Modeling non-Gaussian spatial data

Andrew O. Finley¹ and Alan Gelfand²

¹ Department of Forestry, Michigan State University, Lansing, Michigan.
² Department of Statistical Science, Duke University, Durham, North Carolina.

September 9, 2014

Spatial Generalized Linear Models

Often data sets preclude Gaussian modeling: \( y(s) \) 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)

First stage: \( y(s_i) \) are conditionally independent given \( \beta \) and \( w(s_i) \). Here we use a canonical link function, say \( g(E[y(s_i)]) = \eta(s_i) = x(s_i)\beta + w(s_i) \).

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 \( \epsilon(s) \)

We are modeling 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 modeling is attractive for spatial explanation in the mean
- First stage spatial modeling more appropriate to encourage proximate observations to be similar.

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 \( y(s) \) is a binary variable, with
\[
y(s) = \begin{cases} 
1 & \text{if inventory plot is forested} \\
0 & \text{if inventory plot is not forested} 
\end{cases}
\]

Observed covariates include the coinciding pixel values for 3 dates of 30 × 30 m resolution Landsat imagery.

Illustration from:
Binary spatial regression: forest/non-forest

- We fit a generalized linear model where
  \[ y(s_i) \sim Bernoulli(p(s_i)), \quad \text{logit}(p(s_i)) = x(s_i)\beta + w(s_i). \]

- 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:
  \[
  \ln(p(\Omega | y)) \propto -\left(\sigma_a + 1 + \frac{n}{2}\right) \ln(\sigma^2) - \frac{\sigma_a}{\sigma^2} \ln(|R(\phi)|) - \frac{1}{2\sigma^2} w' R(\phi)^{-1} w
  + \sum_{i=1}^{n} y(s_i) \{x(s_i)' \beta + w(s_i)\} - \sum_{i=1}^{n} \ln(1 + \exp(x(s_i)' \beta + w(s_i)))
  + \ln(\sigma^2) + \ln(\phi - \phi_a) + \ln(\phi_b - \phi).
  \]

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 ( (\theta_0) )</td>
<td>82.39 (49.56, 120.46)</td>
</tr>
<tr>
<td>AprilTC1 ( (\theta_1) )</td>
<td>-0.27 (-0.45, -0.11)</td>
</tr>
<tr>
<td>AprilTC2 ( (\theta_2) )</td>
<td>0.17 (0.07, 0.29)</td>
</tr>
<tr>
<td>AprilTC3 ( (\theta_3) )</td>
<td>-0.24 (-0.43, -0.08)</td>
</tr>
<tr>
<td>JulyTC1 ( (\theta_4) )</td>
<td>-0.04 (-0.25, 0.17)</td>
</tr>
<tr>
<td>JulyTC2 ( (\theta_5) )</td>
<td>0.09 (0.01, 0.19)</td>
</tr>
<tr>
<td>JulyTC3 ( (\theta_6) )</td>
<td>0.01 (0.15, 0.16)</td>
</tr>
<tr>
<td>OctTC1 ( (\theta_7) )</td>
<td>-0.43 (-0.68, -0.22)</td>
</tr>
<tr>
<td>OctTC2 ( (\theta_8) )</td>
<td>-0.03 (0.19, 0.14)</td>
</tr>
<tr>
<td>OctTC3 ( (\theta_9) )</td>
<td>-0.26 (-0.46, -0.07)</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>1.358 (0.39, 2.42)</td>
</tr>
<tr>
<td>( \phi )</td>
<td>0.00182 (0.00065, 0.0032)</td>
</tr>
<tr>
<td>( \log(0.05)/\phi ) meters</td>
<td>1644.19 (932.33, 4606.7)</td>
</tr>
</tbody>
</table>

Covariates and proximity to observed FIA plot will contribute to increase precision of prediction.