Introduction to Applied Bayesian Analysis in Wildlife Ecology

Jeffrey W. Doser

May 11, 2024

Bonus material: Hierarchical spatial models

What is spatial data?

What is spatial data?

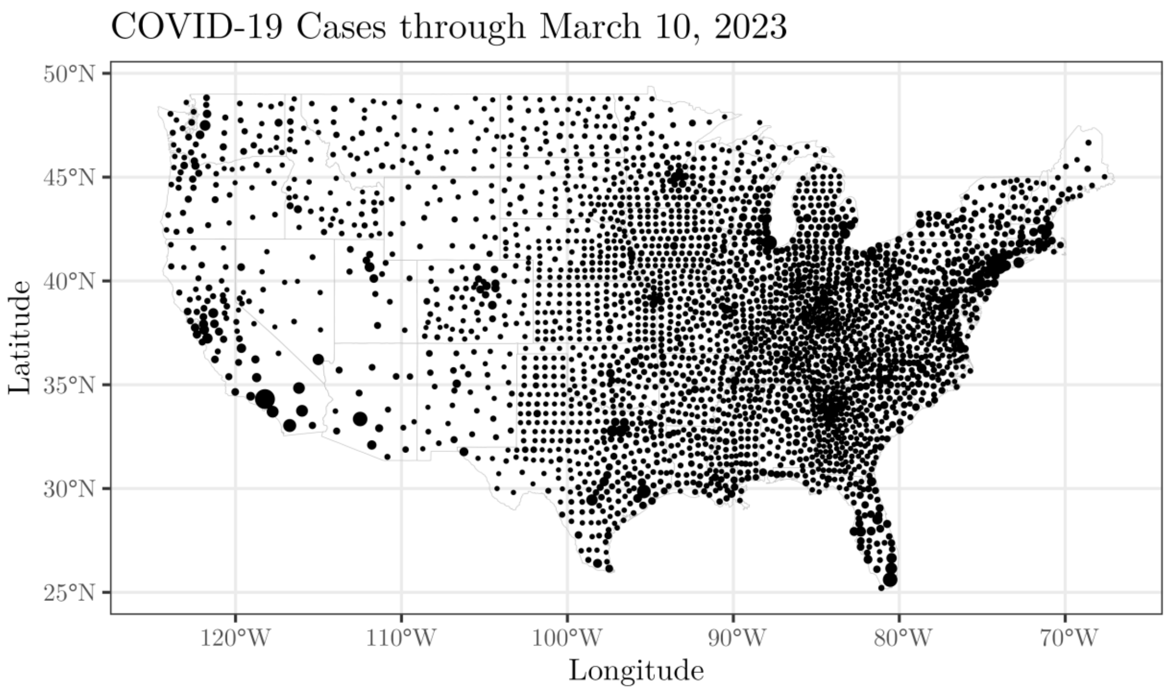
Any type of data that relates to a specific geographical area or location

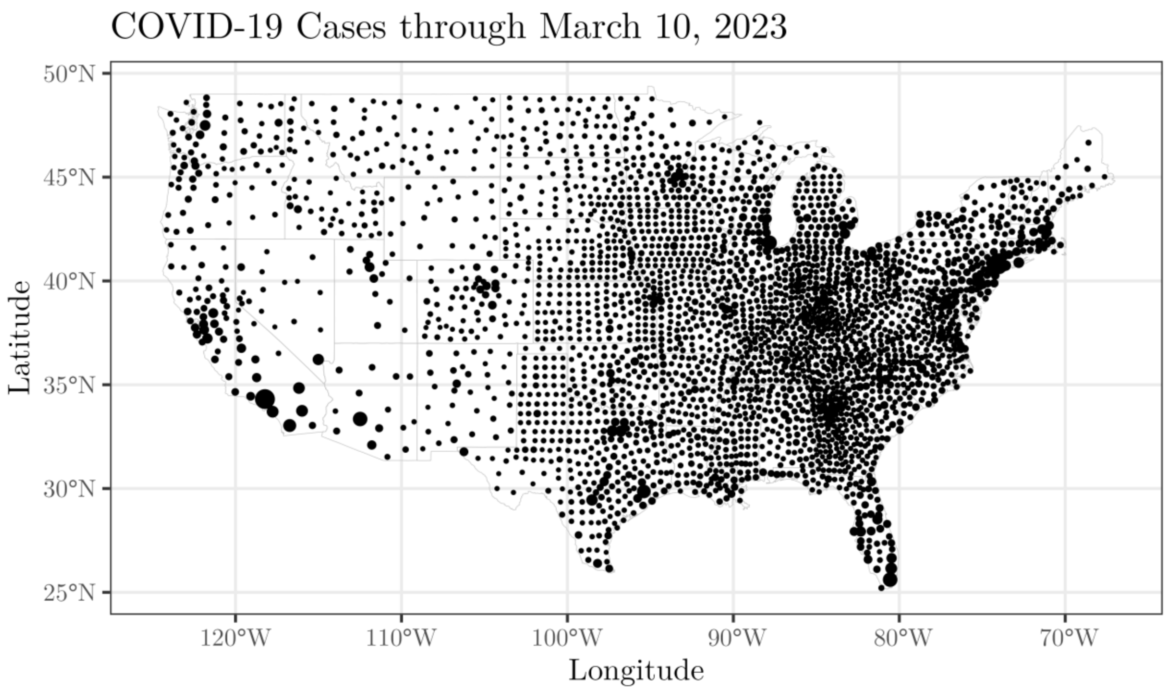
What is spatial data?

Any type of data that relates to a specific geographical area or location

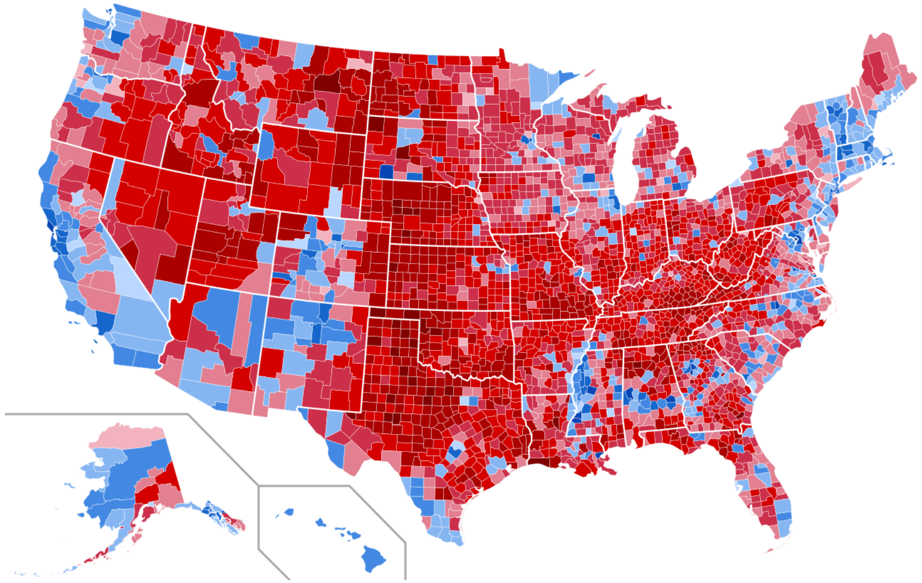
First Law of Geography

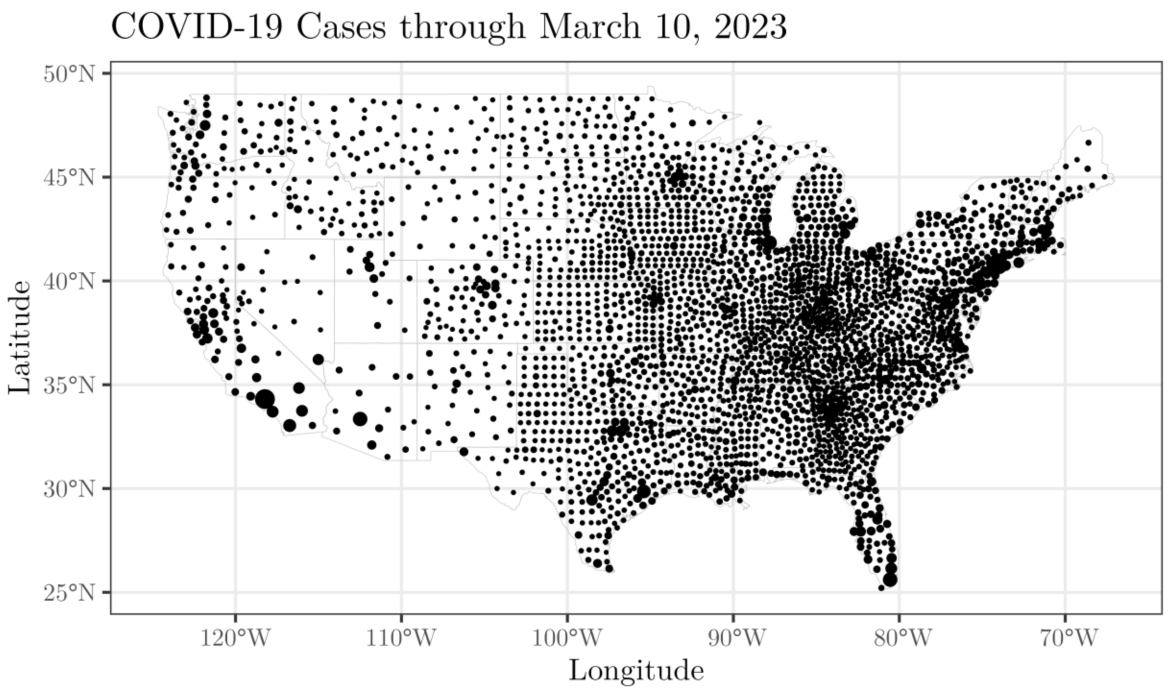
"Everything is related to everything else, but near things are more related than distant things." - Waldo Tobler


COVID-19 Cases through March 10, 2023
50°N
45°N
40°N
35°N
30°N
25°N
Latitude
120°W
110°W
100°W
90°W
80°W
70°W
Longitude

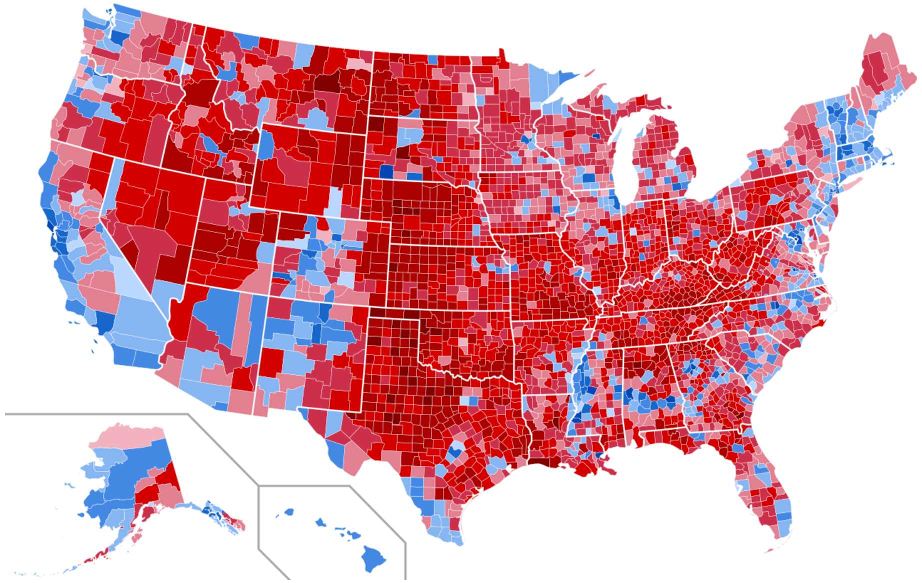


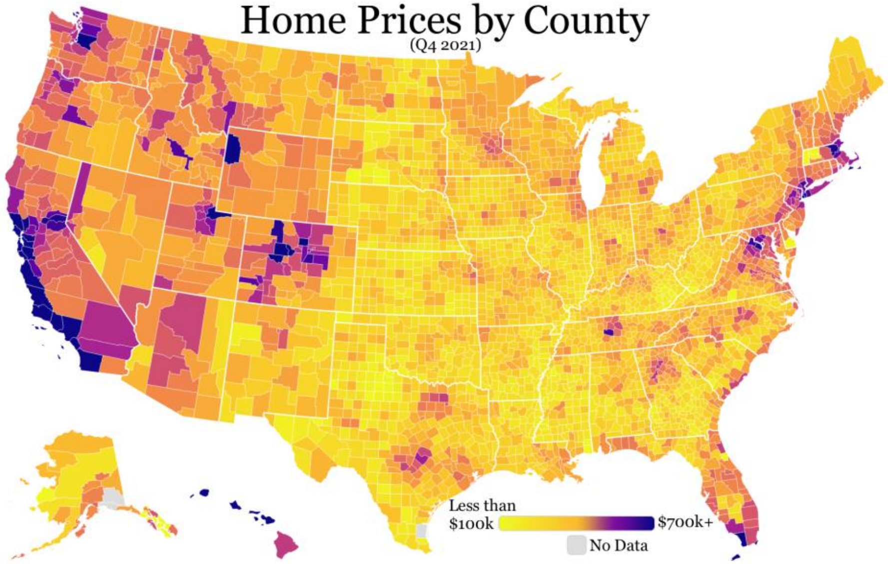
2020 US presidential election map

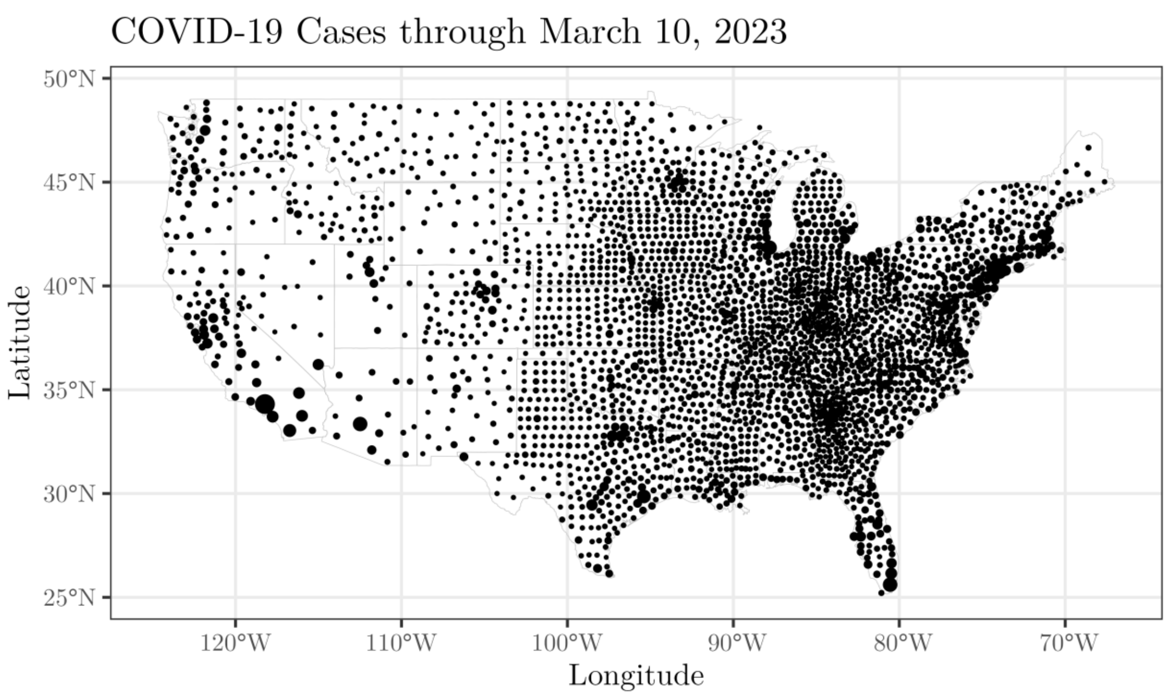




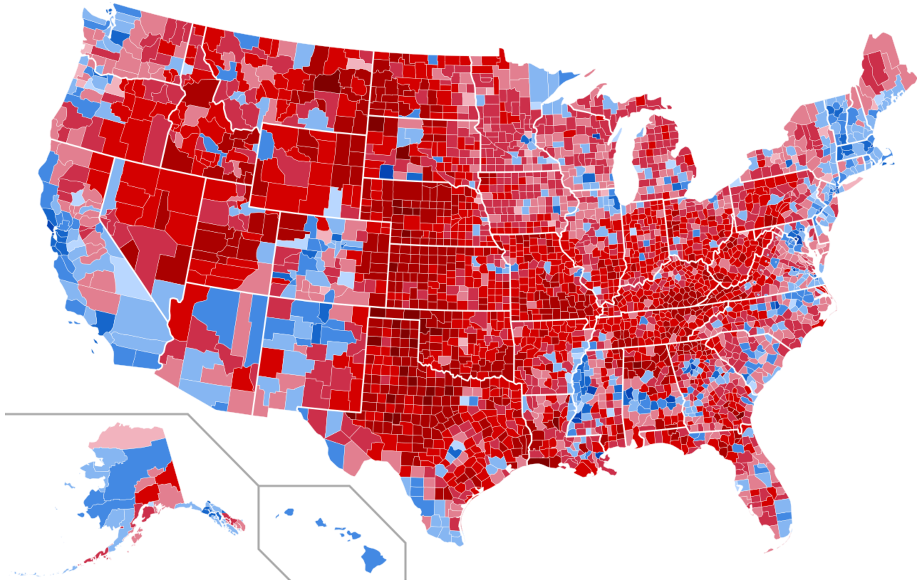
2020 US presidential election map

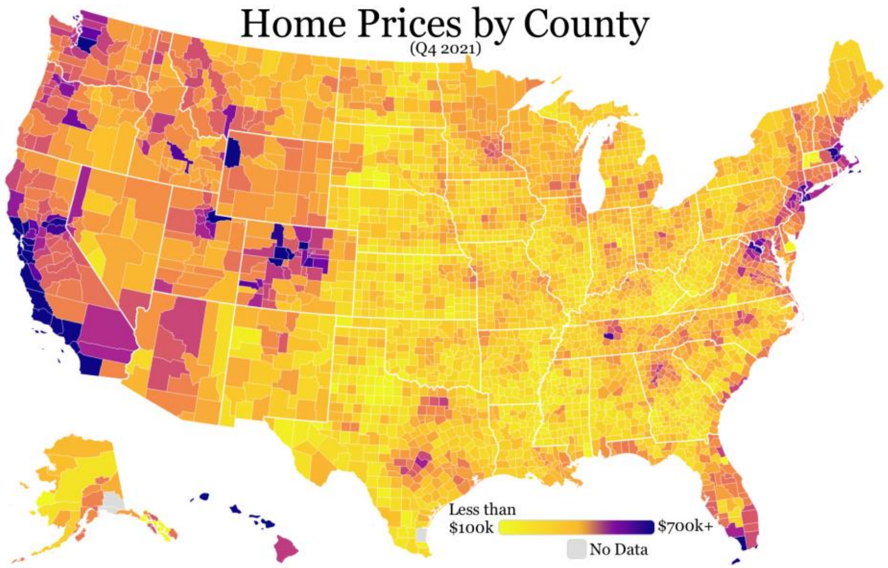


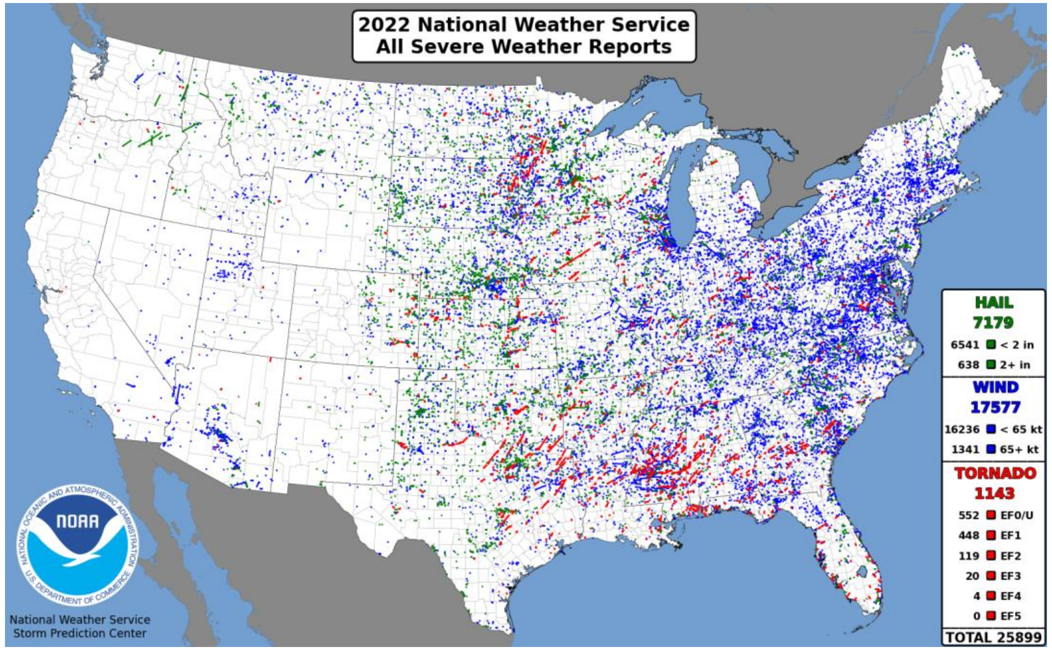




2020 US presidential election map

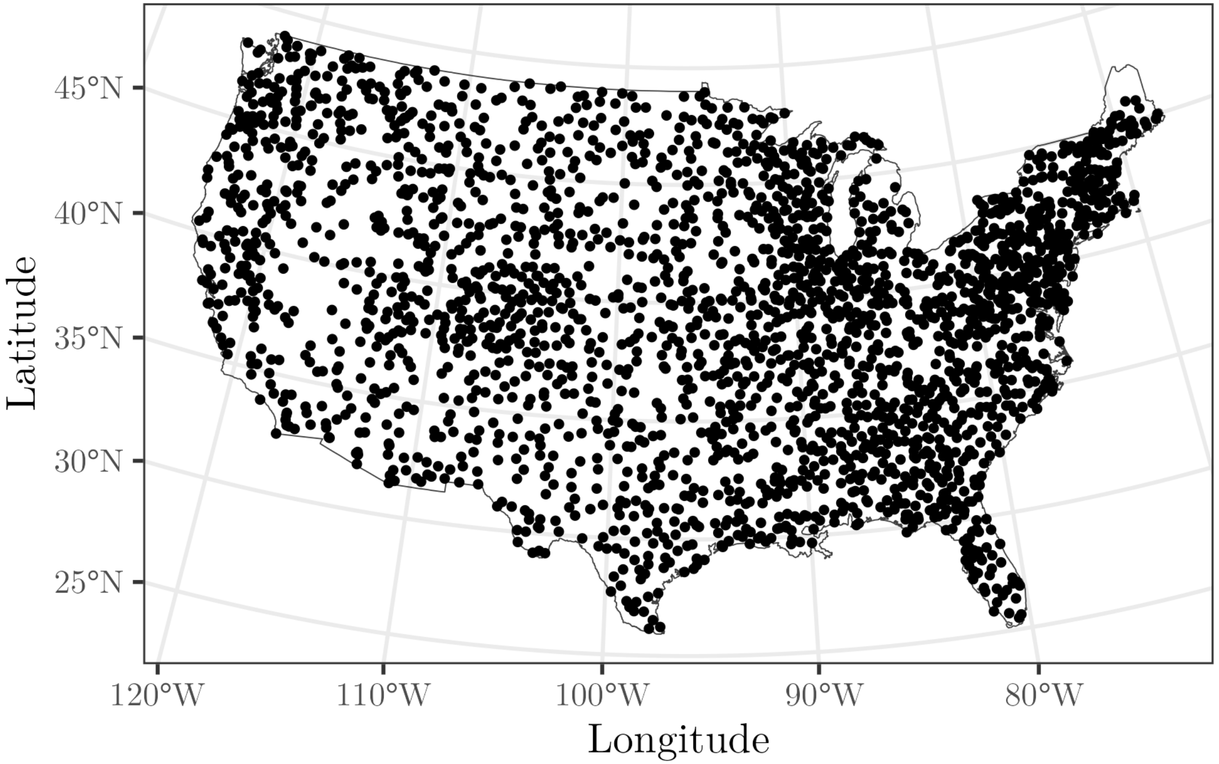






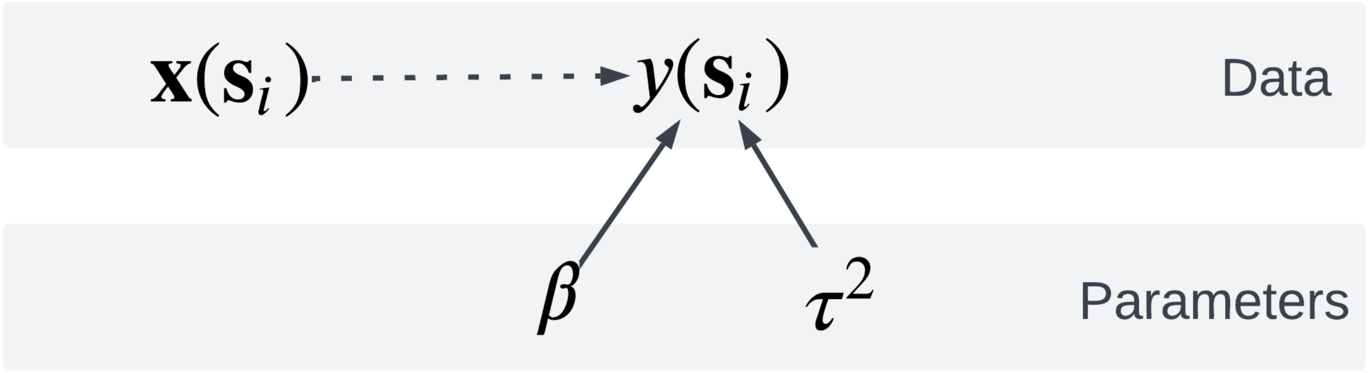
Spatial data

- Some variable is measured at each location
- Spatial coordinates for each location
- Goals:
  - Prediction
  - Inference (e.g., estimate relationships with covariates)



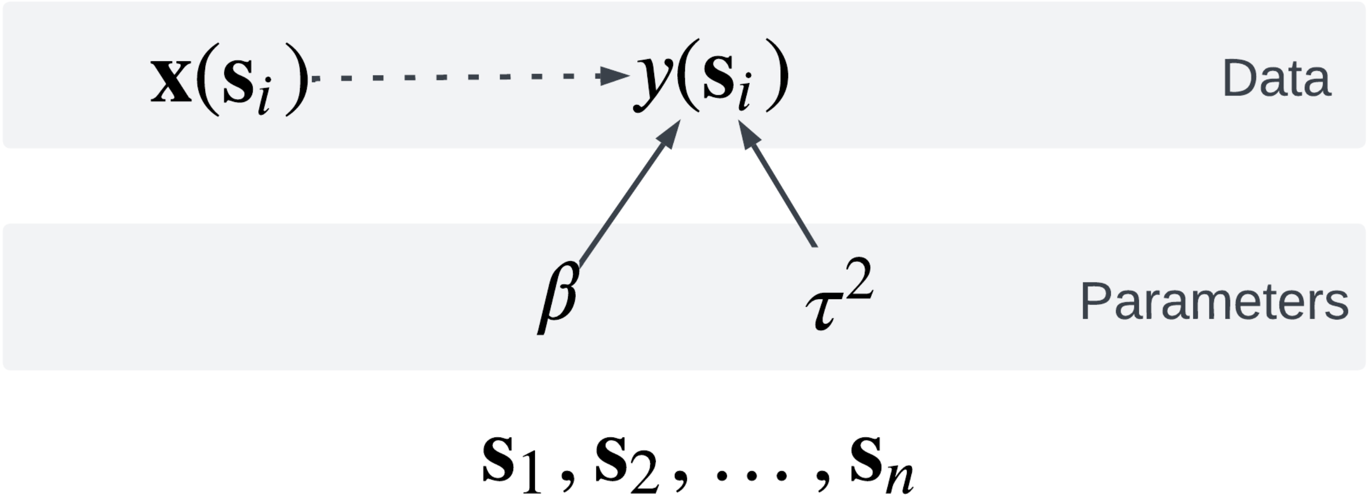
Generalized linear model (GLM)

Generalized linear model (GLM)



$$\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n$$

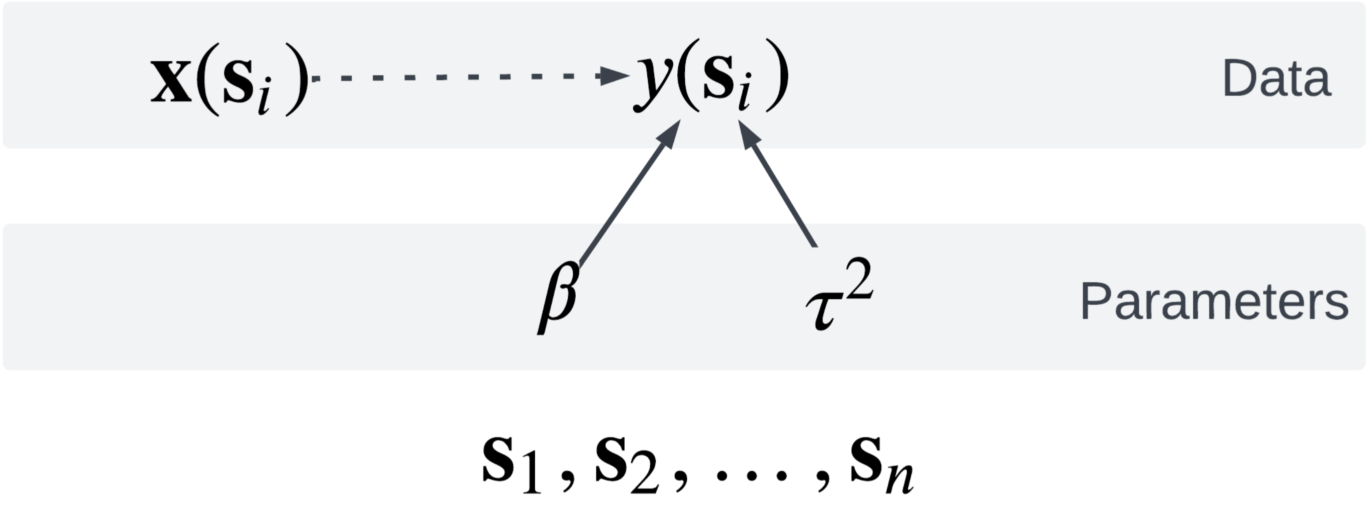
Generalized linear model (GLM)



$$y(\mathbf{s}_i) \sim f(\mu(\mathbf{s}_i), \tau^2)$$

$$g(\mu(\mathbf{s}_i)) = \mathbf{x}(\mathbf{s}_i)^\top \beta$$

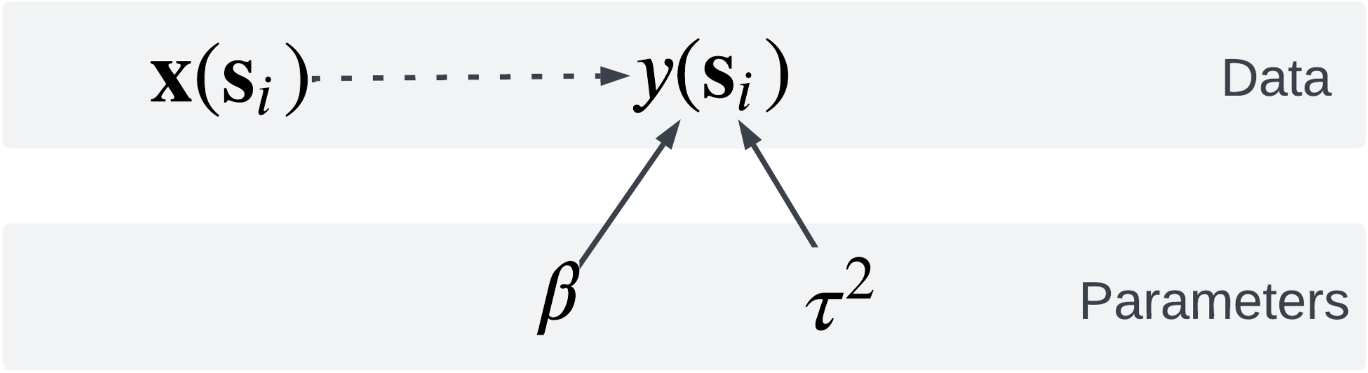
Generalized linear model (GLM)



$$y(\mathbf{s}_i) \sim f(\mu(\mathbf{s}_i), \tau^2)$$
$$g(\mu(\mathbf{s}_i)) = \mathbf{x}(\mathbf{s}_i)^\top \beta$$

- Each $y(\mathbf{s}_i)$ is conditionally independent
- *Spatial data routinely violate the independence assumption*

Generalized linear model (GLM)



$$y(\mathbf{s}_i) \sim f(\mu(\mathbf{s}_i), \tau^2)$$

$$g(\mu(\mathbf{s}_i)) = \mathbf{x}(\mathbf{s}_i)^\top \beta$$

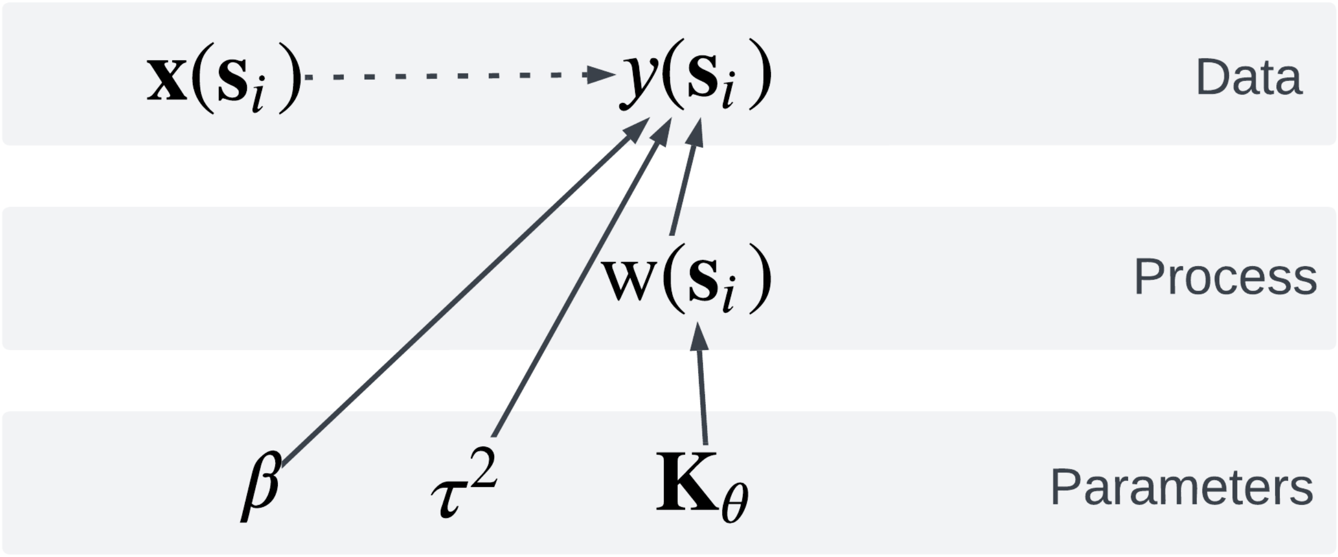
$\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n$

Spatial coordinates of each of the n spatial locations

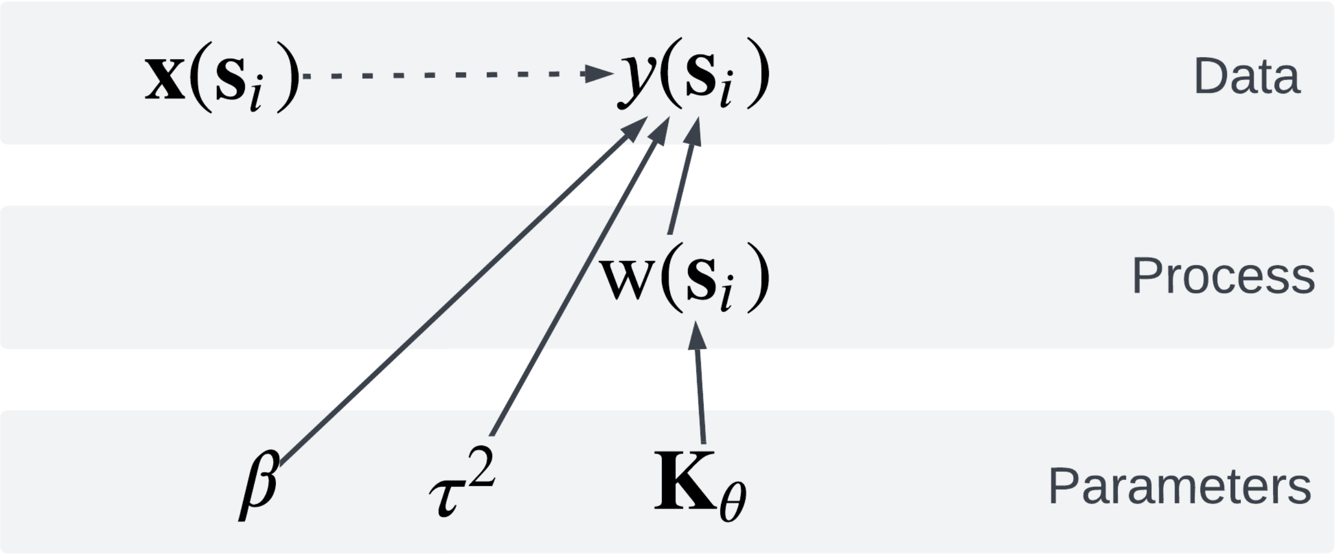
- Each $y(\mathbf{s}_i)$ is conditionally independent
- *Spatial data routinely violate the independence assumption*

Hierarchical spatial model

Hierarchical spatial model



Hierarchical spatial model



Example with n = 5 sites

$$\begin{bmatrix} k_{1,1} & k_{1,2} & k_{1,3} & k_{1,4} & k_{1,5} \\ k_{2,1} & k_{2,2} & k_{2,3} & k_{2,4} & k_{2,5} \\ k_{3,1} & k_{3,2} & k_{3,3} & k_{3,4} & k_{3,5} \\ k_{4,1} & k_{4,2} & k_{4,3} & k_{4,4} & k_{4,5} \\ k_{5,1} & k_{5,2} & k_{5,3} & k_{5,4} & k_{5,5} \end{bmatrix}$$

Covariance between site 5 and site 1

Hierarchical spatial model

$$y(\mathbf{s}_i) \sim f(\mu(\mathbf{s}_i), \tau^2)$$
$$g(\mu(\mathbf{s}_i)) = \mathbf{x}(\mathbf{s}_i)^\top \boldsymbol{\beta} + \mathrm{w}(\mathbf{s}_i)$$
$$\mathbf{w} \sim \mathrm{GP}(\mathbf{0}, \mathbf{K}_\theta)$$

- Add in spatial random effects **w** to account for spatial autocorrelation
- $\mathbf{K}_\theta$ is a n x n spatial covariance matrix.

Residual spatial autocorrelation

Residual spatial autocorrelation

- Spatial correlation in the linear predictor that is not explained by covariates in the model

Residual spatial autocorrelation

- Spatial correlation in the linear predictor that is not explained by covariates in the model
- Often arises from missing/unavailable covariates

Residual spatial autocorrelation

- Spatial correlation in the linear predictor that is not explained by covariates in the model
- Often arises from missing/unavailable covariates
- Can lead to bias if unaddressed

Residual spatial autocorrelation

- Spatial correlation in the linear predictor that is not explained by covariates in the model
- Often arises from missing/unavailable covariates
- Can lead to bias if unaddressed
- Account for using spatial random effects
  - Each site has a local adjustment
  - The local adjustments are given a spatial structure
  - Estimated parameters: spatial variance and spatial decay

Gaussian process

- "Gold standard" for modeling spatial data

Gaussian process

- "Gold standard" for modeling spatial data
- Covariance between two sites is determined by:
  - Distance between the sites
  - A covariance function

Gaussian process

- "Gold standard" for modeling spatial data
- Covariance between two sites is determined by:
  - Distance between the sites
  - A covariance function
- The software we will use supports four covariance functions: **exponential**, Gaussian, spherical, Matérn

Gaussian process

- "Gold standard" for modeling spatial data
- Covariance between two sites is determined by:
  - Distance between the sites
  - A covariance function
- The software we will use supports four covariance functions: **exponential**, Gaussian, spherical, Matérn
- Covariance between site A and site B using exponential covariance function:

$$\boldsymbol{\Sigma}(d_{A,B}, \sigma^2, \phi) = \sigma^2 \exp(-\phi d_{A,B})$$

Intuition on spatial covariance

$$\boldsymbol{\Sigma}(d_{A,B}, \sigma^2, \phi) = \sigma^2 \exp(-\phi d_{A,B})$$

$d_{A,B}$ Euclidean (linear) distance between site A and B

Intuition on spatial covariance

$$\mathbf{\Sigma}(d_{A,B}, \sigma^2, \phi) = \sigma^2 \mathrm{exp}(-\phi d_{A,B})$$

$d_{A,B}$ Euclidean (linear) distance between site A and B

σ^2 Spatial variance. Controls magnitude of random effects

Intuition on spatial covariance

$$\mathbf{\Sigma}(d_{A,B}, \sigma^2, \phi) = \sigma^2 \exp(-\phi d_{A,B})$$

$d_{A,B}$ Euclidean (linear) distance between site A and B

σ^2 Spatial variance. Controls magnitude of random effects

ϕ Spatial decay. Controls how quickly the correlation between sites decays across space

Intuition on spatial covariance

$$\boldsymbol{\Sigma}(d_{A,B}, \sigma^2, \phi) = \sigma^2 \exp(-\phi d_{A,B})$$

$d_{A,B}$ Euclidean (linear) distance between site A and B

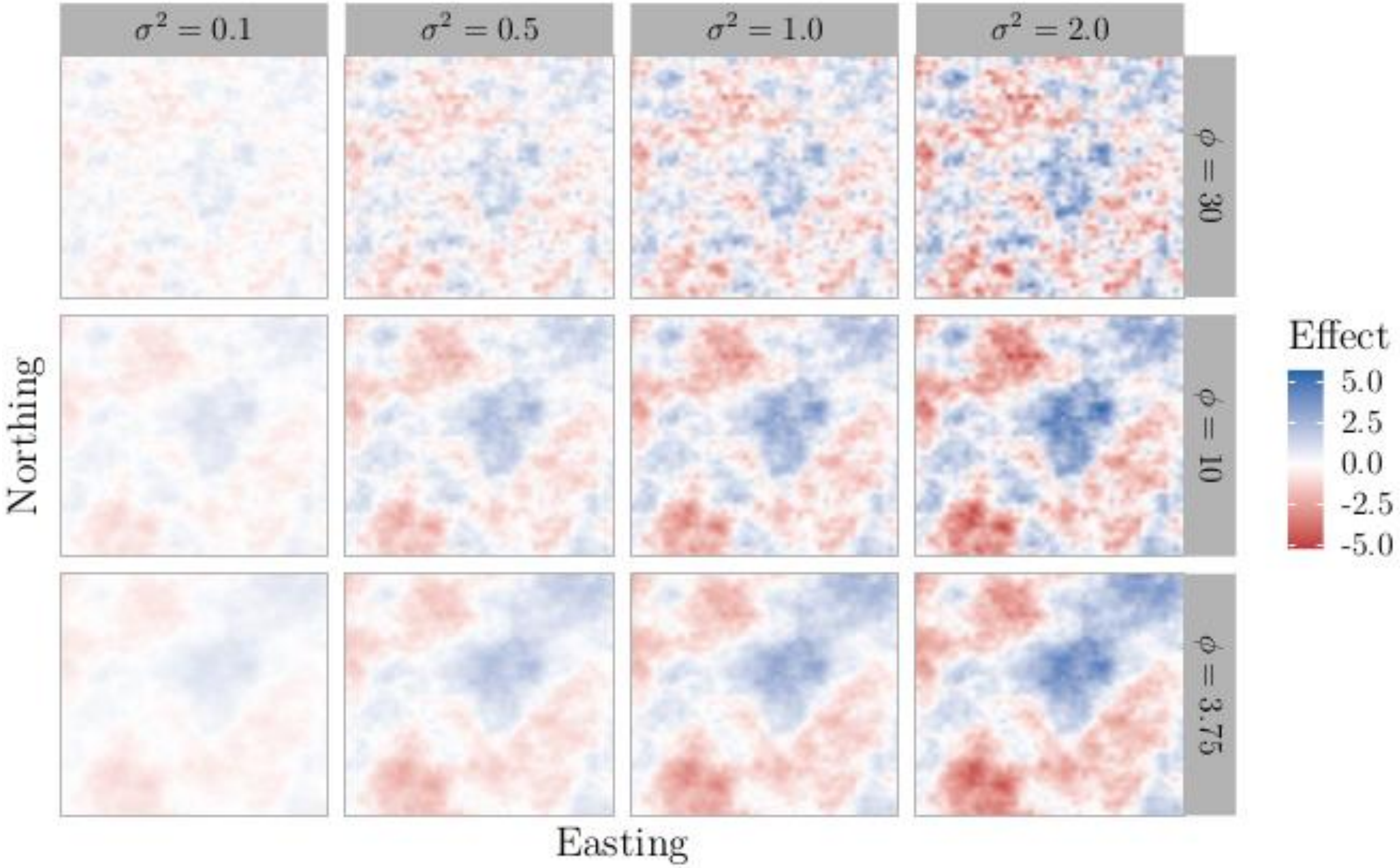
σ^2 Spatial variance. Controls magnitude of random effects

ϕ Spatial decay. Controls how quickly the correlation between sites decays across space

$\frac{3}{\phi}$ "Effective spatial range" when using an exponential covariance function. This is the distance at which the spatial correlation between two sites is essentially negligible (0.05)

Intuition on spatial covariance

$$\mathbf{\Sigma}(d_{A,B}, \sigma^2, \phi) = \sigma^2 \exp(-\phi d_{A,B})$$



R packages we will use

- `spOccupancy`
  - Doser et al. (2022) *MEE*
  - Fits a variety of spatial Bayesian models for estimating species distributions
- `spAbundance`
  - Doser et al. (2024) *MEE*
  - Fits a variety of spatial Bayesian models
  - Focus on estimating abundance, but spatial GLMMs can work with anything

Gaussian process

- Flexible approach to account for spatial autocorrelation

Gaussian process

- Flexible approach to account for spatial autocorrelation
- But... becomes extremely slow as the number of sites increases
- Not practical for data sets with hundreds of data points, let alone thousands.
- Computational bottleneck: dealing with a large, dense J x J matrix
- The "big n" problem
- Need a more efficient approach...

Nearest Neighbor Gaussian Processes (NNGPs)

- Heaton et al. (2019): overview of approaches to model big spatial data

Nearest Neighbor Gaussian Processes (NNGPs)

- Heaton et al. (2019): overview of approaches to model big spatial data
- Our approach: NNGPs (Datta et al. 2016, Finley et al. 2019). Same parameters/interpretation as full GP

Nearest Neighbor Gaussian Processes (NNGPs)

- Heaton et al. (2019): overview of approaches to model big spatial data
- Our approach: NNGPs (Datta et al. 2016, Finley et al. 2019). Same parameters/interpretation as full GP
- Conceptually:
    1. Order the spatial locations (e.g., along the x-axis)

Nearest Neighbor Gaussian Processes (NNGPs)

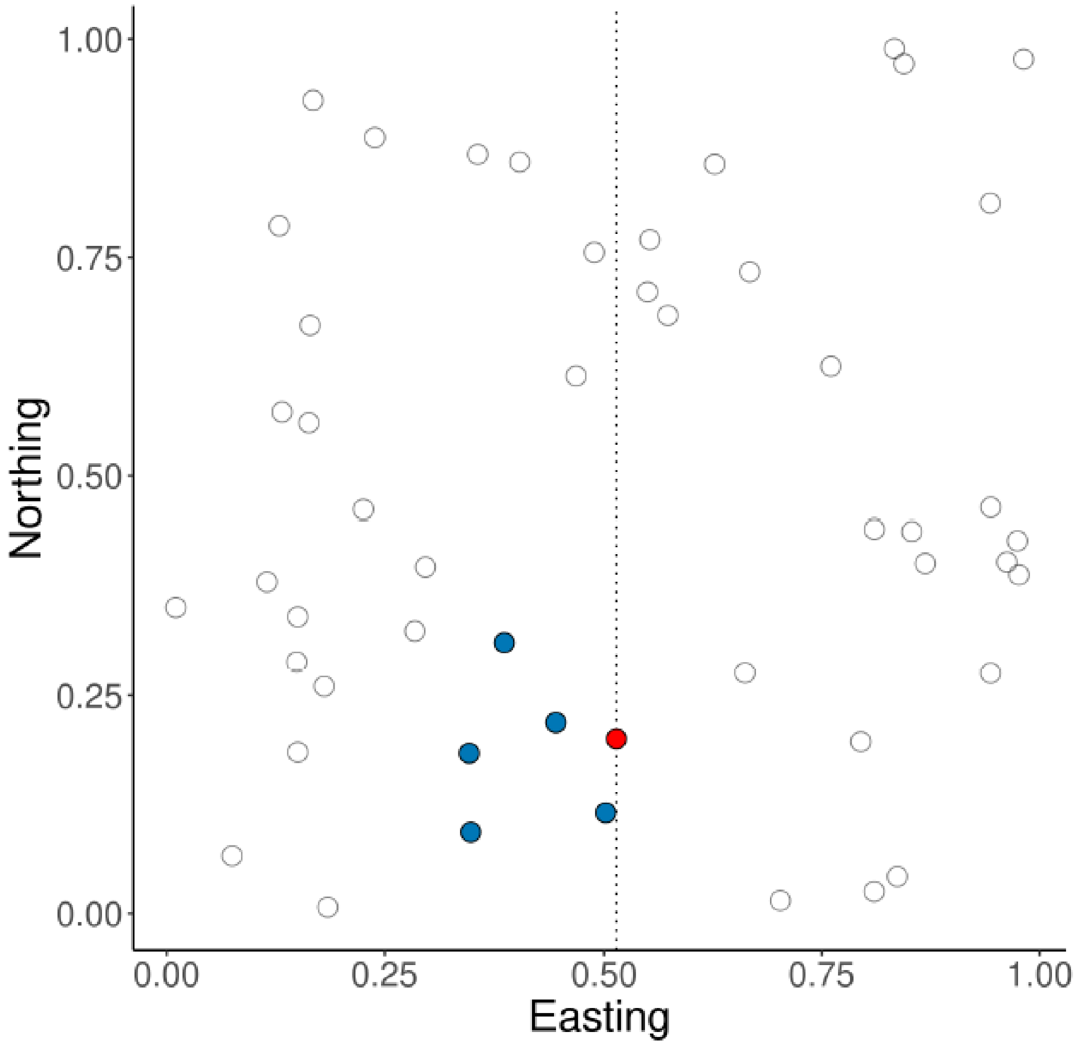
- Heaton et al. (2019): overview of approaches to model big spatial data
- Our approach: NNGPs (Datta et al. 2016, Finley et al. 2019). Same parameters/interpretation as full GP
- Conceptually:
  1. Order the spatial locations (e.g., along the x-axis)
  2. Determine the *m* nearest neighbors (subject to ordering) based on Euclidean (linear) distance

Nearest Neighbor Gaussian Processes (NNGPs)

- Heaton et al. (2019): overview of approaches to model big spatial data
- Our approach: NNGPs (Datta et al. 2016, Finley et al. 2019). Same parameters/interpretation as full GP
- Conceptually:
  1. Order the spatial locations (e.g., along the x-axis)
  2. Determine the *m* nearest neighbors (subject to ordering) based on Euclidean (linear) distance
  3. The spatial random effect at each site only depends on values of its *m* nearest neighbors and is conditionally independent of all other values

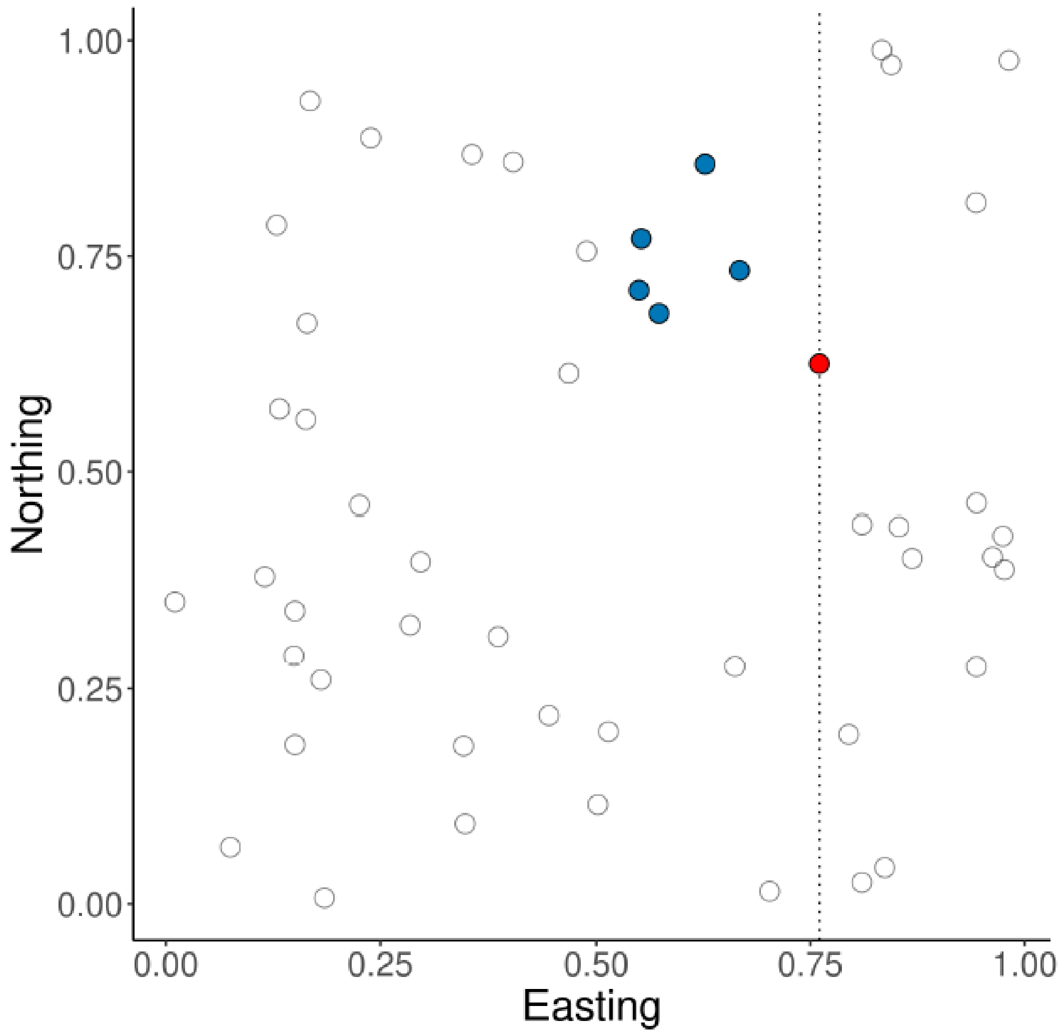
Choosing the neighbors

- spOccupancy and spAbundance order sites along the horizontal axis (i.e., Easting)
- Example: NNGP with 5 neighbors
- Red point denotes the current site
- Blue points denote sites in the "neighbor set"



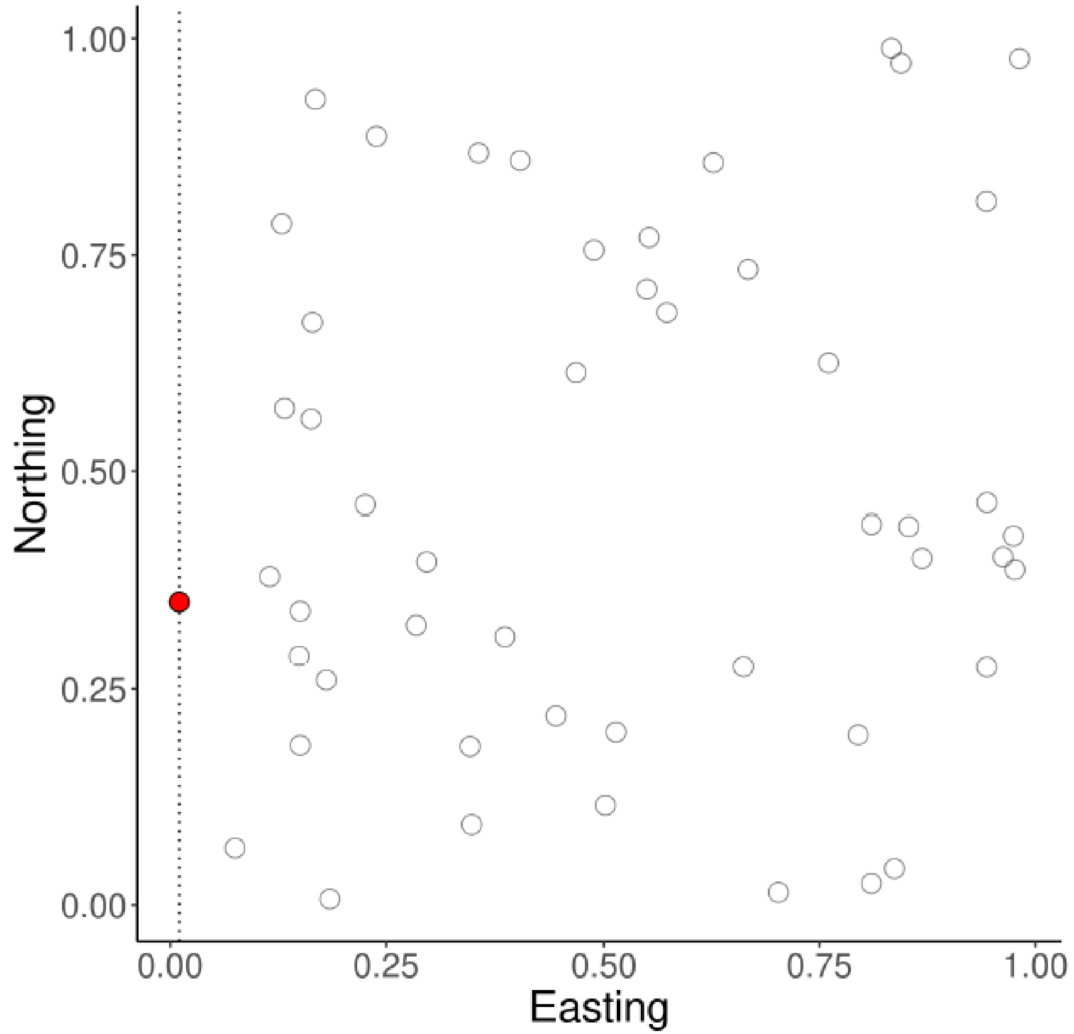
Choosing the neighbors

- spOccupancy and spAbundance order sites along the horizontal axis (i.e., Easting)
- Example: NNGP with 5 neighbors
- Red point denotes the current site
- Blue points denote sites in the "neighbor set"

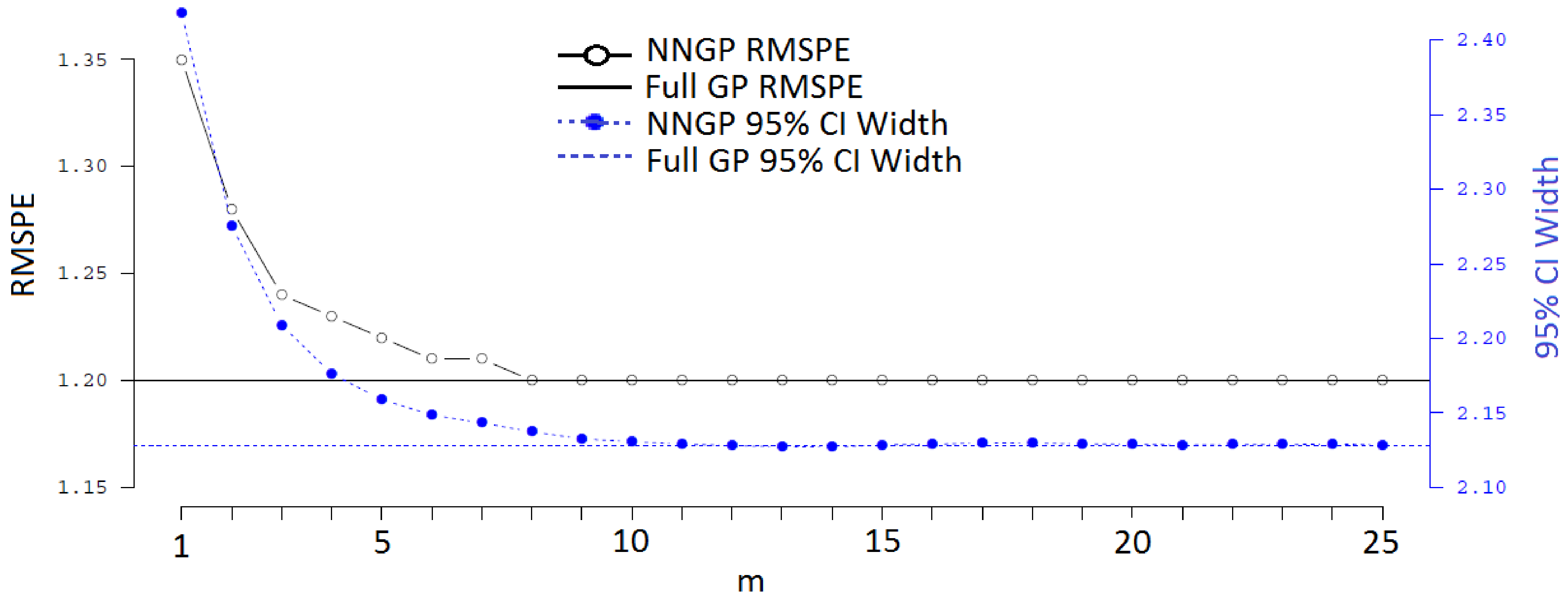


Choosing the neighbors

- spOccupancy and spAbundance order sites along the horizontal axis (i.e., Easting)
- Example: NNGP with 5 neighbors
- Red point denotes the current site
- Blue points denote sites in the "neighbor set"

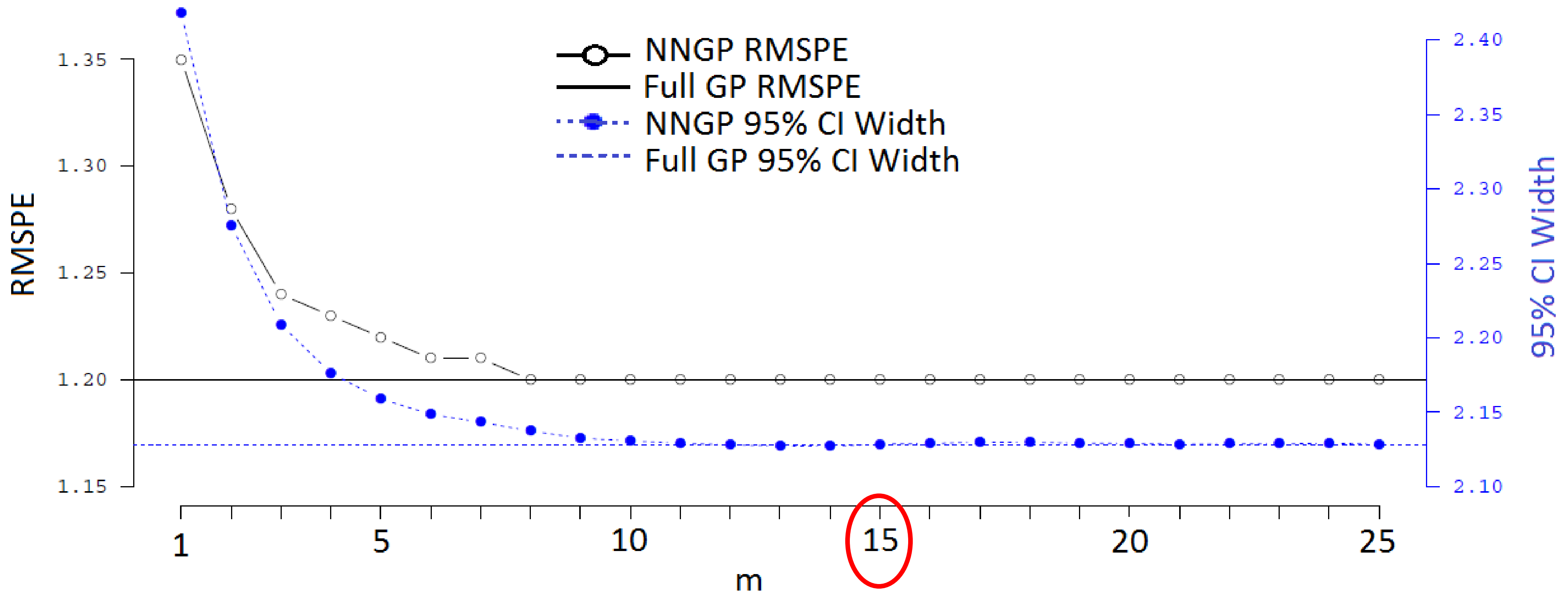


How many neighbors?



Datta et al. (2016) JASA

How many neighbors?



- m= 15 neighbors is often adequate (software default)
- Can compare smaller m using WAIC

Datta et al. (2016) JASA

The "big n" problem in more detail

- Estimating the parameters in our model requires calculating:

$$p(\mathbf{w}) \propto -\frac{1}{2}\log(\boxed{\det(\mathbf{K}_\theta)}) - \frac{1}{2}\mathbf{w}^\top \boxed{\mathbf{K}_\theta^{-1}} \mathbf{w}$$

The "big n" problem in more detail

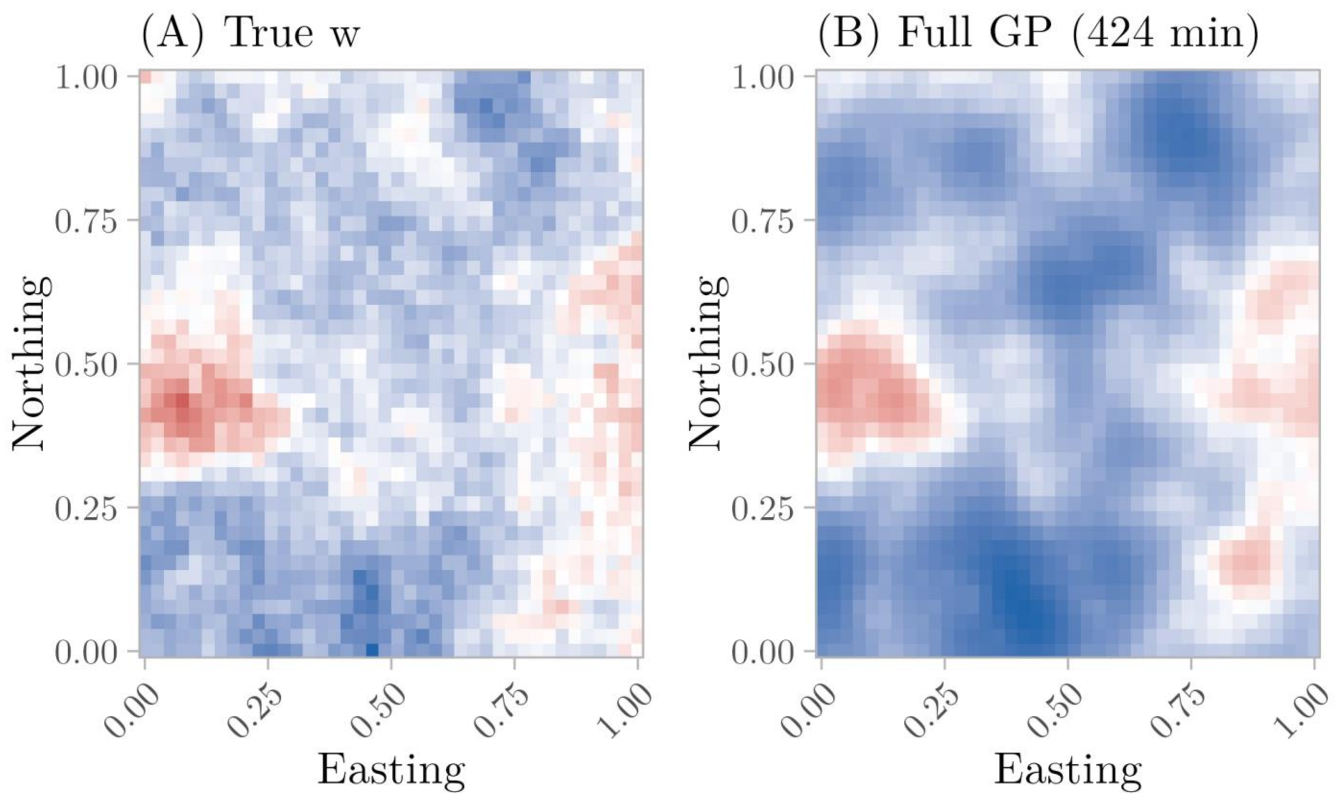
- Estimating the parameters in our model requires calculating:

$$p(\mathbf{w}) \propto -\frac{1}{2}\log\left(\boxed{\det(\mathbf{K}_\theta)}\right) - \frac{1}{2}\mathbf{w}^\top \boxed{\mathbf{K}_\theta^{-1}}\mathbf{w}$$

- Storage requirements: O(n^2) flops
- $\mathbf{K}_\theta$ is a dense matrix
- Computational complexity: $O(n^3)$ flops
- Computationally infeasible when *n* is even moderately large (e.g., 500)

Simulation comparison

- *n = 1600,* Bayesian estimation using MCMC



NNGPs in more detail

NNGPs in more detail

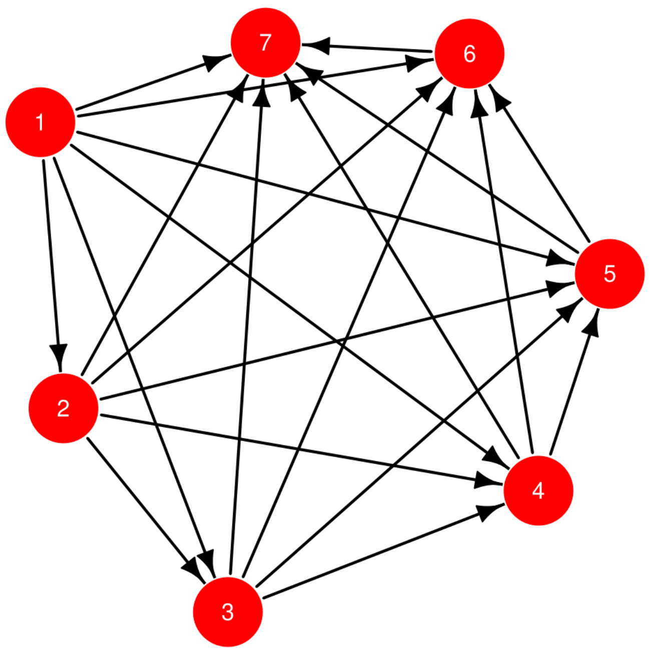
- Use a *sparse* matrix to approximate the dense Gaussian Process covariance matrix $\mathbf{K}_\theta$ (Datta et al. 2016)
- Based on rewriting the GP for **w** as a product of conditional densities.

NNGPs in more detail

- Use a *sparse* matrix to approximate the dense Gaussian Process covariance matrix $\mathbf{K}_\theta$ (Datta et al. 2016)
- Based on rewriting the GP for **w** as a product of conditional densities.
- $\mathbf{w} \sim \text{Normal}(\mathbf{0}, \mathbf{K}_\theta)$ is equivalent to

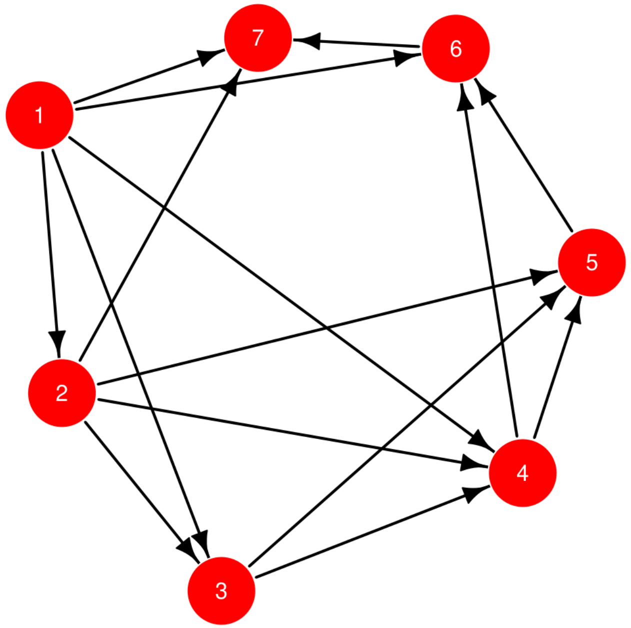
$$p(w_1)p(w_2 \mid w_1)p(w_3 \mid w_1, w_2) \cdots p(w_n \mid w_1, w_2, \ldots, w_{n-1})$$

Introducing sparsity via graphical models



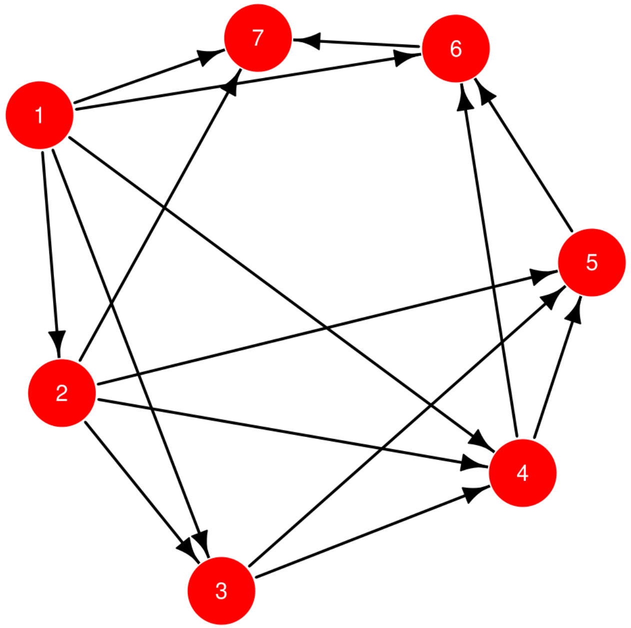
$$p(w_1)p(w_2 \mid w_1)p(w_3 \mid w_1, w_2)p(w_4 \mid w_1, w_2, w_3)$$
$$\times\ p(w_5 \mid w_1, w_2, w_3, w_4)p(w_6 \mid w_1, w_2, \ldots, w_5)$$
$$\times\ p(w_7 \mid w_1, w_2, \ldots, w_6)\ .$$

Introducing sparsity via graphical models



$$p(w_1)p(w_2 \mid w_1)p(w_3 \mid w_1, w_2)p(w_4 \mid w_1, w_2, w_3)$$
$$\times\ p(w_5 \mid \cancel{w_1}, w_2, w_3, w_4)p(w_6 \mid w_1, \cancel{w_2}, \cancel{w_3}, w_4, w_5)$$
$$\times\ p(w_7 \mid w_1, w_2, \cancel{w_3}, \cancel{w_4}, \cancel{w_5}, w_6)$$

Introducing sparsity via graphical models

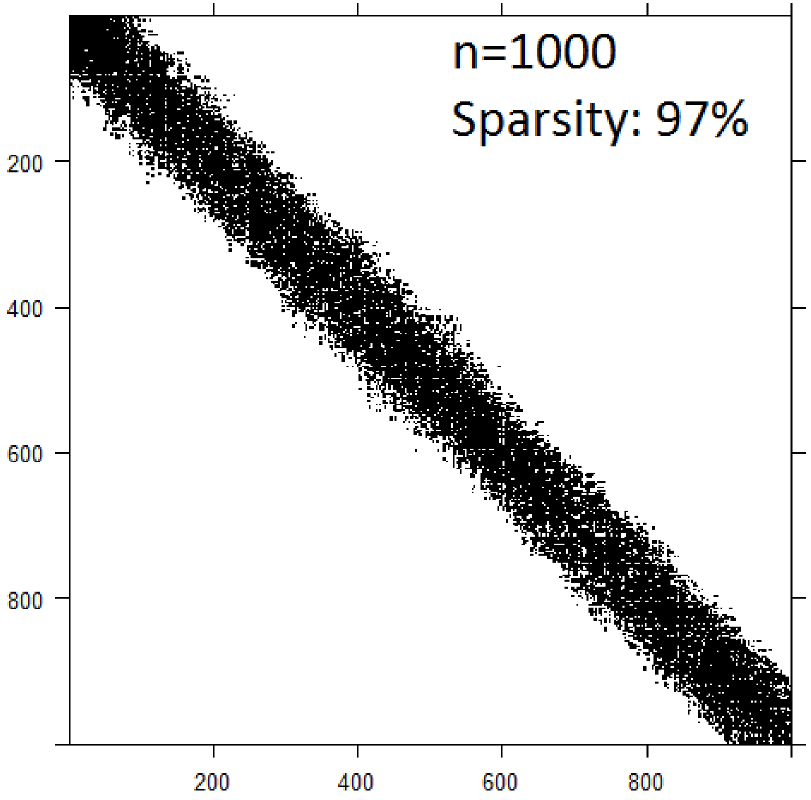


Size of neighbor set is ≤ h

$$p(w_1)p(w_2 \mid w_1)p(w_3 \mid w_1, w_2)p(w_4 \mid w_1, w_2, w_3)$$
$$\times\, p(w_5 \mid \cancel{w_1}, w_2, w_3, w_4)p(w_6 \mid w_1, \cancel{w_2}, \cancel{w_3}, w_4, w_5)$$
$$\times\, p(w_7 \mid w_1, w_2, \cancel{w_3}, \cancel{w_4}, \cancel{w_5}, w_6)$$

Sparsity via NNGP

$$N(\mathbf{w} \mid \mathbf{0}, \mathbf{K}_\theta) \approx N(\mathbf{w} \mid \mathbf{0}, \tilde{\boldsymbol{K}}_\theta)$$



$$\tilde{\boldsymbol{K}}_\theta^{-1}$$

Spatial GLMM with NNGP

$$
\begin{aligned}
y(\mathbf{s}_i) &\sim f(\mu(\mathbf{s}_i), \sigma^2) \\
g(\mu(\mathbf{s}_i)) &= \mathbf{x}(\mathbf{s}_i)^\top \boldsymbol{\beta} + \mathrm{w}(\mathbf{s}_i) \\
\mathbf{w} &\sim \mathrm{NNGP}(\mathbf{0}, \tilde{\boldsymbol{K}}_\theta)
\end{aligned}
$$

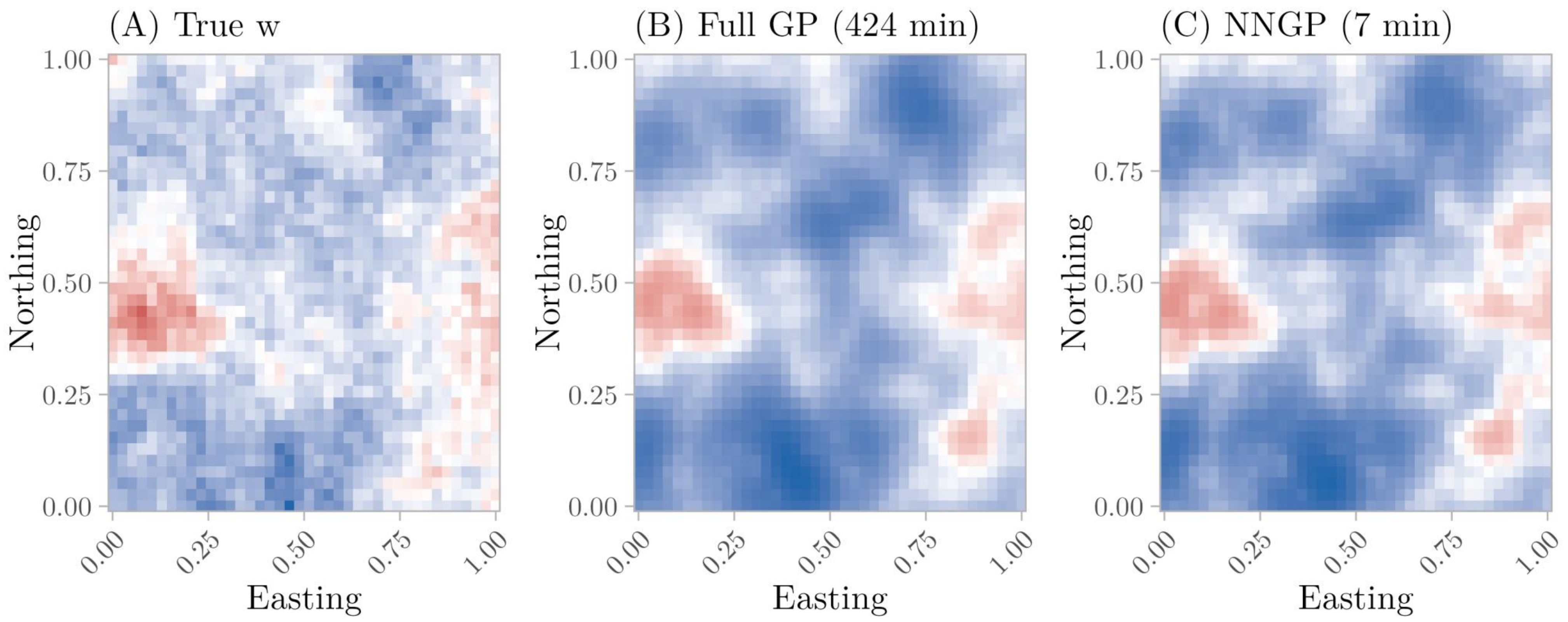
- Model is estimated in a hierarchical Bayesian framework using Markov chain Monte Carlo (MCMC)

Storage and computation

- Storage
  - Never need to store n x n distance matrix
  - Stores smaller h x h matrices
  - Total storage requirements: $O(nh^2)$
- Computation:
  - Only involves inverting small h x h matrices
  - $O(nh^3)$ flops
- Since h << n, NNGP offers great scalability for big spatial data

Simulation comparison

- *n = 1600*, Bayesian estimation using MCMC



Pros/cons of spatial models

Pros

- More accurate predictions
- More accurate uncertainty estimates
- Provide insights on underlying drivers
- Generate new hypotheses

Pros/cons of spatial models

Pros

- More accurate predictions
- More accurate uncertainty estimates
- Provide insights on underlying drivers
- Generate new hypotheses

Cons

- Slower (but NNGPs help a lot!)
- Spatial confounding (Hanks et al. 2015, Mäkinen et al. 2022)
- More data hungry

Example: Hierarchical spatial model

Hooded warbler relative abundance

7-spatial-GLMMs.R