Spatio-Temporal Modelling of Precipitation using Gaussian Markov Random Fields

Johan Lindström & Finn Lindgren

Centre for Mathematical Sciences
Lund University

TIES
Kelowna June 12, 2008

Overview

Background
- African Sahel
- Precipitation

Gaussian Markov random fields
- Approximation of Matérn fields

Model
- Basis functions for the mean
- Model Graph

Estimation
- Markov chain Monte Carlo
- Computational burden

Results
- Future work

African Sahel

- A semi-arid region directly south of the Sahara desert.
- The region suffered a severe drought from the late 1960’s through the early 1980’s.
- Recent studies indicate a vegetation recovery
- The main limiting factor to vegetation growth in the region is (lack of) water.

Precipitation

Yearly precipitation, in metres, for the 443 stations that have reported measurements for 1982. Data from the Global Historical Climatology Network (NOAA, 2007).
Gaussian Markov Random Fields (GMRF:s)

- A Gaussian Markov random field (GMRF) is a Gaussian random field with a Markov property.
- The neighbours $N_i$ to a point $s_i$ are the points that in some sense are close to $s_i$.
- The Gaussian random field $x \in N(\mu, Q^{-1})$ has a joint distribution that satisfies
  \[ p(x_i|\{x_j : j \neq i\}) = p(x_i|\{x_j : j \in N_i\}). \]
  
  \[ j \notin N_i \iff x_i \perp x_j \mid \{x_k : k \notin \{i, j\}\} \iff Q_{i,j} = 0. \]
- Fast algorithms that utilise the sparsity of $Q$ exist (c-package GMRFlib by Rue, 2007)

See Rue and Held (2005) for extensive details on GMRF:s.

Approximating Gaussian Markov random fields

- A GMRF might be computationally effective but it is difficult in construct precision matrices that result in reasonable covariance functions for the underlying Gaussian fields.
- Rue and Tjelmeland (2002) created GMRF:s on rectangular grids in that approximate Gaussian fields with a wide class of covariance functions.
- Having the field defined only on a regular grid leads to issues with mapping the observations to the grid points.
- An alternative that allows construction of the GMRF directly on a set of irregular points would be preferable.

Approximation of Matérn fields

- As noted by Whittle (1963), the solutions, in $\mathbb{R}^d$, to
  \[ (x^2 - \Delta)^{\alpha/2} x(s) = \mathcal{E}(s), \]
  are Gaussian random fields with covariances that correspond to the Matérn family
  \[ C(x(s), x(s + \tau)) \propto (x\|\tau\|)^\alpha K_\alpha(x\|\tau\|), \]
  \[ \alpha = \nu + d/2. \]
- This was used by Lindgren and Rue (2007) to construct GMRF:s that approximate fields with Matérn covariance for $\alpha \in \mathbb{Z}^+$. 
- Lindgren and Rue further argued that solutions to the SPDE can be used as a natural generalisation of Matérn covariances on general manifolds.

Approximation of Matérn fields (cont.)

- The precision matrix of the approximating GMRF is found using the finite element method on a triangulation of irregularly spaced points.
- The resulting GMRF is defined on the points of the triangulation, making it suitable for modelling fields that are observed at irregular locations.
Model

- Assume an underlying Gaussian field
  \( X \in \mathcal{N}(\mu, Q^{-1}) \).

- Transform the precipitation data
  \( y = \tilde{y}^{1/3} \cdot (1 - 0.13\tilde{y}^{1/3}) \).

- The transformed data is modelled as Gaussian observations of the underlying field
  \( Y(s_i, t)|X \in \mathcal{N}(X(s_i, t), \sigma^2) \).

Model – Temporal dependence

- The time dependence is modelled using an AR(1) process
  \[
  X_t - \mu = a(X_{t-1} - \mu) + \eta_t, \quad \eta_t \in \mathcal{N}(0, Q_S(x^2, \chi)^{-1}), \quad X_1 \in \mathcal{N}(\mu, Q_S^{-1}/(1 - a^2)).
  \]

- Expectation is modelled using a set of basis functions \( \mu = B\theta \).

- Resulting distribution for X
  \[
  X \in \mathcal{N}(1 \otimes B\theta, (Q_t \otimes Q_s)^{-1}).
  \]

- Priors for the parameters \( \phi = \{x^2, \chi, a, \sigma^2, \theta\} \).

Basis functions for the mean

Left: Latitude is modelled using a broken linear trend.
Right: In addition to the broken trend 15 B-spline surface basis functions are used.

Model graph

Directed acyclic graph describing the resulting hierarchical model.
We want to obtain posterior distributions for the parameters and the underlying field given observations.

To estimate the posterior of the field given observations we use an MCMC-approach.

It is possible to integrate out $\theta$ and $X$ reducing the problem to four unknown parameters $\phi = \{x^2, \chi, a, \sigma^2\}$.

The posterior for $p(\phi|Y)$ can be calculated,

$$p(\phi|Y) \propto \exp \left( -\frac{Y^T Y}{2\sigma^2} + \frac{b^T \hat{Q}^{-1} b}{2} + \frac{b_0^T \hat{Q}_0^{-1} b_0}{2} \right) \cdot \left( \frac{|Q_S|^T (1 - a^2) N}{|\hat{Q}| \hat{Q}_0 |\sigma^2 I|} \right)^{1/2} p(\phi),$$

where

$$b = \frac{A^T Y}{\sigma^2}, \quad \hat{Q} = (Q_T \otimes Q_S) + \frac{A^T A}{\sigma^2},$$

$$b_0 = Q_0 \mu_0 + (Q_T 1 \otimes Q_S B)^T \hat{Q}^{-1} b,$$

$$\hat{Q}_0 = Q_0 + (Q_T 1 \otimes Q_S B)^T \hat{Q}^{-1} \left( \frac{A^T A}{\sigma^2} \right) (1 \otimes B).$$

First we transform the parameters to obtain parameters valid on $\mathbb{R}$:

$$\tilde{\sigma}^2 = \log(\sigma^2), \quad \tilde{a} = \log(1 + a) - \log(1 - a),$$

$$\tilde{x}^2 = \log(x^2), \quad \tilde{\chi} = \log(\chi).$$

Construct a proposal distribution by using a random walk proposal on the transformed variable space.

However initial runs on a reduced datasets with coarser triangulation showed strong dependencies in the posterior distribution, leading us to a correlated proposal distribution:

$$\tilde{\psi}_{\text{new}} \in N \left( \tilde{\psi}_{\text{old}}, \Sigma_{\text{proposal}} \right).$$

Two-dimensional histograms illustrating the dependence between the components of $(\psi|Y)$ before (left pane) and after (right pane) the transformation.
Markov chain Monte Carlo (cont.)

Parameter trajectories for the MCMC-simulation.

Results – Interpolation

Interpolated yearly precipitation, in metres, for 1982.

Results – Model verification

Model verification using observations from 10 stations left out of the original analysis.

Computational burden

- The dominating cost for each MCMC iteration is calculation of the Cholesky factorisation of $Q$.
- Inverting a full covariance matrix is $\mathcal{O}(n^3)$,
- Given a spatial GMRF on a lattice with $n$ points the Cholesky factor is $\mathcal{O}(n^{3/2})$,
- Given a spatio-temporal GMRF on a lattice with $n$ points the Cholesky factor is $\mathcal{O}(n^2)$.

How bad is the additional burden for the temporal dependence?

<table>
<thead>
<tr>
<th>$n$</th>
<th>Spatial</th>
<th>Spatio-temporal</th>
</tr>
</thead>
<tbody>
<tr>
<td>2039</td>
<td>0.05 s</td>
<td>—</td>
</tr>
<tr>
<td>15 \cdot 2039 = 30585</td>
<td>2.48 s</td>
<td>46.8 s</td>
</tr>
</tbody>
</table>

Johan Lindström - johanl@maths.lth.se

Spatio-Temporal Modelling of Precipitation using GMRF:s
Future work

- Avoid the MCMC by using optimisation and numerical integration (c-code exists, see Rue and Martino, 2007).
- Use temporal basis functions with spatially dependent coefficients.
- Investigate non-stationary formulations of the GMRF:s field.
- Include altitude as a covariate.
- Investigate dependence between precipitation and vegetation.

Bibliography


www.maths.lth.se/matstat/staff/johanl/papers/LUTFMS5074.pdf