Point-referenced

S.Lan

Spatial
Problems

Elements of
point-referenced
modeling

Spatial process
model

Exploratory data
analysis (EDA)

Classical spatial
prediction

Lecture 2   Point-referenced Data Models

Shiwei Lan[1]

[1]School of Mathematical and Statistical Sciences
Arizona State University

STP598   Spatiotemporal Analysis
Fall 2020

Point-referenced

S.Lan

Spatial
Problems

Elements of
point-referenced
modeling

Spatial process
model

Exploratory data
analysis (EDA)

Classical spatial
prediction

Table of Contents

Point-referenced

S.Lan

Spatial
Problems

Elements of
point-referenced
modeling

Spatial process
model

Exploratory data
analysis (EDA)

Classical spatial
prediction

RISING AND FALLING NEW CORONAVIRUS CASES
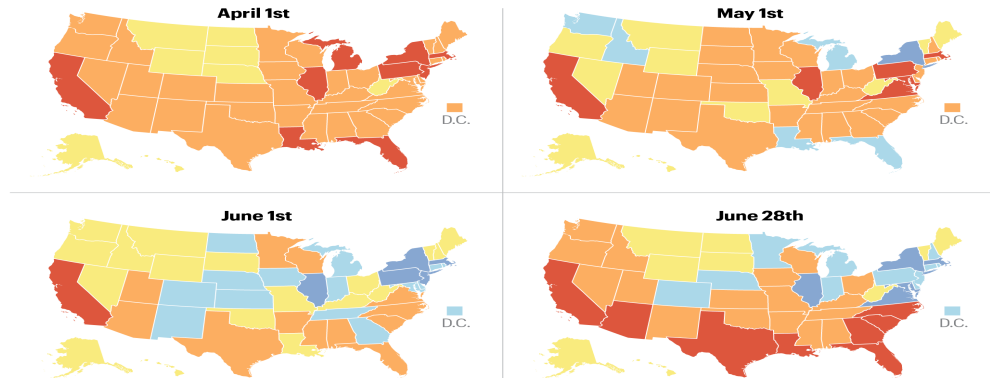## CHANGE IN DAILY NUMBER OF NEW CASES

STRONG INCREASE   INCREASE   FLAT   DECREASE   STRONG DECREASE

April 1st   May 1st

June 1st   June 28th

SEVEN-DAY AVERAGE OF NEW CASES. "STRONG" CHANGE: IN EXCESS OF 500 CASES; "FLAT": +/- 25
SOURCE: N.Y. TIMES COMPILATION OF STATE AND LOCAL GOVERNMENTS AND HEALTH DEPARTMENTS DATA

FORTUNE

Table of Contents

Point-referenced

S.Lan

Spatial
Problems

Elements of
point-referenced
modeling

Spatial process
model

Exploratory data
analysis (EDA)

Classical spatial
prediction

1. Spatial Problems

2. Elements of point-referenced modeling

3. Spatial process model

4. Exploratory data analysis (EDA)

5. Classical spatial prediction

Assume the spatial process $Y(s)$ has a mean, $\mu(s) = \mathrm{E}[Y(s)]$, and the variance of $Y(s)$ exists for all $s \in D$.

Definition (Gaussian Process)

The process $Y(s)$ is said to be Gaussian if, for any $n \geq 1$ and any set of sites $\{s_1, \cdots, s_n\}$, $Y = \{Y(s_1), \cdots, Y(s_n)\}$ has a multivariate normal distribution.

Definition ((strict) stationarity)

A process is said to be strictly stationary if, for any given $n \geq 1$, any set of n sites $\{s_1, \cdots, s_n\}$ and any $h \in \mathbb{R}^r$, the distribution of $(Y(s_1), \cdots, Y(s_n))$ is the same as that of $(Y(s_1 + h), \cdots, Y(s_n + h))$.

Definition (weak stationarity)

A spatial process $Y(\mathbf{s})$ is said to be weakly stationary if $\mu(\mathbf{s}) \equiv \mu$ (?) and

$$\text{Cov}(Y(\mathbf{s}), Y(\mathbf{s} + \mathbf{h})) = C(\mathbf{h}) \tag{1}$$

for all $\mathbf{h} \in \mathbb{R}^r$ such that $\mathbf{s}, \mathbf{s} + \mathbf{h} \in D$.

- Stationarity $\Rightarrow$ weak stationarity? Weak stationarity $\Rightarrow$ stationarity?
- How about Gaussian process?

Definition (intrinsic stationarity)

A spatial process $Y(\mathbf{s})$ is said to be intrinsically stationary if
$E[Y(\mathbf{s}+\mathbf{h}) - Y(\mathbf{s})] = 0$ *and*

$$\mathrm{E}[Y(\mathbf{s}+\mathbf{h}) - Y(\mathbf{s})]^2 = \mathrm{Var}(Y(\mathbf{s}+\mathbf{h}) - Y(\mathbf{s})) = 2\gamma(\mathbf{h}) \qquad (2)$$

- $2\gamma(\mathbf{h})$ is called *variogram* and $\gamma(\mathbf{h})$ is named *semivariogram*.
- We have

$$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h}) \qquad (3)$$

- Can we recover C from γ?
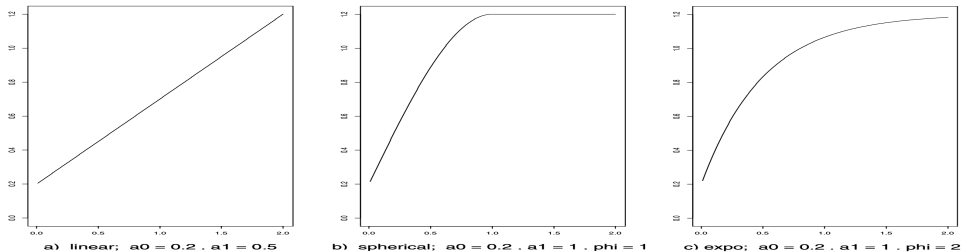- Weak stationarity $\Rightarrow$ intrinsic stationarity? Intrinsic stationarity $\Rightarrow$ weak stationarity?

Figure 2.1 *Theoretical semivariograms for three models: (a) linear, (b) spherical, and (c) exponential.*

For any set of locations $\{s_1, \cdots, s_n\}$ and any constants $a_1, \cdots, a_n$ such that $\sum_i a_i = 0$.

$$\sum_i \sum_j a_i a_j \gamma(s_i - s_j) \leq 0 \qquad (4)$$

Point-referenced

S.Lan

Spatial
Problems

Elements of
point-referenced
modeling

Spatial process
model

Exploratory data
analysis (EDA)

Classical spatial
prediction

Isotropy

Definition (isotropy)

A spatial process $Y(\mathbf{s})$ is said to be isotropic if the semivariogram function $\gamma(\mathbf{h})$ depends upon the separation vector only through its length $\|\mathbf{h}\|$, i.e.

$$\gamma(\mathbf{h}) = \gamma(\|\mathbf{h}\|) \tag{5}$$

Otherwise we say it is anisotropic.

- Linear:

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 t & \text{if } t > 0, \tau^2 > 0, \sigma^2 > 0 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

Isotropy

Point-referenced

S.Lan

Spatial
Problems

Elements of
point-referenced
modeling

Spatial process
model

Exploratory data
analysis (EDA)

Classical spatial
prediction

- Spherical:

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 & \text{if } t \geq 1/\phi \\ \tau^2 + \sigma^2 \left\{ \frac{3\phi t}{2} - \frac{1}{2}(\phi t)^3 \right\} & \text{if } 0 < t \leq 1/\phi \\ 0 & \text{otherwise} \end{cases} \qquad (7)$$

- Exponential:

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi t)) & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases} \qquad (8)$$

- Gaussian:

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi^2 t^2)) & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases} \qquad (9)$$

Variograms

Point-referenced

S.Lan

Spatial Problems

Elements of point-referenced modeling

Spatial process model

Exploratory data analysis (EDA)

Classical spatial prediction

model	Variogram, $\gamma(t)$
Linear	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 t & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$
Spherical	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 & \text{if } t \geq 1/\phi \\ \tau^2 + \sigma^2 \left[\frac{3}{2}\phi t - \frac{1}{2}(\phi t)^3\right] & \text{if } 0 < t \leq 1/\phi \\ 0 & \text{otherwise} \end{cases}$
Exponential	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi t)) & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$
Powered exponential	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-|\phi t|^p)) & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$
Gaussian	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi^2 t^2)) & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$
Rational quadratic	$\gamma(t) = \begin{cases} \tau^2 + \frac{\sigma^2 t^2}{(1+\phi t^2)} & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$
Wave	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2(1 - \frac{\sin(\phi t)}{\phi t}) & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$
Power law	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 t^\lambda & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$
Matérn	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left[1 - \frac{(2\sqrt{\nu}t\phi)^\nu}{2^{\nu-1}\Gamma(\nu)} K_\nu(2\sqrt{\nu}t\phi)\right] & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$
Matérn at $\nu = 3/2$	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left[1 - (1 + \phi t)\exp\left(-\phi t\right)\right] & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$

Table 2.2 *Summary of variograms for common parametric isotropic models.*

Covariances

Model	Covariance function, $C(t)$
Linear	$C(t)$ does not exist
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	$C(t) = \begin{cases} \sigma^2 \exp(-|\phi t|^p) & \text{if } t > 0 \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$
Gaussian	$C(t) = \begin{cases} \sigma^2 \exp(-\phi^2 t^2) & \text{if } t > 0 \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$
Rational quadratic	$C(t) = \begin{cases} \sigma^2 \left(1 - \frac{t^2}{(1+\phi t^2)} \right) & \text{if } t > 0 \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$
Wave	$C(t) = \begin{cases} \sigma^2 \frac{\sin(\phi t)}{\phi t} & \text{if } t > 0 \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$
Power law	$C(t)$ does not exist
Matérn	$C(t) = \begin{cases} \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(2\sqrt{\nu} t \phi \right)^\nu K_\nu \left(2\sqrt{\nu} t \phi \right) & \text{if } t > 0 \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$
Matérn at $\nu = 3/2$	$C(t) = \begin{cases} \sigma^2 \left(1 + \phi t \right) \exp\left(-\phi t \right) & \text{if } t > 0 \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$

Table 2.1 *Summary of covariance functions (covariograms) for common parametric isotropic models.*

- Matheron (1963) proposed the *empirical semivariogram* as an estimator

$$\widehat{\gamma}(t) = \frac{1}{2N(t)} \sum_{(s_i, s_j) \in N(t)} [Y(s_i) - Y(s_j)]^2 \qquad (10)$$

where for some grid $0 < t_1 < \cdots t_k$, let $I_k = (t_{k-1}, t_k)$,

$$N(t_k) = \{(s_i, s_j) : \|s_i - s_j\| \in I_k\}, \quad k = 1, \cdots, K \qquad (11)$$

- Cressie and Hawkins (1980) proposed a robustified estimate

$$\widehat{\gamma}(t) = \frac{1}{2N(t)} \sum_{(s_i, s_j) \in N(t)} |Y(s_i) - Y(s_j)|^{1/2} \qquad (12)$$

Table of Contents

1. Spatial Problems

2. Elements of point-referenced modeling

3. Spatial process model

4. Exploratory data analysis (EDA)

5. Classical spatial prediction

Covariance functions

- We require a 'valid' covariance $c(h) = \mathrm{Cov}(Y(s), Y(s+h))$ such that for any finite sits $s_1, \cdots, s_n$ and for any $a_1, \cdots, a_n$,

$$\mathrm{Var}\left[\sum_i a_i Y(s_i)\right] = \sum_{i,j} a_{i,j} \mathrm{Cov}(Y(s_i), Y(s_j)) = \sum_{i,j} a_{i,j} c(s_i - s_j) \geq 0 \quad (13)$$

- To verify the positivity of the covariance function, we have the following *Bochner's Theorem*.

Theorem

$c(h)$ *is positive definite if and only if*

$$c(h) = \int \cos(w^T h) G(dw) \quad (14)$$

where G is a bounded, positive, symmetric about 0 measure in $\mathbb{R}^r$.

- Note $c(0) = \int G(d\mathrm{w})$ becomes a normalizing constant.
- $G(d\mathrm{w})/c(0)$ is referred to as the *spectral distribution* that induces $c(\mathrm{h})$.
- If $G(d\mathrm{w})$ has a density with respect to Lebesgue measure, i.e. $G(d\mathrm{w}) = g(\mathrm{w})d\mathrm{w}$, then $g(\mathrm{w})/c(0)$ is referred to as *spectral density*.
- Due to the symmetry of G around 0, we have $c(\mathrm{h})$ as a characteristic function of G:

$$c(\mathrm{h}) = \int \exp\{i\mathrm{w}^T\mathrm{h}\}G(d\mathrm{w}) \tag{15}$$

- Denote the Fourier transform of $c(\mathrm{h})$ as $\widehat{c}(\mathrm{w}) = \int \exp\{-i\mathrm{w}^T\mathrm{h}\}c(\mathrm{h})d\mathrm{h}$. Then we have the spectral density as $g(\mathrm{w}) = (2\pi)^{-r}\widehat{c}(\mathrm{w})/c(0)$.

Point-referenced

S.Lan

Spatial
Problems

Elements of
point-referenced
modeling

Spatial process
model

Exploratory data
analysis (EDA)

Classical spatial
prediction

Construction of covariance/correlation functions

There are many ways to construct correlation functions from existing ones.

- *mixing*: Suppose $c_1, \cdots, c_m$ are valid correlation function in $\mathbb{R}^r$, then $c(\mathsf{h}) = \sum_{i=1}^{m} p_i c_i(\mathsf{h})$ is also a valid correlation function for $p_i > 0$ and $\sum_{i=1}^{m} p_i = 1$.
- *product*: $c(\mathsf{h}) = \prod_{i=1}^{m} c_i(\mathsf{h})$ is also a valid correlation function.
- *convolution*: $c_{12}(\mathsf{h}) = \int c_1(\mathsf{h} - \mathsf{t}) c_2(\mathsf{t}) d\mathsf{t}$.

We could also construction anisotropic correlation c from isotropic one ρ:

$$c(\mathsf{s} - \mathsf{s}') = \sigma^2 \rho((\mathsf{s} - \mathsf{s}')^T B(\mathsf{s} - \mathsf{s}')) \tag{16}$$

where B is a positive definite matrix in $\mathbb{R}^r$. This is also called *geometric anisotropy*.

Table of Contents

Point-referenced

S.Lan

Spatial
Problems

Elements of
point-referenced
modeling

Spatial process
model

Exploratory data
analysis (EDA)

Classical spatial
prediction

Point-referenced

S.Lan

Spatial
Problems

Elements of
point-referenced
modeling

Spatial process
model

Exploratory data
analysis (EDA)

Classical spatial
prediction

Exploratory data analysis (EDA)

- Exploratory data analysis (EDA) tools are routines for analyzing one- and two-sample data sets, regression studies, etc..
- For continuous data, the starting point is the *first law of geostatistics* that decomposes data into mean and error.
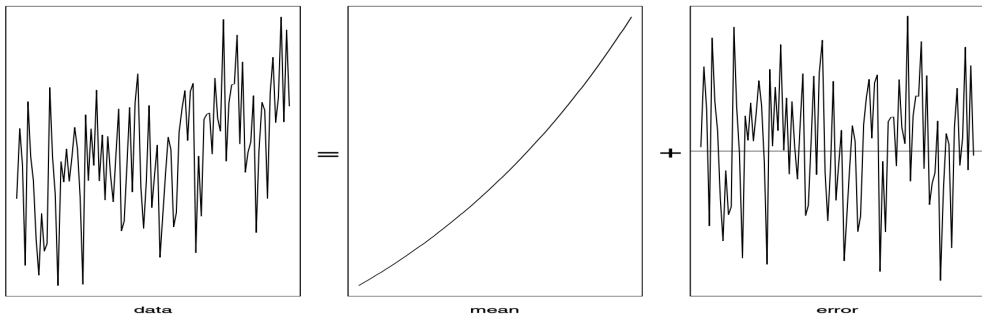- EDA tools exam both first- and second-order behavior.

Figure 2.2 *Illustration of the first law of geostatistics.*

Point-referenced

S.Lan

Spatial
Problems

Elements of
point-referenced
modeling

Spatial process
model

Exploratory data
analysis (EDA)

Classical spatial
prediction

Exploratory data analysis (EDA)

- To examine how *regular* the arrangement of the points is, we could use `drop line`, `surface plot`, or a smoothed summary `contour boxplot`.
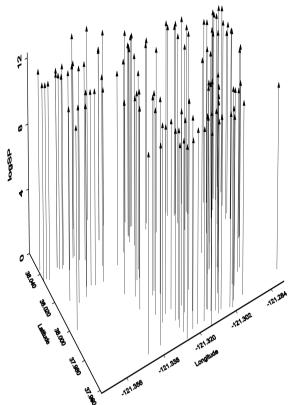

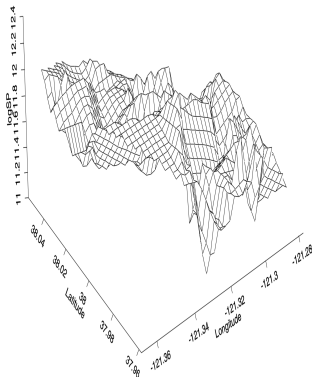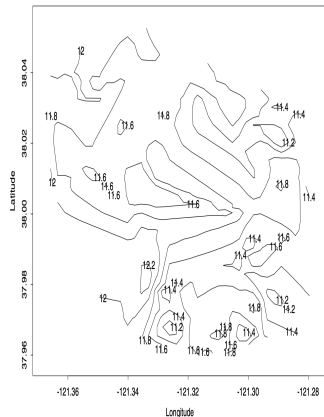Figure 2.3 *Illustrative three-dimensional "drop line" scatterplot, scallop data.*

Figure 2.4 *Illustrative three-dimensional surface ("perspective") plot, Stockton real estate data.*

Figure 2.5 *Illustrative contour plot, Stockton real estate data.*

ASU

Point-referenced

S.Lan

Spatial
Problems

Elements of
point-referenced
modeling

Spatial process
model

Exploratory data
analysis (EDA)

Classical spatial
prediction

Exploratory data analysis (EDA)

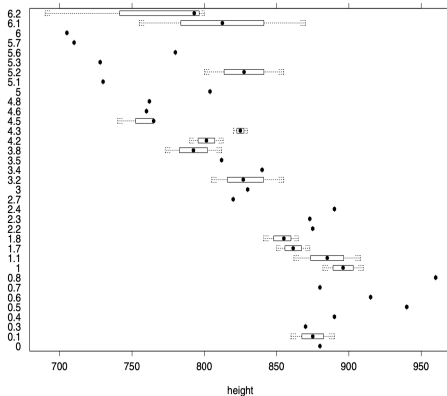- To reveal more information about the trend, $\mu(s)$, we could use `row boxplot`, and `column plot`.

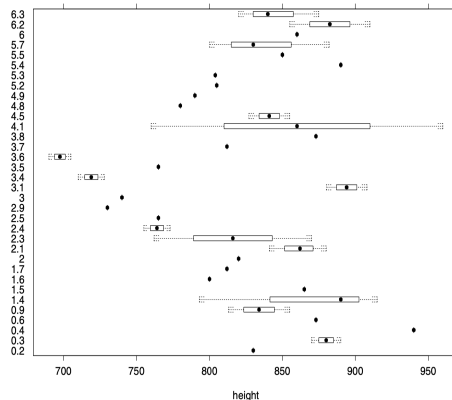Figure 2.6 *Illustrative row box plots, Diggle and Ribeiro (2002) surface elevation data.*

Figure 2.7 *Illustrative column box plots, Diggle and Ribeiro (2002) surface elevation data.*

Exploratory data analysis (EDA)

Point-referenced

S.Lan

Spatial
Problems

Elements of
point-referenced
modeling

Spatial process
model

Exploratory data
analysis (EDA)

Classical spatial
prediction

- To access small-scale behavior, we could use an empirical (nonparametric) covariance estimate

$$\widehat{c}(t_k) = \frac{1}{N_k} \sum_{(s_i, s_j) \in N(t_k)} (Y(s_i) - \bar{Y})(Y(s_j) - \bar{Y}) \qquad (17)$$

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

- On a regular grid or binning, we could create 'same-lag' scatter plots, i.e. $Y(s_i + he)$ vs $Y(s_i)$ for a fixed h and a fixed e to reveal the presence of anisotropy and nonstationarity.

- We could use a 'slide-window' type of investigation for the trend and pattern. Suppose we attach a neighborhood to each point. We could compute the sample mean, sample variance or sample correlation coefficient using all the points in the neightborhood.

Exploratory data analysis (EDA)

Point-referenced

S.Lan

Spatial
Problems

Elements of
point-referenced
modeling

Spatial process
model

Exploratory data
analysis (EDA)

Classical spatial
prediction

- To assess the anisotropy, we use *empirical semivariogram contour (ESC)* plot.
- For each of $\frac{N(N-1)}{2}$ pairs of sties in $\mathbb{R}^2$, calculate h_x and h_y, the separate distances along each axis. Restrict $h_y \geq 0$ and aggregate these distances into rectangular bins B_{ij}.
- Calculate empirical semivariogram values for (i, j)th bin:

$$\gamma_{ij}^* = \frac{1}{2N_{B_{ij}}} \sum_{(k,l):(\mathsf{s}_k - \mathsf{s}_l) \in B_{ij}} \left(Y(\mathsf{s}_k) - Y(\mathsf{s}_l) \right)^2 \tag{18}$$

where $N_{B_{ij}}$ is the number of sites in bin B_{ij}.
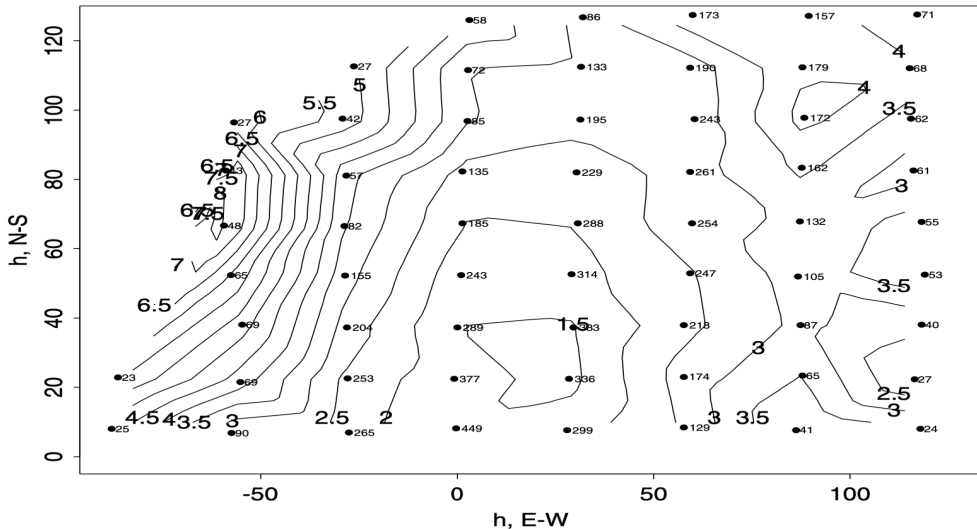
Exploratory data analysis (EDA)

Table of Contents

- *Kriging*, named by Matheron (1963), in honor of D.G. Krige (1951).
- Given observations of a random field $Y = (Y(s_1), \cdots, Y(s_n))$, how to predict the variable Y at a site s_0 where it ha not been observed?
- Suppose we model

$$Y = X\beta + \epsilon, \quad \epsilon \sim N(0, \Sigma) \tag{19}$$

- We have

$$\mathrm{E}[(Y(s_0) - f(y))^2 | y] = \mathrm{E}[(Y(s_0) - \mathrm{E}[Y(s_0)|y])^2 | y] + (\mathrm{E}[Y(s_0)|y] - f(y))^2 \tag{20}$$

- $f(y) = \mathrm{E}[Y(s_0)|y]$, the posterior mean, is the best solution!

Classical spatial prediction

Point-referenced

S.Lan

Spatial
Problems

Elements of
point-referenced
modeling

Spatial process
model

Exploratory data
analysis (EDA)

Classical spatial
prediction

- Let
$$\Sigma = \sigma^2 H(\phi) + \tau^2 I \tag{21}$$
where $(H(\phi))_{i,j} = \rho(\phi; d_{ij})$.

- Using the property of multivariate normal $[Y(s_0), y]$, we obtain
$$\begin{aligned}
\mathrm{E}[Y(s_0)|y] &= x_0^T \boldsymbol{\beta} + \boldsymbol{\gamma}^T \Sigma^{-1}(y - X\boldsymbol{\beta}) \\
\mathrm{Var}[Y(s_0)|y] &= \sigma_0^2 + \tau^2 - \boldsymbol{\gamma}^T \Sigma^{-1}\boldsymbol{\gamma}
\end{aligned} \tag{22}$$
where $x_0 = x(s_0)$, and $\boldsymbol{\gamma}^T = (\sigma^2 \rho(\phi; d_{01}), \cdots, \sigma^2 \rho(\phi; d_{0n}))$.

- In practice, we need estimate parameters using data. Note we modify $f(y)$

$$\widehat{f(y)} = x_0^T \widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\gamma}}^T \widehat{\Sigma}^{-1} (y - X\widehat{\boldsymbol{\beta}}) \tag{23}$$

where $\widehat{\boldsymbol{\gamma}} = (\hat{\sigma}^2 \rho(\hat{\phi}; d_{01,}), \cdots, \hat{\sigma}^2 \rho(\hat{\phi}; d_{0n,}))$, $\widehat{\boldsymbol{\beta}} = (X^T\widehat{\Sigma}^{-1}X)^{-1}X^T\widehat{\Sigma}^{-1}y$, and $\widehat{\Sigma} = \hat{\sigma}^2 H(\hat{\phi})$.

- Thus $\widehat{f(y)}$ can be written as linear function of y, $\widehat{f(y)} = \boldsymbol{\lambda}^T y$, where

$$\boldsymbol{\lambda} = \widehat{\Sigma}^{-1}\widehat{\boldsymbol{\gamma}} + \widehat{\Sigma}^{-1}X(X^T\widehat{\Sigma}^{-1}X)^{-1}(x_0 - X^T\widehat{\Sigma}^{-1}\widehat{\boldsymbol{\gamma}}) \tag{24}$$