Hierarchical Modeling for Univariate Spatial Data

Sudipto Banerjee¹, Bradley P. Carlin¹ and Alan E. Gelfand²

¹Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota, U.S.A.
²Department of Statistical Sciences, Duke University, Durham, North Carolina.

August 02, 2014

Univariate spatial models

Algorithmic Modeling

- Spatial surface observed at finite set of locations \( S = \{s_1, s_2, \ldots, s_n\} \)
- Tessellate the spatial domain (data locations as vertices)
- Fit an interpolating polynomial:
  \[ f(s) = \sum_i w_i(S; s)f(s_i) \]
  “Interpolate” by reading off \( f(s_0) \).
- Includes: triangulation, weighted averages, geographically weighted regression (GWR)
- Issues: Sensitivity to tessellations, choices of interpolators, limited inference, numerical error analysis

Simple linear model

- Response: \( Y(s) \) at location \( s \)
- Mean: \( \mu = x^T(s)\beta \)
- Error: \( \epsilon(s) \overset{iid}{\sim} N(0, \tau^2) \)

Simple linear model

- Assumptions regarding \( \epsilon(s) \):
  - \( \epsilon(s) \overset{iid}{\sim} N(0, \tau^2) \)
  - \( \epsilon(s_i) \) and \( \epsilon(s_j) \) are uncorrelated for all \( i \neq j \)
### Univariate spatial models

**Sources of variation**

- **Univariate spatial regression**
  - **Hierarchical modeling**
    - **First stage:**
      \[
      y | \beta, w, \tau^2 \sim \prod_{i=1}^{n} N(Y(s_i), x^T(s_i)\beta + w(s_i), \tau^2)
      \]
    - **Second stage:**
      \[
      w | \sigma^2, \phi \sim N(0, \sigma^2 R(\phi))
      \]
    - **Third stage:**
      - Priors on \( \Omega = (\beta, \tau^2, \sigma^2, \phi) \)
      - Marginalized likelihood:
        \[
        y | \Omega \sim N(X\beta + \tau^2 \Sigma + \tau^2 I)
        \]
    - **Note:** Spatial process parameterizes \( \Sigma \):
      \[
      y = X\beta + \epsilon, \epsilon \sim N(0, \Sigma), \Sigma = \sigma^2 R(\phi) + \tau^2 I
      \]

- **Simple linear model + random spatial effects**
  - **Response:** \( Y(s) \) at some site
  - **Mean:** \( \mu = x^T(s)\beta \)
  - **Spatial random effects:** \( w(s) \sim GP(0, \sigma^2 \rho(\phi; \|s_1 - s_2\|)) \)
  - **Non-spatial variance:** \( \epsilon(s) \sim N(0, \tau^2) \)
  - Interpretation as pure error, measurement error, replication error, microscale error

- **Bayesian Computations**
  - **Choice:** Fit \( y | \Omega \times [\Omega] \) or \( y | \beta, w, \tau^2 \times [w | \sigma^2, \phi] \times [\Omega] \).
  - **Conditional model:**
    - conjugate full conditionals for \( \beta, \sigma^2, \tau^2 \) and \( w \) – easier to program.
  - **Marginalized model:**
    - need Metropolis or Slice sampling for \( \sigma^2, \tau^2 \) and \( \phi \). Harder to program.
    - But, reduced parameter space \( \Rightarrow \) faster convergence
  - \( \sigma^2 R(\phi) + \tau^2 I \) is more stable than \( \sigma^2 R(\phi) \).
  - **But what about \( R^{-1}(\phi) \) ?? EXPENSIVE!**

**Realization of a Gaussian process:**
- Changing \( \phi \) and holding \( \sigma^2 = 1 \):
  \[
  w \sim N(0, \sigma^2 R(\phi)), \quad R(\phi) = [\rho(\phi; |s_i - s_j|)]_{i,j=1}^{n}
  \]
- **Correlation model for \( R(\phi) \):**
  - e.g., exponential decay
  \[
  \rho(\phi; t) = \exp(-\phi t) \text{ if } t > 0.
  \]
- **Effective range,**
  \[
  t_0 = \ln(0.05)/\phi \approx 3/\phi
  \]
- **Other valid models** e.g., Gaussian, Spherical, Matérn.

- **Spatial Gaussian processes (GP):**
  - Say \( w(s) \sim GP(0, \sigma^2 R(\phi)) \) and
  \[
  \text{Cov}(w(s_1), w(s_2)) = \sigma^2 \rho(\phi; \|s_1 - s_2\|)\]
  - Let \( w = [w(s_1), \ldots, w(s_n)]^T \), then
  \[
  w \sim N(0, \sigma^2 R(\phi)), \quad R(\phi) = [\rho(\phi; \|s_i - s_j\|)]_{i,j=1}^{n}
  \]

- **Univariate spatial models**
  - \( w(s) \sim N(0, \sigma^2 R(\phi)) \) provides complex spatial dependence through simple structured dependence.
  
  E.g., anisotropic Matérn correlation function:
  \[
  \rho(s_1, s_2, \phi) = (1\Gamma(\nu)^2) \left( 2\sqrt{d_{ij}} \right)^\nu (2\sqrt{d_{ij}})^\nu \]
  where
  \[
  d_{ij} = (s_i - s_j)^T \Sigma^{-1} (s_i - s_j), \quad \Sigma = \Omega \Theta \Omega^T
  \]
  Thus, \( \phi = (\nu, \psi, \Lambda) \).

**Bayesian Computations**
- **Choice:** Fit \( y | [\Omega] \times [\Omega] \) or \( y | \beta, w, \tau^2 \times [w | \sigma^2, \phi] \times [\Omega] \).
- **Conditional model:**
  - conjugate full conditionals for \( \beta, \sigma^2, \tau^2 \) and \( w \) – easier to program.
- **Marginalized model:**
  - need Metropolis or Slice sampling for \( \sigma^2, \tau^2 \) and \( \phi \). Harder to program.
  - But, reduced parameter space \( \Rightarrow \) faster convergence
  - \( \sigma^2 R(\phi) + \tau^2 I \) is more stable than \( \sigma^2 R(\phi) \).
- **But what about \( R^{-1}(\phi) \) ?? EXPENSIVE!**

**Realization of a Gaussian process:**
- Changing \( \phi \) and holding \( \sigma^2 = 1 \):
  \[
  w \sim N(0, \sigma^2 R(\phi)), \quad R(\phi) = [\rho(\phi; |s_i - s_j|)]_{i,j=1}^{n}
  \]
- **Correlation model for \( R(\phi) \):**
  - e.g., exponential decay
  \[
  \rho(\phi; t) = \exp(-\phi t) \text{ if } t > 0.
  \]
- **Effective range,**
  \[
  t_0 = \ln(0.05)/\phi \approx 3/\phi
  \]
- **Other valid models** e.g., Gaussian, Spherical, Matérn.

**Spatial Gaussian processes (GP):**
- Say \( w(s) \sim GP(0, \sigma^2 R(\phi)) \) and
  \[
  \text{Cov}(w(s_1), w(s_2)) = \sigma^2 \rho(\phi; \|s_1 - s_2\|)\]
- Let \( w = [w(s_1), \ldots, w(s_n)]^T \), then
  \[
  w \sim N(0, \sigma^2 R(\phi)), \quad R(\phi) = [\rho(\phi; \|s_i - s_j\|)]_{i,j=1}^{n}
  \]
Univariate spatial models

Where are the $w$'s?

- Interest often lies in the spatial surface $w(y)$.

- They are recovered from

$$w(y, X) = \int [w(\Omega, y, X) \times \Omega] d\Omega$$

using posterior samples:

- Obtain $\Omega^{(1)}, \ldots, \Omega^{(G)} \sim [\Omega | y, X]$.
- For each $\Omega^{(g)}$, draw $w^{(g)} \sim [w | \Omega^{(g)}, y, X]$.

**NOTE:** With Gaussian likelihoods $[w | \Omega, y, X]$ is also Gaussian. With other likelihoods this may not be a standard distribution; conditional updating scheme is preferred.

Often we need to predict $Y(s)$ at a new set of locations $\{\tilde{s}_0, \ldots, \tilde{s}_m\}$ with associated predictor matrix $\tilde{X}$.

- Sample from predictive distribution:

$$[\tilde{y}(s) | y, X] = \int [\tilde{y}(s) | y, \Omega, X, \tilde{X}] d\Omega$$

$$= \int [\tilde{y}(s) | y, \Omega, X, \tilde{X}] \times [\Omega | y, X] d\Omega,$$

$[\tilde{y}(s), \Omega, X, \tilde{X}]$ is multivariate normal. Sampling scheme:

- Obtain $\Omega^{(1)}, \ldots, \Omega^{(G)} \sim [\Omega | y, X]$.
- For each $\Omega^{(g)}$, draw $\tilde{y}^{(g)} \sim [\tilde{y} | \Omega^{(g)}, y, X]$.

Residual plot: $[w(s) | y]$
Modeling temperature: 50 locations in Colorado.

Simple spatial regression model:

$$Y(s) = X'(s) \beta + w(s) + \epsilon(s)$$

$$w(s) \sim GP(0, \sigma^2 \rho(\cdot; \phi, \nu)); \epsilon(s) \sim N(0, \tau^2)$$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>50% (2.5%, 97.5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.827 (2.131, 3.866)</td>
</tr>
<tr>
<td>Elevation</td>
<td>-0.426 (-0.527, -0.333)</td>
</tr>
<tr>
<td>Precipitation</td>
<td>0.037 (0.002, 0.072)</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.134 (0.051, 1.245)</td>
</tr>
<tr>
<td>$\phi$</td>
<td>7.39E-3 (4.71E-3, 51.21E-3)</td>
</tr>
<tr>
<td>Range</td>
<td>278.2 (38.8, 476.3)</td>
</tr>
<tr>
<td>$\tau^2$</td>
<td>0.051 (0.022, 0.092)</td>
</tr>
</tbody>
</table>

Spatial Generalized Linear Models

Often data sets preclude Gaussian modeling; in fact, $Y(s)$ 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

Spatial GLM, Diggle Tawn and Moyeed (1998)
### First stage: Y(s) are conditionally independent given β and w(s), so f(Y(s)|β, w(s), γ) equals

\[ h(y(s), γ) \exp (γ[y(s), y(\hat{s})] - ψ(y(s), \hat{s})) \]

where \( g(E(Y(s))) = η(s) = x^T(s)β + w(s) \) (canonical link function) and γ is a dispersion parameter.

### Second stage: Model w(s) as a Gaussian process:

\[ w \sim \mathcal{N}(0, σ^2 R(φ)) \]

### Third stage: Priors and hyperpriors.

- No process for Y(s), only a valid joint distribution
- Not sensible to add a pure error term e(s)

---

### Comments

- 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.