Chapter 5 - Exploratory methods for spatio-temporal data

Odd Kolbjørnsen and Geir Storvik

March 20. 2017

Last time: Point processes

- Point process (random locations, random count, random marks)
- Density of a point process (given the count)
- Poisson process
  - Intensity
  - Homogeneous/ In homogeneous
- Other processes
  - Random intensity (Log Gaussian Cox Process)
  - Clustering (Parent-child,e.g. Neyman-Scott)
  - Repulsion (Markov point process, Strauss - Hard core process)
- Inference
  - Kernel estimate of intensity
  - L-function (K-function)
  - Edge effects

Spatial part extra by Petter Abrahamsen today

Conditional simulation of Gaussian random fields.

- When is this useful?
- Conditioning to linear events using Kriging.
- Conditioning to nonlinear observations.

Marked Point processes

- Extensions of the Boolean model from the book.
- Complex marks

Spatio-temporal data

Types

- Observation networks
- Remote-sensing platforms
- Data-assimilation algorithms (output from computer models)

Characteristics

- Large data set
- Often repeated measurements in same spatial locations at new times
- Dependencies across a large range of spatial and temporal scales

Key challenge: Extract information
Important tool: Exploratory analysis

Visualization

Plotting data/ presenting results is an important part of job as a statistician. It helps you understand the data, and is essential in communication with other research disciplines.

Script Visualization `SST_VIS.R` on webpage

- Spatial maps over time
- Animations
- Space (1-D)/Time-plot
- Time-series plots of particular features
- Empirical covariance/correlation functions

Empirical orthogonal function (EOF) motivation

Recall K-L expansion (Karhunen-Loève)

$$Z(\mathbf{s}) = \sum_{i=1}^{\infty} z_i \, \psi_i(\mathbf{s}) \;\; , \;\; z_i \sim \text{ independent, } N(0, \lambda_i^2)$$

$$C_Z(\mathbf{s}, \mathbf{r}) = \sum_{i=1}^{\infty} \lambda_i^2 \, \psi_i(\mathbf{s}) \psi_i(\mathbf{r})$$

Add time dependency to the coefficients:(notation different from book)

$$Z(\mathbf{s}, t) = \sum_{i=1}^{\infty} z_i(t) \, \psi_i(\mathbf{s}) \;\; , \;\; z_i(t) \sim \text{ independent in index i, dependent in time,}$$

$$z_i(t) \sim N(0, \lambda_i^2), \;\; \text{Cov}(z_i(t_1), z_j(t_2)) = I(i = j) C_i(t_1, t_2)$$

Discrete version

Eigen vector expansion of process.

$$Z = \sum_{i=1}^{m} z_i \psi_i \ , \ z_i \sim \text{ independent, } N(0, \lambda_i^2)$$

Eigenvector decomposition of covariance matrix.

$$\mathbf{C}_Z = \mathbf{\Psi}\mathbf{\Lambda}^2\mathbf{\Psi}^T, \ \ \mathbf{\Lambda}^2 = \text{diag}\{\lambda_i^2\} \ \ \mathbf{\Psi} = [\psi_i, ..., \psi_m]$$

Under suitable conditions on the time series we can estimate the time lag zero spatial correlation function from the repeated time samples of the process.

Eigen vector decomposition

- Linear combination $a_t = \psi' \mathbf{Z}_t$ describing most of the variability
- Scaling problem: Restrict ψ to $\psi'\psi = 1$
- Note: $\text{var}[a_t] = \psi' \mathbf{C} \psi$
- Mathematical formulation: Find $\max \psi' \mathbf{C} \psi$ subject to $\psi'\psi = 1$
- Mathematical solution: ψ is the eigenvector of $\mathbf{C}$ corresponding to the largest eigenvalue, ψ_1.

- Continue: Find $\max \psi' \mathbf{C} \psi$ subject to $\psi'\psi = 1$ *and* $\psi'\psi_1 = 0$
- Solution: ψ is the eigenvector of $\mathbf{C}$ corresponding to the second largest eigenvalue, ψ_2.
- ...

Empirical orthogonal function (EOF) in practice

- Data $\mathbf{Z}_t = (Z_t(\mathbf{s}_1), ..., Z_t(\mathbf{s}_m))'$
- Assume $\text{Cov}[Z_t(\mathbf{s}_i), Z_t(\mathbf{s}_j)] = C_{ij}$, i.e. independent of time.

- Estimate covariance function:

$$\widehat{\mathbf{C}} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{Z}_t - \bar{\mathbf{Z}})(\mathbf{Z}_t - \bar{\mathbf{Z}})'$$

$$\bar{\mathbf{Z}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_t$$

- Compute eigen vectors of empirical time lag zero correlation.
- Principal components / EOF

EOF when $m > T$

- $\widehat{\mathbf{C}} = \frac{1}{T}\sum_t(\mathbf{Z}_t - \bar{\mathbf{Z}})(\mathbf{Z}_t - \bar{\mathbf{Z}})'$ only have $T-1$ non-zero eigenvalues
- Calculation of eigenvectors numerically unstable
- Linear algebra: $\mathbf{W}$ $m \times T$ matrix. Then
  - $\mathbf{W}\mathbf{W}'$ (dim $m \times m$) and $\mathbf{W}'\mathbf{W}$ (dim $T \times T$) have the same eigenvalues
  - If $\boldsymbol{\xi}_i$ is ith eigenvector of $\mathbf{W}\mathbf{W}'$, then $\psi_i = \mathbf{W}\boldsymbol{\xi}_i/\sqrt{\boldsymbol{\xi}_i'\mathbf{W}'\mathbf{W}\boldsymbol{\xi}_i}$ is the ith eigenvector of $\mathbf{W}'\mathbf{W}$
- Define

$$
\begin{aligned}
\widetilde{\mathbf{Z}_t} =& \mathbf{Z}_t - \bar{\mathbf{Z}}\\
\widetilde{\mathbf{Z}} =& (\widetilde{\mathbf{Z}}_1, ..., \widetilde{\mathbf{Z}}_T), \quad m \times T \text{ matrix}\\
\widehat{\mathbf{C}} =& \frac{1}{T}\widehat{\mathbf{Z}}\widehat{\mathbf{Z}}', \quad m \times m \text{ matrix}\\
\mathbf{A} =& \frac{1}{T}\widehat{\mathbf{Z}}'\widehat{\mathbf{Z}}', \quad T \times T \text{ matrix}
\end{aligned}
$$

- Calculate eigenvalues/vectors from $\mathbf{A}$

EOF-Temporal formulation

- Data $\mathbf{Z}(\mathbf{s}_i) = (Z_1(\mathbf{s}_i), ..., Z_T(\mathbf{s}_i))'$
- Assume $\text{Cov}[Z_t(\mathbf{s}_i), Z_v(\mathbf{s}_i)] = \Gamma_{tv}$, i.e. independent of space.
- Linear combination $a(\mathbf{s}_i) = \boldsymbol{\psi}'\mathbf{Z}(\mathbf{s}_i)$ describing most of the variability?
- Scaling problem: Restrict $\boldsymbol{\psi}$ to $\boldsymbol{\psi}'\boldsymbol{\psi} = 1$
- Note: $\text{var}[a(\mathbf{s}_i)] = \boldsymbol{\psi}'\boldsymbol{\Gamma}\boldsymbol{\psi}$
- Solution: $\boldsymbol{\psi}$ is the eigenvector of $\boldsymbol{\Gamma}$ corresponding to the second largest eigenvalue, $\boldsymbol{\psi}_2$.
- ...
- In practice:
  $\widehat{\mathbf{C}} = m^{-1}\sum_i(\mathbf{Z}(\mathbf{s}_i) - \bar{\mathbf{Z}})(\mathbf{Z}(\mathbf{s}_i) - \bar{\mathbf{Z}})'$,
  $\bar{\mathbf{Z}} = m^{-1}\sum_i \mathbf{Z}(\mathbf{s}_i)$
- EOF/Principal components in time

Testing for spatio-temporal structure

Testing for spatial structure: Moran's I (sec 4.2):

$$I \equiv \frac{\mathbf{Z}'\mathbf{W}\mathbf{Z}}{\mathbf{Z}'\mathbf{Z}} = \frac{\sum_{ij} w_{ij} Z(\mathbf{s}_i) Z(\mathbf{s}_j)}{\sum_i Z(\mathbf{s}_i)^2}$$

Space-Time Index (STI):

$$STI \equiv \frac{\sum_{t=2}^{T} \sum_{i=1}^{m} \sum_{j=1}^{m} w_{ij,t-1} Z(\mathbf{s}_i;t) Z(\mathbf{s}_j;t-1)}{m(T-1)(\sum_{t=2}^{T} \sum_{i=1}^{m} \sum_{j=1}^{m} w_{ij,t-1})(\sum_{t=1}^{T} \sum_{i=1}^{m} Z(\mathbf{s}_i;t)^2)}$$

Testing by permutation (page 303):

- Randomly permute the values of $\{Z(\mathbf{s}_i;t)\}$ and calculate STI
- Compare STI from original data with simulated values

STI for SST-data

$$w_{ij} = \begin{cases} 1/d_{ij} & d_{ij} \leq 4 \\ 0 & d_{ij} > 4 \end{cases}$$