Space-time modelling: A case study

Johan Lindström¹,²

¹Centre for Mathematical Sciences
Lund University

²Department of Statistics
University of Washington

Danish Technical University Aqua
Charlottenlund
October 17, 2012

Overview

Lecture 1: Spatial modelling
Lecture 2: Gaussian Markov Random Fields
Lecture 3: Spatio-Temporal modelling — A case study
  1. Spatio-temporal frameworks
  2. Examples of Spatio-temporal modelling
  3. Modelling NOₓ in Los Angeles — A case study

Spatio-temporal modelling

Spatio-temporal models typically fall into one of two main categories:

1. Spatial fields evolving in time
2. Spatially varying time series

The modelling strategy should be based on the available data, the scientific question and computational considerations.

Lots of recent work, less available “off-the-shelf” methods/packages.

Spatio-temporal modelling — Examples

► A class of covariance functions for space-time data (Gneiting, 2002; Fuentes et al., 2008).
► Modelling of PM₂.₅ using a series of spatially correlated fields (Paciorek et al., 2009).
► A separable space-time model formulated using GMRF:s (Cameletti et al., 2012).
► Physics based modelling of rainfall, used to postprocess forecasts (Sigrist et al., 2012).
► Dynamic model for coupled environmental variables (Ippoliti et al., 2012).
The MESA Air study

- The Multi-Ethnic Study of Atherosclerosis (MESA) is a large study of cardiovascular diseases.
- It follows more than 6,000 people from six communities:
  - Baltimore
  - Chicago
  - Los Angeles
  - Minneapolis–Saint Paul
  - New York
  - Winston–Salem
- MESA Air is an EPA study of how air pollution affects cardiovascular diseases.
- The primary pollutants in the MESA Air study are PM$_{2.5}$ and NO$_x$ (this case).
- See Szpiro et al. (2010); Sampson et al. (2011); Lindström et al. (2011)

Available data — Los Angeles

<table>
<thead>
<tr>
<th>Type of site</th>
<th>sites</th>
<th>Start date</th>
<th>End date</th>
<th>Nbr. of obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQS</td>
<td>20</td>
<td>1999–01–27</td>
<td>2009–10–07</td>
<td>4180</td>
</tr>
<tr>
<td>MESA fixed</td>
<td>5</td>
<td>2005–12–07</td>
<td>2009–07–01</td>
<td>399</td>
</tr>
<tr>
<td>MESA home</td>
<td>84</td>
<td>2006–05–24</td>
<td>2008–02–13</td>
<td>155</td>
</tr>
<tr>
<td>MESA snapshot$^1$</td>
<td>177</td>
<td>2006–07–05</td>
<td>2007–01–31</td>
<td>449</td>
</tr>
</tbody>
</table>


Available data — Los Angeles

Glendora 60370016

Lynwood 60371301

Costa Mesa 60590007

A Home close to Lynwood 60371301
We model the logarithm of each 2-week average as
\[ y(s,t) = \sum_{i=1}^{m} \beta_i(s)f_i(t) + \nu(s,t). \]

\( f_i(t) \) Smooth temporal trends with \( f_1(t) \equiv 1 \) and \( f_2(t), \ldots, f_m(t) \) mean zero.
\( \beta_i(s) \) Spatially varying coefficients for the temporal trends.
\( \nu(s,t) \) Residuals, modelled as a mean zero Gaussian field that is independent in time but has spatial structure.

The smooth temporal trends, \( f_i(t) \) are compute using a singular value decomposition of the data matrix, \( Y \) (see Fuentes et al., 2006, and the computer exercise).

\[ \beta_i(s) \in \mathbb{N}(X_i \alpha_i, \Sigma_{\beta_i}(\theta_B)) \]
\( X_i \) Design matrices, that includes geographical covariates (different for each \( i \)).
\( \alpha_i \) Regression coefficients.
\( \Sigma_{\beta_i} \) Covariance matrix describing additional spatial dependence not captured by the geographical covariates.
\( \theta_B \) Parameters of the covariance matrices.

\[ \nu(s,t) \in \mathbb{N}(0, \Sigma_{\nu}(\theta_\nu)) \]
\( \Sigma_{\nu} \) Block diagonal covariance matrix for the residuals.
\( \theta_\nu \) Parameters of the residual covariance matrix.

Writing the model on matrix form we obtain
\[ Y = FB + \nu \]
where
\[ B \in \mathbb{N} \begin{bmatrix} X_1 & 0 & 0 & \cdots & 0 \\ 0 & X_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & X_m \end{bmatrix}, \begin{bmatrix} \Sigma_{\beta_1}(\theta_B) & 0 & 0 \\ 0 & \Sigma_{\beta_2}(\theta_B) & 0 \\ \vdots & \vdots & \ddots \\ 0 & 0 & 0 & \Sigma_{\beta_m}(\theta_B) \end{bmatrix} \]

and
\[ \nu \in \mathbb{N} \begin{bmatrix} \Sigma_{t=1,\nu}(\theta_\nu) & 0 & \cdots \\ 0 & \Sigma_{t=2,\nu}(\theta_\nu) & \cdots \end{bmatrix} \]

Both \( B \) and \( \nu \) are Gaussian and we have
\[ [Y|\theta_B, \theta_\nu, \alpha] \in \mathbb{N}(FX\alpha, \Sigma_{\nu}(\theta_\nu) + F\Sigma_{\beta}(\theta_B)F^\top), \]

Our model is now
\[ [Y|\theta_B, \theta_\nu, \alpha] \in \mathbb{N}(FX\alpha, \Sigma_{\nu}(\theta_\nu) + F\Sigma_{\beta}(\theta_B)F^\top), \]

and parameters can be estimated by maximising the log-likelihood
\[ l(\theta_B, \theta_\nu, \alpha|Y). \]

Estimation of the above model is computationally expensive and we reduce the computational cost by:
1. Use **profile likelihood** to reduce the number of parameters in the log-likelihood.
2. Utilise the block structure in \( \Sigma_B \) and \( \Sigma_\nu \) to reduce the computational burden.
Likelihood simplifications

- Matrix algebra can be used to “simplify” the likelihood (Harville, 1997; Petersen and Pedersen, 2008).
- As an example we study the determinant of the log-likelihood

\[ \log |\Sigma + F\Sigma_B F^T| = \log |\Sigma| + \log |\Sigma_B| + \log |\Sigma_B^{-1} + F^T\Sigma_B^{-1} F| \]

This may not seem simpler but:
1. \(\Sigma + F\Sigma_B F^T\) is dense \(N \times N\)-matrix, and computing the determinant requires \(O(N^3)\) operations.
2. \(\Sigma\) and \(\Sigma_B\) are both block diagonal, with “small” blocks.
3. \(\Sigma_B^{-1} + F^T\Sigma_B^{-1} F\) is a dense \(mn \times nm\)-matrix. Computing the determinant requires \(O(m^3n^3)\) operations, with \(mn \ll N\).

Where:
- \(N\) Total number of observations.
- \(n\) Total number of observed sites.
- \(m\) Number of temporal basis functions (incl. intercept).

Computational issues

**Computer time for evaluation of the profile log-likelihood**

![Graph showing computer time for evaluation of the profile log-likelihood](image)

**Predicted average NO\(_x\) concentration — Los Angeles**

![Map showing predicted average NO\(_x\) concentration](image)
Model validation — NO\textsubscript{x} in Los Angeles

![Graphs showing NO\textsubscript{x} levels in various locations over time.]

LUNCH
(Questions?)

Bibliography I


Bibliography II


