAGe is one of three divisions within the computational laboratory

- Summer rainfall
- Spatial statistics with bumps
- LatticeKrig
- Connections
- IMAGe Activities

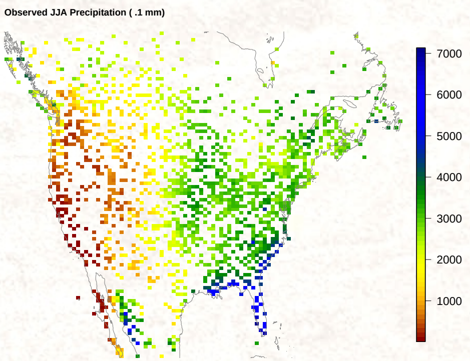
Observed mean summer precipitation

1720 stations reporting, "mean" for 1950-2010



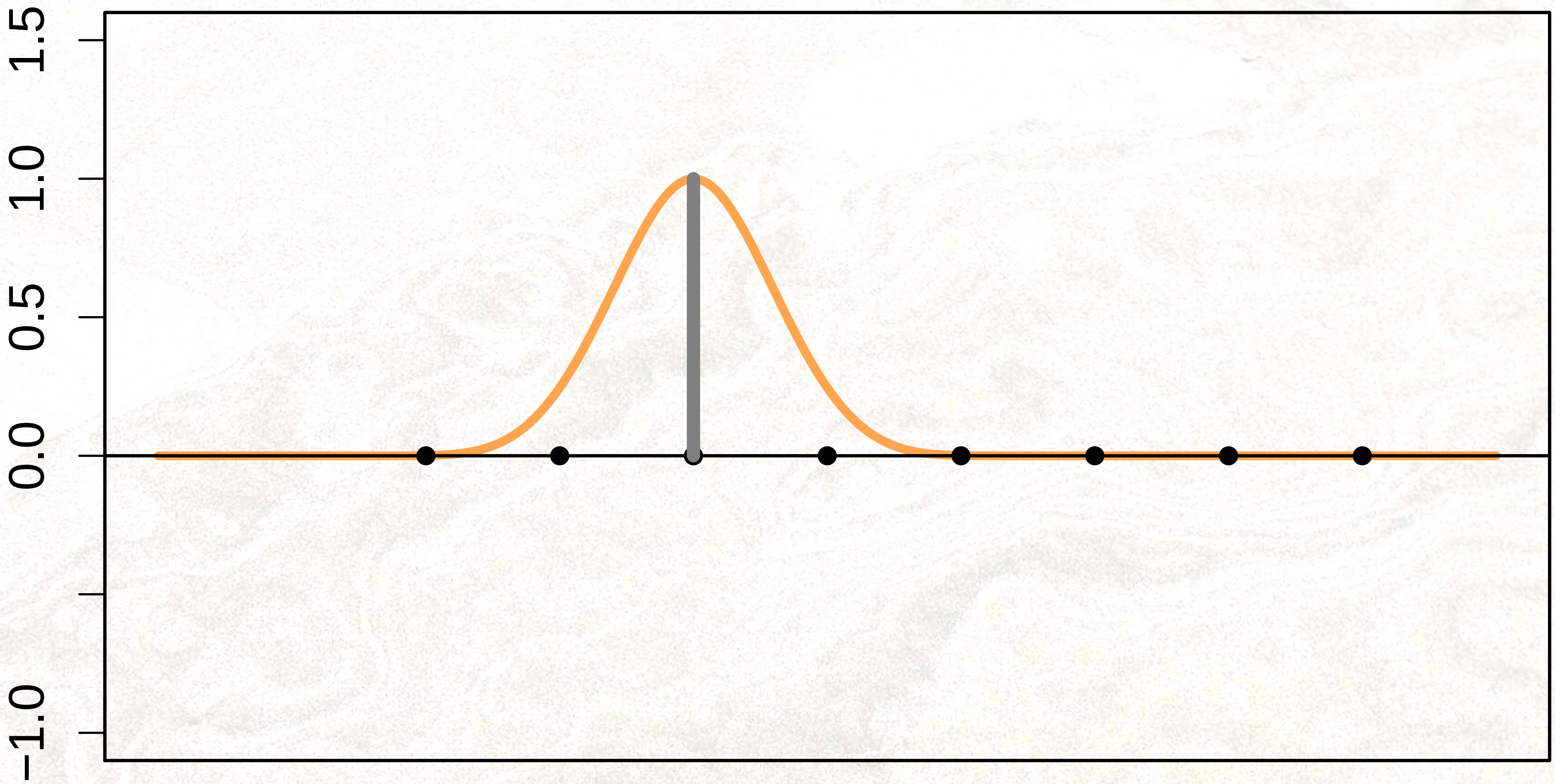
The statistical problem

What is the summer rainfall at places where there is no data?



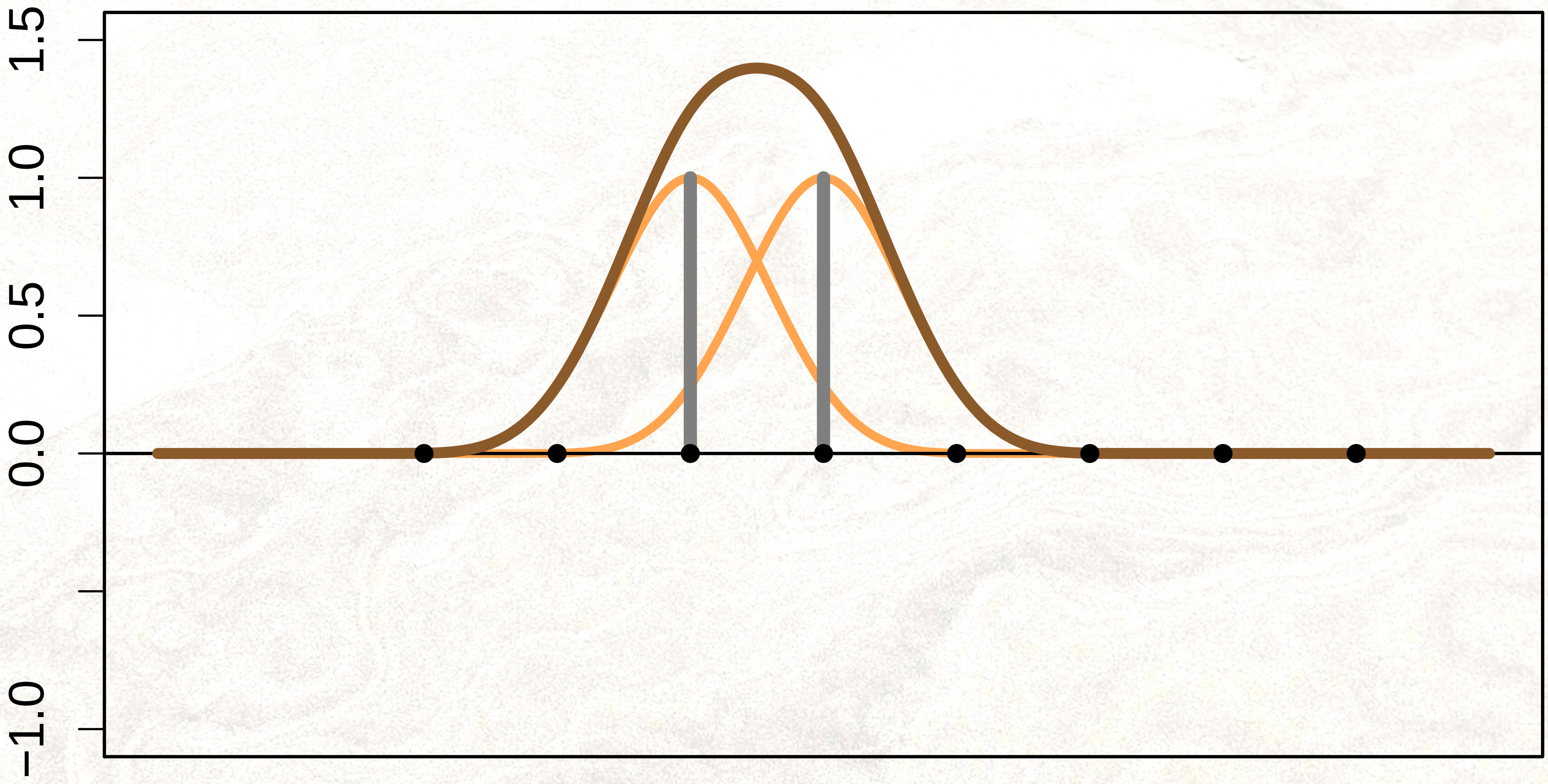
What is the uncertainty in the estimates?

Building a curve from bumps



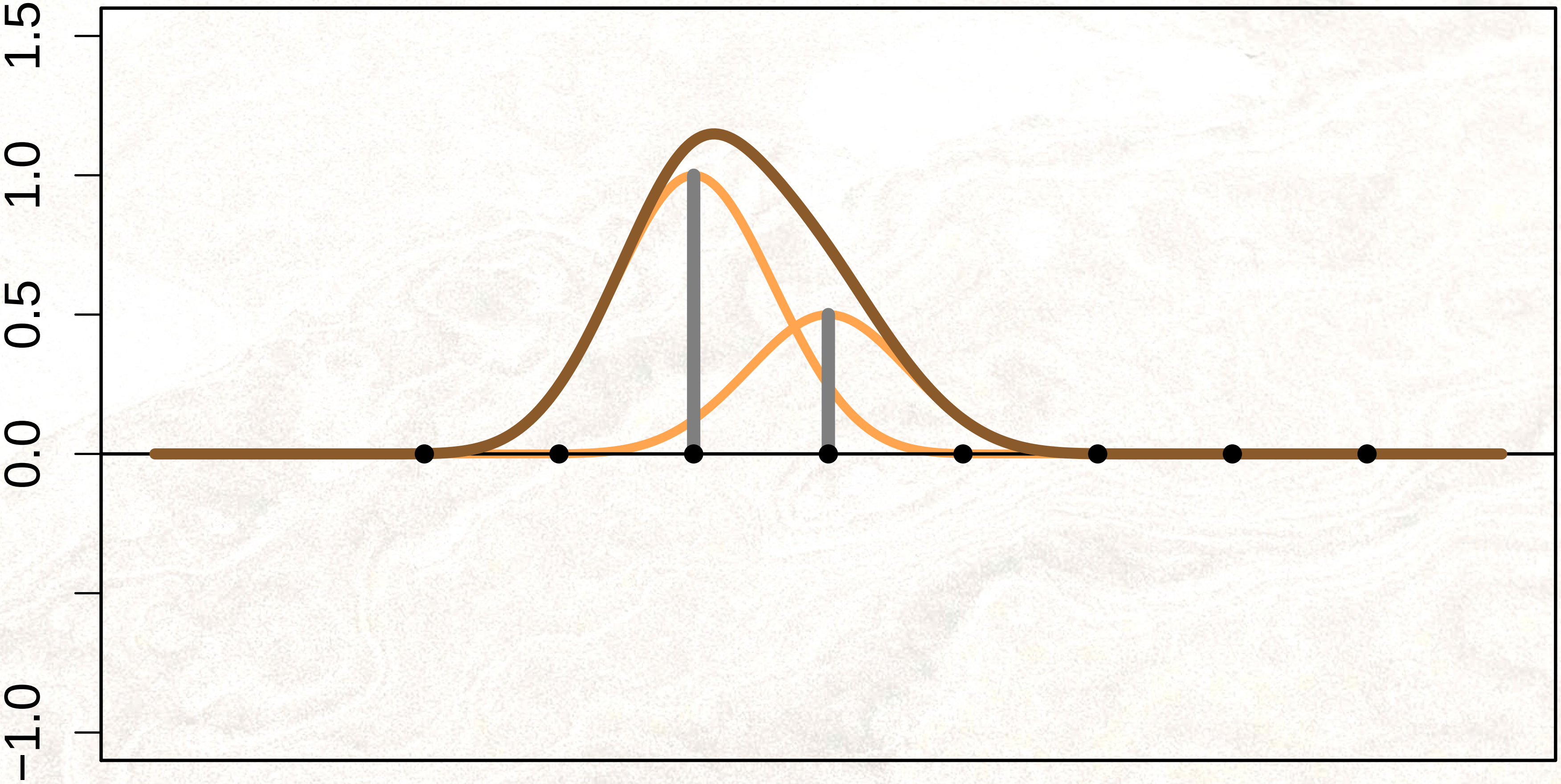
Single bump

Building a curve from bumps



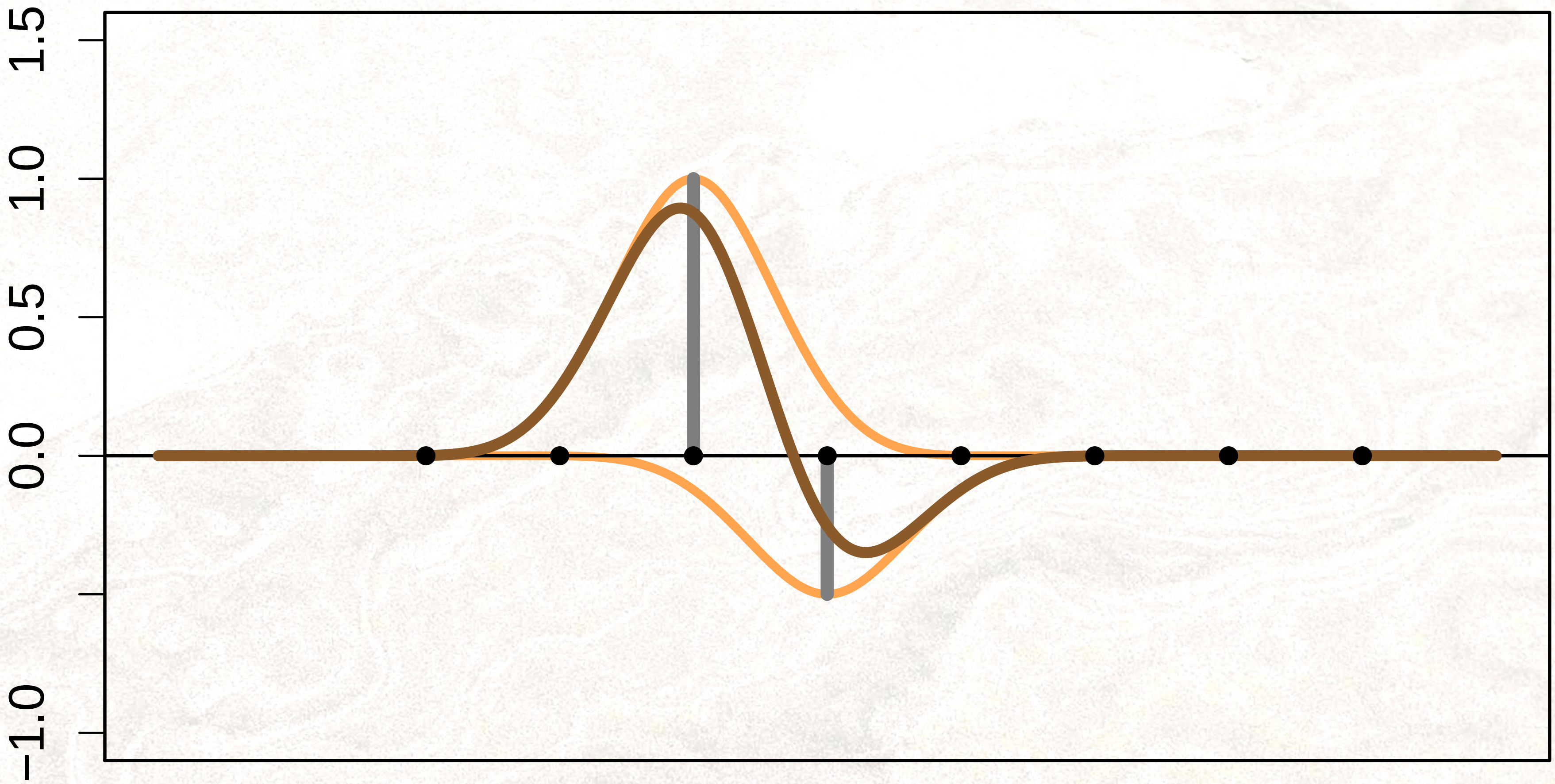
Two bumps same height

Building a curve from bumps



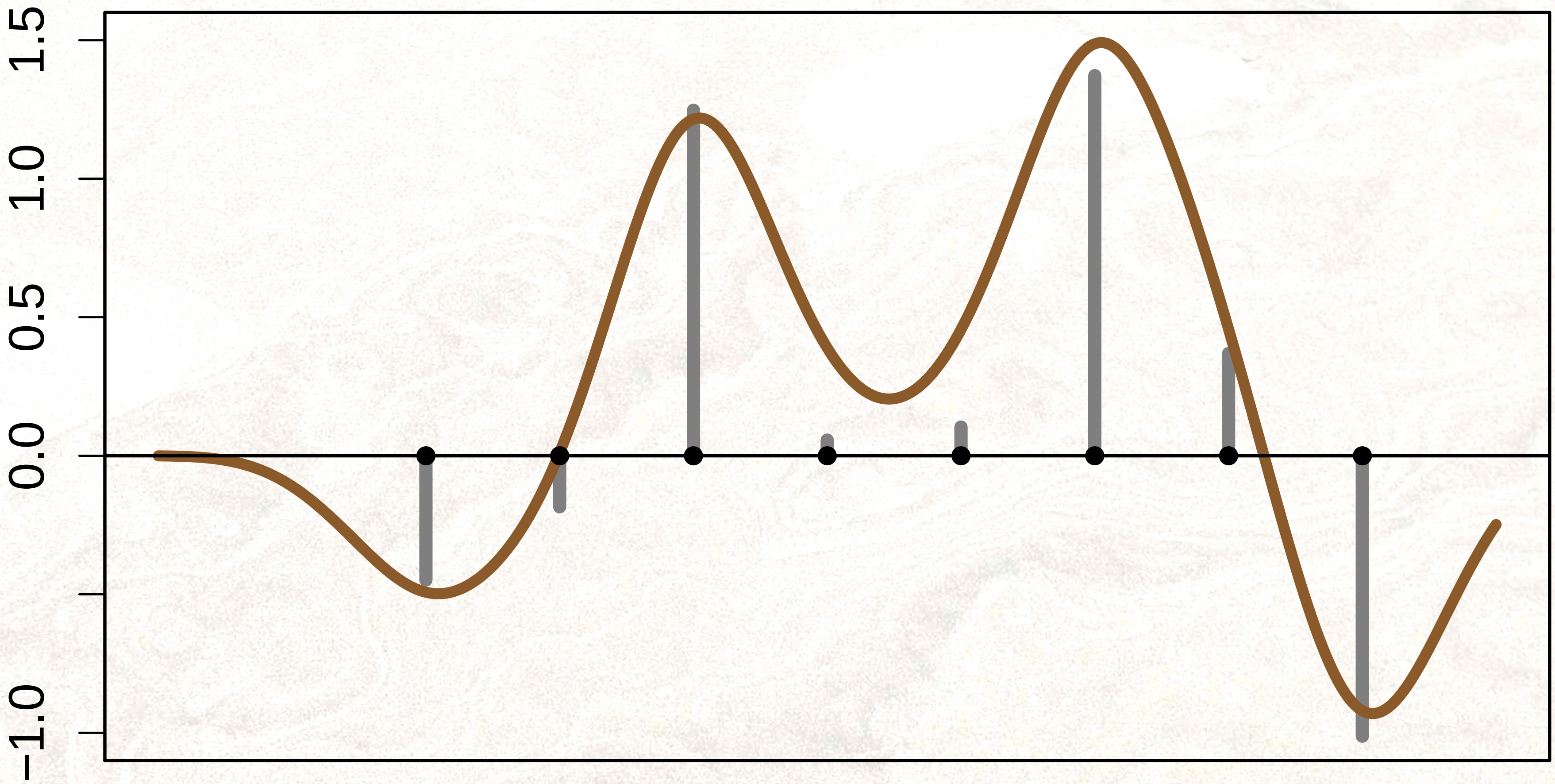
Two bumps different heights

Building a curve from bumps



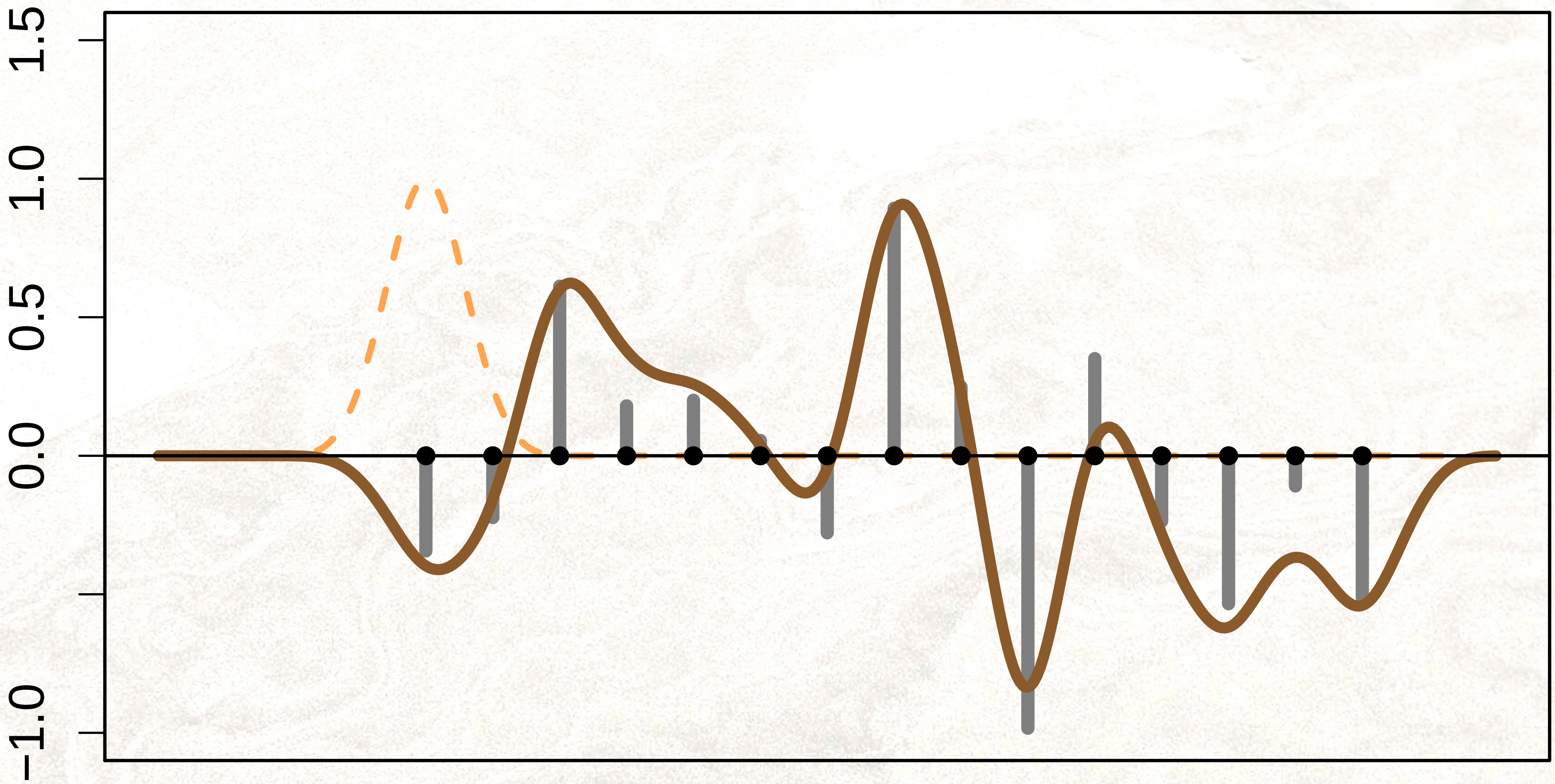
Two bumps different heights

Building a curve from bumps



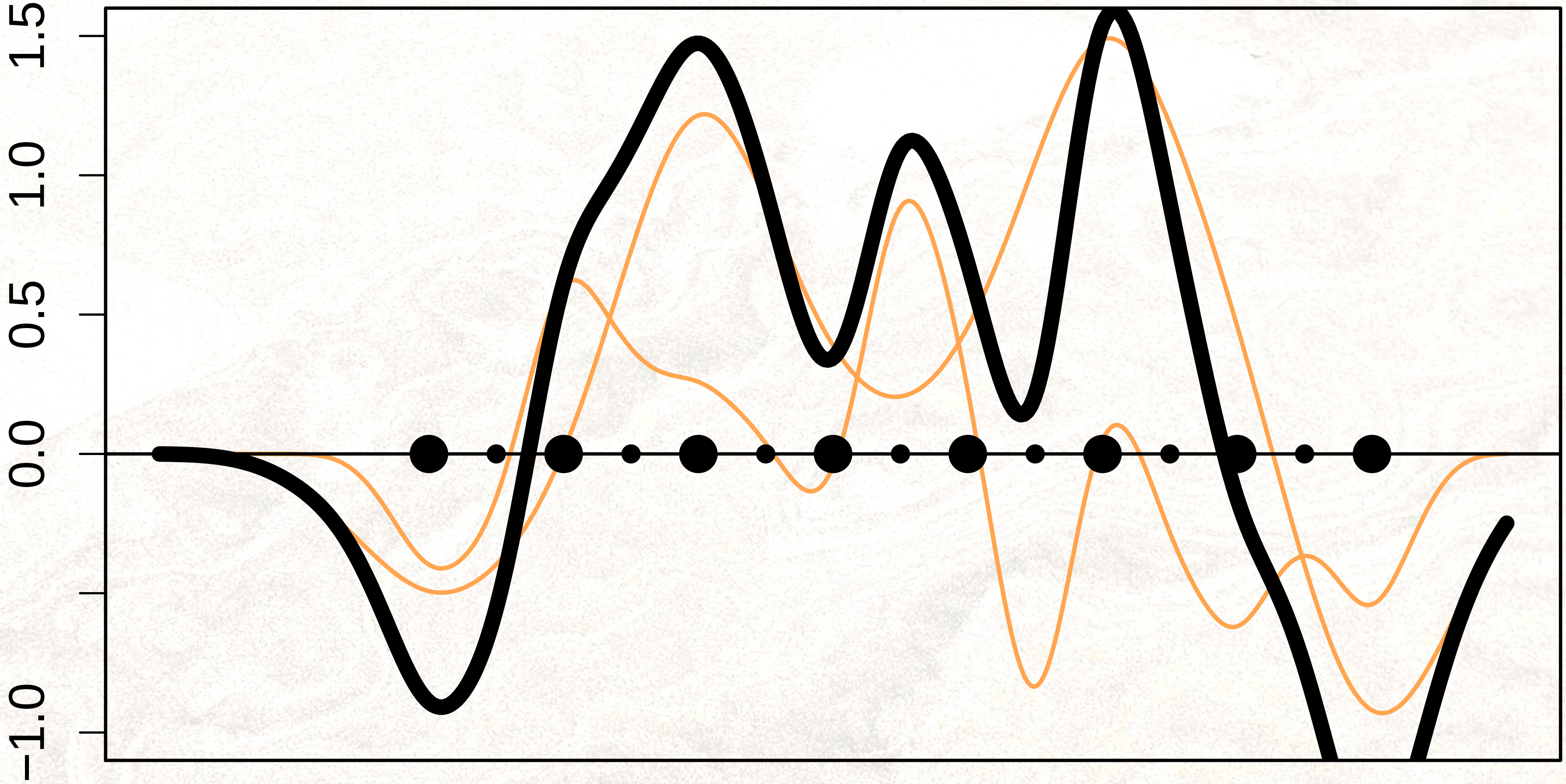
Eight bumps – all different heights

Building a curve from bumps



16 bumps – all different heights

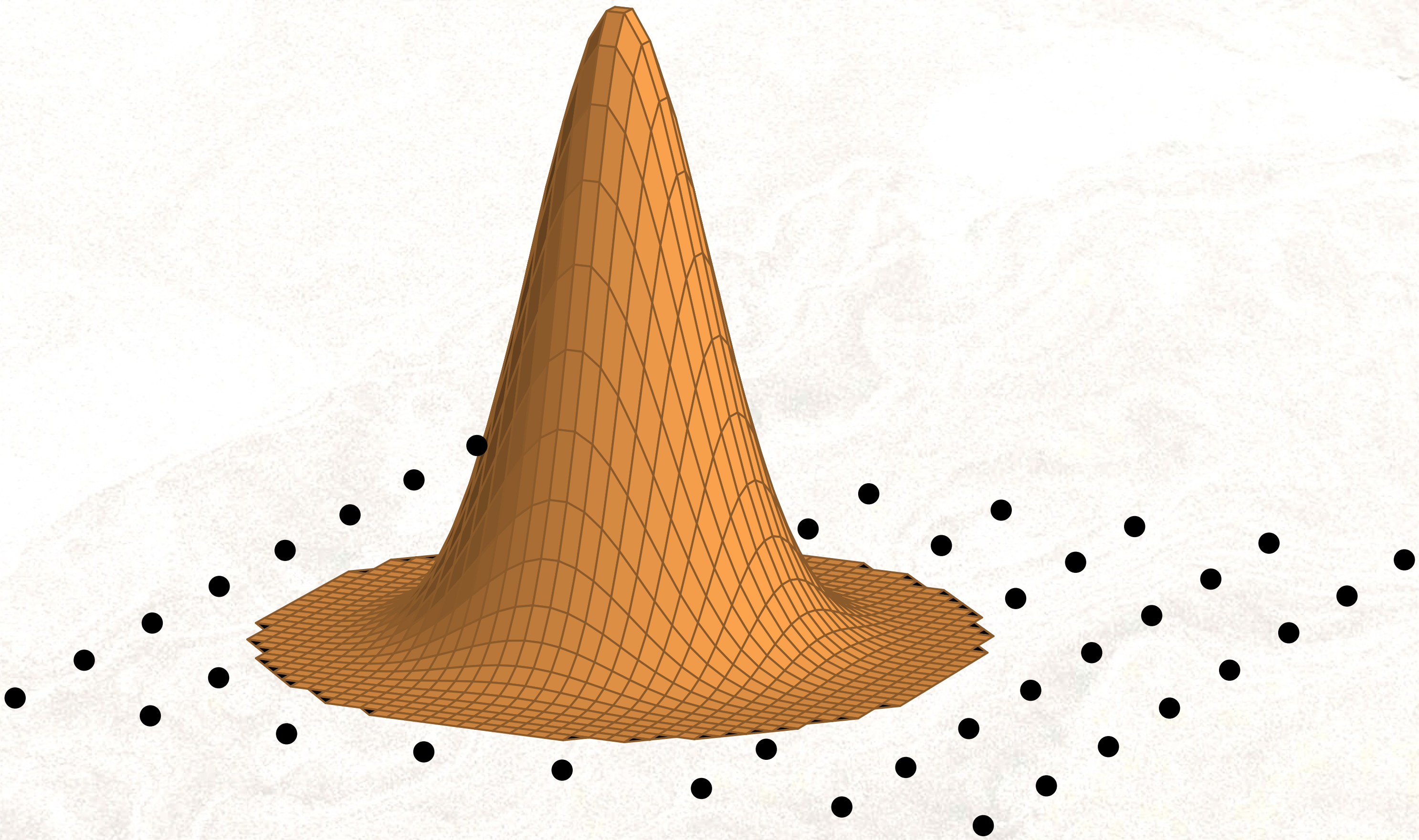
Building a curve from bumps



Adding them together

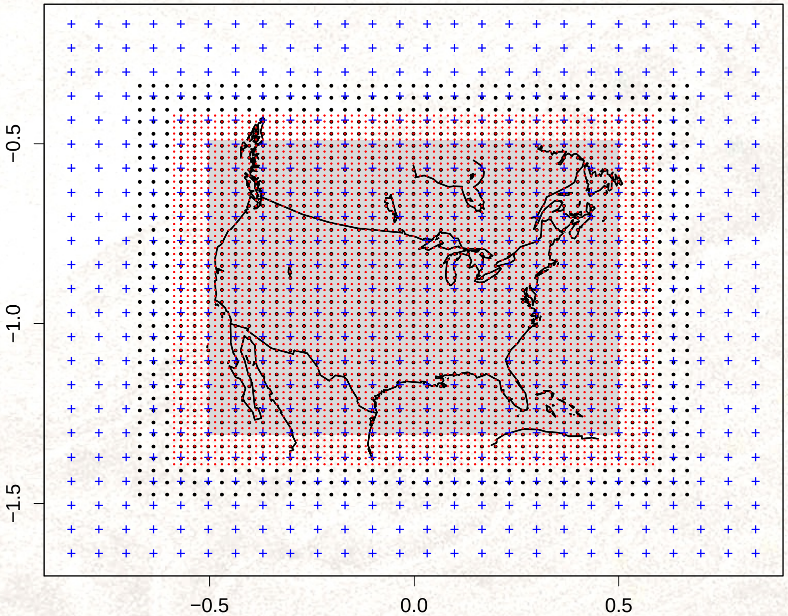
bumps = basis functions, bump heights = coefficients

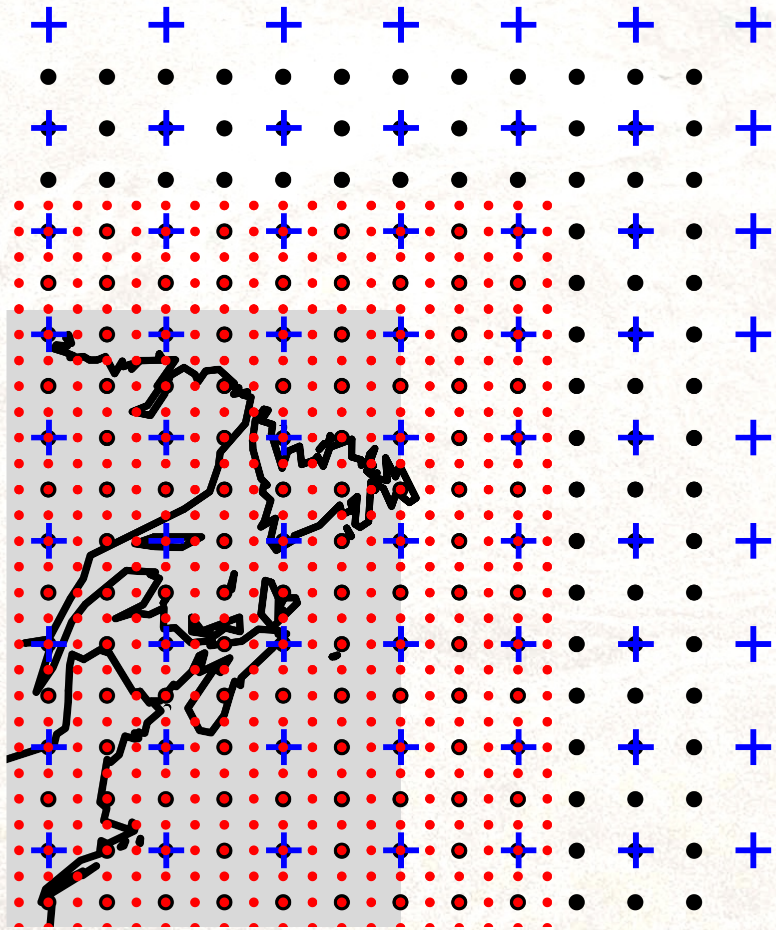
Going to two dimensions



Example of a 2-d bump

The lattice for the climate data





About 4000 total lattice points

Kriging

Danie G. Krige

South African Mining Engineer who pioneered the field of geostatistics.

Kriging

Methodology for estimating a surface based on irregular observations.

Justified by reasonable assumptions on the unknown surface.

Balancing two features

A cost function

(fit of the surface to the data) + (roughness of the surface)

- Want a surface that tracks the observations but is not overly rough and irregular.

Minimizing cost $\equiv$ Kriging

- Involves picking good coefficients for the basis functions
i.e. choosing how much sand to dump at each lattice point.

For math types:

$$\min_{\boldsymbol{c}} (\boldsymbol{y} - X\boldsymbol{c})^T (\boldsymbol{y} - X\boldsymbol{c}) + \boldsymbol{c}^T \boldsymbol{Q} \boldsymbol{c}$$

$\boldsymbol{y}$ the data, X matrix of basis functions,
$\boldsymbol{c}$ coefficients, Q roughness matrix.

For the statisticians

Negative, log posterior

(fit of the surface to the data) + (roughness of the surface)

+ (penalty for parameters)

$$\min_{\boldsymbol{c},\theta} (\boldsymbol{y} - X\boldsymbol{c})^T (\boldsymbol{y} - X\boldsymbol{c}) + \boldsymbol{c}^T \boldsymbol{Q}_\theta \boldsymbol{c} - log|\boldsymbol{Q}_\theta| + \text{stuff}$$

$\boldsymbol{y}$ the data, X matrix of basis functions, $\boldsymbol{c}$ coefficients,
Q_θ inverse covariance matrix, θ statistical parameters .

Contours of cost function around minimum used to describe the uncertainty

More about the roughness penalty

Some coefficients:

$$\begin{array}{ccccc} . & . & . & . & . \\ . & . & c_1 & . & . \\ . & c_2 & c_* & c_3 & . \\ . & . & c_4 & . & . \\ . & . & . & . & . \end{array}$$

Some weights:

$$\begin{array}{ccccc} . & . & . & . & . \\ . & . & -1/4 & . & . \\ . & -1/4 & \alpha & -1/4 & . \\ . & . & -1/4 & . & . \\ . & . & . & . & . \end{array}$$

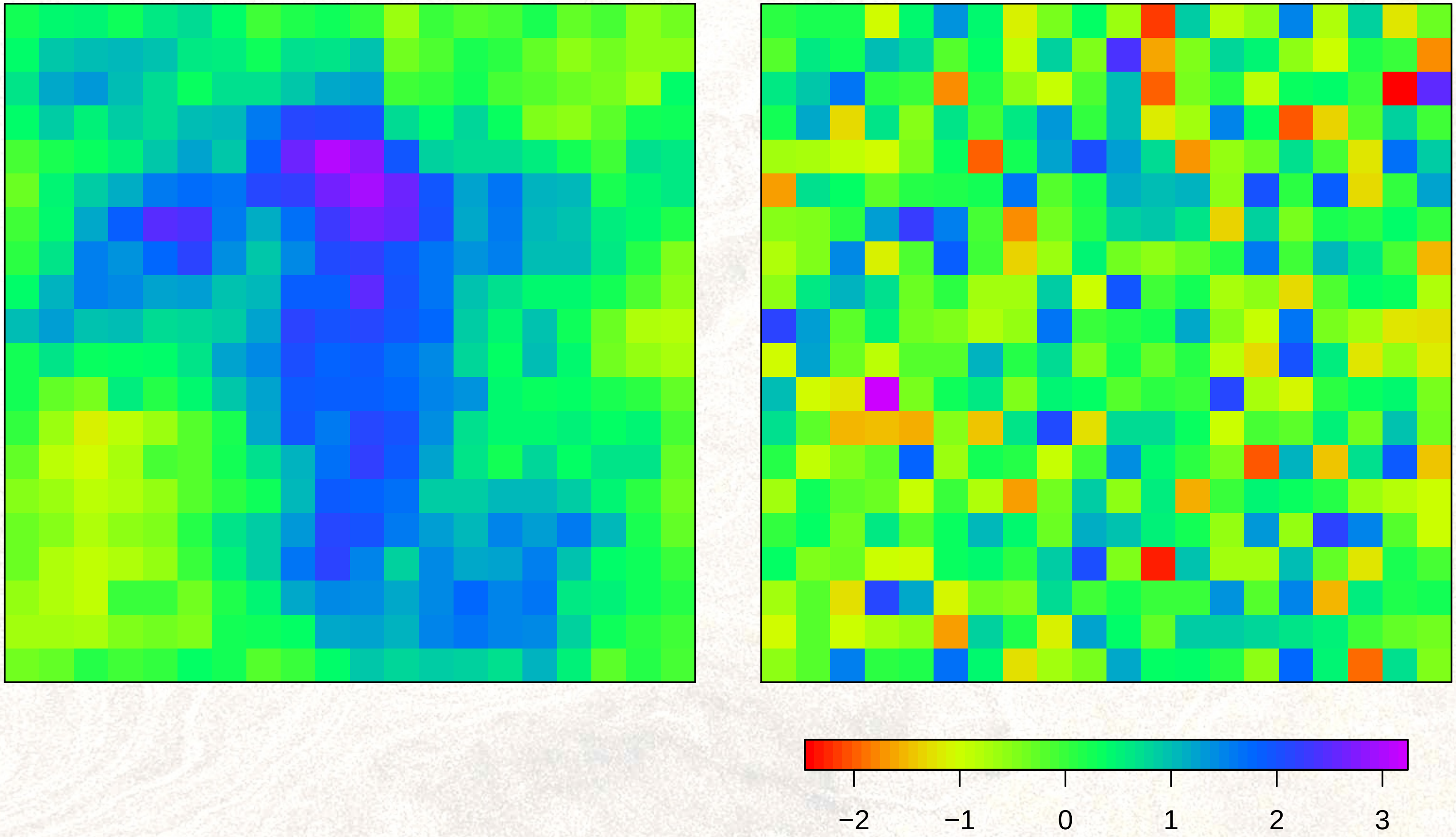
The filter:

$$\alpha c_* - 1/4\,(c_1 + c_2 + c_3 + c_4) = \text{white noise}$$

- α needs to be greater than 1.
- A simple discretization of the Laplacian. $\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$
- Roughness penalty is the sum of squares of the filtered coefficients.

Filtering coefficients

Coefficients on the lattice *Applying the filter*

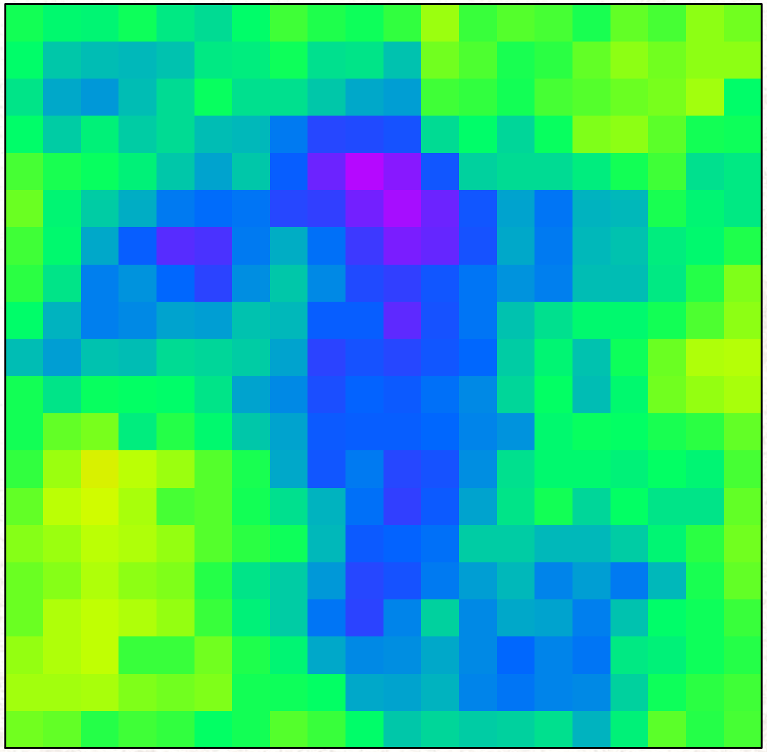


$$c_* \rightarrow \alpha c_* - 1/4\,(c_1 + c_2 + c_3 + c_4)$$

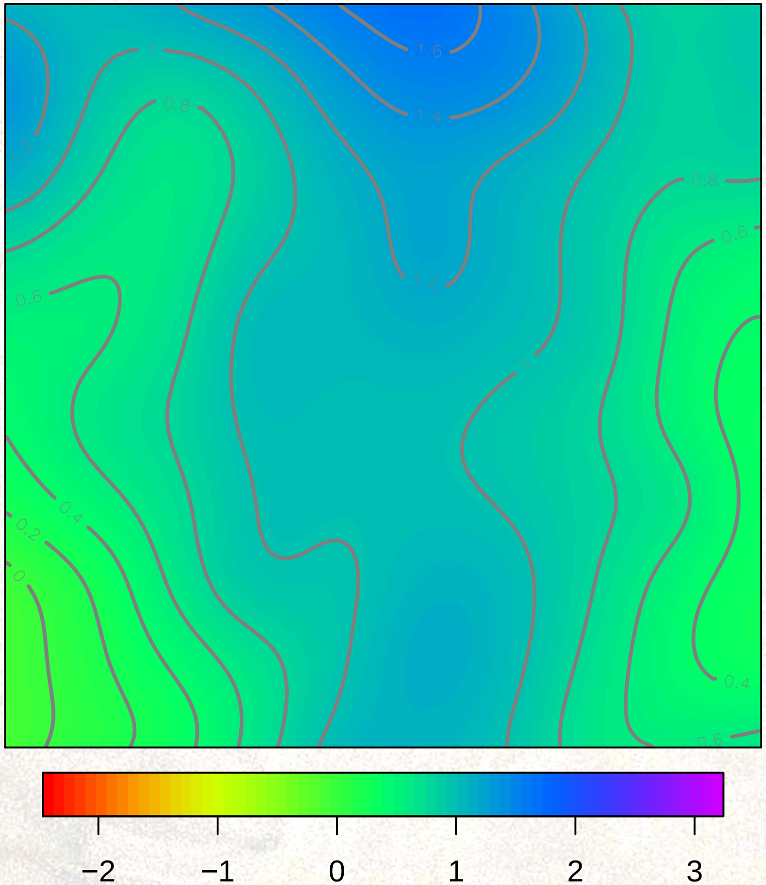
$\alpha = 1.0025$

Applying the basis functions

Coefficients on the lattice

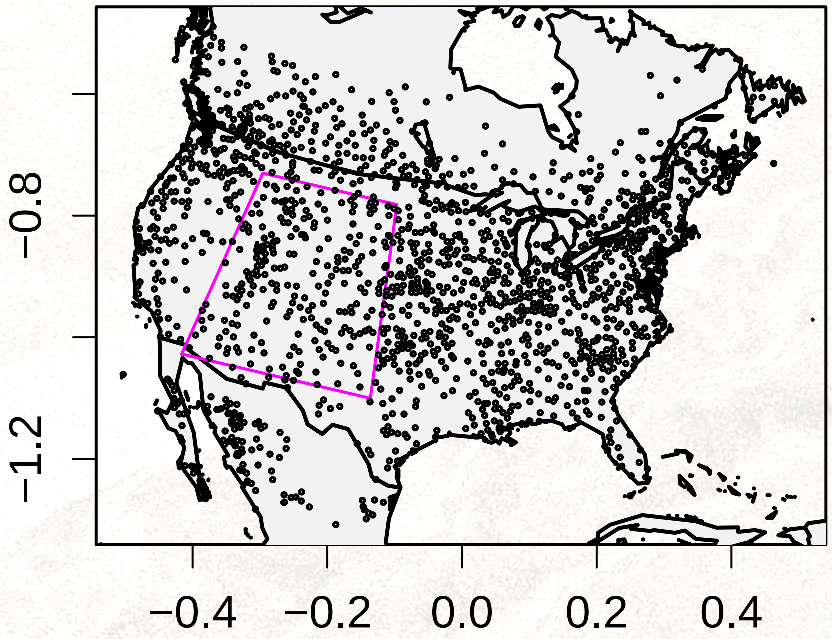


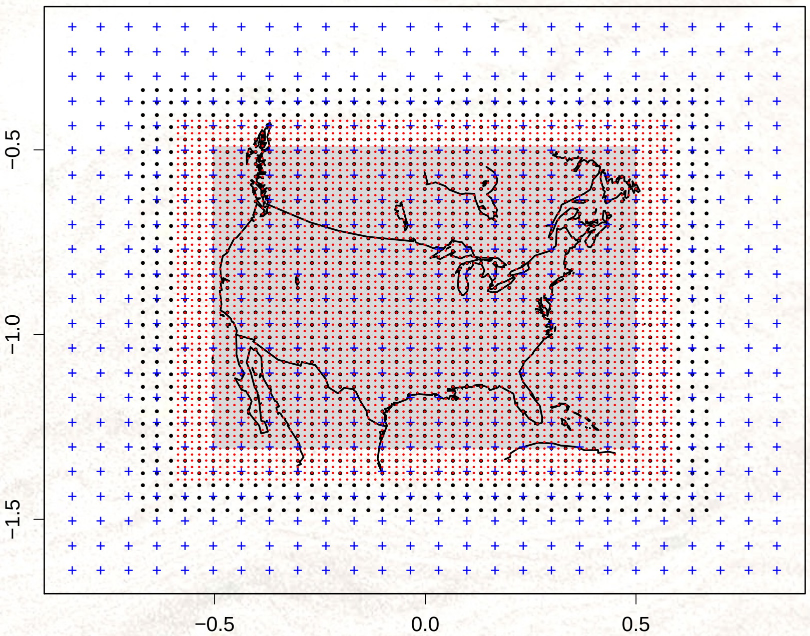
Expanding with basis functions



$$c_k \to \sum \phi_k(x) c_k$$

Back to rainfall observations



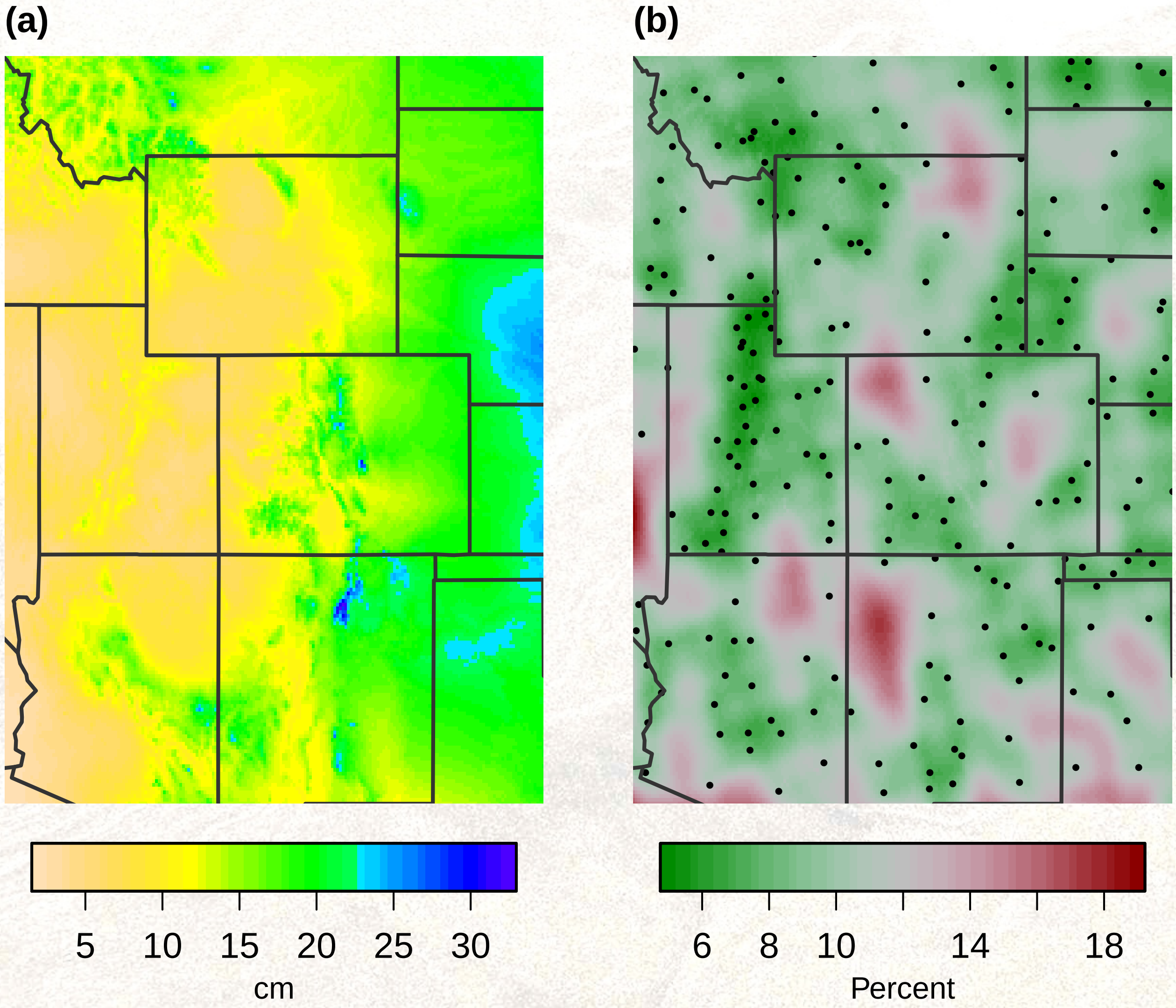


Three levels of resolution

- ≈ 4000 basis functions total.
- statistical parameters found by maximum likelihood
- coefficients found by "kriging"
- uncertainty found by Monte Carlo ensemble
- includes linear adjustment for elevation

Estimated summer rainfall

Predicted JJA rainfall (cm) Pointwise standard errors (percent)



Summary

- **Computational efficiency gained by compact basis functions and sparse precision matrix.**

- **Multi-resolution can approximate standard covariance families (e.g. Matern)**

- **Easy to generate uncertainty measures.**

See `LatticeKrig` contributed package in R

Connections

- Supercomputing
- Data assimilation
- Uncertainty in pattern scaling

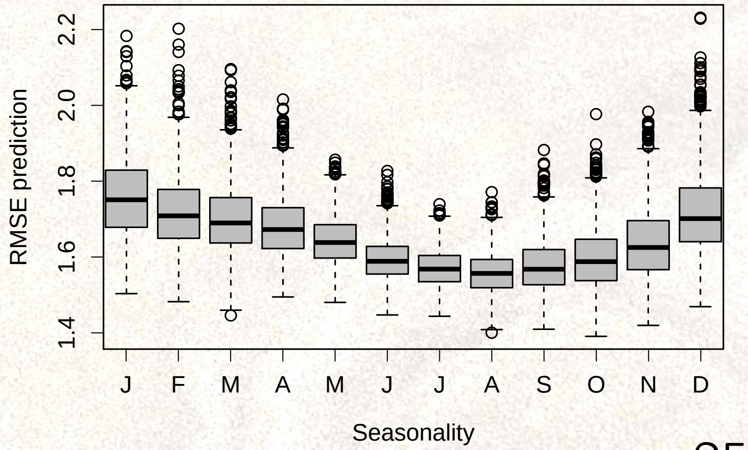
Interactive supercomputing

What would a statistician do with 10 seconds of Yellowstone?

- Run separate R session on a 1000 cores
- Analyze different parts of data in parallel
- Use the same code that runs on a laptop!

50 years of daily temperatures for N America

- About 15,000 days, each with several thousand locations
- Spatial prediction error depends on season



Other applications

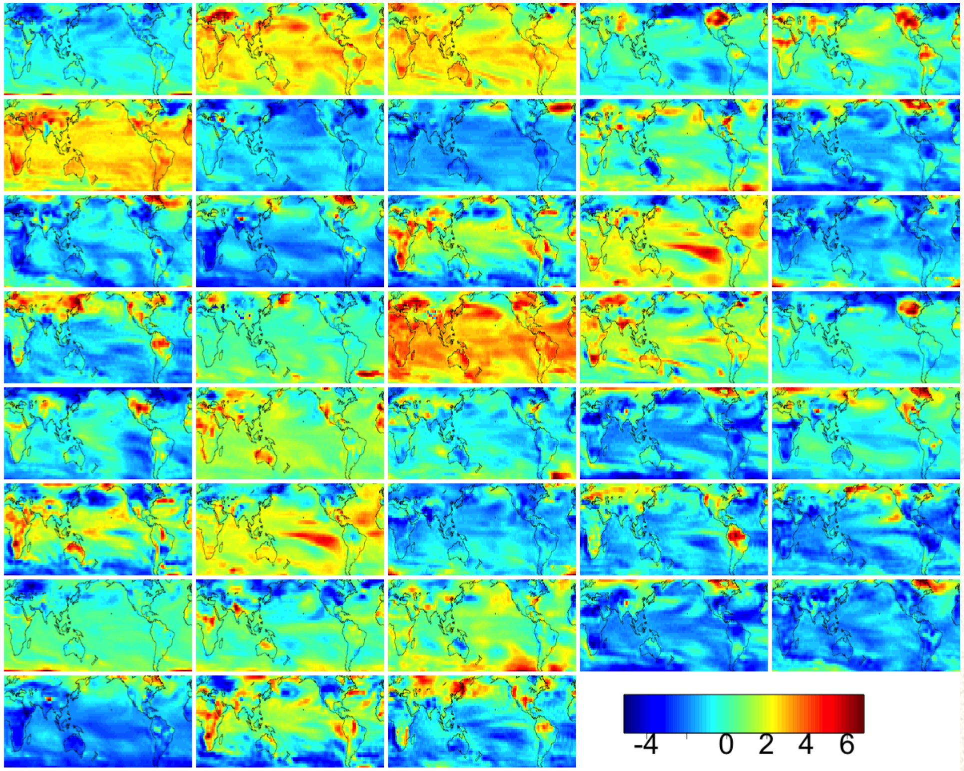
Represent covariance information in data assimilation

Compact way to blend variational and ensemble methods.

Represent uncertainty in multimodel climate experiments

Efficiently generate ensembles for integrated assessment models.

CMIP3 temperature residuals from pattern scaling (2 C warming).



Activities

Some events at IMAGe

Partial Differential Equations on the Sphere	April, 2014
Pattern Scaling, Climate Emulators and Scenarios.	April, 2014
Understanding Climate Change from Data	June, 2014
Summer Program: The Surface Temperature Initiative	July, 2014
Uncertainty in climate change research	July, 2014
Graduate Workshop on Environmental Data Analytics	July, 2014
Workshop on Climate Informatics	Sep., 2014

Analysis for large data, S. Sain and D. Nychka, Term A, CU-Boulder

Summary

- Efficient and flexible statistical models for large spatial data
- Community software (R) for laptops through supers
- Extensions to assimilation and for climate model emulation

Thank you!