What’s for today

We will discuss several methods for efficient handling of large data:

1. Toeplitz structure
2. Circulant structure
3. Covariance tapering
4. Likelihood approximation
5. Reduced rank approach
6. Predictive process
7. Adjustments to predictive process
8. Markov random field
Write the covariance matrix of $Z$ as $\Sigma = SKS^T(+\sigma^2V)$: $n \times n$

- Here $\sigma^2$ is nugget and $V$ is a diagonal matrix

- $K$ consists of unknown parameters, it is a positive definite $r \times r$ matrix

- $S$: $n \times r$ consists of basis functions

- $S(u) = (S_1(u), \ldots, S_r(u))^T$

- We have $C(u, v) = S(u)^T KS(v)$

- Therefore, we have $\text{Cov}\{Z(s_0), Z\} = S(s_0)^T KS^T$

- The point is, to invert $\Sigma$, we only need to invert $r \times r$ matrices
Fixed rank kriging (Cressie and Johannesson, JRSS B 2008)

- Note
  \[ \Sigma^{-1} = \sigma^{-1} V^{-1/2} \{ I + (\sigma^{-1} V^{-1/2} S) K (\sigma^{-1} V^{-1/2} S)^T \}^{-1} \sigma^{-1} V^{-1/2} \]

- Then using Sherman-Morrison-Woodbury formula:
  \[ (I + PKP^T)^{-1} = I - P(K^{-1} + P^T P)^{-1} P^T \]

- we get
  \[ \Sigma^{-1} = (\sigma^2 V)^{-1} - (\sigma^2 V)^{-1} S \{ K^{-1} + S^T (\sigma^2 V)^{-1} S \}^{-1} S^T (\sigma^2 V)^{-1} \]

- They recommend to use multiresolutional basis functions
In their analysis of total column ozone levels, they had 173405 observations (irregularly spaced).

They chose $r = 396$.

Their basis functions are from $S_{j(l)}(u) = \{1 - (\|u - v_{j(l)}\|/r_l)^2\}^2$, where $v_{j(l)}$ is one of the center points of the $l$th resolution, $l = 1, 2, 3$.

$r_l = 1.5$ (shortest arc distance between center points of the $l$th resolution).
Figure 2: Center points of 3 resolutions of DGG
Predictive process

- Banerjee et al. JRSS B, 2008 and other papers
- Consider a fixed set of “knots” $S^* = \{s_1^*, \ldots, s_m^*\}$. The BLUP of $w(s)$ conditional on $w^*$ is given by:

$$\tilde{w}(s) = E\{w(s)|w^*\} = \text{Cov}(w(s), w^*)\text{Var}^{-1}(w^*)w^*$$

- $\tilde{w}(s)$ is referred to as the predictive process and is used to approximate $w(s)$
- Captures large scale dependence well
- Fails to capture local/small scale dependence accurately and thus leads to biased parameter estimations and errors in prediction
A NEW APPROXIMATION METHOD

- Consider the following decomposition of \( w(s) \),

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

- \( \tilde{w}(s) \): a (multivariate) predictive process with
  \[
  \Gamma_{\tilde{w}}(s, s') = \text{Cov}(w(s), w^*)\text{Var}^{-1}(w^*)\text{Cov}(w^*, w(s'))
  \]
- \( w(s) - \tilde{w}(s) \): the residual process

Then the (cross) covariance is given by:

\[
  \Gamma_w(s, s') = \Gamma_{\tilde{w}}(s, s') + (\Gamma_w(s, s') - \Gamma_{\tilde{w}}(s, s'))
\]

- Using predictive process alone to approximate \( w(s) \), we discard the residual process \( w(s) - \tilde{w}(s) \) entirely and thus fail to capture the (small-scale) dependence it carries!
- Solution: approximates the residual process \( w(s) - \tilde{w}(s) \) (small-scale) while still maintaining computational efficiency.
The full-scale approximation (FSA)

\[
\mathbf{w}^\dagger(s) = \tilde{\mathbf{w}}(s) + \tilde{\mathbf{e}}(s)
\]

- \(\tilde{\mathbf{w}}(s)\) captures large scale spatial dependence
- \(\tilde{\mathbf{e}}(s)\) captures small scale spatial dependence that is unexplained by \(\tilde{\mathbf{w}}(s)\).

\[
\Gamma_{\tilde{\mathbf{e}}}(s, s') = [\Gamma_{\mathbf{w}}(s, s'; \theta) - \Gamma_{\tilde{\mathbf{w}}}(s, s'; \theta)] \circ \mathcal{K}(s, s')
\]

- FSA-Block: \(\mathcal{K}(s, s') = \mathbf{1}_{K \times K}\) if \(s\) and \(s'\) belong to the same subregion, and \(\mathcal{K}(s, s') = \mathbf{0}_{K \times K}\) otherwise.
- FSA-Taper: \(\mathcal{K}(s, s') = \bigoplus_{r=1}^{R} K_{taper, r}(s, s'; \gamma_r)\)
The data likelihood: $\mathbf{Y} \sim \text{MVN}(\mathbf{X}\beta, \Sigma_{\tilde{w}} + \Sigma_{\tilde{\epsilon}} + \mathbf{I}_n \otimes \Sigma_\epsilon)$

- $\Sigma_{\tilde{w}} = C_w(S, S^*; \theta) C_w^*(\theta) C_w^T(S, S^*; \theta)$
  - $\Sigma_{\tilde{w}}$ is a reduced rank matrix
  - $C_w(S, S^*; \theta) = \text{Cov}(w(S), w(S^*))$, $nR \times mR$
  - $C_w^* = \text{Var}(w(S^*))$, $mR \times mR$

- Here $R$ is the number of processes. So for the univariate case, $R = 1$

- $\Sigma_{\tilde{\epsilon}} = [\Gamma_{\tilde{\epsilon}}(\mathbf{s}_i, \mathbf{s}_j; \theta)]_{i,j=1}^n$
  - $\Sigma_{\tilde{\epsilon}}$ is a block diagonal matrix (FSA-Block) or a sparse matrix (FSA-Taper)
Use Sherman-Woodbury-Morrison formula for the inverse

Determinant computation:

\[
det(\Sigma_{\tilde{w}} + \Sigma_{\tilde{\epsilon}} + I_n \otimes \Sigma_\epsilon) \\
= \det\{C_w^* + C_w(S, S^*)^T (\Sigma_{\tilde{\epsilon}} + I_n \otimes \Sigma_\epsilon)^{-1} C_w(S, S^*)\} \\
\times \{\det(C_w^*)\}^{-1} \det(\Sigma_{\tilde{\epsilon}} + I_n \otimes \Sigma_\epsilon).
\]
Likelihood-based inference proceeds through maximum likelihood or restricted maximum likelihood methods.

With Bayesian hierarchical models, assign prior distributions to Ω. Draw samples \( p(\Omega | \mathbf{Y}) \propto P(\Omega)P(\mathbf{Y} | \Omega) \) using MCMC.

Spatial prediction (cokriging) is natural. Under the Bayesian inference framework, draw samples of \( \mathbf{Y}(s_0) \) from the predictive distribution \( P(\mathbf{Y}(s_0) | \Omega, \mathbf{Y}) \) by composition.
Climate model errors: Correlation between climate models 1 and 3

Corr of Model 1&3 (FSA)

Std(Corr) of Model 1&3 (FSA)

Corr of Model 1&3 (PP)

Std(Corr) of Model 1&3 (PP)
Let $Z = (Z_1, \ldots, Z_n)^T$ be a Gaussian random field with mean $\mu$ and covariance $\Sigma$

The precision matrix is denoted by $Q = \Sigma^{-1}$

We say $Z$ is a Gaussian Markov random field with respect to the labeled undirected graph $G = (\mathcal{V}, \mathcal{E})$ if the nodes are $\mathcal{V} = \{1, \ldots, n\}$ and the edges $\mathcal{E} = \{\{i, j\} \in \mathcal{V} \times \mathcal{V} : Q_{ij} \neq 0 \ & \ i \neq j\}$

If $\{i, j\} \in \mathcal{E}$, then $i$ and $j$ are neighbors

Conditional independence: $Z_i$ and $Z_j$ are conditionally independent given $Z_{-ij}$ if and only if $Q_{ij} = 0, i \neq j$

We have $V(Z_i|Z_{-i}) = 1/Q_{ii}$, $E(Z_i|Z_{-i}) = \mu_i - \frac{1}{Q_{ii}} \sum_{i,j: \text{neighbors}} Q_{ij}(Z_j - \mu_j)$, $\text{Cor}(Z_i, Z_j|Z_{-ij}) = -\frac{Q_{ij}}{\sqrt{Q_{ii}Q_{jj}}}$
In most cases, only $O(n)$ terms of $Q$ are nonzero and $O(n^2)$ terms are zero.

- There is open source C-library GMRFLib for efficient computation.
- For more details, see the work of Håvard Rue (Norwegian University of Science and Technology).