Advanced Bayesian Computation Weeks 6 & 7

Rajarshi Guhaniyogi
Winter 2018

February 23, 2018

Linear Model Co-regionalization

- Let $\boldsymbol{A} = ((a_{ij}))_{i,j=1}^{m}$, such that the entries above lower triangle are all zero.
- Assign $a_{ij} \sim N(0,1)$ and $\log(a_{ii}) \sim N(0,1)$ for $i > j$.
- Assume $v_k(\boldsymbol{s}) \sim GP(0, C_{1,\phi_k})$ independently.
- ϕ_k's are assigned uniform priors as before.
- It is customarily assumed that $\boldsymbol{\Psi} = diag(\psi_1, ..., \psi_m)$, $\psi_1, ..., \psi_m > 0$.
- $\psi_k \sim IG(a, b)$ a-priori, for $k = 1, ..., m$.
- Entries of $\boldsymbol{B}$ are assigned normal priors.

Multivariate Matern

- Need to define $Cov(\boldsymbol{w}(\boldsymbol{s}), \boldsymbol{w}(\boldsymbol{s}'))$, where $h = ||\boldsymbol{s} - \boldsymbol{s}'||$.
- Amounts to defining $Cov(w_k(\boldsymbol{s}), w_k(\boldsymbol{s}'))$ and $Cov(w_k(\boldsymbol{s}), w_j(\boldsymbol{s}'))$.
- $Cov(w_k(\boldsymbol{s}), w_k(\boldsymbol{s}')) = M(h|\nu_k, a_k)$, $Cov(w_k(\boldsymbol{s}), w_j(\boldsymbol{s}')) = \alpha_{kj} M(h|\nu_{kj}, a_{kj})$.
- Parsimonious Matern kernel defines $a_{kj} = a_k = a$ and $\nu_{kj} = \frac{\nu_k + \nu_j}{2}$.
- $\alpha_{kj} = \zeta_{kj} \lambda_k \lambda_j$, $\lambda_k, \lambda_j > 0$
- $\zeta_{kj} = \eta_{kj} \frac{\Gamma(\nu_k + d/2)^{1/2}}{\Gamma(\nu_k)^{1/2}} \frac{\Gamma(\nu_j + d/2)^{1/2}}{\Gamma(\nu_j)^{1/2}} \frac{\Gamma((\nu_k + \nu_j)/2)}{\Gamma((\nu_k + \nu_j)/2 + d/2)}$.
- $\eta_{kk} = 1$ and $((\eta_{kj}))_{k,j=1}^{m}$ is non-negative definite.
- This ensures a valid multivariate cross covariance matrix.

Gaussian process High Dimensional Regression

- Consider modeling the situation $y = f(\boldsymbol{x}) + \epsilon$, where $\boldsymbol{x} \in \mathbb{R}^p$ where p is large.
- We have already seen in the theoretical study of Gaussian process that the convergence rate of the fitted function to the truth is $n^{-s/(2s+p)}(\log(n))^c$, where c is a constant, s is the smoothness of the true surface, p is the number of predictors and n is the sample size.
- When p is large, convergence rate suffers a lot.
- We land in a hopeless situation (similar to the linear regression case) where we need to add some assumption in the true regression model to be able to recover it.
- Assumption: The regression function is affected only by a few covariates.
- Alternatively, one assumes that the covariates lie in a lower dimensional noisy manifold.

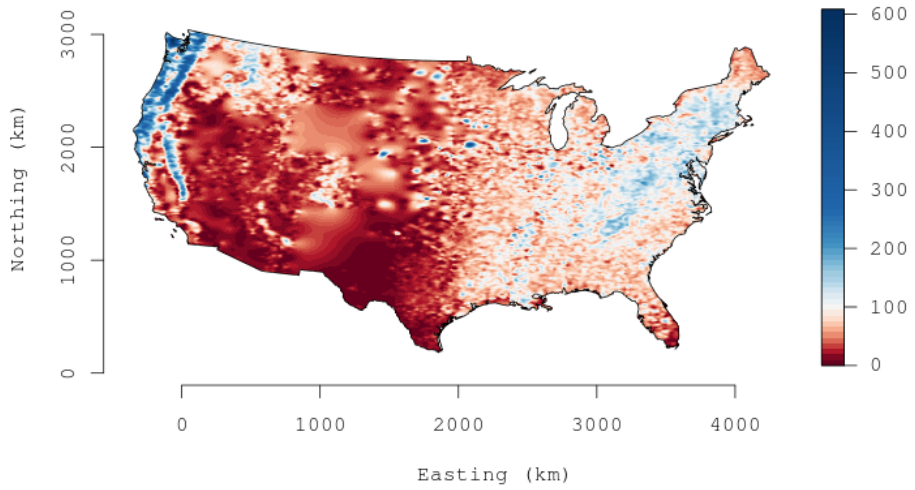
Variable selection in Gaussian process

- Variable selection in Gaussian process is a very hard and unsolved problem.
- One idea was to employ a variant of the squared exponential covariance kernel.
- They define the covariance kernel $C_{\sigma^2,\phi,\boldsymbol{\lambda}}(\boldsymbol{x},\boldsymbol{x}') = \sigma^2 \exp(-\sum_{j=1}^{p} \lambda_j (x_j - x_j')^2)$.
- Use some penalty to select important λ_j's.
- This is known as the automatic relevance determination (ARD).
- Full Bayesian updating of λ_j's by adding a spike and slab prior and computing the posterior does not seems to work very well.

Big data, large sample size n and Gaussian Process

- In many machine learning or environmental applications number of predictors is small.
- Sample size is massive.
- Important data applications.
- It is a wide area with different strategies applied to different models.
- We will see a few strategies.

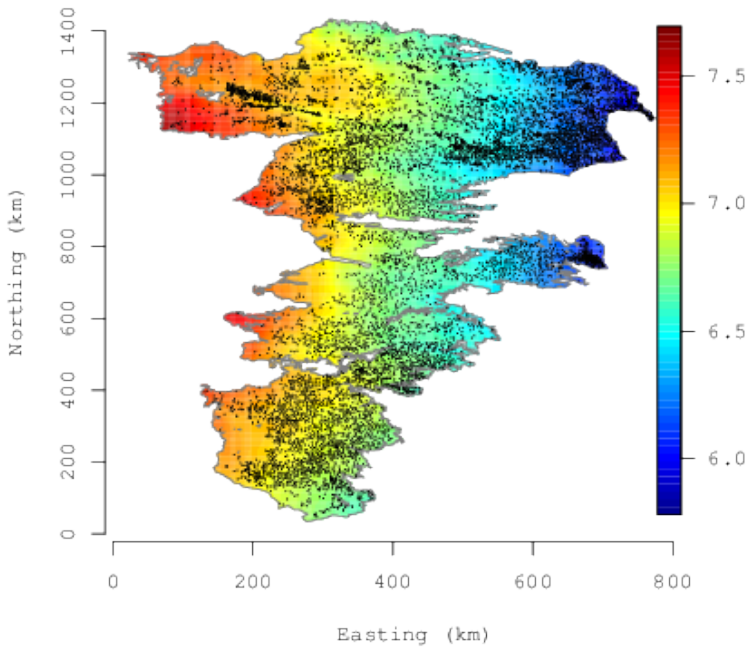
Forest Biomass Data



- Biomass is the biological material in a living or recently dead organism.
- Prediction of forest biomass is important to understand current carbon stock and flux, bio-feedstock for emerging bio-economies, and impact of deforestation.
- Forest Inventory and Analysis (FIA) under USDA collects data on Biomass regurlarly.

Depth of water table data between 1935 to 2013.

(a) High Plain Aquifer



(b) Depth of water table

Modeling of spatio-temporal trend of water table

Problem with Gaussian processes with big data

- $\boldsymbol{y} = (y_1, ..., y_n)'$ are observed data and $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ are the corresponding predictors.
- Let $\boldsymbol{X} = [\boldsymbol{x}_1 : \cdots : \boldsymbol{x}_n]'$ be the predictor matrix.
- Model: $\boldsymbol{y} \sim N(\mu \mathbf{1}_n, \boldsymbol{C}_{\boldsymbol{\theta}} + \tau^2 \boldsymbol{I})$.
- Estimating parameters $\boldsymbol{\theta}$ from the likelihood

$$-\frac{1}{2}\log(det(\boldsymbol{C}_{\boldsymbol{\theta}} + \tau^2 \boldsymbol{I})) - \frac{1}{2}(\boldsymbol{y} - \mu \mathbf{1}_n)'(\boldsymbol{C}_{\boldsymbol{\theta}} + \tau^2 \boldsymbol{I})^{-1}(\boldsymbol{y} - \mu \mathbf{1}_n)$$

Challenges

- Store $\boldsymbol{C}_{\boldsymbol{\theta}} + \tau^2 \boldsymbol{I}$
- Compute $\texttt{Chol}(\boldsymbol{C}_{\boldsymbol{\theta}} + \tau^2 \boldsymbol{I}) = \boldsymbol{L}\boldsymbol{L}'$.

Low Rank Model

- Approximate $\boldsymbol{C_\theta} \approx \boldsymbol{B_\theta} \boldsymbol{C_\theta^{*-1}} \boldsymbol{B_\theta'} + \boldsymbol{D_\theta}$
- $\boldsymbol{B_\theta}$ is the $n \times r$ spatial basis matrix $r << n$.
- $\boldsymbol{C_\theta^*}$ is an $r \times r$ spatial covariance matrix.
- $\boldsymbol{D_\theta}$ is either sparse or diagonal.
- Different choices of basis functions leads to different low rank models.
- The computational complexity $O(r^3 + nr^2) \leq O(n^3)$.

Kernel Convolution

- Use a finite basis representation to approximate a Gaussian process.
- $f(\boldsymbol{x}) = \int K(\boldsymbol{x} - \boldsymbol{z}, \boldsymbol{\theta}) du(\boldsymbol{z})$, where $u(\boldsymbol{z})$ is a white noise process.
- This is called Kernel convolution of white noise processes and is widely used in environmental applications.
- An approximation of the $f(\boldsymbol{x})$ by a finite sum is given by

$$f(\boldsymbol{x}) \approx \sum_{j=1}^{J} K(\boldsymbol{x} - \boldsymbol{z}_j^*, \boldsymbol{\theta}) u_j,$$

  $\boldsymbol{z}_1^*, ..., \boldsymbol{z}_J^*$ are "knot" points in $\mathbb{R}^p$ and $u_1, ..., u_J \sim N(0, \sigma^2)$.
- $J << n$, i.e. number knots is much lesser than the sample size.
- Choice of $K(\cdot, \boldsymbol{\theta})$ is important for the method to work.
- In the spatial context this idea was proposed in 2001 by David Higdon.

Kernel Convolution

- Model

$$y = \mu + \sum_{j=1}^{J} K(\boldsymbol{x} - \boldsymbol{z}_j^*, \boldsymbol{\theta}) u_j + \epsilon, \ \epsilon \sim N(0, \tau^2).$$

- With data $(y_1, \boldsymbol{x}_1), ..., (y_n, \boldsymbol{x}_n)$, the data equation is given by

$$\boldsymbol{y} = \mu \mathbf{1}_n + K\boldsymbol{u} + \boldsymbol{\epsilon}, \ \ \boldsymbol{\epsilon} \sim N(0, \tau^2 \boldsymbol{I})$$

  where the ith row of $\boldsymbol{K}$ is given by $(K(\boldsymbol{x}_i - \boldsymbol{z}_1^*, \boldsymbol{\theta}), ..., K(\boldsymbol{x}_i - \boldsymbol{z}_J^*, \boldsymbol{\theta}))'$.
- $\boldsymbol{u} = (u_1, ..., u_J)' \sim N(0\sigma^2 \boldsymbol{I})$.
- Marginalizing over $\boldsymbol{u}$, the likelihood is $N(\boldsymbol{y}|\mu\mathbf{1}_n, \sigma^2 \boldsymbol{K}\boldsymbol{K}' + \tau^2 \boldsymbol{I})$.
- More generally, if one assumes $u_1, ..., u_J$ are correlated and $\boldsymbol{u} \sim N(0, \sigma^2 \boldsymbol{D})$ then the likelihood is $N(\boldsymbol{y}|\mu\mathbf{1}_n, \sigma^2 \boldsymbol{K}\boldsymbol{D}\boldsymbol{K}' + \tau^2 \boldsymbol{I})$.

Posterior distribution

$$p(\mu, \sigma^2, \tau^2) \propto N((\boldsymbol{y}|\mu \mathbf{1}_n, \sigma^2 \boldsymbol{K}\boldsymbol{D}\boldsymbol{K}' + \tau^2 \boldsymbol{I}) \times p(\mu) \times p(\tau^2) \times p(\sigma^2).$$

- $\boldsymbol{K}$ is a $n \times J$ matrix.
- Inverting the $n \times n$ matrix $\boldsymbol{H} = \sigma^2 \boldsymbol{K}\boldsymbol{D}\boldsymbol{K}' + \tau^2 \boldsymbol{I}$ can be made easy with simple trick.
- This technique is called Sherman-Woodbury-Morrison matrix identity.
- The main computational cost comes from inverting a $J \times J$ matrix.

Posterior predictive distribution and estimation of regression function

- Posterior distribution of $\boldsymbol{u}$ can be obtained easily.
- Note that $p(\boldsymbol{u}|-) \propto N(\boldsymbol{y}|\mu \mathbf{1}_n + \boldsymbol{K}\boldsymbol{u}, \tau^2 \boldsymbol{I}) \times N(\boldsymbol{u}|\mathbf{0}, \sigma^2 \boldsymbol{D})$.
- $\boldsymbol{u}|-$ follows a multivariate normal distribution.
- Prediction at new predictor $\boldsymbol{x}$ can be obtained as before.
- $y \sim N(\mu + \sum_{j=1}^{J} K(\boldsymbol{x} - \boldsymbol{z}_j^*, \boldsymbol{\theta}) u_j, \tau^2)$. Draw posterior predictive samples from this distribution corresponding to every posterior sample of the parameters.

Choice of the kernel function

- Choice of the kernel function plays important role in the kernel convolution method.
- Popular choices are radial basis functions, wavelet basis functions etc.
- There can be drastic differences with different choices of the basis function.
- Can the kernel choice be automated?

Predictive Process Model (Banerjee et al., 2008)/Sparse Gaussian Process

- Consider a set of n^* knots $\mathscr{S}^* = \{\boldsymbol{s}_1^*, ..., \boldsymbol{s}_{n^*}^*\}$ with $n^* << n$.

Predictive Process Model (Banerjee et al., 2008)/Sparse Gaussian Process

- Consider a set of n^* knots $\mathscr{S}^* = \{\boldsymbol{s}_1^*, ..., \boldsymbol{s}_{n^*}^*\}$ with $n^* << n$.
- Process realization over knots:
  $\boldsymbol{w}^* = (w(\boldsymbol{s}_1), ..., w(\boldsymbol{s}_{n^*}^*))' \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{R}^*)$.

Predictive Process Model (Banerjee et al., 2008)/Sparse Gaussian Process

- Consider a set of n^* knots $\mathscr{S}^* = \{\boldsymbol{s}_1^*, ..., \boldsymbol{s}_{n^*}^*\}$ with $n^* << n$.
- Process realization over knots:
  $\boldsymbol{w}^* = (w(\boldsymbol{s}_1), ..., w(\boldsymbol{s}_{n^*}^*))' \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{R}^*)$.
- $\boldsymbol{R}^* = [\rho(\boldsymbol{s}_i^*, \boldsymbol{s}_j^*; \phi)]$ is an $n^* \times n^*$.

Predictive Process Model (Banerjee et al., 2008)/Sparse Gaussian Process

- Consider a set of n^* knots $\mathscr{S}^* = \{\boldsymbol{s}_1^*, ..., \boldsymbol{s}_{n^*}^*\}$ with $n^* << n$.
- Process realization over knots:
  $\boldsymbol{w}^* = (w(\boldsymbol{s}_1), ..., w(\boldsymbol{s}_{n^*}^*))' \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{R}^*)$.
- $\boldsymbol{R}^* = [\rho(\boldsymbol{s}_i^*, \boldsymbol{s}_j^*; \phi)]$ is an $n^* \times n^*$.
- Correlation between $w(\boldsymbol{s})$ and $\boldsymbol{w}^*$: $\boldsymbol{r}(\boldsymbol{s}) = [\rho(\boldsymbol{s}, \boldsymbol{s}_i^*; \phi)]$.

Predictive Process Model (Banerjee et al., 2008)/Sparse Gaussian Process

- Consider a set of n^* knots $\mathscr{S}^* = \{\boldsymbol{s}_1^*, ..., \boldsymbol{s}_{n^*}^*\}$ with $n^* << n$.
- Process realization over knots:
  $\boldsymbol{w}^* = (w(\boldsymbol{s}_1), ..., w(\boldsymbol{s}_{n^*}^*))' \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{R}^*)$.
- $\boldsymbol{R}^* = [\rho(\boldsymbol{s}_i^*, \boldsymbol{s}_j^*; \phi)]$ is an $n^* \times n^*$.
- Correlation between $w(\boldsymbol{s})$ and $\boldsymbol{w}^*$: $\boldsymbol{r}(\boldsymbol{s}) = [\rho(\boldsymbol{s}, \boldsymbol{s}_i^*; \phi)]$.
- predictive process: $\tilde{w}(\boldsymbol{s}) = \mathrm{E}[w(\boldsymbol{s})|\boldsymbol{w}^*] = \boldsymbol{r}(\boldsymbol{s})' \boldsymbol{R}^{*-1} \boldsymbol{w}^*$

Predictive Process Model (Banerjee et al., 2008)/Sparse Gaussian Process

- Consider a set of n^* knots $\mathscr{S}^* = \{\boldsymbol{s}_1^*, ..., \boldsymbol{s}_{n^*}^*\}$ with $n^* << n$.
- Process realization over knots:
  $\boldsymbol{w}^* = (w(\boldsymbol{s}_1), ..., w(\boldsymbol{s}_{n^*}^*))' \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{R}^*)$.
- $\boldsymbol{R}^* = [\rho(\boldsymbol{s}_i^*, \boldsymbol{s}_j^*; \phi)]$ is an $n^* \times n^*$.
- Correlation between $w(\boldsymbol{s})$ and $\boldsymbol{w}^*$: $\boldsymbol{r}(\boldsymbol{s}) = [\rho(\boldsymbol{s}, \boldsymbol{s}_i^*; \phi)]$.
- predictive process: $\tilde{w}(\boldsymbol{s}) = \mathrm{E}[w(\boldsymbol{s})|\boldsymbol{w}^*] = \boldsymbol{r}(\boldsymbol{s})' \boldsymbol{R}^{*-1} \boldsymbol{w}^*$

predictive process model

$$y(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s})'\beta + \tilde{w}(\boldsymbol{s}) + \epsilon(\boldsymbol{s}),\ \epsilon(\boldsymbol{s}) \sim N(0, \tau^2)$$

Predictive Process Model (Banerjee et al., 2008)/Sparse Gaussian Process

- Consider a set of n^* knots $\mathscr{S}^* = \{\boldsymbol{s}_1^*, ..., \boldsymbol{s}_{n^*}^*\}$ with $n^* << n$.
- Process realization over knots:
  $\boldsymbol{w}^* = (w(\boldsymbol{s}_1), ..., w(\boldsymbol{s}_{n^*}^*))' \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{R}^*)$.

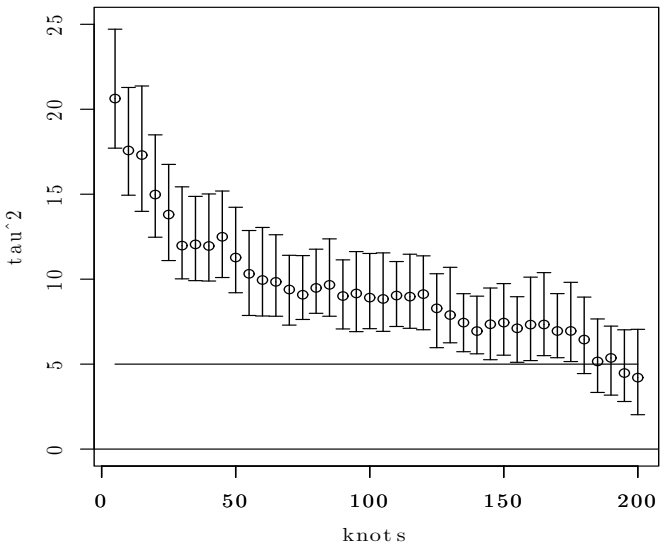
Dimension reduction comes from applying Sherman-Woodbury-Morrison matrix identity

Only inverting an $n^* \times n^*$ matrix inversion is required at each stage of MCMC.

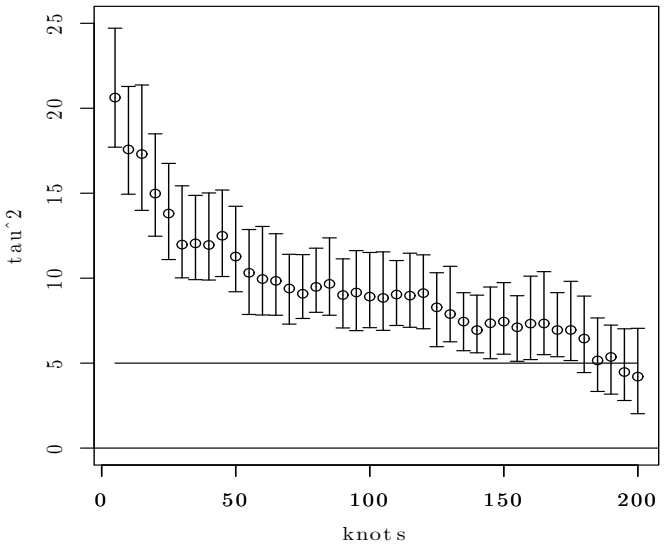
predictive process model

$$y(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s})'\beta + \tilde{w}(\boldsymbol{s}) + \epsilon(\boldsymbol{s}),\ \epsilon(\boldsymbol{s}) \sim N(0, \tau^2)$$

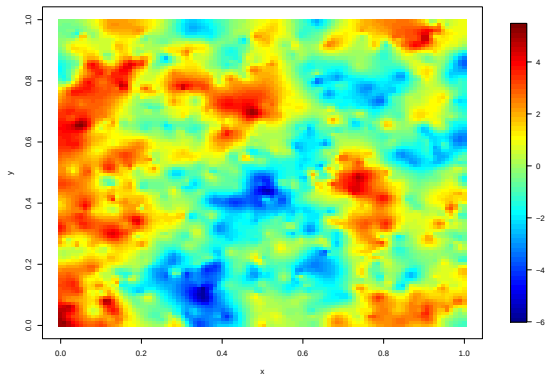
Issues with PP: Overestimated Nugget & Smoothed Spatial Surface



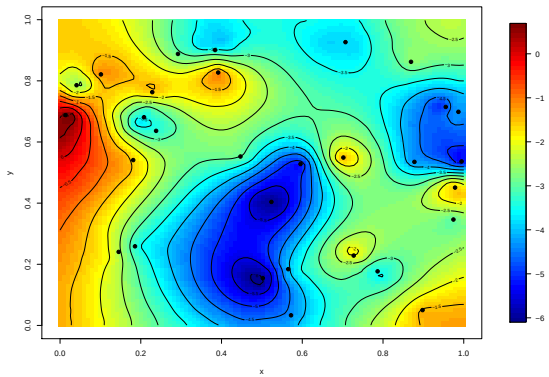
Issues with PP: Overestimated Nugget & Smoothed Spatial Surface



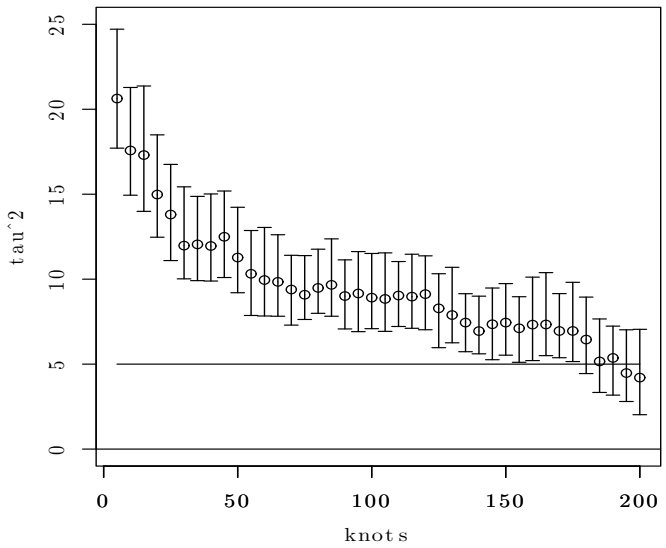
True Spatial Surface



Predictive Process Surface



Issues with PP: Overestimated Nugget & Smoothed Spatial Surface



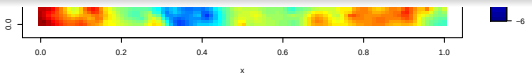
True Spatial Surface

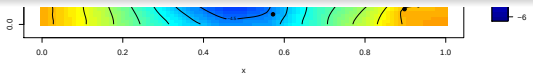
Predictive Process Surface

Quantified underestimation

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

$$\tilde{\epsilon}(\boldsymbol{s}) = w(\boldsymbol{s}) - \tilde{w}(\boldsymbol{s}) \sim GP(0, C_{\tilde{\epsilon}}(\boldsymbol{s}_1, \boldsymbol{s}_2; \boldsymbol{\theta}))$$





Model Based Bias Adjustments

▸ TPP slide

$$y(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s})'\boldsymbol{\beta} + \tilde{w}(\boldsymbol{s}) + \tilde{\epsilon}(\boldsymbol{s}) + \epsilon(\boldsymbol{s})$$

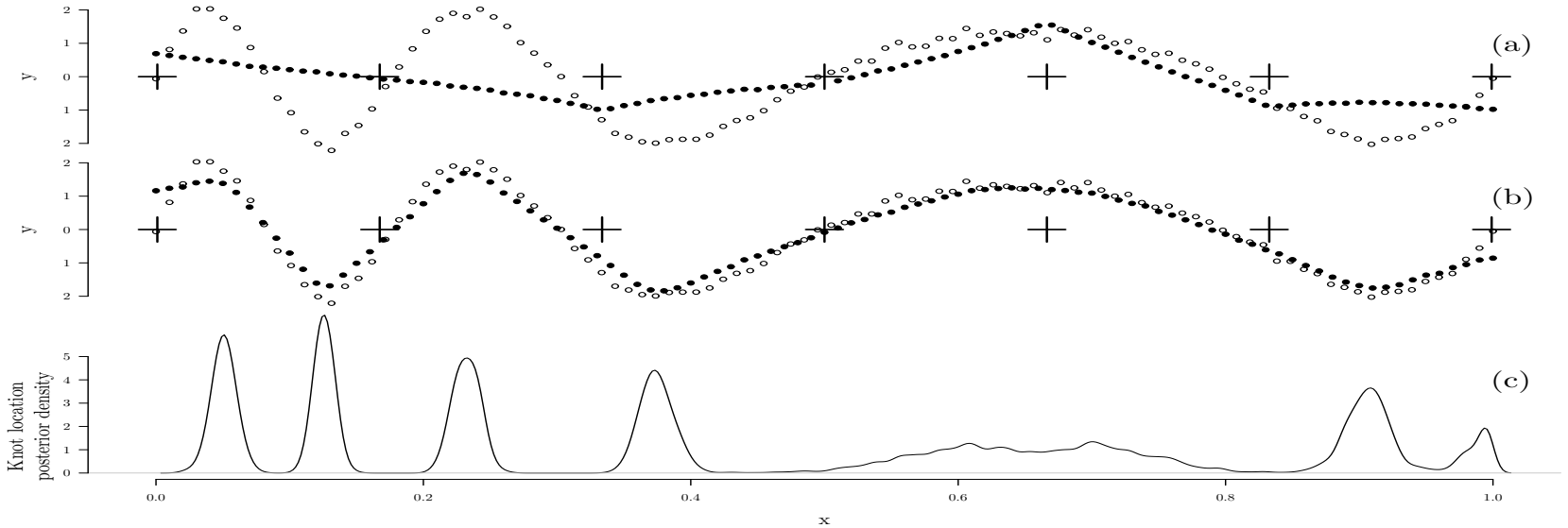
Modified predictive process (Finley et al., 2009)

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

Results

	True	Non-spatial	PP	Modified PP
β_0	8.25	8.26 (8.15 , 8.27)	10.83 (9.29 , 12.60)	9.21 (7.83 , 10.97)
σ^2	6	–	8.95 (2.68 , 15.81)	5.07 (3.44 , 7.32)
τ^2	0.5	3.59 (3.30 , 3.88)	2.20 (2.02 , 2.40)	.73 (.39 , 1.17)
ϕ	4	–	2.78 (2.32 , 3.62)	2.73 (2.23 , 5.38)
G	–	3959.95	2397.21	347.16
P	–	3943.83	2502.70	1471.05
D	–	7903.79	4899.91	1818.22
p_D	–	1.95	31.79	731.42
DIC	–	2509.32	2000.50	1628.88

Knots are Learning from The Data



- Observed data ∘ drawn from normal distribution with varying frequency sine function mean and .01 variance
- knots are assigned U(0,1) prior.

+ naive knot starting locations

• posterior predictive means of 100 new locations

Knot Selection in Predictive Process, Guhaniyogi et al., 2012

Hierarchical adaptive predictive process model with random knots

$$[\boldsymbol{\beta}, \boldsymbol{\theta}_1, \mathscr{S}^*, \boldsymbol{\theta}_2 \,|\, \boldsymbol{y}, \mathscr{S}, n^*] \propto \times[\,\boldsymbol{\theta}_1\,] \times N(\boldsymbol{\beta} \,|\, \boldsymbol{\mu}_\beta, \Sigma_\beta)$$
$$\times N\left(\boldsymbol{y} \,|\, \boldsymbol{X}\boldsymbol{\beta}, \Sigma_{\boldsymbol{y}}\left(\mathscr{S}^*, \mathscr{S}; \boldsymbol{\theta}_1\right)\right) .$$

- Process parameter in data likelihood

Knot Selection in Predictive Process, Guhaniyogi et al., 2012

Hierarchical adaptive predictive process model with random knots

$$[\boldsymbol{\beta}, \boldsymbol{\theta}_1, \mathscr{S}^*, \boldsymbol{\theta}_2 \,|\, \boldsymbol{y}, \mathscr{S}, n^*] \propto [\,\boldsymbol{\theta}_2\,] \times [\,\boldsymbol{\theta}_1\,] \times N(\boldsymbol{\beta} \,|\, \boldsymbol{\mu}_\beta, \Sigma_\beta)$$
$$\times [\mathscr{S}^* \,|\, \boldsymbol{\theta}_2\,] \times N\left(\boldsymbol{y} \,|\, \boldsymbol{X}\boldsymbol{\beta}, \Sigma_{\boldsymbol{y}}\left(\mathscr{S}^*, \mathscr{S}; \boldsymbol{\theta}_1\right)\right) .$$

- Process parameter in data likelihood

Knot Selection in Predictive Process, Guhaniyogi et al., 2012

Hierarchical adaptive predictive process model with random knots

$$[\boldsymbol{\beta}, \boldsymbol{\theta}_1, \mathscr{S}^*, \boldsymbol{\theta}_2 \,|\, \boldsymbol{y}, \mathscr{S}, n^*] \propto [\,\boldsymbol{\theta}_2\,] \times [\,\boldsymbol{\theta}_1\,] \times N(\boldsymbol{\beta} \,|\, \boldsymbol{\mu}_\beta, \Sigma_\beta)$$
$$\times [\mathscr{S}^* \,|\, \boldsymbol{\theta}_2\,] \times N\left(\boldsymbol{y} \,|\, \boldsymbol{X}\boldsymbol{\beta}, \Sigma_{\boldsymbol{y}}\left(\mathscr{S}^*, \mathscr{S}; \boldsymbol{\theta}_1\right)\right) .$$

- Process parameter in data likelihood
- Parameters in the intensity surface for the knots.

Knot Selection in Predictive Process, Guhaniyogi et al., 2012

Hierarchical adaptive predictive process model with random knots

$$[\boldsymbol{\beta}, \boldsymbol{\theta}_1, \mathscr{S}^*, \boldsymbol{\theta}_2 \,|\, \boldsymbol{y}, \mathscr{S}, n^*] \propto [\,\boldsymbol{\theta}_2\,] \times [\,\boldsymbol{\theta}_1\,] \times N(\boldsymbol{\beta} \,|\, \boldsymbol{\mu}_\beta, \Sigma_\beta)$$
$$\times [\mathscr{S}^* \,|\, \boldsymbol{\theta}_2\,] \times N\left(\boldsymbol{y} \,|\, \boldsymbol{X}\boldsymbol{\beta}, \Sigma_{\boldsymbol{y}}\left(\mathscr{S}^*, \mathscr{S}; \boldsymbol{\theta}_1\right)\right).$$

- Process parameter in data likelihood
- Parameters in the intensity surface for the knots.

- Idea of modeling knots is applicable to any knot based low rank model.

Modelling the intensity surface for knots

$$[\mathscr{S}^* \mid \eta_D, n^*] = \prod_{i=1}^{n^*} \frac{\eta(\boldsymbol{s}_i^*)}{\int_D \eta(\boldsymbol{s}) d\boldsymbol{s}}$$

Strategies to Model Knots

Modelling the intensity surface for knots

$$[\mathscr{S}^* \mid \eta_D, n^*] = \prod_{i=1}^{n^*} \frac{\eta(\boldsymbol{s}_i^*)}{\int_D \eta(\boldsymbol{s}) d\boldsymbol{s}}$$

- Parametric option:

$$\log \eta(\boldsymbol{s}; \boldsymbol{\theta}_2) = \frac{1}{m} \sum_{j=1}^{m} N_{2D}(\boldsymbol{s} \mid \boldsymbol{u}_j, \Sigma_\eta) \ ,$$

Modelling the intensity surface for knots

$$[\mathscr{S}^* \mid \eta_D, n^*] = \prod_{i=1}^{n^*} \frac{\eta(\boldsymbol{s}_i^*)}{\int_D \eta(\boldsymbol{s}) d\boldsymbol{s}}$$

- Parametric option:

$$\log \eta(\boldsymbol{s}; \boldsymbol{\theta}_2) = \frac{1}{m} \sum_{j=1}^{m} N_{2D}(\boldsymbol{s} \mid \boldsymbol{u}_j, \Sigma_\eta) \ ,$$

- A log-Gaussian approach:

$$\eta(\boldsymbol{s}) = \exp\{\alpha w_2(\boldsymbol{s})\}; \ w_2(\boldsymbol{s}) \sim GP(0, \rho_2(\cdot; \phi_2)) \ ,$$

Strategies to Model Knots

Modelling the intensity surface for knots

$$[\mathscr{S}^* \mid \eta_D, n^*] = \prod_{i=1}^{n^*} \frac{\eta(\boldsymbol{s}_i^*)}{\int_D \eta(\boldsymbol{s}) d\boldsymbol{s}}$$

- Parametric option:

$$\log \eta(\boldsymbol{s}; \boldsymbol{\theta}_2) = \frac{1}{m} \sum_{j=1}^{m} N_{2D}(\boldsymbol{s} \mid \boldsymbol{u}_j, \Sigma_\eta) \ ,$$

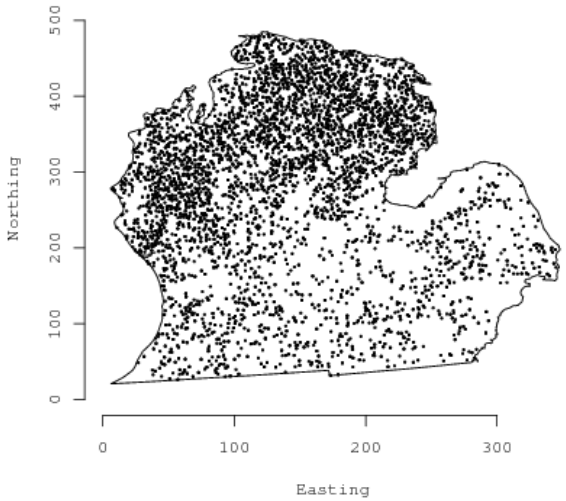
- A log-Gaussian approach:

$$\eta(\boldsymbol{s}) = \exp\{\alpha w_2(\boldsymbol{s})\}; \ w_2(\boldsymbol{s}) \sim GP(0, \rho_2(\cdot; \phi_2)) \ ,$$

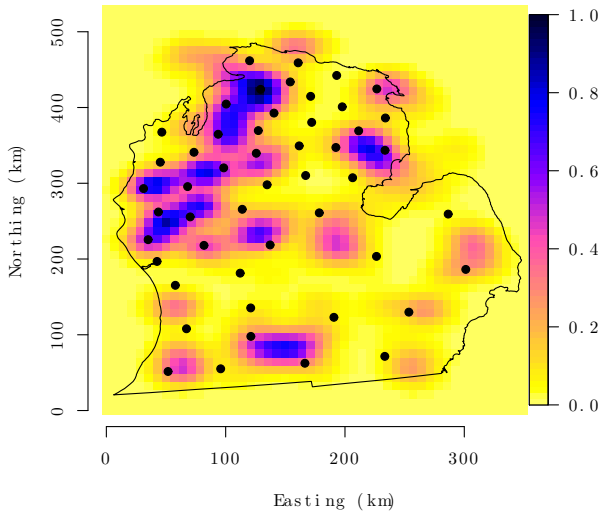
- Any other random measure on $[\mathscr{S}^* \mid n^*]$.

Estimated Spatial Surface: PP vs. APP

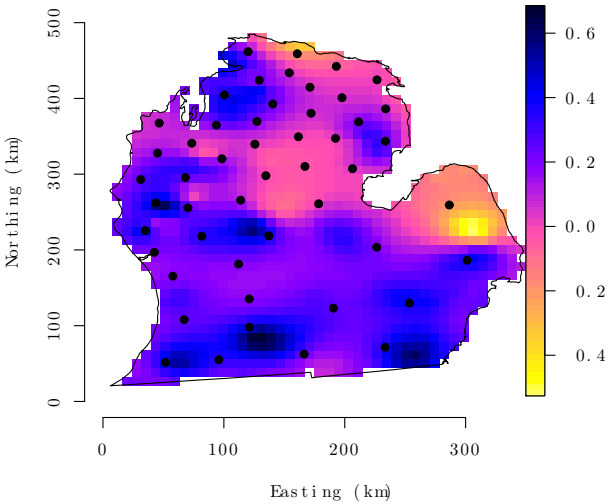
data locations



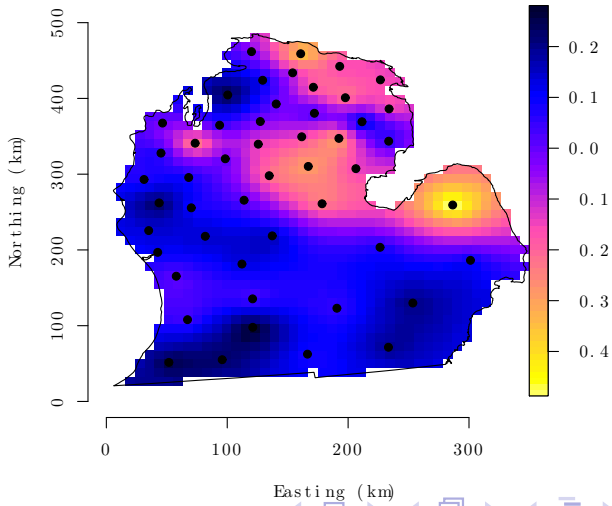
Knot density



APP: residual surface



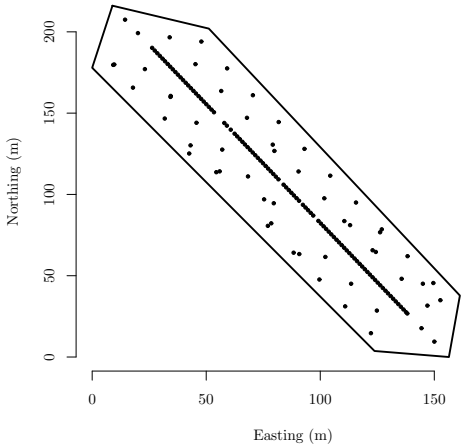
PP: residual surface



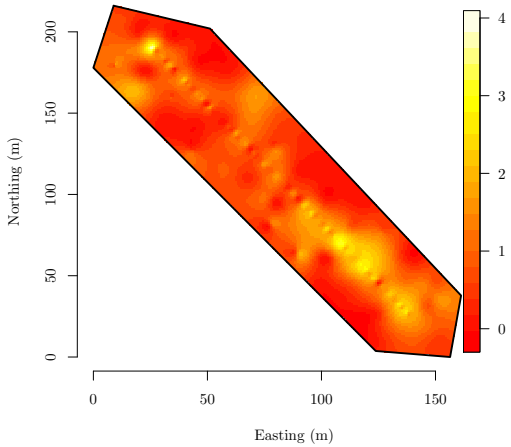
Motivating Dataset on Soil Nutrients

LMC

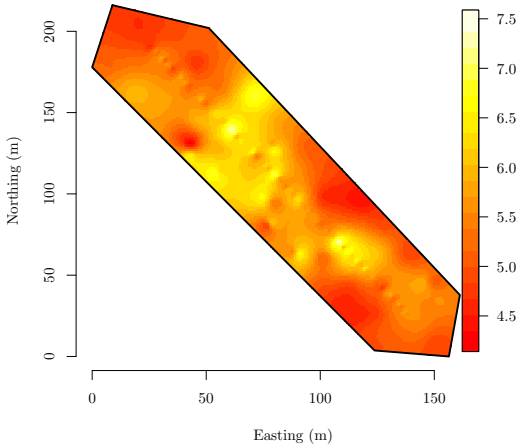
Soil sample locations



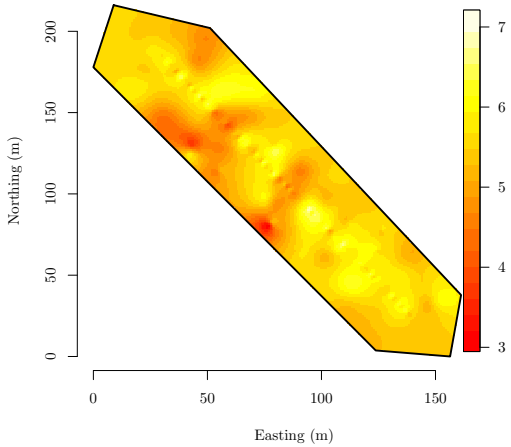
Phosphorus (P)



Sum of base cations (SBC)



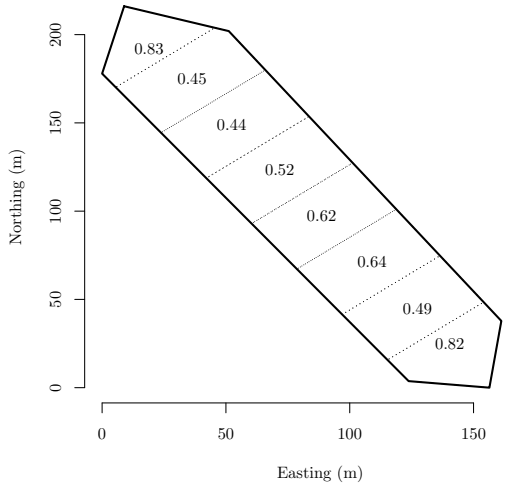
Sum of nitrogen (SN)



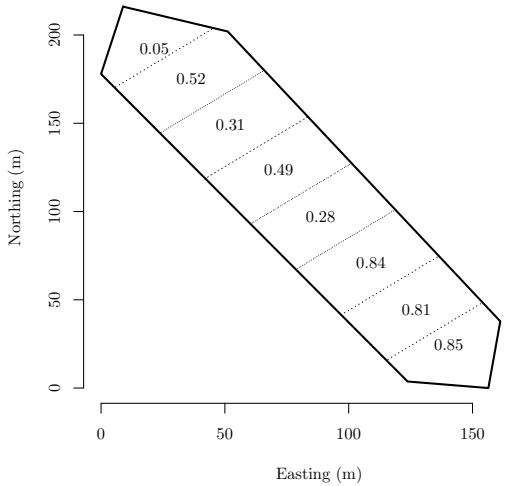
Correlations Between Soil Nutrients are Space Varying

sub-domain correlation among the soil nutrient outcomes

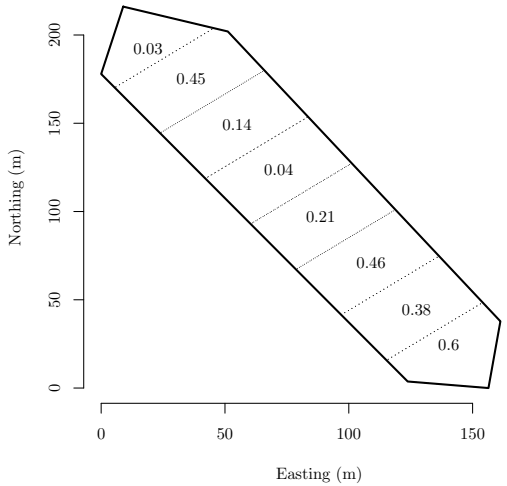
$\rho_{P,SBC}$



$\rho_{P,SN}$



$\rho_{SN,SBC}$



Multivariate Spatial Modeling

▸ Figure:Soil Nutrients

$$\boldsymbol{Y}(\boldsymbol{s}) = \boldsymbol{\mu}(\boldsymbol{s}) + \boldsymbol{W}(\boldsymbol{s}) + \boldsymbol{\epsilon}(\boldsymbol{s})$$

- Multivariate response
- Fixed effects
- Multivariate spatial function
- Multivariate error

$\boldsymbol{Y}(\boldsymbol{s}) = (y_1(\boldsymbol{s}), ..., y_m(\boldsymbol{s}))'$ $\boldsymbol{W}(\boldsymbol{s}) = (w_1(\boldsymbol{s}), ..., w_m(\boldsymbol{s}))'$

$$\boldsymbol{\epsilon}(\boldsymbol{s}) = (\epsilon_1(\boldsymbol{s}), ..., \epsilon_m(\boldsymbol{s}))' \overset{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Psi})$$

In our case $m = 3$

Linear Model Coregionalization

Constructive approach following factor analysis ideas:

$$\begin{pmatrix} w_1(\boldsymbol{s}) \\ \vdots \\ \vdots \\ w_m(\boldsymbol{s}) \end{pmatrix} = \begin{pmatrix} a_{11} & 0 & \cdots & \cdots & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \cdots & \cdots & \cdots & a_{mm} \end{pmatrix} \begin{pmatrix} v_1(\boldsymbol{s}) \\ \vdots \\ \vdots \\ v_m(\boldsymbol{s}) \end{pmatrix}$$

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

controls correlation in w_i's

$v_i(\cdot) \overset{ind}{\sim} GP(\cdot, \cdot; \boldsymbol{\theta}_1)$

Space Varying Linear Model Coregionalization

$$\begin{pmatrix} w_1(\boldsymbol{s}) \\ \vdots \\ \vdots \\ w_m(\boldsymbol{s}) \end{pmatrix} = \begin{pmatrix} a_{11}(\boldsymbol{s}) & 0 & \cdots & \cdots & 0 \\ a_{21}(\boldsymbol{s}) & a_{22}(\boldsymbol{s}) & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1}(\boldsymbol{s}) & \cdots & \cdots & \cdots & a_{mm}(\boldsymbol{s}) \end{pmatrix} \begin{pmatrix} v_1(\boldsymbol{s}) \\ \vdots \\ \vdots \\ v_m(\boldsymbol{s}) \end{pmatrix}$$

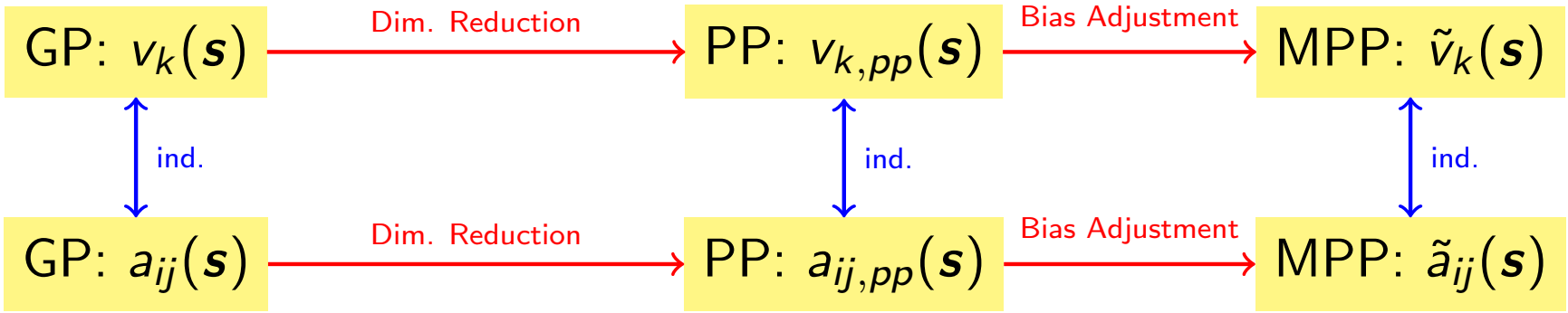
$\boldsymbol{W}(\boldsymbol{s})$

$a_{ij}(\cdot) \overset{ind}{\sim} GP(\cdot, \cdot; \boldsymbol{\theta}_{1,a})$

$v_i(\cdot) \overset{ind}{\sim} GP(\cdot, \cdot; \boldsymbol{\theta}_1)$

- Big-N problem is encountered in updating $\boldsymbol{v}_i = (v_i(\boldsymbol{s}_1), ..., v_i(\boldsymbol{s}_n))'$, $\boldsymbol{a}_{ij} = (a_{ij}(\boldsymbol{s}_1), ..., a_{ij}(\boldsymbol{s}_n))'$.

Dimension Reduction



Dimension Reduced Model

$$\tilde{\boldsymbol{W}}(\boldsymbol{s}) = \tilde{\boldsymbol{A}}(\boldsymbol{s})\tilde{\boldsymbol{v}}(\boldsymbol{s})$$

$$\begin{pmatrix} \tilde{w}_1(\boldsymbol{s}) \\ \vdots \\ \vdots \\ \tilde{w}_m(\boldsymbol{s}) \end{pmatrix} = \begin{pmatrix} \tilde{a}_{11}(\boldsymbol{s}) & 0 & \cdots & \cdots & 0 \\ \tilde{a}_{21}(\boldsymbol{s}) & \tilde{a}_{22}(\boldsymbol{s}) & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \tilde{a}_{m1}(\boldsymbol{s}) & \cdots & \cdots & \cdots & \tilde{a}_{mm}(\boldsymbol{s}) \end{pmatrix} \begin{pmatrix} \tilde{v}_1(\boldsymbol{s}) \\ \vdots \\ \vdots \\ \tilde{v}_m(\boldsymbol{s}) \end{pmatrix}$$

$\tilde{\boldsymbol{W}}(\boldsymbol{s})$

$\tilde{a}_{ij}(\cdot) \overset{ind}{\sim} MPP(\cdot, \cdot; \boldsymbol{\theta}_{1,a})$

$\tilde{v}_i(\cdot) \overset{ind}{\sim} MPP(\cdot, \cdot; \boldsymbol{\theta}_1)$

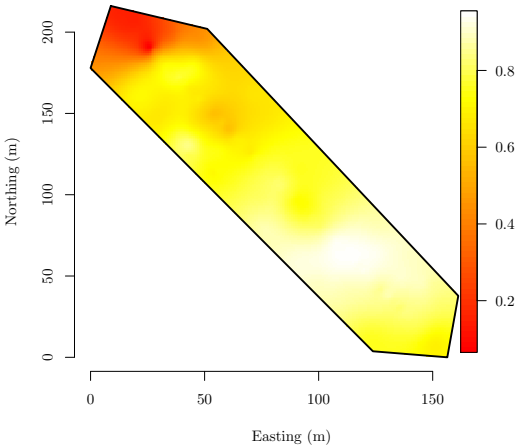
Soil Nutrient Data Analysis

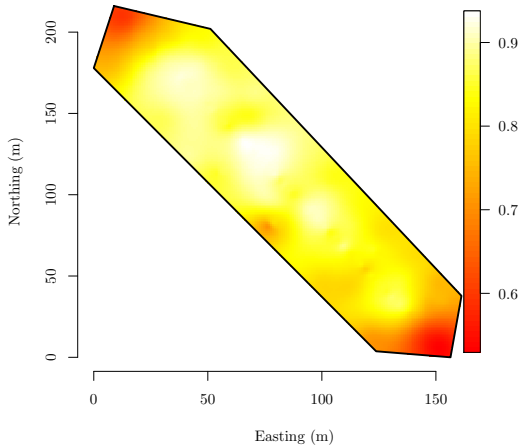
1. Modified predictive process models with varying number of knots are fitted.

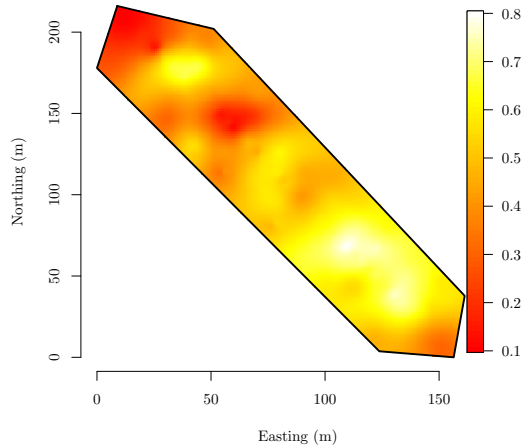
	Stationary	Non-stationary full	Non-stationary MPP 26 knots
G	39.45	28.02	24.4
P	92.62	79.9	77.28
D	132.07	107.92	101.68

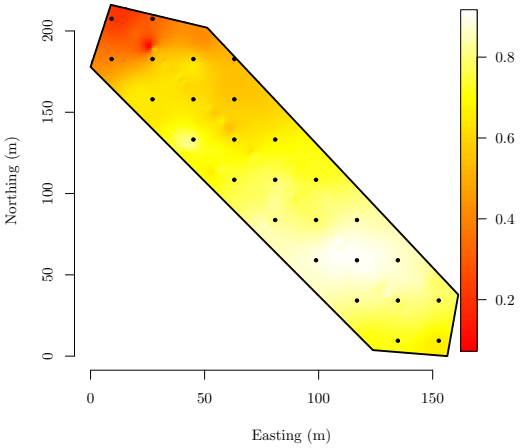
Advantages over space varying full model

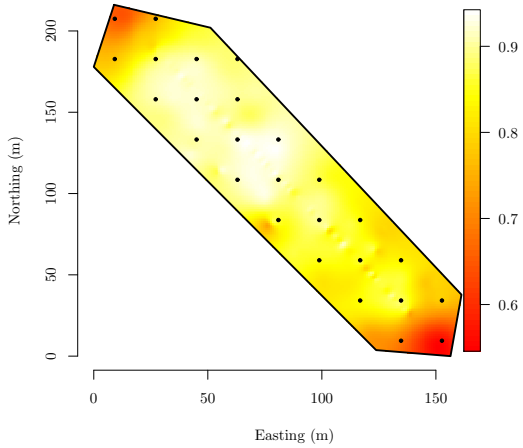
1. 5 hours of running time for modified predictive process, 21 hours of running time for full model.
2. Better model fit.

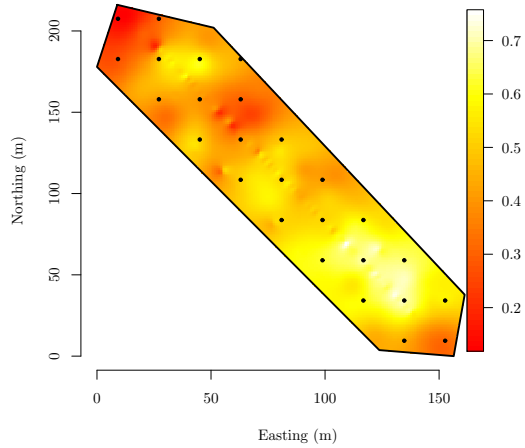

Northing (m)
0
50
100
150
200
Easting (m)
0
50
100
150
0.8
0.6
0.4
0.2

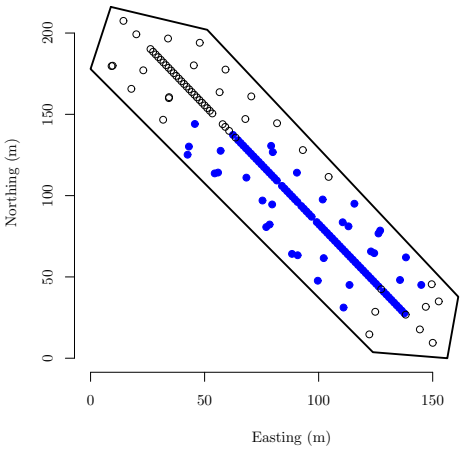

Northing (m)
0
50
100
150
200
Easting (m)
0
50
100
150
0.9
0.8
0.7
0.6

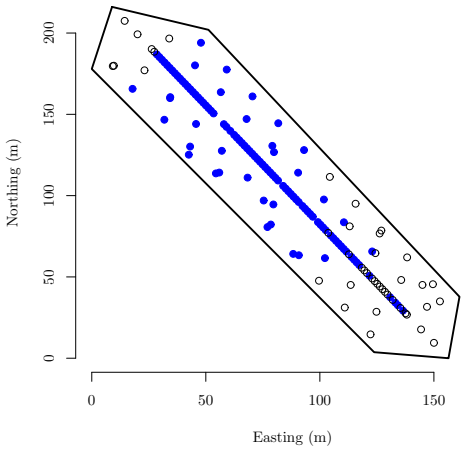

Northing (m)
0
50
100
150
200
Easting (m)
0
50
100
150
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1

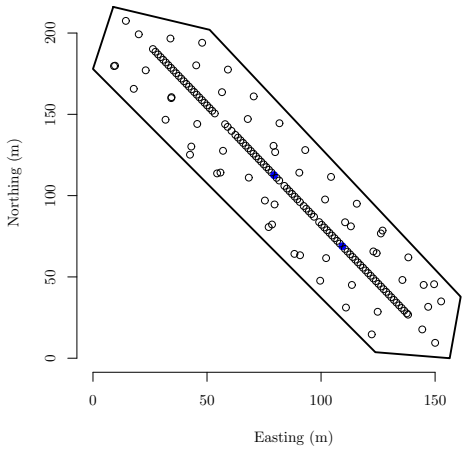

Northing (m)
0
50
100
150
200
Easting (m)
0
50
100
150
0.8
0.6
0.4
0.2


Northing (m)
0
50
100
150
200
Easting (m)
0
50
100
150
0.9
0.8
0.7
0.6


Northing (m)
0
50
100
150
200
Easting (m)
0
50
100
150
0.7
0.6
0.5
0.4
0.3
0.2







- Observed points with statistically significant correlations given in blue.
- Left to right $\rho(\boldsymbol{s})_{P,SBC}$, $\rho(\boldsymbol{s})_{P,SN}$ and $\rho(\boldsymbol{s})_{SN,SBC}$.
- P, SN are highly correlated; so are P and SBC.