Basis Function Models for Nonstationary Spatial Modeling

Yanan Jia, Mark Risser, Abhijoy Saha

with Peter Craigmile, Tom Santner

October 14, 2014

GENERAL OVERVIEW

- Assumption of stationarity implies the statistical association between 2 points is only a function of the distance between them.
- For nonstationary modeling, we use basis functions to estimate the spatial covariance structure $\boldsymbol{C}(\boldsymbol{s_1}, \boldsymbol{s_2})$.
- The earliest modeling strategy was based on decomposing the spatial process in terms of empirical orthogonal functions (EOFs) (discussed later)
- Holland et. al. (1998) represented the spatial covariance function as the sum of stationary isotropic spatial covariance model and a finite decomposition in terms of EOFs.
- Nychka et. al. (2002) used wavelet basis function decomposition with computational focus on large problems.
- Pintore and Holmes (2004) worked with both Karhunen – Loéve and Fourier expansions.

REPRESENTING FUNCTIONS BY BASIS FUNCTIONS

- A basis function system is a set of known functions $\boldsymbol{\phi_k}$ that are mathematically independent of each other.

- We can approximate arbitrarily well any function by taking a weighted sum or linear combination of a sufficiently large number $\boldsymbol{K}$ of these functions.

- Some familiar basis function systems include the classic orthogonal polynomial sequence (Hermite polynomials)

$$\mathbf{1},\ \boldsymbol{t}^{\mathbf{1}},\ \boldsymbol{t}^{\mathbf{2}}-\mathbf{1},\ \boldsymbol{t}^{\mathbf{3}}-\mathbf{3}\boldsymbol{t},\ \boldsymbol{t}^{\mathbf{4}}-\mathbf{6}\boldsymbol{t}^{\mathbf{2}}+\mathbf{3}, \ldots$$

that is used to construct power series, or the Fourier series system

$$\mathbf{1}, \mathbf{sin}(\boldsymbol{\omega t}), \mathbf{cos}(\boldsymbol{\omega t}), \mathbf{sin}(\mathbf{2}\boldsymbol{\omega t}), \mathbf{cos}(\mathbf{2}\boldsymbol{\omega t}), \ldots, \mathbf{sin}(\boldsymbol{k\omega t}), \mathbf{cos}(\boldsymbol{k\omega t}), \ldots$$

REPRESENTING FUNCTIONS BY BASIS FUNCTIONS

- Basis function procedures represent a function $\boldsymbol{x}$ which depends on $\boldsymbol{t}$ by a linear expansion

$$\boldsymbol{x}(\boldsymbol{t}) = \sum_{k=1}^{K} \lambda_k \phi_k(\boldsymbol{t})$$

in terms of $\boldsymbol{K}$ known basis functions.

- Let

$\boldsymbol{\Lambda}$ denote the vector of length $\boldsymbol{K}$ of the coefficients $\boldsymbol{\lambda_k}$ and
$\boldsymbol{\Phi}$ denote the functional vector whose elements are basis functions $\boldsymbol{\phi_k}$.

We can express $\boldsymbol{x}(\boldsymbol{t})$ in matrix notation as

$$\boldsymbol{x} = \boldsymbol{\Lambda}'\boldsymbol{\Phi} = \boldsymbol{\Phi}'\boldsymbol{\Lambda}$$

- Expressing the covariance function in terms of linear basis functions, allows us to perform calculations using well known methods of matrix algebra.

REPRESENTING FUNCTIONS BY BASIS FUNCTIONS

- Thus basis expansion methods represent infinite dimensional functions within the finite dimensional framework of vectors.
- Choice of the basis system $\boldsymbol{\Phi}$ is also important.
- The degree to which the data is smoothed as opposed to interpolated is determined by the number of basis functions $\boldsymbol{K}$.
- Accordingly, the basis system is not defined by a fixed number $\boldsymbol{K}$ of parameters, but $\boldsymbol{K}$ itself is viewed as a parameter which is chosen according to the characteristics of the data.
- Basis functions allow us to store information about functions, and gives us the flexibility that is needed for efficient computational algorithms for analyzing large spatial data sets.

REPRESENTING FUNCTIONS BY BASIS FUNCTIONS

- The fundamental principle behind decomposing the spatial covariance function for a nonstationary process is to find a suitable parametric setup which facilitates us to recover the initial empirical covariance matrix obtained from the observations or some smooth edition of it.

- This enables us to evaluate the covariance function between any two points outside of the measurement set.

- The Karhunen – Loéve expansion of a covariance function is a spectral decomposition using orthogonal basis functions, namely the eigen vectors.

- We can also use non-orthogonal basis function, i.e. wavelet basis in place of the eigen functions and relax the condition to allow some correlation between the previously assumed independent coefficients of the actual process.

Empirical orthogonal function

A GENERAL MODELING FRAMEWORK

- Let $Z(\cdot)$ be a realization of a spatial stochastic process defined for all $\mathbf{s} \in \mathcal{D} \subset \mathbb{R}^{\mathbf{d}}$, where d is typically equal to 2 or 3
- For all $\mathbf{s} \in \mathcal{D}$, let

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + Y(\mathbf{s}) + \epsilon(\mathbf{s})$$

  - $Y(\cdot)$ is a mean-zero latent spatial (Gaussian) process.
  - $Y(\cdot)$ is a nonstationary spatial process with covariance function $C(\mathbf{s}_1, \mathbf{s}_2) = cov(Y(\mathbf{s}_1), Y(\mathbf{s}_2))$

Modeling $C(\mathbf{s}_1, \mathbf{s}_2)$ using Empirical orthogonal function (EOF)

Empirical orthogonal function

Idea: decompose the spatial covariance function in terms of basis functions

$$\mathbf{Y} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{\Sigma})$$

- Eigendecomposition
$$\mathbf{\Sigma} = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^\mathsf{T},$$
where $\mathbf{\Phi}$ is the square $(n \times n)$ matrix whose ith column is the eigenvector ϕ_k of $\mathbf{\Sigma}$ and $\mathbf{\Lambda}$ is the diagonal matrix whose diagonal elements are the corresponding eigenvalues λ_i.
- In place of $\mathbf{Y}$
$$\mathbf{Y} = \mathbf{\Phi}\mathbf{\Lambda}^{1/2}\boldsymbol{\alpha} = \sum_{i=1}^{n} \sqrt{\lambda_i}\alpha_i\phi_i,$$
where the α_i are iid $\mathcal{N}(0, 1)$.

Empirical orthogonal function

- For a process with a given covariance function there exists a unique orthogonal expansion of the process.
- **Karhunen-Loéve expansion**

$$Y(\mathbf{s}) = \sum_{k=1}^{\infty} \sqrt{\lambda_k}\alpha_k \phi_k(\mathbf{s})$$

where, for example, the $\{\alpha_k\}$ are iid $\mathcal{N}(0,1)$.

- ϕ_k are the eigenfunctions and λ_k are the eigenvalues of the Fredholm integral equation of the second kind:

$$\int_{\mathcal{D}} C(\mathbf{s}_i, \mathbf{s}_j)\phi_k(\mathbf{s}_i) d\mathbf{s}_i = \lambda_k \phi_k(\mathbf{s}_j).$$

Empirical orthogonal function

Karhunen-Loéve expansion

$$Y(\mathbf{s}) = \sum_{k=1}^{\infty} \sqrt{\lambda_k}\alpha_k\phi_k(\mathbf{s})$$

where the $\{\alpha_k\}$ are iid $\mathcal{N}(0,1)$.

Covariance function:

$$\begin{aligned}
C_Y(\mathbf{s}_i, \mathbf{s}_j) &= cov(Y(\mathbf{s}_i), Y(\mathbf{s}_j)) \\
&= \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} \sqrt{\lambda_k}\phi_k(\mathbf{s}_i)\sqrt{\lambda_l}\phi_l(\mathbf{s}_j)cov(\alpha_k, \alpha_l) \\
&= \sum_{k=1}^{\infty} \lambda_k\phi_k(\mathbf{s}_i)\phi_k(\mathbf{s}_j)
\end{aligned}$$

Empirical orthogonal function

In the discrete case, with a finite set of data points (n).

- Karhunen-Loéve expansion is equivalent to a principal component analysis
- Fredholm equation is the analog of the matrix eigenvector equation.

$$\mathbf{S}\boldsymbol{\Phi} = \boldsymbol{\Phi}\boldsymbol{\Lambda}$$

where $\boldsymbol{\Phi} = (\phi_k(\mathbf{s}_i))_{i,k=1,\dots,n}$ is the matrix of eigenvectors, and $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues λ_i.

$$Y(\mathbf{s}_i) = \sum_{k=1}^{n} \sqrt{\lambda_k}\alpha_k\phi_k(\mathbf{s}_i)$$

$$C_Y(\mathbf{s}_i, \mathbf{s}_j) = \sum_{k=1}^{n} \lambda_k\phi_k(\mathbf{s}_i)\phi_k(\mathbf{s}_j)$$

Empirical orthogonal function

In the discrete case, with a finite set of data points (n).

- The eigenvalues λ_k are a **non-parametric** function of k.
- To allow for the spectrum (eigenvalues) to evolve over space, we tempering it by taking it to some power $\eta(\mathbf{s})$ at each location $\mathbf{s}$, where $\eta(\mathbf{s})$ is some smooth function of $\mathbf{s}$.
- By heating or cooling the spectrum at each location, one is able to control the amount of smoothness induced by the model.

$$Y(\mathbf{s}) = \sum_{k=1}^{n} {\lambda_k}^{\eta(\mathbf{s})/2} \alpha_k \phi_k(\mathbf{s})$$

$$C_Y(\mathbf{s}_i, \mathbf{s}_j) = \sum_{k=1}^{n} {\lambda_k}^{\eta(\mathbf{s}_i)/2} {\lambda_k}^{\eta(\mathbf{s}_j)/2} \phi_k(\mathbf{s}_i) \phi_k(\mathbf{s}_j)$$

Empirical orthogonal function

Prediction

- Evaluate the covariance function between any two points in the space.
- Extend the eigenvectors $\{\phi_i = (\phi_i(\mathbf{s}_1), \ldots, \phi_i(\mathbf{s}_n))\}_{i=1,\ldots,n}$ to eigenfunctions $\{\tilde{\phi}(\mathbf{s})\}_i, i = 1, ..., n$ defined for all $\mathbf{s} \in \mathcal{D}$.

The integration formulae methodology:

$$\tilde{\phi}_i(s) = 1/\lambda_i \sum_{j=1}^{n} C(s, s_j)\phi_i(s_j)$$

where $C(\cdot, \cdot)$ is the stationary matrix.

Compute the covariance between any two points using $\tilde{\phi}_i(\cdot)$, $i = 1, ..., n$.

Empirical orthogonal function

Dimension reduction

- Select $m << n$ points from the initial data set and fit a non-stationary model to these m points using only $p < m$ eigenvectors.
  - Obtain m locations $c_i, i = 1, ..., m$ in the field using k-means clustering on the full data set.
  - For each cluster i, choose the data point which is closest to c_i.
- Approximating the full stationary covariance matrix to extend the p eigenvectors.

Advantages of using EOFs:

- naturally nonstationary

Disadvantages of using EOFs:

- prediction
- measurement error

Non-orthogonal basis functions

Recall: a Gaussian process can be written in terms of eigenvalue/eigenfunction decomposition of the covariance function C as

$$Y(\mathbf{s}) = \sum_{k=1}^{\infty} \sqrt{\lambda_k}\alpha_k \phi_k(\mathbf{s})$$

where, for example, the $\{\alpha_k\}$ are iid $\mathcal{N}(0, 1)$.

New idea (Nychka et al., 2002):

1. Use (non-orthogonal) basis functions in place of the eigenfunctions
2. Allow some correlation among the $\{\alpha_k\}$

Discretized model (Nychka et al., 2002)

Let $\mathbf{Y} \in \mathcal{R}^m$ by the values of the spatial field on a fine, rectangular grid, such that

$$Cov(\mathbf{Y}) = \mathbf{\Sigma} = \mathbf{\Psi D \Psi}' \quad \longleftrightarrow \quad \mathbf{Y} = \mathbf{\Psi H a}$$

where $\mathbf{H} = \mathbf{D}^{1/2}$.

- The columns of $\mathbf{\Psi}$ are individual basis functions evaluated on the grid but stacked as a single vector.

Eigen-decomposition hints at an alternative way of building the covariance: specify basis functions (which define $\mathbf{\Psi}$) and $\mathbf{H}$.

Details:

- If $\mathbf{\Psi}$ is not orthogonal and $\mathbf{D}$ is not diagonal, this is a problem: $\mathbf{\Sigma}$ is $m \times m$; $m^2 \approx 1,000,000$
- Solution: Choose a multi resolution basis for $\mathbf{\Psi}$ to allow for fast recursive algorithms in calculating $\mathbf{\Psi}^{-1}$; restrict $\mathbf{H}$ to be sparse

Multiresolution bases (Nychka et al., 2002)

Generate a basis for expanding the covariance using repeated translations of a few fixed functions. Rewrite

$$Y(\mathbf{s}) = \sum_i \sum_j \sqrt{\lambda_{ij}} \alpha_{ij} \phi_{ij}(\mathbf{s})$$

where i refers to a scaling and j a shift (grid locations). For example, in one dimension, for **fixed** constants $\{a_i\}$ and $\{b_j\}$:

$$\phi_{ij}(s) = \frac{1}{\sqrt{a_i}} \phi\left(\frac{s - b_j}{a_i}\right)$$

- The "parent" function $\phi(\cdot)$ could be any wavelet function with local support.
- These basis functions lend well to nonstationary fields because stochastic properties can be controlled locally.

Multiresolution bases

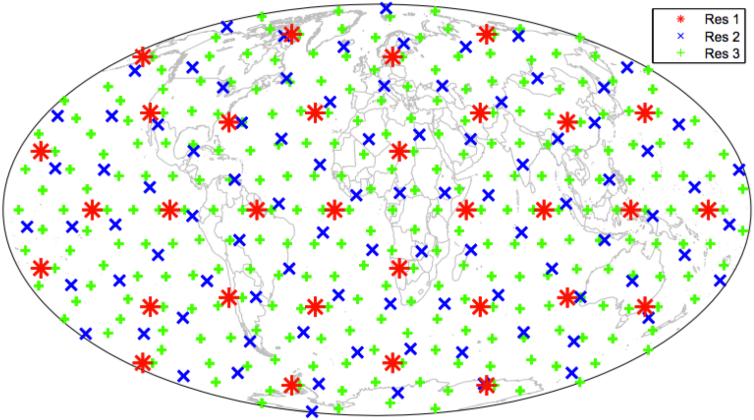
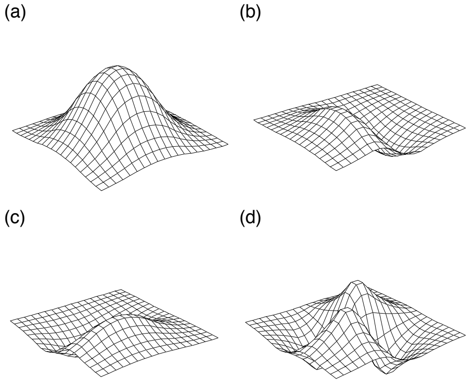


Figure 5: The locations of the centers of the 396 basis functions of the three resolutions.

Multiresolution bases (Nychka et al., 2002)

In two dimensions: use four "parent" functions:



Estimating $\boldsymbol{\Sigma}$

Assume we have independent replicates over $t = 1, \ldots, T$ time points; collect into a $m \times T$ matrix

$$\mathbf{Y} = [\, \mathbf{Y}_1 \, \mathbf{Y}_2 \, \cdots \, \mathbf{Y}_T \,]$$

such that each $\mathbf{Y}_t$ has covariance $\boldsymbol{\Sigma}$. Then

$$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Psi}\widehat{\mathbf{D}}\boldsymbol{\Psi}' = (1/T)\mathbf{Y}\mathbf{Y}' \quad \longleftrightarrow \quad \widehat{\mathbf{D}} = (1/T)(\boldsymbol{\Psi}^{-1}\mathbf{Y})(\boldsymbol{\Psi}^{-1}\mathbf{Y})'$$

- $\widehat{\mathbf{D}}$ is the sample covariance of the basis function coefficients

Estimating $\boldsymbol{\Sigma}$

Recall: we want sparse $\mathbf{D}$ and $\mathbf{H} = \mathbf{D}^{1/2}$

Solve for $\widehat{\mathbf{H}}$ through a singular value decomposition of $\boldsymbol{\Psi}^{-1}\mathbf{Y}$. Let

$$\mathbf{V}\boldsymbol{\Lambda}\mathbf{U}' = (1/\sqrt{T})(\boldsymbol{\Psi}^{-1}\mathbf{Y}) \quad \longleftrightarrow \quad \widehat{\mathbf{H}} = \mathbf{V}\boldsymbol{\Lambda}^{1/2}\mathbf{V}$$

Enforce sparsity in $\widehat{\mathbf{H}}$: for all $k \neq l$,

$$\widehat{H}_{kl} = \begin{cases} \widehat{H}_{kl}, & \text{if } |\widehat{H}_{kl}| > \delta \\ 0 & \text{otherwise.} \end{cases}$$

This gives a good approximation even if $\widehat{\mathbf{H}}$ has only 5-10% nonzero elements.

Summary

Pros: fast way to estimate a nonstationary covariance structure

Cons: problems when

1. data is not observed on a complete grid
2. data is located at irregularly spaced locations

Replicates are required
What about prediction?

Bibliography I

Nychka, D., Wikle, C., and Royle, J. A. (2002). Multiresolution models for nonstationary spatial covariance functions. *Statistical Modelling*, 2(4):315–331.

Pintore, A. and Holmes, C. (2004). A dimension-reduction approach for spectral tempering using empirical orthogonal functions. In Leuangthong, O. and Deutsch, C., editors, *Geostatistics Banff 2004*, volume 14 of *Quantitative Geology and Geostatistics*, pages 1007–1015. Springer Netherlands.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer.