Modeling complex spatial dependencies: Low-rank cross-covariance models

Rajarshi Guhaniyogi[1], Andrew O. Finley[2] and Sudipto Banerjee[3]

[1] Department of Statistics and Decision Sciences,
Duke University, Durham, North Carolina, U.S.A.

[2] Department of Forestry & Department of Geography,
Michigan State University, East Lansing, Michigan, U.S.A

[3] Biostatistics, School of Public Health,
University of Minnesota, Twin Cities, Minnesota, U.S.A.

June 23, 2014

Challenges in spatial data analysis

- Our ability to collect, manage, and use spatial-temporal data is rapidly evolving.
- Interdisciplinary works leads to more complex questions $\longrightarrow$ complicated statistical models.
- Data-rich environments provide extraordinary opportunities to understand the complexity of large datasets.
- Major challenges:
  - Understand the complex inferential questions;
  - Construct (perhaps) complex statistical models, but that are interpretable and identifiable ("valid").
  - Overcome computational bottlenecks in implementation.

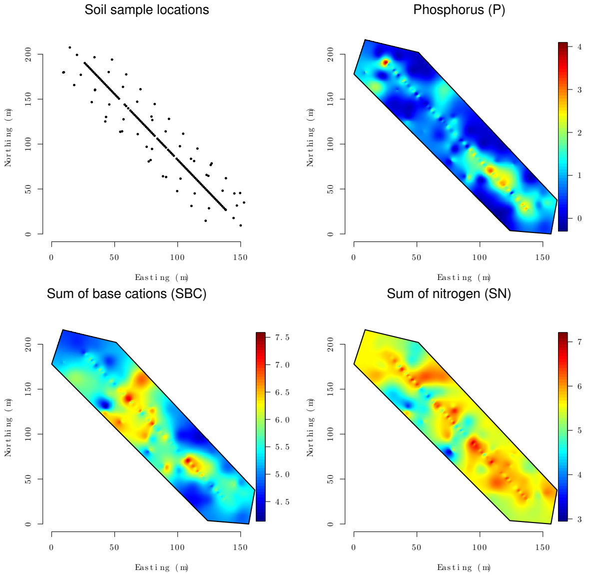
Point-referenced spatial data often arise as multivariate measurements at each location.

Examples:

- environmental monitoring: stations yield measurements on ozone, NO_2, CO, SO_2 and PM
- community ecology: assemblage of plant species due to water, nutrients, temperature, and light requirements
- forestry: measurements of stand characteristics age, total biomass, and average tree diameter.
- atmospheric modeling: at a given location we observe surface temperature, precipitation and wind speed

Dependence between outcomes within a given location and across proximate locations.

La Selva Biological Station soil data



Predictors include subject's access to environmental resources e.g., water, other nutrients, light.

Our objectives:

- predict soil nutrients for each tree's location (i.e., to serve as competition model covariates)
- document how nutrients co-vary in these tropical soils

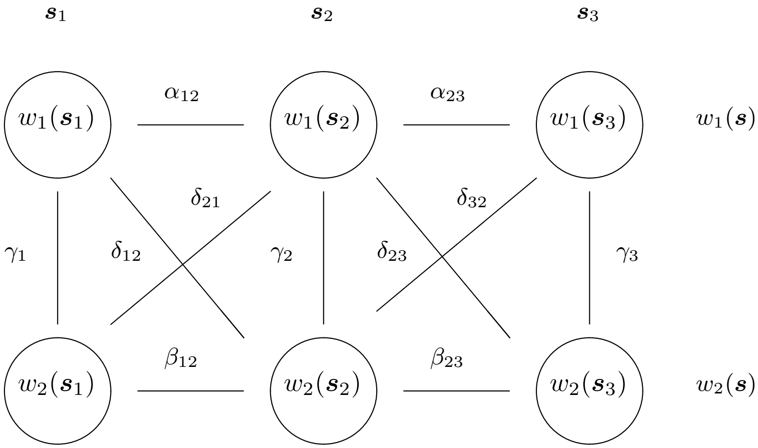
Predictors include subject's access to environmental resources e.g., water, other nutrients, light.

Our objectives:

- predict soil nutrients for each tree's location (i.e., to serve as competition model covariates)
- document how nutrients co-vary in these tropical soils

Data from La Selva Biological Station in Costa Rica:

- soil samples n =251
- three soil nutrients measured at each location


$\boldsymbol{s}_1$
$\boldsymbol{s}_2$
$\boldsymbol{s}_3$
$w_1(\boldsymbol{s}_1)$
α_{12}
$w_1(\boldsymbol{s}_2)$
α_{23}
$w_1(\boldsymbol{s}_3)$
$w_1(\boldsymbol{s})$
δ_{21}
δ_{32}
γ_1
δ_{12}
γ_2
δ_{23}
γ_3
β_{12}
β_{23}
$w_2(\boldsymbol{s}_1)$
$w_2(\boldsymbol{s}_2)$
$w_2(\boldsymbol{s}_3)$
$w_2(\boldsymbol{s})$

$$\text{var}\left\{\begin{array}{c} w_1(\boldsymbol{s}_1) \\ w_2(\boldsymbol{s}_1) \\ w_1(\boldsymbol{s}_2) \\ w_2(\boldsymbol{s}_2) \\ w_1(\boldsymbol{s}_3) \\ w_2(\boldsymbol{s}_3) \end{array}\right\} = \begin{pmatrix} * & \gamma_1 & \alpha_{12} & \delta_{12} & \alpha_{13} & \delta_{13} \\ \gamma_1 & * & \delta_{21} & \beta_{12} & \delta_{31} & \beta_{13} \\ \alpha_{12} & \delta_{21} & * & \gamma_2 & \alpha_{23} & \delta_{23} \\ \delta_{12} & \beta_{12} & \gamma_2 & * & \delta_{32} & \beta_{23} \\ \alpha_{13} & \delta_{31} & \alpha_{23} & \delta_{32} & * & \gamma_3 \\ \delta_{13} & \beta_{13} & \delta_{23} & \beta_{23} & \gamma_3 & * \end{pmatrix}$$

Conditional independence (graphical) models

- Can be computationally beneficial – introduce sparsity.
- Cond. indep. models may NOT be process models.
- Consider a Cond. indep. model for n sites:

$$[Y_1, \ldots, Y_n]_1$$

- Consider the observation from a "new" node, say Y_0. Form the distribution:

$$[Y_0, Y_1, \ldots, Y_n]_2$$

- Unfortunately:

$$\int [Y_0, Y_1, \ldots, Y_n]_2 = \int [Y_1, \ldots, Y_n]_2 \neq [Y_1, \ldots, Y_n]_1$$

- Not suitable for predictions at all. Inappropriate for continuous topologies.

The spatial interpolation problem

- Unknown signal $w(\cdot)$ observed over $\mathscr{S} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\} \subset \Re^d$.
- We seek an $f(\cdot)$ to agree with $w(\cdot)$ on $\mathscr{S}$.

$$f(\boldsymbol{s}) = b_1(\boldsymbol{s})\beta_1 + b_2(\boldsymbol{s})\beta_2 + \cdots + b_n(\boldsymbol{s})\beta_n = \boldsymbol{b}(\boldsymbol{s})'\boldsymbol{\beta}\ .$$

- Find β's such that $f(\boldsymbol{s}_i) = w(\boldsymbol{s}_i)$ for $\boldsymbol{s}_i \in \mathscr{S}$:

$$\begin{bmatrix} b_1(\boldsymbol{s}_1) & b_2(\boldsymbol{s}_1) & \cdots & b_n(\boldsymbol{s}_1) \\ b_1(\boldsymbol{s}_2) & b_2(\boldsymbol{s}_2) & \cdots & b_n(\boldsymbol{s}_2) \\ \vdots & \vdots & \ddots & \vdots \\ b_1(\boldsymbol{s}_n) & b_2(\boldsymbol{s}_n) & \cdots & b_n(\boldsymbol{s}_n) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} = \begin{bmatrix} w(\boldsymbol{s}_1) \\ w(\boldsymbol{s}_2) \\ \vdots \\ w(\boldsymbol{s}_n) \end{bmatrix} .$$

$$\boldsymbol{B\beta} = \boldsymbol{w}$$

- When $\boldsymbol{B}^{-1}$ exists: $f(\boldsymbol{s}) = \boldsymbol{b}(\boldsymbol{s})'\boldsymbol{B}^{-1}\boldsymbol{w}$.

"Kriging"

- How about constructing a *covariance* matrix as B?

$$b_j(\boldsymbol{s}_i) = \text{cov}\{w(\boldsymbol{s}_i), w(\boldsymbol{s}_j)\} = C_\theta(\boldsymbol{s}_i, \boldsymbol{s}_j)\ .$$

- $C_\theta(\boldsymbol{s}, \boldsymbol{t}) = C_\theta(\boldsymbol{t}, \boldsymbol{s})$ is a real-valued *covariance function*: For *any* $\mathscr{S} \subseteq \Re^d$,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} u_i C_\theta(\boldsymbol{s}_i, \boldsymbol{s}_j) u_j > 0 \quad \forall \quad u_i,\, u_j \in \Re \setminus \{0\}\ .$$

- Then $\boldsymbol{B} = \text{var}\{\boldsymbol{w}\}$ is symmetric, positive definite and

$$f(\boldsymbol{s}) = \text{cov}\{w(\boldsymbol{s}), \boldsymbol{w}\}'\text{var}\{\boldsymbol{w}\}^{-1}\boldsymbol{w}\ .$$

Covariance functions

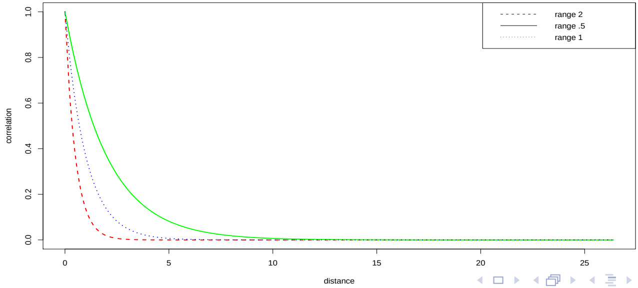
- Stationary: $C_\theta(\boldsymbol{s},\boldsymbol{t}) = C_\theta(\boldsymbol{t}-\boldsymbol{s})$. Isotropy: $C_\theta(\boldsymbol{s},\boldsymbol{t}) = C_\theta(\|\boldsymbol{t}-\boldsymbol{s}\|)$.
- Bochner: Covariance function $\Leftrightarrow$ characteristic function.

Matérn correlation:

$$C_\theta(\boldsymbol{s},\boldsymbol{t}) = \frac{\sigma^2}{2^{\phi_2-1}\Gamma(\phi_2)}(\|\boldsymbol{t}-\boldsymbol{s}\|\phi_1)^{\phi_2}\kappa_{\phi_2}(\|\boldsymbol{t}-\boldsymbol{s}\|;\phi_1)$$

$\phi_1 \rightarrow$ controls how fast correlation decays

$\phi_2 \rightarrow$ controls smoothness of the spatial surface



The multivariate spatial interpolation problem

- Now $\boldsymbol{w}(\boldsymbol{s})$ is an $m \times 1$ vector $\boldsymbol{s} \in \Re^d$.
- We seek an $m \times 1$ function $\boldsymbol{f}(\cdot)$ to agree with $\boldsymbol{w}(\cdot)$ on $\mathscr{S}$.

$$\boldsymbol{f}(\boldsymbol{s}) = \boldsymbol{B}_1(\boldsymbol{s})\boldsymbol{\beta}_1 + \boldsymbol{B}_2(\boldsymbol{s})\boldsymbol{\beta}_2 + \cdots + \boldsymbol{B}_n(\boldsymbol{s})\boldsymbol{\beta}_n = \boldsymbol{B}(\boldsymbol{s})'\boldsymbol{\beta}\ .$$

- Find β's such that $\boldsymbol{f}(\boldsymbol{s}_i) = \boldsymbol{w}(\boldsymbol{s}_i)$ for $\boldsymbol{s}_i \in \mathscr{S}$:

$$\begin{bmatrix} \boldsymbol{B}_1(\boldsymbol{s}_1) & \boldsymbol{B}_2(\boldsymbol{s}_1) & \cdots & \boldsymbol{B}_n(\boldsymbol{s}_1) \\ \boldsymbol{B}_1(\boldsymbol{s}_2) & \boldsymbol{B}_2(\boldsymbol{s}_2) & \cdots & \boldsymbol{B}_n(\boldsymbol{s}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{B}_1(\boldsymbol{s}_n) & \boldsymbol{B}_2(\boldsymbol{s}_n) & \cdots & \boldsymbol{B}_n(\boldsymbol{s}_n) \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{w}(\boldsymbol{s}_1) \\ \boldsymbol{w}(\boldsymbol{s}_2) \\ \vdots \\ \boldsymbol{w}(\boldsymbol{s}_n) \end{bmatrix} .$$

$$\boldsymbol{B}\boldsymbol{\beta} = \boldsymbol{w}$$

- But when will B^{-1} exist? Harder problem as $B(s_i)$'s are matrix functions.

"Multivariate Kriging"

- The analogue from the univariate case:

$$\boldsymbol{B}_j(\boldsymbol{s}) = \text{cov}\{\boldsymbol{w}(\boldsymbol{s}), \boldsymbol{w}(\boldsymbol{s}_j)\} = \boldsymbol{C}_\theta(\boldsymbol{s}, \boldsymbol{s}_j) = \{\text{cov}\{w_k(\boldsymbol{s}), w_l(\boldsymbol{s}_j)\}\} \ .$$

- $\boldsymbol{C}_\theta(\boldsymbol{s}, \boldsymbol{t}) = \boldsymbol{C}_\theta(\boldsymbol{t}, \boldsymbol{s})'$ is a *matrix*-valued *cross-covariance function*.
- For *any* $\mathscr{S} \subseteq \Re^d$,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{u}_i \boldsymbol{C}_\theta(\boldsymbol{s}_i, \boldsymbol{s}_j)\boldsymbol{u}_j > \mathbf{0} \quad \forall \quad \boldsymbol{u}_i,\, \boldsymbol{u}_j \in \Re^d \setminus \{0\} \ .$$

- $\boldsymbol{B} = \text{var}\{\boldsymbol{w}\}$ must be symmetric, positive definite, whereupon

$$\boldsymbol{f}(\boldsymbol{s}) = \text{cov}\{\boldsymbol{w}(\boldsymbol{s}), \boldsymbol{w}\}'\text{var}\{\boldsymbol{w}\}^{-1}\boldsymbol{w} \ .$$

So why not become fully stochastic?

- Assume that $\boldsymbol{w}(\boldsymbol{s})$ is an $m \times 1$ *multivariate spatial process*

$$\boldsymbol{w}(\boldsymbol{s}) \sim GP\left(\mathbf{0}, \boldsymbol{C}_\theta(\cdot)\right); \quad \boldsymbol{C}_\theta(\boldsymbol{s}, \boldsymbol{t}) = \{\text{cov}\{w_i(\boldsymbol{s}), w_j(\boldsymbol{t})\}\} .$$

- For any $\mathscr{S} = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n\} \subset \Re^d$, let $\boldsymbol{C}_w(\theta) = \{\boldsymbol{C}_\theta(\boldsymbol{s}_i, \boldsymbol{s}_j)\}$.

$$\boldsymbol{w} = (\boldsymbol{w}(\boldsymbol{s}_1)', \boldsymbol{w}(\boldsymbol{s}_2)', \ldots, \boldsymbol{w}(\boldsymbol{s}_n)')' \sim N(\mathbf{0}, \boldsymbol{C}_w(\theta));$$

- Spatial interpolation:

$$\text{E}[\boldsymbol{w}(\boldsymbol{s}) \,|\, \boldsymbol{w}] = \text{cov}\{\boldsymbol{w}(\boldsymbol{s}), \boldsymbol{w}\}'\text{var}\{\boldsymbol{w}\}^{-1}\boldsymbol{w} = \boldsymbol{f}(\boldsymbol{s}) .$$

Hierarchical Spatial model

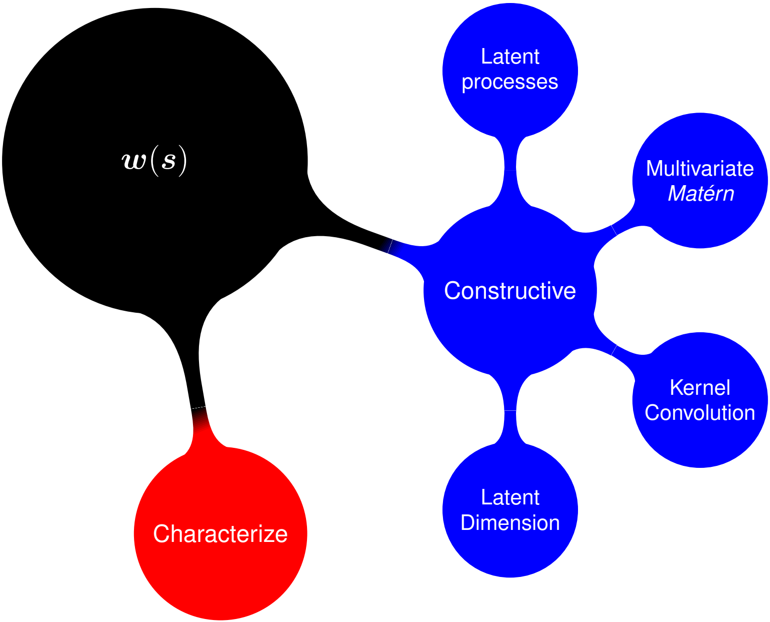
$$p(\boldsymbol{\theta}, \boldsymbol{\Psi}, \boldsymbol{\beta}, \boldsymbol{w} \,|\, \boldsymbol{y}) \propto p(\boldsymbol{\theta}) \times IW(\boldsymbol{\Psi} \,|\, a_\psi, \boldsymbol{S}_\psi) \times N(\boldsymbol{\beta} \,|\, \boldsymbol{\mu}_\beta, \Sigma_\beta)$$

$$\times N(\boldsymbol{w} \,|\, \mathbf{0}, \boldsymbol{C}_w(\boldsymbol{\theta})) \times \prod_{i=1}^{n} N_m(\boldsymbol{y}(\boldsymbol{s}_i) \,|\, \boldsymbol{X}(\boldsymbol{s}_i)'\boldsymbol{\beta} + \boldsymbol{w}(\boldsymbol{s}_i), \boldsymbol{\Psi})$$

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

Two primary issues.

- How do we construct valid matrix-valued cross-covariance functions?
- What if n is LARGE? How do we tackle $C_w(\theta)^{-1}$ (an $mn \times mn$ matrix)?


$w(s)$
Latent processes
Multivariate *Matérn*
Constructive
Kernel Convolution
Latent Dimension
Characterize

Constructive approach using latent variables

$$\begin{pmatrix} w_1(\boldsymbol{s}) \\ w_2(\boldsymbol{s}) \\ \vdots \\ w_m(\boldsymbol{s}) \end{pmatrix} = \begin{pmatrix} l_{11} & 0 & \dots & 0 \\ l_{21} & l_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{m1} & l_{m2} & \dots & l_{mm} \end{pmatrix} \begin{pmatrix} v_1(\boldsymbol{s}) \\ v_2(\boldsymbol{s}) \\ \vdots \\ v_m(\boldsymbol{s}) \end{pmatrix}$$

$\boldsymbol{w}(\boldsymbol{s})$

controls correlation in w_i's

$v_i(\cdot) \overset{ind}{\sim} GP(0, C_v(\theta_{v_i}))$

Spatially-varying (nonstationary) cross-covariances

$$\begin{pmatrix} w_1(\boldsymbol{s}) \\ w_2(\boldsymbol{s}) \\ \vdots \\ w_m(\boldsymbol{s}) \end{pmatrix} = \begin{pmatrix} l_{11}(\boldsymbol{s}) & 0 & \dots & 0 \\ l_{21}(\boldsymbol{s}) & l_{22}(\boldsymbol{s}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{m1}(\boldsymbol{s}) & l_{m2}(\boldsymbol{s}) & \dots & l_{mm}(\boldsymbol{s}) \end{pmatrix} \begin{pmatrix} v_1(\boldsymbol{s}) \\ v_2(\boldsymbol{s}) \\ \vdots \\ v_m(\boldsymbol{s}) \end{pmatrix}$$

$\boldsymbol{w}(\boldsymbol{s})$

$l_{ij}(\cdot) \overset{ind}{\sim} GP(0; C_w(\cdot; \boldsymbol{\theta}_{l_{ij}}))$

$v_i(\cdot) \overset{ind}{\sim} GP(0, C_v(\cdot; \boldsymbol{\theta}_{v_i}))$

$$\boldsymbol{w}(\boldsymbol{s}) = \boldsymbol{L}(\boldsymbol{s})\boldsymbol{v}(\boldsymbol{s})$$

$$\boldsymbol{C_w}(\boldsymbol{s},\boldsymbol{t}) = \boldsymbol{L}(\boldsymbol{s})\boldsymbol{C_v}(\boldsymbol{s},\boldsymbol{t})\boldsymbol{L}(\boldsymbol{t})'$$

$v_i(\boldsymbol{s}) \stackrel{ind}{\sim} GP(0, \rho_i(\cdot))$ and $\text{var}\{v_i(\boldsymbol{s})\} = 1$ for all $\boldsymbol{s}$.

$$\text{cov}(v_i(\boldsymbol{s}), v_j(\boldsymbol{t})) = \begin{cases} \rho_i(\boldsymbol{s},\boldsymbol{t};\boldsymbol{\theta}_i) & \text{if} \quad i = j \\ 0 \quad \text{if} \quad i \neq j \;, \end{cases}$$

$$\boldsymbol{C_v}(\boldsymbol{s},\boldsymbol{t}) = \begin{cases} \text{diag}\{\rho_i(\boldsymbol{s},\boldsymbol{t};\boldsymbol{\theta}_i)\} & \text{if} \quad \boldsymbol{s} \neq \boldsymbol{t} \\ \boldsymbol{I}_m \quad \text{if} \quad \boldsymbol{s} = \boldsymbol{t} \end{cases}$$

$$\boldsymbol{C_w}(\boldsymbol{s},\boldsymbol{s}) = \boldsymbol{L}(\boldsymbol{s})\boldsymbol{L}(\boldsymbol{s})' \Longrightarrow \boldsymbol{L}(\boldsymbol{s}) = \texttt{chol}(\boldsymbol{C_w}(\boldsymbol{s},\boldsymbol{s})) \;.$$

Dimension reduction

What if n is LARGE? How do we tackle $\boldsymbol{C}_w(\theta)^{-1}$ (an $mn \times mn$ matrix)?

- Covariance tapering (Furrer et al. 2006; Zhang and Du, 2007; Du et al. 2009; Kaufman et al., 2009)
- Spectral domain: (Fuentes 2007; Paciorek, 2007)
- Approximations using cond. indep. (Vecchia 1988; Stein et al. 2004; Rue et al. (2003))
- low-rank approaches (Wahba, 1990; Higdon, 2002; Lin et al., 2000; Paciorek, 2007; Rasmussen & Williams, 2006; Tokdar et al., 2007, 2011; Stein 2007, 2008; Cressie & Johannesson, 2008; Banerjee et al., 2008)

Low rank interpolation

- Cannot handle interpolations over $\mathscr{S}$.
- Interpolate over a smaller set of n^* locations, say $\mathscr{S}^* = \{\boldsymbol{s}_1^*, \boldsymbol{s}_2^*, \ldots, \boldsymbol{s}_{n^*}^*\}$ and $n^* << n$.

$$\tilde{\boldsymbol{f}}(\boldsymbol{s}) = \boldsymbol{B}_1^*(\boldsymbol{s})\boldsymbol{\beta}_1^* + \boldsymbol{B}_2^*(\boldsymbol{s})\boldsymbol{\beta}_2^* + \cdots + \boldsymbol{B}_{n^*}^*(\boldsymbol{s})\beta_{n^*}^* = \boldsymbol{B}^*(\boldsymbol{s})'\boldsymbol{\beta}^* \ .$$

- Find $\boldsymbol{\beta}$'s such that $\boldsymbol{f}(\boldsymbol{s}_i) = \boldsymbol{w}(\boldsymbol{s}_i)$ for $\boldsymbol{s}_i \in \mathscr{S}^*$:

$$\begin{bmatrix} \boldsymbol{B}_1^*(\boldsymbol{s}_1^*) & \boldsymbol{B}_2^*(\boldsymbol{s}_1^*) & \cdots & \boldsymbol{B}_{n^*}^*(\boldsymbol{s}_1^*) \\ \boldsymbol{B}_1^*(\boldsymbol{s}_2^*) & \boldsymbol{B}_2^*(\boldsymbol{s}_2^*) & \cdots & \boldsymbol{B}_{n^*}^*(\boldsymbol{s}_2^*) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{B}_1^*(\boldsymbol{s}_{n^*}) & \boldsymbol{B}_2^*(\boldsymbol{s}_{n^*}) & \cdots & \boldsymbol{B}_{n^*}^*(\boldsymbol{s}_{n^*}^*) \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1^* \\ \boldsymbol{\beta}_2^* \\ \vdots \\ \boldsymbol{\beta}_{n^*}^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{w}(\boldsymbol{s}_1^*) \\ \boldsymbol{w}(\boldsymbol{s}_2^*) \\ \vdots \\ \boldsymbol{w}(\boldsymbol{s}_{n^*}^*) \end{bmatrix} .$$

$$\boldsymbol{B}^*\boldsymbol{\beta}^* = \boldsymbol{w}^*$$

- When $\boldsymbol{B}^{*-1}$ exists: $\tilde{\boldsymbol{f}}(\boldsymbol{s}) = \boldsymbol{B}^*(\boldsymbol{s})'\boldsymbol{B}^{*-1}\boldsymbol{w}^*$.

Low rank kriging

- We set the basis functions as:

$$\boldsymbol{B}_j^*(\boldsymbol{s}) = \text{cov}\{\boldsymbol{w}(\boldsymbol{s}), \boldsymbol{w}(\boldsymbol{s}_j^*)\} = \boldsymbol{C}_\theta(\boldsymbol{s}, \boldsymbol{s}_j^*) \ .$$

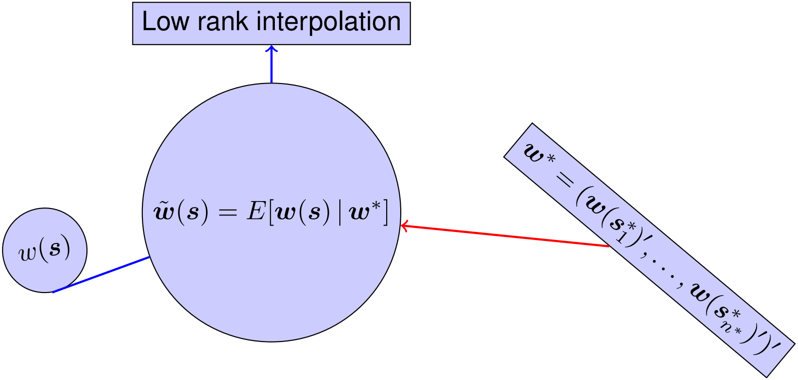
- Note: $\boldsymbol{B}^* = \text{var}\{\boldsymbol{w}^*\}$ is symmetric, positive definite and

$$\tilde{\boldsymbol{f}}(\boldsymbol{s}) = \text{cov}\{\boldsymbol{w}(\boldsymbol{s}), \boldsymbol{w}^*\}'\text{var}\{\boldsymbol{w}^*\}^{-1}\boldsymbol{w}^* \ .$$

- Inversion required for $\boldsymbol{B}^* = \text{var}\{\boldsymbol{w}^*\}$, which is $n^* \times n^*$.

Low rank Gaussian process

- Does low rank interpolation correspond to a "low rank" spatial process?
- Call $\boldsymbol{w}(\boldsymbol{s}) \sim GP_m\left(\mathbf{0}, \boldsymbol{C}_\theta(\cdot)\right)$ the *parent process*
- For $\mathscr{S}^* = \{\boldsymbol{s}_1^*, \boldsymbol{s}_2^*, \ldots, \boldsymbol{s}_{n^*}^*\}$, let $\boldsymbol{C}_{\boldsymbol{w}^*}^*(\theta) = \left\{\boldsymbol{C}_\theta(\boldsymbol{s}_i^*, \boldsymbol{s}_j^*)\right\}$:
$$\boldsymbol{w}^* = (\boldsymbol{w}(\boldsymbol{s}_1^*)', \boldsymbol{w}(\boldsymbol{s}_2^*)', \ldots, \boldsymbol{w}(\boldsymbol{s}_{n^*}^*)')' \sim N(\mathbf{0}, \boldsymbol{C}_{\boldsymbol{w}}^*(\theta))$$
- The *predictive process* derived from $w(\boldsymbol{s})$ is:
$$\tilde{\boldsymbol{w}}(\boldsymbol{s}) = \mathsf{E}[\boldsymbol{w}(\boldsymbol{s}) \,|\, \boldsymbol{w}^*] = \mathsf{cov}\{\boldsymbol{w}(\boldsymbol{s}), \boldsymbol{w}^*\}'\mathsf{var}\{\boldsymbol{w}^*\}^{-1}\boldsymbol{w}^* = \tilde{\boldsymbol{f}}(\boldsymbol{s})\ .$$
- $\tilde{\boldsymbol{w}}(\boldsymbol{s})$ is a *degenerate* Gaussian process delivering dimension-reduction.

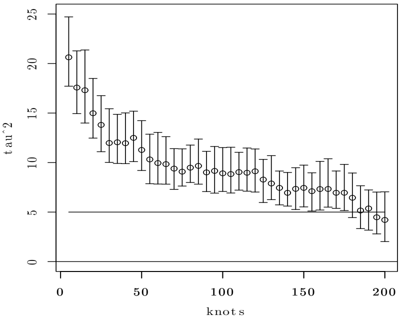


Hierarchical predictive process models

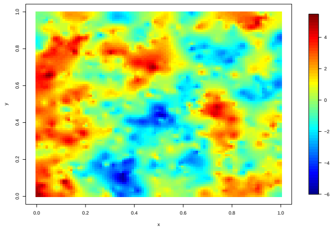
$$
\begin{aligned}
p(\boldsymbol{\theta}, \boldsymbol{\Psi}, \boldsymbol{\beta}, \boldsymbol{w}^* \,|\, \boldsymbol{y}) \propto\; & p(\boldsymbol{\theta}) \times IW(\boldsymbol{\Psi} \,|\, a_\psi, \boldsymbol{S}_\psi) \times N(\boldsymbol{\beta} \,|\, \boldsymbol{\mu}_\beta, \Sigma_\beta) \\
& \times N(\boldsymbol{w}^* \,|\, \mathbf{0}, \boldsymbol{C}_w^*(\boldsymbol{\theta})) \times \prod_{i=1}^{n} N_m(\boldsymbol{y}(\boldsymbol{s}_i) \,|\, \boldsymbol{X}(\boldsymbol{s}_i)'\boldsymbol{\beta} + \tilde{\boldsymbol{w}}(\boldsymbol{s}_i), \boldsymbol{\Psi}).
\end{aligned}
$$

How do we choose $\mathscr{S}^*$

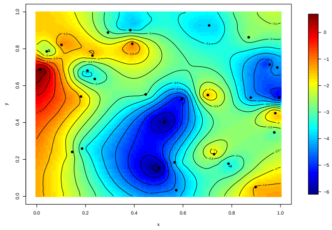
- Knot selection: Regular grid? More knots near locations we have sampled more?
- Formal spatial design paradigm: maximize information metrics.
- Geometric considerations: space-filling designs; various clustering algorithms
- Adaptive modeling of knots using point processes (Guhaniyogi et al., 2011).
- Compared performance of estimation of range and smoothness by varying knot size.
- Usually inference is quite robust to $\mathscr{S}^*$.
- More important to capture loss of variability due to low rank approximation.
- Seamlessly adapts to multivariate and spatiotemporal settings.



Parent process surface



Predictive process surface



Systemic under-estimation:

Systematic under-estimation

$$\begin{aligned}\text{var}\{w(\boldsymbol{s})\} &= \text{var}\{\text{E}[w(\boldsymbol{s}) \mid \boldsymbol{w}^*]\} + \text{E}\{\text{var}[w(\boldsymbol{s}) \mid \boldsymbol{w}^*]\} \\ &\geq \text{var}\{\text{E}[w(\boldsymbol{s}) \mid \boldsymbol{w}^*]\} = \text{var}\{\tilde{w}(\boldsymbol{s})\}.\end{aligned}$$

- Orthogonal decomposition:

$$\text{var}\{w(\boldsymbol{s})\} = \text{var}\{\tilde{w}(\boldsymbol{s})\} + \text{var}\{w(\boldsymbol{s}) - \tilde{w}(\boldsymbol{s})\}$$

- $\tilde{\epsilon}(\boldsymbol{s}) = w(\boldsymbol{s}) - \tilde{w}(\boldsymbol{s}) \sim GP(0, C_{\tilde{\epsilon}}(\boldsymbol{s}, \boldsymbol{t}; \boldsymbol{\theta}))$:

$$C_{\tilde{\epsilon}}(\boldsymbol{s}, \boldsymbol{t}; \boldsymbol{\theta}) = C(\boldsymbol{s}, \boldsymbol{t}; \boldsymbol{\theta}) - \boldsymbol{c}(\boldsymbol{s}; \boldsymbol{\theta})' \boldsymbol{C}^*(\boldsymbol{\theta})^{-1} \boldsymbol{c}(\boldsymbol{s}; \boldsymbol{\theta})\ .$$

Model-based non-degenerate structures

- Modified (non-degenerate) predictive process:

$$\tilde{\epsilon}(\boldsymbol{s}_i) \overset{iid}{\sim} N(0, \delta^2(\boldsymbol{s}_i; \boldsymbol{\theta}))\,; \quad \delta^2(\boldsymbol{s}; \boldsymbol{\theta}_1) = C_{\tilde{\epsilon}}(\boldsymbol{s}, \boldsymbol{s}; \boldsymbol{\theta})\ .$$

Tapered adjustment

$$\begin{aligned} \tilde{\epsilon}(\boldsymbol{s}) &\sim GP(0, C_{tap}(\boldsymbol{s}, \boldsymbol{t})) \\ C_{tap}(\boldsymbol{s}, \boldsymbol{t}; \boldsymbol{\theta}) &= C_{\tilde{\epsilon}}(\boldsymbol{s}, \boldsymbol{t}; \boldsymbol{\theta}) C_{\nu}(\|\boldsymbol{s} - \boldsymbol{t}\|; \boldsymbol{\theta})\ , \end{aligned}$$

- $C_{\nu}(\|\boldsymbol{s} - \boldsymbol{t}\|; \boldsymbol{\theta})$ is a compactly supported correlation function on $[0, \nu]$.

$$\begin{aligned} \nu &= 0 \Rightarrow \text{modified predictive process} \\ \nu &= \infty \Rightarrow \text{parent spatial process} \end{aligned}$$

Formal theory for oversmoothing by low-rank processes.

- Mean square continuity and differentiability at $\boldsymbol{s}_0$ of a process $w(\cdot)$ requires existence of some vector $\nabla w(\boldsymbol{s}_0)$ with,

$$\lim_{\boldsymbol{s}\to\boldsymbol{s}_0} E\left(w(\boldsymbol{s}) - w(\boldsymbol{s}_0)\right)^2 = 0$$

$$\lim_{h\to 0} E\left(\frac{w(\boldsymbol{s}_0 + h\boldsymbol{u}) - w(\boldsymbol{s}_0)}{h} - \langle \nabla w(\boldsymbol{s}_0), \boldsymbol{u}\rangle\right)^2 = 0$$

With Matérn correlation function for the parent process:

1. Predictive process is infinitely mean square differentiable except at the set of knot points $\mathscr{S}^*$.
2. Modified predictive process is not mean square continuous at any point.
3. Tapered predictive process can have exactly the same degree of smoothness as the parent process.

Low rank cross-covariances

$$
\begin{tikzcd}
\text{GP: } v_k(\boldsymbol{s}) \arrow[rr, "\text{Dim. Reduction}"] \arrow[d, leftrightarrow, "\text{ind.}"] & & \text{PP: } v_{k,pp}(\boldsymbol{s}) \arrow[r, "\text{Non degenerate}"] \arrow[d, leftrightarrow, "\text{ind.}"] & \text{MPP: } \tilde{v}_k(\boldsymbol{s}) \arrow[d, leftrightarrow, "\text{ind.}"] \\
\text{GP: } l_{ij}(\boldsymbol{s}) \arrow[rr, "\text{Dim. Reduction}"] & & \text{PP: } l_{ij,pp}(\boldsymbol{s}) \arrow[r, "\text{Non degenerate}"] & \text{MPP: } \tilde{l}_{ij}(\boldsymbol{s})
\end{tikzcd}
$$

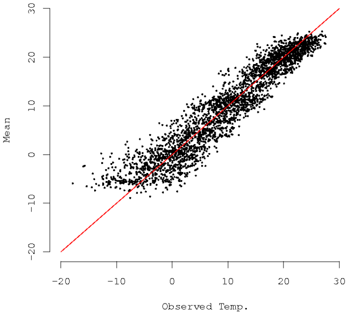
$$\tilde{\boldsymbol{w}}(\boldsymbol{s}) = \tilde{\boldsymbol{L}}(\boldsymbol{s})\tilde{\boldsymbol{v}}(\boldsymbol{s})$$

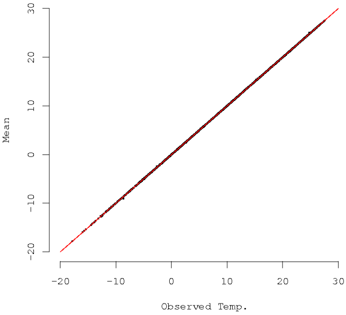
$$\boldsymbol{w}_{mpp}(\boldsymbol{s}) = \boldsymbol{L}_{pp}(\boldsymbol{s})\boldsymbol{v}_{pp}(\boldsymbol{s}) + \tilde{\boldsymbol{\epsilon}}(\boldsymbol{s})$$

Parameter credible intervals, 50 (2.5 97.5) percentiles, for soil nutrient data analysis candidate models.

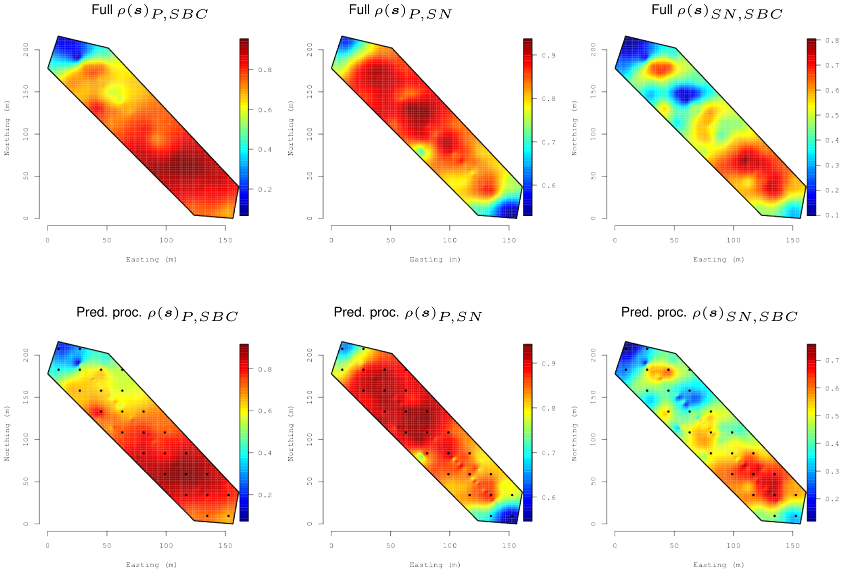
Parameter	Stationary	Non-stationary: Full	Non-stationary: Predictive process 26
$\beta_{0,P}$	0.71 (0.26, 1.35)	0.66 (0.22, 1.05)	0.64 (0.33, 1.20)
$\beta_{0,SBC}$	5.38 (5.03, 6.08)	5.18 (4.86, 5.49)	5.16 (4.83, 5.40)
$\beta_{0,SN}$	5.42 (4.97, 5.86)	5.53 (5.30, 5.78)	5.53 (5.31, 5.73)
$\sigma^2_{P,P}$	0.92 (0.52, 2.29)	0.20 (0.09, 0.53)	0.22 (0.08, 0.57)
$\sigma^2_{SBC,P}$	0.47 (0.25, 1.23)	0.24 (0.10, 0.63)	0.21 (0.10, 0.54)
$\sigma^2_{SN,P}$	0.49 (0.26, 1.25)	0.20 (0.09, 0.50)	0.23 (0.11, 0.75)
$\sigma^2_{SBC,SBC}$	0.44 (0.27, 1.08)	0.54 (0.18, 1.64)	0.36 (0.13, 1.01)
$\sigma^2_{SN,SBC}$	0.19 (0.06, 0.51)	0.14 (0.06, 0.36)	0.15 (0.07, 0.38)
$\sigma^2_{SN,SN}$	0.39 (0.19, 1.08)	1.85 (0.62, 6.11)	1.77 (0.41, 10.38)
ϕ_a	–	0.0135 (0.0125, 0.0173)	0.0134 (0.0125, 0.0170)
ϕ_w	0.0499 (0.0165, 0.0873)	0.0371 (0.0180, 0.0737)	0.0284 (0.0133, 0.0603)
Eff. range$_a$ m	–	222.13 (173.32, 238.97)	224.36 (176.57, 239.17)
Eff. range$_w$ m	60.04 (34.31, 181.08)	80.68 (40.66, 166.33)	105.64 (49.65, 225.05)
τ^2_P	0.21 (0.14, 0.30)	0.19 (0.13, 0.28)	0.19 (0.13, 0.28)
τ^2_{SBC}	0.07 (0.05, 0.11)	0.06 (0.04, 0.09)	0.06 (0.04, 0.09)
τ^2_{SN}	0.15 (0.11, 0.21)	0.11 (0.07, 0.16)	0.09 (0.06, 0.14)

Model Assessment



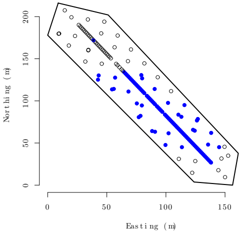


Non-stationary – full versus predictive process

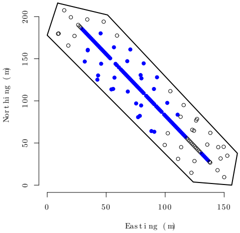


Non-stationary – full versus predictive process, $\rho(\boldsymbol{s})$ sig. at 0.05 level (●) positive, (●) negative

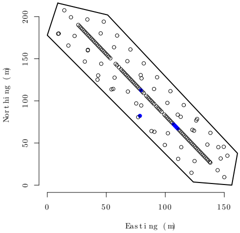
Full $\rho(\boldsymbol{s})_{P,SBC}$



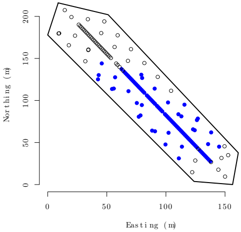
Full $\rho(\boldsymbol{s})_{P,SN}$



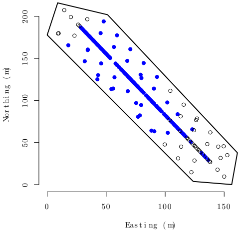
Full $\rho(\boldsymbol{s})_{SN,SBC}$



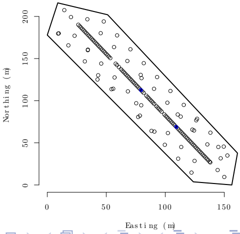
Pred. proc. $\rho(\boldsymbol{s})_{P,SBC}$



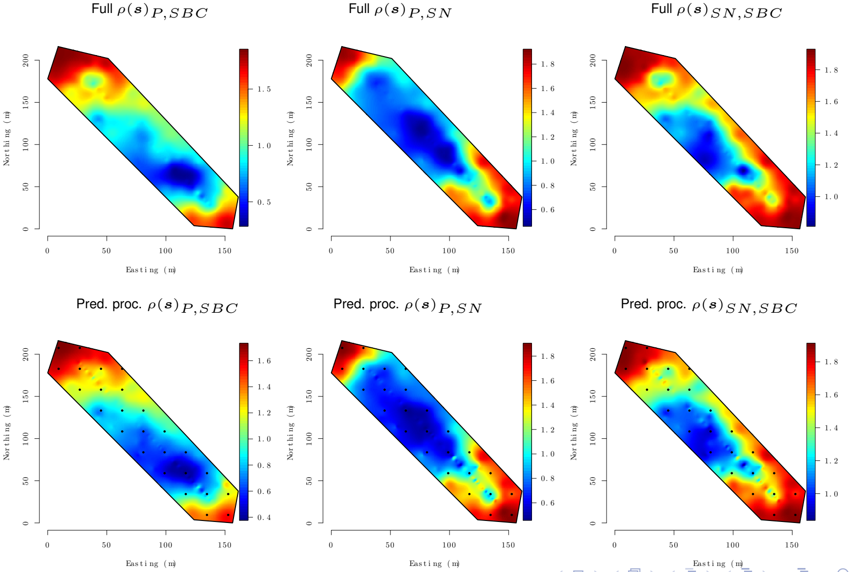
Pred. proc. $\rho(\boldsymbol{s})_{P,SN}$



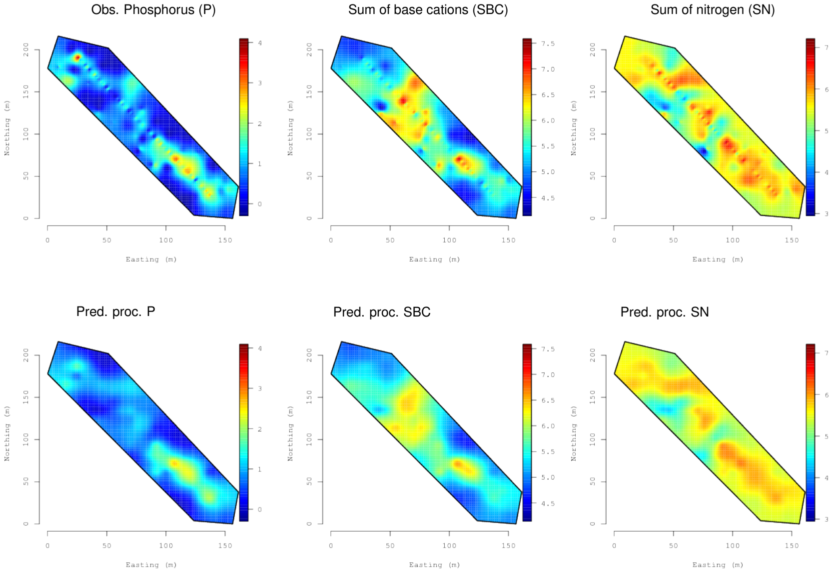
Pred. proc. $\rho(\boldsymbol{s})_{SN,SBC}$



Non-stationary – full versus predictive process, $\rho(\boldsymbol{s})$ range between 0.025-0.975 CI



Non-stationary – observed (interpolated) versus predictive process (predicted)



Summary

Challenge – to meet spatial modeling needs:

- Predictive process models for large datasets and complex models
    - Use some model-based adjustment to compensate for over-smoothing;
    - stochastically model the knots?
    - Tapered adjustment delivers same level of smoothness as parent (Guhaniyogi et al., 2011).
- Computing: C++ with OpenMP/MKL
    - Now available in the `R` package `spBayes`.

This work was supported by:

NSF Grant DMS 0706870.

Thank you!