Lecture 3. Geostatistical Data

Spatial Big Data Analysis with GIS

Korean Statistical Society, Winter School, February 24, 2023

Stochastic processes and random fields

- A stochastic process is a family or collection of random variables and the members of it can be identified or indexed according to some metric.
- We call a spatial process, $Z(\mathbf{s}), \mathbf{s} \in D, D \subset \mathbb{R}^d$, a random field. Here, D is a fixed subset of d-dimensional Euclidean space.
- While the process may exist at all possible $\mathbf{s} \in D$, it is observed only at a finite set of n locations, say $\mathbf{s}_1, \ldots, \mathbf{s}_n$.
- Example: Level of pollution $Z(\mathbf{s})$ is observed at monitoring sites $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$, but we may be interested in pollution levels across the entire levels.

Autocovariance

- The covariance function of $Z(\mathbf{s})$ is defined as

$$\begin{aligned} C(\mathbf{s}_1, \mathbf{s}_2) &= \text{Cov}\{Z(\mathbf{s}_1), Z(\mathbf{s}_2)\} \\ &= E[\{Z(\mathbf{s}_1) - \mu(\mathbf{s}_1)\}\{Z(\mathbf{s}_2) - \mu(\mathbf{s}_2)\}] \end{aligned}$$

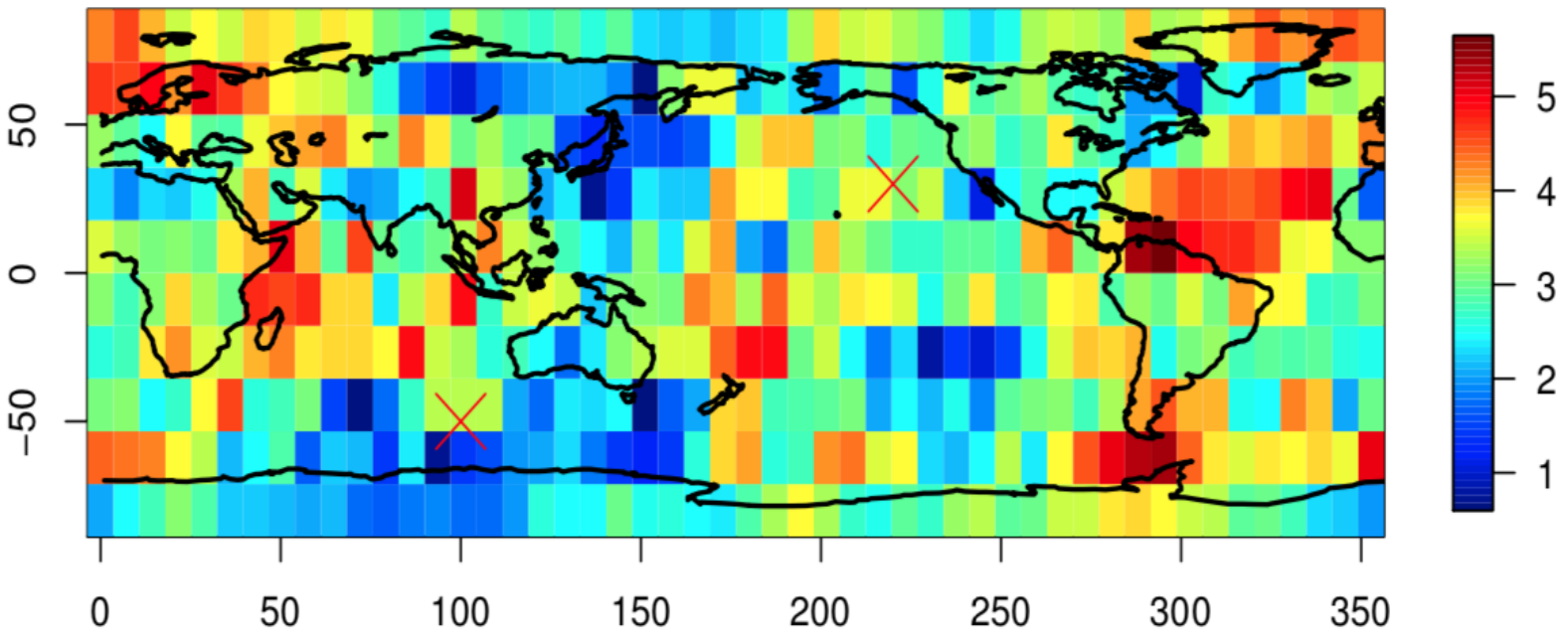
  if $\mu(\mathbf{s}) = E\{Z(\mathbf{s})\}$.

- It is called autocovariance since it is covariance of the process with itself but at two different locations.

  Note: For spatial data you have a sample size 1!

Autocovariance (conti-)

- Suppose you have the following data:

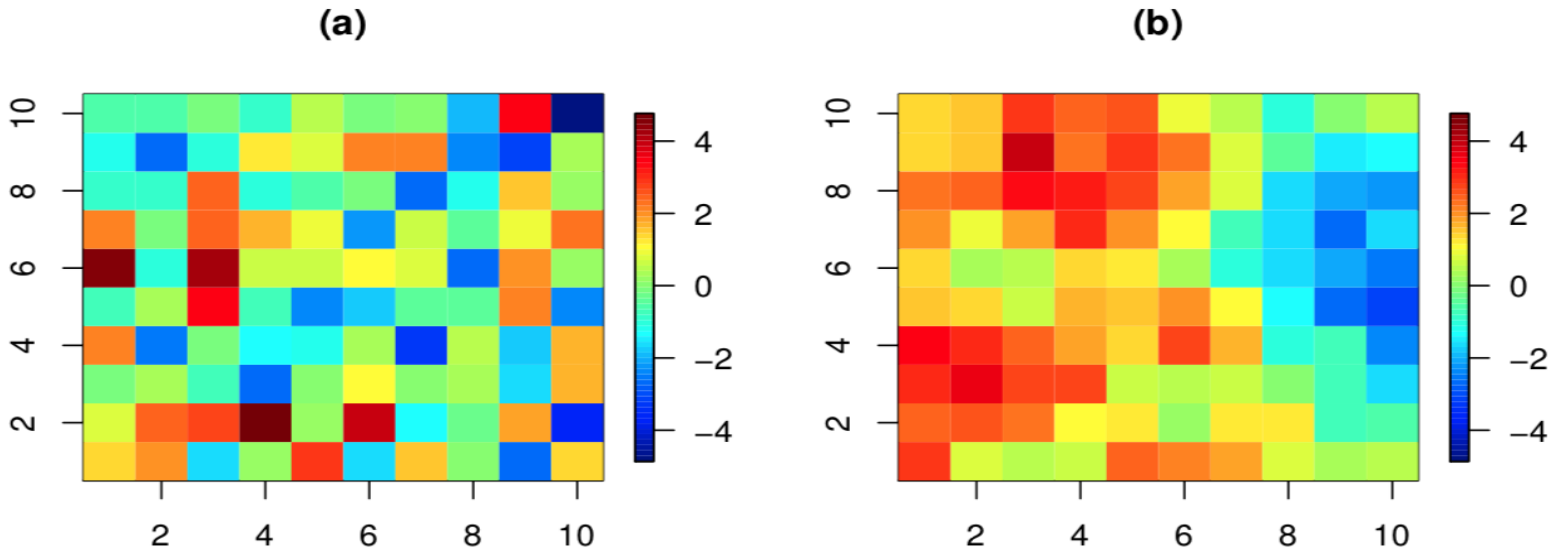


  - What is the domain of your spatial process?
  - How do you calculate the covariance between the two locations?

Autocovariance (conti-)

- With one sample, to estimate the autocovariance, we need certain assumptions on the covariance structure.
- For example,
  - (Weak) stationarity: $C(\mathbf{s}_1, \mathbf{s}_2) = C_1(\mathbf{s}_1 - \mathbf{s}_2)$ for a value covariance function C_1.
  - isotropy: $C(\mathbf{s}_1, \mathbf{s}_2) = C_2(\|\mathbf{s}_1 - \mathbf{s}_2\|)$ for a valid covariance function C_2.
- There are certain requirements that a covariance function should satisfy "**positive definiteness**".

Autocovariance (conti-)



- Both (a) and (b) give random fields on the domain $[0, 10] \times [0, 10]$, but in (a), every point is independent of each other and in (b), nearby points have correlations.
- In (a), if one pixel is missing, what would be the “best" guess for that missing pixel? How about (b)?

Covariance function of a random field

- For a real random field Z on D with $E\{Z(\mathbf{s})^2\} < \infty$ for all $\mathbf{s} \in D$, the covariance function $C(\mathbf{s}_1, \mathbf{s}_2) = \text{Cov}\{Z(\mathbf{s}_1), Z(\mathbf{s}_2)\}$ should satisfy

$$\sum_{j,k=1}^{n} c_j c_k C(\mathbf{s}_j, \mathbf{s}_k) \geq 0$$

  for all finite n, all $\mathbf{s}_1, \ldots, \mathbf{s}_n \in D$, and all real $c_1, \ldots, c_n$.

- We call such function C *positive definite* (more accurately *positive semi-definite or non-negative definite*).

- Positive definiteness is the basic requirement for a function to be a valid covariance function.

Stationarity

- We need to make certain assumptions to overcome “one sample" problem.
- A random field Z is called *weakly stationary* if it has finite second moments, its mean function is constant, and its covariance function satisfies the following:

$$\text{Cov}\{Z(\mathbf{s}_1), Z(\mathbf{s}_2)\} = C(\mathbf{s}_1 - \mathbf{s}_2)$$

- That is, stationarity means the covariance is translation invariant.
- Sometimes we assume **strict stationarity**, which requires the probability distribution to be transition invariant.
- Gaussian process is strict stationary if and only if it is weakly stationary.

Isotropy

- A random field Z is called **weakly isotropic** if it has finite second moments, its mean function is constant, and its covariance function satisfies the following:

$$\text{Cov}\{Z(\mathbf{s}_1), Z(\mathbf{s}_2)\} = C(\|\mathbf{s}_1 - \mathbf{s}_2\|)$$

- That is, isotropy means that the covariance is translation and rotation invariant.

- Note that if a random field is isotropic, then it is stationary.

Anisotropy

- A random field is **anisotropic** if it is not isotropic.
- However, there is a special kind of anisotropy. If a process is isotropic after a linear transformation of coordinates, we say the random field is *geometrically anisotropic*.
- That is, for a nonsingular matrix A, if $Z(A\mathbf{s})$ is isotropic, we say Z is geometrically anisotropic.

Property of autocovariance

- Suppose Z is weakly stationary on $\mathbb{R}^d$ with autocovariance function C.
- Then, C should satisfy
  1. $C(\mathbf{0}) \geq 0$
  2. $C(\mathbf{h}) = C(-\mathbf{h})$
  3. $|C(\mathbf{h})| \leq C(\mathbf{0})$

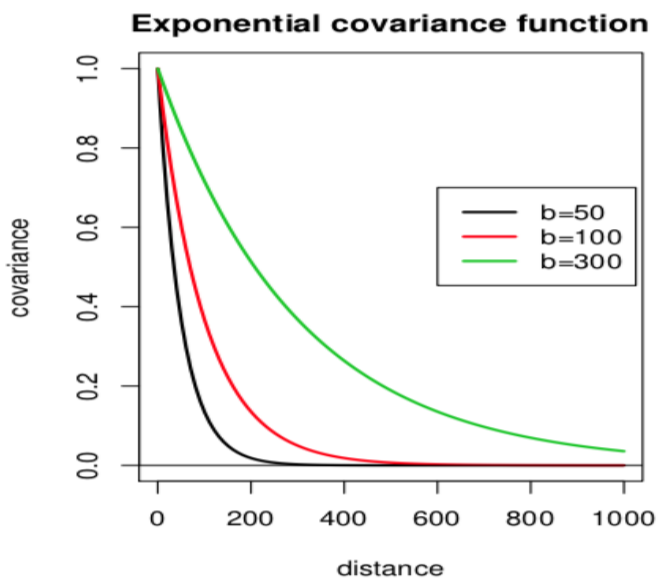
Examples of isotropic autocovariance functions

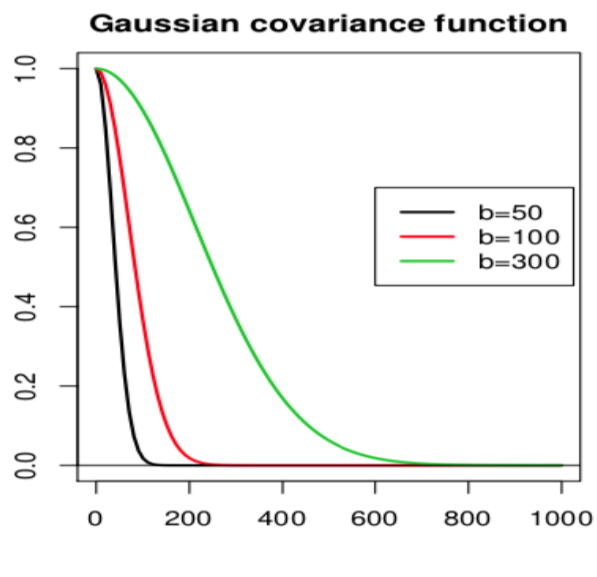
- Exponential covariance function:

$$\text{Cov}\{Z(\mathbf{s}_1), Z(\mathbf{s}_2)\} = a \exp(-\|\mathbf{s}_1 - \mathbf{s}_2\|/b)$$

- Gaussian covariance function:

$$\text{Cov}\{Z(\mathbf{s}_1), Z(\mathbf{s}_2)\} = a \exp\{-(\|\mathbf{s}_1 - \mathbf{s}_2\|/b)^2\}$$





Matérn covariance function

- A flexible and commonly used class of autocovariance functions.
- There are different parameterizations of this function:
  - $C(x) = \frac{\sqrt{\pi}\phi}{2^{\nu-1}\Gamma(\nu+1/2)\alpha^{2\nu}}(\alpha x)^{\nu}\mathcal{K}_{\nu}(\alpha x)$
  - $C(x) = \alpha(x/\beta)^{\nu}\mathcal{K}_{\nu}(x/\beta)$ where $\mathcal{K}_{\nu}$ is a modified Bessel function (in `R`: `besselK`).
  - The parameters ϕ, α, ν, β should be positive.
- This class is positive definite on $\mathbb{R}^d$ for any integer $d \geq 1$.
- Due to Matérn (1960) but practical advantages pointed out recently by Handcock and Stein (1993) and Stein (1999).

Matérn covariance function (conti-)

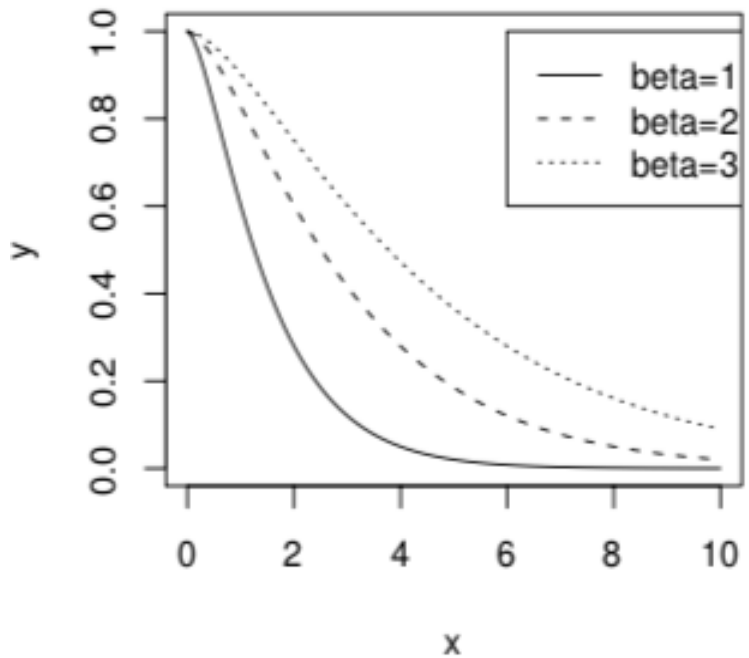
- Consider the following parameterization:

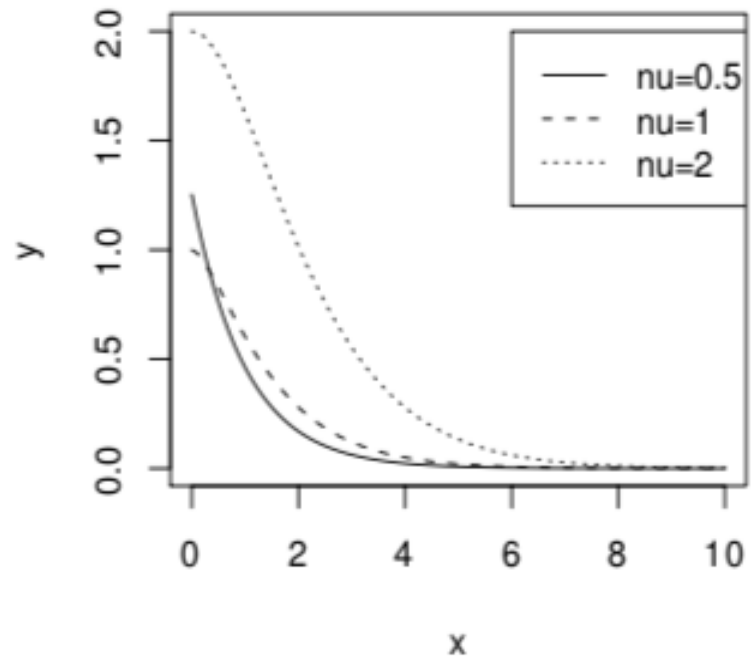
$$C(x) = \alpha(x/\beta)^{\nu}\mathcal{K}_{\nu}(x/\beta)$$

- α: “sill", β: “spatial range", and ν: “smoothness parameter".
- The larger ν, the smoother Z.
- Z is m-times mean square differentiable (C: $2m$-times differentiable) if and only if $\nu > m$.
- $\nu = 1/2 \Rightarrow$ exponential; $\nu \to \infty \Rightarrow$ Gaussian covariance function.
- $\nu = m + 1/2$, $m \in \mathbb{N}$: C is $e^{-\alpha x}$ times polynomial(x) of degree m.

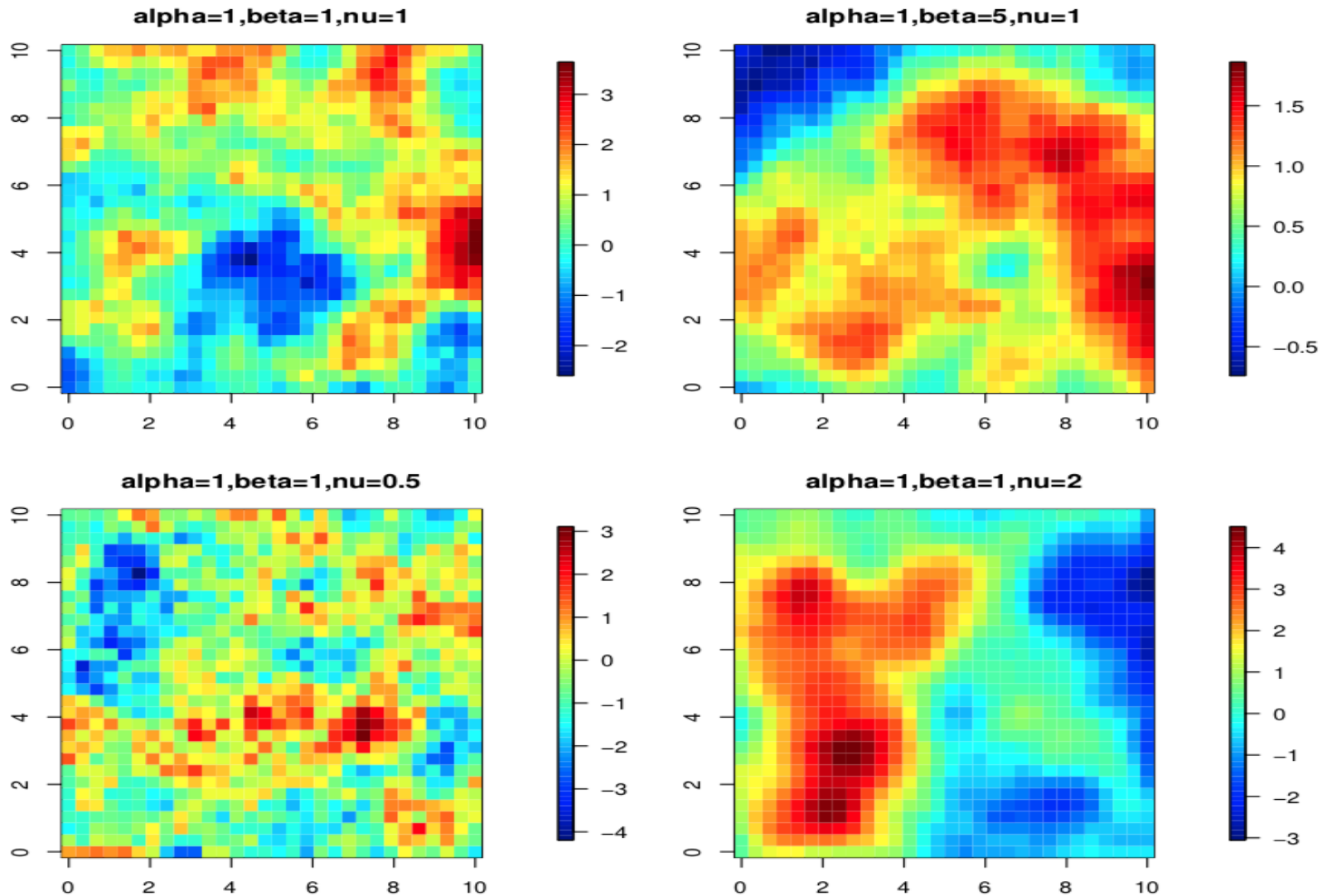
Matérn covariance function (conti-)

- The first picture has $\alpha = 1$ and $\nu = 1$.
- The second picture has $\alpha = 1$ and $\beta = 1$.





Simulated random field using Matérn covariance function

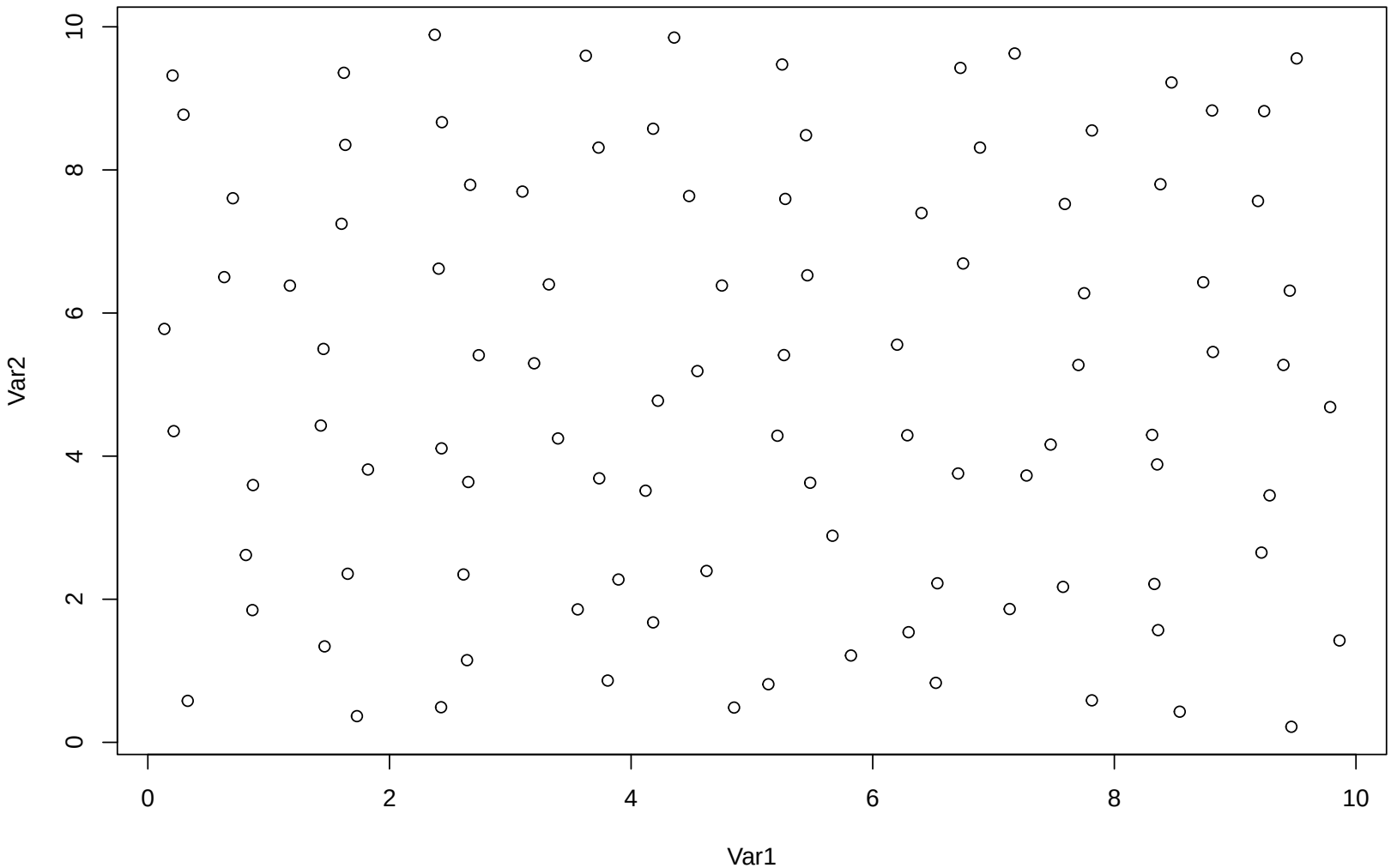


Computation of Matérn covariance function using R

- Use "`matern`" (`geoR`) or "`besselK`" (base) command to generate covariances.
- One thing to be careful about (with "`besselK`") is that, often times, when the distance is zero or close to zero, the command gives "NaN".
- When you create a covariance matrix, make sure your diagonals have actual numbers, not "NaN".
- `R` to simulate irregular locations in $[0, 10] \times [0, 10]$ not too close of each other:

```
# 100 observations
nx <- 10; n <- nx^2
grid <- expand.grid(1:nx,1:nx)
set.seed(123)
xy <- matrix(runif(n*2,-0.4,0.4),n,2)
loc <- 10*(grid-0.5+xy)/nx
```

100 locations in $[0, 10]^2$



Variogram

Consider a stationary (or isotropic) random field Z with a covariance function C.

- $\text{Var}\{Z(\mathbf{s}_1) - Z(\mathbf{s}_2)\} = 2C(\mathbf{0}) - 2C(\mathbf{s}_1 - \mathbf{s}_2) = 2\gamma(\mathbf{s}_1 - \mathbf{s}_2)$.

  $2\gamma(\cdot)$ is called a **variogram** and $\gamma(\cdot)$ is called a semivariogram.

- If Z is isotropic, then $\gamma(\mathbf{h}) = \gamma_0(\|\mathbf{h}\|)$.

  Note: if mean of Z is constant, then $\text{Var}\{Z(\mathbf{s}_1) - Z(\mathbf{s}_2)\} = E[\{Z(\mathbf{s}_1) - Z(\mathbf{s}_2)\}^2]$.

- If Z with $E\{Z(\mathbf{s}_1) - Z(\mathbf{s}_2)\} = 0$ and $\text{Var}\{Z(\mathbf{s}_1) - Z(\mathbf{s}_2)\} = 2\gamma(\mathbf{s}_1 - \mathbf{s}_2)$, then intrinsic stationary (more general than weak stationarity; variance may not exist).

Variogram models

Some isotropic variogram models (let $h = \|\mathbf{s}_1 - \mathbf{s}_2\|$):

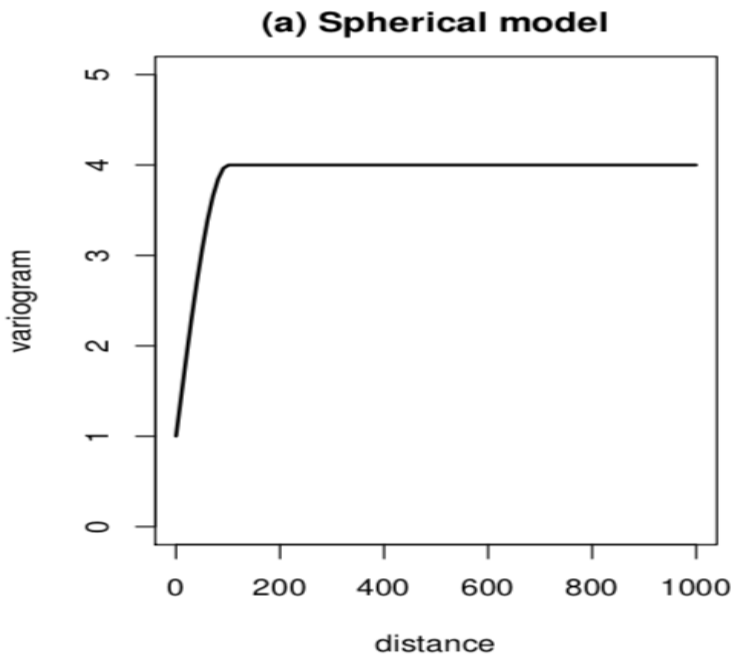
- Spherical model (valid in $\mathbb{R}^d, d = 1, 2, 3$):

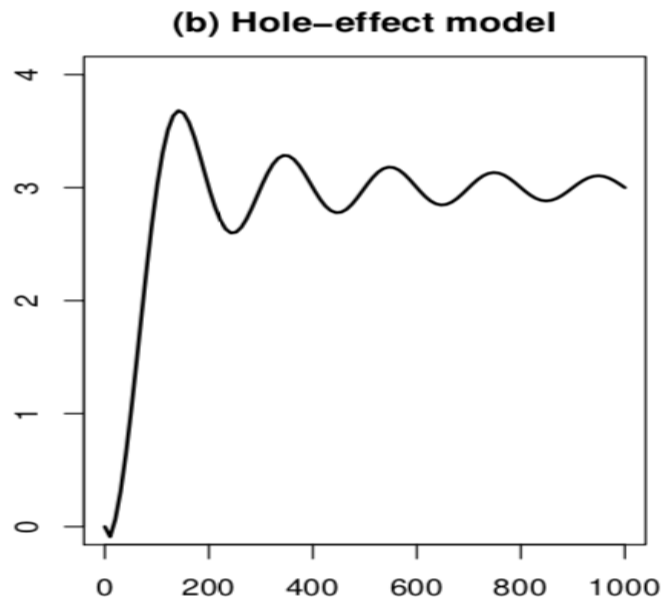
$$\gamma(h) = \begin{cases} 0 & , h = 0 \\ c_0 + c_s\{\frac{3}{2}(h/a_s) - \frac{1}{2}(h/a_s)^3\} & , 0 < h \leq a_s \\ c_0 + c_s & , h \geq a_s \end{cases}$$

- Wave (hole-effect) model (valid in $\mathbb{R}^d, d = 1, 2, 3$):

$$\gamma(h) = \begin{cases} 0 & , h = 0 \\ c_0 + c_w\{1 - a_w \sin(h/a_w)/h)\} & , h > 0 \end{cases}$$

Variogram models (conti-)





Variogram terminology

- **Nugget effect** gives the variation of the process at a finer scale than the smallest distance measured. Formally, we define nugget effect as non-zero constant c_0 such that $c_0 = \lim_{h \to 0^+} \gamma(h)$. It accounts for discontinuity at 0 (attributed to measurement error).

- **Sill** is the $\lim_{h \to \infty} \gamma(h)$ and the difference between the sill and the nugget (sill-nugget) is called the 'partial sill'.

- **Range** is h at which $\gamma(h)$ first attains the sill (applied only to the spherical variogram).

How do we plot variogram?

- Suppose you have the data points $Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)$ and assume isotropy for the data.
- First plot squared differences of the data for every possible pairs against the distance between the two locations.

  This plot is called a *variogram cloud*.
- Then, you bin the distance and in each bin, you take average of the squared differences and average of distances.

  Empirical variogram estimator (Matheron, 1962):

$$2\hat{\gamma}(\mathbf{h}) = \frac{1}{N_h} \sum_{\mathbf{s}_i - \mathbf{s}_j = \mathbf{h}} \{Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\}^2$$

  Sometimes you may look at the quantiles instead of the mean.

Covariance functions versus variograms

- Should we use covariance functions or variograms?
- Cressie (1991) argues in favor of variograms:
  1. Variogram is defined in cases when covariance function is not.
  2. Classical estimation of variograms is more robust.
- In practice, modeling often done via covariance functions (more intuitive) even though weak stationarity is not usually verified.

Parametric modeling (variograms, covariance functions)

- There are several ways of estimating parametric models.
  1. Method of moments
  2. Maximum likelihood estimation
  3. Least squares method
- The empirical variogram is a method of moments estimate:

$$2\hat{\gamma}(x) = \frac{1}{|N(x)|} \sum_{(\mathbf{s}_i, \mathbf{s}_j)} \{Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\}^2$$

where $N(x)$ is a set of $(\mathbf{s}_i, \mathbf{s}_j)$ pairs that are at a distance of x apart from each other.

Maximum likelihood estimation (MLE)

- Suppose a random variable X has a density (pdf) $f(x;\theta)$ such that $F(x;\theta) = P(X \leq x) = \int_{-\infty}^{x} f(y;\theta)dy$.
- We observe n data points $x_1, \ldots, x_n$, a random sample of size n from the distribution of X.
- We consider the likelihood $L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$ and $\hat{\theta}$ that maximizes $L(\theta)$ is the maximum likelihood estimator of θ.
- An example for $X \sim N(\mu, 1)$ with $x_1, \ldots, x_n$. What is the MLE for μ?

MLE for spatial data

- If we observe $\mathbf{Z} = \{Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)\}^\top$ and if we assume $\mathbf{Z}(\mathbf{s}) \sim N(\mu\mathbf{1}, \Sigma(\boldsymbol{\theta}))$.
- We have the likelihood function as follows $L(\boldsymbol{\theta}) = \frac{1}{(2\pi)^{n/2}\det(\Sigma(\boldsymbol{\theta}))^{1/2}} \exp\{-(\mathbf{Z}-\mu\mathbf{1})^\top \Sigma^{-1}(\boldsymbol{\theta})(\mathbf{Z}-\mu\mathbf{1})/2\}$.
- In modeling $\Sigma(\boldsymbol{\theta})$, we may use one of the covariance functions (e.g., Matérn, exponential, spherical, etc.).

MLE for spatial data: Practical Issues

- In practice, we may maximize $L(\boldsymbol{\theta})$ by numerical optimization. You may maximize it analytically, however, if you use the Matérn covariance function, getting derivatives with respect to the smoothness parameter ν is tricky.

- In `R`, there are two functions, `nlm()` and `optim()` for numerical optimizations.

- Note that, in numerical optimization, we should transform the parameters in the log-likelihood functions. Some parameters are non-negative, so it is necessary to specify the lower limits (and/or upper limits) for the parameters.

Least squares method

This is a quick overview of the Least Squares method.

- Suppose we have two vectors of data $\mathbf{X} = (X_1, \ldots, X_n)^\top$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$.
- Consider the linear regression problem of $\mathbf{Y}$ on $\mathbf{X}$.
- The model is $\mathbf{Y} = a\mathbf{X} + b\mathbf{1} + \mathbf{e}$, where a and b are parameters to be estimated and we assume $\mathbf{e} \sim N(0, \sigma^2 I_n)$.
- Then, we find a and b that minimizes the sum of squares $\sum_{i=1}^n (Y_i - a - bX_i)^2$.
- This is called the Least Squares (LS) method.

Estimating variograms using LS

- Suppose $2\gamma(h;\boldsymbol{\theta})$ is the true variogram and we have empirical variogram values at n distance lags: $2\hat{\gamma}(h_i), i = 1, \ldots, n$.
- We consider a model $\hat{\gamma}(h) = \gamma(h;\boldsymbol{\theta}) + e(h)$ and we assume e is mean zero. The covariance structure of e:
  - Since we have n "observed" variograms at distance lags $h_i, i = 1, \ldots, n$, we need to consider a covariance matrix of e with dimension $n \times n$.
  - Are $e(h_i)$ and $e(h_j)$ for $i \neq j$ independent? How can we calculate cov$\{e(h_i), e(h_j)\}$?
- If we denote the covariance matrix of e by $R(\boldsymbol{\theta})$, we need to find $\boldsymbol{\theta}$ that minimizes

$$\{\hat{\gamma}(h) - \gamma(h;\boldsymbol{\theta})\}^\top R(\boldsymbol{\theta})^{-1}\{\hat{\gamma}(h) - \gamma(h;\boldsymbol{\theta})\}.$$

Simplest things to do: OLS

- Ordinary least squares (OLS)
- We pretend that $e(h_i)$'s are independent and have a constant variance, $R(\boldsymbol{\theta}) = \phi \mathbf{I}$.
- Then, the problem reduces to finding $\boldsymbol{\theta}$ that minimizes

$$\{\hat{\gamma}(h) - \gamma(h; \boldsymbol{\theta})\}^\top \{\hat{\gamma}(h) - \gamma(h; \boldsymbol{\theta})\}.$$

Simplest things to do: WLS

- Weighted least squares (WLS)
- We pretend that $e(h_i)$'s are independent but we replace the diagonal of $R(\boldsymbol{\theta})$ by

$$Var\{2\hat{\gamma}(h_i)\} \approx 2\frac{\{2\gamma(h_i)\}^2}{|N(h_i)|}.$$

  where $|N(h_i)|$ denotes the number of observations in the bin that corresponds to the lag h_i.

- Some papers in the literature show that the estimators from OLS and WLS are more or less equally efficient (or equally bad).

More sophisticated thing to do: GLS

- GLS stands for generalized least squares.
- WLS is still not good since it ignores the off diagonals of $R(\boldsymbol{\theta})$ while there should be clearly dependence between $e(h_i)$ and $e(h_j), i \neq j$.
- If we denote $T_{ij} = Z(\mathbf{s}_i) - Z(\mathbf{s}_j)$ and $h_{ij} = |\mathbf{s}_i - \mathbf{s}_j|$, then we can show that

  $$\text{Cov}(T_{ij}^2, T_{kl}^2) = 2\{\gamma(h_{il}; \boldsymbol{\theta}) + \gamma(h_{jk}; \boldsymbol{\theta}) - \gamma(h_{jl}; \boldsymbol{\theta}) - \gamma(h_{ik}; \boldsymbol{\theta})\}^2.$$

- While clearly GLS should perform better than OLS or WLS, there is a great computational difficulty.
- In practice, people commonly use OLS or WLS, e.g., Genton (1998).

Applying OLS and WLS to spatial data

- You can use the function `variofit` and other functions in the packages `geoR`.

  They let you do OLS or WLS and you can control the size of the bins.

  These functions may not work as you expect though.

- In general, in terms of efficiency OLS$<$WLS$<$GLS but in terms of ease of use GLS$<$WLS$<$OLS.

- OLS can be implemented immediately (simple optimization) but WLS and GLS require estimation of $R(\boldsymbol{\theta})$. Also no guarantee that solution exists.

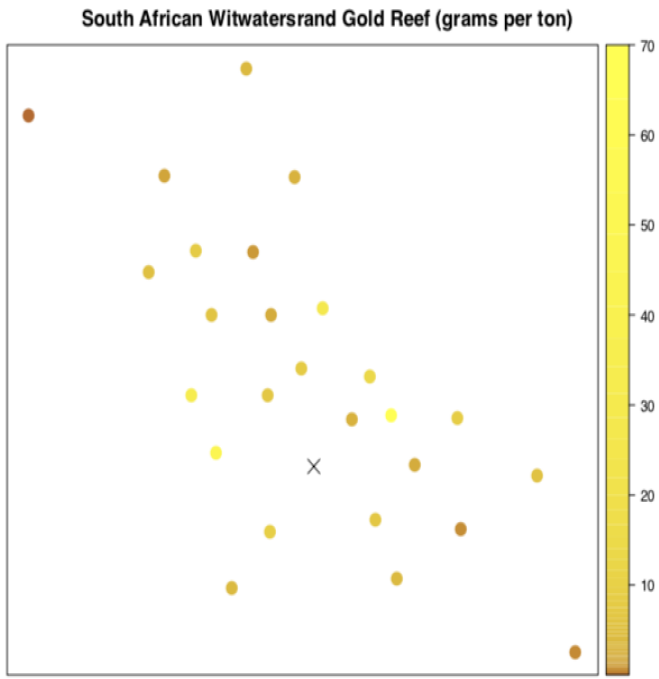
Kriging

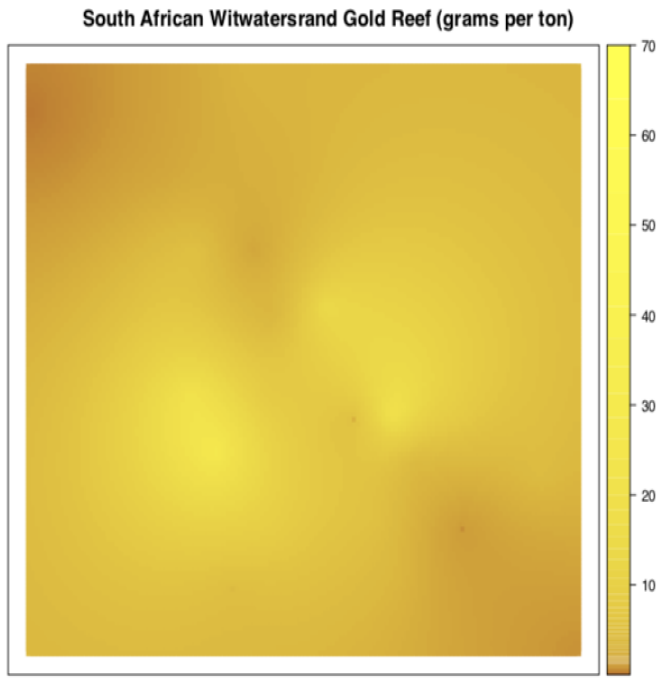
- **Kriging** is a name for **Best Linear Unbiased Prediction** for Spatial Statistics.
- Kriging is named for Danie Krige (1919-2013), a South African mining engineer.

  He was trying to predict gold grades at the Witwatersrand reef complex.

  The prediction method that he used was the one just discussed.
- For these historical reasons, spatial prediction is often called **geostatistics**.

Witwatersrand gold data





General concept of Kriging

- Suppose a random field Z has the following representation

$$Z(\mathbf{s}) = \mathbf{m}(\mathbf{s})^\top \boldsymbol{\beta} + \epsilon(\mathbf{s}), \ \epsilon(\mathbf{s}) \text{ from } N(0, \Sigma),$$

  where $\mathbf{m}$ is known, Σ is known, and $\boldsymbol{\beta}$ is unknown.

- Note that due to the spatial dependence, Σ is not a diagonal matrix in general.

- Suppose we observe the random field at locations $\mathbf{s}_i, i = 1, \ldots, n$.

- Now we want to predict the value of Z at a new location, $\mathbf{s}_0$. This is a general problem of Kriging.

- There are many different types of Kriging, mainly due to different mean structures.

General concept of Kriging (conti-)

- Our estimator for $Z(\mathbf{s}_0)$ can be expressed as a linear combination of the observations

$$\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^{n} \lambda_i Z(\mathbf{s}_i).$$

- To estimate λ_i's we give two constraints:
  - Unbiasedness: $E\{\hat{Z}(\mathbf{s}_0)\} = \mathbf{m}(\mathbf{s}_0)^\top \boldsymbol{\beta} = \sum_{i=1}^{n} \lambda_i E\{Z(\mathbf{s}_0)\} = \sum_{i=1}^{n} \lambda_i \mathbf{m}(\mathbf{s}_i)^\top \boldsymbol{\beta}$
  - Best/optimal (MSPE): minimize $E[\{Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0)|\mathbf{Z}\}^2]$
- Simple Kriging: when the mean is known or when the mean is zero (also called objective analysis in Atmospheric sciences).
- Ordinary Kriging: when the mean is an unknown constant.
- Universal Kriging: for more general $\mathbf{m}$.

Expression for Kriging estimator

- If we denote the Kriging estimator as $\boldsymbol{\lambda}^\top \mathbf{Z}$ (i.e., $\boldsymbol{\lambda}$ is a $n \times 1$ vector with λ_i's as its elements), we can show that $\boldsymbol{\lambda} = \{K^{-1} - K^{-1}M(M^\top K^{-1} M)^{-1} M^\top K^{-1}\}\mathbf{k} + K^{-1}M(M^\top K^{-1}M)^{-1}\mathbf{m}(\mathbf{s}_0)$ where $M = (\mathbf{m}(\mathbf{s}_1), \ldots, \mathbf{m}(\mathbf{s}_n)), \mathbf{k} = \text{Cov}\{\mathbf{Z}, Z(\mathbf{s}_0)\}$, and $K = \text{Cov}(\mathbf{Z}, \mathbf{Z}^\top)$.
- The estimator for $\boldsymbol{\beta}$ is given by $\hat{\boldsymbol{\beta}} = (M^\top K^{-1} M)^{-1} M^\top K^{-1} \mathbf{Z}$ (generalized least squares estimator of $\boldsymbol{\beta}$).
- Then, $\boldsymbol{\lambda}^\top \mathbf{Z} = \mathbf{k}^\top K^{-1}(\mathbf{Z} - M\hat{\boldsymbol{\beta}}) + \mathbf{m}(\mathbf{s}_0)^\top \hat{\boldsymbol{\beta}}$.
- MSPE: $k_0 - \mathbf{k}^\top K^{-1}\mathbf{k} + \gamma^\top (M^\top K^{-1} M)^{-1}\gamma$ where $\gamma = \mathbf{m}(\mathbf{s}_0) - M^\top K^{-1}\mathbf{k}$ and $k_0 = C(\mathbf{s}_0, \mathbf{s}_0)$.
- Here, we assume that M is of full rank and K is nonsingular.

Conditional expectation as optimal predictor

- Now, to obtain the optimal predictor, we need to calculate $\hat{Z}(\mathbf{s}_0) = E\{Z(\mathbf{s}_0)|\mathbf{Z}\}$.
- **Unconditional**: Let $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$ where Y_i are iid from $N(0,1)$ and Σ is an $n \times n$ covariance matrix with $\Sigma_{ij} = \text{Cov}\{Z(\mathbf{s}_i), Z(\mathbf{s}_j)\}$. Then, $\mathbf{Z} = \boldsymbol{\mu} + \Sigma^{1/2}\mathbf{Y}$ is $N_n(\boldsymbol{\mu}, \Sigma)$. Here, $\Sigma^{1/2}$ from spectral decomposition or Choleski decomposition of Σ.

Conditional expectation as optimal predictor (conti-)

- **Conditional**: if

$$\begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix} \sim N_{n+m}\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

  then,
  $(\mathbf{Z}_2|\mathbf{Z}_1 = \mathbf{z}_1) \sim N_m(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{z}_1 - \boldsymbol{\mu}_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$.

- $\hat{Z}(\mathbf{s}_0) = E\{Z(\mathbf{s}_0)|\mathbf{Z}\} = E(\mathbf{Z}_2|\mathbf{Z}_1 = \mathbf{z}_1)$. Since $\mathbf{z}_1$ is observed (known) and all $\boldsymbol{\mu}_i$ and Σ_{ij} for $i, j = 1, 2$ are assumed to be known, we can immediately calculate the optimal predictor of the spatial process at a new location using this result.

Conditional expectation with unknown parameters

- Spatial model parameter $(\boldsymbol{\beta}, \boldsymbol{\theta})$ are unknown and are to be estimated.
  - We can use a wide variety of methods to estimate $(\boldsymbol{\beta}, \boldsymbol{\theta})$ such as Maximum Likelihood (ML), Restricted ML (REML), WLS etc.
  - 'Plug-in' approach can use same result as before, except replacing $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ with estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$, respectively.

    Note: We are not accounting for variability of the estimate $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ when we use a 'plug-in' method.

Kriging in R

- Each package, `fields` and `geoR`, has many functions that do Kriging, plotting Krigged surface, etc. (e.g., `fields` has a function called `Krig`).
- For these functions, you have to specify the covariance model (some of the known models are implemented).
- For more complex covariance models, you can code yourself (this is usually what statisticians do).

Reference

- Cressie, N. Statistics for Spatial Data. Wiley. Chapter 1.
- Banerjee, S., Carlin, B., and Gelfand, A. Hierarchical Modeling and Analysis for Spatial Data (2nd). CRC Press.
- Jun, M., Genton, M. G., and Jeong, J. Lecture Notes for Spatial Statistics. UH, KAUST, and HYU.