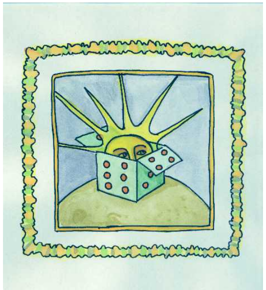
Large, nonstationary spatial fields

Douglas Nychka, Andy Royle and Chris Wikle

Geophysical Statistics Project,
National Center for Atmospheric Research
`http://www.cgd.ucar.edu/stats/`

- Spatial process estimates and covariance models
- Conditional simulation
- Multiresolution bases - decimation
- An example: Ozone model output for the Midwest



Why are we doing this?

Many geophysical/biological processes are nonstationary over large area

- meteorological variables: precipitation
- pollutants: ambient ozone
- human health: disease incidence

Analyzing numerical model output

Given the output of a geophysical model, the covariance function of a simulated field is a useful summary of the model behavior. Also this covariance can be used for sampling designs and spatial prediction when the observed data is sparse.

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 build off of weighted leastsquares as a primitive.

Spatial Models (continued)

Discretize the problem

$\boldsymbol{z}$ be the field values on a large, regular 2-d grid (and stacked as a vector).

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

Locations are discretized to the nearest grid point.

Basis functions

If we find the eigenvector/eigenvalue decomposition for Σ.
$\Sigma = \Psi D \Psi^T$, $\Psi\Psi^T = I$ and D diagonal.
then $\boldsymbol{z} = \Psi \boldsymbol{a}$ where $\boldsymbol{a}$ is random with $COV(\boldsymbol{a}) = D$.
If $D = H^2$ then $\boldsymbol{a} = H\boldsymbol{e}$ where $\boldsymbol{e}$ $iidN(0,1)$.

Some key ideas

The eigen decomposition suggests an alternative way of building the covariance by specifying the basis functions and D.

But ....

Ψ need not be orthogonal and D need not be diagonal.

The main constraint is that the spatial estimator be computable.

Observational model

Divide up $\boldsymbol{z}$ into two pieces:

$$\boldsymbol{z} = \begin{pmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \end{pmatrix} = \begin{pmatrix} observed \\ grid \end{pmatrix}$$

The goal is to estimate $\boldsymbol{z}_2$ using $\boldsymbol{z}_1$.

A more realistic observation model

$$\boldsymbol{y} = J\boldsymbol{z} + \boldsymbol{e}$$

J is a known "observational functional" such as incidence matrices of ones and zeroes for irregularly spaced data or weighted averages. $COV(\boldsymbol{e}) = \sigma^2 I$ (where part of the variability may be due to discretization error.)

Estimation and Inference

So what do we do with the covariance once we have it?

Find the conditional distribution of unobserved given observed. $[\boldsymbol{z}_2|\boldsymbol{z}_1]$.

Under normality, conditional distribution of $\boldsymbol{z}_2$ for fixed covariance is Gaussian.

Conditional mean

$$\hat{\boldsymbol{z}}_2 = COV(\boldsymbol{z}_2, \boldsymbol{z}_1)\left[COV(\boldsymbol{z}_1, \boldsymbol{z}_1)\right]^{-1}\boldsymbol{z}_1 = \Sigma_{2,1}\Sigma_{1,1}^{-1}\boldsymbol{z}_1$$

Conditional covariance

$$\Sigma_{2,2} - \Sigma_{2,1}\Sigma_{1,1}^{-1}\Sigma_{1,2}$$

The problems

Σ_{11} and Σ_{21} can big! ... And so are Ψ and J !

Solutions

1. Use an iterative method for the solution of large linear systems to find the conditional mean.
2. Estimate variability by generating samples from the conditional distribution. This can be done by reusing the code for the finding the conditional mean.

A Diversion: Conditional Simulation

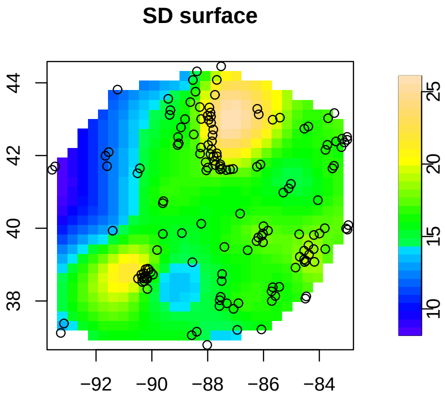
We want to generate a sample from

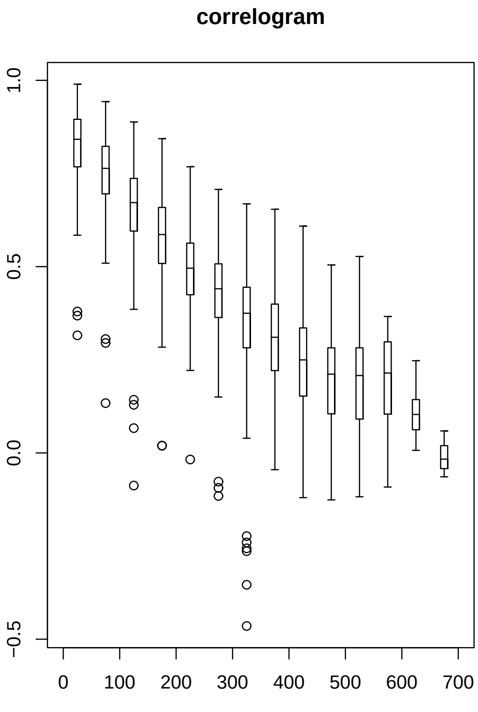
$$MN(\Sigma_{2,1}\Sigma_{1,1}^{-1}\mathbf{z}_1, \Sigma_{2,2} - \Sigma_{2,1}\Sigma_{1,1}^{-1}\Sigma_{1,2})$$

Trick is to reuse the basic Kriging estimator

1. Find $\hat{\mathbf{z}}_2 = \Sigma_{2,1}\Sigma_{1,1}^{-1}\mathbf{z}_1$ (just do this once)
2. Generate a synthetic field $\mathbf{z}^* = (\mathbf{z}_1^*, \mathbf{z}_2^*)$. (using square root of Σ)
3. $\mathbf{e} = \mathbf{z}_2^* - \Sigma_{2,1}\Sigma_{1,1}^{-1}\mathbf{z}_1^*$ (error in prediction using this bogus data)
4. Conditional field: $[\mathbf{z}_2|\mathbf{z}_1] = \hat{\mathbf{z}}_2 + \mathbf{e}$

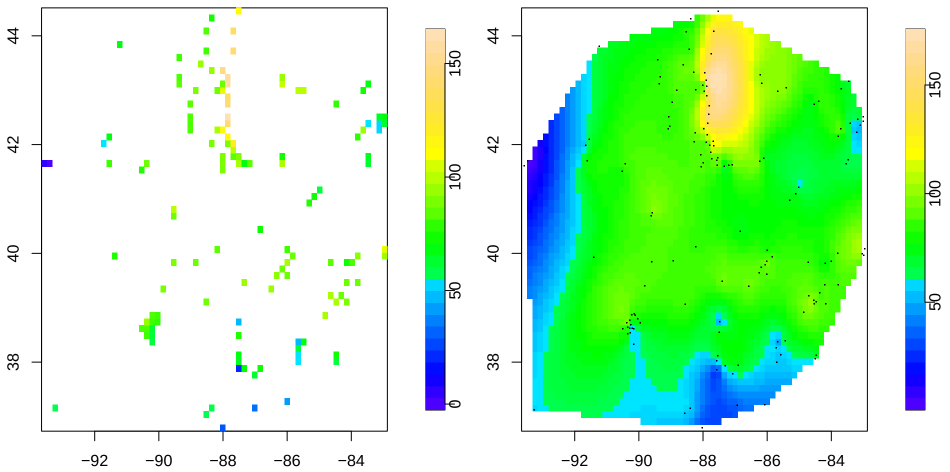
Example using 8 hour average Ozone



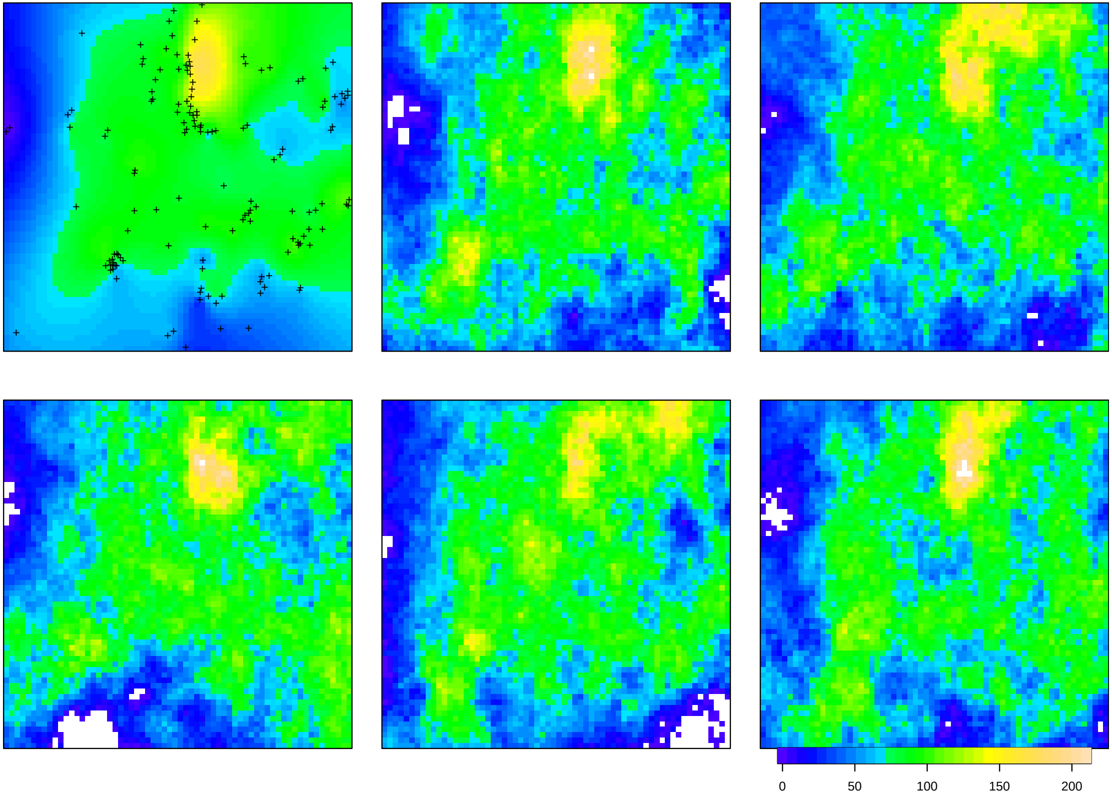


Kriging surface for day 16

Data and the conditional mean surface



Mean surface and 5 draws from the ”posterior”



Remarks about computing

Iterative methods demand that one can multiply Σ with a vector efficiently. For irregularly spaced data, do the whole multiplication (Σ) and just keep the parts you need ($\Sigma_{1,1}$)

Recall that $\Sigma = \Psi D \Psi^T$ so fast multiplication of Σ can be done by fast multiplications of Ψ, D and Ψ^T.

Sampling the conditional distribution requires that one can simulate a process at all grid points. e.g. Need $\Sigma^{1/2} = \Psi D^{1/2}$.

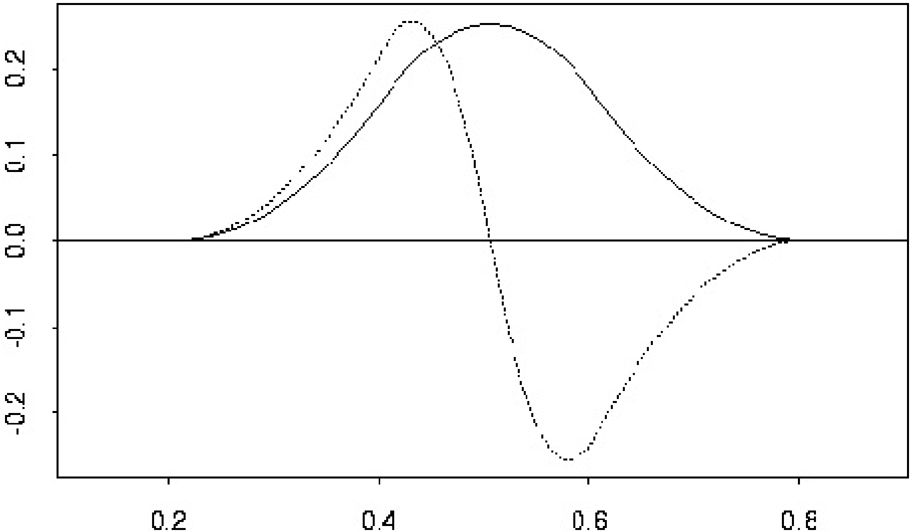
Estimating D is easier if we have Ψ^{-1}.

Multiresolution Bases

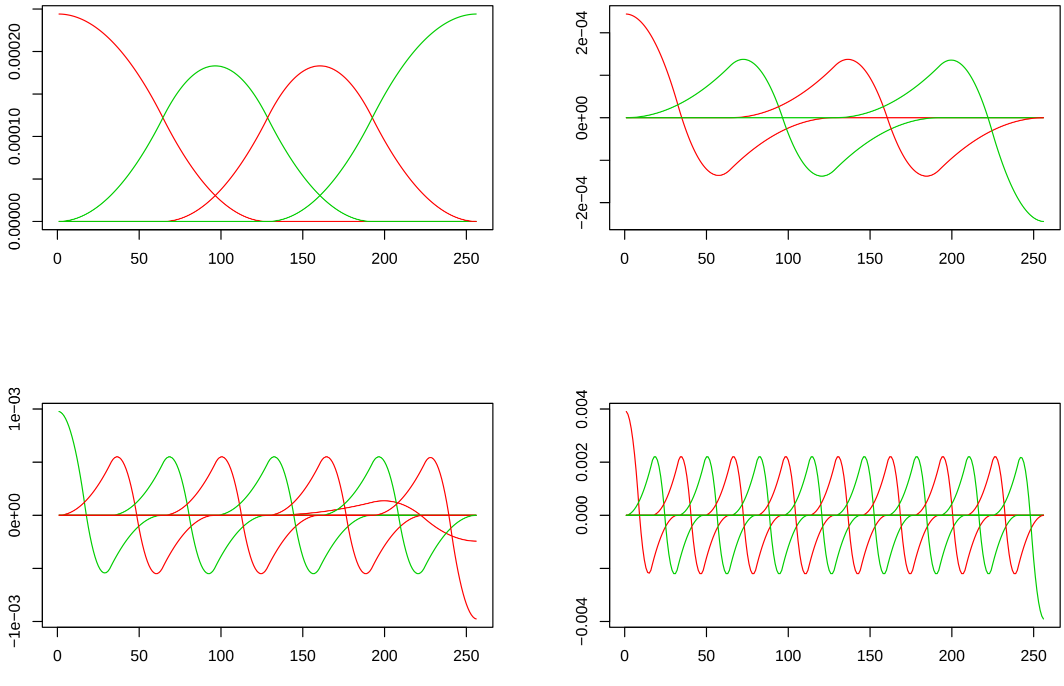
- translations and scalings of a few fixed functions
- have local support
- defined recursively
- $\Psi \boldsymbol{z}$ and $\Psi^T \boldsymbol{z}$ can be multiplied quickly

Key idea: Because the basis functions are locally supported changing the variances and covariances of individual coefficients will only have a local impact.

1-d Mother and father wavelets



Building up a family (Starting with 4)



Approximating other covariances

(No data yet!)

If Σ is any covariance matrix for the grid points.

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

One can always find a D or H that will work.

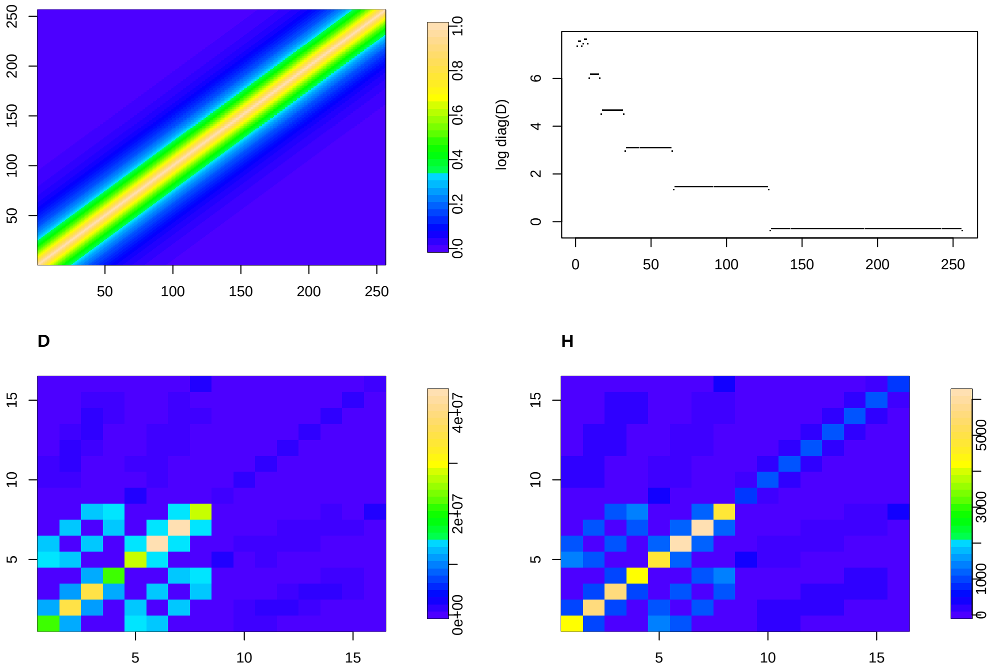
Also

$$D = \Psi^{-1} \Sigma \big(\Psi^T\big)^{-1}$$

However, the decomposition is only *useful* if we can find D ... or an H and it is close to diagonal.

1-d exponential example (range of 8 on [0,256])

What do D and H look like?



Sparseness of H

Decorrelation

For $H = D^{1/2}$ want H to be sparse. But wavelets *decorrelate* a spatial signal so we can expect many elements of H to be near zero.

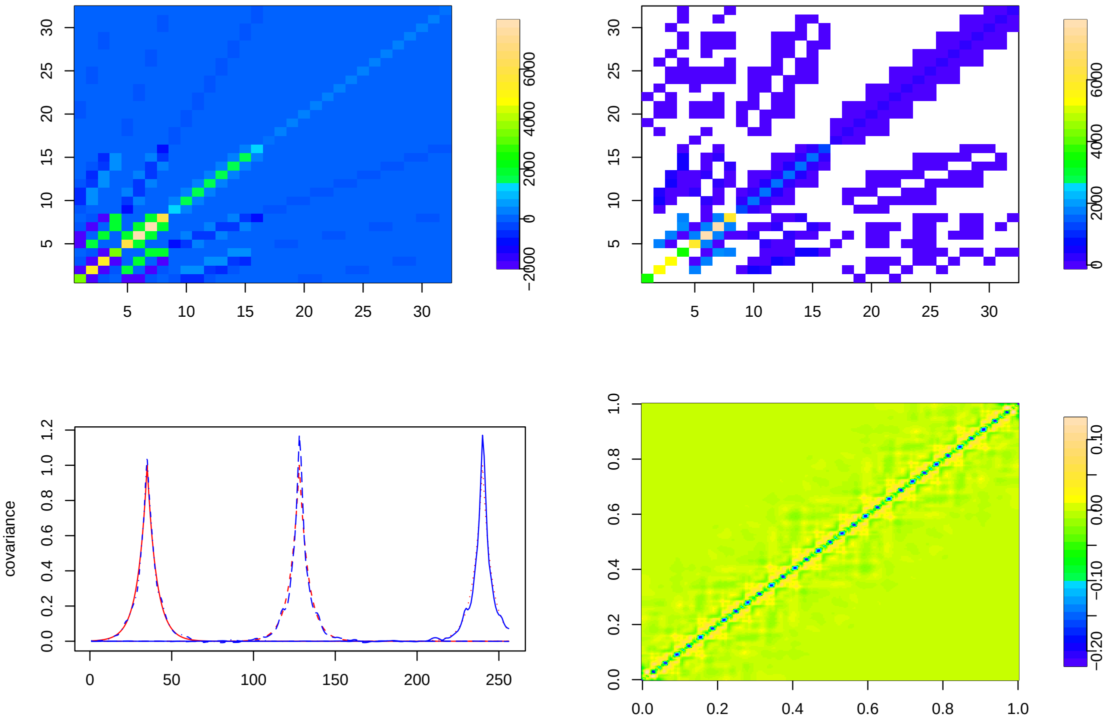
Decimation of H

Only keep the large values of H.

$$\widehat{H}_{i,j} = \begin{array}{ll} H_{i,j} & \text{if } |H_{i,j}| > \tau \\ 0 & \text{otherwise} \end{array}$$

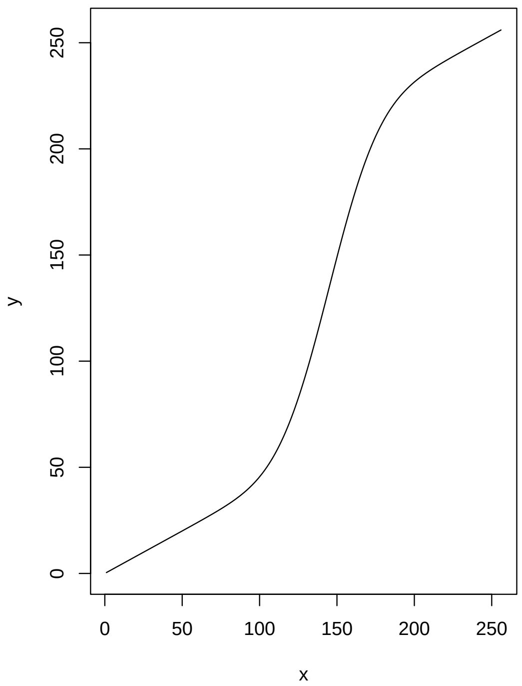
Zero out all off diagonal elements for basis functions that are below a set resolution.

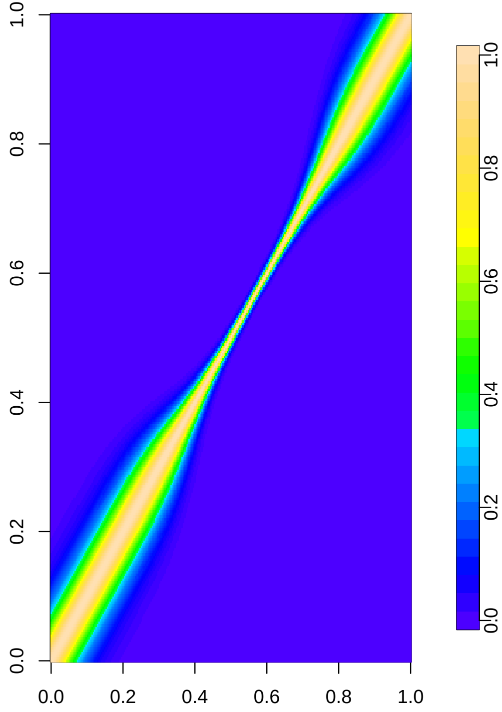
Approximation by wavelets: 98 % decimation of H!



Nonstationary, 1-d deformation

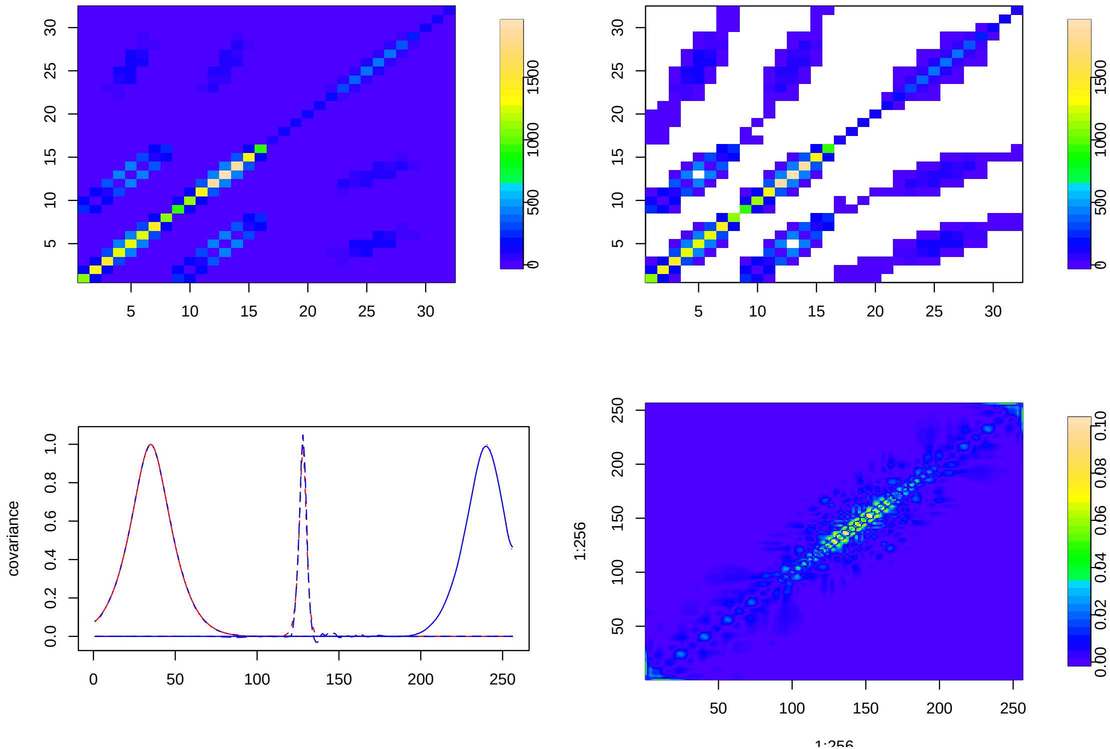
The deformation and the 1-d covariance matrix.



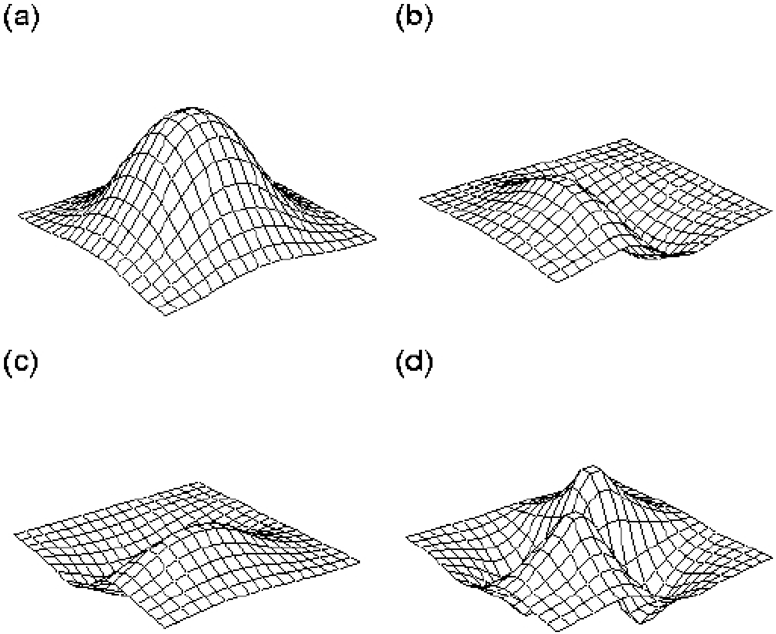


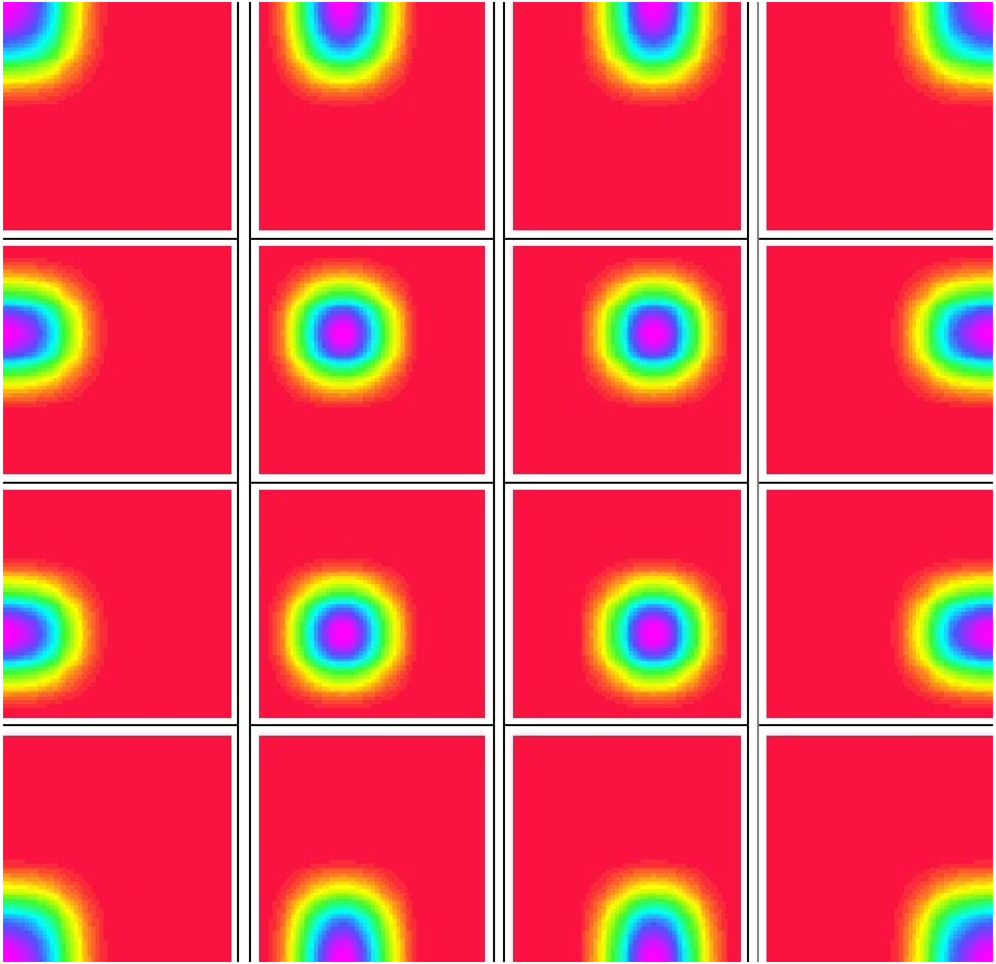
Approximation by wavelets 98 % decimation of H!

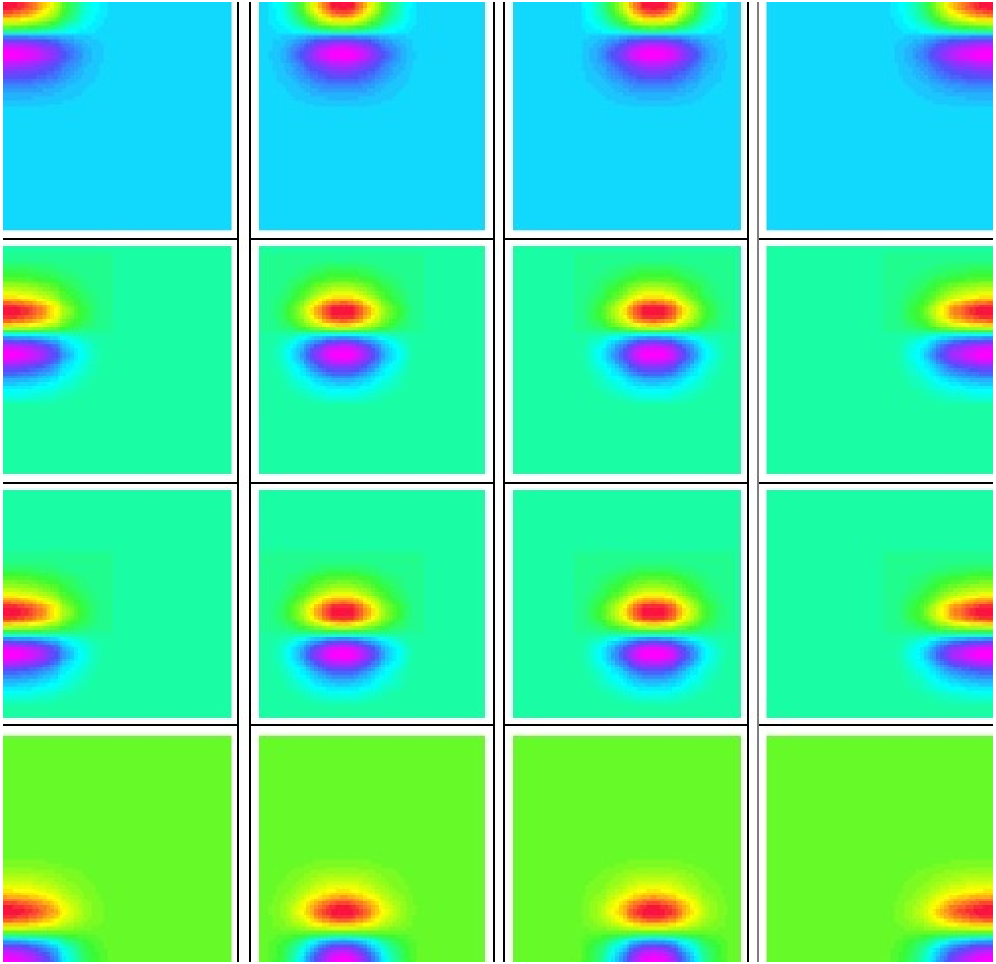
H matrix, Decimated H, Some covariance traces and the errors.

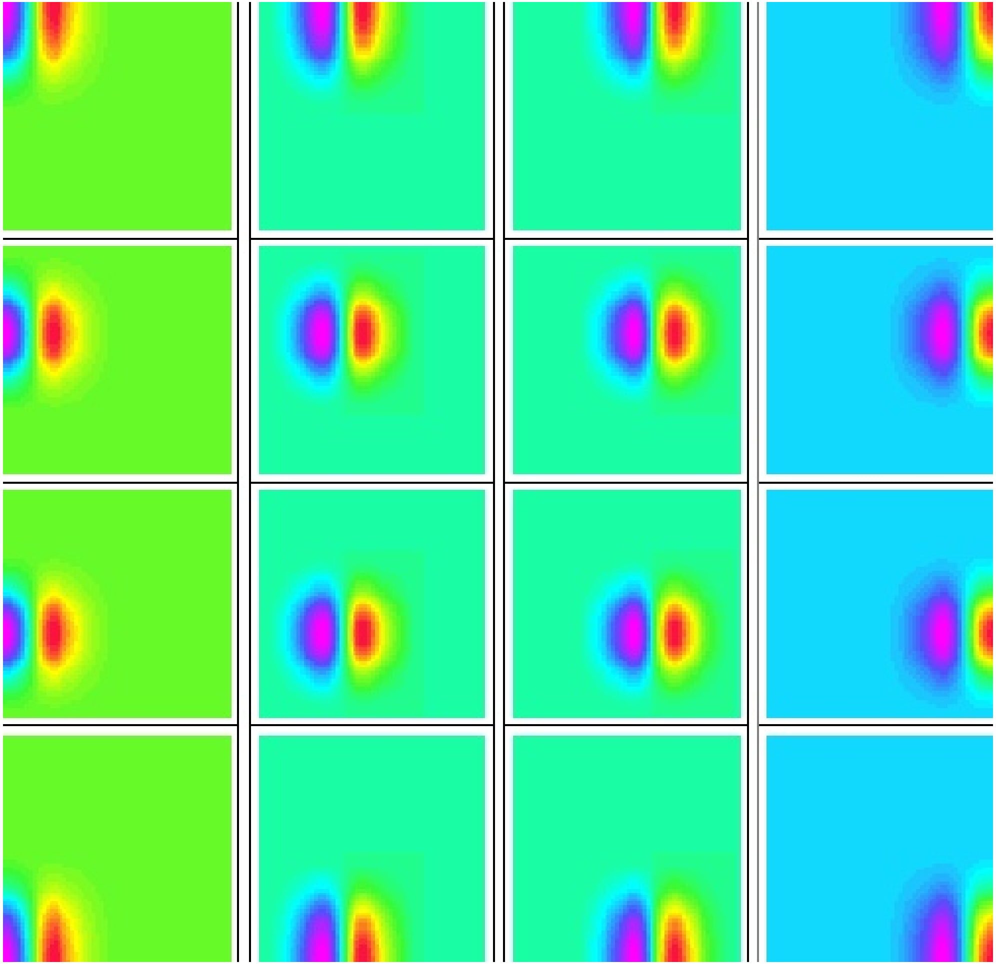


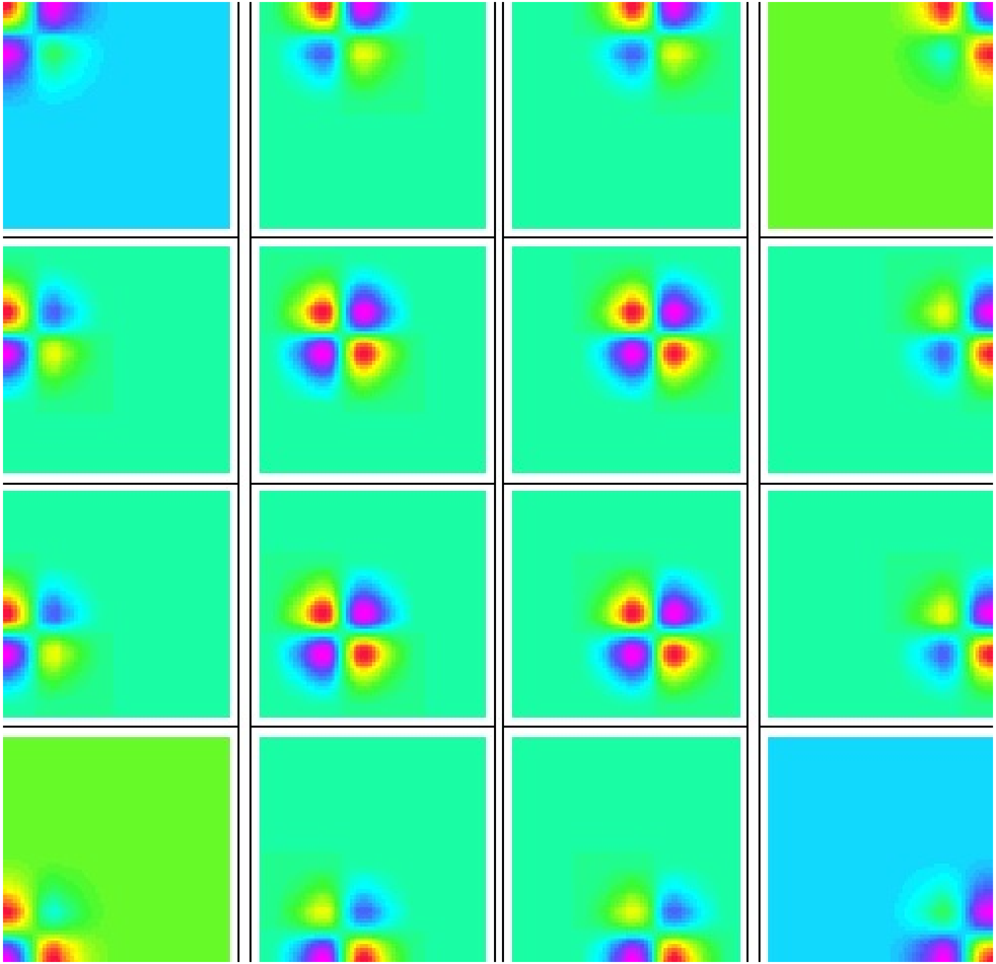
Two dimensions: tensor products of father and mothers

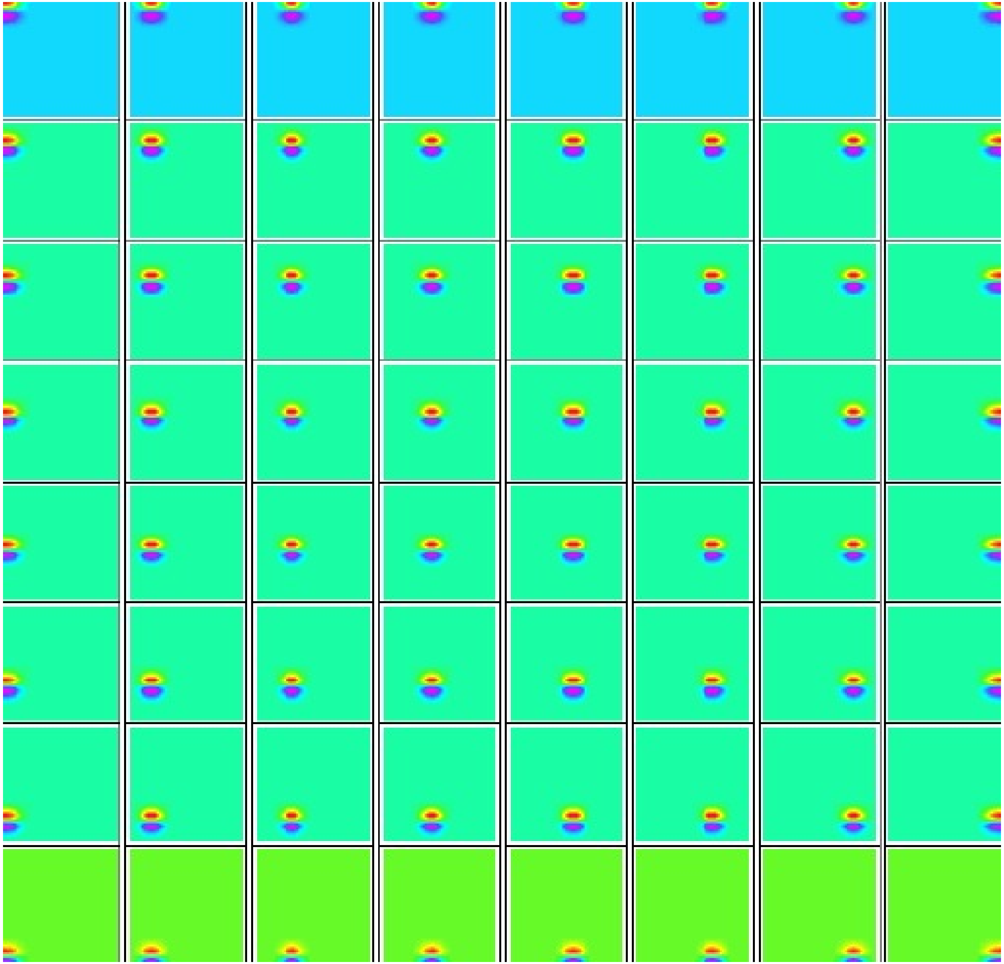












So what do we do when we have data?

Sample estimates of H

With gridded 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.

$\boldsymbol{Z}$ matrix with rows= space and columns = time

$$\widehat{\Sigma} = \boldsymbol{Z}\boldsymbol{Z}^T$$

or

$$\widehat{D} = (\Psi^{-1}\boldsymbol{Z})(\Psi^{-1}\boldsymbol{Z})^T$$

Decimation

The sparseness of H guarantees that we do not have to look (or compute) many off-diagonal elements.

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

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

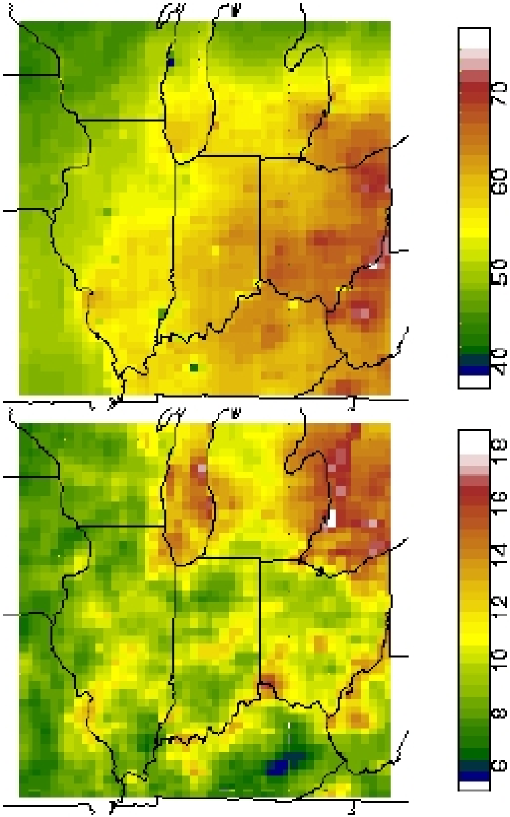
For the ROM example we just decimate the leading block of H (12×12) by 90% and retain diagonal elements for the rest.

Regional Oxidant Model (ROM)

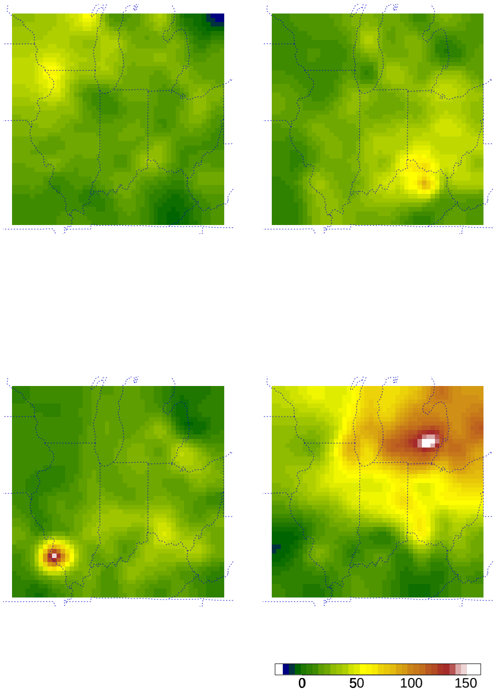
Atmospheric chemistry model that determines ozone formation and transport based on sources of pollution and meteorological conditions.

Model output:
8 hour daily average ozone
48×48 grid centered on Illinois and Ohio,
grid box size: 25km
79 days in June-August 1987, this was a period of high summer ozone

Mean and sd fields for ROM output 8 hour average ozone



Estimated covariance/correlation function at four locations (with 90 % decimation of H)



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.
- The most important future work is to extend the covariance estimate to irregularly sampled data.