Multi-resolution Based Nonstationary Covariance Modeling: Monte-Carlo EM approach

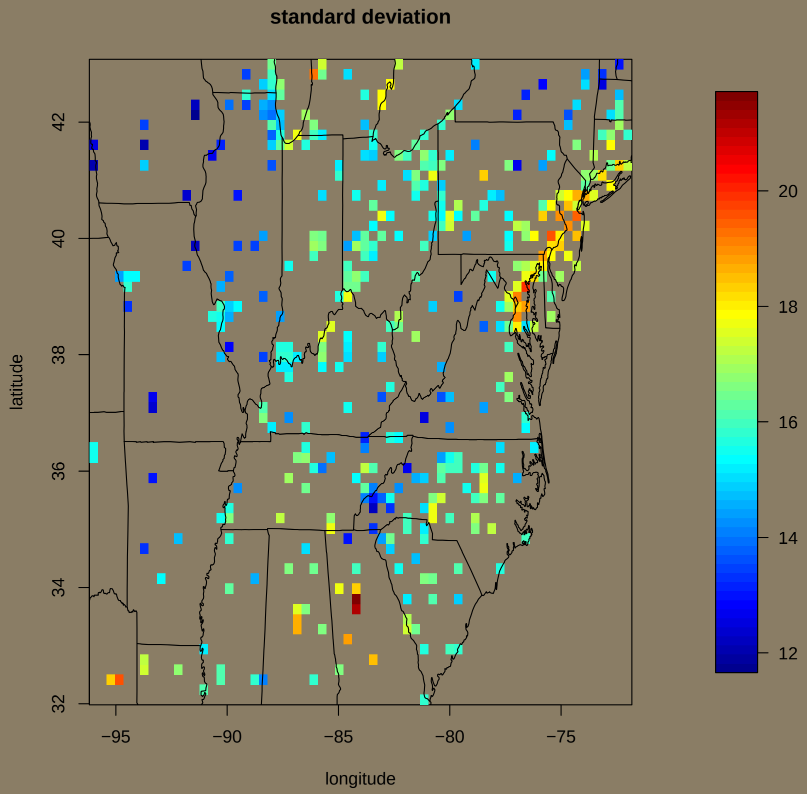
- Sparse Wavelet Estimator
- EM for irregular data
- Ozone example

IMAGe
Institute for Mathematics Applied to Geosciences

Supported by the National Science Foundation *Random Matrix Workshop, May 2007*

Surface Ozone

standard deviation



- O_3 is one of six common pollutants
- EPA's national air quality standards (80 ppb)
- 1997 Data Set
  - 364 locations on 48×48 grid
  - 184 days from May to Oct

Motivation and Goal

- Motivation:

  Flexible Nonstationary Covariance Model

  Gaussian Model in Spatial Statistics

  - Kriging (geostatistics)

    [e.g., Higdon et al., 1999; Fuentes, 2001; Fuentes and Smith, 2001; Nychka et al., 2003; Sampson and Guttorp, 1992; Anderes and Stein, 2005]

  - Variational and OI methods (data assimilation) [e.g., Purser et al., 2005; Gaspari et al., 2006]

- Goal and Challenges: Computational efficiency

  - Irregularly distributed observational data
  - Large data set

Multi-resolution representaion

Gaussian stochastic process $f(x)$

$$f(x) = \sum_{\nu=1}^{N} \alpha_\nu w_\nu(x)$$

α_ν **are Gaussian, basis $\{w_\nu\}$ is known and fixed.**

By vectors and matrices

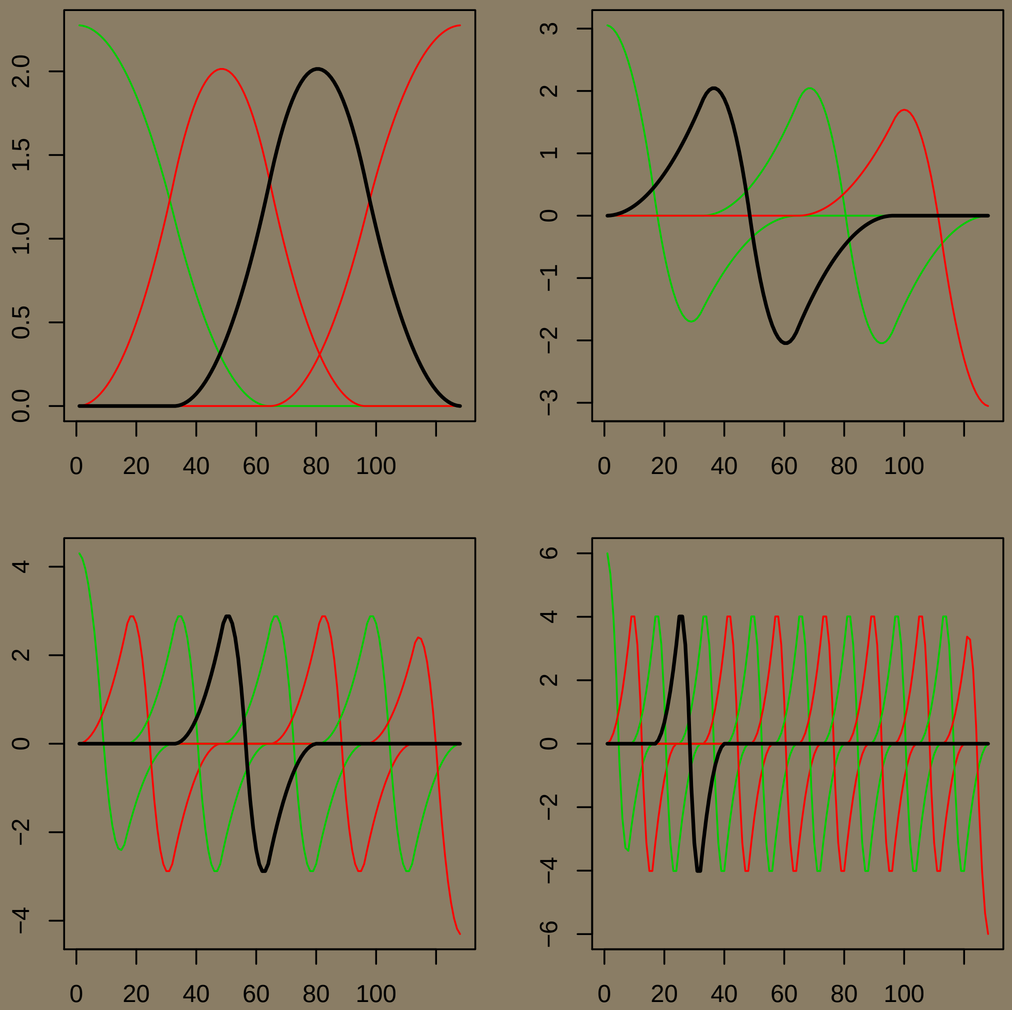
$$f = \mathcal{W}\alpha$$

Giving values of f on a fine grid.

A 1-d Wavelet basis

$$f(x) = \sum_{k=1}^{2^J} a_{J,k}\phi\left(2^J x - k\right) + \sum_{j=J}^{\infty}\sum_{k=1}^{2^j} b_{j,k}\psi\left(2^j x - k\right)$$

A 1-d local, but nonorthogonal basis



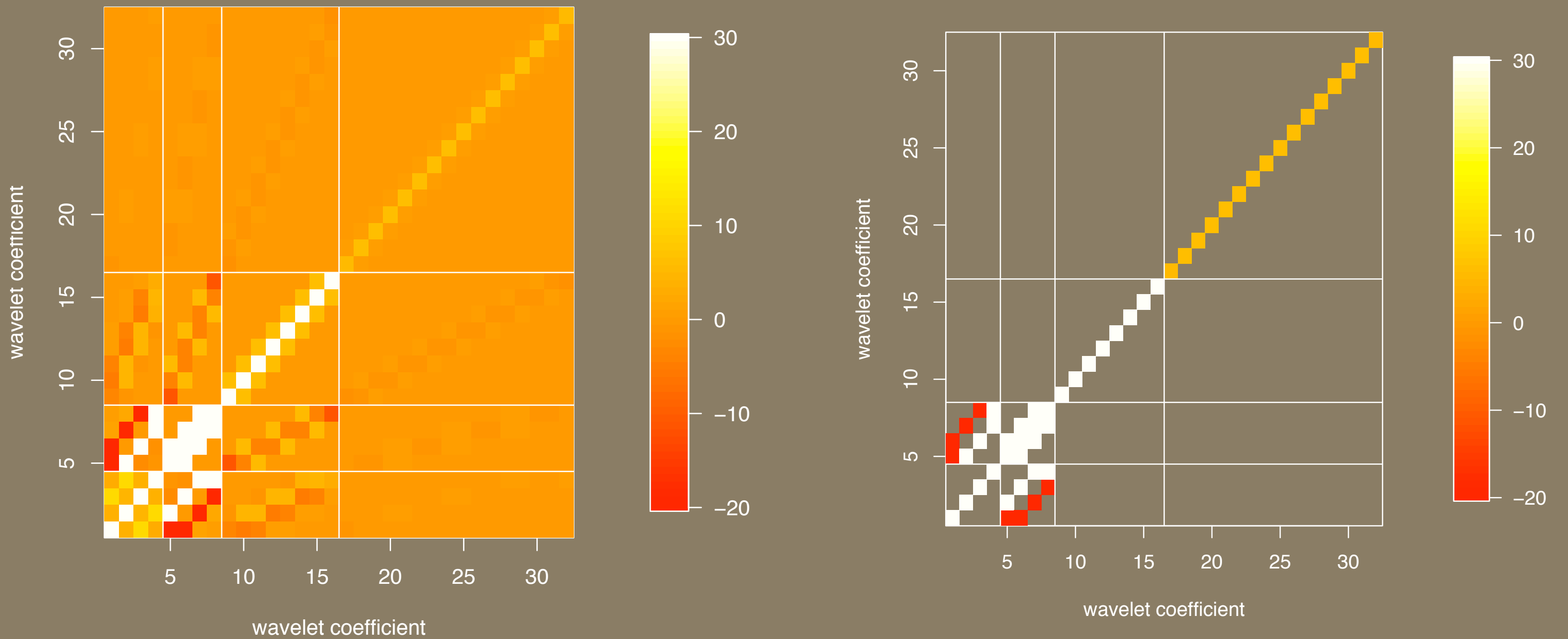
Nonparametric Model

$$\Sigma = \mathcal{W} H^2 \mathcal{W}^T$$

$$H = \begin{bmatrix} H_{00} & H_{01} \\ H_{10} & H_{11} \end{bmatrix} \approx \begin{bmatrix} \tilde{H}_{00} & 0 \\ 0 & \tilde{H}_1 \end{bmatrix}$$

where $\tilde{H}_{00}$ **is thresholded and** $\tilde{H}_1 = \textbf{diag}(H_{11})$

Enforced sparsity in H



Covariance Estimator

Gaussian Model

$$\mathrm{f} = \begin{pmatrix} \mathrm{f}_1 \\ \mathrm{f}_2 \end{pmatrix} \sim \mathcal{N}(0, \Sigma_\theta) \qquad \text{where} \quad \Sigma_\theta = \mathcal{W}\tilde{H}^2(\theta)\mathcal{W}^T.$$

Monte-Carlo EM

$$\begin{aligned} Q(\theta, \theta^*) &= E[\mathcal{L}(\mathrm{f}, \theta) | \mathrm{f}_1, \theta^*] \\ &\approx \frac{1}{N} \sum_{n=1}^{N} \mathcal{L}\left(\begin{pmatrix} \mathrm{f}_1 \\ \mathrm{f}_2^{(n)} \end{pmatrix}, \theta \right) \end{aligned}$$

MC sampling:

$$\mathrm{f}_2^{(n)} \sim [\ \mathrm{f}_2 | \mathrm{f}_1, \mathcal{W}\tilde{H}^2(\theta^*)\mathcal{W}^T\ ]$$

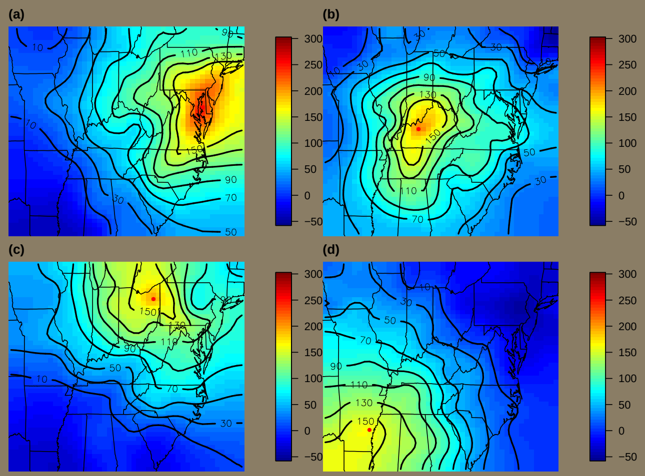
Smoothed MC EM

$$Q(\theta, \theta^*) + p(\theta)$$

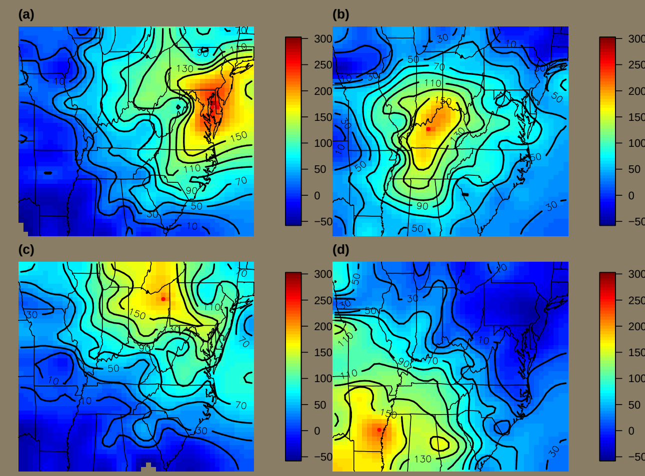
where p is a roughness penalty — or log prior.

Different thresholding in $\tilde{H}$

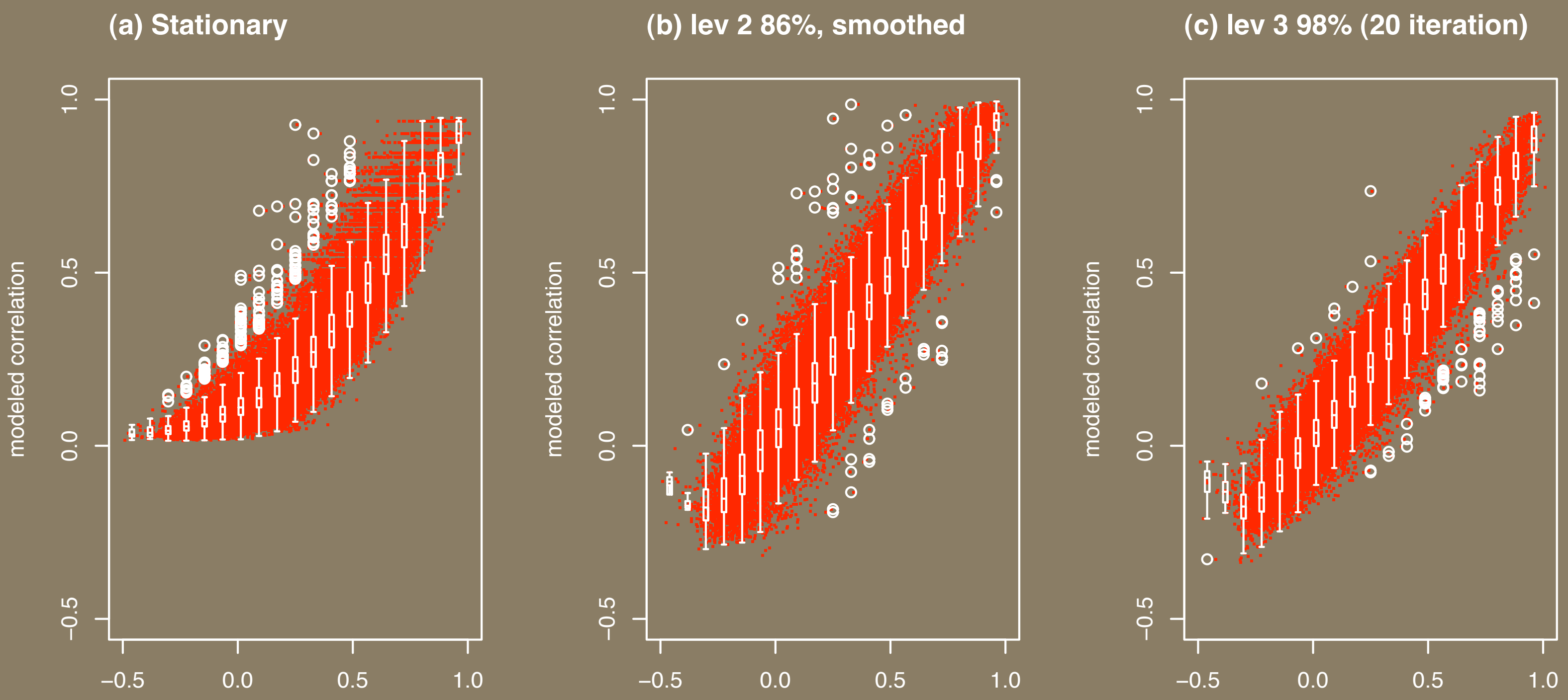
Lev 2, 86.08% (5045)



Lev 3, 98% (8363)



Cross-validation



Summary and Future Work

- Flexible nonstationary covariance model $\mathcal{W}\tilde{H}^2\mathcal{W}^T$
- Theory to support sparsity in $\tilde{H}$
- Practical estimator (Monte-Carlo EM) to handle the incomplete data
- Examples using surface ozone data
- Application to the Polar Ionosphere
  - Aurora image data ($\sim$ 100K)
  - Prior covariance for ionospheric data assimilation