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, Lansing Michigan, U.S.A.

October 15, 2012
Point-referenced spatial data often come as multivariate measurements at each location.
Point-referenced spatial data often come as \textit{multivariate measurements} at each location.

\textbf{Examples:}

- Environmental monitoring: stations yield measurements on ozone, NO, CO, and PM$_{2.5}$.
Point-referenced spatial data often come as multivariate measurements at each location.

Examples:

- **Environmental monitoring**: stations yield measurements on ozone, NO, CO, and PM$_{2.5}$.
- **Community ecology**: assemblages of plant species due to water availability, temperature, and light requirements.
Point-referenced spatial data often come as multivariate measurements at each location.

Examples:

- **Environmental monitoring**: stations yield measurements on ozone, NO, CO, and PM$_{2.5}$.
- **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.
Point-referenced spatial data often come as multivariate measurements at each location.

Examples:

- **Environmental monitoring**: stations yield measurements on ozone, NO, CO, and PM$_{2.5}$.
- **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.
Point-referenced spatial data often come as multivariate measurements at each location.

Examples:

- **Environmental monitoring**: stations yield measurements on ozone, NO, CO, and PM$_{2.5}$.
- **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.

We anticipate dependence between measurements.
Point-referenced spatial data often come as multivariate measurements at each location.

Examples:

- **Environmental monitoring**: stations yield measurements on ozone, NO, CO, and PM$_{2.5}$.
- **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.

We anticipate dependence between measurements at a particular location.
Point-referenced spatial data often come as multivariate measurements at each location.

Examples:

- Environmental monitoring: stations yield measurements on ozone, NO, CO, and PM$_{2.5}$.
- 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.

We anticipate dependence between measurements

- at a particular location
- across locations
Each location contains $m$ spatial regressions

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

Mean: $\mu(s) = [\mu_k(s)]_{k=1}^m = [x_k^T(s)\beta_k]_{k=1}^m$

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

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

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 $Diag(\tau_k^2)_{k=1}^m$. 

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$\Psi$ is an $m \times m$ p.d. matrix, e.g. usually $Diag(\tau_k^2)_{k=1}^m$.
- \( w(s) \sim MVGP(0, \Gamma_w(\cdot)) \) with

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

- Example: with \( m = 2 \)

\[
\Gamma_w(s, s') = \begin{pmatrix}
Cov(w_1(s), w_1(s')) & Cov(w_1(s), w_2(s')) \\
Cov(w_2(s), w_1(s')) & Cov(w_2(s), w_2(s'))
\end{pmatrix}
\]

- For finite set of locations \( \mathcal{S} = \{s_1, \ldots, s_n\} \):

\[
Var \left( [w(s_i)]_{i=1}^n \right) = \Sigma_w = [\Gamma_w(s_i, s_j)]_{i,j=1}^n
\]
Properties:

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

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 \( s \neq s' \).
Moving average or kernel convolution of a process:

Let $Z(s) \sim GP(0, \rho(\cdot))$. 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:

$$[\Gamma_w(s - s')]_{i,j} = \int \int \kappa_i(s - s' + u)\kappa_j(u')\rho(u - u')du du'$$

Convolution of Covariance Functions:

$\rho_1, \rho_2, \ldots \rho_m$ are valid covariance functions. Form:

$$[\Gamma_w(s - s')]_{i,j} = \int \rho_i(s - s' - t)\rho_j(t)dt.$$
Constructive approach

- Let \( v_k(s) \sim GP(0, \rho_k(s, s')) \), for \( k = 1, \ldots, m \) be \( m \) independent GP’s with unit variance.
- Form the simple multivariate process \( \mathbf{v}(s) = [v_k(s)]_{k=1}^m \):
  \[
  \mathbf{v}(s) \sim MVGP(\mathbf{0}, \Gamma_v(\cdot, \cdot))
  \]
  with \( \Gamma_v(s, s') = \text{Diag}(\rho_k(s, s'))_{k=1}^m \).
- Assume \( \mathbf{w}(s) = A(s)\mathbf{v}(s) \) arises as a space-varying linear transformation of \( \mathbf{v}(s) \). Then:
  \[
  \Gamma_w(s, s') = A(s)\Gamma_v(s, s')A^T(s')
  \]
is a valid cross-covariance function.
Constructive approach, contd.

- When $\mathbf{s} = \mathbf{s}'$, $\Gamma_v(\mathbf{s}, \mathbf{s}) = I_m$, so:
  \[ \Gamma_w(\mathbf{s}, \mathbf{s}) = A(\mathbf{s})A^T(\mathbf{s}) \]

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

- $A(\mathbf{s})$ is unknown!
  - Should we first model $A(\mathbf{s})$ to obtain $\Gamma_w(\mathbf{s}, \mathbf{s})$?
  - Or should we model $\Gamma_w(\mathbf{s}, \mathbf{s}')$ first and derive $A(\mathbf{s})$?
  - $A(\mathbf{s})$ is completely determined from within-site associations.
Constructive approach, contd.

- If $A(s) = A$:  
  - $w(s)$ is stationary when $v(s)$ is.
  - $\Gamma_w(s, s')$ is symmetric.
  - $\Gamma_v(s, s') = \rho(s, s')I_m \Rightarrow \Gamma_w = \rho(s, s')AA^T$

- Last specification is called intrinsic and leads to separable models:
  $\Sigma_w = H(\phi) \otimes \Lambda; \Lambda = AA^T$
Let $y = [Y(s_i)]_{i=1}^n$ and $w = [W(s_i)]_{i=1}^n$.

First stage:

$$y \mid \beta, w, \Psi \sim \prod_{i=1}^n MVN \left( Y(s_i) \mid X(s_i)^T \beta + w(s_i), \Psi \right)$$
Let \( y = [Y(s_i)]_{i=1}^{n} \) and \( w = [W(s_i)]_{i=1}^{n} \).

**First stage:**

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

**Second stage:**

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

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

**First stage:**

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

**Second stage:**

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

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

**Third stage:** Priors on \( \Omega = (\beta, \Psi, \Phi) \).
Let $y = \left[ Y(s_i) \right]_{i=1}^{n}$ and $w = \left[ W(s_i) \right]_{i=1}^{n}$.

First stage:

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

Second stage:

$$w \mid \theta \sim MVN(0, \Sigma_w(\Phi))$$

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

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

Marginalized likelihood:

$$y \mid \beta, \theta, \Psi \sim MVN \left( X\beta, \Sigma_w(\Phi) + I \otimes \Psi \right)$$
Choice: Fit as $\mathbf{y} | \Omega \times [\Omega]$ or as $\mathbf{y} | \beta, \mathbf{w}, \Psi \times [\mathbf{w} | \Phi] \times [\Omega]$. 
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.
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 $O(n^3)$. 

11
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 $w|y$. 

NOTE: With Gaussian likelihoods $w|Ω,y,X$ is also Gaussian. With other likelihoods this may not be easy and often the conditional updating scheme is preferred.
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:
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 $\Omega^{(1)}, \ldots, \Omega^{(G)} \sim [\Omega|y, X]$
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 \( \Omega^{(1)}, \ldots, \Omega^{(G)} \sim [\Omega | y, X] \)
- For each \( \Omega^{(g)} \), draw \( w^{(g)} \sim [w | \Omega^{(g)}, y, X] \)
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 \( \Omega^{(1)}, \ldots, \Omega^{(G)} \sim [\Omega|y, X] \)
- 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.
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 $\tilde{y}^{(g)} \sim [\tilde{y}|y, \Omega^{(g)}, X, \tilde{X}]$. 
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 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)
Bartlett Experimental Forest

Image provided by www.fs.fed.us/ne/durham/4155/bartlett
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
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 $1090 \times 60$. 
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 $Diag(\Psi) = \tau^2$
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 $\phi$, $\text{Diag}(\Psi)$
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 $\phi$, $\text{Diag}(\Psi)$
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 $\phi$, $\text{Diag}(\Psi)$
4. $\text{Diag}(K)$, different $\phi$, $\text{Diag}(\Psi)$

Illustration  Bartlett Experimental Forest

UNL Department of Statistics Spatio-temporal Workshop
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 $\phi$, $\text{Diag}(\Psi)$

4. $\text{Diag}(K)$, different $\phi$, $\text{Diag}(\Psi)$

5. $K$, different $\phi$, $\text{Diag}(\Psi)$
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 $Diag(\Psi) = \tau^2$

2. $Diag(K)$, same $\phi$, $Diag(\Psi)$

3. $K$, same $\phi$, $Diag(\Psi)$

4. $Diag(K)$, different $\phi$, $Diag(\Psi)$

5. $K$, different $\phi$, $Diag(\Psi)$

6. $K$, different $\phi$, $\Psi$
Model comparison

Deviance Information Criterion (DIC):

\[
D(\Omega) = -2 \log L(Data \mid \Omega)
\]

\[
\overline{D(\Omega)} = E_{\Omega \mid Y}[D(\Omega)]
\]

\[
p_D = \overline{D(\Omega)} - D(\bar{\Omega}); \quad \bar{\Omega} = E_{\Omega \mid Y}[\Omega]
\]

\[
DIC = \overline{D(\Omega)} + p_D.
\]

Lower DIC is better.
Model comparison

Deviance Information Criterion (DIC):

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

\[ \overline{D(\Omega)} = E_{\Omega | Y}[D(\Omega)] \]

\[ p_D = \overline{D(\Omega)} - D(\bar{\Omega}); \quad \bar{\Omega} = E_{\Omega | Y}[\bar{\Omega}] \]

\[ DIC = \overline{D(\Omega)} + p_D. \]

Lower DIC is better.

<table>
<thead>
<tr>
<th>Model</th>
<th>( p_D )</th>
<th>DIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>35</td>
<td>8559</td>
</tr>
<tr>
<td>2</td>
<td>35</td>
<td>8543</td>
</tr>
<tr>
<td>3</td>
<td>35</td>
<td>8521</td>
</tr>
<tr>
<td>4</td>
<td>33</td>
<td>8536</td>
</tr>
<tr>
<td>5</td>
<td>34</td>
<td>8505</td>
</tr>
<tr>
<td>6</td>
<td>38</td>
<td>8507</td>
</tr>
</tbody>
</table>
Selected model

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

Illustration  Bartlett Experimental Forest
\[ E[\mathbf{w} | \text{Data}] \]
\[ E[\mathbf{w} | \text{Data}] \]

\[ E[\mathbf{w}^* | \text{Data}] \]
\[ E[Y^* \mid Data] \]

Illustration

Validation plots

BE

EH

RM

SM

YB
$$E[Y^* | Data]$$

$$P(2.5 < Y^* < 97.5 | Data)$$
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
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
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