Lecture 2 Point-referenced Data Models

Shiwei Lan[1]

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

STP598 Spatiotemporal Analysis
Fall 2021

Table of Contents

Spatial Problems

Spatiotemporal Problems

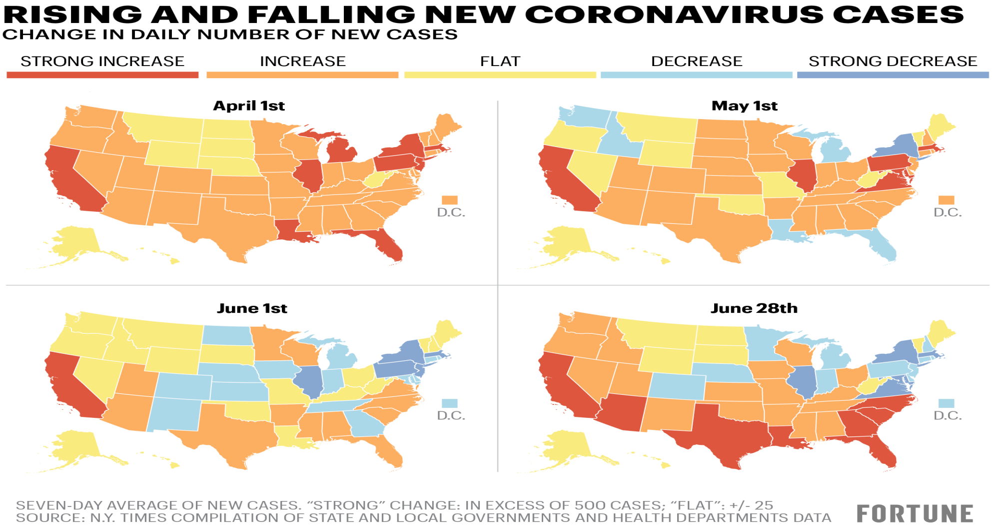

Table of Contents

Elements of Point-referenced Modeling

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

Definition (Gaussian Process)

The process $Y(\mathbf{s})$ is said to be Gaussian if, for any $n \geq 1$ and any set of sites $\{\mathbf{s}_1, \cdots, \mathbf{s}_n\}$, $\mathbf{Y} = \{Y(\mathbf{s}_1), \cdots, Y(\mathbf{s}_n)\}$ 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 $\{\mathbf{s}_1, \cdots, \mathbf{s}_n\}$ and any $\mathbf{h} \in \mathbb{R}^r$, the distribution of $(Y(\mathbf{s}_1), \cdots, Y(\mathbf{s}_n))$ is the same as that of $(Y(\mathbf{s}_1 + \mathbf{h}), \cdots, Y(\mathbf{s}_n + \mathbf{h}))$.

Stationarity

Definition (weak stationarity)

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

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

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?

Stationarity

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*

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

- $2\gamma(\mathbf{h})$ is called *variogram* and $\gamma(\mathbf{h})$ is named *semivariogram*.
- We have
$$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h}) \tag{3}$$
- Can we recover C from γ?
- Weak stationarity $\Rightarrow$ intrinsic stationarity? Intrinsic stationarity $\Rightarrow$ weak stationarity?

Variogram

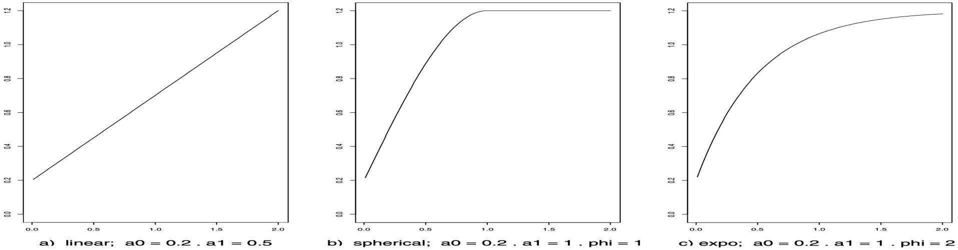

Figure 2.1 *Theoretical semivariograms for three models: (a) linear, (b) spherical, and (c) exponential.*

For any set of locations $\{\mathbf{s}_1, \cdots, \mathbf{s}_n\}$ and any constants $a_1, \cdots, a_n$ such that $\sum_i a_i = 0$.

$$\sum_i \sum_j a_i a_j \gamma(\mathbf{s}_i - \mathbf{s}_j) \leq 0 \tag{4}$$

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

- Spherical:

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

- Exponential:

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

- Gaussian:

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

Variograms

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	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\lvert\phi t\rvert^p)) & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$
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 $\nu = 3/2$	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left[1 - (1 + \phi t)\exp(-\phi t)\right] & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$

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	$C(t) = \begin{cases} \sigma^2 \exp(-\lvert\phi t\rvert^p) & \text{if } t > 0 \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$
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	$C(t) = \begin{cases} \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (2\sqrt{\nu} t\phi)^\nu K_\nu(2\sqrt{\nu} t\phi) & \text{if } t > 0 \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$
Matérn at $\nu = 3/2$	$C(t) = \begin{cases} \sigma^2 (1 + \phi t) \exp(-\phi t) & \text{if } t > 0 \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases}$

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

Empirical semivariogram

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

$$\widehat{\gamma}(t) = \frac{1}{2N(t)} \sum_{(\mathbf{s}_i,\mathbf{s}_j)\in N(t)} [Y(\mathbf{s}_i) - Y(\mathbf{s}_j)]^2 \tag{10}$$

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

$$N(t_k) = \{(\mathbf{s}_i, \mathbf{s}_j) : \|\mathbf{s}_i - \mathbf{s}_j\| \in I_k\}, \quad k = 1, \cdots, K \tag{11}$$

- Cressie and Hawkins (1980) proposed a robustified estimate

$$\widehat{\gamma}(t) = \frac{1}{2N(t)} \sum_{(\mathbf{s}_i,\mathbf{s}_j)\in N(t)} |Y(\mathbf{s}_i) - Y(\mathbf{s}_j)|^{1/2} \tag{12}$$

Table of Contents

Covariance functions

- We require a 'valid' covariance $c(\mathbf{h}) = \mathrm{Cov}(Y(\mathbf{s}), Y(\mathbf{s}+\mathbf{h}))$ such that for any finite sits $\mathbf{s}_1, \cdots, \mathbf{s}_n$ and for any $a_1, \cdots, a_n$,

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

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

Theorem

$c(\mathbf{h})$ is positive definite if and only if

$$c(\mathbf{h}) = \int \cos(\mathbf{w}^T \mathbf{h}) G(d\mathbf{w}) \quad (14)$$

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

Spectra

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

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

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

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 $c(\mathbf{h}) = \sum_{i=1}^m p_i c_i(\mathbf{h})$ is also a valid correlation function for $p_i > 0$ and $\sum_{i=1}^m p_i = 1$.
- *product*: $c(\mathbf{h}) = \prod_{i=1}^m c_i(\mathbf{h})$ is also a valid correlation function.
- *convolution*: $c_{12}(\mathbf{h}) = \int c_1(\mathbf{h} - \mathbf{t}) c_2(\mathbf{t}) d\mathbf{t}$.

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

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

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

Table of Contents

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.

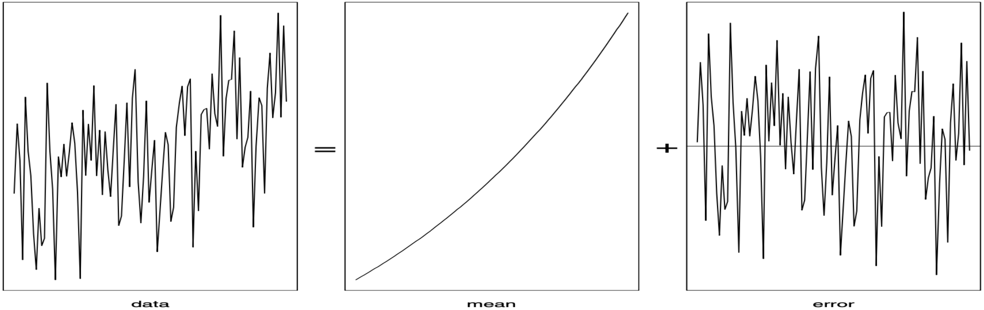

Figure 2.2 *Illustration of the first law of geostatistics.*

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`.

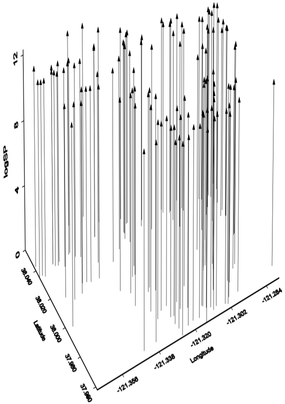

Figure 2.3 *Illustrative three-dimensional "drop line" scatterplot, scallop data.*

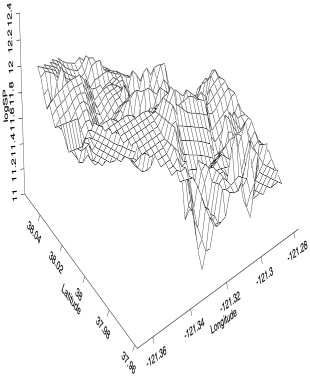

Figure 2.4 *Illustrative three-dimensional surface ("perspective") plot, Stockton real estate data.*

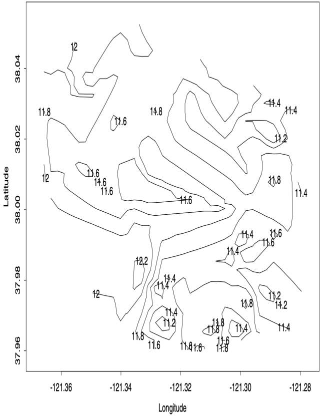

Figure 2.5 *Illustrative contour plot, Stockton real estate data.*

Exploratory data analysis (EDA)

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

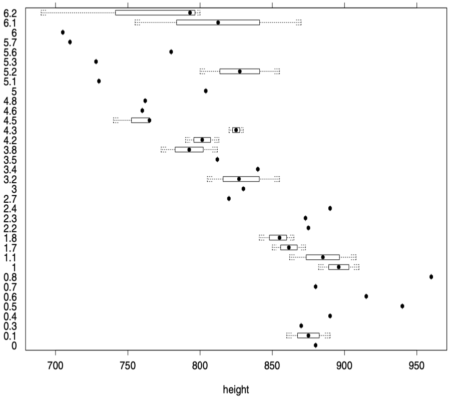

Figure 2.6 *Illustrative row box plots, Diggle and Ribeiro (2002) surface elevation data.*

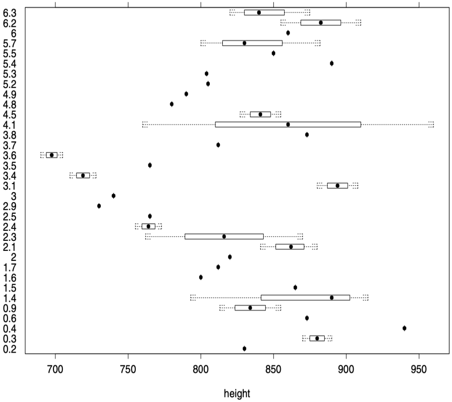

Figure 2.7 *Illustrative column box plots, Diggle and Ribeiro (2002) surface elevation data.*

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

$$\widehat{c}(t_k) = \frac{1}{N_k} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(t_k)} (Y(\mathbf{s}_i) - \bar{Y})(Y(\mathbf{s}_j) - \bar{Y}) \tag{17}$$

  where $N(t_k) = \{(\mathbf{s}_i, \mathbf{s}_j) : \|\mathbf{s}_i - \mathbf{s}_j\| \in I_k\}$ for $k = 1, \cdots, K$.
- On a regular grid or binning, we could create 'same-lag' scatter plots, i.e. $Y(\mathbf{s}_i + h\mathbf{e})$ vs $Y(\mathbf{s}_i)$ for a fixed h and a fixed $\mathbf{e}$ 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 neighborhood.

Exploratory data analysis (EDA)

- 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:

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

where $N_{B_{ij}}$ is the number of sites in bin B_{ij}.

Exploratory data analysis (EDA)

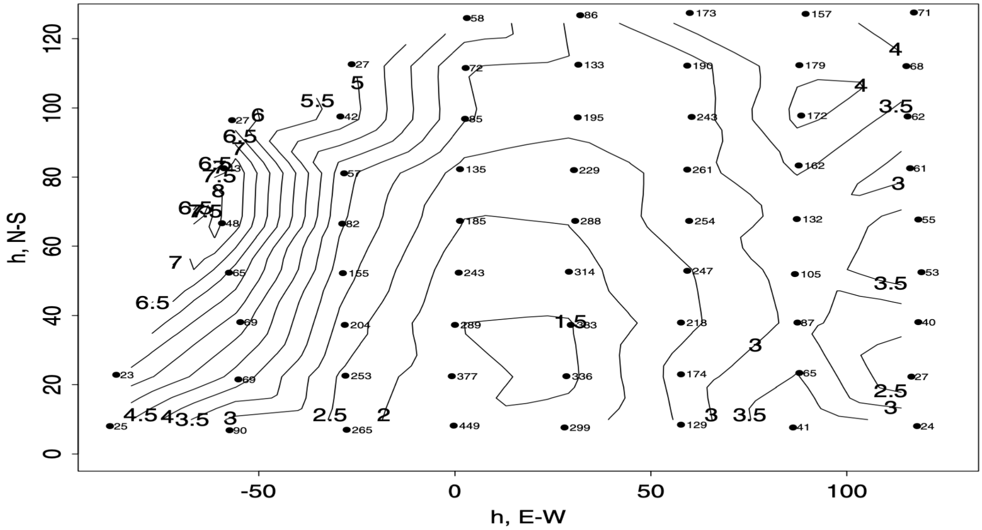

Table of Contents

Classical spatial prediction

- *Kriging*, named by Matheron (1963), in honor of D.G. Krige (1951).
- Given observations of a random field $\mathbf{Y} = (Y(\mathbf{s}_1), \cdots, Y(\mathbf{s}_n))$, how to predict the variable Y at a site $\mathbf{s}_0$ where it ha not been observed?
- Suppose we model

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

- We have

$$\mathrm{E}[(Y(\mathbf{s}_0) - f(\mathbf{y}))^2|\mathbf{y}] = \mathrm{E}[(Y(\mathbf{s}_0) - \mathrm{E}[Y(\mathbf{s}_0)|\mathbf{y}])^2|\mathbf{y}] + (\mathrm{E}[Y(\mathbf{s}_0)|\mathbf{y}] - f(\mathbf{y}))^2 \tag{20}$$

- $f(\mathbf{y}) = \mathrm{E}[Y(\mathbf{s}_0)|\mathbf{y}]$, the posterior mean, is the best solution!

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 $[Y(\mathbf{s}_0), \mathbf{y}]$, we obtain
$$\begin{aligned} \mathrm{E}[Y(\mathbf{s}_0)|\mathbf{y}] &= \mathbf{x}_0^T \boldsymbol{\beta} + \boldsymbol{\gamma}^T \boldsymbol{\Sigma}^{-1}(\mathbf{y} - X\boldsymbol{\beta}) \\ \mathrm{Var}[Y(\mathbf{s}_0)|\mathbf{y}] &= \sigma_0^2 + \tau^2 - \boldsymbol{\gamma}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma} \end{aligned} \tag{22}$$
where $\mathbf{x}_0 = \mathbf{x}(\mathbf{s}_0)$, and $\boldsymbol{\gamma}^T = (\sigma^2 \rho(\phi; d_{01}), \cdots, \sigma^2 \rho(\phi; d_{0n}))$.

Classical spatial prediction

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

$$\widehat{f(\mathbf{y})} = \mathbf{x}_0^T \widehat{\beta} + \widehat{\gamma}^T \widehat{\mathbf{\Sigma}}^{-1} (\mathbf{y} - X\widehat{\beta}) \tag{23}$$

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

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

$$\boldsymbol{\lambda} = \widehat{\mathbf{\Sigma}}^{-1} \widehat{\gamma} + \widehat{\mathbf{\Sigma}}^{-1} X (X^T \widehat{\mathbf{\Sigma}}^{-1} X)^{-1} (\mathbf{x}_0 - X^T \widehat{\mathbf{\Sigma}}^{-1} \widehat{\gamma}) \tag{24}$$