

Introduction to Geostatistics

Abhi Datta¹, Sudipto Banerjee² and Andrew O. Finley³
July 31, 2017

¹Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland.

²Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles.

³Departments of Forestry and Geography, Michigan State University, East Lansing, Michigan.

- Course materials available at <https://abhirupdatta.github.io/spatstatJSM2017/>

What is spatial data?

- Any data with some geographical information

2

What is spatial data?

- Any data with some geographical information
- Common sources of spatial data: climatology, forestry, ecology, environmental health, disease epidemiology, real estate marketing etc
 - have many important predictors and response variables
 - are often presented as maps

2

What is spatial data?

- Any data with some geographical information
- Common sources of spatial data: climatology, forestry, ecology, environmental health, disease epidemiology, real estate 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:
 - Neuroimaging (data for each voxel in the brain)
 - Genetics (position along a chromosome)

2

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

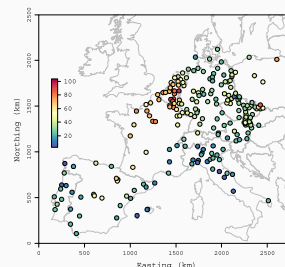


Figure: Pollutant levels in Europe in March, 2009

3

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 $\{Y(s) : s \in D\}$, D is a subset in Euclidean space.
- **Example**: $Y(s)$ is a **pollutant level** at site s
- **Conceptually**: Pollutant level exists at all possible sites
- **Practically**: Data will be a partial realization of a spatial process – observed at $\{s_1, \dots, s_n\}$
- **Statistical objectives**: **Inference** about the process $Y(s)$; **predict** at new locations.
- **Remarkable**: Can learn about entire $Y(s)$ surface. The **key**: Structured dependence

4

Exploratory data analysis (EDA): Plotting the data

- A typical setup: Data observed at n locations $\{s_1, \dots, s_n\}$
- At each s_i we observe the response $y(s_i)$ and a $p \times 1$ vector of covariates $x(s_i)'$
- **Surface plots** of the data often helps to understand spatial patterns

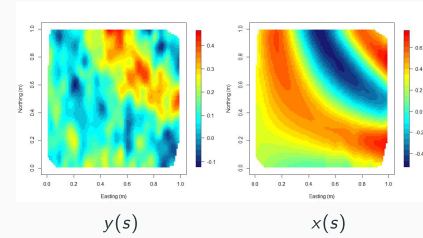


Figure: Response and covariate surface plots for Dataset 1

5

What's so special about spatial?

- Linear regression model: $y(s_i) = x(s_i)'\beta + \epsilon(s_i)$
- $\epsilon(s_i)$ are iid $N(0, \tau^2)$ errors
- $y = (y(s_1), y(s_2), \dots, y(s_n))'$; $X = (x(s_1)', x(s_2)', \dots, x(s_n)')$
- **Inference**: $\hat{\beta} = (X'X)^{-1}X'Y \sim N(\beta, \tau^2(X'X)^{-1})$
- **Prediction** at new location s_0 : $\widehat{y(s_0)} = x(s_0)'\hat{\beta}$
- Although the data is spatial, this is an **ordinary linear regression** model

6

Residual plots

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

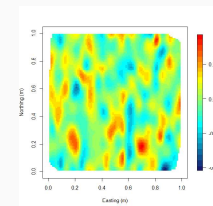


Figure: Residual plot for Dataset 1 after linear regression on $x(s)$

7

Residual plots

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

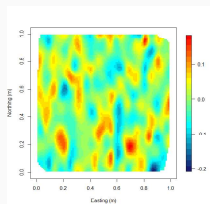


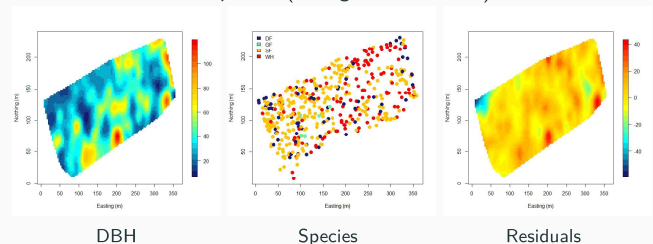
Figure: Residual plot for Dataset 1 after linear regression on $x(s)$

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

7

Western Experimental Forestry (WEF) data

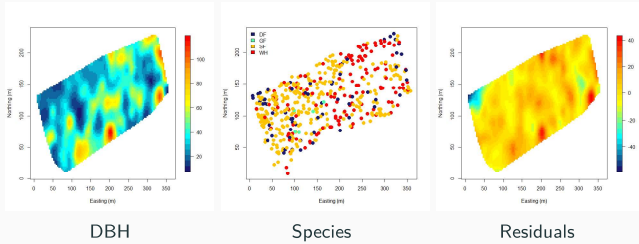
- Data consist of a census of all trees in a 10 ha. stand in Oregon
- Response of interest: Diameter at breast height (DBH)
- Covariate: Tree species (Categorical variable)



8

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)



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

8

More EDA

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

9

More EDA

- 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

9

More EDA

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

9

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:

$$N(t_k) = \{(s_i, s_j) : ||s_i - s_j|| \in I_k\}, k = 1, \dots, K.$$

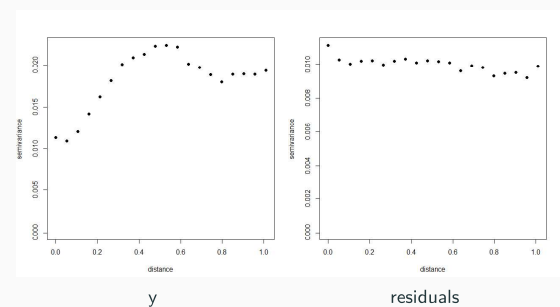
- **Empirical semivariogram**:

$$\gamma(t_k) = \frac{1}{2|N(t_k)|} \sum_{s_i, s_j \in N(t_k)} (Y(s_i) - Y(s_j))^2$$

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

10

Empirical variogram: Data 1

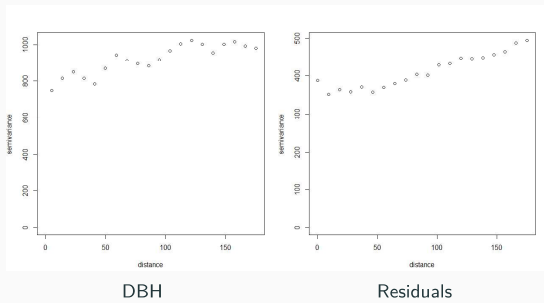


- Residuals display little spatial variation

11

Empirical variograms: WEF data

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



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

12

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:

$$y(s) = x(s)' \beta + w(s) + \epsilon(s) \text{ for all } s \in D$$
- How to choose the function $w(\cdot)$?
- Since we want to predict at any location over the entire domain D , this choice will amount to choosing a **surface** $w(s)$
- How to do this ?

13

Gaussian Processes (GPs)

- One popular approach to **model** $w(s)$ is via Gaussian Processes (GP)
- The collection of random variables $\{w(s) \mid s \in D\}$ is a GP if
 - it is a **valid** stochastic process
 - all finite dimensional densities $\{w(s_1), \dots, w(s_n)\}$ follow multivariate Gaussian distribution
- A GP is completely characterized by a mean function $m(s)$ and a covariance function $C(\cdot, \cdot)$
- Advantage:** **Likelihood** based inference.
 $w = (w(s_1), \dots, w(s_n))' \sim N(m, C)$ where
 $m = (m(s_1), \dots, m(s_n))'$ and $C = C(s_i, s_j)$

14

Valid covariance functions and isotropy

- $C(\cdot, \cdot)$ needs to be **valid**. For all n and all $\{s_1, s_2, \dots, s_n\}$, the resulting covariance matrix $C(s_i, s_j)$ for $(w(s_1), w(s_2), \dots, w(s_n))$ must be positive definite
- So, $C(\cdot, \cdot)$ needs to be a **positive definite** function
- Simplifying assumptions:
 - Stationarity:** $C(s_1, s_2)$ only depends on $h = s_1 - s_2$ (and is denoted by $C(h)$)
 - Isotropic:** $C(h) = C(\|h\|)$
 - 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 .

15

Some common isotropic covariance functions

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

16

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.

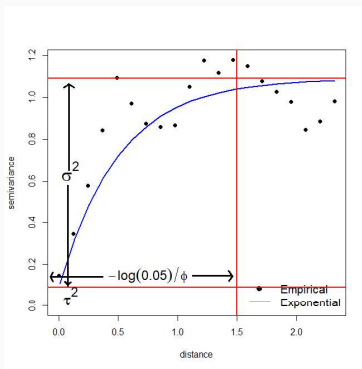
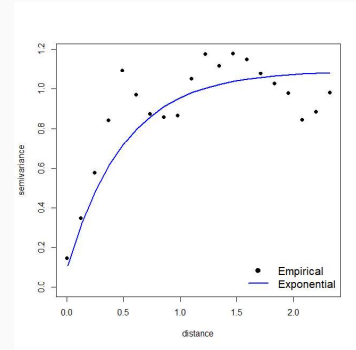
17

- **Recall:** Empirical semivariogram:

$$\gamma(t_k) = \frac{1}{2|N(t_k)|} \sum_{s_i, s_j \in N(t_k)} (Y(s_i) - Y(s_j))^2$$
- For any stationary GP,

$$E(Y(s+h) - Y(s))^2 / 2 = C(0) - C(h) = \gamma(h)$$
- $\gamma(h)$ is the **semivariogram** corresponding to the covariance function $C(h)$
- **Example:** For exponential GP,

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



- The Matérn 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

- **Goal:** Given observations $w = (w(s_1), w(s_2), \dots, w(s_n))'$, predict $w(s_0)$ for a new location s_0
- If $w(s)$ is modeled as a GP, then $(w(s_0), w(s_1), \dots, w(s_n))'$ jointly follow multivariate normal distribution
- $w(s_0) | w$ follows a normal distribution with
 - Mean (**kriging estimator**): $m(s_0) + c' C^{-1}(w - m)$
 - where $m = E(w)$, $C = Cov(w)$, $c = Cov(w, w(s_0))$
 - Variance: $C(s_0, s_0) - c' C^{-1} c$
- The GP formulation gives the **full predictive distribution** of $w(s_0) | w$

Spatial linear model

$$y(s) = x(s)' \beta + w(s) + \epsilon(s)$$

- $w(s)$ modeled as $GP(0, C(\cdot | \theta))$ (usually without a nugget)
- $\epsilon(s) \stackrel{iid}{\sim} N(0, \tau^2)$ contributes to the nugget
- Under isotropy: $C(s+h, s) = \sigma^2 R(\|h\|; \phi)$
- $w = (w(s_1), \dots, w(s_n))' \sim N(0, \sigma^2 R(\phi))$ where $R(\phi) = \sigma^2 (R(\|s_i - s_j\|; \phi))$
- $y = (y(s_1), \dots, y(s_n))' \sim N(X\beta, \sigma^2 R(\phi) + \tau^2 I)$

Parameter estimation

- $y = (y(s_1), \dots, y(s_n))' \sim N(X\beta, \sigma^2 R(\phi) + \tau^2 I)$
- We can obtain MLEs of parameters $\beta, \tau^2, \sigma^2, \phi$ 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

23

Model comparison

- For k total parameters and sample size n :
 - **AIC**: $2k - 2 \log(l(y | \hat{\beta}, \hat{\theta}, \hat{\tau}^2))$
 - **BIC**: $\log(n)k - 2 \log(l(y | \hat{\beta}, \hat{\theta}, \hat{\tau}^2))$
- 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

24

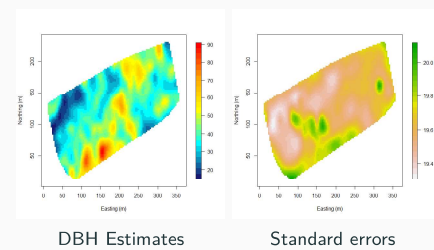
Back to WEF data

Table: Model comparison

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

25

WEF data: Kriged surfaces



26

Summary

- 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

27