SPATIO-TEMPORAL WEATHER MODELS

Chris Glasbey & Dave Allcroft

Biomathematics and Statistics Scotland
QUESTION 1: If the Pentland Hills were covered with solar panels how would energy output vary?

Solar radiation was measured at 10 sites every 10 minutes for a month
All data
QUESTION 2: What can be done if rainfall is needed at a finer spatial scale than recorded?

40² km squares

⇒

8² km squares

scale

disaggregation
PLAN

1. Solar radiation
   - transformation to normality
   - STARMA model
   - results

2. Rainfall

3. Summary
1. Solar radiation - transformation to normality

\[
\text{Clearness index} = 100 \times \frac{\text{observed radiation}}{\text{theoretical maximum radiation}}
\]
Sample auto- and cross-correlations of normalised clearness index (\(Y\)):
1. Solar radiation - STARMA model

For the spatio-temporal process \( Y \), we could

- model \( V \) directly, perhaps using a Matern correlation function
- use a GMRF (Gaussian Markov random field) in space and time
- use a STARMA (spatio-temporal auto-regressive moving average) model

To simulate solar radiation forwards in time, STARMA models are the natural choice
The simplest STARMA model is a STAR(1) process:

\[ Y_t = \Phi Y_{t-1} + \epsilon_t \quad t = \ldots, -1, 0, 1, \ldots \]

On an \( n \times n \) lattice, vectors are of length \( n^2 \)

In particular, for a first-order neighbourhood

\[ \Phi_{ni+j, nk+l} \equiv \Phi_{ij, kl} = \begin{cases} 
\phi_0 & \text{if } i = k, j = l \\
\phi_1 & \text{if } |i - k| + |j - l| = 1 \\
0 & \text{otherwise}
\end{cases} \]

\[ \epsilon_t \sim N(0, \Lambda), \text{ independent at each } t, \text{ and we use a Matern process} \]

\[ \Lambda_{ij, kl} \propto \left( \frac{\sqrt{(i - k)^2 + (j - l)^2}}{\beta_1} \right)^{\beta_2} K_{\beta_2} \left( \frac{\sqrt{(i - k)^2 + (j - l)^2}}{\beta_1} \right) \]
1. Solar radiation - results

Inter-site distances: 3-20km

Area of interest approximately: 10×10km

So use 64×64 toroidal grid, with 1km spacing between lattice points

STAR(1) process fitted by minimising sum of squares of differences between sample and expected correlations

<table>
<thead>
<tr>
<th>neighbourhood order</th>
<th>r.m.s. difference</th>
<th>$\hat{\phi}_0$</th>
<th>$\hat{\phi}_1$</th>
<th>$\hat{\phi}_2$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0785</td>
<td>0.953</td>
<td></td>
<td></td>
<td>248</td>
<td>0.1819</td>
</tr>
<tr>
<td>1</td>
<td>0.0709</td>
<td>0.116</td>
<td>0.210</td>
<td></td>
<td>3115</td>
<td>0.0258</td>
</tr>
<tr>
<td>2</td>
<td>0.0704</td>
<td>0.125</td>
<td>0.056</td>
<td>0.151</td>
<td>1587</td>
<td>0.0254</td>
</tr>
</tbody>
</table>
Fit of STAR(1) process with first-order neighbourhood:
Simulation of clearness index on $64^2$ km grid at one time point:
Average over $10^2$ km square from initial low value:
single realisation, median and 95% limits from 999 independent simulations
PLAN

1. Solar radiation
   • transformation to normality
   • STARMA model
   • results

2. Rainfall
   • transformation to normality
   • GMRF model
   • results

3. Summary
2. Rainfall - transformation to normality

We have 12 hours $\times$ 1200km $\times$ 600km of data for Arkansas USA

We will build a model using fine-resolution data

then use it to disaggregate data at a coarser scale and see how well we recover the fine scale
Rainfall data are far from normal

- 91% of values are zero
- Non-zero values have skewed distribution:

\[ \Rightarrow \]

Latent Gaussian
Latent Gaussian \((Y)\)

\[ \downarrow \]

Rainfall
Estimation of autocorrelation

Numerically maximise bivariate likelihood of pairs of points a given space-time lag \((k, l, s)\) apart:

\[
p(Y_{ijt}, Y_{i-k,j-l,t-s}) = \begin{cases} 
\Phi_2(\alpha_0, \alpha_0, \rho) & \text{if } Y_{ijt} = Y_{i-k,j-l,t-s} = \ast \\
\phi(Y_{ijt}) \Phi \left( \frac{\alpha_0 - \rho Y_{ijt}}{\sqrt{1-\rho^2}} \right) & \text{if only } Y_{i-k,j-l,t-s} = \ast \\
\phi_2(Y_{ijt}, Y_{i-k,j-l,t-s}, \rho) & \text{otherwise}
\end{cases}
\]
Autocorrelation estimates

**Time lag $s = 0$**

<table>
<thead>
<tr>
<th>Time lag $s$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.57</td>
<td>.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.68</td>
<td>.62</td>
<td>.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.83</td>
<td>.73</td>
<td>.65</td>
<td>.58</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1.</td>
<td>.89</td>
<td>.75</td>
<td>.66</td>
<td>.59</td>
</tr>
</tbody>
</table>

**Time lag $s = 1$**

<table>
<thead>
<tr>
<th>Time lag $s$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.50</td>
<td>.47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.57</td>
<td>.53</td>
<td>.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.63</td>
<td>.59</td>
<td>.55</td>
<td>.51</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>.68</td>
<td>.65</td>
<td>.60</td>
<td>.55</td>
<td>.51</td>
</tr>
</tbody>
</table>

0 1 2 3 4
2. Rainfall - GMRF model

Again for the spatio-temporal process we could

- model $V$ directly
- use a GMRF in space and time
- use a STARMA model

To disaggregate rainfall, which requires simulation from conditional distributions, GMRFs are most suitable
For a GMRF

\[ \text{prob}(Y) \propto \exp\left[-\frac{1}{2}Y^T Q Y\right] \]

where \( Q \) is the precision matrix, with non-zero entries specifying the conditional dependencies between elements in \( Y \).

For example, a \( 3 \times 3 \times 3 \) neighbourhood:

\[
\begin{array}{ccc}
\text{t-1} & \text{t} & \text{t+1} \\
\hline
& & \\
& & \\
\end{array}
\]

requires 5 parameters, if we allow for symmetries.
2. Rainfall - results

Fit my minimising weighted sum of squares of differences in correlations

For neighbourhood size $5 \times 5 \times 3$:

- Time lag $s = 0$
- Time lag $s = 1$
Simulation

- Gibbs sampling to update blocks of $5 \times 5$ pixels. Full conditional distribution is multivariate normal:

$$
\begin{pmatrix}
Y_A \\
Y_B
\end{pmatrix}
\sim \text{MVN}
\left(
\begin{pmatrix}
0 \\
0
\end{pmatrix},
\begin{pmatrix}
Q_{AA} & Q_{AB} \\
Q_{BA} & Q_{BB}
\end{pmatrix}^{-1}
\right)
$$

- Need to constrain the 25 simulated values ($Y_A$) such that total rain in block corresponds to observed

- Update all blocks in turn. Burn-in 500 iterations, then run 5000 iterations. Mixing quite slow due to high correlations
Examples of disaggregation

Observed hour 3  Simulation 1  Simulation 2  scale
3. Summary

We have developed spatio-temporal models for solar radiation and rainfall.

In both cases we had to transform to marginal normality.

For solar radiation the aim was simulation, so we used a STARMA model.
For rainfall the aim was disaggregation, so we used a GMRF model.

In both cases we simplified computations by approximating space by a torus, and fitted by minimising differences between observed and expected correlations.

Further details are in Allcroft & Glasbey (2003) and Glasbey & Allcroft (2007), at: http://www.bioss.sari.ac.uk/staff/chris.html