NC STATE UNIVERSITY

ST533 Midterm 2 Group Project

Bayesian Predictive Process

Ziyang Jiang, Tyler Schappe, Mingkang Zheng

Predictive Process (PP) approach

Spatial regression model: $Y(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \underline{w(\boldsymbol{s})} + \varepsilon(\boldsymbol{s})$

spatial process (parent process)

↓ Replace $w(\boldsymbol{s})$ with $\tilde{w}(\boldsymbol{s})$

Predictive process model: $Y(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \underline{\tilde{w}(\boldsymbol{s})} + \varepsilon(\boldsymbol{s})$

predictive process

- $\tilde{w}(\boldsymbol{s})$ is <u>approximation</u> of $w(\boldsymbol{s})$
- $\tilde{w}(\boldsymbol{s})$ is derived from $w(\boldsymbol{s})$

How to get $\tilde{w}(\boldsymbol{s})$ *?*

Select "knots"

N observed locations: $\boldsymbol{s} = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \dots, \boldsymbol{s}_N\}$

↓ Lower-dimensional subspace($K << N$)

K "Knot" locations: $\boldsymbol{s}^\star = \{\boldsymbol{s}_1^\star, \boldsymbol{s}_2^\star, \dots, \boldsymbol{s}_K^\star\}$

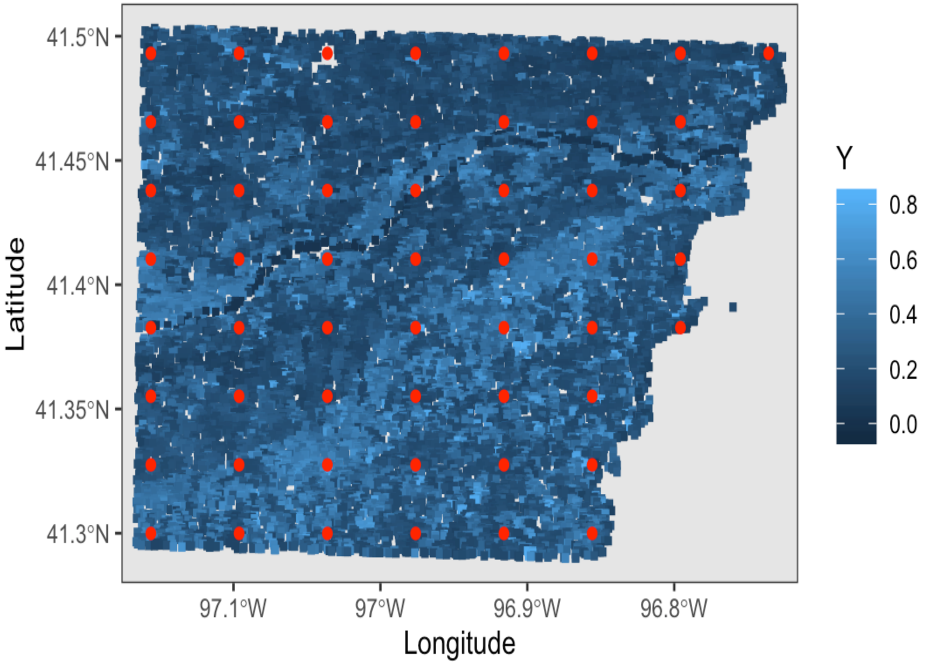
- Knot structure

Regular grids(which we will use), Lattice plus close pairs, Lattice plus infill, etc.

$\boldsymbol{s}^\star$ may or may not form a subset of the entire collection of $\boldsymbol{s}$

- Knot sizes (K)

Consider computational cost and sensitivity to choice of K.



Regular grids structure

Mathematical derivation of $\widetilde{w}(\boldsymbol{s})$

- Parent process $w(\boldsymbol{s}) \sim GP\ \{\ 0, \mathbb{C}(\boldsymbol{s}, \boldsymbol{s}')\ \}$, where $\mathbb{C}(\boldsymbol{s}, \boldsymbol{s}') = \sigma_w^2 \cdot \rho(\boldsymbol{s}, \boldsymbol{s}')$ (parent cov function)
- Realization of parent process over the knots($\boldsymbol{s}^\star$) is: $\boldsymbol{w}^\star = (\ w(\boldsymbol{s}_1^\star), w(\boldsymbol{s}_2^\star), \ldots, w(\boldsymbol{s}_K^\star)\ )'$
- $\boldsymbol{w}^\star \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{w}^\star})$, where $\boldsymbol{\Sigma}_{\boldsymbol{w}^\star} = \begin{bmatrix} \mathbb{C}(\boldsymbol{s}_1^\star, \boldsymbol{s}_1^\star) & \cdots & \mathbb{C}(\boldsymbol{s}_1^\star, \boldsymbol{s}_K^\star) \\ \vdots & \ddots & \vdots \\ \mathbb{C}(\boldsymbol{s}_K^\star, \boldsymbol{s}_1^\star) & \cdots & \mathbb{C}(\boldsymbol{s}_K^\star, \boldsymbol{s}_K^\star) \end{bmatrix}$ is a K x K covariance matrix

Spatial interpolant (that leads to kriging)

Use realization of $w(\boldsymbol{s})$ over K knots($s^\star$) to estimate the realization of $w(\boldsymbol{s})$ over N observed locations($\boldsymbol{s}$)

$$w(\boldsymbol{s}) \approx \widetilde{w}(\boldsymbol{s}) = E[w(\boldsymbol{s})|\boldsymbol{w}^\star] = \mathbb{C}'(\boldsymbol{s}, \boldsymbol{s}^\star)\boldsymbol{\Sigma}_{\boldsymbol{w}^\star}^{-1}\boldsymbol{w}^\star$$

- Here $\mathbb{C}(\boldsymbol{s}, \boldsymbol{s}^\star) = (\ \mathbb{C}(\boldsymbol{s}, \boldsymbol{s}_1^\star), \mathbb{C}(\boldsymbol{s}, \boldsymbol{s}_2^\star), \ldots, \mathbb{C}(\boldsymbol{s}, \boldsymbol{s}_K^\star)\ )'$
- $\widetilde{w}(\boldsymbol{s})$ is a spatially varying linear transformation of $\boldsymbol{w}^\star$

Predictive Process (PP) approach

Predictive process model: $Y(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \underline{\widetilde{w}(\boldsymbol{s})} + \varepsilon(\boldsymbol{s})$

- Predictive process $\widetilde{w}(\boldsymbol{s}) = \mathbb{C}'(\boldsymbol{s}, \boldsymbol{s}^\star)\boldsymbol{\Sigma}_{\boldsymbol{w}^\star}^{-1}\boldsymbol{w}^\star$ *$\widetilde{w}(\boldsymbol{s})$ is completely determined by $\mathbb{C}(\cdot,\cdot)$ and $\boldsymbol{s}^\star$
- $\widetilde{w}(\boldsymbol{s}) \sim GP\ \{\, 0,\ \tilde{\mathbb{C}}(\cdot)\, \}$, where $\tilde{\mathbb{C}}(\cdot) = \mathbb{C}'(\boldsymbol{s}, \boldsymbol{s}^\star)\boldsymbol{\Sigma}_{\boldsymbol{w}^\star}^{-1}\mathbb{C}'(\boldsymbol{s}, \boldsymbol{s}^\star)$
- Attractive advantages

  1. The PP approach is unaltered when considering modeling complexities such as anisotropy, non-stationarity or multivariate processes.

  2. Only need to calculate the inverse and determinant of a $K \times K$ matrix, which Results in massive computational savings when *K*<<*N*.

- What can be improved?

Modified Predictive Process

- According to the Finley(2009) paper, the predictive process $\tilde{w}(s)$ underestimates the variance of the parent process $w(s)$

$$\mathbb{V}ar\left(\tilde{w}(\boldsymbol{s})\right) = \mathbb{C}'(\boldsymbol{s},\boldsymbol{s}^{\star})\boldsymbol{\Sigma}_{\boldsymbol{w}^{\star}}^{-1}\mathbb{C}'(\boldsymbol{s},\boldsymbol{s}^{\star}) \boxed{\leq} \sigma_w^2 = \mathbb{V}ar(w(\boldsymbol{s}))$$

Add $\xi(\boldsymbol{s}) \overset{independent}{\sim} \mathcal{N}(\,0, \sigma_w^2 - \mathbb{C}'(\boldsymbol{s},\boldsymbol{s}^{\star})\boldsymbol{\Sigma}_{\boldsymbol{w}^{\star}}^{-1}\mathbb{C}'(\boldsymbol{s},\boldsymbol{s}^{\star})\,)$

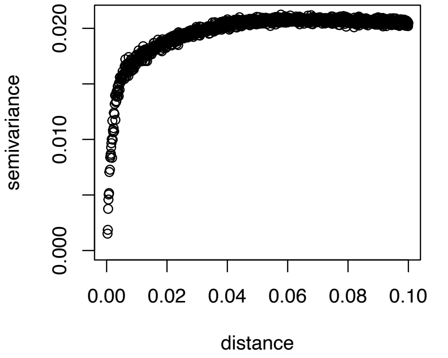
Make sure $\mathbb{V}ar\left(\tilde{w}(s) + \xi(\boldsymbol{s})\right) \boxed{=} \sigma_w^2 = \mathbb{V}ar(w(\boldsymbol{s}))$

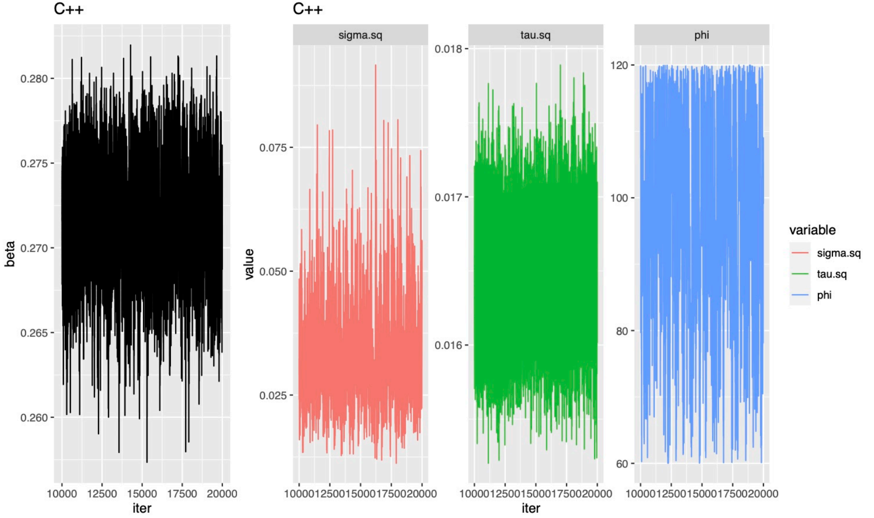
- “Modified” predictive process model: $Y(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \tilde{w}(\boldsymbol{s}) + \xi(\boldsymbol{s}) + \varepsilon(\boldsymbol{s})$

In the following analysis, “modified” predictive process will be used.

Implementation

1. spBayes package in R
   - Finley, Banerjee, and Gelfand
   - spLM function
     - Univariate, Gaussian process
     - Expontential, Matern covariance functions
     - (Modified) predictive process models
   - Gibbs sampler with Metropolis steps
     - Marginalized likelihood, integrate out ***w****
2. Custom C++ Script from Heaton paper Github
   - OpenMP multithreading
   - MUCH faster run times (4-5x)
   - **Untrustworthy results!**
   - Did not pursue further

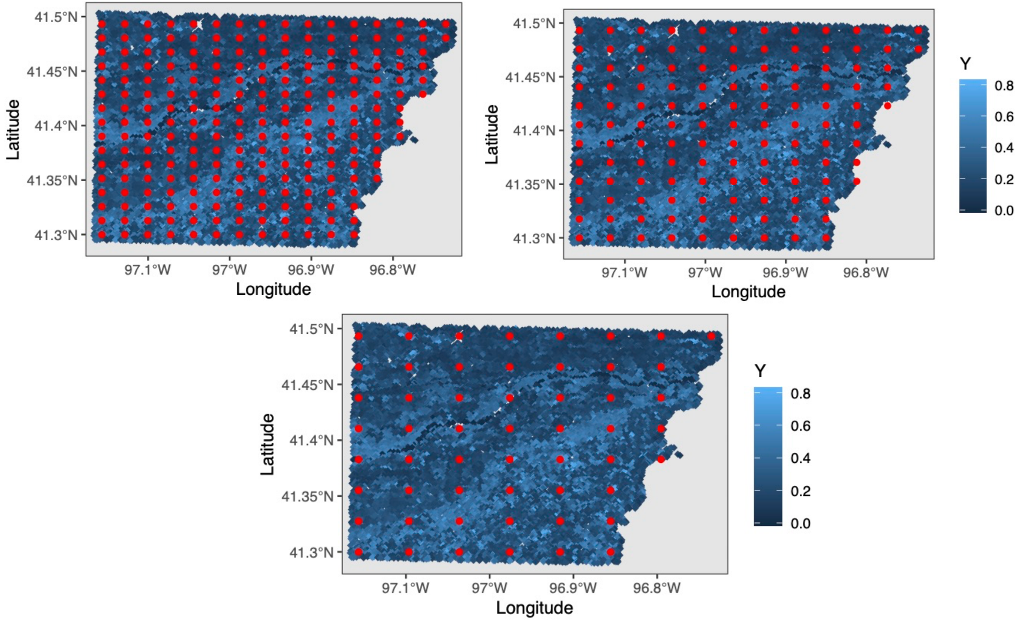




Method Comparison Setup

Method	Covariance Function	Number of knots
Bayesian Predictive Process	Matern	64
		144
		256
	Exponential	64
		256
MLE	Matern	-
	Exponential	-

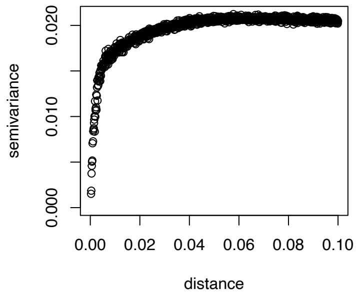
- Knots
  - Regular grid
    - 8 x 8
    - 12 x 12
    - 16 x 16
  - Eliminated knots outside of spatial range

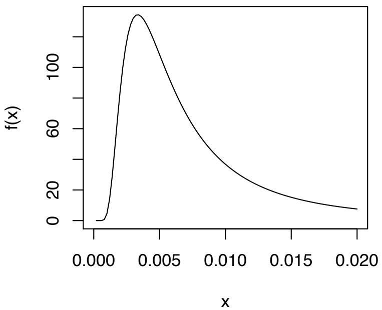


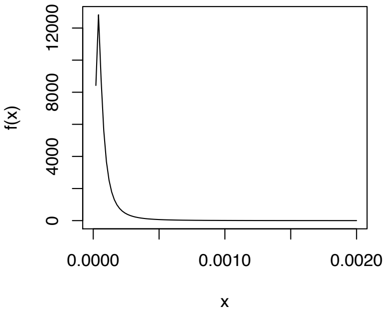
spBayes Implementation Details

- **Model fitting: spLM**
  - Predictors: Longitude + Latitude (centered)
  - Starting values
    - ϕ : 60 (inverse)
    - σ^2: 0.05 Var[Y_{Full}]
    - τ^2: 0.05 Var[Y_{full}]
    - ν: 0.5
  - Tuning parameters
    - ϕ : 3
    - σ^2: 0.1
    - τ^2: 0.1
    - ν: 0.1
  - Priors
    - $\beta_p \sim MVNormal(\mu, \Sigma)$
    - $\mu = [0, 0, 0]$
    - $\Sigma = \begin{bmatrix} 100 & 0 & 0 \\ 0 & 100 & 0 \\ 0 & 0 & 100 \end{bmatrix}$
    - $\phi \sim Uniform\left(\frac{1}{2*\max[dist]}, \frac{1}{0.01*\max[dist]}\right)$
    - $\sigma^2 \sim InvGamma(2, 0.01)$
    - $\tau^2 \sim InvGamma(2, 0.0001)$
    - $\nu^2 \sim Uniform(0.001, 2)$
  - Multicore implementation
    - 'doparallel' package
    - Multiple models simultaneously

- **Prediction: spPredict**
  - Complete uncertainty quantification
  - Propagates parameter uncertainty to predictions
  - Took longer than model fitting
    - Up to 8 hours
- 95% highest density intervals

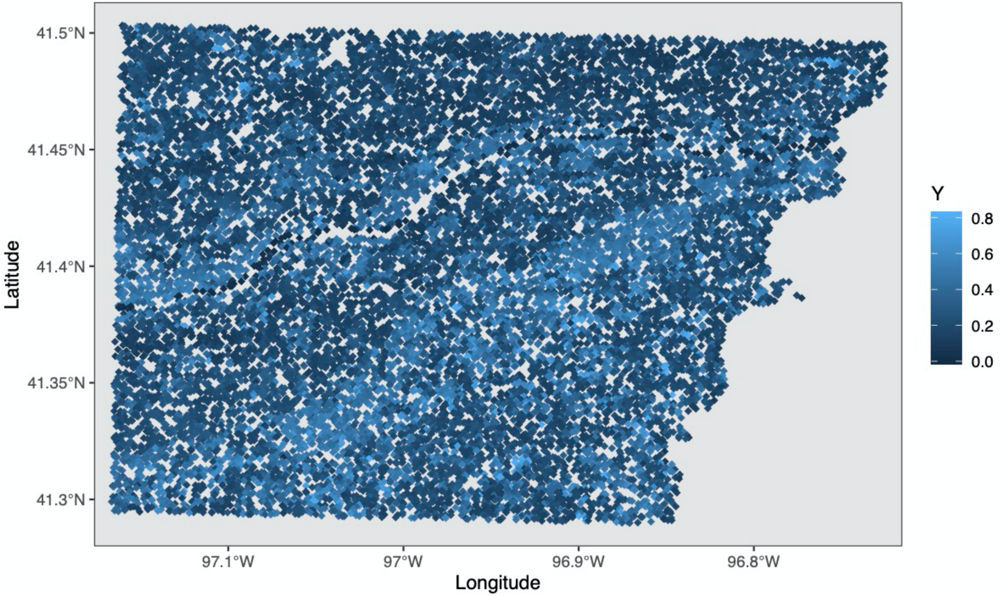






MLE Implementation Details

- **Model fitting: likfit**
  - Predictors: Longitude + Latitude
  - Starting values
    - **ϕ**: 0.05/3
    - $\boldsymbol{\sigma}^2$: 0.02
    - $\boldsymbol{\tau}^2$: 0
- **Prediction: krige.conv**
  - Inputs:
    - MLE estimates
    - Prediction locations
    - Covariates at prediction locations
  - ”Plug-in” Kriging predictions
- Time consuming!
  - Only able to complete up to n=6000

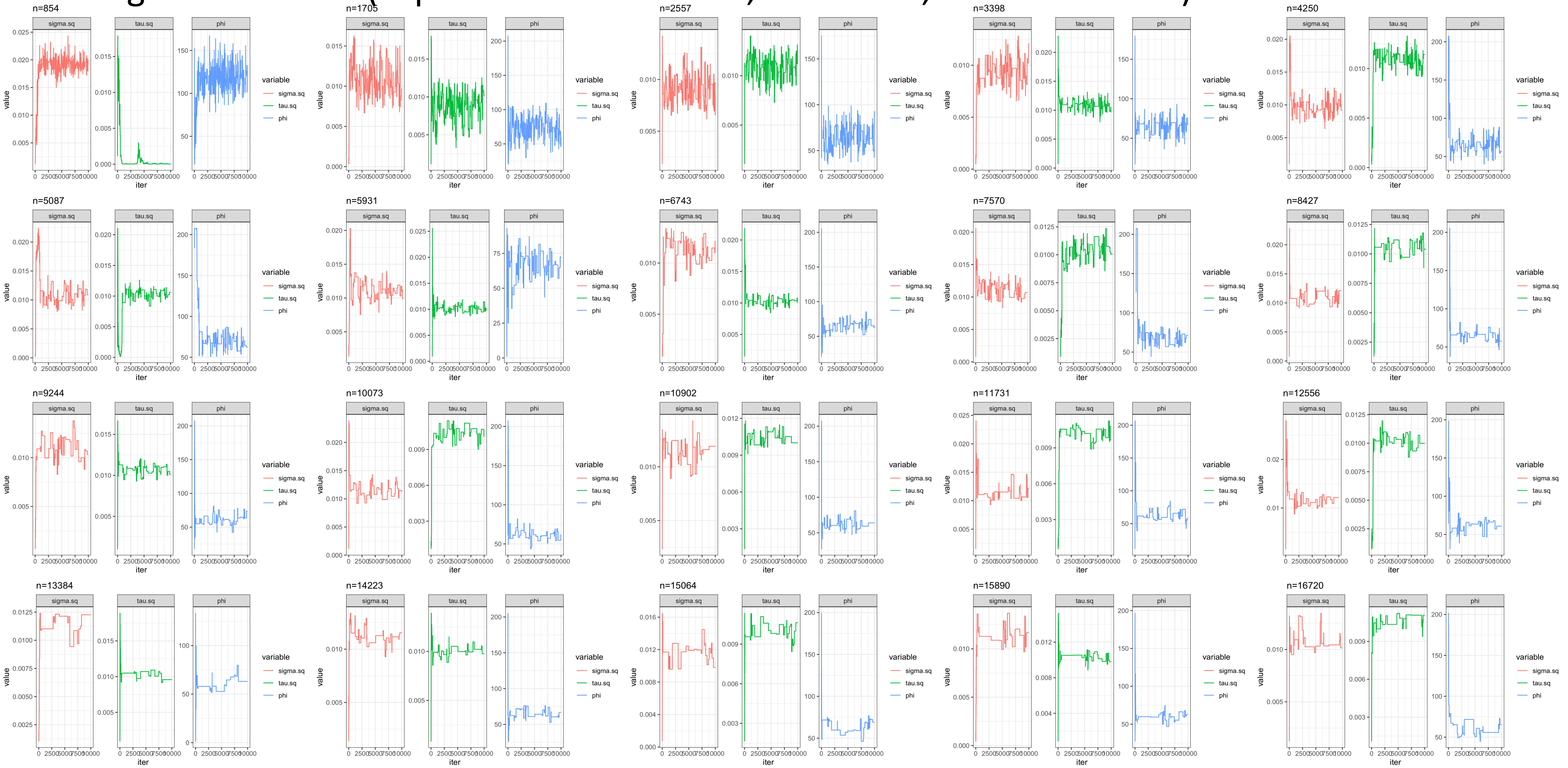


Results

- Convergence of parameters
- Comparison of MSE
- Comparison of run time of model fitting

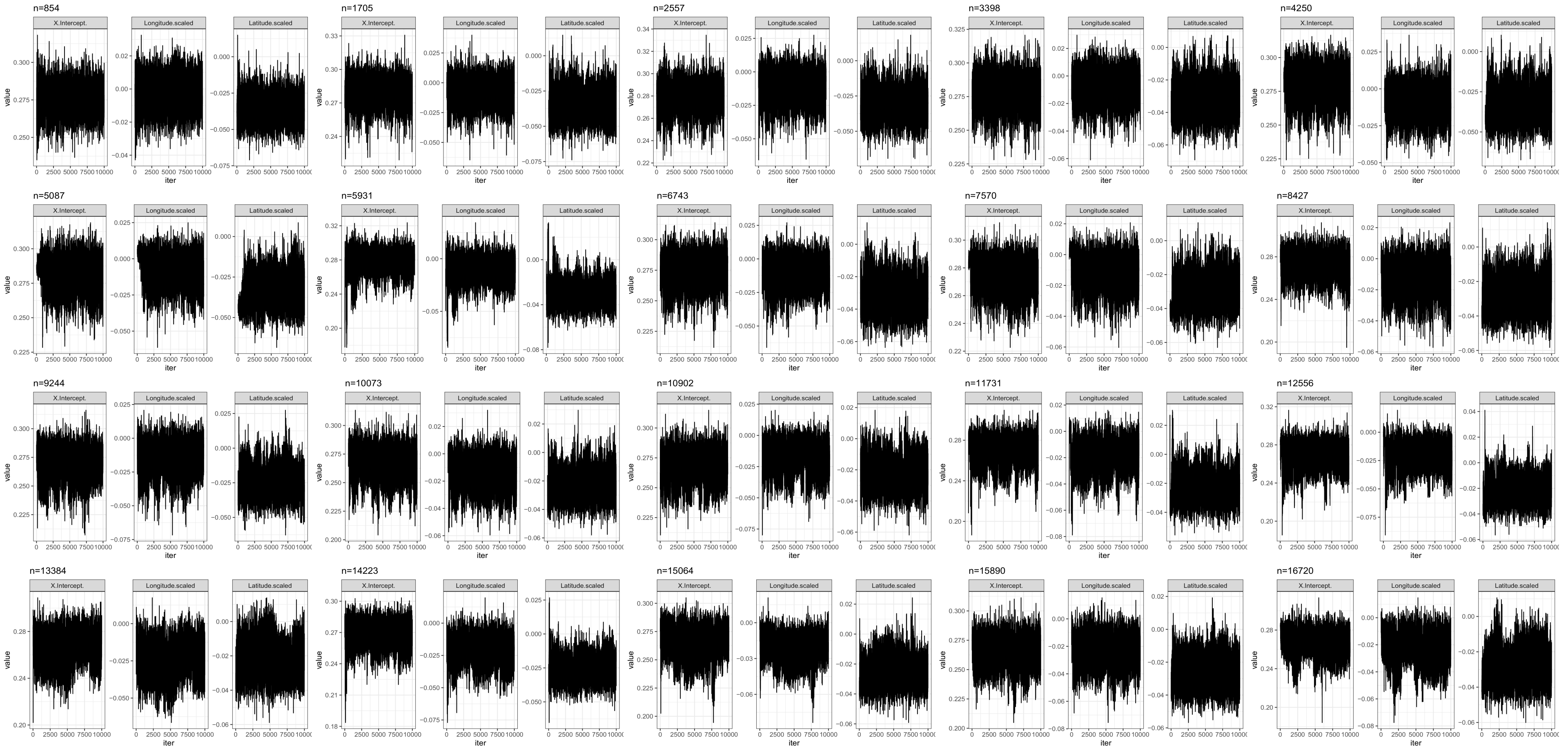
Mingkang Zheng

Convergence of theta (exponential covariance, 256 knots, 5000 iterations)

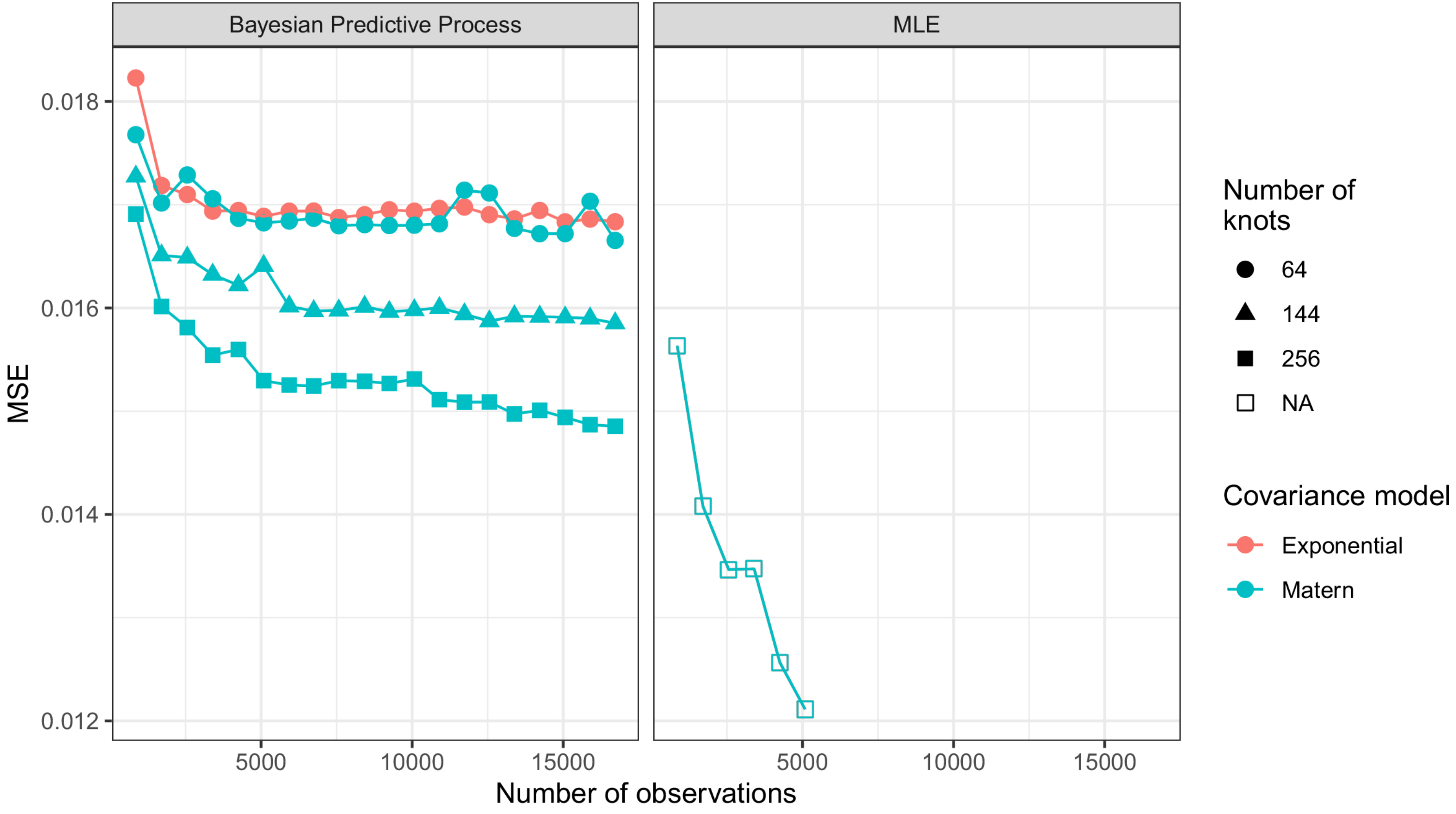


Mingkang Zheng

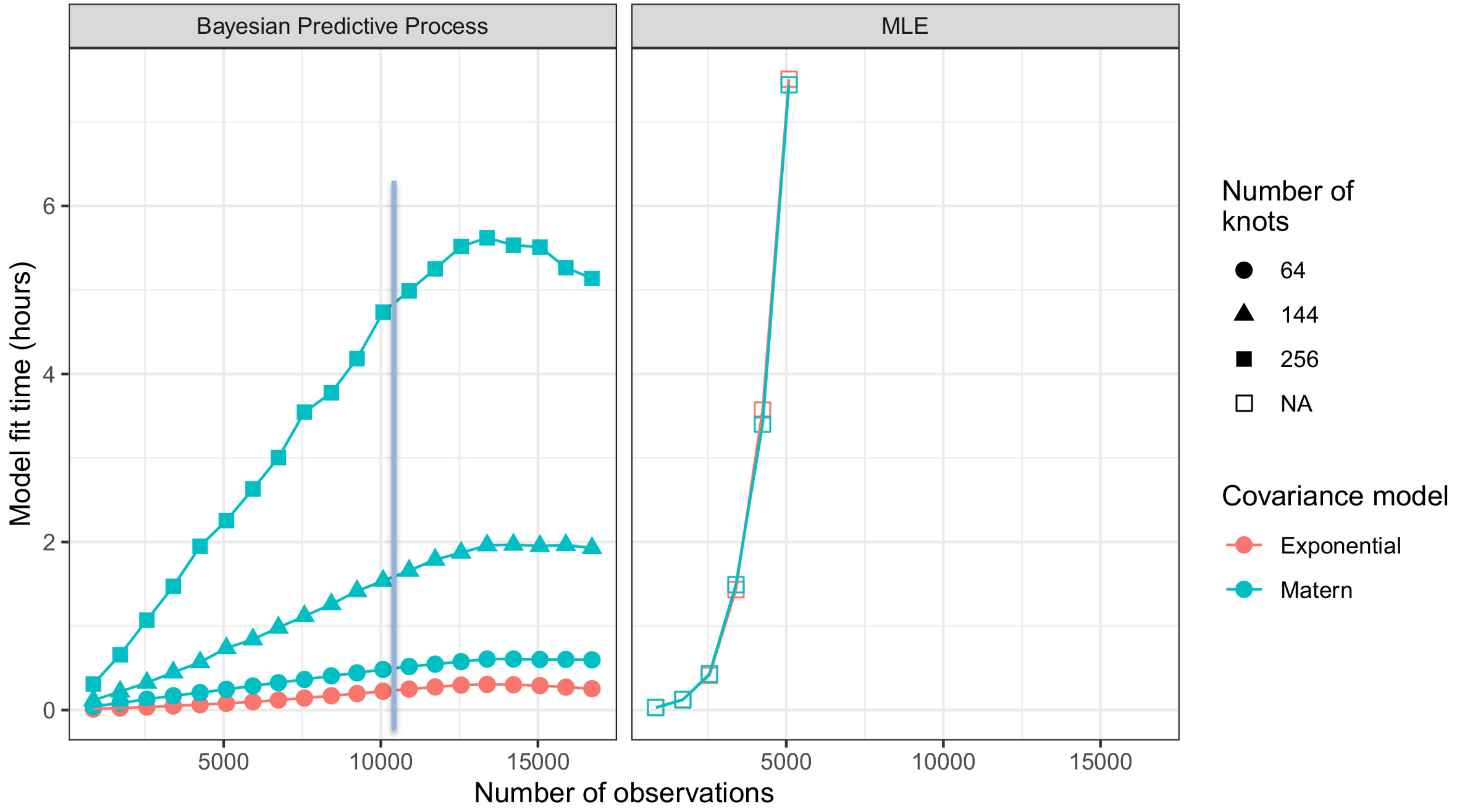
Convergence of beta (exponential covariance, 256 knots, 5000 iterations)



Comparison of MSE—performance of Kriging prediction



Comparison of run time of model fitting



References

Banerjee, S., Gelfand, A.E., Finley, A.O. and Sang, H., 2008. Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(4), pp.825-848.

Finley, A.O., Banerjee, S. and Gelfand, A.E., 2013. spBayes for large univariate and multivariate point-referenced spatio-temporal data models. *arXiv preprint arXiv:1310.8192*.

Finley, A.O., Sang, H., Banerjee, S. and Gelfand, A.E., 2009. Improving the performance of predictive process modeling for large datasets. *Computational statistics & data analysis*, *53*(8), pp.2873-2884.