Spatial Statistics: Point-referenced Data

Samuel I. Berchuck
STA 440L, Duke University

March 7, 2023

(Slides were adapted from notes by Joshua Warren, PhD)

Goals of a Point-referenced Analysis

1. Estimation and explanation

Goals of a Point-referenced Analysis

1. Estimation and explanation
   - Typical regression parameter estimation

Goals of a Point-referenced Analysis

1. Estimation and explanation
   - Typical regression parameter estimation
     - How does temperature change across the domain (large-scale)?

Goals of a Point-referenced Analysis

1. Estimation and explanation
   - Typical regression parameter estimation
     - How does temperature change across the domain (large-scale)?
2. Prediction at unobserved locations

Goals of a Point-referenced Analysis

1. Estimation and explanation
   - Typical regression parameter estimation
     - How does temperature change across the domain (large-scale)?
2. Prediction at unobserved locations
   - Original development of spatial methods

Goals of a Point-referenced Analysis

1. Estimation and explanation
    - Typical regression parameter estimation
        - How does temperature change across the domain (large-scale)?
2. Prediction at unobserved locations
    - Original development of spatial methods
    - Kriging named after D.G. Krige (mining applications)

Goals of a Point-referenced Analysis

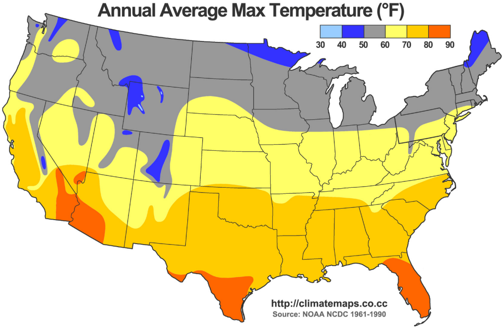
1. Estimation and explanation
   - Typical regression parameter estimation
     - How does temperature change across the domain (large-scale)?
2. Prediction at unobserved locations
   - Original development of spatial methods
   - Kriging named after D.G. Krige (mining applications)
3. Design issues

Goals of a Point-referenced Analysis

1. Estimation and explanation
   - Typical regression parameter estimation
     - How does temperature change across the domain (large-scale)?
2. Prediction at unobserved locations
   - Original development of spatial methods
   - Kriging named after D.G. Krige (mining applications)
3. Design issues
   - Where to put a new air pollution monitor to optimize future prediction criteria?

Goals of a Point-referenced Analysis

- Estimation and Explanation



Goals of a Point-referenced Analysis

- Spatial Prediction

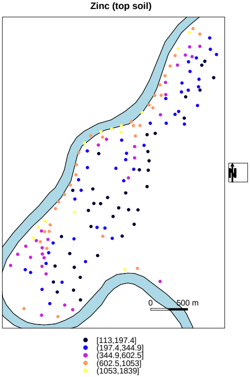


Figure: Observed Data

Goals of a Point-referenced Analysis

- Spatial Prediction

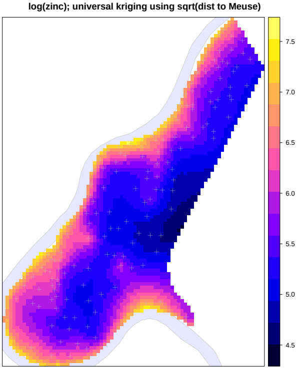


Figure: Predictions

Goals of a Point-referenced Analysis

- Spatial Prediction

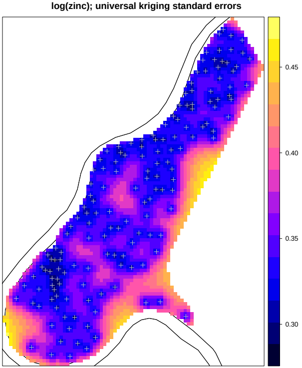


Figure: Standard Errors

Point-referenced Modeling

- Observations closer in space tend to be more similar

Point-referenced Modeling

- Observations closer in space tend to be more similar
- Common regression models assume independence among observations

Point-referenced Modeling

- Observations closer in space tend to be more similar
- Common regression models assume independence among observations
  - Not a valid assumption here, especially at short distances

Point-referenced Modeling

- Observations closer in space tend to be more similar
- Common regression models assume independence among observations
  - Not a valid assumption here, especially at short distances
- Multivariate normal distribution with valid spatial covariance function used in Bayesian modeling

Point-referenced Modeling

- Observations closer in space tend to be more similar
- Common regression models assume independence among observations
  - Not a valid assumption here, especially at short distances
- Multivariate normal distribution with valid spatial covariance function used in Bayesian modeling
  - Spatial covariance describes how observations are correlated based on their proximity to each other

Point-referenced Modeling

- Observations closer in space tend to be more similar
- Common regression models assume independence among observations
  - Not a valid assumption here, especially at short distances
- Multivariate normal distribution with valid spatial covariance function used in Bayesian modeling
  - Spatial covariance describes how observations are correlated based on their proximity to each other
- Advanced models built on similar ideas

Point-referenced Modeling

- Observations closer in space tend to be more similar
- Common regression models assume independence among observations
  - Not a valid assumption here, especially at short distances
- Multivariate normal distribution with valid spatial covariance function used in Bayesian modeling
  - Spatial covariance describes how observations are correlated based on their proximity to each other
- Advanced models built on similar ideas
  - Latent processes often used

Point-referenced Modeling

$$Y(\boldsymbol{s}) = \mu(\boldsymbol{s}) + e(\boldsymbol{s});\ \boldsymbol{s} \in \boldsymbol{D}$$

- $\mu(\boldsymbol{s})$: deterministic large scale trend

Point-referenced Modeling

$$Y(\boldsymbol{s}) = \mu(\boldsymbol{s}) + e(\boldsymbol{s})\,; \;\; \boldsymbol{s} \in \boldsymbol{D}$$

- $\mu(\boldsymbol{s})$: deterministic large scale trend
- $e(\boldsymbol{s})$: error term, small scale structure

Point-referenced Modeling

$$Y(\boldsymbol{s}) = \mu(\boldsymbol{s}) + e(\boldsymbol{s})\,;\ \ \boldsymbol{s} \in \boldsymbol{D}$$

- $\mu(\boldsymbol{s})$: deterministic large scale trend
- $e(\boldsymbol{s})$: error term, small scale structure
- Observe $\{Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n)\} \in \mathbb{R}^p$ from the domain $\boldsymbol{D}$

Point-referenced Modeling

$$Y(\boldsymbol{s}) = \mu(\boldsymbol{s}) + e(\boldsymbol{s});\ \boldsymbol{s} \in \boldsymbol{D}$$

- $\mu(\boldsymbol{s})$: deterministic large scale trend
- $e(\boldsymbol{s})$: error term, small scale structure
- Observe $\{Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n)\} \in \mathbb{R}^p$ from the domain $\boldsymbol{D}$
  - $p > 1$ in spatial setting (usually 2 or 3)

Point-referenced Modeling

$$Y(\boldsymbol{s}) = \mu(\boldsymbol{s}) + e(\boldsymbol{s})\,;\;\; \boldsymbol{s} \in \boldsymbol{D}$$

- $\mu(\boldsymbol{s})$: deterministic large scale trend
- $e(\boldsymbol{s})$: error term, small scale structure
- Observe $\{Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n)\} \in \mathbb{R}^p$ from the domain $\boldsymbol{D}$
  - $p > 1$ in spatial setting (usually 2 or 3)
  - $p = 1$ is a time series (spatial process on a line)

Point-referenced Modeling

$$Y(\boldsymbol{s}) = \mu(\boldsymbol{s}) + e(\boldsymbol{s});\ \boldsymbol{s} \in \boldsymbol{D}$$

- $\mu(\boldsymbol{s})$: deterministic large scale trend
- $e(\boldsymbol{s})$: error term, small scale structure
- Observe $\{Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n)\} \in \mathbb{R}^p$ from the domain $\boldsymbol{D}$
  - $p > 1$ in spatial setting (usually 2 or 3)
  - $p = 1$ is a time series (spatial process on a line)
  - Inference based on this partial realization of the spatial process

Point-referenced Modeling

Large Scale Structure

- $\mu(\boldsymbol{s})$ often modeled as a function of components of $\boldsymbol{s}$; $\boldsymbol{s} = (x, y)$ or $\boldsymbol{s} = (\text{lat}, \text{long})$ in two dimensions

Point-referenced Modeling

Large Scale Structure

- $\mu(\boldsymbol{s})$ often modeled as a function of components of $\boldsymbol{s}$; $\boldsymbol{s} = (x, y)$ or $\boldsymbol{s} = (\text{lat}, \text{long})$ in two dimensions
- Choice of the form of $\mu(\boldsymbol{s})$ depends on exploratory analysis

Point-referenced Modeling

Large Scale Structure

- $\mu(\boldsymbol{s})$ often modeled as a function of components of $\boldsymbol{s}$; $\boldsymbol{s} = (x, y)$ or $\boldsymbol{s} = (\text{lat}, \text{long})$ in two dimensions
- Choice of the form of $\mu(\boldsymbol{s})$ depends on exploratory analysis
- Often polynomial form used such that $\mu(\boldsymbol{s}) = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 xy + \beta_4 x^2 + \beta_5 y^2 + \ldots$

Point-referenced Modeling

Large Scale Structure

- $\mu(\boldsymbol{s})$ often modeled as a function of components of $\boldsymbol{s}$; $\boldsymbol{s} = (x, y)$ or $\boldsymbol{s} = (\text{lat}, \text{long})$ in two dimensions
- Choice of the form of $\mu(\boldsymbol{s})$ depends on exploratory analysis
- Often polynomial form used such that $\mu(\boldsymbol{s}) = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 xy + \beta_4 x^2 + \beta_5 y^2 + \ldots$
- Goal: Remove trend without overfitting

Point-referenced Modeling

Small Scale Structure

- $e(\boldsymbol{s}) = w(\boldsymbol{s}) + \epsilon(\boldsymbol{s})$

Point-referenced Modeling

Small Scale Structure

- $e(\boldsymbol{s}) = w(\boldsymbol{s}) + \epsilon(\boldsymbol{s})$
- Nugget effect: $\epsilon(\boldsymbol{s})$

Point-referenced Modeling

Small Scale Structure

- $e(\boldsymbol{s}) = w(\boldsymbol{s}) + \epsilon(\boldsymbol{s})$
- Nugget effect: $\epsilon(\boldsymbol{s})$
  - Measurement error of the process

Point-referenced Modeling

Small Scale Structure

- $e(\boldsymbol{s}) = w(\boldsymbol{s}) + \epsilon(\boldsymbol{s})$
- Nugget effect: $\epsilon(\boldsymbol{s})$
  - Measurement error of the process
- $w(\boldsymbol{s})$ is the remaining error (purely spatial)

Point-referenced Modeling

Small Scale Structure

- $e(\boldsymbol{s}) = w(\boldsymbol{s}) + \epsilon(\boldsymbol{s})$
- Nugget effect: $\epsilon(\boldsymbol{s})$
  - Measurement error of the process
- $w(\boldsymbol{s})$ is the remaining error (purely spatial)
  - Zero mean random process with spatial covariance function

Point-referenced Modeling

Small Scale Structure

- $e(\boldsymbol{s}) = w(\boldsymbol{s}) + \epsilon(\boldsymbol{s})$
- Nugget effect: $\epsilon(\boldsymbol{s})$
  - Measurement error of the process
- $w(\boldsymbol{s})$ is the remaining error (purely spatial)
  - Zero mean random process with spatial covariance function
- Choice of spatial covariance/semivariogram function based on sample semivariogram analysis

Point-referenced Modeling

Common Modeling Assumptions for $w(\boldsymbol{s})$

- Stationarity: Constant mean and $\mathrm{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}')\} = \mathrm{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}+\boldsymbol{h})\} = C(\boldsymbol{h})$

Point-referenced Modeling

Common Modeling Assumptions for $w(\boldsymbol{s})$

- Stationarity: Constant mean and $\text{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}')\} = \text{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}+\boldsymbol{h})\} = C(\boldsymbol{h})$
  - $w(\boldsymbol{s})$ has a constant mean of zero

Point-referenced Modeling

Common Modeling Assumptions for $w(\boldsymbol{s})$

- Stationarity: Constant mean and
  $\text{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}')\} = \text{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}+\boldsymbol{h})\} = C(\boldsymbol{h})$
  - $w(\boldsymbol{s})$ has a constant mean of zero
  - Same $\boldsymbol{h}$ vector leads to same covariance

Point-referenced Modeling

Common Modeling Assumptions for $w(\boldsymbol{s})$

- Stationarity: Constant mean and
  $\text{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}')\} = \text{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}+\boldsymbol{h})\} = C(\boldsymbol{h})$
  - $w(\boldsymbol{s})$ has a constant mean of zero
  - Same $\boldsymbol{h}$ vector leads to same covariance
- Isotropy:

Point-referenced Modeling

Common Modeling Assumptions for $w(\boldsymbol{s})$

- Stationarity: Constant mean and $\text{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}')\} = \text{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}+\boldsymbol{h})\} = C(\boldsymbol{h})$
  - $w(\boldsymbol{s})$ has a constant mean of zero
  - Same $\boldsymbol{h}$ vector leads to same covariance
- Isotropy:
  - $\text{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}')\} = \text{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}+\boldsymbol{h})\} = C(\|\boldsymbol{h}\|)$

Point-referenced Modeling

Common Modeling Assumptions for $w(\boldsymbol{s})$

- Stationarity: Constant mean and $\text{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}')\} = \text{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}+\boldsymbol{h})\} = C(\boldsymbol{h})$
  - $w(\boldsymbol{s})$ has a constant mean of zero
  - Same $\boldsymbol{h}$ vector leads to same covariance
- Isotropy:
  - $\text{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}')\} = \text{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}+\boldsymbol{h})\} = C(\|\boldsymbol{h}\|)$
  - Covariance depends only on Euclidean distance between the locations

Point-referenced Modeling

Common Modeling Assumptions for $w(\boldsymbol{s})$

- Stationarity: Constant mean and $\text{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}')\} = \text{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}+\boldsymbol{h})\} = C(\boldsymbol{h})$
  - $w(\boldsymbol{s})$ has a constant mean of zero
  - Same $\boldsymbol{h}$ vector leads to same covariance
- Isotropy:
  - $\text{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}')\} = \text{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}+\boldsymbol{h})\} = C(\|\boldsymbol{h}\|)$
  - Covariance depends only on Euclidean distance between the locations
  - **Most common assumption in applied spatial modeling!**

Point-referenced Modeling

Common Covariance Functions (Isotropic)

- Exponential:
  - $C(\|\boldsymbol{h}\|) = \begin{cases} \sigma^2 \exp\{-\phi \|\boldsymbol{h}\|\} & \|\boldsymbol{h}\| > 0 \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$
- $\text{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}+\boldsymbol{h})\} \to 0$ as $\|h\|$ gets large
- $\text{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}+\boldsymbol{h})\} = \tau^2 + \sigma^2$ when $\|h\| = 0$

Point-referenced Modeling

Common Covariance Functions (Isotropic)

- Gaussian:
  - $C(\|\boldsymbol{h}\|) = \begin{cases} \sigma^2 \exp\left\{-\phi^2 \|\boldsymbol{h}\|^2\right\} & \|\boldsymbol{h}\| > 0 \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$
- $\text{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}+\boldsymbol{h})\} \to 0$ as $\|h\|$ gets large
- $\text{Cov}\{w(\boldsymbol{s}), w(\boldsymbol{s}+\boldsymbol{h})\} = \tau^2 + \sigma^2$ when $\|h\| = 0$

Point-referenced Modeling

Common Covariance Functions (Isotropic)

- Spherical:
  - $C\left(\|\boldsymbol{h}\|\right) = \begin{cases} 0 & \|\boldsymbol{h}\| > \frac{1}{\phi} \\ \sigma^2\left\{1 - \frac{3\phi\|\boldsymbol{h}\|}{2} + \frac{1}{2}\left(\phi\|\boldsymbol{h}\|\right)^3\right\} & 0 < \|\boldsymbol{h}\| \le \frac{1}{\phi} \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$
- $\text{Cov}\left\{w\left(\boldsymbol{s}\right), w\left(\boldsymbol{s} + \boldsymbol{h}\right)\right\} = 0$ for $\|h\| \ge \frac{1}{\phi}$
- $\text{Cov}\left\{w\left(\boldsymbol{s}\right), w\left(\boldsymbol{s} + \boldsymbol{h}\right)\right\} = \tau^2 + \sigma^2$ when $\|h\| = 0$
- Independence after certain distance

Point-referenced Modeling

Covariance Function Parameters (Isotropic Case)

- Nugget: τ^2; Measurement error variance

Point-referenced Modeling

Covariance Function Parameters (Isotropic Case)

- Nugget: τ^2; Measurement error variance
- Partial sill: σ^2 ; Spatial process variance

Point-referenced Modeling

Covariance Function Parameters (Isotropic Case)

- Nugget: τ^2; Measurement error variance
- Partial sill: σ^2 ; Spatial process variance
  - Sill: $\tau^2 + \sigma^2$ (total variance)

Point-referenced Modeling

Covariance Function Parameters (Isotropic Case)

- Nugget: τ^2; Measurement error variance
- Partial sill: σ^2 ; Spatial process variance
  - Sill: $\tau^2 + \sigma^2$ (total variance)
- Decay Parameter: ϕ; Describes the strength of spatial correlation

Point-referenced Modeling

Covariance Function Parameters (Isotropic Case)

- Nugget: τ^2; Measurement error variance
- Partial sill: σ^2 ; Spatial process variance
  - Sill: $\tau^2 + \sigma^2$ (total variance)
- Decay Parameter: ϕ; Describes the strength of spatial correlation
  - Range: Distance at which the correlation is zero (spherical)

Point-referenced Modeling

Covariance Function Parameters (Isotropic Case)

- Nugget: τ^2; Measurement error variance
- Partial sill: σ^2 ; Spatial process variance
  - Sill: $\tau^2 + \sigma^2$ (total variance)
- Decay Parameter: ϕ; Describes the strength of spatial correlation
  - Range: Distance at which the correlation is zero (spherical)
  - Effective Range: Distance at which the correlation is 0.05 (most other structures)

Point-referenced Modeling

Topics Not Discussed Here:

- How to choose the appropriate spatial covariance matrix?

Point-referenced Modeling

Topics Not Discussed Here:

- How to choose the appropriate spatial covariance matrix?
  - Frequentist spatial course covers this extensively

Point-referenced Modeling

Topics Not Discussed Here:

- How to choose the appropriate spatial covariance matrix?
  - Frequentist spatial course covers this extensively
- How to handle anisotropy?

Point-referenced Modeling

Topics Not Discussed Here:

- How to choose the appropriate spatial covariance matrix?
  - Frequentist spatial course covers this extensively
- How to handle anisotropy?
- How to handle nonstationarity?

Point-referenced Modeling

Topics Not Discussed Here:

- How to choose the appropriate spatial covariance matrix?
  - Frequentist spatial course covers this extensively
- How to handle anisotropy?
- How to handle nonstationarity?
  - Flexible Bayesian models are ideal in these settings

Point-referenced Modeling

Benefits of Fully Bayesian Analysis of Spatial Data:

- Classic frequentist methods ignore the uncertainty in the estimated covariance matrix (assume parameters are known)

Point-referenced Modeling

Benefits of Fully Bayesian Analysis of Spatial Data:

- Classic frequentist methods ignore the uncertainty in the estimated covariance matrix (assume parameters are known)
    - Including ϕ, σ^2, and τ^2

Point-referenced Modeling

Benefits of Fully Bayesian Analysis of Spatial Data:

- Classic frequentist methods ignore the uncertainty in the estimated covariance matrix (assume parameters are known)
  - Including ϕ, σ^2, and τ^2
- Bayesian analysis accounts for this uncertainty and leads to improved estimation/prediction of the process

Point-referenced Modeling

Benefits of Fully Bayesian Analysis of Spatial Data:

- Classic frequentist methods ignore the uncertainty in the estimated covariance matrix (assume parameters are known)
  - Including ϕ, σ^2, and τ^2
- Bayesian analysis accounts for this uncertainty and leads to improved estimation/prediction of the process
  - Prior distributions placed on these parameters

Point-referenced Modeling

Benefits of Fully Bayesian Analysis of Spatial Data:

- Classic frequentist methods ignore the uncertainty in the estimated covariance matrix (assume parameters are known)
  - Including ϕ, σ^2, and τ^2
- Bayesian analysis accounts for this uncertainty and leads to improved estimation/prediction of the process
  - Prior distributions placed on these parameters
- Efficient MCMC algorithms available to fit basic models

Point-referenced Modeling

Benefits of Fully Bayesian Analysis of Spatial Data:

- Classic frequentist methods ignore the uncertainty in the estimated covariance matrix (assume parameters are known)
  - Including ϕ, σ^2, and τ^2
- Bayesian analysis accounts for this uncertainty and leads to improved estimation/prediction of the process
  - Prior distributions placed on these parameters
- Efficient MCMC algorithms available to fit basic models
- Note: Proc Mixed and Glimmix can handle frequentist spatial models using MLE (can be very slow)

Bayesian Model Fitting

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

$$\boldsymbol{s}_i \in \boldsymbol{D}; i = 1, \ldots, n$$

- $Y(\boldsymbol{s}_i)$: observation at location $\boldsymbol{s}_i$

Bayesian Model Fitting

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

$$\boldsymbol{s}_i \in \boldsymbol{D}; i = 1, \ldots, n$$

- $Y(\boldsymbol{s}_i)$: observation at location $\boldsymbol{s}_i$
- $\mathbf{x}(\boldsymbol{s}_i)$: vector of site specific covariates

Bayesian Model Fitting

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

$$\boldsymbol{s}_i \in \boldsymbol{D}; i = 1, \ldots, n$$

- $Y(\boldsymbol{s}_i)$: observation at location $\boldsymbol{s}_i$
- $\mathbf{x}(\boldsymbol{s}_i)$: vector of site specific covariates
- $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$: vector of regression parameters

Bayesian Model Fitting

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

$$\boldsymbol{s}_i \in \boldsymbol{D}; i = 1, \ldots, n$$

- $Y(\boldsymbol{s}_i)$: observation at location $\boldsymbol{s}_i$
- $\mathbf{x}(\boldsymbol{s}_i)$: vector of site specific covariates
- $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$: vector of regression parameters
- $\epsilon(\boldsymbol{s}_i) \overset{iid}{\sim} \mathrm{N}(0, \sigma_\epsilon^2)$: σ_ϵ^2 is nugget effect

Bayesian Model Fitting

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

$$\boldsymbol{s}_i \in \boldsymbol{D}; i = 1, \ldots, n$$

- $Y(\boldsymbol{s}_i)$: observation at location $\boldsymbol{s}_i$
- $\mathbf{x}(\boldsymbol{s}_i)$: vector of site specific covariates
- $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$: vector of regression parameters
- $\epsilon(\boldsymbol{s}_i) \overset{iid}{\sim} \mathrm{N}(0, \sigma_\epsilon^2)$: σ_ϵ^2 is nugget effect
- $w(\boldsymbol{s}_i)$: spatially correlated "random effect" (in frequentist terms)

Bayesian Model Fitting

Prior and Hyperprior Distributions:

$$\boldsymbol{w} = \{w(\boldsymbol{s}_1), \ldots, w(\boldsymbol{s}_n)\}^T \sim \text{MVN}\left(\boldsymbol{0}_n, \sigma_w^2 \Sigma(\phi)\right)$$

- $\Sigma(\phi)$: spatial correlation matrix which depends on unknown parameter(s) ϕ

Bayesian Model Fitting

Prior and Hyperprior Distributions:

$$\boldsymbol{w} = \{w(\boldsymbol{s}_1), \ldots, w(\boldsymbol{s}_n)\}^T \sim \text{MVN}\left(\boldsymbol{0}_n, \sigma_w^2 \Sigma(\phi)\right)$$

- $\Sigma(\phi)$: spatial correlation matrix which depends on unknown parameter(s) ϕ
- $\sigma_\epsilon^2 \sim$ Inverse Gamma or Uniform

Bayesian Model Fitting

Prior and Hyperprior Distributions:

$$\boldsymbol{w} = \{w(\boldsymbol{s}_1), \ldots, w(\boldsymbol{s}_n)\}^T \sim \mathrm{MVN}\left(\mathbf{0}_n, \sigma_w^2 \Sigma(\phi)\right)$$

- $\Sigma(\phi)$: spatial correlation matrix which depends on unknown parameter(s) ϕ
- $\sigma_\epsilon^2 \sim$ Inverse Gamma or Uniform
- $\sigma_w^2 \sim$ Inverse Gamma or Uniform

Bayesian Model Fitting

Prior and Hyperprior Distributions:

$$\boldsymbol{w} = \{w(\boldsymbol{s}_1), \ldots, w(\boldsymbol{s}_n)\}^T \sim \text{MVN}\left(\mathbf{0}_n, \sigma_w^2 \Sigma(\phi)\right)$$

- $\Sigma(\phi)$: spatial correlation matrix which depends on unknown parameter(s) ϕ
- $\sigma_\epsilon^2 \sim$ Inverse Gamma or Uniform
- $\sigma_w^2 \sim$ Inverse Gamma or Uniform
- $\boldsymbol{\beta}$: flat prior $\propto 1$ or MVN$\left(\mathbf{0}_{p+1}, \sigma_\beta^2 \boldsymbol{I}_{p+1}\right)$; σ_β^2 fixed, large

Bayesian Model Fitting

Prior and Hyperprior Distributions:

$$\boldsymbol{w} = \{w(\boldsymbol{s}_1), \ldots, w(\boldsymbol{s}_n)\}^T \sim \text{MVN}\left(\mathbf{0}_n, \sigma_w^2 \Sigma(\phi)\right)$$

- $\Sigma(\phi)$: spatial correlation matrix which depends on unknown parameter(s) ϕ
- $\sigma_\epsilon^2 \sim$ Inverse Gamma or Uniform
- $\sigma_w^2 \sim$ Inverse Gamma or Uniform
- $\boldsymbol{\beta}$: flat prior $\propto 1$ or $\text{MVN}\left(\mathbf{0}_{p+1}, \sigma_\beta^2 I_{p+1}\right)$; σ_β^2 fixed, large
- ϕ: typically only a single parameter; Uniform or Gamma $(\phi > 0)$

Bayesian Model Fitting

- **Fitting the Model (Two Options):**

Bayesian Model Fitting

- **Fitting the Model (Two Options):**
    1. Conditional on the spatial effects ($w(\boldsymbol{s}_i)$)

Bayesian Model Fitting

- **Fitting the Model (Two Options):**
  1. Conditional on the spatial effects $\left(w\left(\boldsymbol{s}_i\right)\right)$
  2. Unconditionally, marginalizing over the spatial effects

Bayesian Model Fitting

- **Fitting the Model (Two Options):**
  1. Conditional on the spatial effects $(w(\boldsymbol{s}_i))$
  2. Unconditionally, marginalizing over the spatial effects
- Pros and cons to implementing each option

Fitting the Model: Option 1

- $\boldsymbol{Y}|\beta, \boldsymbol{w}, \sigma_\epsilon^2 \sim \text{MVN}\left(X\beta + \boldsymbol{w}, \sigma_\epsilon^2 I_n\right)$

Fitting the Model: Option 1

- $\boldsymbol{Y}|\beta, \boldsymbol{w}, \sigma_\epsilon^2 \sim \text{MVN}\left(X\beta + \boldsymbol{w}, \sigma_\epsilon^2 I_n\right)$
    - $\boldsymbol{Y} = \left\{Y\left(\boldsymbol{s}_1\right), \ldots, Y\left(\boldsymbol{s}_n\right)\right\}^T$

Fitting the Model: Option 1

- $\boldsymbol{Y}|\beta, \boldsymbol{w}, \sigma_\epsilon^2 \sim \text{MVN}\left(X\beta + \boldsymbol{w}, \sigma_\epsilon^2 I_n\right)$
  - $\boldsymbol{Y} = \left\{Y\left(\boldsymbol{s}_1\right), \ldots, Y\left(\boldsymbol{s}_n\right)\right\}^T$
  - $X = \left\{x\left(\boldsymbol{s}_1\right), \ldots, x\left(\boldsymbol{s}_n\right)\right\}^T$

Fitting the Model: Option 1

- $\boldsymbol{Y}|\beta, \boldsymbol{w}, \sigma_\epsilon^2 \sim \text{MVN}\left(X\beta + \boldsymbol{w}, \sigma_\epsilon^2 I_n\right)$
    - $\boldsymbol{Y} = \left\{Y\left(\boldsymbol{s}_1\right), \ldots, Y\left(\boldsymbol{s}_n\right)\right\}^T$
    - $X = \left\{x\left(\boldsymbol{s}_1\right), \ldots, x\left(\boldsymbol{s}_n\right)\right\}^T$
- We must update the $\boldsymbol{w}$ vector each iteration since we condition on it

Fitting the Model: Option 1

Pros:

- Closed form for σ_ϵ^2 and σ_w^2, allowing for Gibbs sampling.

Cons:

Fitting the Model: Option 1

Pros:

- Closed form for σ_ϵ^2 and σ_w^2, allowing for Gibbs sampling.
- Automatically obtain posterior samples from the spatial effects $\boldsymbol{w}$ which may be of interest in some studies.

Cons:

Fitting the Model: Option 1

Pros:

- Closed form for σ_ϵ^2 and σ_w^2, allowing for Gibbs sampling.
- Automatically obtain posterior samples from the spatial effects $\boldsymbol{w}$ which may be of interest in some studies.

Cons:

- Updating ϕ can sometimes be difficult when the spatial correlation is strong

Fitting the Model: Option 1

Pros:

- Closed form for σ_ϵ^2 and σ_w^2, allowing for Gibbs sampling.
- Automatically obtain posterior samples from the spatial effects $\boldsymbol{w}$ which may be of interest in some studies.

Cons:

- Updating ϕ can sometimes be difficult when the spatial correlation is strong
    - No nugget effect on the diagonal to stabilize the spatial covariance matrix.

Geostatistical Modeling

Fitting the Model: Option 2

- $\boldsymbol{Y}|\boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_w^2, \phi \sim \text{MVN}\left(X\boldsymbol{\beta}, \sigma_\epsilon^2 I_n + \sigma_w^2 \Sigma\left(\phi\right)\right)$

Geostatistical Modeling

Fitting the Model: Option 2

- $\boldsymbol{Y} | \boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_w^2, \phi \sim \text{MVN}\left(X\boldsymbol{\beta}, \sigma_\epsilon^2 I_n + \sigma_w^2 \Sigma(\phi)\right)$
  - $\Sigma(\phi)$ is the spatial correlation matrix.

Geostatistical Modeling

Fitting the Model: Option 2

- $\boldsymbol{Y}|\boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_w^2, \phi \sim \text{MVN}\left(X\boldsymbol{\beta}, \sigma_\epsilon^2 \boldsymbol{I_n} + \sigma_w^2 \Sigma(\phi)\right)$
    - $\Sigma(\phi)$ is the spatial correlation matrix.
    - Full covariance matrix has nugget effect (σ_ϵ^2) on the diagonal

Geostatistical Modeling

Fitting the Model: Option 2

- $\boldsymbol{Y}|\boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_w^2, \phi \sim \text{MVN}\left(X\boldsymbol{\beta}, \sigma_\epsilon^2 \boldsymbol{I_n} + \sigma_w^2 \Sigma(\phi)\right)$
  - $\Sigma(\phi)$ is the spatial correlation matrix.
  - Full covariance matrix has nugget effect (σ_ϵ^2) on the diagonal
- We no longer update the $\boldsymbol{w}$ vector each iteration.

Fitting the Model: Option 2

Pros:

- The process of updating ϕ is improved due to the nugget effect.

Cons:

Fitting the Model: Option 2

Pros:

- The process of updating ϕ is improved due to the nugget effect.
- Less parameters to update (no $w\left(\boldsymbol{s}_i\right)$ parameters).

Cons:

Fitting the Model: Option 2

Pros:

- The process of updating ϕ is improved due to the nugget effect.
- Less parameters to update (no $w(\boldsymbol{s}_i)$ parameters).

Cons:

- No closed form for σ_ϵ^2 and σ_w^2, must use Metropolis-Hastings algorithm.

Fitting the Model: Option 2

Pros:

- The process of updating ϕ is improved due to the nugget effect.
- Less parameters to update (no $w(\boldsymbol{s}_i)$ parameters).

Cons:

- No closed form for σ_ϵ^2 and σ_w^2, must use Metropolis-Hastings algorithm.
- No longer automatically obtain posterior samples from the spatial effects.

Which Option to Choose?:

Most often we choose to implement Option 2 to fit the model

- Better behaved covariance matrix

Which Option to Choose?:

Most often we choose to implement Option 2 to fit the model

- Better behaved covariance matrix
- Reduced parameter space often improves convergence

What about $\boldsymbol{w}$?

- Can we still get posterior samples from $w(\boldsymbol{s}_i)$ using Option 2?

What about $\boldsymbol{w}$?

- Can we still get posterior samples from $w(\boldsymbol{s}_i)$ using Option 2?
- Actually pretty easy to obtain posterior samples from the spatial effects given the available posterior output

What about $\boldsymbol{w}$?

- Can we still get posterior samples from $w(\boldsymbol{s}_i)$ using Option 2?
- Actually pretty easy to obtain posterior samples from the spatial effects given the available posterior output
- $f(\boldsymbol{w}|\boldsymbol{Y}) = \int\int\int\int f\left(\boldsymbol{w}, \boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_w^2, \phi|\boldsymbol{Y}\right) d\boldsymbol{\beta} d\sigma_\epsilon^2 d\sigma_w^2 d\phi$

What about $\boldsymbol{w}$?

- Can we still get posterior samples from $w(\boldsymbol{s}_i)$ using Option 2?
- Actually pretty easy to obtain posterior samples from the spatial effects given the available posterior output
- $f(\boldsymbol{w}|\boldsymbol{Y}) = \int\int\int\int f\left(\boldsymbol{w}, \boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_w^2, \phi|\boldsymbol{Y}\right) d\boldsymbol{\beta} d\sigma_\epsilon^2 d\sigma_w^2 d\phi$
- $= \int\int\int\int f\left(\boldsymbol{w}|\boldsymbol{Y}, \boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_w^2, \phi\right) f\left(\boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_w^2, \phi|\boldsymbol{Y}\right) d\boldsymbol{\beta} d\sigma_\epsilon^2 d\sigma_w^2 d\phi$

What about $\boldsymbol{w}$?

- Can we still get posterior samples from $w(\boldsymbol{s}_i)$ using Option 2?
- Actually pretty easy to obtain posterior samples from the spatial effects given the available posterior output
- $f(\boldsymbol{w}|\boldsymbol{Y}) = \int\int\int\int f\left(\boldsymbol{w}, \boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_w^2, \phi|\boldsymbol{Y}\right) d\boldsymbol{\beta} d\sigma_\epsilon^2 d\sigma_w^2 d\phi$
- $=$ $\int\int\int\int f\left(\boldsymbol{w}|\boldsymbol{Y}, \boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_w^2, \phi\right) f\left(\boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_w^2, \phi|\boldsymbol{Y}\right) d\boldsymbol{\beta} d\sigma_\epsilon^2 d\sigma_w^2 d\phi$
- What is the distribution of $f\left(\boldsymbol{w}|\boldsymbol{Y}, \boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_w^2, \phi\right)$?

What about $\boldsymbol{w}$?

For each posterior sample, we can draw a sample from $f(\boldsymbol{w}|\boldsymbol{Y})$ using the specified multivariate normal distribution.

- Composition sampling

What about $\boldsymbol{w}$?

For each posterior sample, we can draw a sample from $f(\boldsymbol{w}|\boldsymbol{Y})$ using the specified multivariate normal distribution.

- Composition sampling
- A one-for-one draw for each posterior sample

What about $\boldsymbol{w}$?

For each posterior sample, we can draw a sample from $f(\boldsymbol{w}|\boldsymbol{Y})$ using the specified multivariate normal distribution.

- Composition sampling
- A one-for-one draw for each posterior sample
- This is why Option 1 isn't necessary

Spatial Prediction:

- Interest in predicting the process at unobserved spatial location(s)

Spatial Prediction:

- Interest in predicting the process at unobserved spatial location(s)
- Interest in $f\left\{Y\left(\boldsymbol{s}_0\right)\middle|\boldsymbol{Y}\right\}$, the posterior predictive distribution of the process at the new location $\boldsymbol{s}_0$

Spatial Prediction:

- Interest in predicting the process at unobserved spatial location(s)
- Interest in $f\left\{Y\left(\boldsymbol{s}_0\right)\middle|\boldsymbol{Y}\right\}$, the posterior predictive distribution of the process at the new location $\boldsymbol{s}_0$
- Very similar to obtaining draws from $f\left(\boldsymbol{w}\middle|\boldsymbol{Y}\right)$

Spatial Prediction:

- $f\left\{Y\left(\boldsymbol{s}_0\right)|\boldsymbol{Y}\right\} = \int\int\int\int f\left\{Y\left(\boldsymbol{s}_0\right), \boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_w^2, \phi|\boldsymbol{Y}\right\} d\boldsymbol{\beta} d\sigma_\epsilon^2 d\sigma_w^2 d\phi$

Spatial Prediction:

- $f\left\{Y\left(\mathbf{s}_0\right)|\mathbf{Y}\right\} = \int\int\int\int f\left\{Y\left(\mathbf{s}_0\right),\boldsymbol{\beta},\sigma^2_\epsilon,\sigma^2_w,\phi|\mathbf{Y}\right\}d\boldsymbol{\beta}d\sigma^2_\epsilon d\sigma^2_w d\phi$

- $= \int\int\int\int f\left\{Y\left(\mathbf{s}_0\right)|\mathbf{Y},\boldsymbol{\beta},\sigma^2_\epsilon,\sigma^2_w,\phi\right\} f\left(\boldsymbol{\beta},\sigma^2_\epsilon,\sigma^2_w,\phi|\mathbf{Y}\right) d\boldsymbol{\beta}d\sigma^2_\epsilon d\sigma^2_w d\phi$

Spatial Prediction:

- $f\left\{Y\left(\mathbf{s}_0\right)|\mathbf{Y}\right\} = \int\int\int\int f\left\{Y\left(\mathbf{s}_0\right), \boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_w^2, \phi|\mathbf{Y}\right\} d\boldsymbol{\beta} d\sigma_\epsilon^2 d\sigma_w^2 d\phi$
- $= \int\int\int\int f\left\{Y\left(\mathbf{s}_0\right)|\mathbf{Y}, \boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_w^2, \phi\right\} f\left(\boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_w^2, \phi|\mathbf{Y}\right) d\boldsymbol{\beta} d\sigma_\epsilon^2 d\sigma_w^2 d\phi$
- What is the distribution of $f\left\{Y\left(\mathbf{s}_0\right)|\mathbf{Y}, \boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_w^2, \phi\right\}$?

Spatial Prediction

- We don't have to predict only at a single location

Spatial Prediction

- We don't have to predict only at a single location
- Can predict for multiple locations simultaneously

Spatial Prediction

- We don't have to predict only at a single location
- Can predict for multiple locations simultaneously
- Presented results are easily extended to the multivariate setting

Spatial Prediction

- We don't have to predict only at a single location
- Can predict for multiple locations simultaneously
- Presented results are easily extended to the multivariate setting
- Interest in $f\left(\boldsymbol{Y}_0 | \boldsymbol{Y}\right)$ where $\boldsymbol{Y}_0 = \left\{Y\left(\boldsymbol{s}_{0,1}\right), \ldots, Y\left(\boldsymbol{s}_{0,m}\right)\right\}^T$