Hierarchical Modelling for non-Gaussian Spatial Data

Sudipto Banerjee\textsuperscript{1} and Andrew O. Finley\textsuperscript{2}

\textsuperscript{1} Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota, U.S.A.
\textsuperscript{2} Department of Forestry & Department of Geography, Michigan State University, Lansing Michigan, U.S.A.

March 3, 2010

\textbf{Spatial Generalized Linear Models}

- First stage: $Y(s_i)$ are conditionally independent given $\beta$ and $w(s_i)$, so $f(y(s_i)|\beta, w(s_i), \gamma)$ equals
  \[ k(y(s_i), \gamma) \exp \left( \gamma \left[ y(s_i) - \psi(\eta(s_i)) \right] \right) \]
  where $\psi(E(Y(s_i))) = \eta(s_i) = x(s_i)\beta + w(s_i)$ (canonical link function) and $\gamma$ is a dispersion parameter.
- Second stage: Model $w(s)$ as a Gaussian process:
  \[ w \sim N(0, \sigma^2 R(\phi)) \]
- Third stage: Priors and hyperpriors.
- No process for $Y(s)$, only a valid joint distribution
- Not sensible to add a pure error term $\epsilon(s)$

\textbf{Comments}

- Often data sets preclude Gaussian modelling: $Y(s)$ may not even be continuous
- Example: $Y(s)$ is a binary or count variable
  - species presence or absence at location $s$
  - species abundance from count at location $s$
  - continuous forest variable is high or low at location $s$
- Replace Gaussian likelihood by exponential family member
- Diggle Tawn and Moyeed (1998)

\textbf{Illustration}

Illustration from:


**Spatial Generalized Linear Models**

We are modelling with spatial random effects
- Introducing these in the transformed mean encourages means of spatial variables at proximate locations to be close to each other
- Marginal spatial dependence is induced between, say, $Y(s)$ and $Y(s')$, but observed $Y(s)$ and $Y(s')$ need not be close to each other
- Second stage spatial modelling is attractive for spatial explanation in the mean
- First stage spatial modelling more appropriate to encourage proximate observations to be close.

**Binary spatial regression: forest/non-forest**

We illustrate a non-Gaussian model for point-referenced spatial data:
- Objective is to make pixel-level prediction of forest/non-forest across the domain.
Illustration
Binary spatial regression: forest/non-forest
We illustrate a non-Gaussian model for point-referenced spatial data:

- Objective is to make pixel-level prediction of forest/non-forest across the domain.
- Data: Observations are from 500 georeferenced USDA Forest Service Forest Inventory and Analysis (FIA) inventory plots within a 32 km radius circle in MN, USA.
- The response $Y(s_i)$ is a binary variable, with
  
  $Y(s_i) = \begin{cases} 
  1 & \text{if inventory plot is forested} \\
  0 & \text{if inventory plot is not forested}
  \end{cases}$

Parameter estimates (posterior medians and upper and lower data:

Assume vague flat for $\beta$, a Uniform$(3/32, 3/0.5)$ prior for $\phi$, and an inverse-Gamma$(2, \cdot)$ prior for $\sigma^2$.

Parameters updated with Metropolis algorithm using target log density:

$$\ln (p(Y | \Omega)) \propto - \left( \phi_0 + 1/2 \right) \ln (\sigma^2) - \phi_0 - 1/2 \ln (|R(\phi)|) - 1/2 w^T R(\phi)^{-1} w$$

$$+ \sum_{i=1}^{n} Y(s_i) \left( x^T(s_i) \beta + w(s_i) \right) + \sum_{i=1}^{n} \left( 1 + \exp(x^T(s_i) \beta + w(s_i)) \right)^{-1} \ln (\sigma^2) + \ln (\phi - \phi_0) + \ln (\sigma_0 - \sigma_0).$$

Binary spatial regression: forest/non-forest

We fit a generalized linear model where

$$Y(s_i) \sim Bernoulli(p(s_i)), \quad \logit(p(s_i)) = x^T(s_i) \beta + w(s_i).$$

Assume vague flat for $\beta$, a Uniform$(3/32, 3/0.5)$ prior for $\phi$, and an inverse-Gamma$(2, \cdot)$ prior for $\sigma^2$.

Parameters updated with Metropolis algorithm using target log density:

$$\ln (p(Y | \Omega)) \propto - \left( \phi_0 + 1/2 \right) \ln (\sigma^2) - \phi_0 - 1/2 \ln (|R(\phi)|) - 1/2 w^T R(\phi)^{-1} w$$

$$+ \sum_{i=1}^{n} Y(s_i) \left( x^T(s_i) \beta + w(s_i) \right) + \sum_{i=1}^{n} \left( 1 + \exp(x^T(s_i) \beta + w(s_i)) \right)^{-1} \ln (\sigma^2) + \ln (\phi - \phi_0) + \ln (\sigma_0 - \sigma_0).$$

Posterior parameter estimates

Parameter estimates (posterior medians and upper and lower 2.5 percentiles):

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimates (50%, 2.5%, 97.5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ($\beta$)</td>
<td>$-2.07 (-4.20, 1.34)$</td>
</tr>
<tr>
<td>AprilTC1 ($\alpha_0$)</td>
<td>$0.07 (0.02, 0.22)$</td>
</tr>
<tr>
<td>AprilTC2 ($\alpha_0$)</td>
<td>$0.07 (0.02, 0.22)$</td>
</tr>
<tr>
<td>AprilTC3 ($\alpha_0$)</td>
<td>$0.07 (0.02, 0.22)$</td>
</tr>
<tr>
<td>JulyTC1 ($\alpha_0$)</td>
<td>$0.07 (0.02, 0.22)$</td>
</tr>
<tr>
<td>JulyTC2 ($\alpha_0$)</td>
<td>$0.07 (0.02, 0.22)$</td>
</tr>
<tr>
<td>JulyTC3 ($\alpha_0$)</td>
<td>$0.07 (0.02, 0.22)$</td>
</tr>
<tr>
<td>OctTC1 ($\alpha_0$)</td>
<td>$0.07 (0.02, 0.22)$</td>
</tr>
<tr>
<td>OctTC2 ($\alpha_0$)</td>
<td>$0.07 (0.02, 0.22)$</td>
</tr>
<tr>
<td>OctTC3 ($\alpha_0$)</td>
<td>$0.07 (0.02, 0.22)$</td>
</tr>
</tbody>
</table>

Covariates and proximity to observed FIA plot will contribute to increase precision of prediction.
Classification of 15 $20 \times 20$ pixel areas (based on visual inspection of imagery) into non-forest ($\bullet$), moderately forest ($\circ$), and forest (no marker).