Hierarchical Modelling for Multivariate 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, Lansiing Michigan, U.S.A.

March 3, 2010

- Point-referenced spatial data often come as multivariate measurements at each location.
- Examples:
  - Environmental monitoring: stations yield measurements on ozone, NO, CO, and PM\textsubscript{2.5}.
  - Community ecology: assemblages of plant species due to water availability, temperature, and light requirements.
  - Forestry: measurements of stand characteristics age, total biomass, and average tree diameter.
  - Atmospheric modeling: at a given site we observe surface temperature, precipitation and wind speed.
Multivariate spatial modelling

Point-referenced spatial data often come as multivariate measurements at each location.

Examples:
- Environmental monitoring: stations yield measurements on ozone, NO, CO, and PM$_{10}$.
- Community ecology: assemblages of plant species due to water availability, temperature, and light requirements.
- Forestry: measurements of stand characteristics age, total biomass, and average tree diameter.
- Atmospheric modeling: at a given site we observe surface temperature, precipitation and wind speed

We anticipate dependence between measurements

Each location contains $m$ spatial regressions

$$Y_k(s) = \mu_k(s) + w_k(s) + \epsilon_k(s), \quad k = 1, \ldots, m.$$  

Mean: $\mu(s) = [\mu_k(s)]_{k=1}^m = [\mathbf{x}^T_i(s)]_{i=1}^m$  

Cov: $\mathbf{w}(s) = [w_k(s)]_{k=1}^m \sim MVGP(0, \Gamma_w(\cdot, \cdot))$  

$$\Gamma_w(s, s') = [\text{Cov}(w_k(s), w_{k'}(s'))]_{k,k'=1}^m$$

Error: $\epsilon(s) = [\epsilon_k(s)]_{k=1}^m \sim MVN(0, \Psi)$

$\Psi$ is an $m \times m$ p.d. matrix, e.g. usually $\text{Diag}(\tau_k^2)_{k=1}^m$.

Properties:
- $\Gamma_w(s', s) = \Gamma_w(s, s')$  
- $\lim_{s \to s'} \Gamma_w(s, s')$ is p.d. and $\Gamma_w(s, s) = \text{Var}(\mathbf{w}(s))$.
- For sites in any finite collection $\mathcal{S} = \{s_1, \ldots, s_n\}$:
  $$\sum_{i=1}^n \sum_{j=1}^n u_i^T \Gamma_w(s_i, s_j) u_j \geq 0 \quad \forall u_i, u_j \in \mathbb{R}^m.$$  

Any valid $\Gamma_w$ must satisfy the above conditions.

The last property implies that $\Sigma_w$ is p.d.  
In complete generality:  
- $\Gamma_w(s, s')$ need not be symmetric.  
- $\Gamma_w(s, s')$ need not be p.d. for $s \neq s'$.  

Each location contains $m$ spatial regressions

$$Y_k(s) = \mu_k(s) + w_k(s) + \epsilon_k(s), \quad k = 1, \ldots, m.$$  

Mean: $\mu(s) = [\mu_k(s)]_{k=1}^m = [\mathbf{x}^T_i(s)]_{i=1}^m$  

Cov: $\mathbf{w}(s) = [w_k(s)]_{k=1}^m \sim MVGP(0, \Gamma_w(\cdot, \cdot))$  

$$\Gamma_w(s, s') = [\text{Cov}(w_k(s), w_{k'}(s'))]_{k,k'=1}^m$$

Error: $\epsilon(s) = [\epsilon_k(s)]_{k=1}^m \sim MVN(0, \Psi)$

$\Psi$ is an $m \times m$ p.d. matrix, e.g. usually $\text{Diag}(\tau_k^2)_{k=1}^m$.  

Properties:
- $\Gamma_w(s', s) = \Gamma_w(s, s')$  
- $\lim_{s \to s'} \Gamma_w(s, s')$ is p.d. and $\Gamma_w(s, s) = \text{Var}(\mathbf{w}(s))$.
- For sites in any finite collection $\mathcal{S} = \{s_1, \ldots, s_n\}$:
  $$\sum_{i=1}^n \sum_{j=1}^n u_i^T \Gamma_w(s_i, s_j) u_j \geq 0 \quad \forall u_i, u_j \in \mathbb{R}^m.$$  

Any valid $\Gamma_w$ must satisfy the above conditions.

The last property implies that $\Sigma_w$ is p.d.  
In complete generality:  
- $\Gamma_w(s, s')$ need not be symmetric.  
- $\Gamma_w(s, s')$ need not be p.d. for $s \neq s'$.  

We anticipate dependence between measurements

at a particular location

across locations
### Multivariate spatial modelling

**Modelling cross-covariances**

- **Moving average or kernel convolution of a process:**
  - Let $Z(s) \sim GP(0, \rho(s,s'))$. Use kernels to form:
    $$w_j(s) = \int \kappa_j(u) Z(s+u) du = \int \kappa_j(s-s') Z(s') ds'$$
  - $\Gamma_w(s-s')$ has $(i,j)$-th element:
    $$[\Gamma_w(s-s')]_{i,j} = \int \kappa_i(s-s' + u) \kappa_j(u') \rho(u \to u') du$$

- **Convolution of Covariance Functions:**
  - $\rho_1, \rho_2, \ldots, \rho_m$ are valid covariance functions. Form:
    $$[\Gamma_w(s-s')]_{i,j} = \int \rho_i(s-s'+u) \rho_j(u) du$$

**Constructive approach**

- **When $s=s'$, $\Gamma_v(s, s) = I_m$, so:**
  - $\Gamma_w(s, s) = \Lambda(s) A^T(s)$

- $\Lambda(s)$ identifies with any square-root of $\Gamma_w(s, s)$. Can be taken as lower-triangular (Cholesky).

- $\Lambda(s)$ is unknown!
  - Should we first model $\Lambda(s)$ to obtain $\Gamma_w(s, s)$?
  - Or should we model $\Gamma_w(s, s)$ first and derive $\Lambda(s)$?
  - $\Lambda(s)$ is completely determined from within-site associations.

**Second stage:**

- Let $\Psi = [\Psi(s)]_{i,j}^{n\times n}$ and $w = [W(s)]_{i,j}^{n\times 1}$.

**First stage:**

$$Y | \beta, w, \Psi \sim \prod_{i=1}^n MVN(Y(s_i) | X(s_i)^T \beta + w(s_i), \Psi)$$

**Second stage:**

$$w | \theta \sim MVN(0, \Sigma_w(\theta))$$

where $\Sigma_w(\theta) = [\Gamma_w(s_j, s_j; \theta)]_{i,j=1}^n$.
Let \( y = [Y(s_i)]_{i=1}^n \) and \( w = [W(s_i)]_{i=1}^n \).

First stage:

\[
y | \beta, w, \Psi \sim \prod_{i=1}^n MVN(Y(s_i) | X(s_i)\beta + w(s_i), \Psi)
\]

Second stage:

\[
w | \theta \sim MVN(0, \Sigma_w(\Phi))
\]

where \( \Sigma_w(\Phi) = [\Gamma_W(s_i, s_j; \Phi)]_{i,j=1}^n \).

Third stage: Priors on \( \Omega = (\beta, \Psi, \Phi) \).

Choice: Fit as \( y | \Omega \times \Omega \) or as \( y | \beta, w, \Psi \times w | \Phi \times \Omega \).

Conditional model:

- Conjugate distributions are available for \( \Psi \) and other variance parameters. Easy to program.

Marginalized model:

- Need Metropolis or Slice sampling for most variance-covariance parameters. Harder to program.
- But reduced parameter space (no \( w \)’s) results in faster convergence
- \( \Sigma_w(\Phi) + I \otimes \Psi \) is more stable than \( \Sigma_w(\Phi) \).

Choice: Fit as \( y | \Omega \times \Omega \) or as \( y | \beta, w, \Psi \times w | \Phi \times \Omega \).

Conditional model:

- Conjugate distributions are available for \( \Psi \) and other variance parameters. Easy to program.

Marginalized model:

- Need Metropolis or Slice sampling for most variance-covariance parameters. Harder to program.
- But reduced parameter space (no \( w \)’s) results in faster convergence
- \( \Sigma_w(\Phi) + I \otimes \Psi \) is more stable than \( \Sigma_w(\Phi) \).
- But what about \( \Sigma_w^{-1}(\Phi) \)?? Matrix inversion is EXPENSIVE \( O(n^3) \).
Recovering 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 \( \Omega^{(1)}, \ldots, \Omega^{(S)} \sim [\Omega y, X] \)
- For each \( \Omega^{(s)} \), draw \( w^{(s)} \sim [w[\Omega^{(s)}, y, X] \)

**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.

Often we need to predict \( Y(s) \) at a new set of locations \( \{s_0, \ldots, s_m\} \) with associated predictor matrix \( \hat{X} \).

- Sample from predictive distribution:

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

\]

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

- Obtain \( \Omega^{(1)}, \ldots, \Omega^{(S)} \sim [\Omega y, X] \)
- For each \( \Omega^{(s)} \), draw \( y^{(s)} \sim [y|\Omega^{(s)}, y, X] \).
Study objectives:
- Evaluate methods for multi-source forest attribute mapping
- Find the "best" model, given the data
- Produce maps of biomass and uncertainty, by tree species

Illustration from:

Slight digression – why we fit a model:
- Association between response and covariates, \( \beta \), (e.g., ecological interpretation)
- Residual spatial and/or non-spatial associations and patterns (i.e., given covariates)
- Subsequent prediction

Study objectives:
- Evaluate methods for multi-source forest attribute mapping
- Find the “best” model, given the data
- Produce maps of biomass and uncertainty, by tree species

Study area:
- USDA FS Bartlett Experimental Forest (BEF), NH
  - 1,053 ha heavily forested
  - Major tree species: American beech (BE), eastern hemlock (EH), red maple (RM), sugar maple (SM), and yellow birch (YB)

Response variables:
- Metric tons of total tree biomass per ha
- Measured on 437 \( \frac{1}{10} \) ha plots
- Models fit using random subset of 218 plots
- Prediction at remaining 219 plots
Candidate models

Each model includes 55 covariates and 5 intercepts, therefore, $X^T$ is $1090 \times 60$.

Different specifications of variance structures:

1. Non-spatial multivariate $\text{Diag}(\Psi) = \tau^2$
2. $\text{Diag}(K)$, same $\phi$, $\text{Diag}(\Psi)$
3. $K$, same $\phi$, $\text{Diag}(\Psi)$

Illustration Bartlett Experimental Forest

Covariates

- DEM derived elevation and slope
- Spring, Summer, Fall Landsat ETM+ Tasseled Cap features (brightness, greenness, wetness)

Illustration Bartlett Experimental Forest

Covariates

- DEM derived elevation and slope
- Spring, Summer, Fall Landsat ETM+ Tasseled Cap features (brightness, greenness, wetness)
Candidate models

Each model includes 55 covariates and 5 intercepts, therefore, \( X^T \) is 1090 \times 60.

Different specifications of variance structures:
- **Non-spatial multivariate** \( \text{Diag}(\Psi) = \tau^2 \)
- \( \text{Diag}(K) \), **same** \( \phi \), \( \text{Diag}(\Psi) \)
- \( K \), **same** \( \phi \), \( \text{Diag}(\Psi) \)
- \( \text{Diag}(K) \), **different** \( \phi \), \( \text{Diag}(\Psi) \)
- \( K \), **different** \( \phi \), \( \text{Diag}(\Psi) \)

Model comparison

Deviance Information Criterion (DIC):

\[
D(\Omega) = -2 \log L(\text{Data} | \Omega)
\]

\[
\bar{D}(\Omega) = E_{\Omega}[D(\Omega)]
\]

\[
pD = D(\Omega) - \bar{D}(\Omega); \quad \bar{\Omega} = E_{\Omega}[\Omega(\Omega)]
\]

\[
\text{DIC} = \bar{D}(\Omega) + pD.
\]

Lower DIC is better.

Selected model

- Model 5: \( K \), **different** \( \phi \), \( \text{Diag}(\Psi) \)
- Parameters: \( K = 15 \), \( \phi = 5 \), \( \text{Diag}(\Psi) = 5 \)

Focus on spatial cross-covariance matrix \( K \) (for brevity).

Posterior inference of \( \text{cor}(K) \), e.g., 50 (2.5, 97.5) percentiles:

<table>
<thead>
<tr>
<th>BE</th>
<th>EH</th>
</tr>
</thead>
<tbody>
<tr>
<td>BE</td>
<td>0.16 (-0.13, 0.21)</td>
</tr>
<tr>
<td>EH</td>
<td>-0.20 (-0.23, -0.15)</td>
</tr>
<tr>
<td>RM</td>
<td>-0.20 (-0.22, -0.17)</td>
</tr>
<tr>
<td>SM</td>
<td>0.07 (0.04, 0.08)</td>
</tr>
</tbody>
</table>

These relationships expressed in mapped random spatial effects, \( w \).
Selected model

- Model 5: $K$, different $\phi$, $\text{Diag}(\Psi)$
- Parameters: $K = 15$, $\phi = 5$, $\text{Diag}(\Psi) = 5$

Focus on spatial cross-covariance matrix $K$ (for brevity).

Posterior inference of $\text{cor}(K)$, e.g., 50 (2.5, 97.5) percentiles:

<table>
<thead>
<tr>
<th></th>
<th>BE</th>
<th>EH</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>BE</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EH</td>
<td>0.16 (0.13, 0.21)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>RM</td>
<td>-0.20 (-0.23, -0.15)</td>
<td>0.45 (0.26, 0.66)</td>
<td>...</td>
</tr>
<tr>
<td>SM</td>
<td>-0.20 (-0.22, -0.17)</td>
<td>-0.12 (-0.16, -0.09)</td>
<td>...</td>
</tr>
<tr>
<td>YB</td>
<td>0.07 (0.04, 0.08)</td>
<td>0.22 (0.20, 0.25)</td>
<td>...</td>
</tr>
</tbody>
</table>

These relationships expressed in mapped random spatial effects, $w$.

Summary

Proposed Bayesian hierarchical spatial methodology:
- Partition sources of uncertainty
  - Provides hypothesis testing
  - Reveal spatial patterns and missing covariates
Summary

Proposed Bayesian hierarchical spatial methodology:
- Partition sources of uncertainty
  - Provides hypothesis testing
  - Reveal spatial patterns and missing covariates
- Allow flexible inference
  - Access parameters’ posterior distribution
  - Access posterior predictive distribution
- Provide consistent prediction of multiple variables
  - Maintains spatial and non-spatial association

Extendable model template:
- Cluster plot sample design – multiresolution models
- Non-continuous response – general linear models
- Obs. over time and space – spatiotemporal models