Hierarchical Modeling for Univariate Spatial Data

Geography 890, Hierarchical Bayesian Models for Environmental Spatial Data Analysis

February 15, 2011
Spatial Domain
Spatial Domain

This is a bad example because it’s not a convex hull, but just go with it ☺️.
Simple linear model

\[ Y(s) = \mu(s) + \epsilon(s), \]

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

\[ Y(s) = \mu(s) + \epsilon(s), \]

Assumptions regarding \( \epsilon(s) \):

- \( \epsilon(s) \) iid \( \sim N(0, \tau^2) \)
Simple linear model

\[ Y(\mathbf{s}) = \mu(\mathbf{s}) + \epsilon(\mathbf{s}), \]

Assumptions regarding \( \epsilon(\mathbf{s}) \):

- \( \epsilon(\mathbf{s}) \overset{iid}{\sim} N(0, \tau^2) \)
- \( \epsilon(\mathbf{s}_i) \) and \( \epsilon(\mathbf{s}_j) \) are uncorrelated for all \( i \neq j \)
Spatial Gaussian processes ($GP$):

- Say $w(s) \sim GP(0, \sigma^2 \rho(\cdot))$ and

$$\text{Cov}(w(s_1), w(s_2)) = \sigma^2 \rho(\phi; \|s_1 - s_2\|)$$
Spatial Gaussian processes ($GP$):

- Say $w(s) \sim GP(0, \sigma^2 \rho(\cdot))$ and

$$Cov(w(s_1), w(s_2)) = \sigma^2 \rho (\phi; \|s_1 - s_2\|)$$

- Let $w = [w(s_i)]_{i=1}^n$, then

$$w \sim N(0, \sigma^2 R(\phi)), \text{ where } R(\phi) = [\rho(\phi; \|s_i - s_j\|)]_{i,j=1}^n$$
Realization of a Gaussian process:

- Changing $\phi$ and holding $\sigma^2 = 1$:

$$w \sim N(0, \sigma^2 R(\phi)), \text{ where}$$

$$R(\phi) = [\rho(\phi; \|s_i - s_j\|)]_{i,j=1}^n$$
Realization of a Gaussian process:

- Changing $\phi$ and holding $\sigma^2 = 1$:
  
  $$
  w \sim N(0, \sigma^2 R(\phi)), \quad \text{where}
  
  R(\phi) = [\rho(\phi; \|s_i - s_j\|)]_{i,j=1}^n
  $$

- Correlation model for $R(\phi)$: e.g., exponential decay
  
  $$
  \rho(\phi; t) = \exp(-\phi t) \text{ if } t > 0.
  $$
Realization of a Gaussian process:

- Changing \( \phi \) and holding \( \sigma^2 = 1 \):

\[ w \sim N(0, \sigma^2 R(\phi)), \text{ where} \]
\[ R(\phi) = [\rho(\phi; \|s_i - s_j\|)]_{i,j=1}^{n} \]

- Correlation model for \( R(\phi) \):
  e.g., exponential decay

\[ \rho(\phi; t) = \exp(-\phi t) \text{ if } t > 0. \]

- Other valid models e.g., Gaussian, Spherical, Matérn.

- Effective range,
  \[ t_0 = \ln(0.05)/\phi \approx 3/\phi \]
$w \sim N(0, \sigma^2_w R(\phi))$ defines complex spatial dependence structures.

E.g., anisotropic Matérn correlation function:

$$
\rho(s_i, s_j; \phi) = \left(1/\Gamma(\nu)2^{\nu-1}\right) \left(2\sqrt{\nu d_{ij}}\right)^\nu \kappa(2\sqrt{\nu d_{ij}}),
$$

where

$$
d_{ij} = (s_i - s_j)' \Sigma^{-1} (s_i - s_j), \quad \Sigma = G(\psi) \Lambda^2 G(\psi)'.
$$

Thus, $\phi = (\nu, \psi, \Lambda)$. 

---

**Simulated**

![Simulated Data](image1.png)

**Predicted**

![Predicted Data](image2.png)
Simple linear model + random spatial effects

\[ Y(s) = \mu(s) + w(s) + \epsilon(s), \]

- **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:** \( \epsilon(s) \overset{iid}{\sim} N(0, \tau^2) \)
Hierarchical modeling

First stage:

\[ y|\beta, w, \tau^2 \sim \prod_{i=1}^{n} N(Y(s_i) | x^T(s_i)\beta + w(s_i), \tau^2) \]
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:**
  \[ w | \sigma^2, \phi \sim N(0, \sigma^2 R(\phi)) \]
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 \( \Omega = (\beta, \tau^2, \sigma^2, \phi) \)
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:
  \[ y | \Omega \sim N(X\beta, \sigma^2 R(\phi) + \tau^2 I) \]
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:**

\[ y | \Omega \sim N(X\beta, \sigma^2 R(\phi) + \tau^2 I) \]

- **Note:** Spatial process parametrizes \( \Sigma \):

\[ y = X\beta + \epsilon, \quad \epsilon \sim N(0, \Sigma), \quad \Sigma = \sigma^2 R(\phi) + \tau^2 I \]
Bayesian Computations

Choice: Fit \([y|\Omega] \times [\Omega]\) or \([y|\beta, w, \tau^2] \times [w|\sigma^2, \phi] \times [\Omega]\).
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 \(\sigma^2, \tau^2\) and \(w\) – easier to program.
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 \(\sigma^2\), \(\tau^2\) 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 \(\sigma^2 R(\phi)\).
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 \( \sigma^2, \tau^2 \) 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!
Where are the $w$’s?

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

NOTE: With Gaussian likelihoods $[w|\Omega, y, X]$ is also Gaussian. With other likelihoods this may not be easy and often the conditional updating scheme is preferred.
Where are the $w$’s?

- Interest often lies in the spatial surface $w|y$.
- They are recovered from

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

using posterior samples:
Where are the $w$'s?

- Interest often lies in the spatial surface $w|y$.
- They are recovered from

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

using posterior samples:
  - Obtain $\Omega^{(1)}, \ldots, \Omega^{(G)} \sim [\Omega|y, X]$
  - For each $\Omega^{(g)}$, draw $w^{(g)} \sim [w|\Omega^{(g)}, y, X]$
Where are the \( w \)'s?

- Interest often lies in the spatial surface \( w|y \).
- They are recovered from

\[
[w|y, X] = \int [w|\Omega, y, X] \times [\Omega|y, X] \, d\Omega
\]

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 easy and often the conditional updating scheme is preferred.
Residual plot: $[w(s) | y]$
Another look: $[w(s)|y]$
Another look: $[w(s) \mid y]$
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:

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

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

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

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

Geography 890
Prediction: Summary of $[Y(s)|y]$
Colorado data illustration

- Modeling temperature: 507 locations in Colorado.
- Simple spatial regression model:

\[ Y(s) = x^T(s)\beta + w(s) + \epsilon(s) \]

- \( w(s) \sim GP(0, \sigma^2 \rho(\cdot; \phi, \nu)) \); \( \epsilon(s) \overset{iid}{\sim} N(0, \tau^2) \)

<table>
<thead>
<tr>
<th>Parameters</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>
Temperature residual map
Elevation map
Residual map with elev. as covariate