Modeling Large Spatial Datasets

Sudipto Banerjee\textsuperscript{1}, Bradley P. Carlin\textsuperscript{1} and Alan E. Gelfand\textsuperscript{2}

\textsuperscript{1}Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota, U.S.A.

\textsuperscript{2}Department of Statistical Sciences, Duke University, Durham, North Carolina.

August 02, 2014
Hierarchical Spatial model

\[ p(\theta) \times p(\Psi) \times N(\beta \mid \mu_\beta, \Sigma_\beta) \times N(w \mid 0, C_w(\theta)) \]
\[ \times \prod_{i=1}^{n} N_{m}(y(s_i) \mid X(s_i)^\top \beta + w(s_i), D(\Psi)) \]

- regression slopes
- spatial random effects from Gaussian process
- nonspatial variability (nugget)
- spatial process parameters (spatial variance, range, smoothness) and
Computational issues.

- We need to evaluate

\[-\frac{1}{2} \log \det (C_w(\theta)) - \frac{1}{2} w^\top C_w(\theta)^{-1} w\]

- What if \( n \) is LARGE? How do we tackle \( C_w(\theta)^{-1} \) and \( \det(C(\theta)) \)?
Approaches to dimension reduction:

- Covariance tapering (Furrer et al. 2006; Zhang and Du, 2007; Du et al. 2009; Kaufman et al., 2009)

- Spectral domain: (Fuentes 2007; Paciorek, 2007)

- Approximate using GMRFs: \textsc{INLA} (Rue et al. 2009; Lindgren et al., 2011)

- Nearest-neighbor models (processes) (Vecchia 1988; Stein et al. 2004; Datta et al., 2014)

- Low-rank approaches (Wahba, 1990; Higdon, 2002; Lin et al., 2000; Kamman & Wand, 2003; Paciorek, 2007; Rasmussen & Williams, 2006; Stein 2007, 2008; Cressie & Johannesson, 2008; Banerjee et al., 2008; 2010; Sang et al., 2011)

\[ S^* = \{s_1^*, s_2^*, \ldots, s_n^*\} \]: a set of “knots”.

\[ w(s) \approx w_{KC}(s) = \sum_{j=1}^{n^*} k(s - s_j^*, \theta_1) u_j \]

\[ u_j \sim iid \sim N(0, 1) \]

Smoothing causes loss in variability:

\[ w(s) - w_{KC}(s) = \int k(s - v, \theta_1) dU(v) - \sum_{j=1}^{n^*} k(s - s_j^*, \theta_1) u_j \approx \sum_{j=n^*+1}^{\infty} k(s - s_j^*, \theta_1) u_j \]

No easy way to quantify this difference with kernel convolutions.
Low rank Gaussian process

- Call $w(s) \sim GP_m(0, C_\theta(\cdot))$ the *parent process*
- For $\mathcal{S}^* = \{s_1^*, s_2^*, \ldots, s_n^*\}$, let $C^*_w(\theta) = \{C_\theta(s_i^*, s_j^*)\}$:
  \[
  w^* = (w(s_1^*)^\top, w(s_2^*)^\top, \ldots, w(s_n^*)^\top)^\top \sim N(0, C^*_w(\theta))
  \]
- The *predictive process* derived from $w(s)$ is:
  \[
  \tilde{w}(s) = \mathbb{E}[w(s) | w^*] = \text{cov}\{w(s), w^*\}^\top \text{var}\{w^*\}^{-1}w^*.
  \]
- $\tilde{w}(s)$ is a *degenerate* Gaussian process delivering dimension-reduction.
Hierarchical predictive process models

\[
\tilde{w}(s) = z(s, \theta)^\top w^*
\]

Low rank interpolation

\[
\begin{align*}
\mathbf{w}^* &= \left(\mathbf{w}(s_1^*)^\top, \ldots, \mathbf{w}(s_{n^*})^\top\right)^\top \\
\end{align*}
\]

Hierarchical predictive process models

\[
p(\theta) \times p(\Psi) \times N(\beta | \mu_\beta, \Sigma_\beta) \times N(w^* | 0, C_w(\theta)) \\
\times \prod_{i=1}^n N_m(y(s_i) | X(s_i)^\top \beta + \tilde{w}(s_i), D(\Psi)).
\]
Hierarchical predictive process models

Parent process surface

Predictive process surface

JSM 2014 Hierarchical Modeling and Analysis for Spatial Data
Hierarchical predictive process models

Systemic under-estimation:

\[
\text{var}\{w(s)\} = \text{var}\{E[w(s) | w^*]\} + E\{\text{var}[w(s) | w^*]\} \\
\geq \text{var}\{E[w(s) | w^*]\} = \text{var}\{\tilde{w}(s)\}.
\]

Orthogonal decomposition:

\[
\text{var}\{w(s)\} = \text{var}\{\tilde{w}(s)\} + \text{var}\{w(s) - \tilde{w}(s)\}
\]

\[
\tilde{\epsilon}(s) = w(s) - \tilde{w}(s) \sim GP(0, C_{\tilde{\epsilon}}(s_1, s_2; \theta_1)):
\]

\[
C_{\tilde{\epsilon}}(s_1, s_2; \theta_1) = C(s_1, s_2; \theta_1) - c(s_1; \theta_1)' C^*(\theta_1)^{-1} c(s_2; \theta_2).
\]