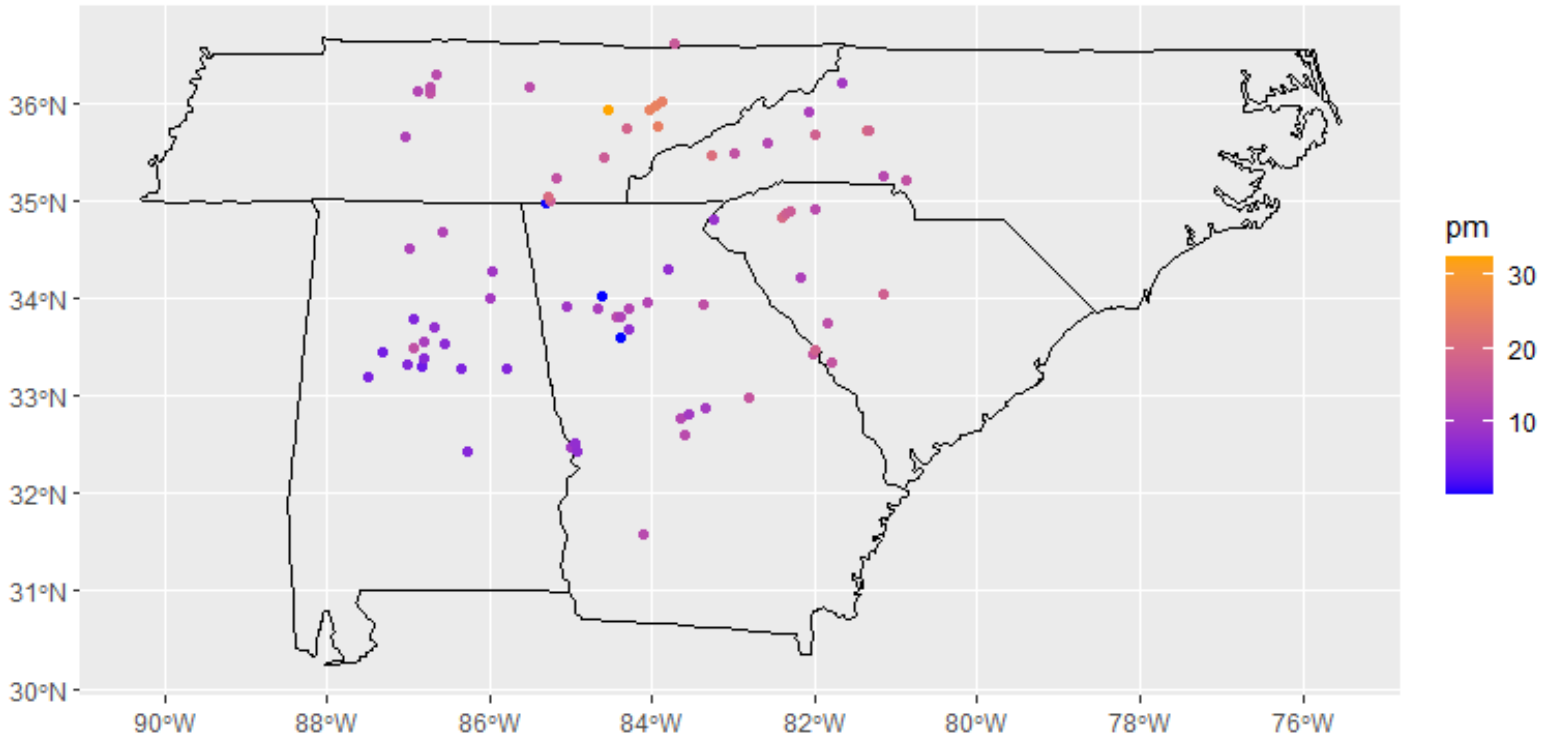
Lecture 6: Modeling Point-Referenced Data

July 10, 2023

Point-Referenced Spatial Data

Average $PM_{2.5}$ concentrations ($\mu g/m^3$) across 78 monitors (2001-2003)

- Observations are associated with point-level locations.



Key Concepts

- Covariance Function
- Gaussian Process
- Geostatistical Regression Model
- Kriging and Spatial Prediction
- Bayesian Kriging

Regression Model for Point-Referenced Spatial Data

We can model the point-level data using linear regression:

$$Y(\mathbf{s}) = \boldsymbol{x}^T(\mathbf{s})\boldsymbol{\beta} + \epsilon(\mathbf{s}) \ .$$

- $\mathbf{s}$ is the spatial location (usually some projected x-y coordinates).
- $Y(\mathbf{s})$ is the observed response variable at location $\mathbf{s}$.
- $\boldsymbol{x}(\mathbf{s})$ is a vector of regression covariates that varies across $\mathbf{s}$.
- $\epsilon(\mathbf{s})$ is the residual variability not explained by $\boldsymbol{x}(\mathbf{s})$.

Goals:

1. Make inference on $\boldsymbol{\beta}$.
2. Predict $Y(\mathbf{s}^*)$ at a new location $\mathbf{s}^*$.

New methods are needed because $\epsilon(\mathbf{s})$'s are correlated across $\mathbf{s}$.

Spatial Stochastic Process

Let $\mathbf{s} = (s_1, s_2)$ denote the spatial location. We are interested in a spatial variable $Y(\mathbf{s})$ over a continuous domain $\mathbf{s} \in \mathbb{R}^2$.

At each location $\mathbf{s}$, $Y(\mathbf{s})$ is a univariate random variable. If we consider a finite sample of spatial locations $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n$, the joint distribution

$$[Y(\mathbf{s}_1), Y(\mathbf{s}_2), \ldots, Y(\mathbf{s}_n)]'$$

is a random vector that incorporates spatial dependency.

- Sets of different spatial locations give different *joint distributions*.
- The observed vector $\mathbf{y} = [y(\mathbf{s}_1), y(\mathbf{s}_2), \ldots, y(\mathbf{s}_n)]'$ represents a single realization from the joint distribution.

In practice, we need some restrictions on the dependency structure of $Y(\mathbf{s})$ across locations for valid inference. It is impossible to consider all possible pairs of locations!

Stationarity and Isotropy

A random field $Y(\mathbf{s})$ is **second-order stationary** if two conditions are satisfied: for all locations $\mathbf{s}$,

1. $E[\,Y(\mathbf{s})\,] = \mu$. constant mean
2. $C(\mathbf{s}+\mathbf{h}, \mathbf{s}) = C(\mathbf{h})$ for any $\mathbf{h} \in \mathbb{R}^2$.

Condition 2 states that the covariance between two locations only depends on the difference in space. It can be re-written as

$$C(\mathbf{s}_1, \mathbf{s}_2) = C(\mathbf{s}_1 - \mathbf{s}_2).$$

Furthermore, a stationary random field is called **isotropic** if the covariance only depends on the distance in space

$$C(\mathbf{s}_1, \mathbf{s}_2) = C(||\mathbf{s}_1 - \mathbf{s}_2||).$$

Isotropic and Stationary Covariance Functions

The spatial dependency is captured by $C(||\mathbf{h}||)$. However only certain functions can provide valid covariance matrices for all finite-dimensional distributions of a random field.

1. $C(0) \geq 0$ because the variance at any $\mathbf{s}$ must be ≥ 0.
2. Positive definite: for any non-zero $n \times 1$ vector $\mathbf{w}$,

$$\sum_i \sum_j w_i w_j C(||\mathbf{s}_i - \mathbf{s}_j||) \geq 0 \ .$$

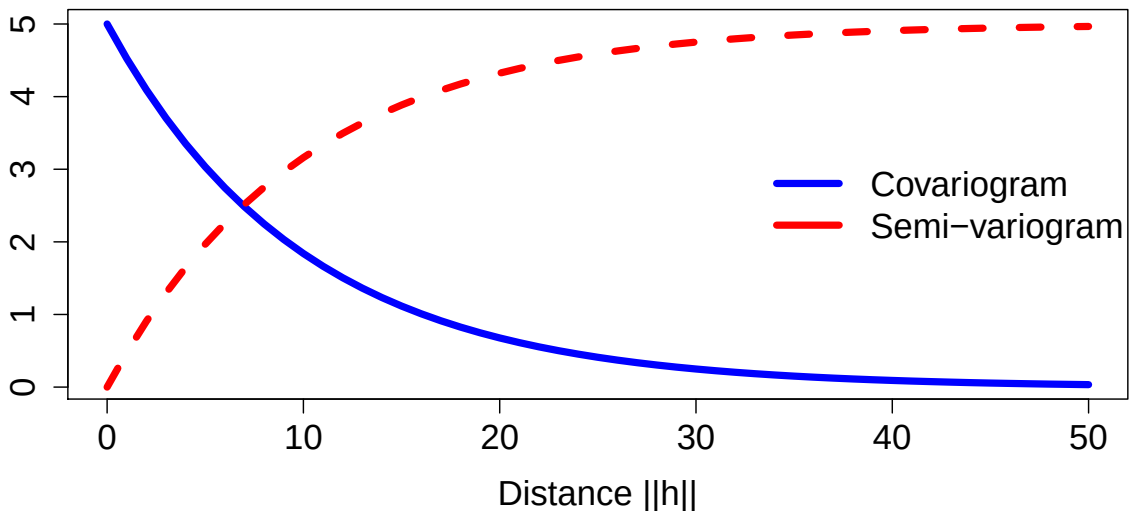
   This is the same as $\mathbf{w}'\Sigma\mathbf{w} \geq 0$, where Σ is the covariance matrix of $Y(\mathbf{s}_1), Y(\mathbf{s}_2), \ldots, Y(\mathbf{s}_n)$.
3. Even function $f(x) = f(-x)$. This is satisfied by isotropy: $C(||\mathbf{h}||) = C(||-\mathbf{h}||)$.
4. A practical assumption for spatial analysis is that $C(||\mathbf{h}||)$ is non-increasing (i.e. larger distance leads to lower correlation).

Example: Exponential Covariance Function

$$C(\mathbf{h};\boldsymbol{\theta}) = \theta_1 \exp\left(-\frac{||\mathbf{h}||}{\theta_2}\right), \quad \theta_1, \theta_2 > 0.$$

- The marginal variance is given by θ_1.
- The correlation decreases as a function of distance. The rate of decrease is determined by θ_2.
- The semi-variogram is given by: $\gamma(\mathbf{h};\boldsymbol{\theta}) = \theta_1 \left[1 - \exp\left(-\frac{||\mathbf{h}||}{\theta_2}\right)\right]$.

Example: $\theta_1 = 5$ and $\theta_2 = 10$



Example: Gaussian Process

A Gaussian spatial process assumes that the finite-dimensional distribution $[Y(\mathbf{s}_1), Y(\mathbf{s}_2), \ldots, Y(\mathbf{s}_n)]'$ is multivariate Normal.

- The distribution is completely specified by its mean and covariance function. It is often used because of its properties for computation, prediction, and inference.

We assume $Y(\mathbf{s})$ is a Gaussian process (GP) with an exponential covariance function with mean zero. Then

$$[Y(\mathbf{s}_1), Y(\mathbf{s}_2), \ldots, Y(\mathbf{s}_n)]' \sim N(\mathbf{0},\ \Sigma)$$

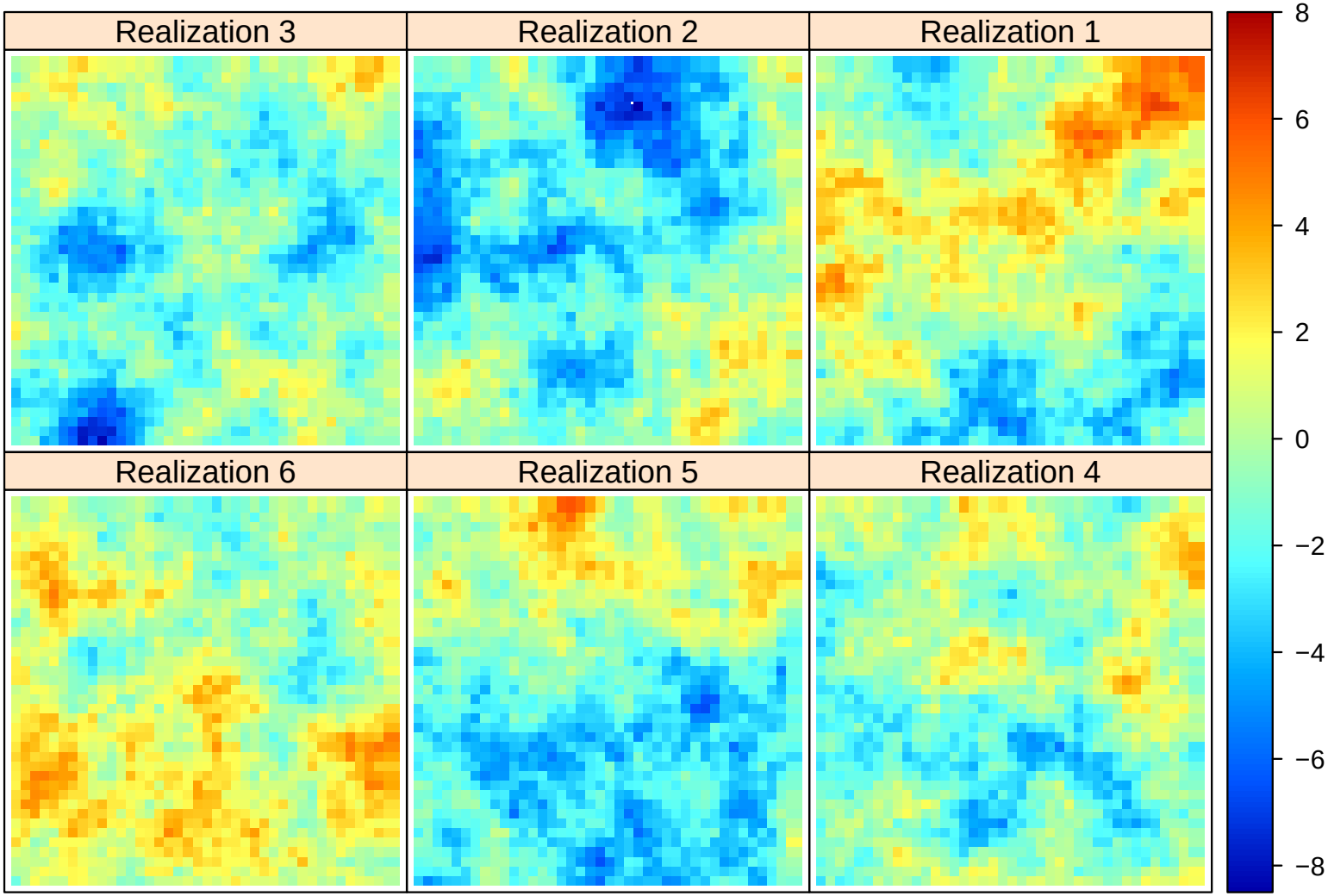
where $\Sigma_{ij} = \theta_1 \exp(-d_{ij}/\theta_2)$ and d_{ij} is the Euclidean distance between $\mathbf{s}_i$ and $\mathbf{s}_j$.

Let $\mathbf{s}_i = (x_i, y_i)$ represent locations over a grid where $x_1 = 1, \ldots, 40$ and $y_1 = 1, \ldots, 40$. We will now examine how θ_1 and θ_2 influence realizations from this GP.

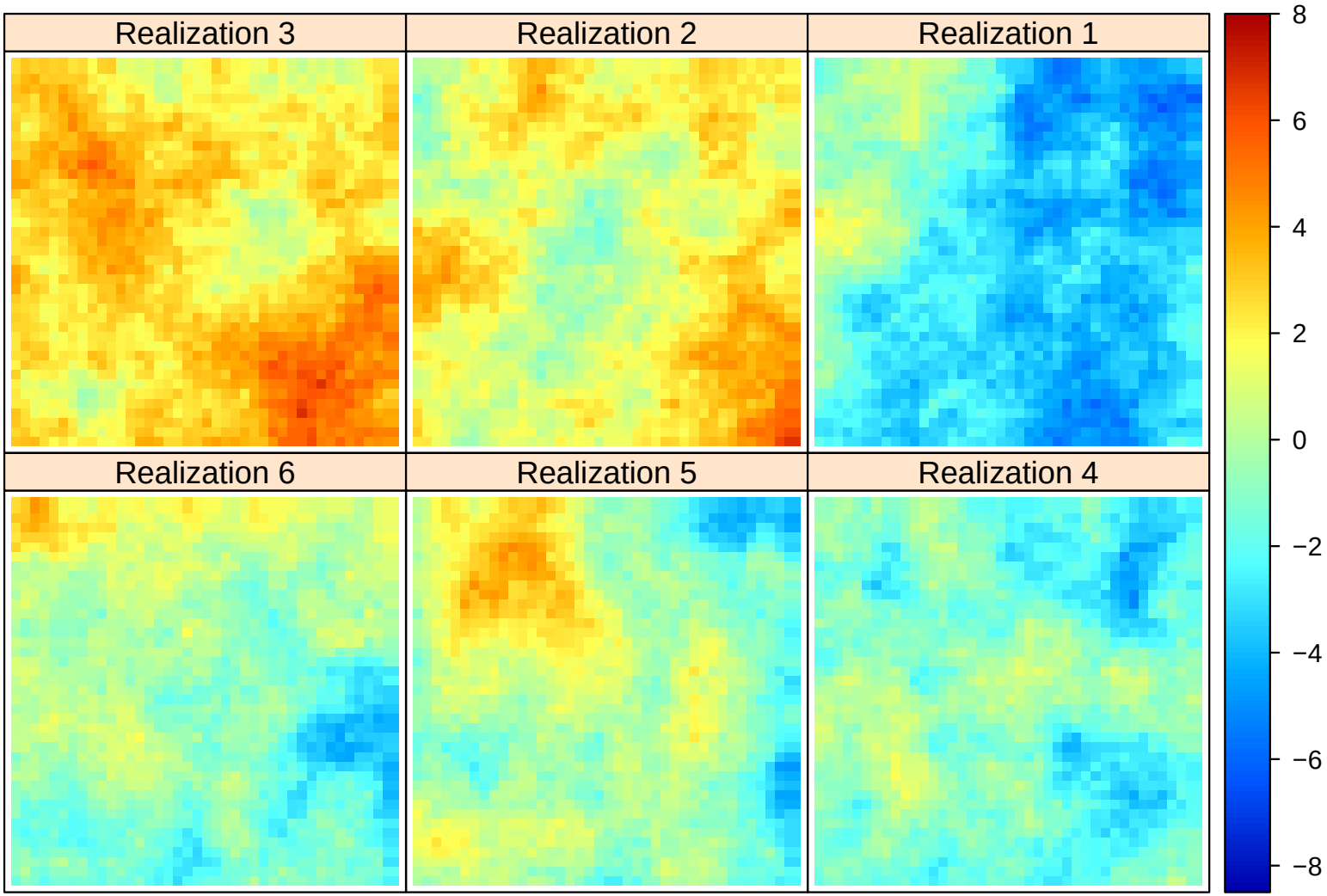
MVN vs Gaussian Process (distinguished by number of points, where GP is for discrete points)

GP $(\theta_1 = 5, \theta_2 = 10)$

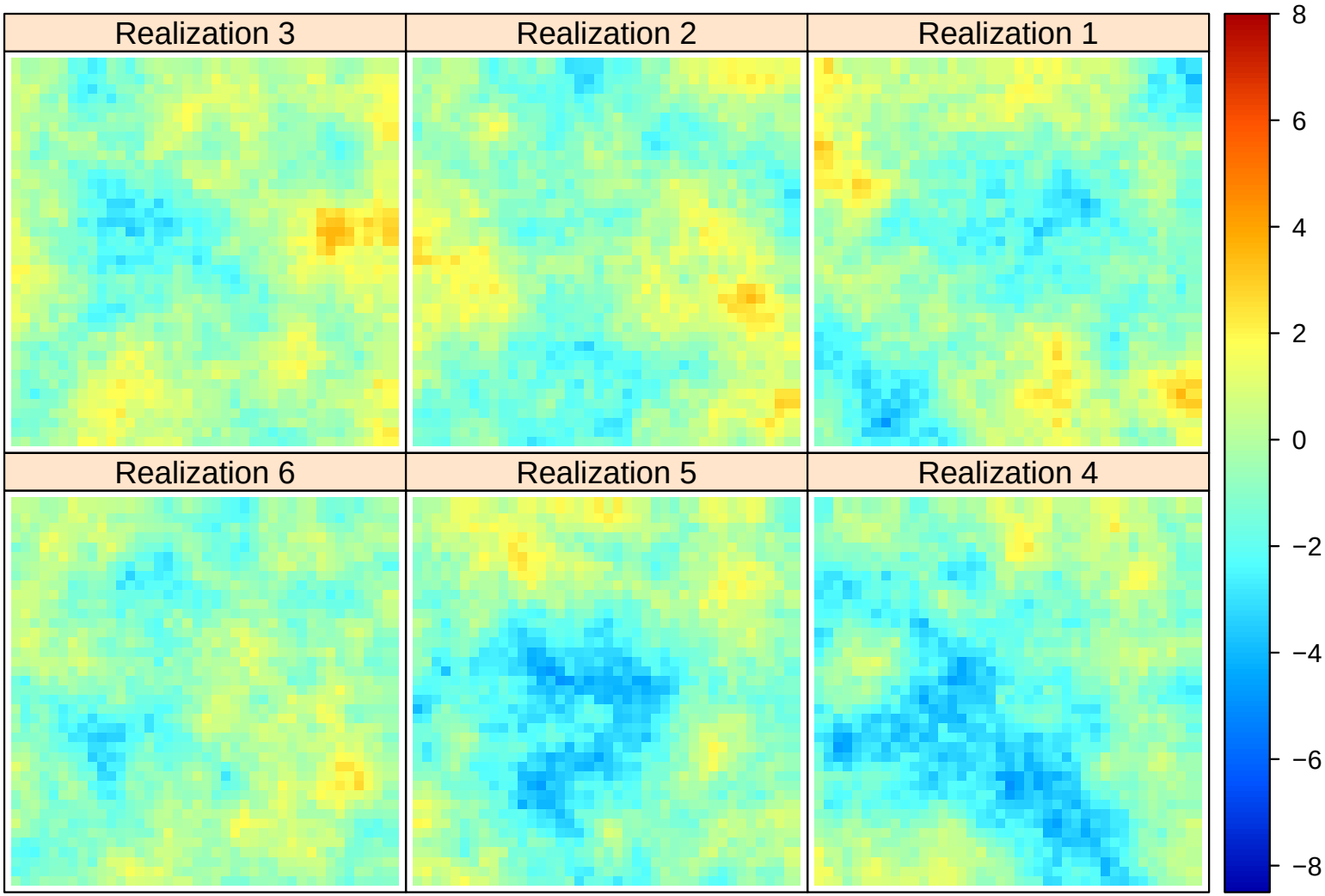
larger theta, larger drop of distance



GP $(\theta_1 = 5, \theta_2 = 30)$



GP $(\theta_1 = 1, \theta_2 = 10)$



Theta1 = measures how variable map is;
marginal variance;
if look at one pixel separately, how variable will it be across different realizations/simulations

Matérn Covariance Function

Another commonly used stationary and isotropic spatial covariance function is the Matèrn:

$$C(\mathbf{h}|\boldsymbol{\theta}) = \theta_1 \frac{1}{\Gamma(\theta_3)2^{\theta_3-1}} \left(\frac{||\mathbf{h}||}{\theta_2}\right)^{\theta_3} K_{\theta_3}\left(\frac{||\mathbf{h}||}{\theta_2}\right)$$

where $\theta_1, \theta_2, \theta_3 > 0$, and $K_{\theta_3}(\cdot)$ is the modified Bessel function of the second kind.

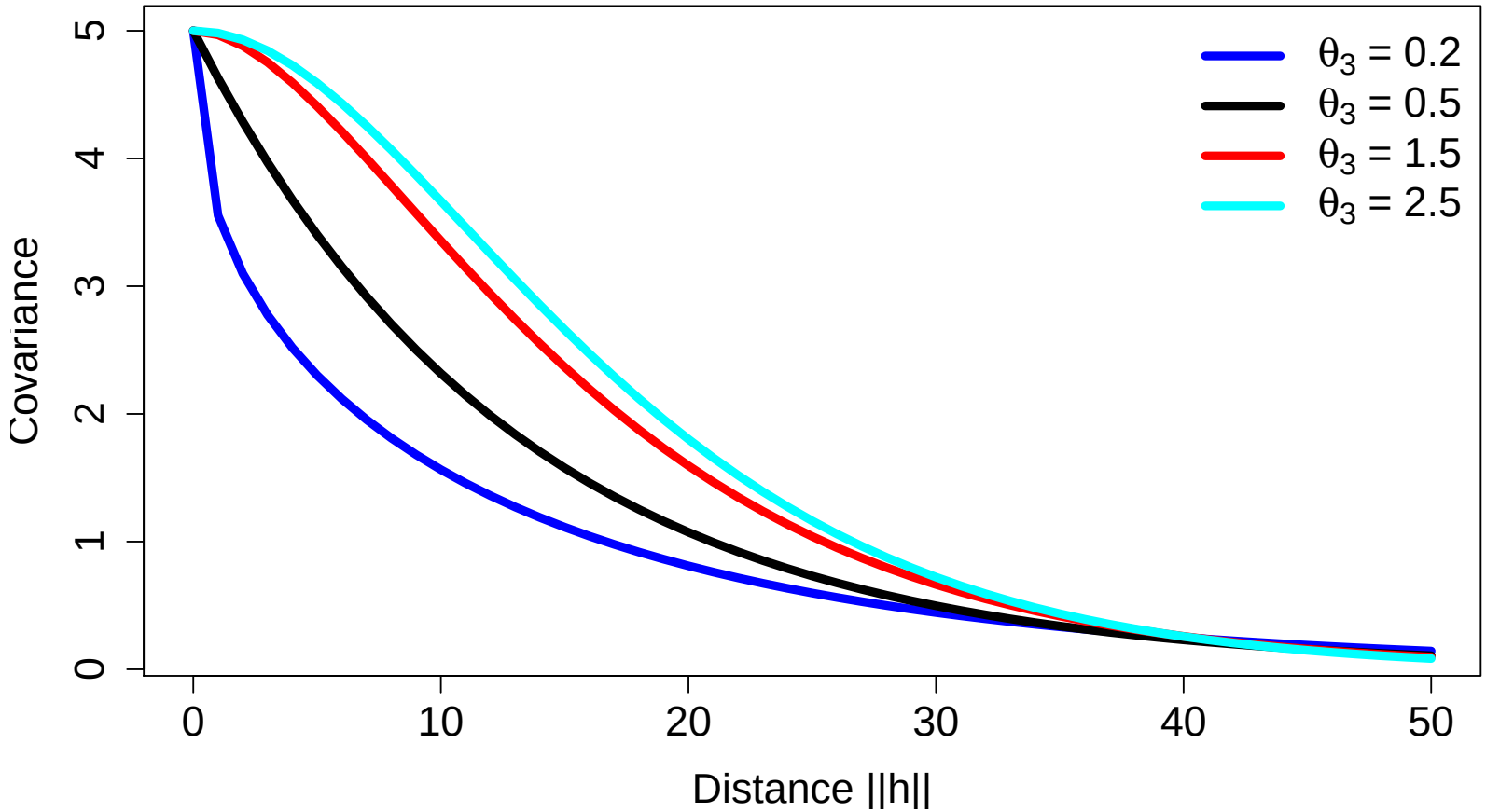
- $\theta_1 =$ marginal variance
- $\theta_2 =$ "scale" parameter
- $\theta_3 =$ "smoothness" parameter

Two special cases:

- $\theta_3 = 1/2$ gives $C(\mathbf{h}|\boldsymbol{\theta}) = \theta_1 \exp(-||\mathbf{h}||/\theta_2)$ "exponential"
- As $\theta_3 \to \infty$ gives $C(\mathbf{h}|\boldsymbol{\theta}) = \theta_1 \exp(-||\mathbf{h}||^2/\tilde{\theta}_2)$ "Gaussian" or "double exponential"

Matérn Covariance Function

$\theta_1 = 5$ and θ_2 is chosen such that the correlation is 0.05 at $||\mathbf{h}|| = 40$.



Smoothness

The covariance function directly impacts how "smooth" the spatial process will be. One common way to measure smoothness is to examine continuity and differentiability of the random variable.

The spatial process $Y(\mathbf{s})$ of a Matérn covariance function is *m-times* differentiable if and only if $\theta_3 > m$.

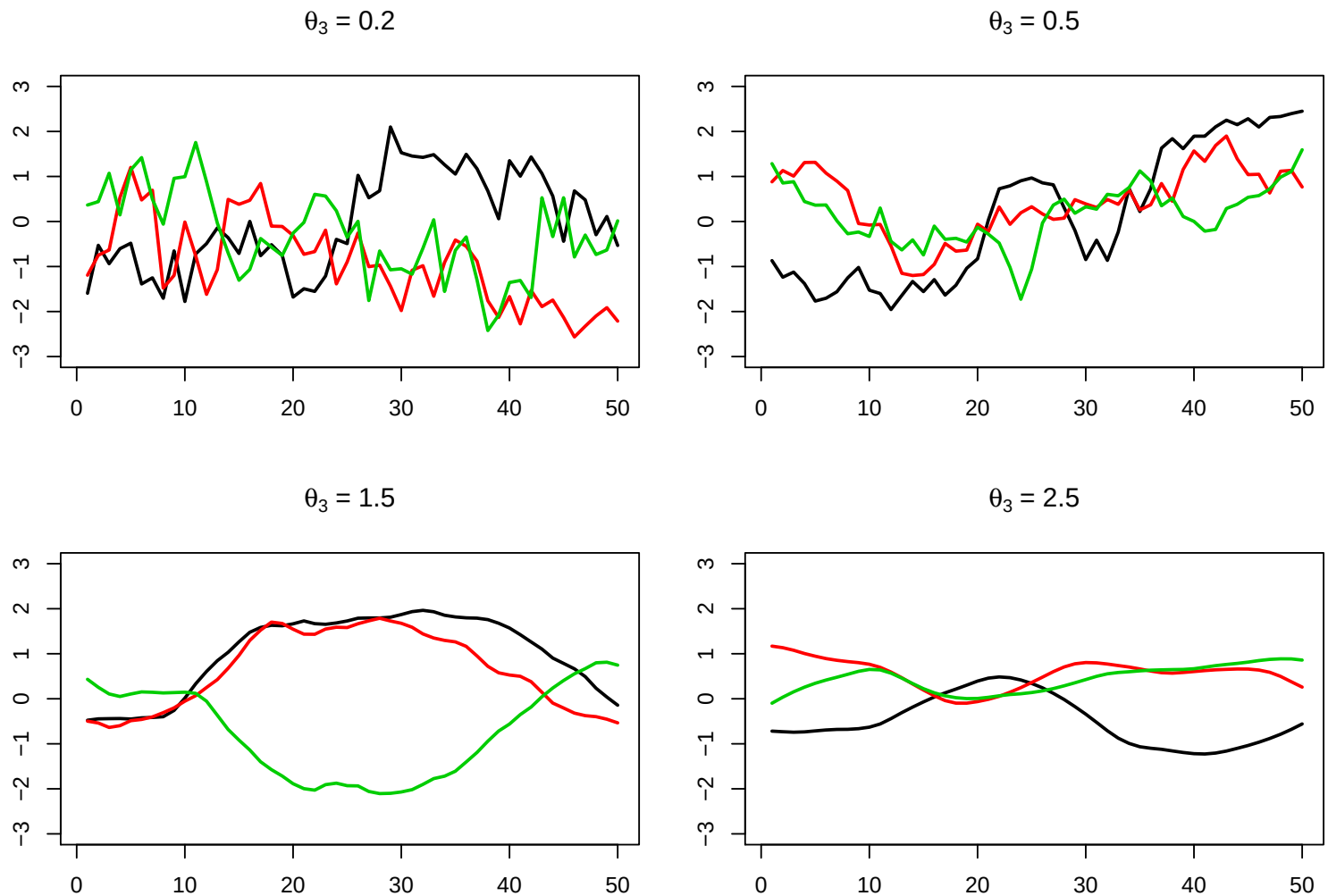
Therefore,

- Exponential covariance function is mean-square continuous.
- Matérn covariance function with $\theta = 1.5$ is once mean-square differentiable.
- Matérn covariance function with $\theta = 2.5$ is twice mean-square differentiable.

We will first examine realizations from a 1-dimensional mean-zero Gaussian process: $Y(t), t \in \mathbb{R}$ with different degree of smoothing.

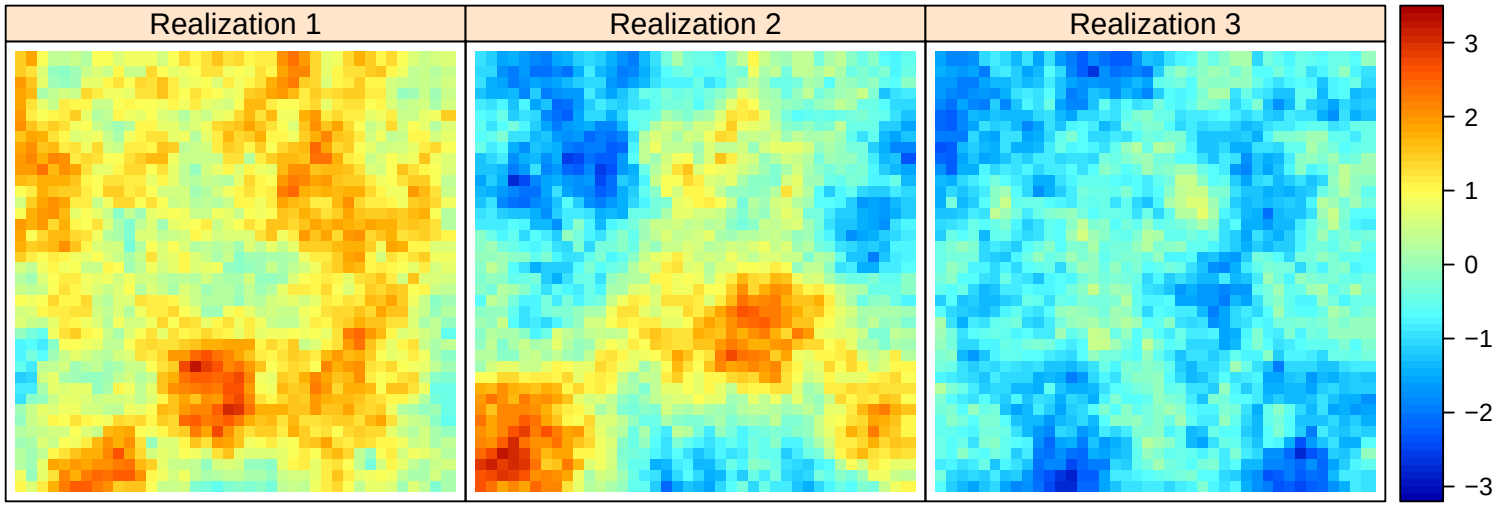
Matérn Covariance Smoothness in 1-D

$\theta_1 = 1$ and θ_2 is chosen such that the correlation is 0.05 at $||\mathbf{h}|| = 40$.

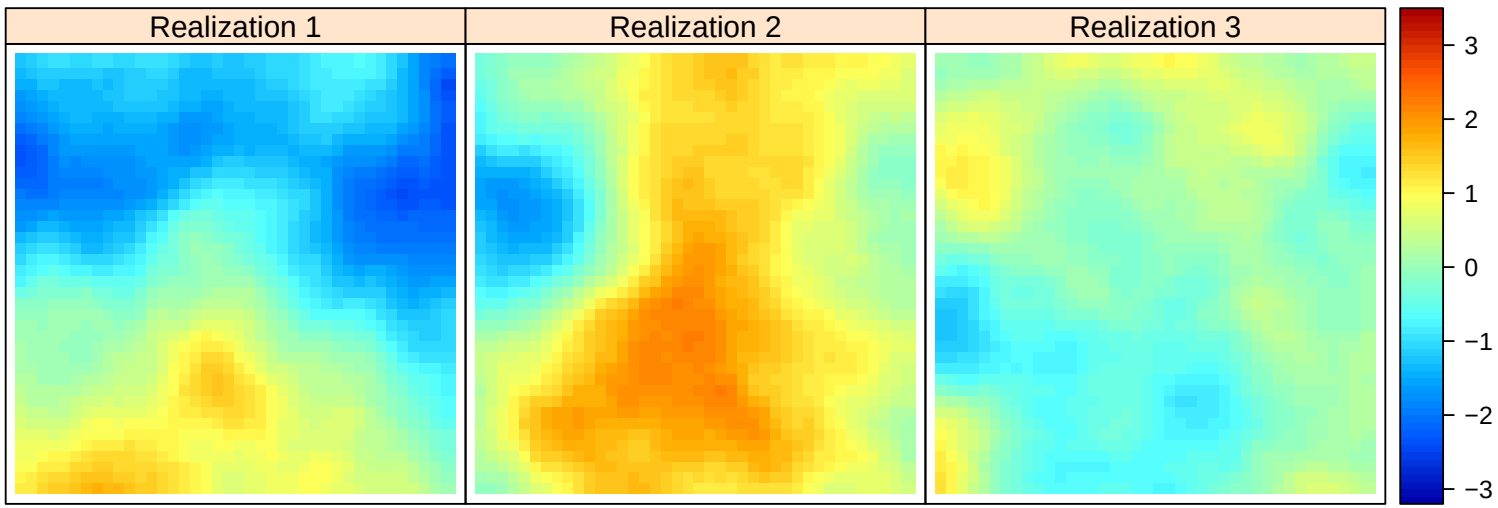


Matérn Covariance Smoothness in 2-D

Matérn ($\theta_1 = 1$, $\theta_2 = 13$, $\theta_3 = 0.5$)



Matérn ($\theta_1 = 1$, $\theta_2 = 8.5$, $\theta_3 = 1.5$)



Considerations for Covariance Function

Two important characteristics of a spatial random field are:

1. Large-scale dependence: when does the correlation between two points fall below a threshold (e.g. 0.05)? Sometimes referred to as the "range" of a covariance function.

2. Smoothness: how quickly does the correlation fall as a function of distance? This is determined by the behavior of the covariance function at $\mathbf{h} = 0$.

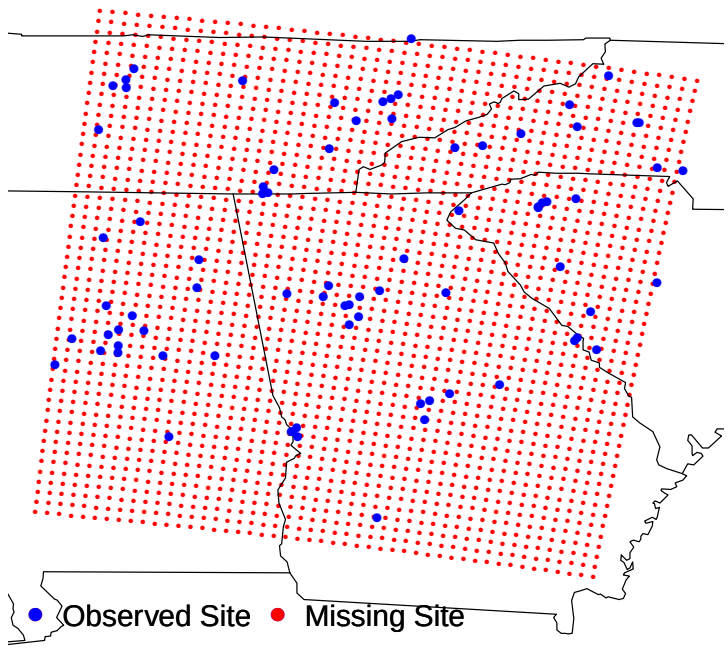
Range and smoothness can play important roles in spatial prediction (i.e. kriging) applications.

In practice, we often select a covariance function family a priori or use exploratory analysis/visualization.

Kriging Prediction

Kriging = Optimal Spatial Prediction

- Let $\mathbf{Y}$ be the vector of observed values at locations $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n$.
- Let W_0 denote the unobserved spatial process at location $\mathbf{s}_0$. We wish to obtain a prediction based on $\mathbf{Y}$. Important to note that the goal is to estimate the spatially-dependent trend.



Kriging

What is *optimal*? We will start with considering a linear predictor:

$$W_0 = \lambda_0 + \lambda_1 Y(\mathbf{s}_i) + \lambda_2 Y(\mathbf{s}_2) + \ldots + \lambda_n Y(\mathbf{s}_n)$$

where $\lambda_0, \ldots, \lambda_n$ are unknown parameters.

If we consider optimality in the sense of mean-square prediction error (MSPE) defined as

$$E[\,\{W_0 - \lambda_0 - \mathbf{L}'\mathbf{Y}\}^2\,]\,,$$

then under simple kriging equations are

$$\hat{\mathbf{W}}_0 = E[\mathbf{W}_0] + \mathbf{C}_{\mathbf{W}_0\mathbf{Y}}\mathbf{C}_{\mathbf{Y}}^{-1}\{\mathbf{Y} - E[\mathbf{Y}]\}\ .$$

$$\text{Var}\,[\hat{\mathbf{W}}_0 - \mathbf{W}_0] = \mathbf{C}_{\mathbf{W}_0} - \mathbf{C}_{\mathbf{W}_0\mathbf{Y}}\mathbf{C}_{\mathbf{Y}}^{-1}\mathbf{C}_{\mathbf{W}_0\mathbf{Y}}^{T}$$

with $\mathbf{C}_{\mathbf{Y}} = \text{Cov}(\mathbf{Y}, \mathbf{Y})\mathbf{C}_{\mathbf{W}_0} = \text{Cov}(\mathbf{W}_0, \mathbf{W}_0)$, and $\mathbf{C}_{\mathbf{W}_0\mathbf{Y}} = \text{Cov}(\mathbf{W}_0, \mathbf{Y})$

We also recognize that if $\mathbf{Y}$ is Normal, the above equations correspond to the conditional mean and variance of $\mathbf{W}_0|\mathbf{Y}$!

Spatial Process Regression Model

Let $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n$ be a finite sample of locations. To model a continuous response variable of interest $Y(\mathbf{s}_i)$, a spatial regression model typically consists of three independent components:

$$Y(\mathbf{s}_i) = \mu(\mathbf{s}_i) + W(\mathbf{s}_i) + \epsilon(\mathbf{s}_i).$$

- Mean Trend $\mu(\mathbf{s}_i)$: it is parameterized as

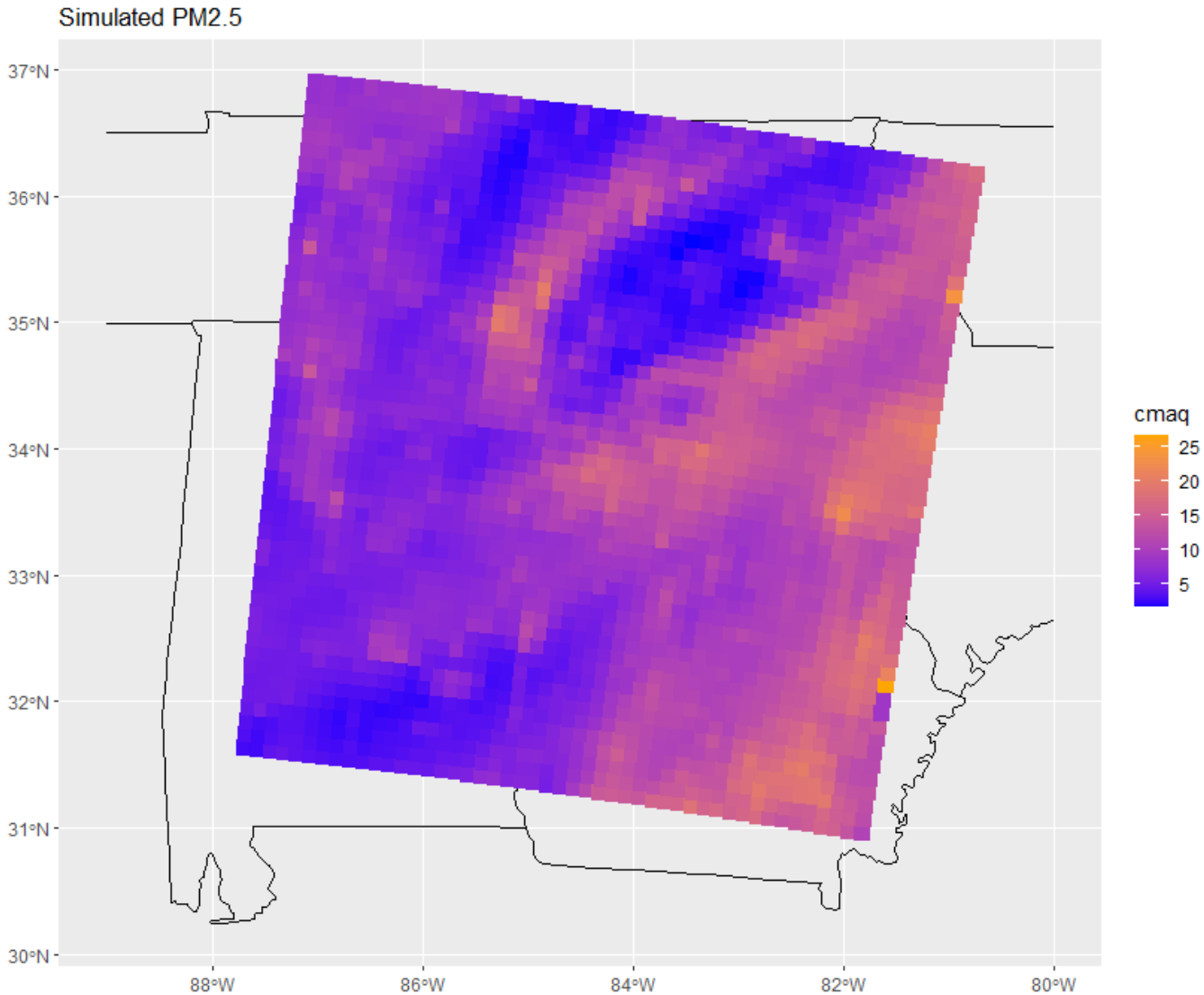
$$\mu(\mathbf{s}_i) = \sum_{j=1}^{J} \beta_j X_j(\mathbf{s})$$

where $X_j(\mathbf{s})$ is the j^{th} spatial covariate and β_j is the corresponding regression coefficient.

This is often called the "fixed-effect" part of the model.

Our Example Predictor

We will consider one spatially-varying covariate to model air pollution (simulation from CMAQ).



Spatial Process Regression Model

- Spatial Process $W(\mathbf{s}_i)$: describes the variation not captured by $\mu(\mathbf{s}_i)$ that exhibits spatial dependence.

  It is typically modeled using a zero-mean Gaussian process:

  $$\mathbf{W} = [W(\mathbf{s}_1), \ldots, W(\mathbf{s}_n)]' \sim GP(0, \Sigma_{\boldsymbol{\theta}})$$

  where $\Sigma_{\boldsymbol{\theta}}$ is an $n \times n$ covariance matrix with elements given by a covariance function $C(d; \boldsymbol{\theta})$, d is the Euclidean distance between two locations, and $\boldsymbol{\theta}$ is a vector of parameters.

  $\boldsymbol{\theta}$ often includes a marginal variance $\tau^2 = C(d = 0; \boldsymbol{\theta})$.

  $\mathbf{W}$ is also called the "spatial random effects" or the "structured" random effects.

- Residual Error $\epsilon(\mathbf{s}_i)$: describes the additional "unstructured" variation

  $$\epsilon(\mathbf{s}_i) \overset{iid}{\sim} N(0, \sigma^2).$$

  Also known as the nugget variance or white-noise process.

Kriging

$$Y(\mathbf{s}_i) = \mathbf{X}(\mathbf{s}_i)'\boldsymbol{\beta} + W(\mathbf{s}_i) + \epsilon(\mathbf{s}_i)$$

Traditional kriging analysis involves the following steps:

1. Identify a covariance function for $W(\mathbf{s}_i)$ and estimate the parameters ($\boldsymbol{\theta}$ and σ^2) associated with $W(\mathbf{s}_i)$ and $\epsilon_i(\mathbf{s}_i)$.
2. Given $\boldsymbol{\theta}$ and σ^2, obtain the weighted least-square estimate of $\boldsymbol{\beta}$.
3. Make predictions at locations of interest.

This algorithm does not require Gaussian assumptions on $W(\mathbf{s}_i)$ and $\epsilon(\mathbf{s}_i)$. However, the multi-stage algorithm means that steps 2 and 3 assume the covariance function is known! We will relax this later with Bayesian kriging.

Spatial Process Regression Model

We first consider a Gaussian covariance function with and without the spatial covariate.

$$\text{Cov}\left(||\mathbf{s}_i - \mathbf{s}_j||\right) = \tau^2 \exp\left(\frac{-||\mathbf{s}_i - \mathbf{s}_j||^2}{\rho^2}\right)$$

Estimate and 95% CI

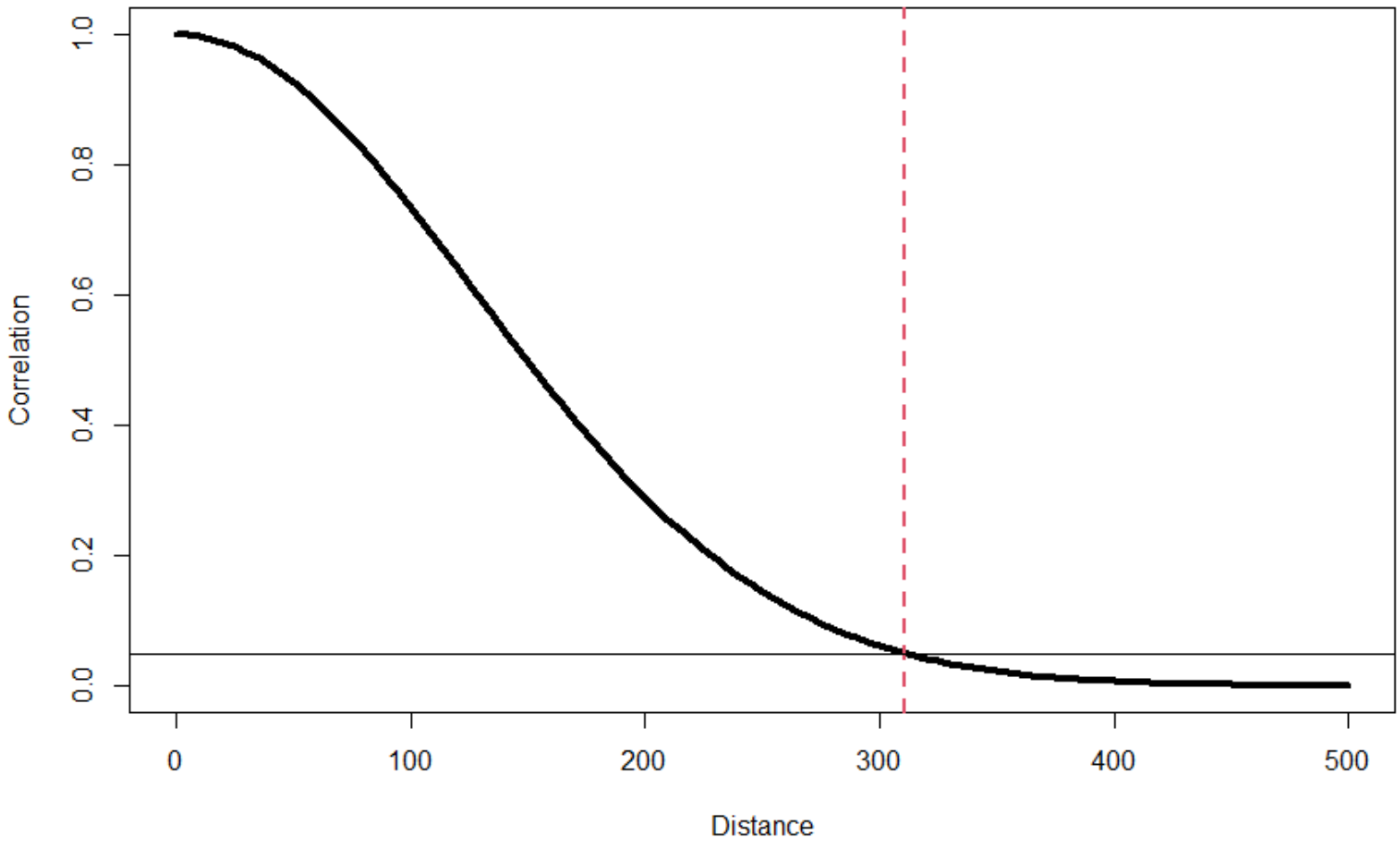
	No covariate	Covariate (CMAQ)
β_0	13.0 (0.85, 17.2)	8.9 (4.34, 13.4)
β_{CMAQ}		0.36 (0.11, 0.61)
τ^2 (partial sill)	18.9	16.4
ρ (range)	204	179
σ^2 (nugget)	15.6	13.4
AIC	457	446

Some evidence that the covariate is useful:

- β_{CMAQ} has 95% CI excluding 0.
- AIC and σ^2 are smaller with the covariate.

Estimated Covariance Function

Residual correlation drops to 0.05 at around 310 km (also known as the effective distance).



A Note on Standard Error Calculation

Let $V_{\boldsymbol{\theta},\sigma^2} = \Sigma_{\boldsymbol{\theta}} + \sigma^2 \mathbf{I}$. Our model-based spatial model is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \tilde{\boldsymbol{\epsilon}} \qquad \tilde{\boldsymbol{\epsilon}} \sim N(\mathbf{0}, V_{\boldsymbol{\theta},\sigma^2}) \, .$$

Recall from the mixed-model framework, the estimate of $\boldsymbol{\beta}$ and its standard error are given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' V^{-1}_{\hat{\boldsymbol{\theta}},\hat{\sigma}^2} \mathbf{X})^{-1} \mathbf{X}' V^{-1}_{\hat{\boldsymbol{\theta}},\hat{\sigma}^2} \mathbf{Y}$$

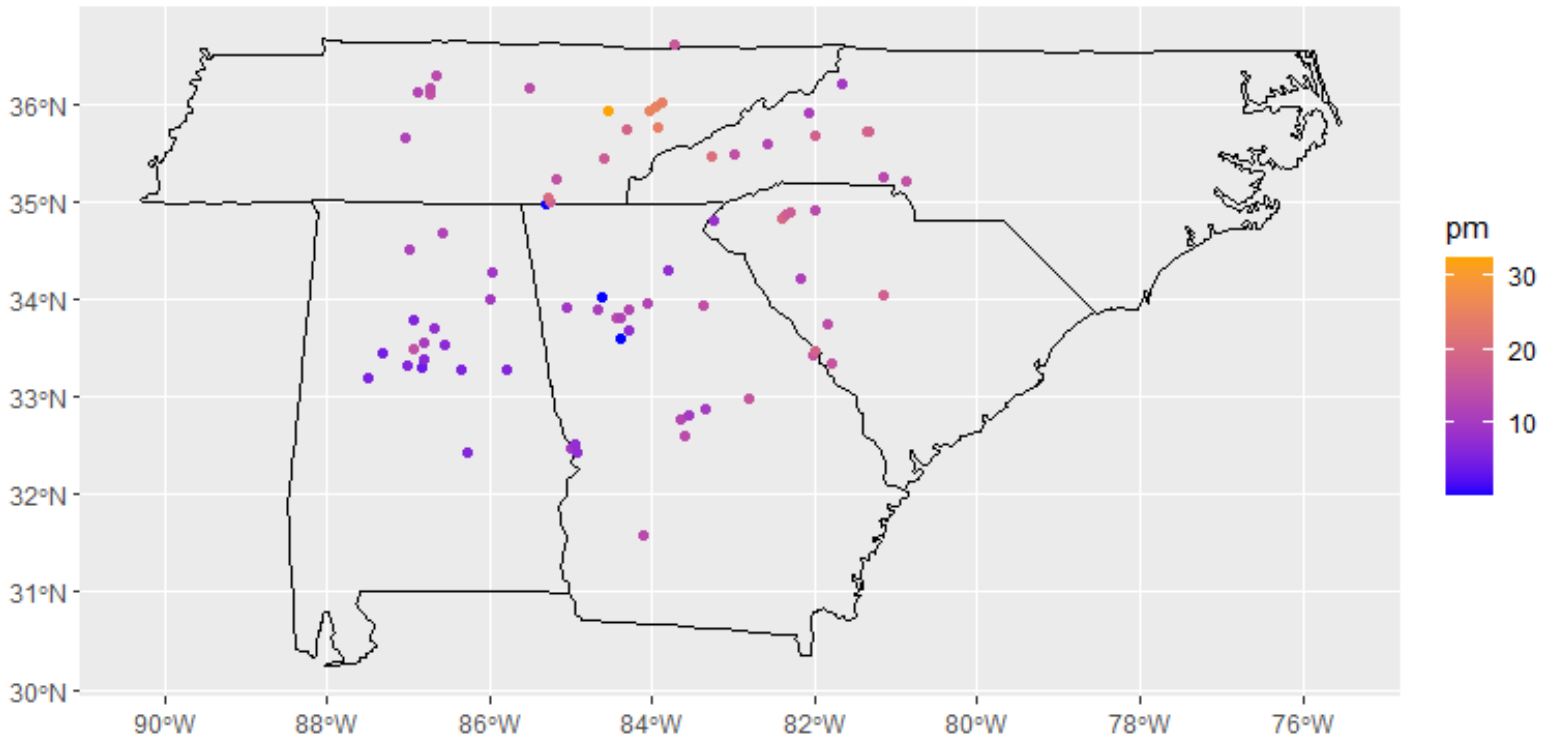
$$\text{Cov}\left[\hat{\boldsymbol{\beta}}\right] = (\mathbf{X}' V^{-1}_{\hat{\boldsymbol{\theta}},\hat{\sigma}^2} \mathbf{X})^{-1}$$

where $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}^2$ are the REML/MLE estimate.

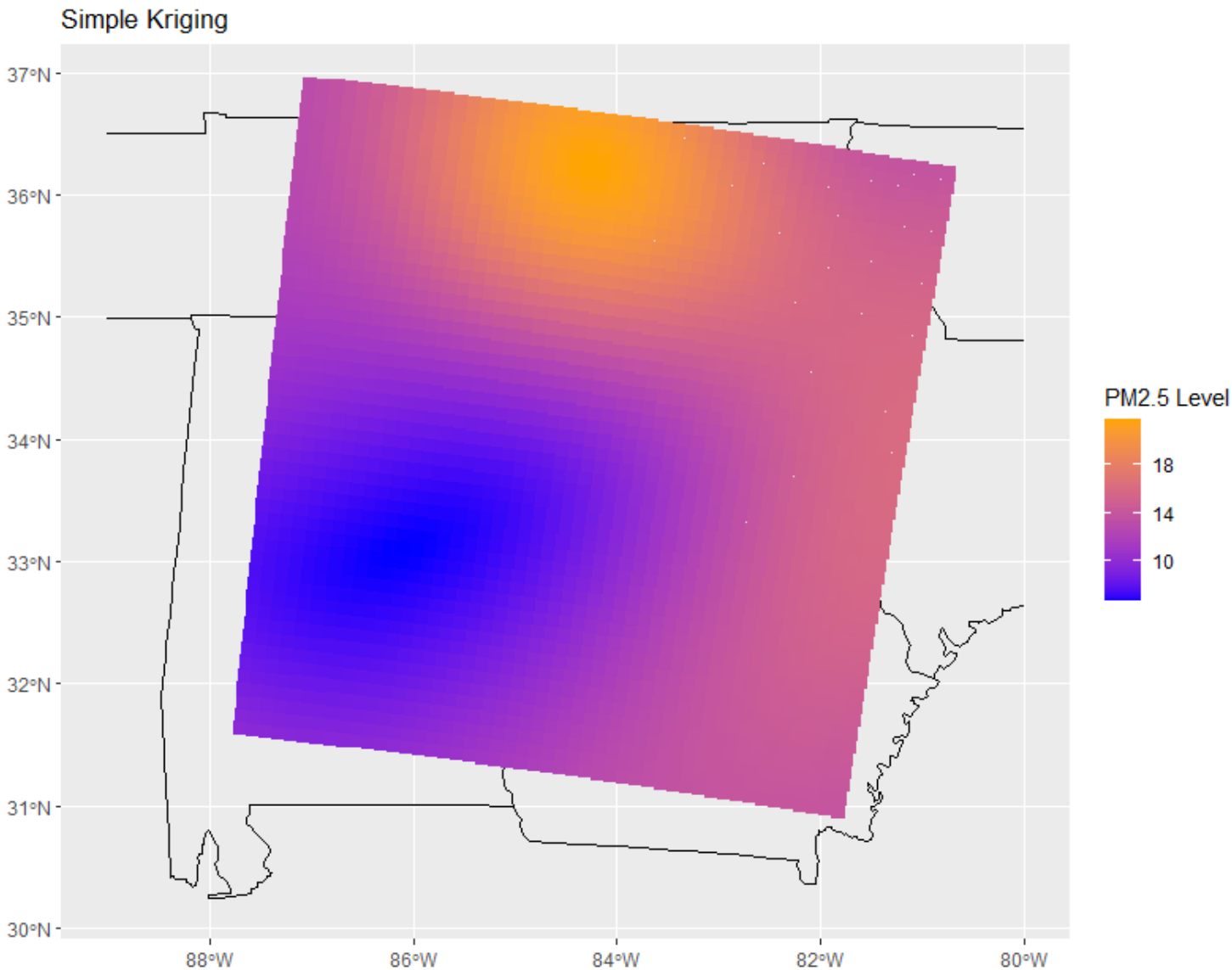
A standard mixed-model analysis assumes $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}^2$ as known. So the uncertainty in $\hat{\beta}$ does not account for uncertainty in $\boldsymbol{\theta}$ and σ^2.

Recall Our Data

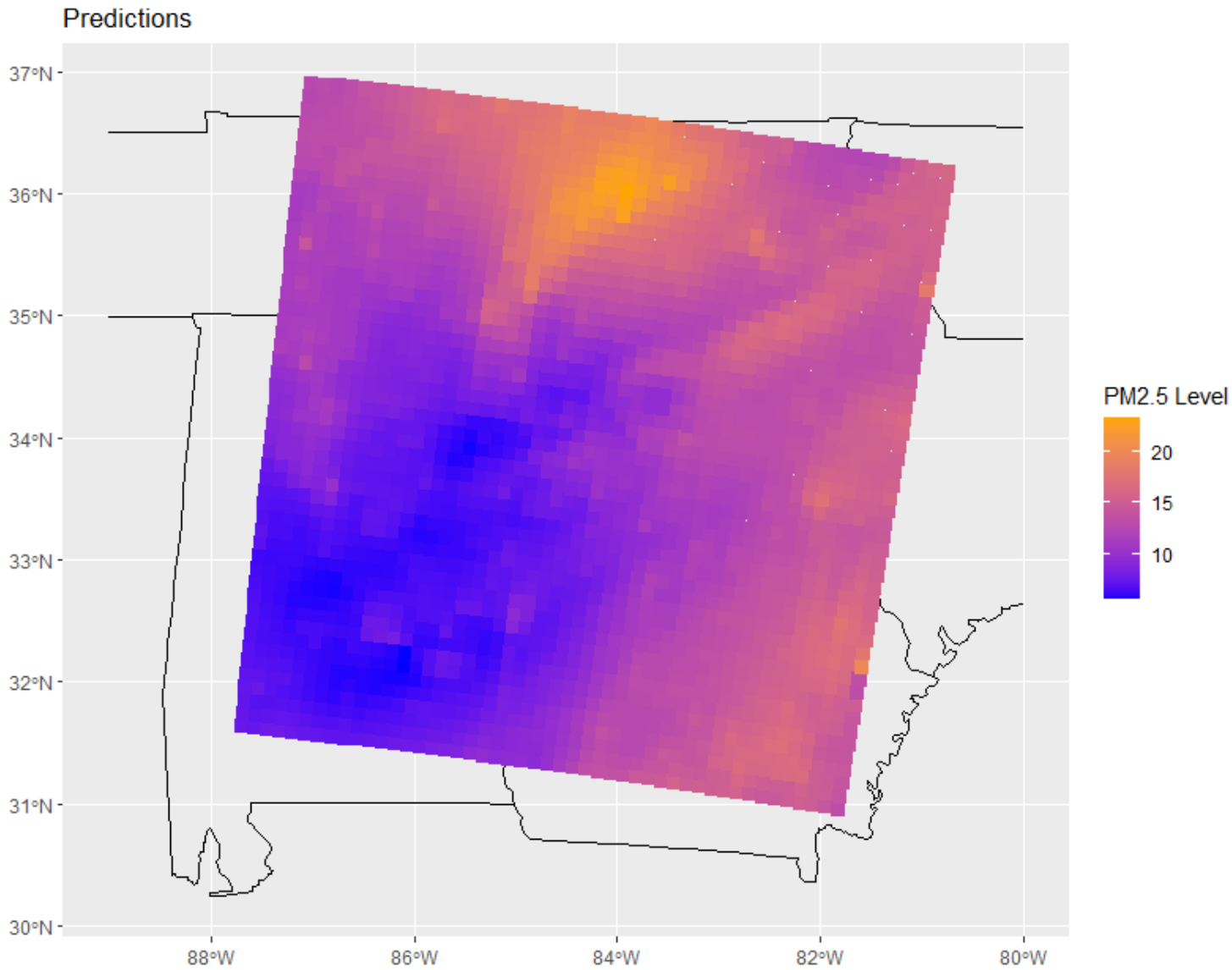
Average $PM_{2.5}$ concentrations ($\mu g/m^3$) across 78 monitors (2001-2003)



Simple Kriging

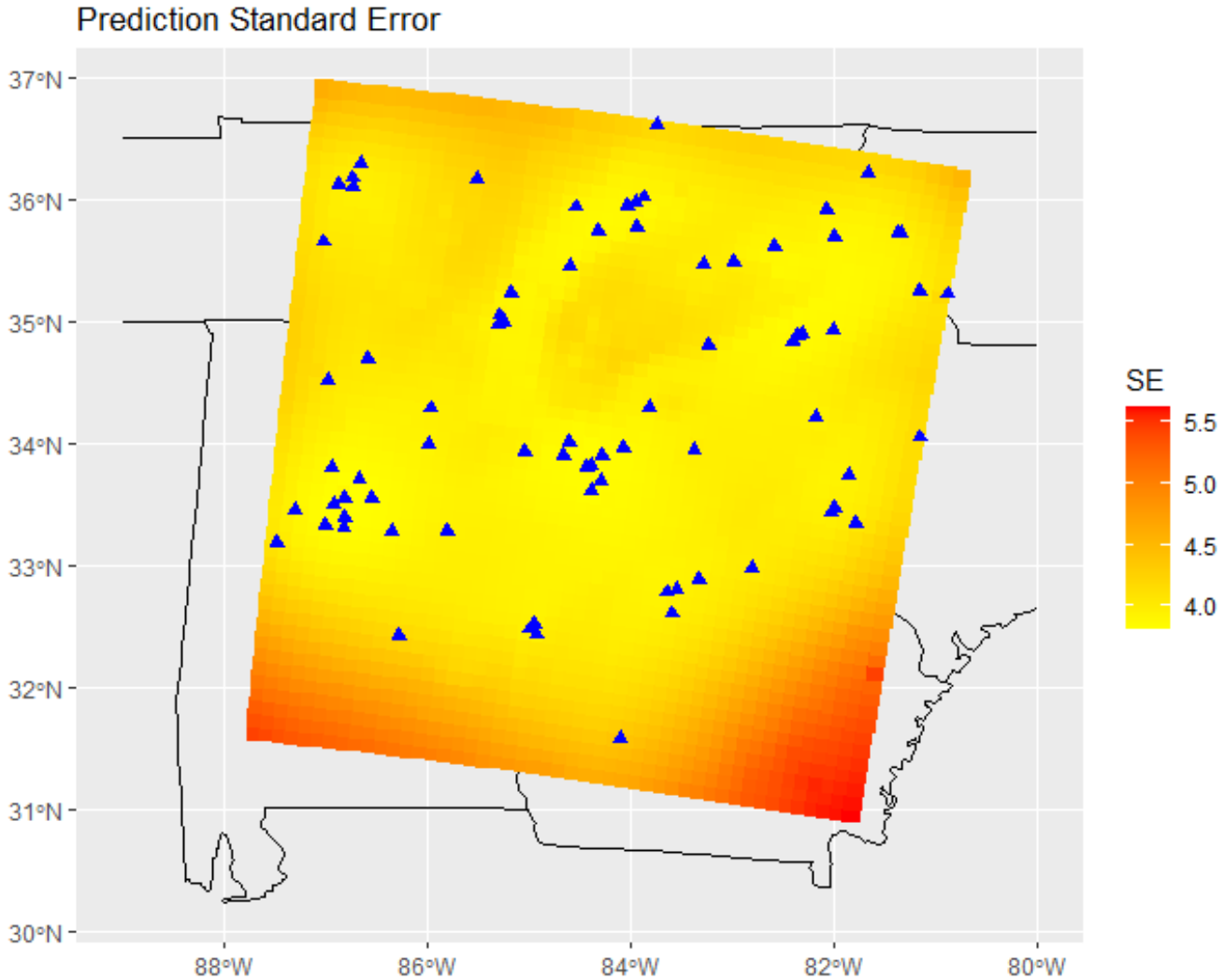


Kriging with Covariate



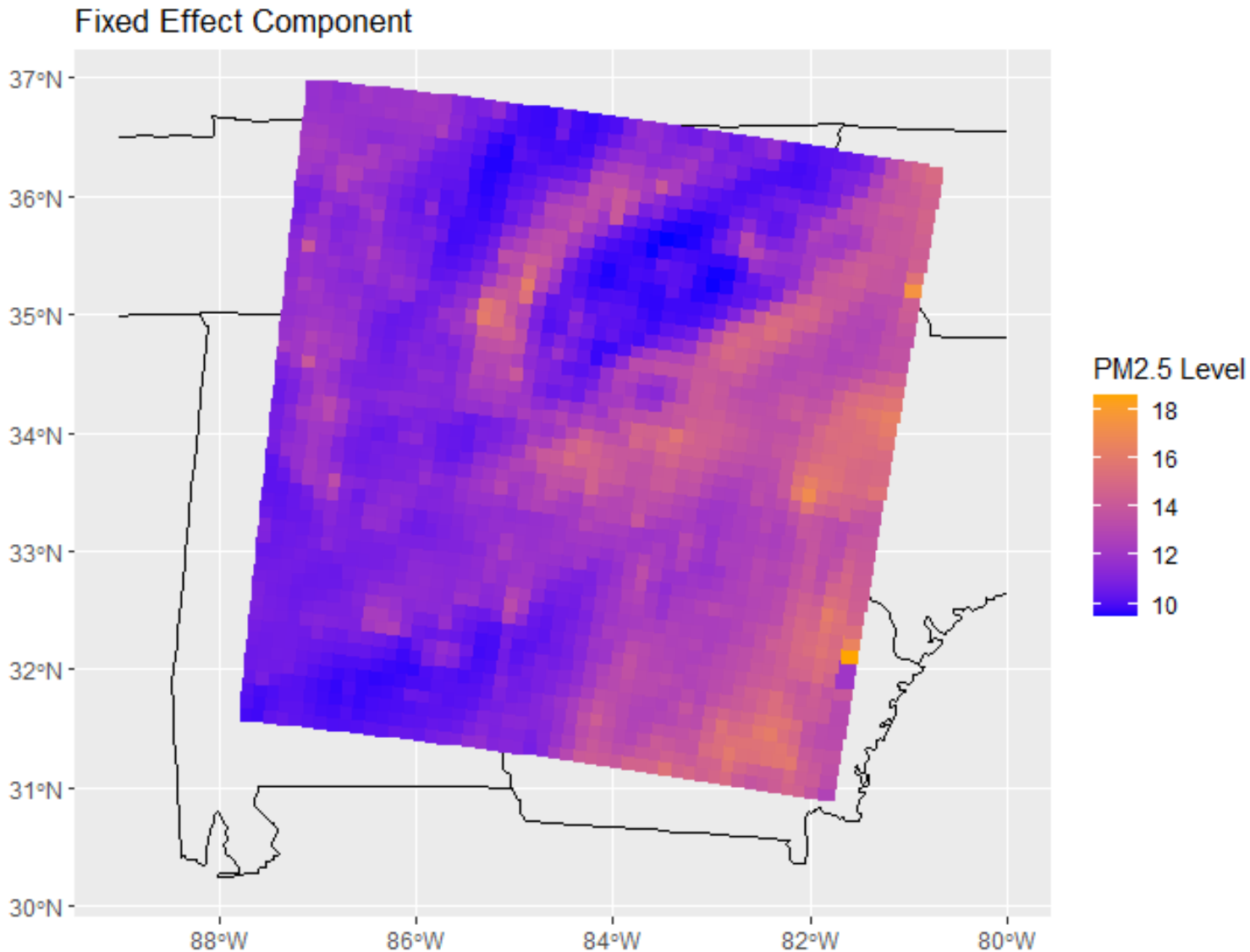
Prediction Error

Monitor locations indicated by triangles.



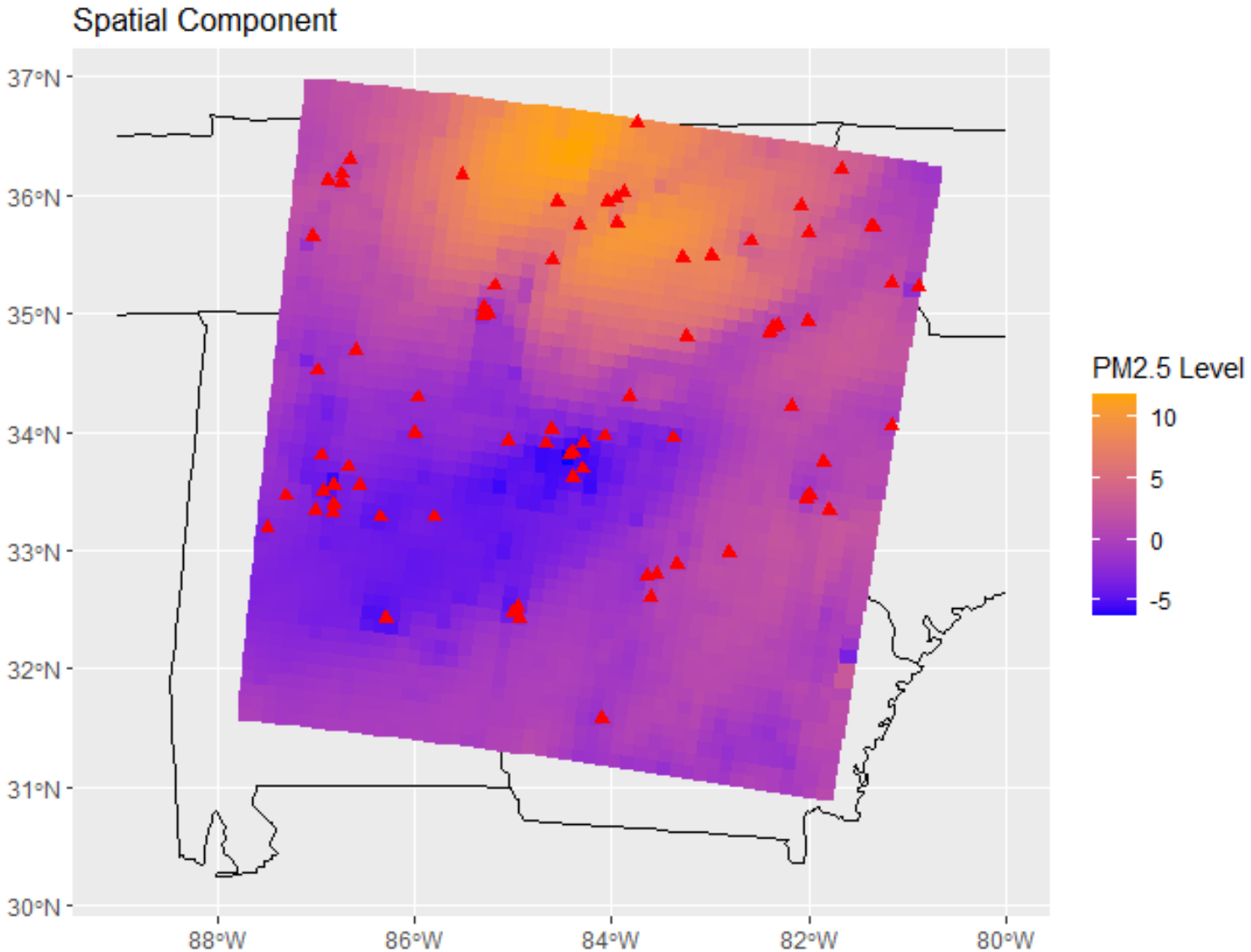
Decomposing the Prediction - Fixed Effect

$$Y(\mathbf{s}_i) = \mathbf{X}(\mathbf{s}_i)'\boldsymbol{\beta} + W(\mathbf{s}_i) + \epsilon(\mathbf{s}_i)$$



Decomposing the Prediction - Spatial Component

$$Y(\mathbf{s}_i) = \mathbf{X}(\mathbf{s}_i)'\boldsymbol{\beta} + W(\mathbf{s}_i) + \epsilon(\mathbf{s}_i)$$



Why Bayesian?

In spatial analysis, treating τ^2, ρ and σ^2 as known may significantly under-estimate the uncertainty.

Advantages

- More appropriate for small sample size.
- When good prediction uncertainty quantification is needed.
- Can incorporate constraints/information on parameters that are difficult to estimate from the data.

Disadvantages

- Computation may be more prohibitive.
- Some situations need to pick prior distributions carefully.

Implementation in geoR

Need to specify prior distributions for ALL model parameters (defaults are in red).

- $\boldsymbol{\beta}$: flat, normal, fixed
- σ^2: reciprocal, uniform, scaled inverse-χ^2, fixed
- τ^2/σ^2 (marginal/residual variance): fixed, uniform, reciprocal
- ϕ (range parameter): discretized uniform, exponential, fixed, squared-reciprocal, reciprocal

Except for the flat and reciprocal priors, one would need to specify additional parameters.

Fixed prior means that assume this parameter is known as the pre-specified value.

Still need to specify a covariance function.

Example: Bayesian Kriging Model

Model:

$$Y(\mathbf{s}_i) = \beta_0 + \beta_1 CMAQ(\mathbf{s}_i) + W(\mathbf{s}_i) + \epsilon(\mathbf{s}_i) \ .$$

$$\mathbf{W} \sim GP(\mathbf{0}, \Sigma_{\boldsymbol{\theta}}) \qquad \epsilon(\mathbf{s}_i) \overset{iid}{\sim} N(0, \sigma^2).$$

$$\Sigma_{\boldsymbol{\theta},jk} = \tau^2 \exp\{-||\mathbf{s}_j - \mathbf{s}_k||^2/\phi^2\}$$

Priors:

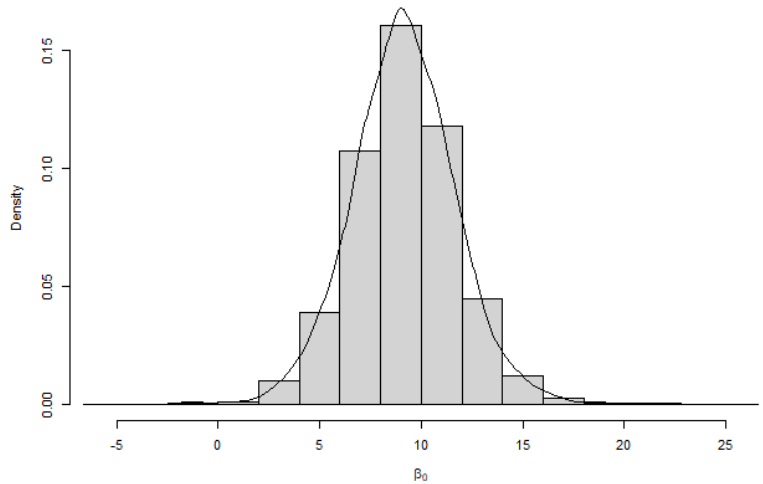
$$\beta_0 \propto 1 \qquad \beta_1 \propto 1$$

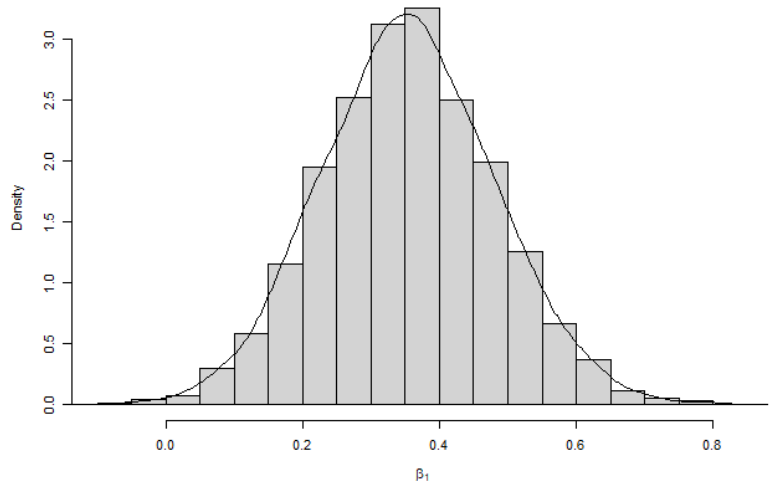
$$\sigma^2 \propto 1/\sigma^2 \qquad \tau^2/\sigma^2 \sim \text{Unif}(0.1, 10) \qquad \phi \sim \text{Unif}(1, 1000)$$

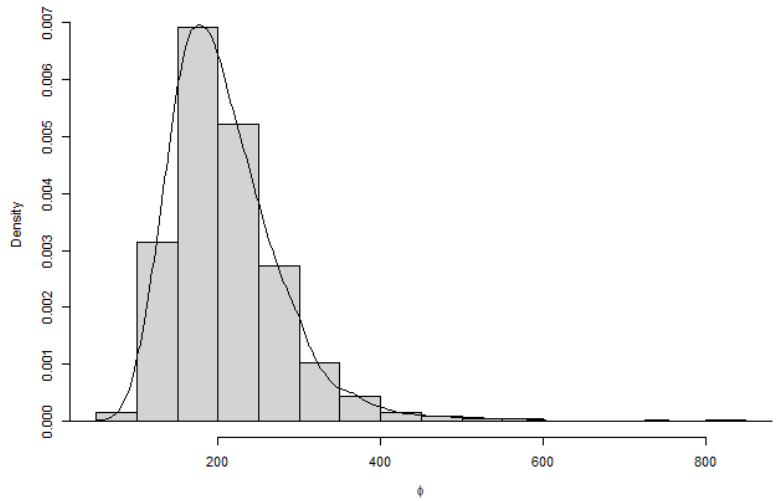
Estimation Precision:

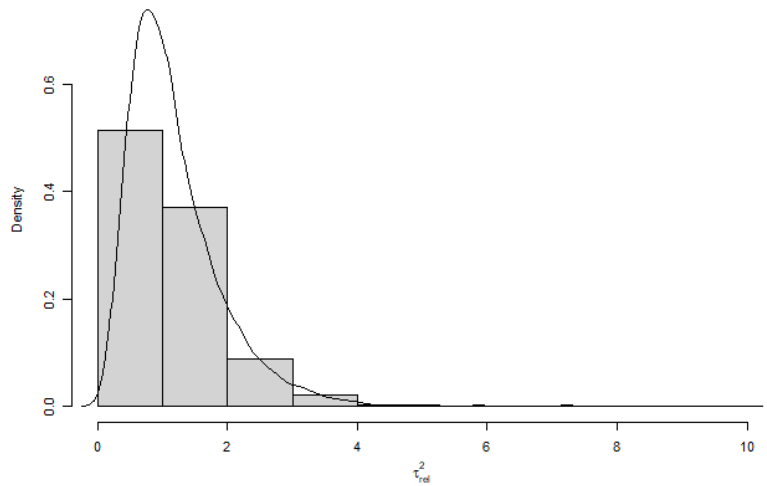
- Total number of posterior samples = 5,000
- Discretization of τ^2/σ^2 by 0.1
- Discretization of ϕ by 5

Posterior Distributions of Parameters



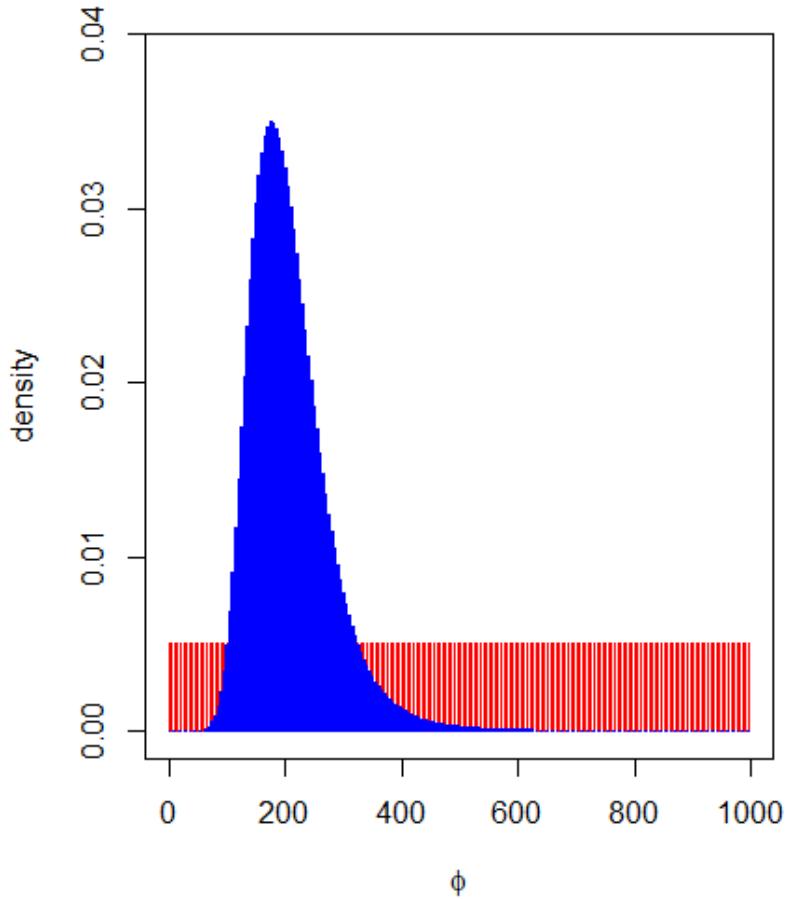


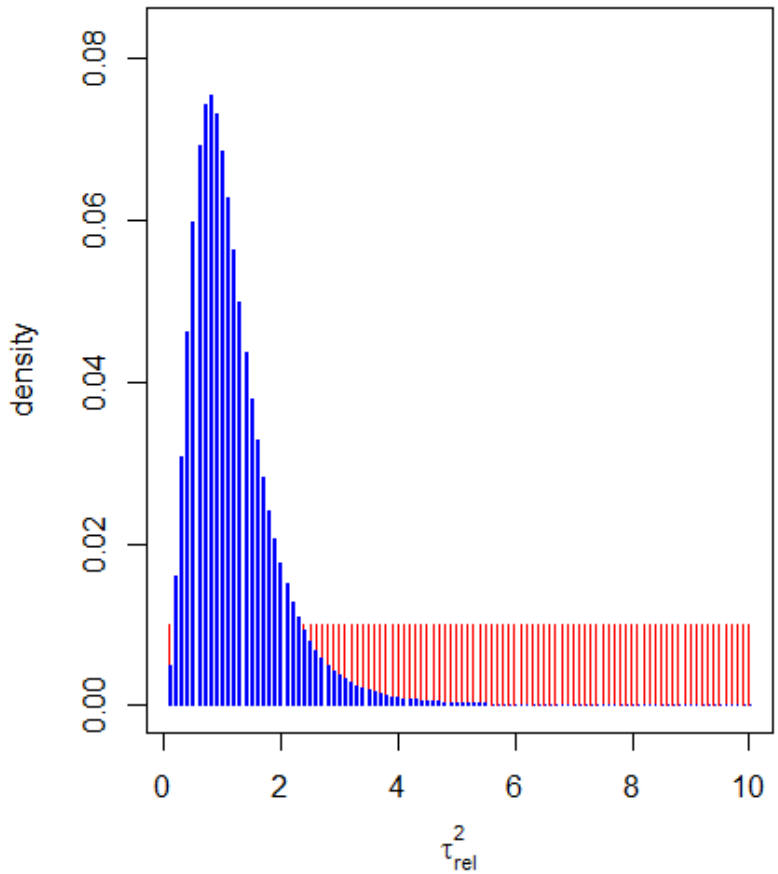




Posterior Distributions of Parameters

These plots show that the data are informative in estimating the spatial parameters when compared to the uniform prior (red).





Estimates Comparison

Estimate and 95% CI

	Bayesian	Non-Bayesian
β_0	9.15 (4.0, 14.6)	8.9 (4.3, 13.4)
β_{CMAQ}	0.35 (0.11, 0.61)	0.36 (0.11, 0.61)
τ^2 (partial sill)	15.4 (10.5, 22.1)	16.4
ρ (range)	209 (111, 381)	179
σ^2 (nugget)	15.4 (10.5, 22.2)	13.4

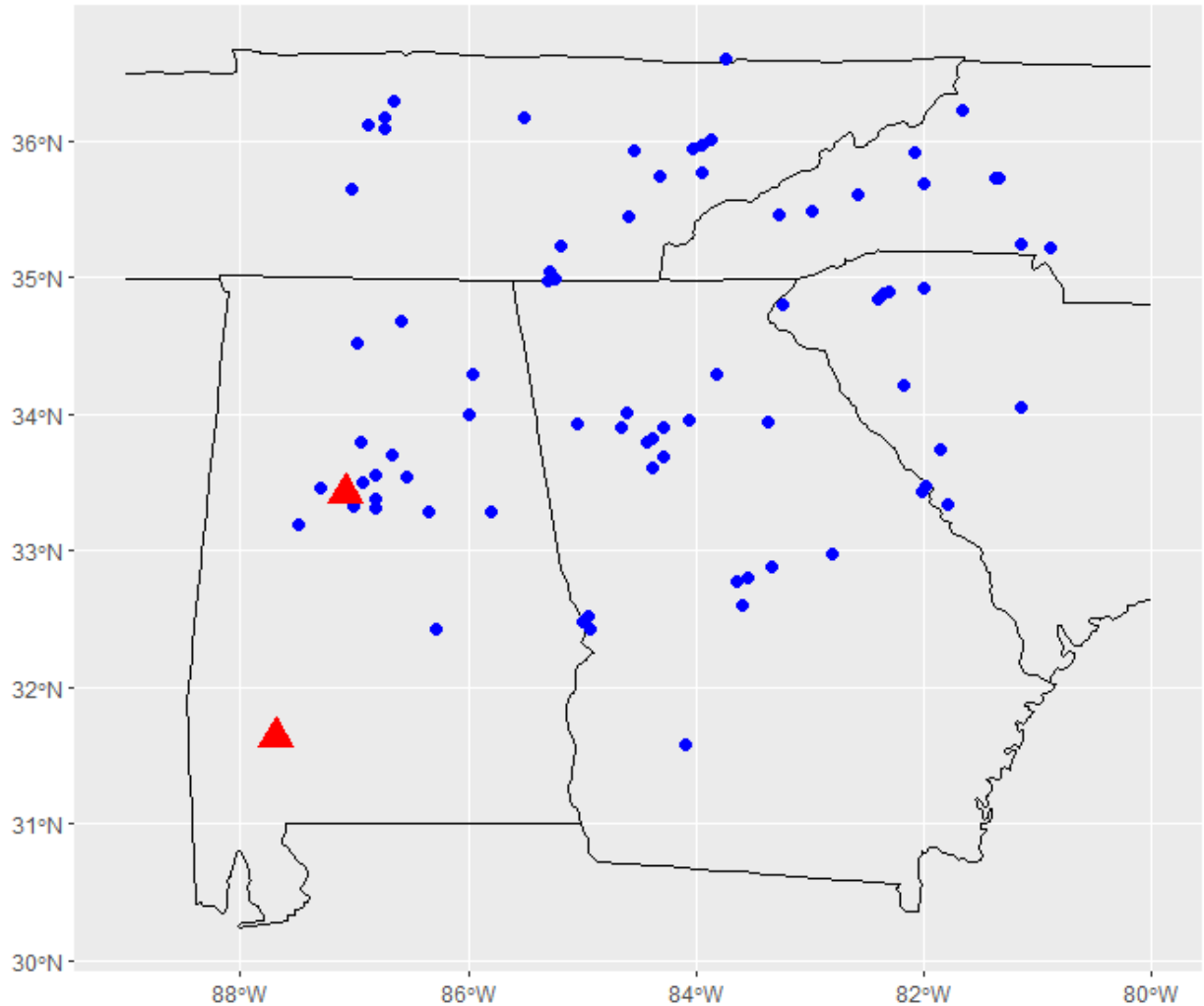
uncertainty intervals for bayesian

Some evidence that the covariate is useful:

- Note the large uncertainties associated with τ^2 and ρ
- The interval for β_0 also increased.

Prediction Test

We will examine predictions at two locations (red): one far away from monitors and one close to monitors.



Posterior Predictive Distributions

Far Location

	Bayesian	Non-Bayesian
Estimate	7.77	7.47
SD/SE	5.84	5.44
95% PI or CI	(-4.2, 18.8)	(-3.08, 18.0)

Close Location

	Bayesian	Non-Bayesian
Estimate	6.89	6.84
SD/SE	4.45	3.85
95% PI or CI	(-2.06, 15.7)	(-0.711, 14.4)

Note:

- Point predictions are very similar.
- Uncertainty at farther location is higher.
- Bayesian estimates give larger standard errors and interval estimates.