A Multiresolution approach to nonstationary and efficient computing

Tomoko Matsuo and Douglas Nychka
National Center for Atmospheric Research

- Some equations, some problems
- Wavelet bases
- Sparsity
- Nonstationarity and EM

Supported by the National Science Foundation DMS

Why are we doing this?

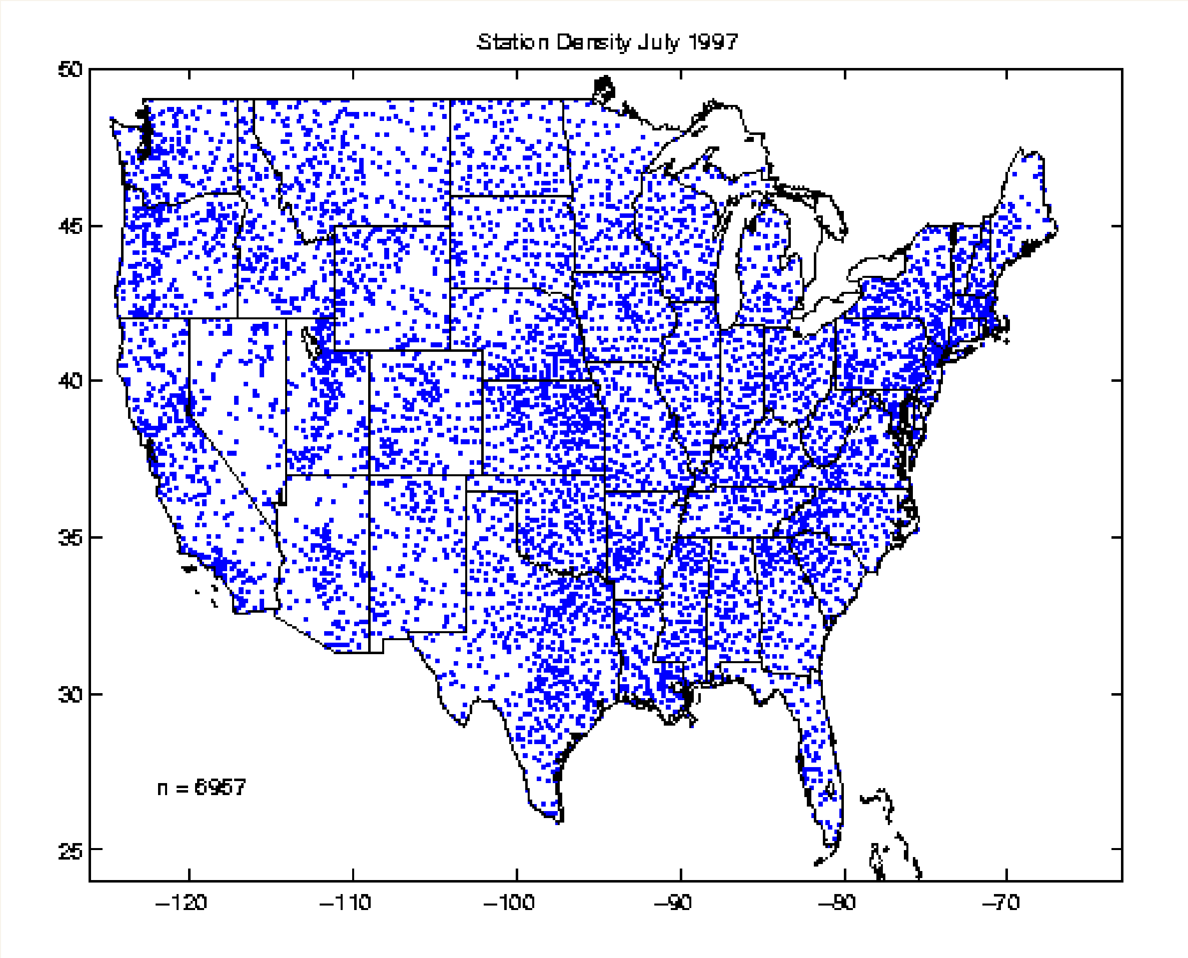
Many geophysical/biological processes are nonstationary over large area

- meteorological variables: precipitation
- forecast errors from a weather prediction
- electric field in the upper atmosphere
- pollutants: ambient ozone
- human health: disease incidence
- remotely sensed measurements

Also many interesting data sets are large.

A large spatial dataset

Reporting stations for monthly precipitation 1997.



Spatial Models

$z(\boldsymbol{x})$, is a random field, e.g. ozone concentration at location $\boldsymbol{x}$,

$$k(\boldsymbol{x}, \boldsymbol{x}') = COV(z(\boldsymbol{x}), z(\boldsymbol{x}'))$$

There are other parts of z that are important:

- $E(z(\boldsymbol{x}))$, fixed effects and covariates
- $z(\boldsymbol{x})$ is not Gaussian
- Copies of $z(\boldsymbol{x})$ observed at different times are correlated, e.g ozone fields for each day.

Spatial Models

$z(\boldsymbol{x})$, is a random field, e.g. ozone concentration at location $\boldsymbol{x}$,

$$k(\boldsymbol{x}, \boldsymbol{x}') = COV(z(\boldsymbol{x}), z(\boldsymbol{x}'))$$

There are other parts of z that are important:

- $E(z(\boldsymbol{x}))$, fixed effects and covariates
- $z(\boldsymbol{x})$ is not Gaussian
- Copies of $z(\boldsymbol{x})$ observed at different times are correlated, e.g ozone fields for each day.

I don't want to talk about these today!

The wavelet/Gaussian model is a platform for more complicated models, just as many methods use weighted least squares as a primitive.

Spatial Models (continued)

Let $\boldsymbol{u}$ be the field values on a large, regular 2-d grid (and stacked as a vector). This is our universe.

$$\Sigma = COV(\boldsymbol{u})$$

Spatial Models (continued)

Let $\boldsymbol{u}$ be the field values on a large, regular 2-d grid (and stacked as a vector). This is our universe.

$$\Sigma = COV(\boldsymbol{u})$$

Observational model

We observe part of $\boldsymbol{u}$, possibly with error.

$$\boldsymbol{Y} = K\boldsymbol{u} + \boldsymbol{e}$$

K is usually sparse , e.g. an incidence matrix of ones and zeroes for irregularly spaced data

$COV(\boldsymbol{e}) = R$
(also sparse, often diagonal)

Spatial Prediction

Assuming $\boldsymbol{Y}$ has zero mean.

$$\hat{\boldsymbol{u}} = \Sigma K^T (K \Sigma K^T + R)^{-1} \boldsymbol{Y}$$

and the covariance of the estimate is

$$\Sigma - \Sigma K^T (K \Sigma K^T + R)^{-1} K \Sigma$$

log likelihood for Σ:

$$-(1/2) log(|(K \Sigma_{\boldsymbol{\theta}} K^T + R)|) - (1/2) Y^T (K \Sigma_{\boldsymbol{\theta}} K + R)^{-1} Y + C$$

Spatial Prediction

Assuming $\boldsymbol{Y}$ has zero mean.

$$\hat{\boldsymbol{u}} = \Sigma K^T (K \Sigma K^T + R)^{-1} \boldsymbol{Y}$$

and the covariance of the estimate is

$$\Sigma - \Sigma K^T (K \Sigma K^T + R)^{-1} K \Sigma$$

log likelihood for Σ:

$$-(1/2) log(|(K \Sigma_{\boldsymbol{\theta}} K^T + R)|) - (1/2) Y^T (K \Sigma_{\boldsymbol{\theta}} K + R)^{-1} Y + C$$

An approximate posterior:

I like to think of the Kriging estimate as based on the conditional multivariate normal distribution of the grid points given the data: $[\boldsymbol{u}|\boldsymbol{Y}]$

The problems

Given a large numbers of irregularly spaced observations and the usual covariance models,

The problems

Given a large numbers of irregularly spaced observations and the usual covariance models,

Finding

- the spatial prediction: equations are hard to compute because $(K\Sigma K^T + R)^{-1}$ is large.

The problems

Given a large numbers of irregularly spaced observations and the usual covariance models,

Finding

- the spatial prediction: equations are hard to compute because $(K\Sigma K^T + R)^{-1}$ is large.
- the covariance formula: hopeless for many grid points

The problems

Given a large numbers of irregularly spaced observations and the usual covariance models,

Finding

- the spatial prediction: equations are hard to compute because $(K\Sigma K^T + R)^{-1}$ is large.
- the covariance formula: hopeless for many grid points
- MLEs for covariance parameters:

The problems

Given a large numbers of irregularly spaced observations and the usual covariance models,

Finding

- the spatial prediction: equations are hard to compute because $(K\Sigma K^T + R)^{-1}$ is large.
- the covariance formula: hopeless for many grid points
- MLEs for covariance parameters:

  Don't even think about it!

Creating a random function

Start with a set of *fixed* basis functions $\{\psi_j\}$

Multiply them times *random* coefficients

Add 'em up.

$$z(x) = \sum_j a_j \psi_j(x)$$

Creating a random function

Start with a set of *fixed* basis functions $\{\psi_j\}$

Multiply them times *random* coefficients

Add 'em up.

$$z(x) = \sum_j a_j \psi_j(x)$$

The choice of the basis $\{\psi_j\}$ and the covariances among the coefficients $\{a_j\}$ are the key to this work.

In matrix/vector notation:

$$\boldsymbol{z} = \Psi \boldsymbol{a}$$

Each column of Ψ is a basis function evaluated on a grid.

In matrix/vector notation:

$$\boldsymbol{z} = \Psi \boldsymbol{a}$$

Each column of Ψ is a basis function evaluated on a grid.

From linearity

$$COV(\boldsymbol{z}) = \Sigma = \Psi COV(\boldsymbol{a}) \Psi^T = \Psi H H^T \Psi^T$$

In matrix/vector notation:

$$\boldsymbol{z} = \Psi \boldsymbol{a}$$

Each column of Ψ is a basis function evaluated on a grid.

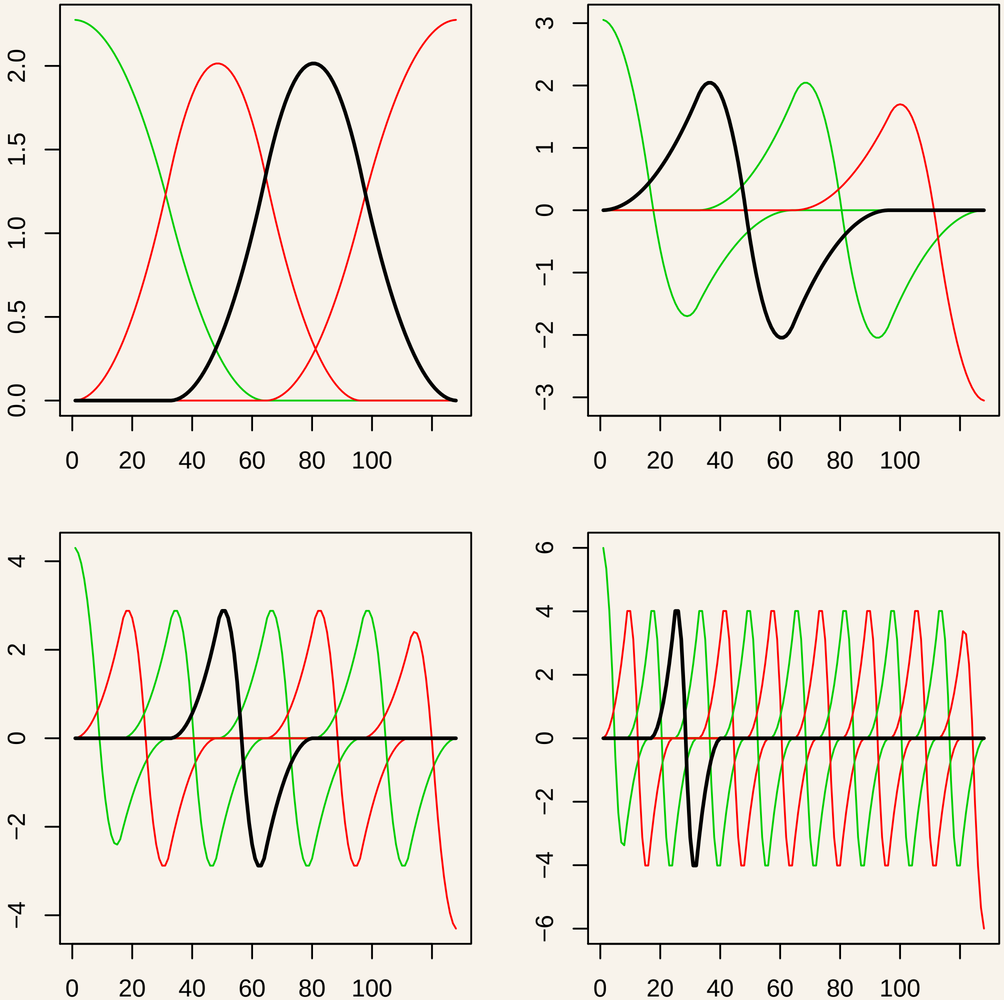
From linearity

$$COV(\boldsymbol{z}) = \Sigma = \Psi COV(\boldsymbol{a}) \Psi^T = \Psi H H^T \Psi^T$$

The big idea:

The basis functions are localized and H is nearly diagonal.

A set of 32 1-d multiresolution basis functions



Bases are organized in levels and are translations and scalings of fixed father and mother wavelets.

Why wavelets?

Reduction in complexity

A complicated random function may have a simpler representation by specifying the covariances among the coefficients instead of directly on the function.

Local support:

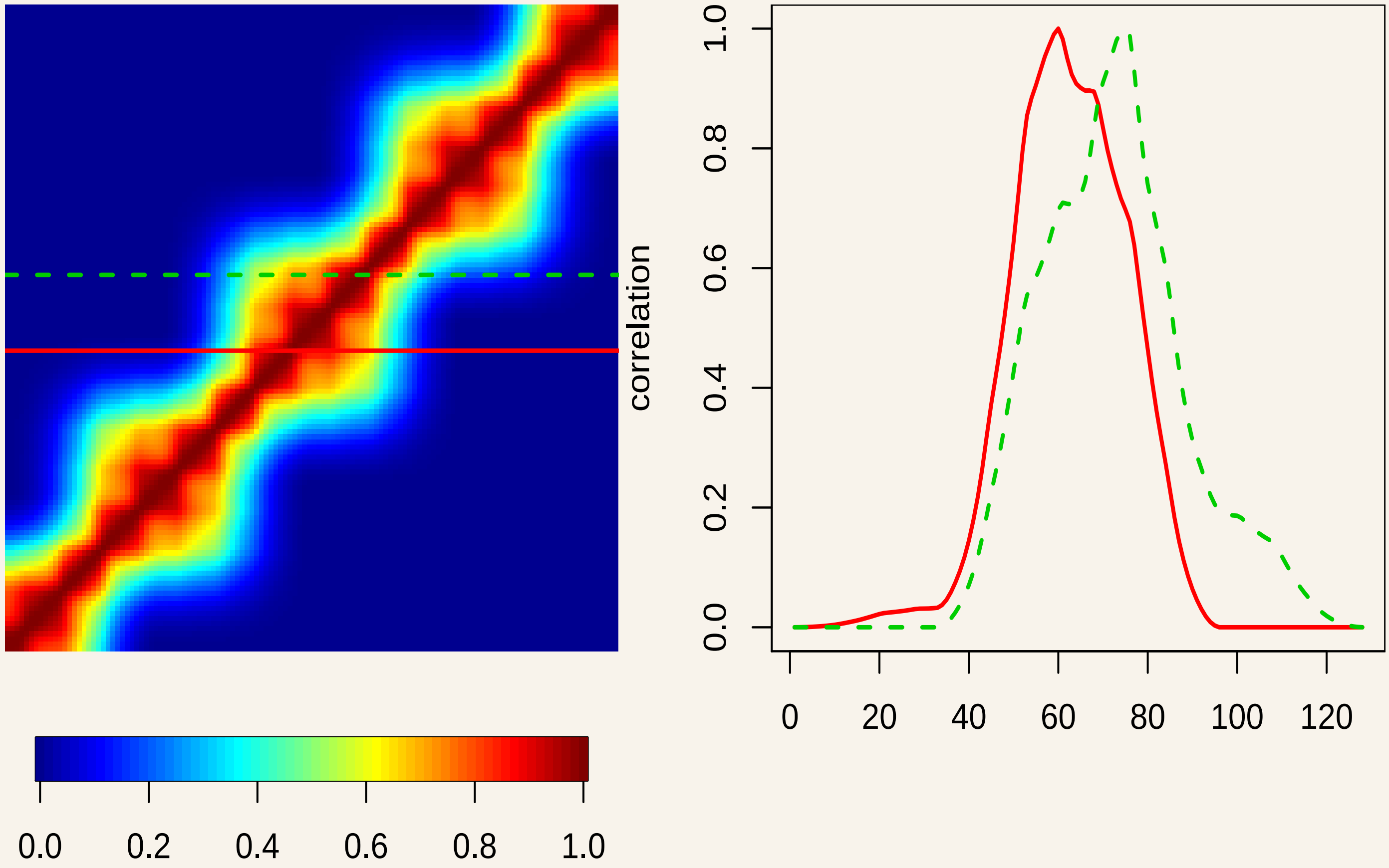
Wavelets can represent nonstationary covariances easily because individual basis functions are associated with specific locations. Local support also leads to fast computation with the basis functions.

Another good thing:

The covariances among the $\{a_j\}$ are sparse. This leads to efficient algorithms for spatial prediction e.g. generating *ensembles* – sampling the posterior distribution.

What can go wrong.

Assuming the coefficients are uncorrelated (H is diagonal):



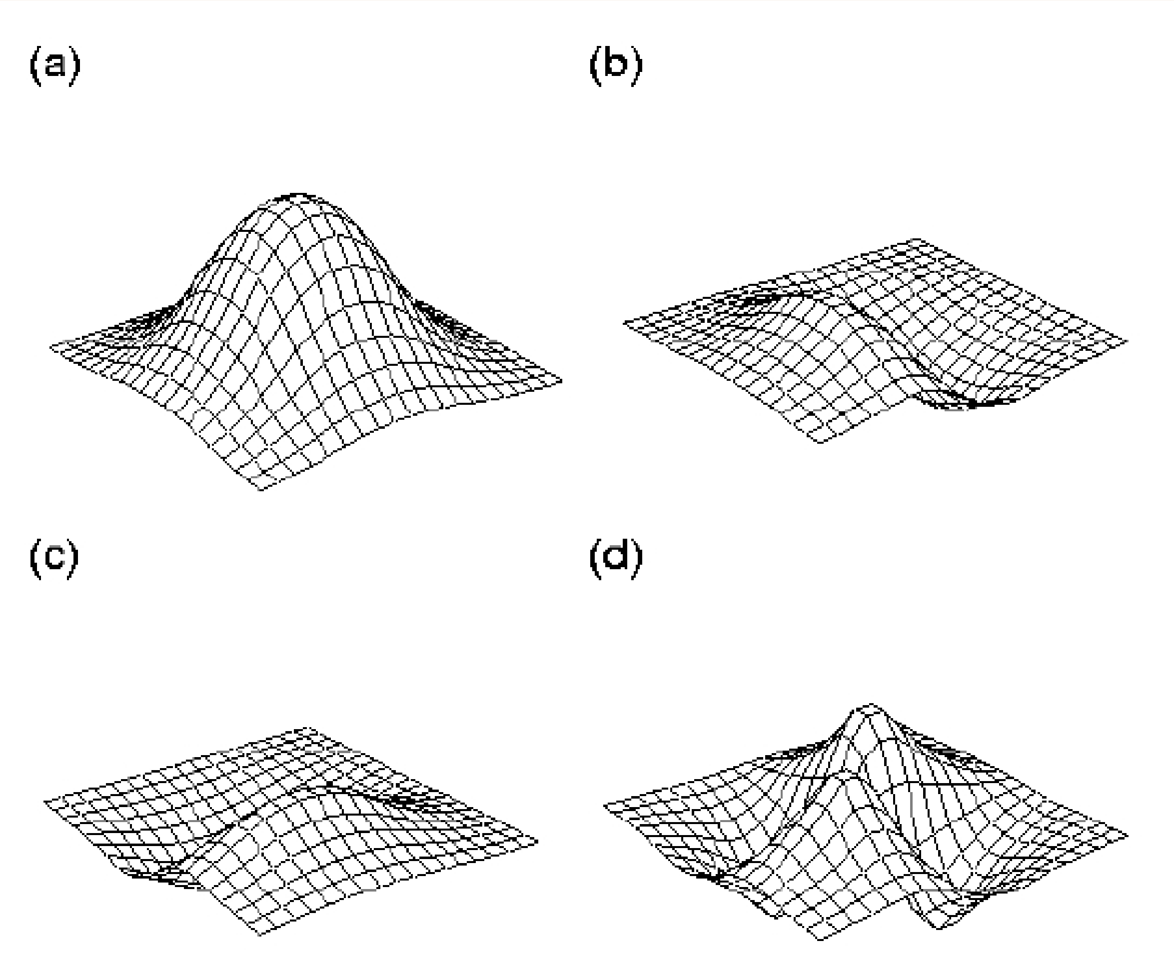
Covariance Models

Covariance of the field, $\Sigma = \Psi H^2 \Psi^T$

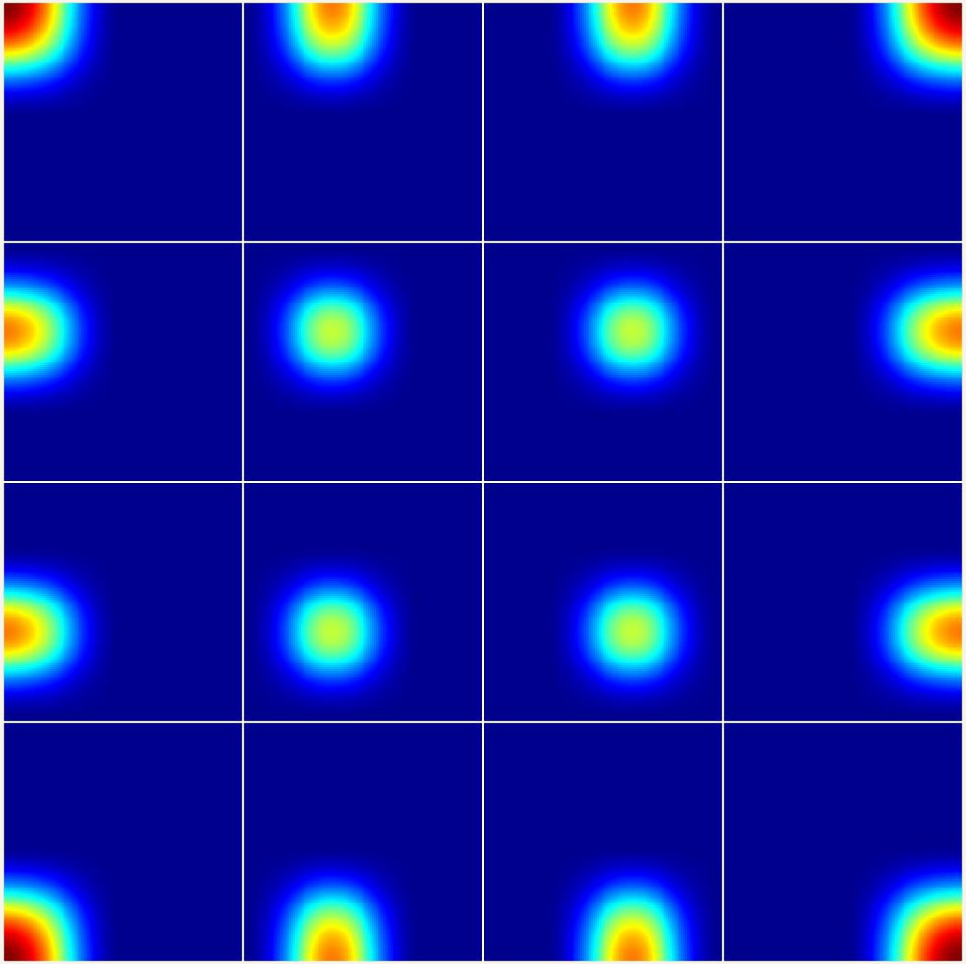
So the goal is to find good choices for H

Two dimensions: (1) father and (3) mothers

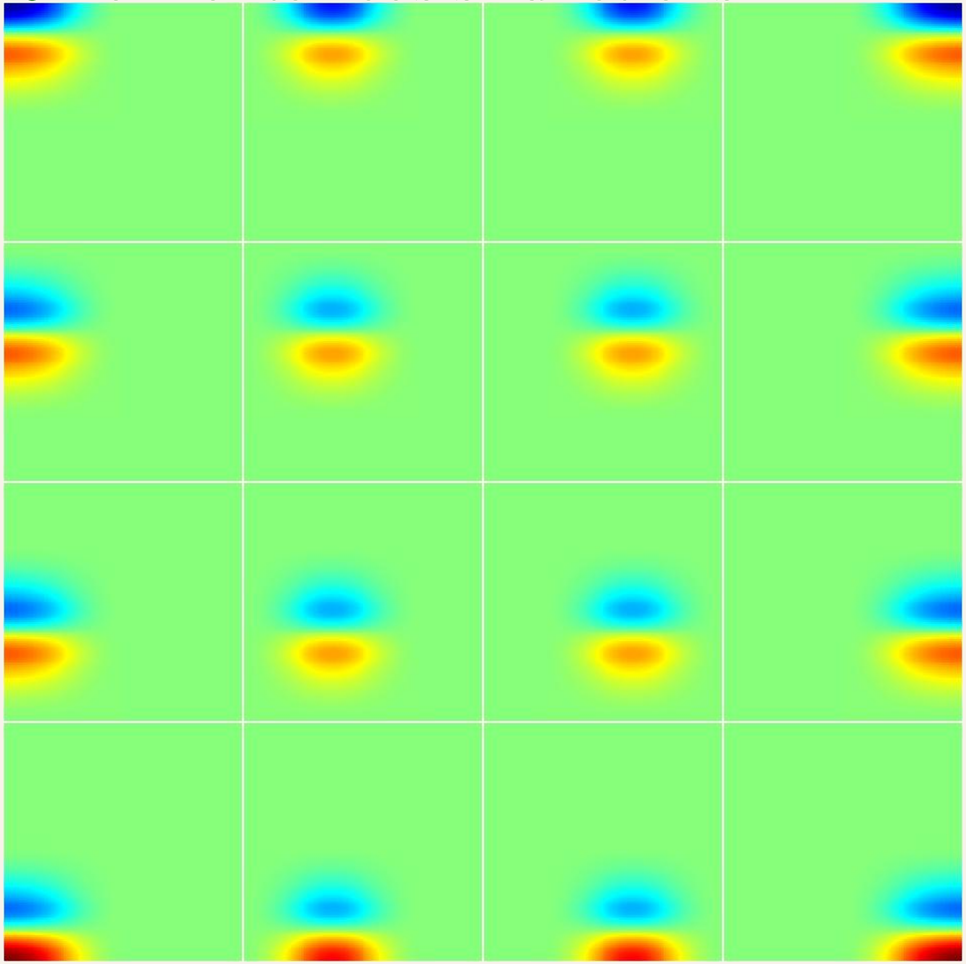
An Old Testament version!



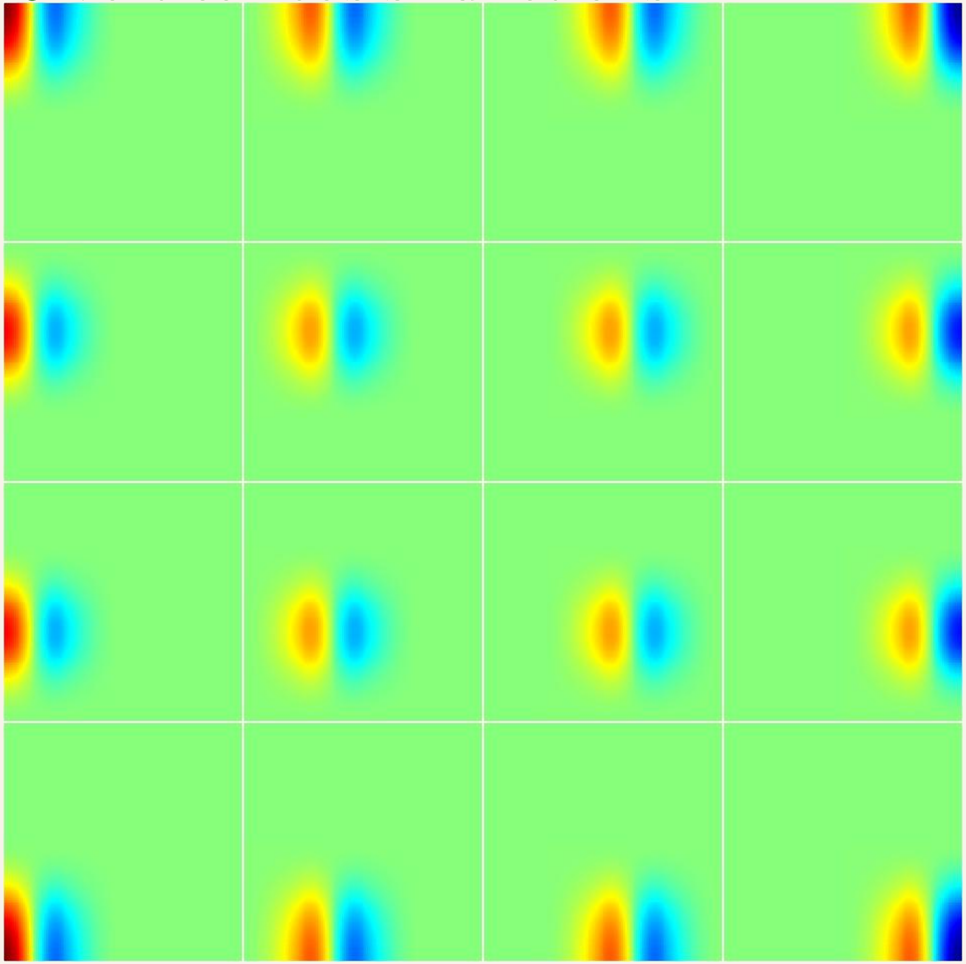
16 Smooth (father) basis functions



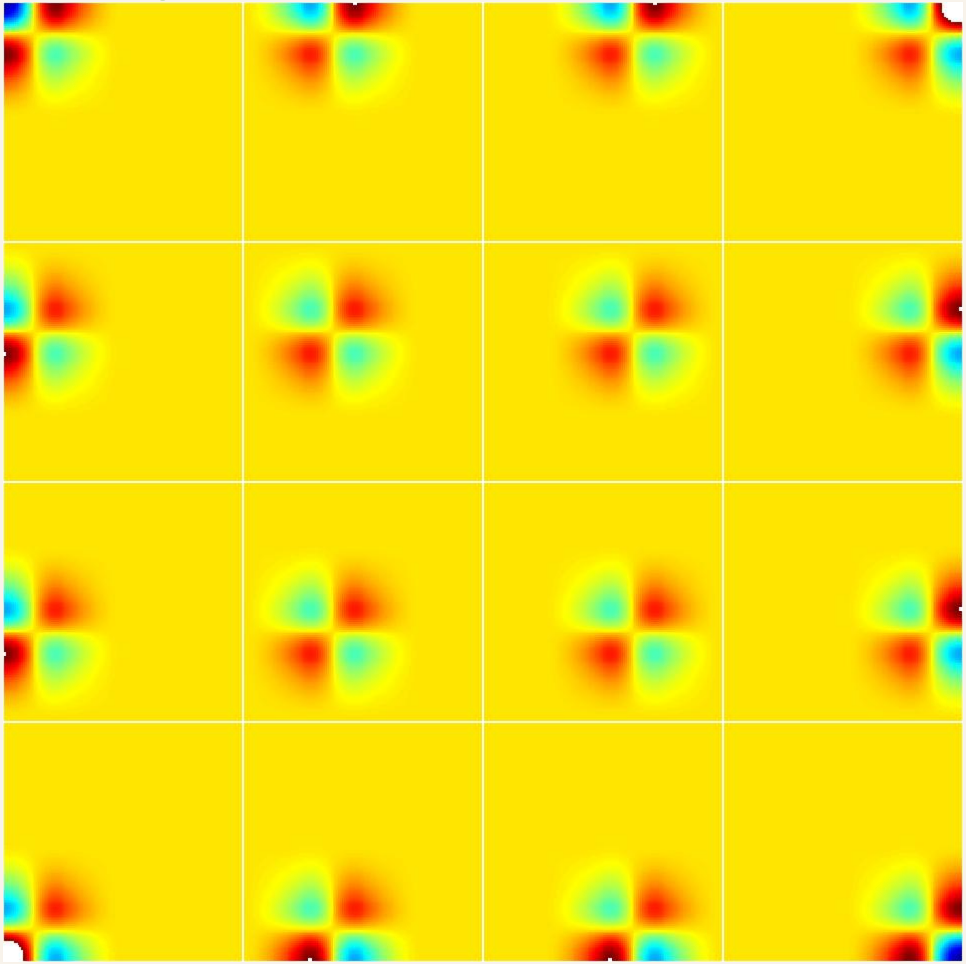
16 horizontal basis functions



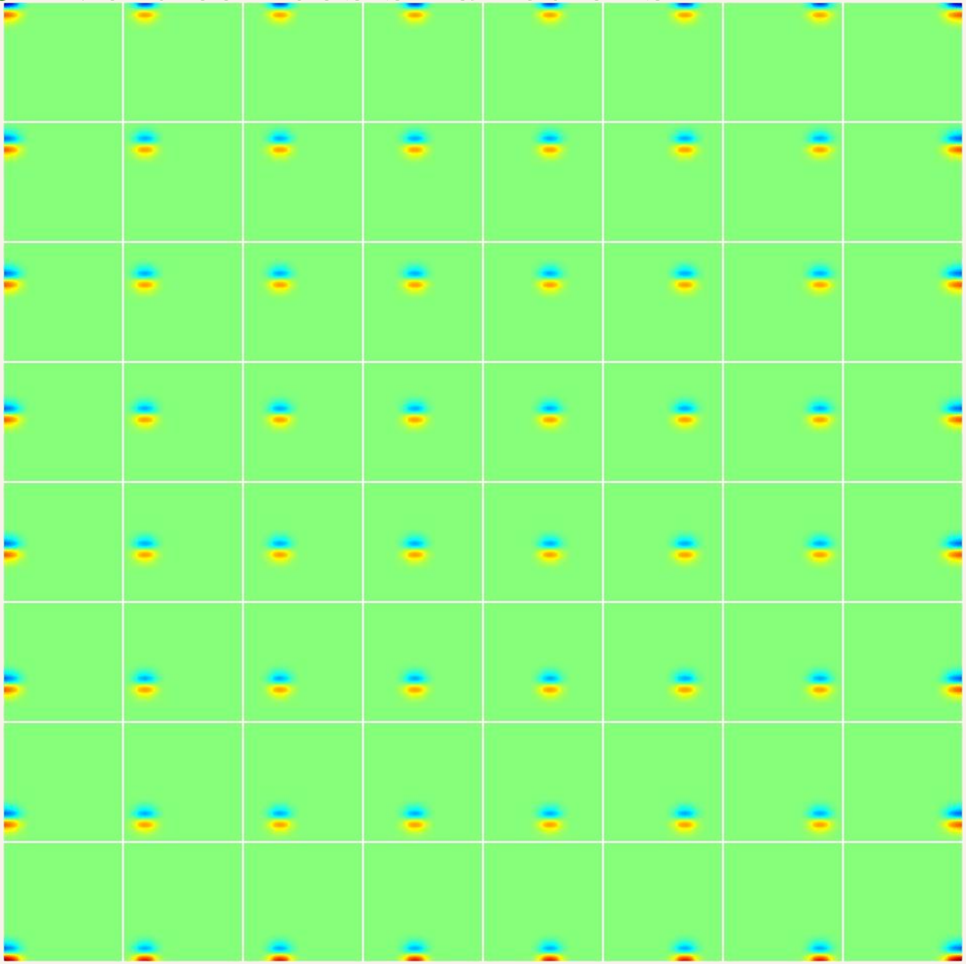
16 vertical basis functions



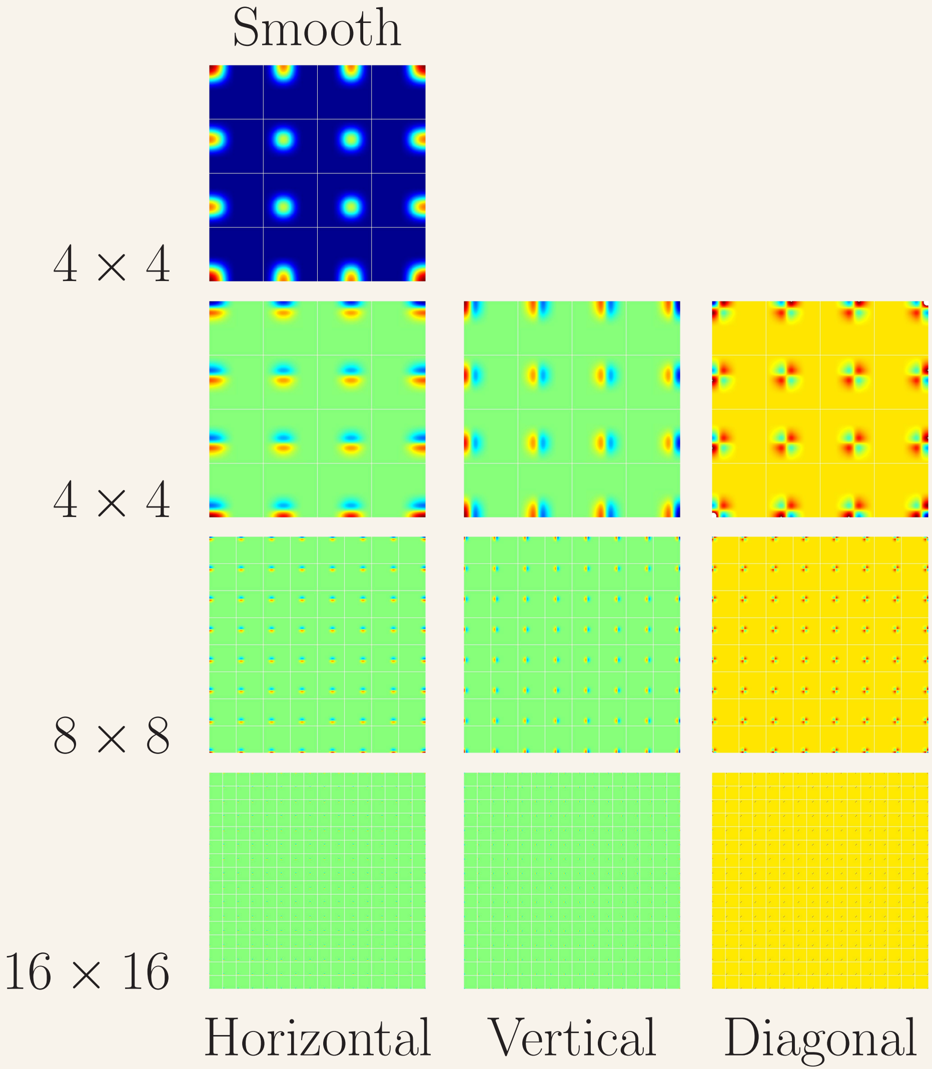
16 diagonal basis functions



64 vertical basis functions



Organizing $1024 = 32 \times 32$ basis functions in a lattice



What does H look like for a Matern Σ?

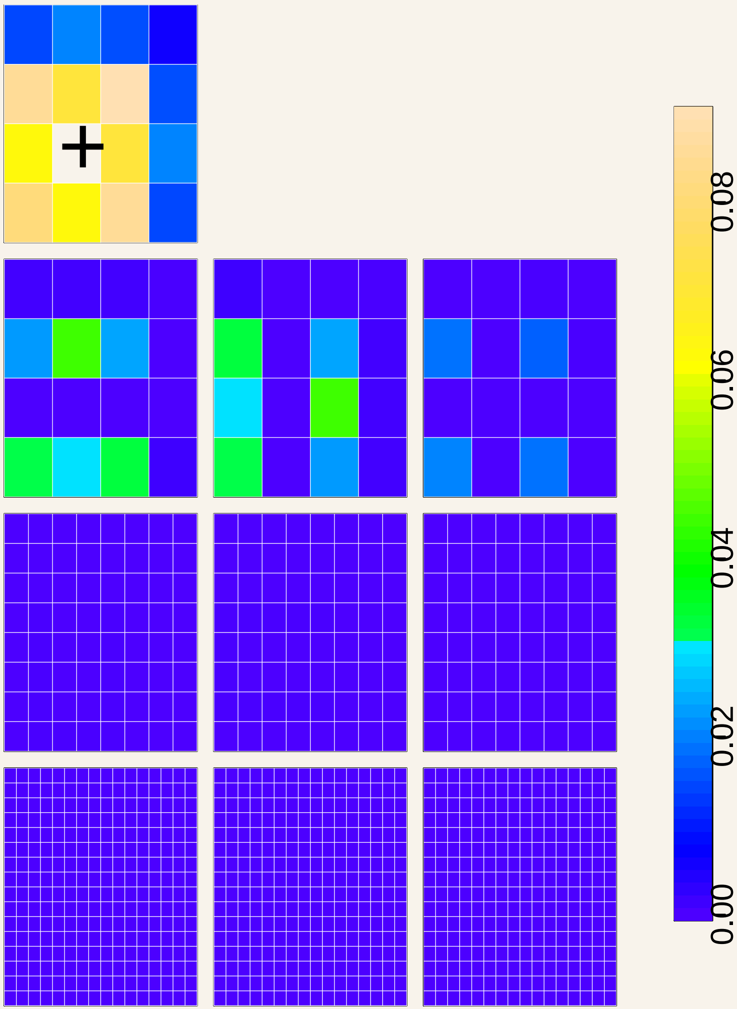
Suppose Σ follows a Matern covariance smoothness 1 and range .2 on a 32×32 grid in $[0, 1] \times [0, 1]$.

$$\Sigma = \Psi H^2 \Psi^T$$

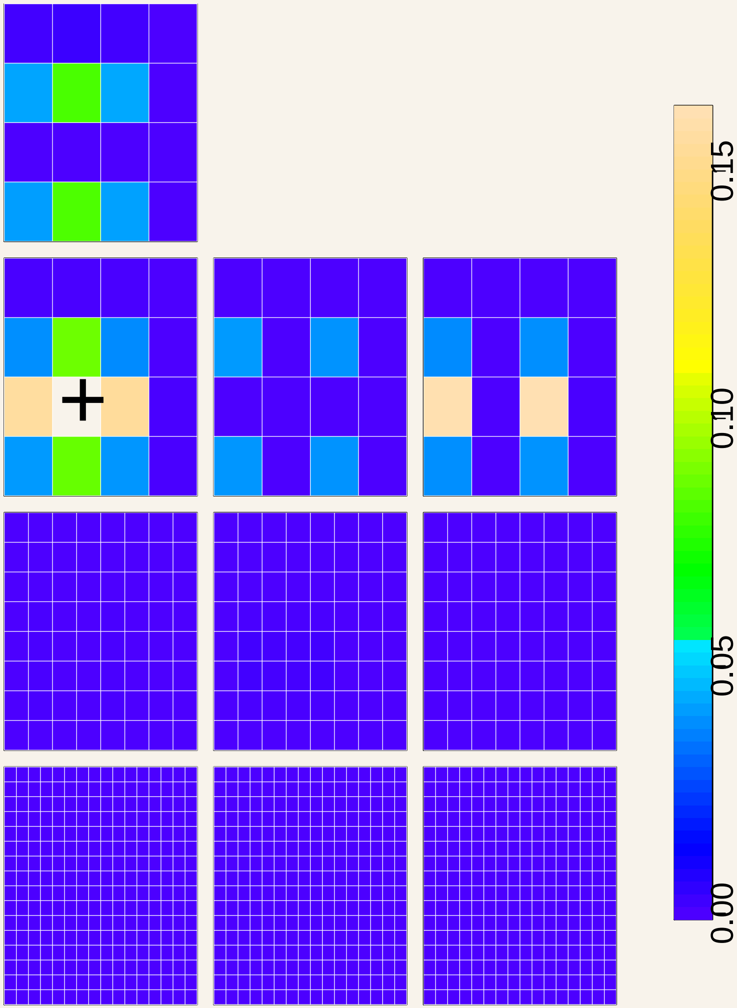
$$H^2 = \Psi^{-1} \Sigma (\Psi^T)^{-1}$$

where H is the *symmetric* square root.

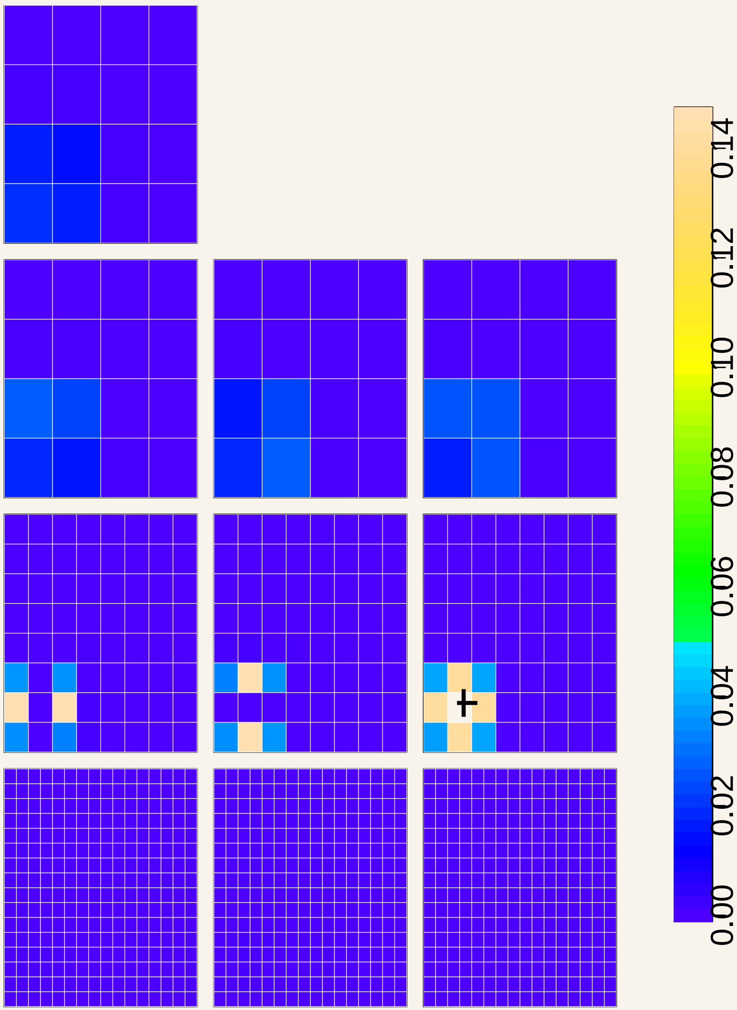
Smooth basis function, (range=.2, smoothness=1)



Horizontal basis function at first level



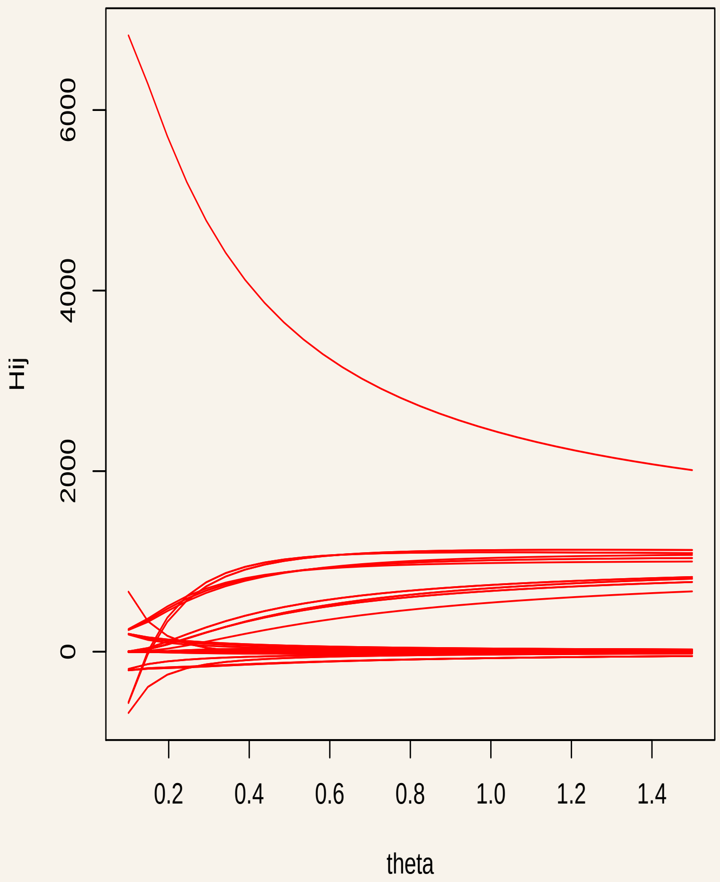
Diagonal basis function at second level



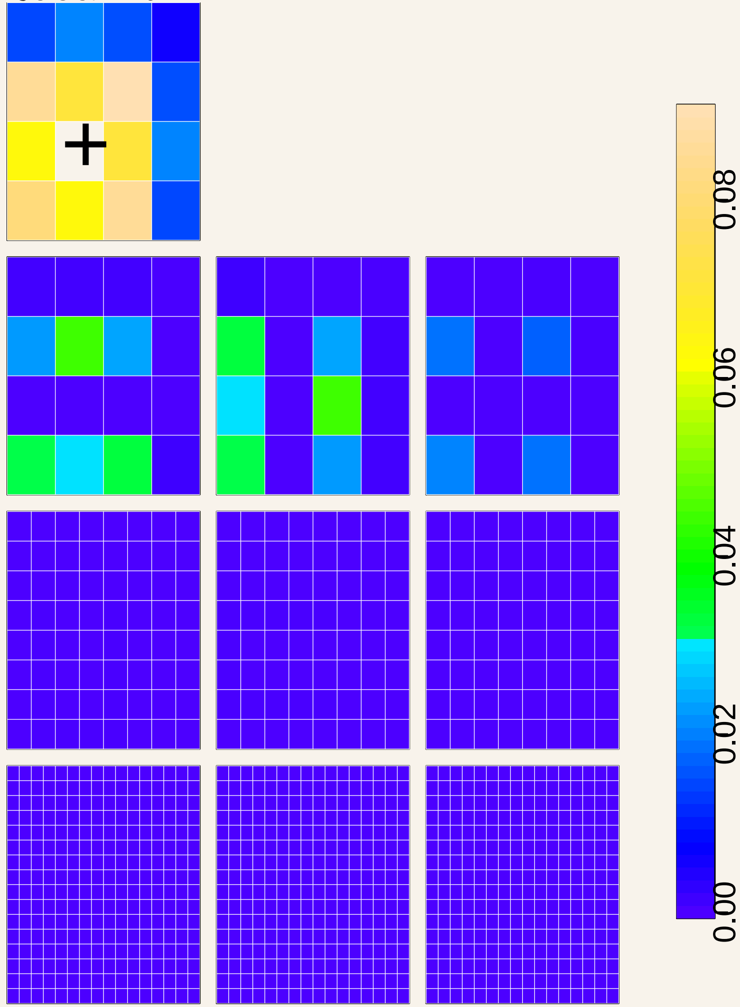
Trace plots of $H_{\boldsymbol{\theta}}$

How do the elements of H change as the range is varied?

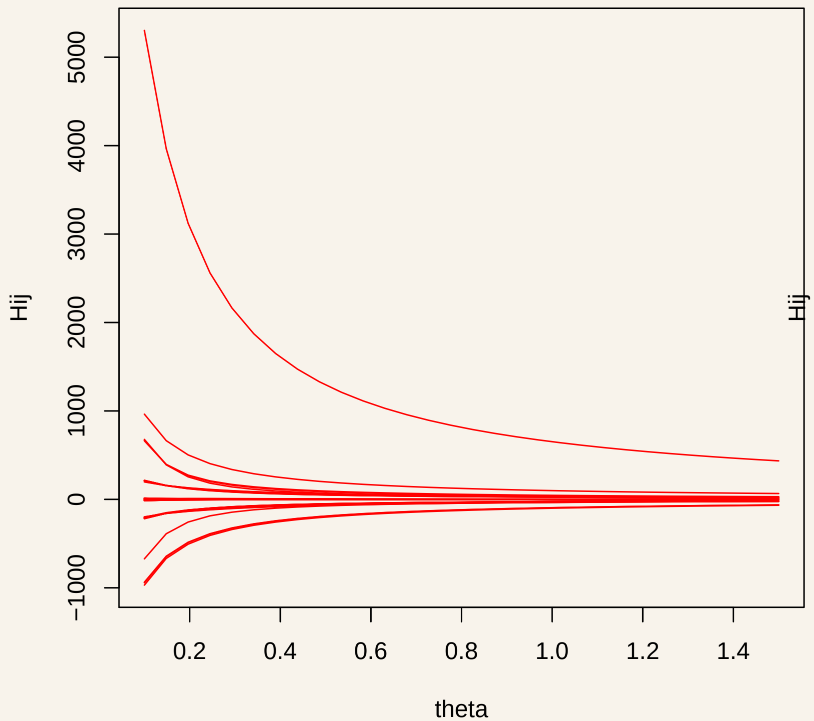
Row of H smooth (2,2) as the range varies

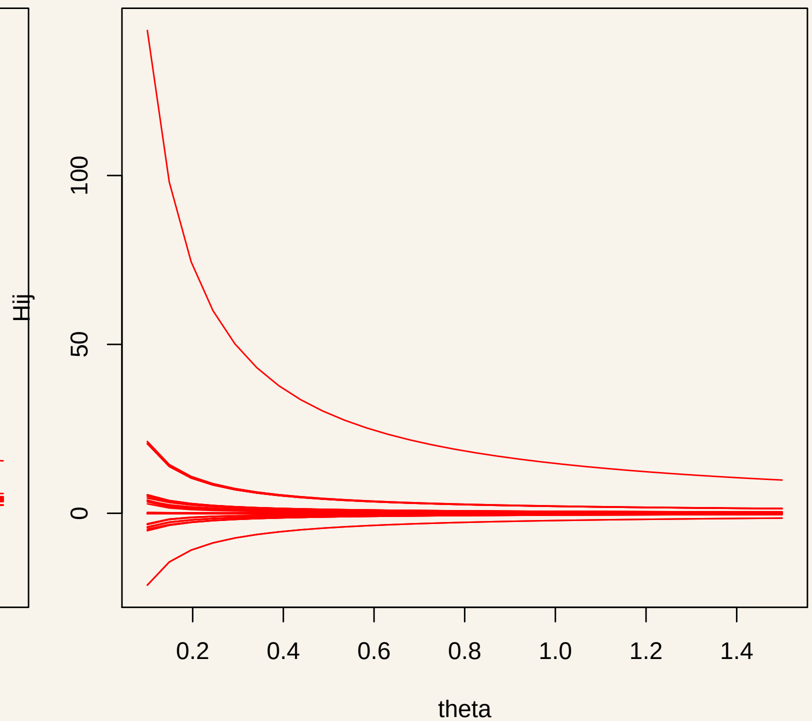


Recall $\theta = .2$



Row of H for mother (2,2)   *Row of H for daughter (2,2)*





Introducing sparseness

Decimation:

Set all small elements of H that are small to zero.

$H_{\theta,\nu}$ Gives an accurate approximation to the Matern family.

H, its inverse and tranposes are now efficient for computing.

The number of nonzero elements goes linearly with the number of observations.

Introducing nonstationarity

We are interested in a spatial field that in the neighborhood of $\boldsymbol{x}$ roughly has a Matern covariance with range parameters $\theta(\boldsymbol{x})$.

This is the *Higdon/Fuentes/Paciorek* model.

Introducing nonstationarity

We are interested in a spatial field that in the neighborhood of $\boldsymbol{x}$ roughly has a Matern covariance with range parameters $\theta(\boldsymbol{x})$.

This is the *Higdon/Fuentes/Paciorek* model.

Modifying rows of H

If a wavelet basis functions is localized around $\boldsymbol{x}$ then the corresponding row of H is based on $\theta(\boldsymbol{x})$.

$$\Sigma_\theta = W H_\theta H_\theta^T W$$

Introducing nonstationarity

We are interested in a spatial field that in the neighborhood of $\boldsymbol{x}$ roughly has a Matern covariance with range parameters $\theta(\boldsymbol{x})$.

This is the *Higdon/Fuentes/Paciorek* model.

Modifying rows of H

If a wavelet basis functions is localized around $\boldsymbol{x}$ then the corresponding row of H is based on $\theta(\boldsymbol{x})$.

$$\Sigma_\theta = W H_\theta H_\theta^T W$$

A Hierarchical Model: (theta is a spatial field)

$$[Y|\Sigma_\theta] \quad [log(\theta)|\gamma] \quad [\gamma]$$

where the likelihood is computationally tractable due to sparseness!

Irregularly spaced locations

Register each location to a grid box.

An EM approach:

Step 0 Start with an initial (possibly nonstationary) covariance model H found from the irregular locations.

E-Step Generate (*random*) complete fields from the conditional distribution using this H.

M-Step Apply a wavelet-based covariance estimate to the complete realizations.

Repeat E and M Steps until covariance estimate converges.

Nonparametric estimates from data.

Sample estimates of H

With complete data and (independent) replications over time, one can get sample estimates of the elements of H. The amount of computation and storage is of the order of the image size and time points, not (image size)^2.

Decimation

The sparseness of H guarantees that we do not have to look at (or even compute) many off-diagonal elements. The elements that are nonzero based on the Matern family are reasonable choice.

Once the elements of $\widehat{H}$ are found one can:

- decimate them
- smooth across "spatially adjacent" entries.
- shrink toward a stationary model

Concluding remarks

- Wavelets provide flexible methods for introducing nonstationary spatial structure at different spatial scales. But they can also reproduce standard spatial models.
- Wavelet bases are well suited for computation with large data sets.