REDUCED RANK SPATIAL COVARIANCE: A MULTIRESOLUTION APPROACH

Orietta Nicolis

Instituto de Estadistica, Universidad de Valparaiso, Chile

June 18, 2014

Universidad de Valparaíso
CHILE
Facultad de Ciencias
Departamento de Estadística

Valparaiso

Summary

1. Reduced Rank Covariance
   - Wavelet multiresolution approach
   - The reduced rank covariance
   - Applications: artificial data, AOT and ozone data
   - Conclusions

2. References

The problem of large data sets...

... we propose a wavelet based method for reducing rank of covariance matrices of non stationary and incomplete data sets.

Examples of large data sets:

- data spatially distributed on a regular grid with many missing data (satellite data: aerosols, radar, ndvi, ecc.);
- irregularly distributed observations (gauge data: ozone, rainfall, etc.)

Basis function representation

The general concept behind this work is the expansion of a random function in terms of basis functions

$$f = W\gamma$$

where W is a matrix of basis and γ is the vector of coefficients.

- **Fourier basis representation** (Panciorek, 2006)
  W is the matrix of orthogonal spectral basis functions, and $\gamma_k = a_k + b_k$, $k = 1, ...,$ is a vector of complex-valued coefficients;
- **Karhunen-Loève decomposition**
  W is a matrix of orthogonal basis and the coefficients γ are independent Gaussian random variables, $\gamma \sim N(0, \Lambda)$, where and $\Lambda = diag(\lambda_1, ..., \lambda_n)$.

- **Multiresolution representation** (Matsuo et al., 2008) In this work
  1. W is a multi-resolution or wavelet basis
  2. the coefficients may not be uncorrelated, $\gamma \sim N(0, D)$, where D can be not orthogonal
  3. Because of the **localized support** of wavelet basis functions, the expansion results in a small number of coefficients with significant correlations.

Outline

1. Reduced Rank Covariance
   - Wavelet multiresolution approach
   - The reduced rank covariance
   - Applications: artificial data, AOT and ozone data
   - Conclusions
2. References

Spatial model

Let **y** be the field values on a large (regular) $2-$D $(N \times N)$-grid (stacked as a vector) with covariance function

$$\Sigma = COV(\mathbf{y}).$$

- Eigen decomposition of the covariance

$$\Sigma = WDW^T = WHH^T W^T$$

where $D = H^2$ and $H = (W^{-1}\Sigma W^T)^{1/2}$.

- Representation of the process

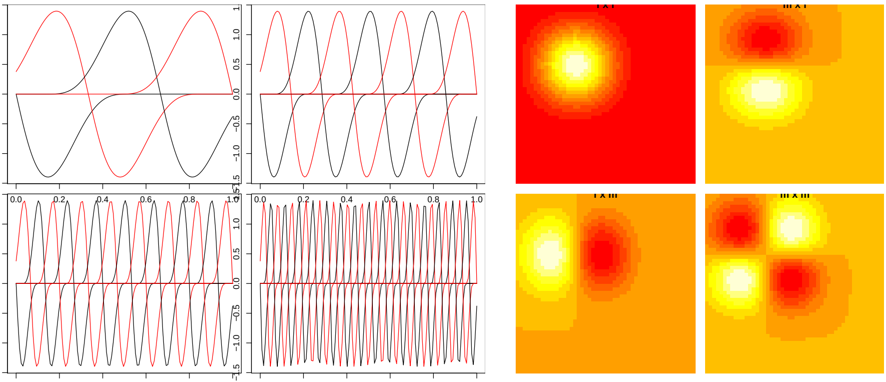
$$\mathbf{y} = WH\gamma$$

where γ is a vector of independent standard normal variables

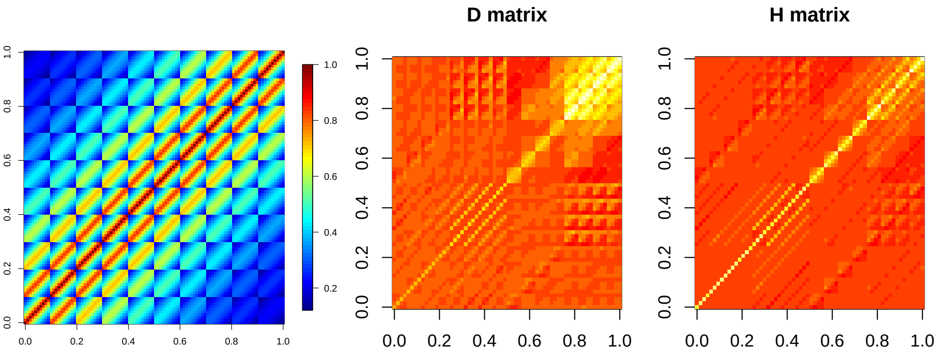
***W* –> need not be orthogonal!**

***D*–> need not be diagonal!**

The matrix W (Kwong and Tang, 1994)



D and H



Observational data

Suppose that the observations $\mathbf{y}$ are samples of a centered Gaussian random field and are composed of two components: the observations at irregularly distributed locations, $\mathbf{y}_o$, and the missing observations, $\mathbf{y}_m$,

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_o \\ \mathbf{y}_m \end{pmatrix} \tag{1}$$

The observational model can be written as

$$\mathbf{y}_o = K\mathbf{y} + \varepsilon$$

where

- $\mathbf{y}_o$ observations
- $\mathbf{y}$ values on a the grid
- K is a incidence matrix
- $\varepsilon \sim MN(0, \sigma^2 I)$

The conditional distribution of y_m given y_o is a multivariate normal with mean

$$\Sigma_{o,m}(\Sigma_{o,o})^{-1}\mathbf{y}_o \tag{2}$$

and variance

$$\Sigma_{m,m} - \Sigma_{m,o}(\Sigma_{o,o})^{-1}\Sigma_{o,m} \tag{3}$$

where

- $\Sigma_{o,m} = W_o H H^T W_m^T$ is the cross-covariance between observed and missing data,
- $\Sigma_{o,o} = W_o H H^T W_o^T + \sigma^2 I$ is covariance of observed data and
- $\Sigma_{m,m} = W_m H H^T W_m^T$ is the covariance of missing data.
- The matrices W_o and W_m are wavelet basis evaluated at the observed and missing data, respectively.

Problem

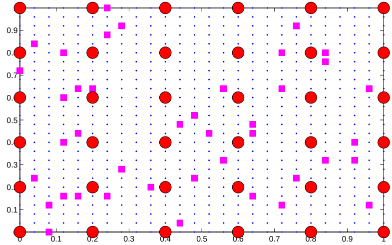
$\Sigma_{o,m}$ and $\Sigma_{o,o}$ very big!!!

Outline

1. Reduced Rank Covariance
   - Wavelet multiresolution approach
   - The reduced rank covariance
   - Applications: artificial data, AOT and ozone data
   - Conclusions
2. References

The reduced rank covariance for Σ (Nicolis and Nychka, 2012).

- The idea is to estimate H on a small sub-grid $\mathcal{G}$ of size $(g \times g)$ starting from a Matérn model and using a MR approach.
- Monte Carlo conditional simulation provide an efficient estimator for the conditional mean and variance.



Conditional simulation

1. Find Kriging prediction on the grid $\mathcal{G}$:

$$\hat{\mathbf{y}}_g = \Sigma_{o,g}(\Sigma_{o,o})^{-1}\mathbf{y}_o,$$

where $\tilde{H}_g = (W_g^{-1}\Sigma_{g,g}W_g^T)^{1/2}$ and $\Sigma_{g,g}$ is stationary covariance model (es. Matern).

2. Generate synthetic "data": $\mathbf{y}^s$ from $\mathbf{y}^s = W\tilde{H}_g a$ with $a \sim N(0,1)$.
3. Simulated Kriging error:

$$\mathbf{u}^* = \mathbf{y}_g - \mathbf{y}_g^s,$$

where $\mathbf{y}_g = W_g\tilde{H}_g a$ and $\mathbf{y}_g^s = \Sigma_{o,g}(\Sigma_{o,o})^{-1}\mathbf{y}_o^s$.

4. Find conditional field $\mathbf{y}_m|\mathbf{y}_o$:

$$\hat{\mathbf{y}}_u = \hat{\mathbf{y}}_g + \mathbf{u}^*.$$

5. Compute the conditional covariance on T replications, $\Sigma_u = COV(\hat{\mathbf{y}}_u)$, using the new $\tilde{H}_g$ in the step 1 of each iteration.

By choosing properly the filter W basis and the levels of resolutions L we obtain the estimation of the conditional mean and variance

$$\Sigma_{o,m}(\Sigma_{o,o})^{-1}\mathbf{y}_o \tag{4}$$

and variance

$$\Sigma_{m,m} - \Sigma_{m,o}(\Sigma_{o,o})^{-1}\Sigma_{o,m} \tag{5}$$

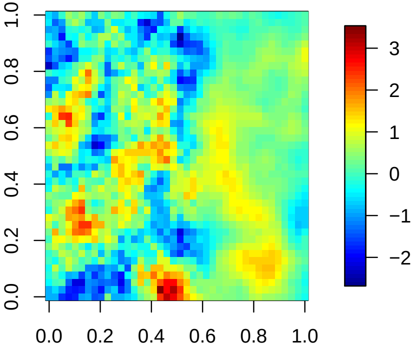
where

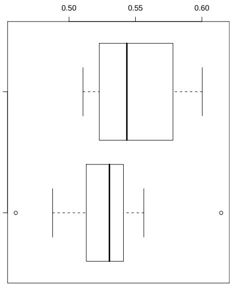
- $\Sigma_{o,m} = W_o H_g H_g^T W_m^T$, $\Sigma_{o,o} = W_o H_g H_g^T W_o^T + \sigma^2 I$ and $\Sigma_{m,m} = W_m H_g H_g^T W_m^T$

Outline

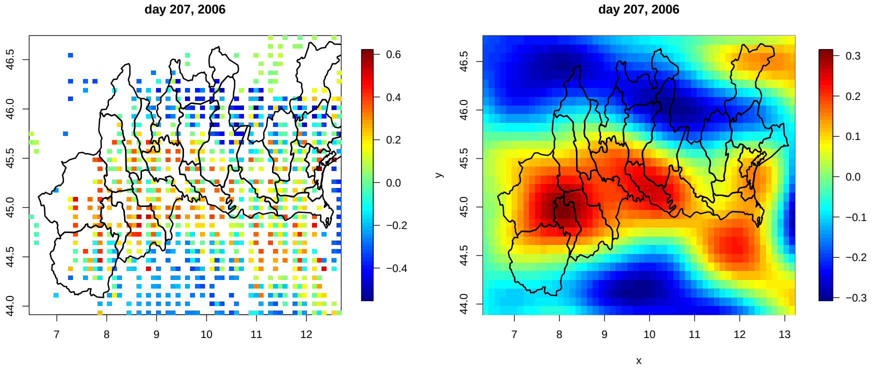
1. Reduced Rank Covariance
   - Wavelet multiresolution approach
   - The reduced rank covariance
   - Applications: artificial data, AOT and ozone data
   - Conclusions
2. References

Non-stationary random field simulated on a 40×40 grid with Matèrn covariance ($\theta = 0.1$ and $\nu = 0.5$)) and 50% of missing values.

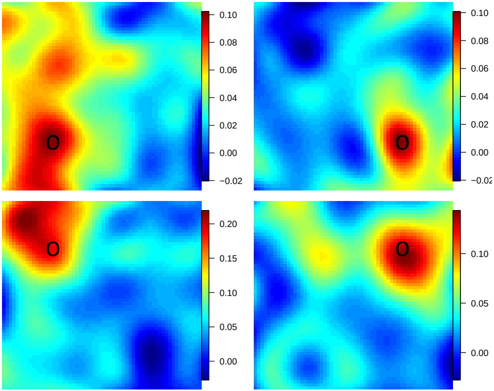




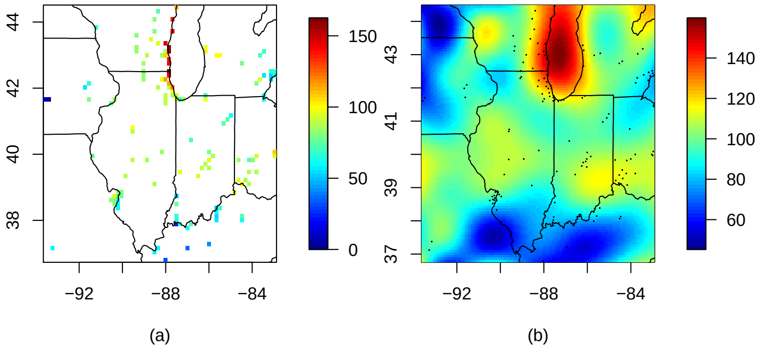
Application to AOT (54×32)



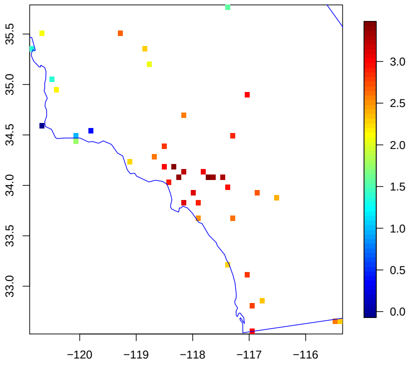
Non-stationary covariance obtained after 5 iterations of MC simulations.



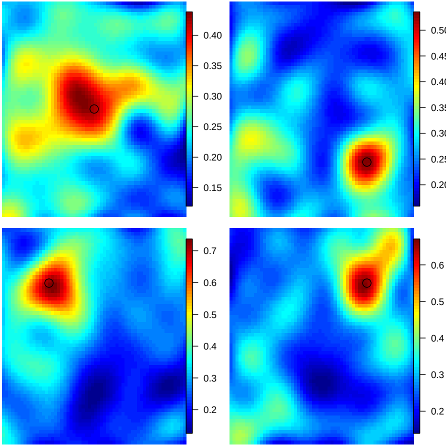
Daily max 8 hour ozone, June 18, 1987



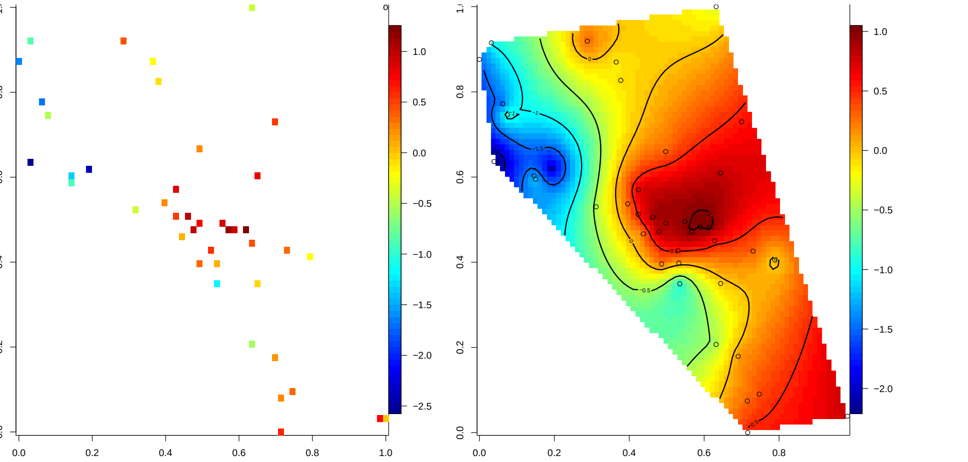
Daily NO2 in California



Covariance NO2 in California



Forecasting NO2 in California



Outline

1. Reduced Rank Covariance
   - Wavelet multiresolution approach
   - The reduced rank covariance
   - Applications: artificial data, AOT and ozone data
   - Conclusions
2. References

Further work

- Find a parametrization for the matrix H_g, depending on the parameters of the Matern covariance, $H_g(\nu, \theta)$, and find the maximum likelihood estimates.
- Consider other basis functions (frames, radial basis etc.).
- Include the multiresolution covariances in spatio-temporal models.
- Extension to multivariate case (for example calibration of aerosol data)

References

- Gneiting, T. (2002) Compactly Supported Correlation Functions. J. of Multivariate Analysis, 83, 493–508
- Matsuo, T., D Nychka and D. Paul (2008) Nonstationary Covariance Modeling for Incomplete Data: Monte Carlo EM approach. In review.
- Nychka, D., Wikle, C. and Royle, K. A. (2003). Multiresolution models for nonstationary spatial covariance functions. Stat. Model., 2, 315–332.
- Nicolis, O., Nychka. D. (2012). Reduced Rank Covariances for the Analysis of Environmental Data. In *Advanced Statistical Methods for the Analysis of Large Data-Sets (Di Ciacco, Coli, and Angulo Ibanez, eds.), Series: Studies in Theoretical and Apllied Statistics, Springer, ISBN 978-3-642-21036-5.*