

Introduction to Geostatistics

Andrew Finley¹ & Jeffrey Doser²

May 15, 2023

¹Department of Forestry, Michigan State University.

²Department of Integrative Biology, Michigan State University.

- Course materials available at
<https://doserjef.github.io/CASANR23-Spatial-Modeling/>

What is spatial data?

- Any data with some geographical information (i.e., spatially indexed)
- Common sources of spatial data: agricultural, climatology, forestry, ecology, environmental health, disease epidemiology, product marketing, etc.
 - have many important predictors and response variables
 - are often presented as maps

What is spatial data?

- Any data with some geographical information (i.e., spatially indexed)
- Common sources of spatial data: agricultural, climatology, forestry, ecology, environmental health, disease epidemiology, product marketing, etc.
 - have many important predictors and response variables
 - are often presented as maps
- Other examples where spatial need not refer to space on earth:
 - Genetics (position along a chromosome)
 - Neuroimaging (data for each voxel in the brain)

Point-referenced spatial data

- Each observation is associated with a location (point)
- Data represents a sample from a continuous spatial domain
- Also referred to as **geocoded** or **geostatistical** data

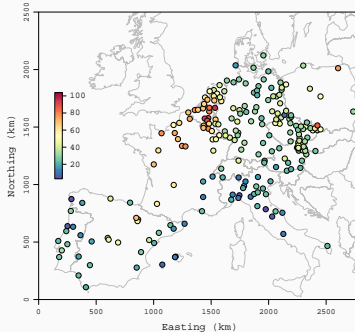


Figure: Pollutant levels in Europe in March, 2009

Point level modeling

- **Point-level modeling** refers to modeling of point-referenced data collected at locations referenced by **coordinates** (e.g., lat-long, Easting-Northing).
- Data from a spatial process $\{Y(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$, \mathcal{D} is a subset in Euclidean space.
- **Example:** $Y(\mathbf{s})$ is a **pollutant level** at site \mathbf{s}
- **Conceptually:** Pollutant level exists at all possible sites
- **Practically:** Data will be a partial realization of a spatial process – observed at $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$
- **Statistical objectives:** **Inference** about the process $Y(\mathbf{s})$; **predict** at new locations.
- **Remarkable:** Can learn about entire $Y(\mathbf{s})$ surface. The **key:** Structured dependence

Exploratory data analysis (EDA): Plotting the data

- A typical setup: Data observed at n locations $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$
- At each \mathbf{s}_i we observe the response $y(\mathbf{s}_i)$ and a $p \times 1$ vector of covariates $\mathbf{x}(\mathbf{s}_i)$
- **Surface plots** of the data often helps to understand spatial patterns

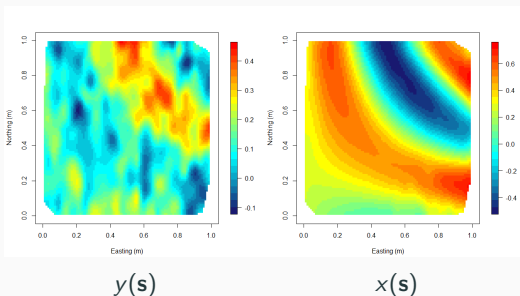


Figure: Response and covariate surface plots for Dataset 1

What's so special about spatial?

- Linear regression model: $y(\mathbf{s}_i) = \mathbf{x}(\mathbf{s}_i)^\top \boldsymbol{\beta} + \epsilon(\mathbf{s}_i)$
- $\epsilon(\mathbf{s}_i)$ are iid $N(0, \tau^2)$ errors
- $\mathbf{y} = (y(\mathbf{s}_1), \dots, y(\mathbf{s}_n))^\top$; $\mathbf{X} = (\mathbf{x}(\mathbf{s}_1)^\top, \dots, \mathbf{x}(\mathbf{s}_n)^\top)^\top$
- Inference: $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \sim N(\boldsymbol{\beta}, \tau^2 (\mathbf{X}^\top \mathbf{X})^{-1})$
- Prediction at new location \mathbf{s}_0 : $\widehat{y(\mathbf{s}_0)} = \mathbf{x}(\mathbf{s}_0)^\top \hat{\boldsymbol{\beta}}$
- Although the data is spatial, this is an ordinary linear regression model

Residual plots

- Surface plots of the residuals ($y(\mathbf{s}) - \widehat{y}(\mathbf{s})$) help to identify any spatial patterns left unexplained by the covariates

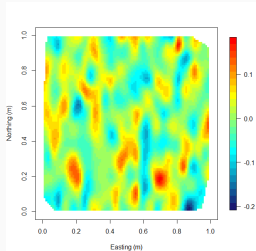


Figure: Residual plot for Dataset 1 after linear regression on $x(\mathbf{s})$

Residual plots

- Surface plots of the residuals ($y(\mathbf{s}) - \widehat{y}(\mathbf{s})$) help to identify any spatial patterns left unexplained by the covariates

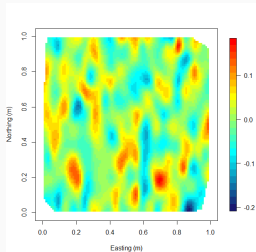
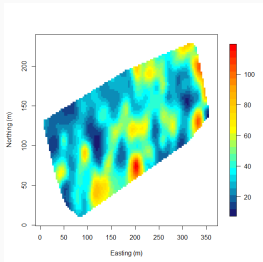


Figure: Residual plot for Dataset 1 after linear regression on $x(\mathbf{s})$

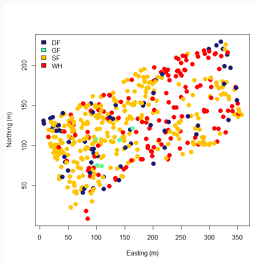
- No evident spatial pattern in plot of the residuals
- The covariate $x(\mathbf{s})$ seem to explain all spatial variation in $y(\mathbf{s})$
- Does a non-spatial regression model always suffice?

Western Experimental Forestry (WEF) data

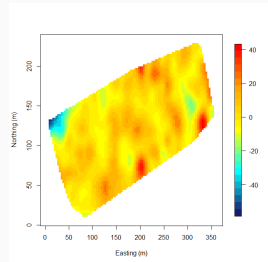
- Data consist of a census of all trees in a 10 ha. stand in Oregon
- Response of interest: Diameter at breast height (DBH)
- Covariate: Tree species (Categorical variable)



DBH



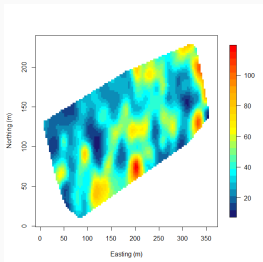
Species



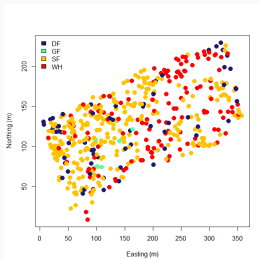
Residuals

Western Experimental Forestry (WEF) data

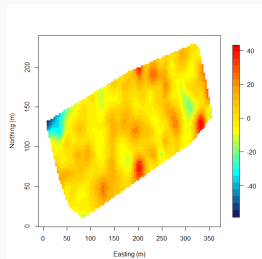
- Data consist of a census of all trees in a 10 ha. stand in Oregon
- Response of interest: Diameter at breast height (DBH)
- Covariate: Tree species (Categorical variable)



DBH



Species



Residuals

- Local spatial patterns in the residual plot
- Simple regression on species seems to be not sufficient

- Besides eyeballing residual surfaces, how to do more formal EDA to identify spatial pattern?

- Besides eyeballing residual surfaces, how to do more formal EDA to identify spatial pattern?

First law of geography

*“Everything is related to everything else, but **near things are more related** than distant things.”* – Waldo Tobler

- Besides eyeballing residual surfaces, how to do more formal EDA to identify spatial pattern?

First law of geography

*“Everything is related to everything else, but **near things are more related** than distant things.”* – Waldo Tobler

- In general $(Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s}))^2$ roughly increasing with $\|\mathbf{h}\|$ will imply a spatial correlation
- Can this be formalized to identify spatial pattern?

Empirical semivariogram

- **Binning:** Make intervals $I_1 = (0, m_1)$, $I_2 = (m_1, m_2)$, and so forth, up to $I_K = (m_{K-1}, m_K)$. Representing each interval by its midpoint t_k , we define:

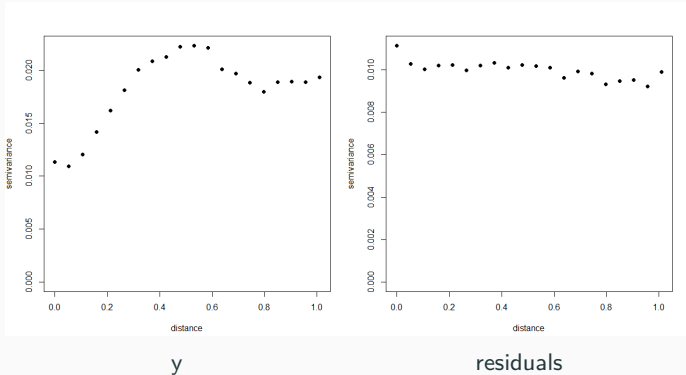
$$N(t_k) = \{(\mathbf{s}_i, \mathbf{s}_j) : \|\mathbf{s}_i - \mathbf{s}_j\| \in I_k\}, k = 1, \dots, K.$$

- **Empirical semivariogram:**

$$\gamma(t_k) = \frac{1}{2|N(t_k)|} \sum_{\mathbf{s}_i, \mathbf{s}_j \in N(t_k)} (Y(\mathbf{s}_i) - Y(\mathbf{s}_j))^2$$

- For spatial data, the $\gamma(t_k)$ is expected to roughly increase with t_k
- A flat semivariogram would suggest little spatial variation

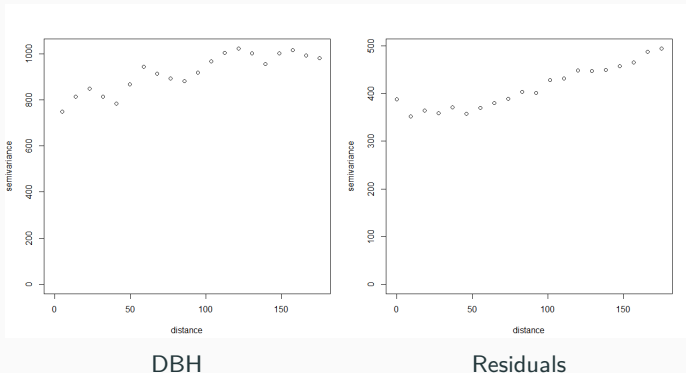
Empirical variogram: Data 1



- Residuals display little spatial variation

Empirical variograms: WEF data

- Regression model: $\text{DBH} \sim \text{Species}$



- Variogram of the residuals confirm **unexplained spatial variation**

Modeling with the locations

- When purely covariate based models does not suffice, one needs to leverage the information from locations
- General model using the locations:
$$y(\mathbf{s}) = \mathbf{x}(\mathbf{s})^\top \boldsymbol{\beta} + w(\mathbf{s}) + \epsilon(\mathbf{s}) \text{ for all } \mathbf{s} \in \mathcal{D}$$
- How to choose the function $w(\cdot)$?
- Since we want to predict at any location over the entire domain \mathcal{D} , this choice will amount to choosing a **surface** $w(\mathbf{s})$
- How should such a surface be chosen?

Gaussian Processes (GPs)

- One popular approach to **model** $w(\mathbf{s})$ is via Gaussian Processes (GP)
- The collection of random variables $\{w(\mathbf{s}) \mid \mathbf{s} \in \mathcal{D}\}$ is a GP if
 - it is a **valid** stochastic process
 - all finite dimensional densities $\{w(\mathbf{s}_1), \dots, w(\mathbf{s}_n)\}$ follow multivariate Gaussian distribution
- A GP is completely characterized by a mean function $m(\mathbf{s})$ and a covariance function $C(\cdot, \cdot)$
- **Advantage:** **Likelihood** based inference.
 $\mathbf{w} = (w(\mathbf{s}_1), \dots, w(\mathbf{s}_n))^T \sim N(\mathbf{m}, \mathbf{C})$ where
 $\mathbf{m} = (m(\mathbf{s}_1), \dots, m(\mathbf{s}_n))^T$ and $\mathbf{C} = C(\mathbf{s}_i, \mathbf{s}_j)$

Valid covariance functions and isotropy

- $C(\cdot, \cdot)$ needs to be **valid**. For all n and all $\{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n\}$, the resulting covariance matrix $C(\mathbf{s}_i, \mathbf{s}_j)$ for $(w(\mathbf{s}_1), w(\mathbf{s}_2), \dots, w(\mathbf{s}_n))$ must be positive definite
- So, $C(\cdot, \cdot)$ needs to be a **positive definite** function
- Simplifying assumptions:
 - **Stationarity**: $C(\mathbf{s}_1, \mathbf{s}_2)$ only depends on $\mathbf{h} = \mathbf{s}_1 - \mathbf{s}_2$ (and is denoted by $C(\mathbf{h})$)
 - **Isotropic**: $C(\mathbf{h}) = C(\|\mathbf{h}\|)$
 - **Anisotropic**: Stationary but not isotropic
- Isotropic models are popular because of their **simplicity**, **interpretability**, and because a number of relatively **simple parametric forms** are available as candidates for C .

Some common isotropic covariance functions

Model	Covariance function, $C(t) = C(h)$
Spherical	$C(t) = \begin{cases} 0 & \text{if } t \geq 1/\phi \\ \sigma^2 \left[1 - \frac{3}{2}\phi t + \frac{1}{2}(\phi t)^3 \right] & \text{if } 0 < t \leq 1/\phi \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$
Exponential	$C(t) = \begin{cases} \sigma^2 \exp(-\phi t) & \text{if } t > 0 \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$
Powered exponential	$C(t) = \begin{cases} \sigma^2 \exp(- \phi t ^p) & \text{if } t > 0 \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$
Matérn at $\nu = 3/2$	$C(t) = \begin{cases} \sigma^2 (1 + \phi t) \exp(-\phi t) & \text{if } t > 0 \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$

Notes on exponential model

$$C(t) = \begin{cases} \tau^2 + \sigma^2 & \text{if } t = 0 \\ \sigma^2 \exp(-\phi t) & \text{if } t > 0 \end{cases} .$$

- We define the **effective range**, t_0 , as the distance at which this correlation has dropped to only 0.05. Setting $\exp(-\phi t_0)$ equal to this value we obtain $t_0 \approx 3/\phi$, since $\log(0.05) \approx -3$.
- The **nugget** τ^2 is often viewed as a “**nonspatial effect variance**,”
- The **partial sill** (σ^2) is viewed as a “**spatial effect variance**.”
- $\sigma^2 + \tau^2$ gives the maximum total variance often referred to as the **sill**
- Note **discontinuity** at 0 due to the nugget. **Intentional!** To account for measurement error or micro-scale variability.

Covariance functions and semivariograms

- **Recall:** Empirical semivariogram:

$$\gamma(t_k) = \frac{1}{2|N(t_k)|} \sum_{\mathbf{s}_i, \mathbf{s}_j \in N(t_k)} (Y(\mathbf{s}_i) - Y(\mathbf{s}_j))^2$$

- For any stationary GP,

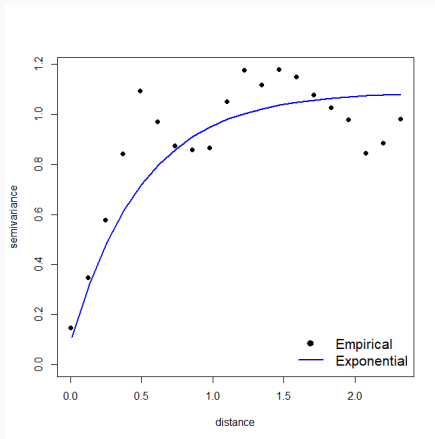
$$E(Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s}))^2/2 = C(\mathbf{0}) - C(\mathbf{h}) = \gamma(\mathbf{h})$$

- $\gamma(\mathbf{h})$ is the **semivariogram** corresponding to the covariance function $C(\mathbf{h})$

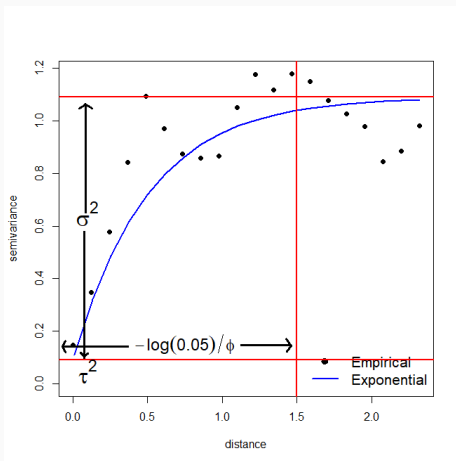
- **Example:** For exponential GP,

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi t)) & \text{if } t > 0 \\ 0 & \text{if } t = 0 \end{cases}, \text{ where } t = \|\mathbf{h}\|$$

Covariance functions and semivariograms



Covariance functions and semivariograms



The Matèrn covariance function

- The Matèrn is a very versatile family:

$$C(t) = \begin{cases} \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (2\sqrt{\nu}t\phi)^\nu K_\nu(2\sqrt{\nu}t\phi) & \text{if } t > 0 \\ \tau^2 + \sigma^2 & \text{if } t = 0 \end{cases}$$

K_ν is the modified Bessel function of order ν (computationally tractable)

- ν is a smoothness parameter controlling process smoothness.
[Remarkable!](#)
- $\nu = 1/2$ gives the exponential covariance function

Kriging: Spatial prediction at new locations

- **Goal:** Given observations $\mathbf{w} = (w(\mathbf{s}_1), w(\mathbf{s}_2), \dots, w(\mathbf{s}_n))^T$, predict $w(\mathbf{s}_0)$ for a new location \mathbf{s}_0
- If $w(\mathbf{s})$ is modeled as a GP, then $(w(\mathbf{s}_0), w(\mathbf{s}_1), \dots, w(\mathbf{s}_n))^T$ jointly follow multivariate normal distribution
- $w(\mathbf{s}_0) | \mathbf{w}$ follows a normal distribution with
 - Mean (**kriging estimator**): $m(\mathbf{s}_0) + \mathbf{c}^T \mathbf{C}^{-1}(\mathbf{w} - \mathbf{m})$
 - where $m = E(\mathbf{w})$, $\mathbf{C} = \text{Cov}(\mathbf{w})$, $\mathbf{c} = \text{Cov}(\mathbf{w}, w(\mathbf{s}_0))$
 - Variance: $\mathbf{C}(\mathbf{s}_0, \mathbf{s}_0) - \mathbf{c}^T \mathbf{C}^{-1} \mathbf{c}$
- The GP formulation gives the **full predictive distribution** of $w(\mathbf{s}_0) | \mathbf{w}$

Spatial linear model

$$y(\mathbf{s}) = \mathbf{x}(\mathbf{s})^\top \boldsymbol{\beta} + w(\mathbf{s}) + \epsilon(\mathbf{s})$$

- $w(\mathbf{s})$ modeled as $GP(0, C(\cdot | \boldsymbol{\theta}))$ (usually without a nugget)
- $\epsilon(\mathbf{s}) \stackrel{\text{iid}}{\sim} N(0, \tau^2)$ contributes to the nugget
- Under isotropy: $C(\mathbf{s} + \mathbf{h}, \mathbf{s}) = \sigma^2 R(\|\mathbf{h}\|; \phi)$
- $\mathbf{w} = (w(\mathbf{s}_1), \dots, w(\mathbf{s}_n))^\top \sim N(\mathbf{0}, \sigma^2 \mathbf{R}(\phi))$ where $\mathbf{R}(\phi) = \sigma^2 (R(\|s_i - s_j\|; \phi))$
- $\mathbf{y} = (y(\mathbf{s}_1), \dots, y(\mathbf{s}_n))^\top \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{R}(\phi) + \tau^2 \mathbf{I})$

Parameter estimation

- $\mathbf{y} = (y(\mathbf{s}_1), \dots, y(\mathbf{s}_n))^T \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{R}(\phi) + \tau^2\mathbf{I})$
- We can obtain MLEs of parameters $\boldsymbol{\beta}, \tau^2, \sigma^2, \phi$ based on the above model and use the estimates to kriging at new locations
- In practice, the likelihood is often very **flat** with respect to the spatial covariance parameters and choice of **initial values** is important
- Initial values can be eyeballed from empirical semivariogram of the residuals from ordinary linear regression
- Estimated parameter values can be used for kriging

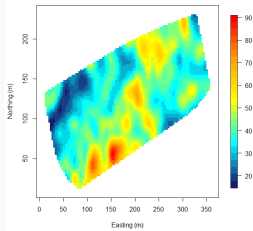
Model comparison

- For k total parameters and sample size n :
 - **AIC**: $2k - 2 \log(l(\mathbf{y} | \hat{\beta}, \hat{\theta}, \hat{\tau}^2))$
 - **BIC**: $\log(n)k - 2 \log(l(\mathbf{y} | \hat{\beta}, \hat{\theta}, \hat{\tau}^2))$
- Prediction based approaches using holdout data:
 - Root Mean Square Predictive Error (**RMSPE**):
$$\sqrt{\frac{1}{n_{out}} \sum_{i=1}^{n_{out}} (y_i - \hat{y}_i)^2}$$
 - Coverage probability (**CP**): $\frac{1}{n_{out}} \sum_{i=1}^{n_{out}} I(y_i \in (\hat{y}_{i,0.025}, \hat{y}_{i,0.975}))$
 - Width of 95% confidence interval (**CIW**):
$$\frac{1}{n_{out}} \sum_{i=1}^{n_{out}} (\hat{y}_{i,0.975} - \hat{y}_{i,0.025})$$
 - The last two approaches compares the distribution of y_i instead of comparing just their point predictions

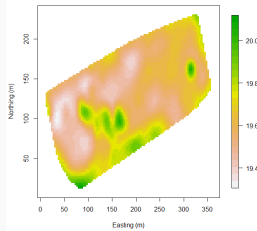
Table: Model comparison

	Spatial	Non-spatial
AIC	4419	4465
BIC	4448	4486
RMSPE	18	21
CP	93	93
CIW	77	82

WEF data: Kriged surfaces



DBH Estimates



Standard errors

- Geostatistics – Analysis of point-referenced spatial data
- Surface plots of data and residuals
- EDA with empirical semivariograms
- Modeling unknown surfaces with Gaussian Processes
- Kriging: Predictions at new locations
- Spatial linear regression using Gaussian Processes