Spatial Statistics with Image Analysis
Lecture 13

Johan Lindström

December 14, 2018

Lecture L13

Home assignment 1
- Covariance functions
- Reconstructions
- Validation

Home assignment 2
- Reconstructions
- Variances
- CAR — $\kappa = 0$

Software

Covariate Selection

- Scatter plots for elevation and square root of precipitation
- Scatter plots for x and y coordinates versus square root of precipitation

Software

Johan Lindström - johanl@maths.lth.se
Covariate Estimates

\[ y_i = \beta_0 + \beta_{\text{elev}} \cdot \text{elev}_i + \beta_{\text{elev}^2} \cdot \text{elev}_i^2 + \beta_x \cdot x_i + \varepsilon_i \]

Suitable model is

Covariance Parameters

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>Matern</th>
<th>Exp.</th>
<th>Gaussian</th>
</tr>
</thead>
<tbody>
<tr>
<td>\sigma^2</td>
<td>—</td>
<td>0.160</td>
<td>0.163</td>
<td>0.134</td>
</tr>
<tr>
<td>\kappa</td>
<td>—</td>
<td>0.063</td>
<td>0.030</td>
<td>46.484</td>
</tr>
<tr>
<td>\nu</td>
<td>—</td>
<td>0.981</td>
<td>0.5</td>
<td>\infty</td>
</tr>
<tr>
<td>\sigma^2 + \sigma^2_{\varepsilon}</td>
<td>0.157</td>
<td>0.000</td>
<td>0.000</td>
<td>0.030</td>
</tr>
<tr>
<td>\sigma^2_{\varepsilon}</td>
<td>0.157</td>
<td>0.160</td>
<td>0.163</td>
<td>0.164</td>
</tr>
</tbody>
</table>
Residuals

For regression all residuals should be
\[ \hat{e} = Y - X\hat{\beta} \quad \hat{e} \in N(0, \sigma^2_e) \]

For the Kriging residuals from the validation data should be
\[ \hat{e} = Y_u - E(Y_u | Y_k, \hat{\theta}) \quad \hat{e} \in N(0, V(Y_u | Y_k, \hat{\theta})) \]

Thus we could check \( \hat{e}_{OLS} \) for normality in a standard normplot. However for the Kriging we need the standardized residuals:
\[ z = \frac{Y_u - E(Y_u | Y_k, \hat{\theta})}{\sqrt{V(Y_u | Y_k, \hat{\theta})}} \]

Validation

- Most Kriging models perform similarly for \( E(Y_u | Y_k, \hat{\theta}) \)
- Differences between Kriging models are larger in \( V(Y_u | Y_k, \hat{\theta}) \)
- Prediction errors are smaller for Kriging.
- The regression has wider confidence bands for prediction.

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>Matern</th>
<th>Gaussian</th>
<th>Exp.</th>
<th>Spherical</th>
<th>Cauchy</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.086</td>
<td>0.024</td>
<td>0.031</td>
<td>0.026</td>
<td>0.027</td>
<td>0.031</td>
</tr>
</tbody>
</table>
>> p = amd(Q_x);
>> tmp = 1:length(p);
>> tmp(p) = tmp;
>> imagesc( reshape(tmp,sz) )

The mean field captures a rather small part of the counts.
Parameters

\[ \begin{array}{|c|ccc|} \hline
\text{Parameter} & \text{CAR} & \text{SAR} & \text{OSC} \\
\hline
\kappa^2 & 7 \cdot 10^{-7} & 0.169 & 0.169 \\
\tau & 0.546 & 0.273 & 0.273 \\
\gamma & & 1 & \\
\hline
\end{array} \]

\[ Q_{\text{CAR}} = \tau (\kappa^2 \mathbf{C} + \mathbf{G}) \]

\[ Q_{\text{SAR}} = \tau (\kappa^4 \mathbf{C} + 2\gamma \kappa^2 \mathbf{G} + \mathbf{G}_2) \]

\[ Q_{\text{OSC}} = \tau (\kappa^4 \mathbf{C} + 2\gamma \kappa^2 \mathbf{G} + \mathbf{G}_2) \]
Variances

To obtain variances for the posterior field we need to compute

\[ V(z|Y) = V([A \ B] \tilde{x}|Y) \approx [A \ B] Q_{xy}^{-1} [A \ B]^T \]

which requires the inverse of a 5003-by-5003 matrix, or the solution of 5003 equations systems.

The inverse takes 2.57s and requires 191MB of memory (Cholesky factor: 0.021s).

For \( \beta \) this is practical since we only need to solve 3 equation systems.

Variance — Simulation

The better option is to:
1. Sample from \( N(0, Q_{xy}) \).
2. Given these sample we compute samples from
   \[ z = [A \ B] \tilde{x} \]
3. Finally we compute variances of the \( z \) samples.

Variance — Timings

\[
\begin{array}{ll}
\text{chol}(Q_{xy}) & 0.021 \text{s} \\
1000 \text{ samples of } \tilde{x} & 0.560 \text{s} \\
\text{Compute } V(z) & 0.038 \text{s} \\
\text{Total} & 0.619 \text{s} \\
\text{chol}(Q_{xy}) & 0.021 \text{s} \\
R_{xy}^{-1} \tilde{A} & 0.435 \text{s} \\
\text{Total} & 0.456 \text{s} \\
A^T \cdot Q_{xy}^{-1} \tilde{A} & 2.589 \text{s}
\end{array}
\]
Reconstructions — Variances

The variance increases for low counts. This can be understood by comparing the variance for Gaussian errors and for the Poisson data.

\[ Q_{x|y} = Q + A^\top Q \varepsilon A \]
\[ Q_\varepsilon = I \sigma_\varepsilon^{-2} \]
\[ Q_{x|y} = Q - A^\top \Delta f A \]
\[ f''(x) = -\exp(z_i) \]
Model for CAR and SAR fields with \( \kappa = 0 \) (IGMRF)

The CAR and SAR fields with \( \kappa = 0 \) are improper GMRF:s (a.k.a. intrinsic GMRF:s — IGMRF) since \( Q_1 = 0 \), i.e. \(|Q| = 0\).

- The resulting fields are invariant to the addition of a constant since

\[
p(x + c_1) \propto \exp \left( -\frac{1}{2} (x + c_1)^\top Q (x + c_1) \right) = \exp \left( -\frac{1}{2} x^\top Q x \right) \propto p(x)
\]

- Since \( p(x + c_1) = p(x) \) we do not need any intercept in \( \mu = B \beta \).

Software

**Matlab:**
- **GPstuff** Gaussian process models for Bayesian analysis, research.cs.aalto.fi/pml/software/gpstuff/

**R:**
- **geoR** Basic spatial statistics.
- **R-INLA** GMRFs and non-Gaussian data. www.r-inla.org/