Hierarchical Modeling for Large Spatial Datasets

Andrew O. Finley

Department of Forestry & Department of Geography, Michigan State University, Lansing Michigan, U.S.A.

October 30, 2013
The Big $n$ issue

Univariate spatial regression

$$y = X\beta + w + \epsilon,$$

- Estimation involves $$(\sigma^2 R(\phi) + \tau^2 I)^{-1},$$ which is $n \times n$.
- Matrix computations occur in each MCMC iteration.
- Known as the “Big-N problem” in geostatistics.
- Approach: Use a model $y = X\beta + Zw^* + \epsilon$. But what $Z$?
Consider “knots” $S^* = \{s_1^*, \ldots, s_n^*\}$ with $n^* << n$.

Let $w^* = \{w(s_i^*)\}_{i=1}^{n^*}$

$Z(\theta) = \{\text{cov}(w(s_i), w(s_j^*))\}' \{\text{var}(w^*)\}^{-1}$ is $n \times n^*$.

Predictive process regression model

$$y = X\beta + Z(\theta)w^* + \epsilon,$$

Fitting requires only $n^* \times n^*$ matrix computations ($n^* << n$).

Key attraction: The above arises as a process model:

$\tilde{w}(s) \sim GP(0, \sigma_w^2 \tilde{\rho}(\cdot; \phi))$ instead of $w(s)$.

$\tilde{\rho}(s_1, s_2; \phi) = \text{cov}(w(s_1), w^*) \text{var}(w^*)^{-1} \text{cov}(w^*, w(s_2))$
Knots: A “Knotty” problem??

- Knot selection: Regular grid? More knots near locations we have sampled more?
- Formal spatial design paradigm: maximize information metrics (Zhu and Stein, 2006; Diggle & Lophaven, 2006)
- Geometric considerations: space-filling designs (Royle & Nychka, 1998); various clustering algorithms
- Compare performance of estimation of range and smoothness by varying knot size.
- Stein (2007, 2008): method may not work for fine-scale spatial data
- Still a popular choice – seamlessly adapts to multivariate and spatiotemporal settings.
Big N problem

Selection of knots

\[ \tau^2 \]

0 50 100 150 200
0 5 10 15 20 25

knots

tau^2

0 50 100 150 200

5

LWF, ZWFH, DVFFA, & DR IBS: Biometry workshop
A rectified predictive process is defined as

$$\tilde{w}_\varepsilon(s) = \tilde{w}(s) + \tilde{\varepsilon}(s),$$

where

$$\tilde{\varepsilon}(s) \overset{\text{indep}}{\sim} N(0, \sigma_w^2 (1 - r(s, \phi)' R^{*-1}(\phi) r(s, \phi))).$$

Maximum likelihood estimates of $\tau^2$:

<table>
<thead>
<tr>
<th># of Knots</th>
<th>Predictive Process</th>
<th>Rectified Predictive Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>1.56941</td>
<td>1.00786</td>
</tr>
<tr>
<td>36</td>
<td>1.65688</td>
<td>1.15386</td>
</tr>
<tr>
<td>64</td>
<td>1.45169</td>
<td>1.08358</td>
</tr>
<tr>
<td>100</td>
<td>1.37916</td>
<td>1.09657</td>
</tr>
<tr>
<td>225</td>
<td>1.27391</td>
<td>1.08985</td>
</tr>
<tr>
<td>400</td>
<td>1.22429</td>
<td>1.09489</td>
</tr>
<tr>
<td>625</td>
<td>1.21127</td>
<td>1.09998</td>
</tr>
<tr>
<td>exact</td>
<td>1.14414</td>
<td>1.14414</td>
</tr>
</tbody>
</table>
For more information, see, e.g.,


Illustration from:

Univariate random effects models


*Quantitative genetics*: studies the inheritance of polygenic traits, focusing upon estimation of additive genetic variance, $\sigma_a^2$, and the heritability $h^2 = \sigma_a^2/\sigma_{Tot}^2$, where the $\sigma_{Tot}^2$ represents the total genetic and unexplained variation.

A high heritability, $h^2$, should result in a larger selection response (i.e., a higher probability for genetic gain in future generations).
Observed trees

Data overview:
- established in 1971 (by Skogforsk)
- partial diallel design of 52 parent trees
- 8,160 planted randomly on 2.2m squares
- 1997 reinventory of 4,970 surviving trees, height, DBH, branch angle, etc.
Genetic effects model:

\[ y_i = x'_i \beta + a_i + d_i + \epsilon_i, \]

- \( a = [a_i]_{i=1}^n \sim MVN(0, \sigma^2_a A) \)
- \( d = [d_i]_{i=1}^n \sim MVN(0, \sigma^2_d D) \)
- \( \epsilon = [\epsilon_i]_{i=1}^n \sim N(0, \tau^2 I_n) \)

\( A \) and \( D \) are fixed relationship matrices (See e.g., Henderson, 1985; Lynch and Walsh, 1998)

Note, genetic variance is further partitioned into additive and the non-additive dominance component \( \sigma^2_d \)
Genetic effects model:

\[ y_i = x_i' \beta + a_i + d_i + \epsilon_i, \]

- Common feature is systematic heterogeneity among observational units (i.e., violation of \( \epsilon \sim N(0, \tau^2 I_n) \))
- Spatial heterogeneity arises from:
  - soil characteristics
  - micro-climates
  - light availability
- Residual correlation among units as a function of distance and/or direction = erroneous parameter estimates (e.g., biased \( h^2 \))
Genetic model results

Parameter credible intervals, 50% (2.5%, 97.5%) for the non-spatial models Scots pine trial.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>72.53 (69.66, 75.08)</td>
<td>72.27 (70.04, 74.57)</td>
</tr>
<tr>
<td>$\sigma_a^2$</td>
<td>31.94 (18.30, 49.85)</td>
<td>25.23 (14.12, 43.96)</td>
</tr>
<tr>
<td>$\sigma_d^2$</td>
<td>–</td>
<td>22.37 (11.24, 40.11)</td>
</tr>
<tr>
<td>$\tau^2$</td>
<td>133.60 (121.18, 144.70)</td>
<td>116.14 (100.51, 127.76)</td>
</tr>
<tr>
<td>$h^2$</td>
<td>0.19 (0.12, 0.28)</td>
<td>0.15 (0.09, 0.26)</td>
</tr>
</tbody>
</table>
Genetic model results, cont’d.

So, \( \epsilon \sim N(0, \tau^2 I_n) \). Consider a spatial model.
Previous approaches to accommodating residual spatial dependence:

- **Manipulate the mean function**
  - constructing covariates using residuals from neighboring units (see e.g., Wilkinson et al., 1983; Besag and Kempton, 1986; Williams, 1986)

- **Geostatistical**
  - spatial process formed $AR(1)_{col} \otimes AR(1)_{row}$ (Martin, 1990; Cullis et al., 1998)
  - classical geostatistical method (Zimmerman and Harville, 1991)

All are computationally feasible, but **ad hoc and/or restrictive** from a modeling perspective.
Spatial model for genetic trials:

\[ y(s_i) = \mathbf{x}'(s_i)\mathbf{\beta} + a_i + d_i + w(s_i) + \epsilon_i, \]

- \( \mathbf{a} = [a_i]_{i=1}^{n} \sim \text{MVN}(\mathbf{0}, \sigma_a^2 \mathbf{A}) \)
- \( \mathbf{d} = [d_i]_{i=1}^{n} \sim \text{MVN}(\mathbf{0}, \sigma_d^2 \mathbf{D}) \)
- \( \mathbf{w} = [w(s_i)]_{i=1}^{n} \sim \text{MVN}(\mathbf{0}, \sigma_w^2 \mathbf{C}(\theta)) \)
- \( \mathbf{\epsilon} = [\epsilon_i]_{i=1}^{n} \sim \text{N}(\mathbf{0}, \tau^2 \mathbf{I}_n) \)

Tools used to estimate parameters:
- Markov chain Monte Carlo (MCMC) - iterative
  - Gibbs sampler (\( \mathbf{\beta}, \mathbf{a}, \mathbf{d}, \mathbf{w} \))
  - Metropolis-Hastings and Slice samplers (\( \theta \))

Here MCMC is computationally infeasible because of Big-N!
Trick to sample genetic effects:

Gibbs draw for random effects, e.g., \( \mathbf{a}_{|\cdot} \sim \text{MVN}(\mu_{a_{|\cdot}}, \Sigma_{a_{|\cdot}}) \), where calculating \( \Sigma_{a_{|\cdot}} = \left[ \frac{1}{\sigma^2_a} \mathbf{A}^{-1} + \frac{1}{\tau^2} \mathbf{I}_n \right]^{-1} \) is computationally expensive!

However \( \mathbf{A} \) and \( \mathbf{D} \) are known, so use initial spectral decomposition i.e., \( \mathbf{A}^{-1} = \mathbf{P}^T \mathbf{\Lambda}^{-1} \mathbf{P} \).

Thus, \( \Sigma_{a_{|\cdot}} = \mathbf{P}^T \left( \frac{1}{\sigma^2_a} \mathbf{\Lambda}^{-1} + \frac{1}{\tau^2} \mathbf{I} \right)^{-1} \mathbf{P} \) to achieve computational benefits.
Unfortunately, this *trick* does not work for \( w \). Rather, we proposed the knot-based *predictive process*.

**Corresponding predictive process model:**

\[
y(s_i) = x'(s_i)\beta + a_i + d_i + \tilde{w}(s_i) + \epsilon_i,
\]

- \( \tilde{w}(s_i) = c(s_i; \theta)^T C(\theta)^{-1}(\theta) w^* \)

where, \( w^* = [w(s^*_i)]_{i=1}^m \sim MVN(0, C^*(\theta)) \) and \( C^*(\theta) = [C(s^*_i, s^*_j; \theta)]_{i,j=1}^m \)

![Illustration](image-url)
\( \tilde{w} \) can accommodate complex spatial dependence structures. E.g., anisotropic Matérn correlation function:

\[
\rho(s_i, s_j; \theta) = \left( \frac{1}{\Gamma(\nu)} 2^{\nu - 1} \right) \left( 2 \sqrt{\nu d_{ij}} \right)^\nu \kappa_\nu \left( 2 \sqrt{\nu d_{ij}} \right),
\]

where

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

Thus, \( \theta = (\nu, \psi, \Lambda) \).
Genetic + spatial effects models

- Candidate spatial models (i.e., specifications of $C^*(\theta)$):
  1. $AR(1)_{col} \otimes AR(1)_{row}$
  2. isotropic Matérn
  3. anisotropic Matérn

- Each model evaluated using 64, 144, and 256 knot grids.

- Model choice using Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002)
Table: Model comparisons using the DIC criterion for the Scots pine dataset.

<table>
<thead>
<tr>
<th>Model</th>
<th>$p_D$</th>
<th>DIC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-spatial</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add.</td>
<td>306.40</td>
<td>15,618.09</td>
</tr>
<tr>
<td>Add. Dom.</td>
<td>555.92</td>
<td>15,547.85</td>
</tr>
<tr>
<td><strong>Spatial Isotropic</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64 Knots</td>
<td>639.77</td>
<td>14,877.51</td>
</tr>
<tr>
<td>144 Knots</td>
<td>739.61</td>
<td>14,814.89</td>
</tr>
<tr>
<td>256 Knots</td>
<td>802.29</td>
<td>14,771.64</td>
</tr>
<tr>
<td><strong>Spatial Anisotropic</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64 Knots</td>
<td>678.82</td>
<td>14,884.13</td>
</tr>
<tr>
<td>144 Knots</td>
<td>748.89</td>
<td>14,823.90</td>
</tr>
<tr>
<td>256 Knots</td>
<td>806.46</td>
<td>14,781.53</td>
</tr>
</tbody>
</table>
Genetic + spatial effects models results

Parameter credible intervals, 50% (2.5%, 97.5%) for the isotropic Matérn and 64 and 256 knots Scots pine trial.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>64 Knots</th>
<th>256 Knots</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta )</td>
<td>72.53 (69.00, 76.05)</td>
<td>74.21 (69.66, 79.66)</td>
</tr>
<tr>
<td>( \sigma_a^2 )</td>
<td>26.87 (17.14, 41.82)</td>
<td>33.03 (18.19, 53.69)</td>
</tr>
<tr>
<td>( \sigma_d^2 )</td>
<td>11.69 (6.00, 34.27)</td>
<td>13.96 (7.65, 27.05)</td>
</tr>
<tr>
<td>( \sigma_w^2 )</td>
<td>41.84 (23.71, 73.34)</td>
<td>50.36 (30.24, 88.10)</td>
</tr>
<tr>
<td>( \tau^2 )</td>
<td>89.55 (72.11, 99.65)</td>
<td>80.75 (67.90, 96.16)</td>
</tr>
<tr>
<td>( \nu )</td>
<td>0.83 (0.31, 1.46)</td>
<td>0.47 (0.26, 1.28)</td>
</tr>
<tr>
<td>( \phi )</td>
<td>0.05 (0.02, 0.09)</td>
<td>0.04 (0.02, 0.09)</td>
</tr>
<tr>
<td>Eff. Range</td>
<td>71.00 (44.66, 127.93)</td>
<td>74.59 (45.22, 129.83)</td>
</tr>
<tr>
<td>( h^2 )</td>
<td>0.21 (0.13, 0.31)</td>
<td>0.25 (0.15, 0.39)</td>
</tr>
</tbody>
</table>

- Decrease in \( \tau^2 \) due to removal of spatial variation, results in increase in \( h^2 \) (i.e., \( \sim 0.25 \) vs. \( \sim 0.15 \) with confounding).
Genetic + spatial effects models results, cont’d.

**Predictive process** – balance model richness with computational feasibility (e.g., $4,970 \times 4,970$ vs. $64 \times 64$).
Summary

Challenge - to meet modeling needs:

- ensure computationally feasible
  - reduce algorithmic complexity = cheap tricks (e.g., spectral decomp. of $A$ prior to MCMC)
  - reduce dimensionality = predictive process

- maintain richness and flexibility
  - focus on the model not how to estimate the parameters = embrace new tools (MCMC) for estimating highly flexible hierarchical models

- truly acknowledge sources of uncertainty
  - propagate uncertainty through hierarchical structures (e.g., recognize uncertainty in $C(\theta)$)