Geostatistical predictions using Kriging and Multivariate modeling

M. PEREIRA

Geosciences and Geoengineering Department
Mines Paris – PSL University
mike.pereira@minesparis.psl.eu

Athens Week - November 2024

■ OUTLINE

I. (Simple and Ordinary) Kriging

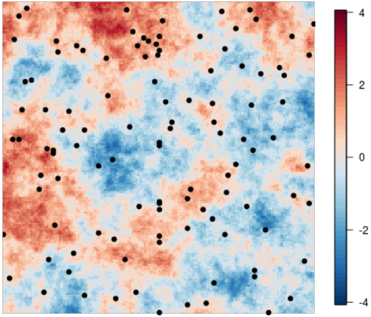
II. Extensions

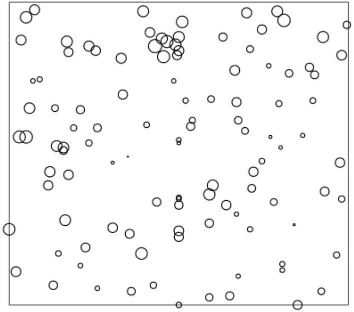
III. Multivariate setting (from H. Wackernagel's lectures)

MOTIVATION

- Data: Observations $\{z(x_i)\}_{i=1}^n$ of a variable z at n locations $x_1, \dots, x_n$ of a domain $\mathcal{D} \subset \mathbb{R}^d$
- Goal: Predict z at a new location $x_0 \in \mathcal{D}$

Start simple: **Linear predictor** $z^*(x_0) = \sum_{i=1}^n \lambda_i z(x_i)$



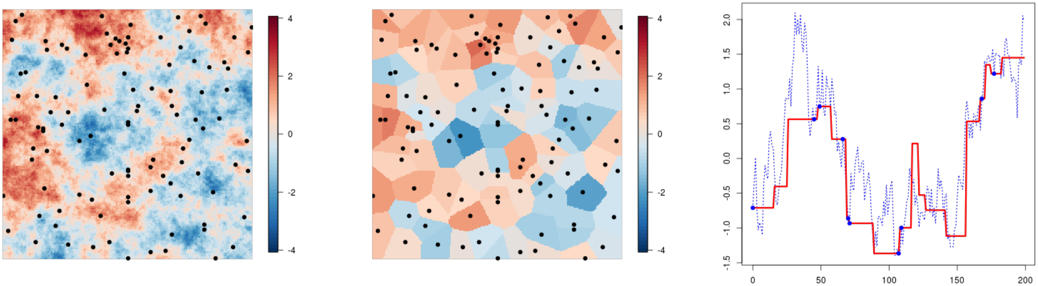


Left: Variable z. Right : Observations

■ MOTIVATION

- Data: Observations $\{z(x_i)\}_{i=1}^{n}$ of a variable z at n locations $x_1, \dots, x_n$ of a domain $\mathcal{D} \subset \mathbb{R}^d$
- Goal: Predict z at a new location $x_0 \in \mathcal{D}$

Nearest neighbor : $z^*(x_0) = z(x_{\mathrm{NN}})$: x_{NN} = nearest neighbor of x_0

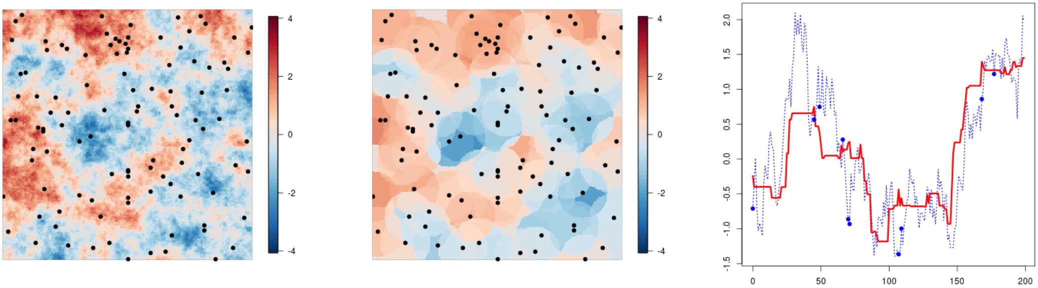


Left: Variable z. Center : Estimation. Right : Estimation (Red) and True values (Blue) on a section.

■ MOTIVATION

- Data: Observations $\{z(x_i)\}_{i=1}^n$ of a variable z at n locations $x_1, \ldots, x_n$ of a domain $\mathcal{D} \subset \mathbb{R}^d$
- Goal: Predict z at a new location $x_0 \in \mathcal{D}$

$$\textbf{Moving average}: \; z^*(x_0) = \sum_{i=1}^{n} \lambda_i z(x_i), \quad \lambda_i = 0 \text{ if } \mathrm{dist}(x, x_o) < R, \quad \sum_{i=1}^{n} \lambda_i = 1$$

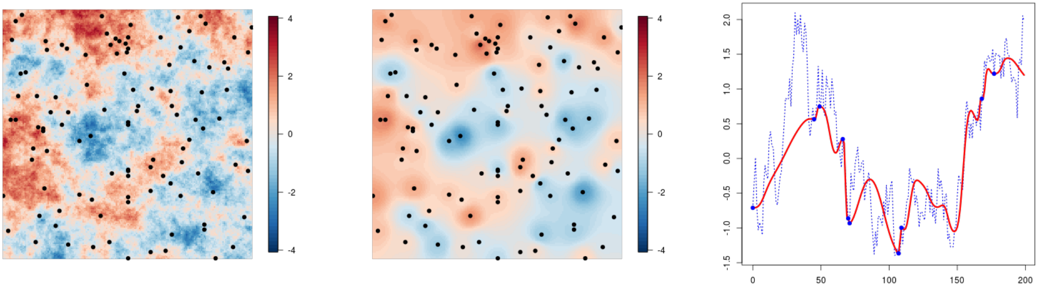


Left: Variable z. Center : Estimation. Right : Estimation (Red) and True values (Blue) on a section.

■ MOTIVATION

- Data: Observations $\{z(x_i)\}_{i=1}^n$ of a variable z at n locations $x_1, \ldots, x_n$ of a domain $\mathcal{D} \subset \mathbb{R}^d$
- Goal: Predict z at a new location $x_0 \in \mathcal{D}$

Inverse distance : $z^*(x_0) = \sum_{i=1}^{n} \lambda_i z(x_i), \quad \lambda_i = \text{dist}(x_0, x_i)^{-1} / \left(\sum_{1 \leq j \leq n} \text{dist}(x_0, x_j)^{-1} \right)$



Left: Variable z. Center : Estimation. Right : Estimation (Red) and True values (Blue) on a section.

MOTIVATION

We would like our predictor to:

- Honor the data
- Be smooth (no visual artifacts)
- Have no bias
- Account for the spatial correlation of the data
- Be “optimal” in some sense

GEOSTATISTICAL SETTING

- **Geostatistical modeling**: over the spatial domain $\mathcal{D} \subset \mathbb{R}^d$

Random Field		Observed variable
$Z : \{Z(x) : x \in \mathcal{D}\}$	Realization $\longrightarrow$	$z : \{z(x) : x \in \mathcal{D}\}$
High correlation		High "similarity"

- Assumptions on Z: Second-order stationary random function, with mean $m \in \mathbb{R}$ and covariance function $C(h)$ given by

$$\boxed{C(h) = \mathrm{Cov}(Z(x), Z(x+h)), \quad h \in \mathbb{R}^d} \quad (\text{ indep. of } x \in \mathcal{D})$$

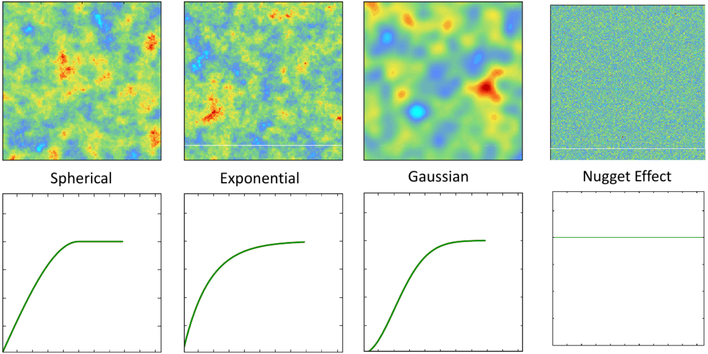
$\rightarrow$ used to model the spatial structure observed on the variable/data

GEOSTATISTICAL SETTING

- In practice, the covariance function C is estimated using its associated variogram γ

$$\gamma(h) = \frac{1}{2}\mathrm{Var}[Z(x+h) - Z(x)] = C(0) - C(h)$$

- Experimental variogram fitted using (a sum of) basic structures



BACK TO THE PREDICTION PROBLEM

- Data: Observations $\{z(x_i)\}_{i=1}^n$ of a variable z at n locations $x_1, \ldots, x_n$ of a domain $\mathcal{D} \subset \mathbb{R}^d$.
- Goal: Predict the value of z at a new location $x_0 \in \mathcal{D}$
- Idea: Start simple, look for a linear predictor $\boxed{z^*(x_0) = \beta + \sum_{i=1}^{n} \lambda_i z(x_i)}$

$\rightarrow$ How do we choose the "best" weights $\beta, \{\lambda_i\}_{i=1}^n$?

BACK TO THE PREDICTION PROBLEM

- Data: Observations $\{z(x_i)\}_{i=1}^n$ of a variable z at n locations $x_1, \dots, x_n$ of a domain $\mathcal{D} \subset \mathbb{R}^d$.
- Goal: Predict the value of z at a new location $x_0 \in \mathcal{D}$
- Idea: Start simple, look for a linear predictor $\boxed{z^*(x_0) = \beta + \sum_{i=1}^n \lambda_i z(x_i)}$

  $\rightarrow$ How do we choose the "best" weights $\beta, \{\lambda_i\}_{i=1}^n$?

- Consider the "randomized" problem $\boxed{Z^*(x_0) = \beta + \sum_{i=1}^n \lambda_i Z(x_i)}$

  and pick the weights $\beta, \{\lambda_i\}_{i=1}^n$ yielding an unbiased predictor AND the smallest error on average

  $\rightarrow$ **B**est **L**inear **U**nbiased **P**redictor (**BLUP**)

SIMPLE KRIGING

<u>*Assumption*</u>: *Z second-order stationary with mean $\underline{m=0}$ and covariance function C*

- **Simple kriging predictor** $Z^{\mathrm{SK}}(x_0)$ of $Z(x_0)$ given $Z(x_1),\dots,Z(x_n)$ = BLUP of $Z(x_0)$:
  - "Linear": $Z^{\mathrm{SK}}(x_0) = \sum_{i=1}^{n} \lambda_i^{\mathrm{SK}} Z(x_i)$, for some weights $\{\lambda_i^{\mathrm{SK}}\}_{i=1}^{n}$
  - "Unbiased": $\mathbb{E}[Z^{\mathrm{SK}}(x_0)] = \mathbb{E}[Z(x_0)] = m = 0$
  - "Best": Error variance (aka kriging variance) $\mathrm{Var}[Z^{\mathrm{SK}}(x_0) - Z(x_0)]$ is minimal

SIMPLE KRIGING

Assumption: Z *second-order stationary with mean* $\underline{m=0}$ *and covariance function* C

- **Simple kriging predictor** $Z^{\mathrm{SK}}(x_0)$ of $Z(x_0)$ given $Z(x_1),\dots,Z(x_n)$ = BLUP of $Z(x_0)$:
  - "Linear": $Z^{\mathrm{SK}}(x_0)=\sum_{i=1}^{n}\lambda_i^{\mathrm{SK}}Z(x_i)$, for some weights $\{\lambda_i^{\mathrm{SK}}\}_{i=1}^{n}$
  - "Unbiased": $\mathbb{E}[Z^{\mathrm{SK}}(x_0)]=\mathbb{E}[Z(x_0)]=m=0$
  - "Best": Error variance (aka kriging variance) $\mathrm{Var}[Z^{\mathrm{SK}}(x_0)-Z(x_0)]$ is minimal

- Take $C_{ij}:=\mathrm{Cov}(Z(x_i),Z(x_j))=C(x_i-x_j)$. The conditions above imply:

$$\forall i\in\{1,\dots,n\},\quad \sum_{j=1}^{n}C_{ij}\lambda_j^{\mathrm{SK}}=C_{i0}$$

The corresponding kriging variance is $\sigma^2_{\mathrm{SK}}(x_0)=\mathrm{Var}[Z^{\mathrm{SK}}(x_0)-Z(x_0)]=\mathrm{C}(0)-\sum_{i=1}^{n}\lambda_i^{\mathrm{SK}}C_{i0}$

SIMPLE KRIGING

- In vectorized form: Recall that $C_{ij} = C(x_i - x_j)$ qnd introduce

$$\boldsymbol{Z} = \begin{pmatrix} Z(x_1) \\ \vdots \\ Z(x_n) \end{pmatrix}, \ \boldsymbol{\lambda}^{\mathrm{SK}} = \begin{pmatrix} \lambda_1^{\mathrm{SK}} \\ \vdots \\ \lambda_n^{\mathrm{SK}} \end{pmatrix}, \ \boldsymbol{c}_0 = \begin{pmatrix} C_{10} \\ \vdots \\ C_{n0} \end{pmatrix}, \boldsymbol{C} = \begin{pmatrix} C_{11} & \dots & C_{1n} \\ \vdots & \ddots & \vdots \\ C_{n1} & \dots & C_{nn} \end{pmatrix}$$

Then,

$$Z^{\mathrm{SK}}(x_0) = (\boldsymbol{\lambda}^{\mathrm{SK}})^T \boldsymbol{Z}$$

where the weights $\boldsymbol{\lambda}^{\mathrm{SK}}$ and kriging variance are obtained as solution of the linear system:

$$\boldsymbol{C}\boldsymbol{\lambda}^{\mathrm{SK}} = \boldsymbol{c}_0 \quad \text{and} \quad \sigma^2_{\mathrm{SK}}(x_0) = C(0) - (\boldsymbol{\lambda}^{\mathrm{SK}})^T \boldsymbol{c}_0$$

$\rightarrow$ **Only depend on the location of the observations!**

SIMPLE KRIGING

- The simple kriging predictor and its variance at the location x_0 can rewritten as

$$Z^{\mathrm{SK}}(x_0) = \boldsymbol{c}_0^T \boldsymbol{C}^{-1} \boldsymbol{Z}, \quad \sigma^2_{\mathrm{SK}}(x_0) = C(0) - \boldsymbol{c}_0^T \boldsymbol{C}^{-1} \boldsymbol{c}_0$$

SIMPLE KRIGING

- The simple kriging predictor and its variance at the location x_0 can rewritten as

$$Z^{\mathrm{SK}}(x_0) = \boldsymbol{c}_0^T \boldsymbol{C}^{-1} \boldsymbol{Z}, \quad \sigma^2_{\mathrm{SK}}(x_0) = C(0) - \boldsymbol{c}_0^T \boldsymbol{C}^{-1} \boldsymbol{c}_0$$

- For multiple target locations $y_1, \ldots, y_p \in \mathcal{D}$ we get

$$\boldsymbol{Z}^{\mathrm{SK}} := \begin{pmatrix} Z^{\mathrm{SK}}(y_1) \\ \vdots \\ Z^{\mathrm{SK}}(y_p) \end{pmatrix} = \boldsymbol{C}_{\mathrm{TD}} \boldsymbol{C}^{-1} \boldsymbol{Z}$$

where $\boldsymbol{C}_{\mathrm{TD}} = \left[C(y_i - x_j) \right]_{\substack{1 \le i \le p \\ 1 \le j \le n}} \begin{pmatrix} C(x_1 - y_1) & \ldots & C(x_n - y_1) \\ \vdots & \ddots & \vdots \\ C(x_1 - y_p) & \ldots & C(x_n - y_p) \end{pmatrix} \in \mathbb{R}^{p \times n}$

The associated kriging variances are the diagonal elements of: $C(0)\boldsymbol{I}_p - \boldsymbol{C}_{\mathrm{TD}} \boldsymbol{C}^{-1} (\boldsymbol{C}_{\mathrm{TD}})^T$

WHAT IF THE MEAN IS NOT ZERO?

Assumption: Z *second-order stationary with mean* $m \in \mathbb{R}$ *and covariance function* C

- Note: $Z - m$ has mean 0
- If the mean of Z is m, simply work with the simple kriging predictors of $Z - m$

$$Z^{\mathrm{SK}}(x_0) - m = \boldsymbol{c}_0^T \boldsymbol{C}^{-1}(\boldsymbol{Z} - m\mathbf{1}_n), \quad \sigma^2_{\mathrm{SK}}(x_0) = C(0) - \boldsymbol{c}_0^T \boldsymbol{C}^{-1}\boldsymbol{c}_0$$

- For multiple target locations $y_1, \dots, y_p \in \mathcal{D}$ we get

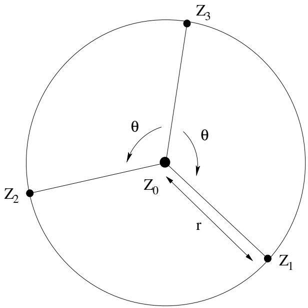
$$\boldsymbol{Z}^{\mathrm{SK}} - m\mathbf{1}_p := \begin{pmatrix} Z^{\mathrm{SK}}(y_1) - m \\ \vdots \\ Z^{\mathrm{SK}}(y_p) - m \end{pmatrix} = \boldsymbol{C}_{\mathrm{TD}}\boldsymbol{C}^{-1}(\boldsymbol{Z} - m\mathbf{1}_n)$$

where $\mathbf{1}_n = (1, \dots, 1)^T \in \mathbb{R}^n$

- (Similar result when $m = m(x)$ depends on location *but is known...*)

EXERCISE

Consider the simple Kriging of Z_0 as a function of Z_1, Z_2, Z_3, with covariance function $C(h) = e^{-|h|}$ and assuming that the mean of Z is m



Write down the kriging system and formula for the kriging variance

WHAT IF THE MEAN IS UNKNOWN?

Goal: Find a linear predictor $\boxed{\widehat{Z}(x_0) = \beta + \sum_{i=1}^{n} \lambda_i Z(x_i)}$ that is unbiased and with minimal error,

without assuming that mean $m = \mathbb{E}[Z(x)]$ is known

WHAT IF THE MEAN IS UNKNOWN?

Goal: Find a linear predictor $\boxed{\widehat{Z}(x_0) = \beta + \sum_{i=1}^{n} \lambda_i Z(x_i)}$ that is unbiased and with minimal error,

without assuming that mean $m = \mathbb{E}[Z(x)]$ is known

- Unbiasedness gives

$$0 = \mathbb{E}[\widehat{Z}(x_0) - Z(x_0)] = \beta + \sum_{i=1}^{n} \lambda_i m - m = \beta + m(\sum_{i=1}^{n} \lambda_i - 1)$$

$\rightarrow$ True for any value of m if we **impose**:

$$\sum_{i=1}^{n} \lambda_i = 1 \quad \text{and} \quad \beta = 0$$

WHAT IF THE MEAN IS UNKNOWN?

Goal: Find a linear predictor $\boxed{\widehat{Z}(x_0) = \beta + \sum_{i=1}^{n} \lambda_i Z(x_i)}$ that is unbiased and with minimal error,

without assuming that mean $m = \mathbb{E}[Z(x)]$ is known

- Unbiasedness gives

$$0 = \mathbb{E}[\widehat{Z}(x_0) - Z(x_0)] = \beta + \sum_{i=1}^{n} \lambda_i m - m = \beta + m(\sum_{i=1}^{n} \lambda_i - 1)$$

$\to$ True for any value of m if we **impose**:

$$\sum_{i=1}^{n} \lambda_i = 1 \quad \text{and} \quad \beta = 0$$

- Consequently, the error variance is given by:

$$\mathrm{Var}[\widehat{Z}(x_0) - Z(x_0)] = C(0) - 2\boldsymbol{\lambda}^T \boldsymbol{c}_0 + \boldsymbol{\lambda}^T \boldsymbol{C} \boldsymbol{\lambda}$$

WHAT IF THE MEAN IS UNKNOWN?

Goal: Find a linear predictor $\boxed{\widehat{Z}(x_0) = \beta + \sum_{i=1}^{n} \lambda_i Z(x_i)}$ that is unbiased and with minimal error,

without assuming that mean $m = \mathbb{E}[Z(x)]$ is known

- The resulting predictor is called the **ordinary kriging predictor**
- **Ordinary kriging predictor** $Z^{\mathrm{OK}}(x_0)$ of $Z(x_0)$ given $Z(x_1), \ldots, Z(x_n)$ = BLUP of $Z(x_0)$ when the mean m of Z is unknown:

$$\boxed{Z^{\mathrm{OK}}(x_0) = \sum_{i=1}^{n} \lambda_i^{\mathrm{OK}} Z(x_i)} = (\boldsymbol{\lambda}^{\mathrm{OK}})^T \boldsymbol{Z}$$

for some weights $\boldsymbol{\lambda}^{\mathrm{OK}} = (\lambda_1^{\mathrm{OK}}, \ldots, \lambda_n^{\mathrm{OK}})^T$ obtained as

$$\boldsymbol{\lambda}^{\mathrm{OK}} = \underset{\boldsymbol{\lambda} \in \mathbb{R}^n}{\arg\min} \left(C(0) - 2\boldsymbol{\lambda}^T \boldsymbol{c}_0 + \boldsymbol{\lambda}^T \boldsymbol{C} \boldsymbol{\lambda} \right) \text{ under the constraint } \boldsymbol{\lambda}^T \mathbf{1} = 1$$

ORDINARY KRIGING

- The weights $\boldsymbol{\lambda}^{\text{OK}}$ of the ordinary kriging predictor $Z^{\text{OK}}(x_0) = (\boldsymbol{\lambda}^{\text{OK}})^T \boldsymbol{Z}$ are defined as

$$\boldsymbol{\lambda}^{\text{OK}} = \underset{\boldsymbol{\lambda} \in \mathbb{R}^n}{\arg\min} \left(C(0) - 2\boldsymbol{\lambda}^T \boldsymbol{c}_0 + \boldsymbol{\lambda}^T \boldsymbol{C} \boldsymbol{\lambda} \right) \text{ under the constraint } \boldsymbol{\lambda}^T \mathbf{1} = 1$$

- The Lagrangian of the system writes

$$L(\boldsymbol{\lambda}, \mu) = C(0) - 2\boldsymbol{\lambda}^T \boldsymbol{c}_0 + \boldsymbol{\lambda}^T \boldsymbol{C} \boldsymbol{\lambda} + 2\mu(\boldsymbol{\lambda}^T \mathbf{1} - 1)$$

  And the KKT conditions then yield the following system satisfied by the solution of the constrained minimization problem

$$\begin{pmatrix} \boldsymbol{C} & \mathbf{1}_n \\ \mathbf{1}_n^T & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda}^{\text{OK}} \\ \mu^{\text{OK}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{c}_0 \\ 1 \end{pmatrix}, \quad \begin{cases} \boldsymbol{C} = [C(x_i - x_j)]_{1 \le i,j \le n} \\ \boldsymbol{c}_0 = [C(x_i - x_0)]_{1 \le i \le n} \end{cases}$$

ORDINARY KRIGING

- The weights $\boldsymbol{\lambda}^{\mathrm{OK}}$ of the ordinary kriging predictor $Z^{\mathrm{OK}}(x_0) = (\boldsymbol{\lambda}^{\mathrm{OK}})^T \boldsymbol{Z}$ are defined as

$$\boldsymbol{\lambda}^{\mathrm{OK}} = \underset{\boldsymbol{\lambda} \in \mathbb{R}^n}{\arg\min} \left(C(0) - 2\boldsymbol{\lambda}^T \boldsymbol{c}_0 + \boldsymbol{\lambda}^T \boldsymbol{C} \boldsymbol{\lambda} \right) \text{ under the constraint } \boldsymbol{\lambda}^T \mathbf{1} = 1$$

- The Lagrangian of the system writes

$$L(\boldsymbol{\lambda}, \mu) = C(0) - 2\boldsymbol{\lambda}^T \boldsymbol{c}_0 + \boldsymbol{\lambda}^T \boldsymbol{C} \boldsymbol{\lambda} + 2\mu(\boldsymbol{\lambda}^T \mathbf{1} - 1)$$

  And the KKT conditions then yield the following system satisfied by the solution of the constrained minimization problem

$$\begin{pmatrix} \boldsymbol{C} & \mathbf{1}_n \\ \mathbf{1}_n^T & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda}^{\mathrm{OK}} \\ \mu^{\mathrm{OK}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{c}_0 \\ 1 \end{pmatrix}, \quad \begin{cases} \boldsymbol{C} = [C(x_i - x_j)]_{1 \le i,j \le n} \\ \boldsymbol{c}_0 = [C(x_i - x_0)]_{1 \le i \le n} \end{cases}$$

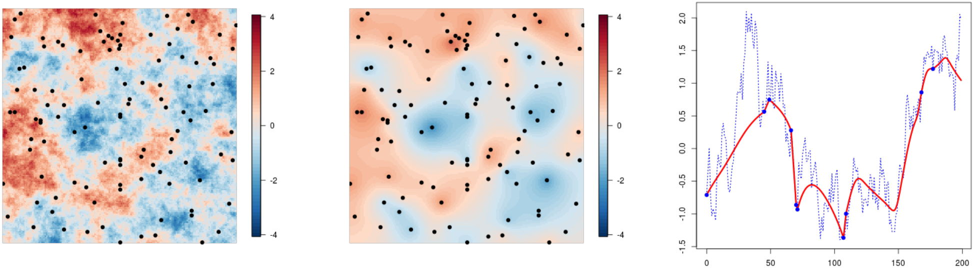
- The associated kriging variance is given by

$$\sigma^2_{\mathrm{OK}}(x_0) = C(0) - (\boldsymbol{\lambda}^{\mathrm{OK}})^T \boldsymbol{c}_0 - \mu^{\mathrm{OK}}$$

- Once again, the kriging weights variance do not depend on the locations of the observations

■ EXAMPLE

Ordinary kriging



Left: Variable z. Center : Estimation. Right : Estimation (Red) and True values (Blue) on a section.

ESTIMATING THE MEAN

- What if we want to estimate the mean m of Z? $\rightarrow$ Idea: Look again for the BLUP!
  - Linear: $\widehat{m} = \beta^{\mathrm{MK}} + \sum_{i=1}^{n} \lambda_i^{\mathrm{MK}} Z(x_i)$, for some weights $\beta^{\mathrm{MK}}, \{\lambda_i^{\mathrm{MK}}\}_{i=1}^{n}$
  - Unbiased: $\mathbb{E}[\widehat{m}] = m$
  - Best: Minimizes the error variance $\mathrm{Var}[\widehat{m} - m]$

■ ESTIMATING THE MEAN

- What if we want to estimate the mean m of Z? $\rightarrow$ Idea: Look again for the BLUP!
  - Linear: $\widehat{m} = \beta^{\mathrm{MK}} + \sum_{i=1}^{n} \lambda_i^{\mathrm{MK}} Z(x_i)$, for some weights $\beta^{\mathrm{MK}}, \{\lambda_i^{\mathrm{MK}}\}_{i=1}^{n}$
  - Unbiased: $\mathbb{E}[\widehat{m}] = m$
  - Best: Minimizes the error variance $\mathrm{Var}[\widehat{m} - m]$

- Computations similar to ordinary kriging yield $\widehat{m} = (\boldsymbol{\lambda}^{\mathrm{MK}})^T \boldsymbol{Z}$ where

$$\boldsymbol{\lambda}^{\mathrm{MK}} = \underset{\boldsymbol{\lambda} \in \mathbb{R}^n}{\arg\min}\, \boldsymbol{\lambda}^T \boldsymbol{C} \boldsymbol{\lambda} \text{ under the constraint } \boldsymbol{\lambda}^T \mathbf{1} = 1$$

is obtained by solving $\begin{pmatrix} \boldsymbol{C} & \mathbf{1}_n \\ \mathbf{1}_n^T & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda}^{\mathrm{MK}} \\ \mu^{\mathrm{MK}} \end{pmatrix} = \begin{pmatrix} \mathbf{0}_n \\ 1 \end{pmatrix}$ and the resulting kriging variance is $\sigma^2_{\mathrm{MK}}(x_0) = -\mu^{\mathrm{MK}}$

- Property: **Ordinary kriging = Simple kriging with mean estimated with BLUP**

PROPERTIES OF SK AND OK PREDICTORS

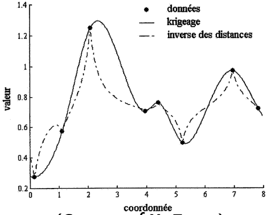
Let $Z^{\mathrm{K}}(x_0) = Z^{\mathrm{SK}}(x_0)$ or $Z^{\mathrm{OK}}(x_0)$ using observations $Z(x_1), \dots, Z(x_n)$ at **fixed** locations

- The kriging weights and variance **only depend on the locations of the data**, not their values
  $\rightarrow$ One of the limits of linear Geostatistics...

- The kriging predictors are **interpolators**, i.e. $\forall i \in \{1, \dots, n\}$, $Z^{\mathrm{K}}(x_i) = Z(x_i)$

  Eg. Simple kriging: If $x_0 = x_1$ then $\boldsymbol{c}_0 =$ First column of $\boldsymbol{C}$ and therefore $\boldsymbol{\lambda}^{\mathrm{SK}} = (1, 0, \dots, 0)$
  $\Rightarrow Z^{\mathrm{SK}}(x_1) = Z(x_1)$

- Kriging weights **can be negative** $\rightarrow$ Allows to predict outside the range of the data



(Courtesy of X. Emery)

LINEARITY OF KRIGING

Let $Z^{\mathrm{K}} = Z^{\mathrm{SK}}$ or Z^{OK} using observations $Z(x_1), \dots, Z(x_n)$ at **fixed** locations

- **Linearity**: Let $\mathcal{L}$ be a linear operator, then

$$[\mathcal{L}(x \mapsto Z(x))]^{\mathrm{K}} = \mathcal{L}(x \mapsto Z^{\mathrm{K}}(x))$$

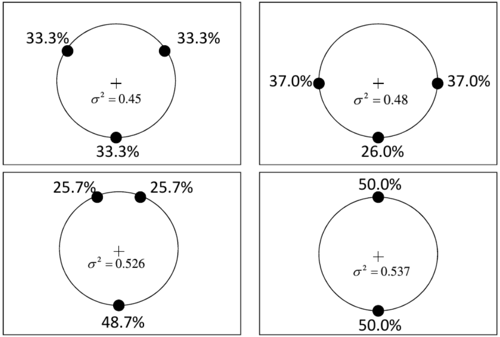
  → Kriging a quantity defined by applying a linear operator to $x \mapsto Z(x)$ is the same as applying $\mathcal{L}$ to the kriging predictor function $x \mapsto Z^{\mathrm{K}}(x)$

  → Consequence of linearity of expectation + Linearity of the equations used to compute kriging weights

- Examples of linear operators
  - $\mathcal{L}(Z) = \sum_{i=1}^{p} \alpha_i Z(y_i)$ for $\alpha_1, \dots, \alpha_p \in \mathbb{R} \Rightarrow (\mathcal{L}(Z))^{\mathrm{K}} = \sum_{i=1}^{p} \alpha_i Z^{\mathrm{K}}(y_i)$
  - $\mathcal{L}(Z) = |V|^{-1} \int_V Z(x) dx$ for $V \subset \mathcal{D}$ → Block kriging
  - $\mathcal{L}(Z) = \nabla Z(x_0)$ → Gradient estimation

PROPERTIES: DECLUSTERING EFFECT

Isotropic spherical with range > radius

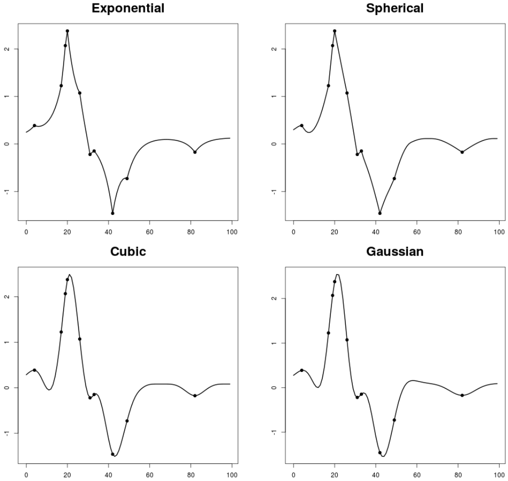


Kriging weights of the point + for different data locations

- The kriging predictor "recognizes" clusters: Cumulated weight of a cluster $\approx$ Weight of the cluster center

- Consequence of linear systems used to compute kriging weights $\rightarrow$ Incorporates information about data location

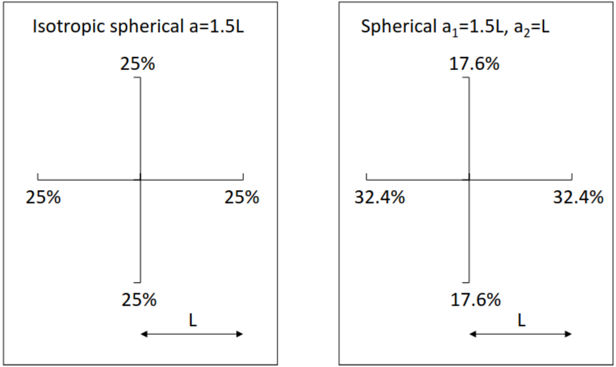
PROPERTIES: INFLUENCE OF THE COVARIANCE CHOICE



Kriging predictors for different model choices (same range and sill)

- Regularity of the covariance affects the regularity of the estimate

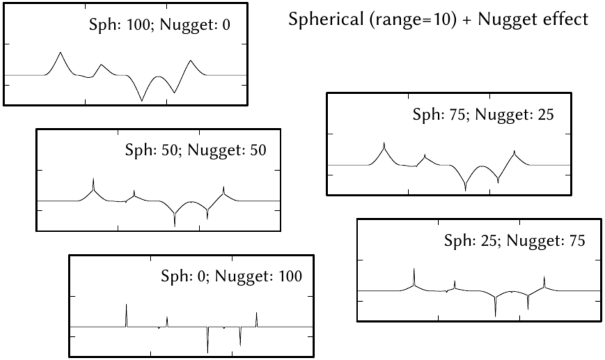
■ PROPERTIES: INFLUENCE OF ANISOTROPY



Kriging weights for an isotropic and an anisotropic model

The anisotropy distributions modifies the weight distribution

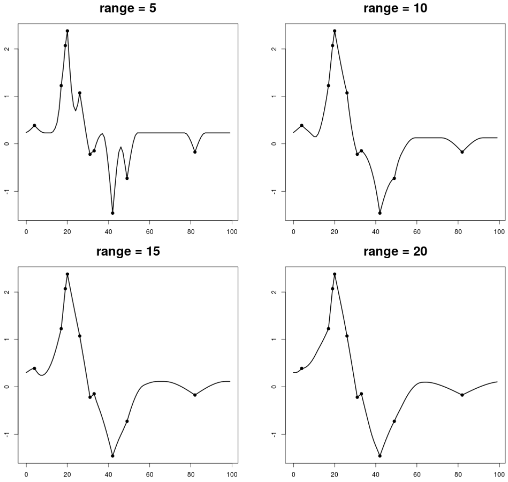
■ PROPERTIES: INFLUENCE OF NUGGET EFFECT



Kriging predictors for various models with varying nugget effect value

The bigger the nugget effect, the smaller the influence of the observation value

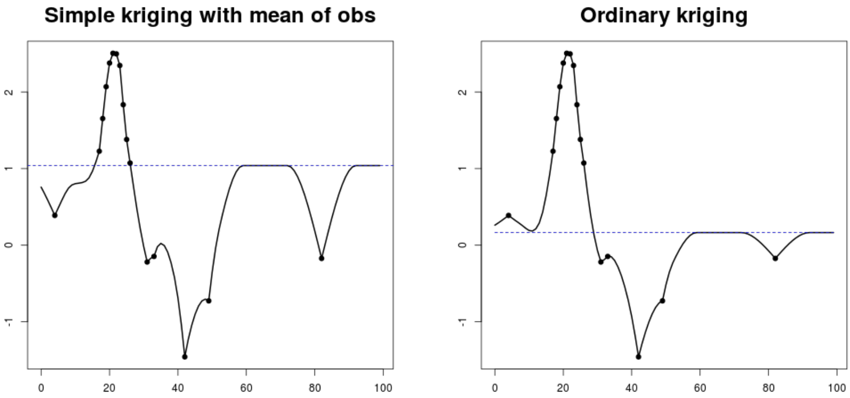
PROPERTIES: INFLUENCE OF THE CORRELATION RANGE



Kriging predictors for different covariance choices
(Spherical covariance, same sill, different ranges)

- The smallest the range, the smallest the influence of the data
- Away from the correlation range, the kriging predictor returns a constant value
  - For simple kriging: the specified mean m of Z
    $\rightarrow$ The mean m has an important role!
  - For ordinary kriging: the BLUP of the mean = Kriging estimate of the mean

■ EXAMPLE: ORDINARY KRIGING VS SIMPLE KRIGING



Away from the data, the ordinary kriging falls back to a declustered mean

OUTLINE

I. (Simple and Ordinary) Kriging

II. Extensions

III. Multivariate setting (from H. Wackernagel's lectures)

WORKING WITH MOVING NEIGHBORHOOD

- Up until now, we work with all the observations $\rightarrow$ Need to solve linear systems of size n $\rightarrow$ Potentially big
  - This is called the **unique neighborhood approach**

- Intuitively, not all points are "useful" for the estimation : Points very far away from the target x_0 won't contribute much

WORKING WITH MOVING NEIGHBORHOOD

- Up until now, we work with all the observations $\rightarrow$ Need to solve linear systems of size n $\rightarrow$ Potentially big
  - This is called the **unique neighborhood approach**

- Intuitively, not all points are "useful" for the estimation : Points very far away from the target x_0 won't contribute much

  Idea: Restrict the observations used in the kriging predictor to those "close" to the target

$\rightarrow$ Smaller system to solve but must be done case by case since the neighborhood moves with the considered target

$\rightarrow$ Size and shape of the neighborhood should depend on the covariance model (correlation range, anisotropy) and on available data

RELAXING THE STATIONARITY ASSUMPTION

- **If for any $x, y \in \mathcal{D}$,** $\text{Cov}(Z(x), Z(y))$ **is finite**, the simple/ordinary kriging predictors (aka BLUP with known/unkown mean) are computed using the same equations but
  - $\rightarrow$ Replace the stationary covariances $C(x-y)$ by $C(x,y) = \text{Cov}(Z(x), Z(y))$ in the kriging systems
  - $\rightarrow$ This implies that $(x,y) \mapsto C(x,y) = \text{Cov}(Z(x), Z(y))$ is known...

RELAXING THE STATIONARITY ASSUMPTION

- **If for any $x, y \in \mathcal{D}$, $\mathrm{Cov}(Z(x), Z(y))$ is finite**, the simple/ordinary kriging predictors (aka BLUP with known/unkown mean) are computed using the same equations but
  - $\rightarrow$ Replace the stationary covariances $C(x-y)$ by $C(x,y) = \mathrm{Cov}(Z(x), Z(y))$ in the kriging systems
  - $\rightarrow$ This implies that $(x,y) \mapsto C(x,y) = \mathrm{Cov}(Z(x), Z(y))$ is known...

- **In the intrinsic case**, i.e. mean-stationary + we can only compute variogram values

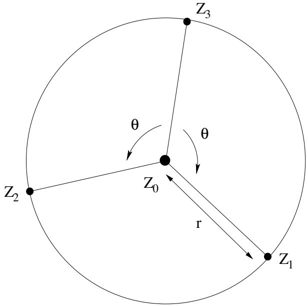
$$\gamma(h) = \frac{1}{2}\mathrm{Var}[Z(x+h) - Z(x)], \quad h \in \mathbb{R}^d$$

the BLUP (called **intrinsic kriging predictor**) is computed using the same equations as the ordinary kriging predictor but

$\rightarrow$ Replace the stationary covariances $C(x-y)$ by $-\gamma(x-y)$ in the kriging systems

■ EXERCISE

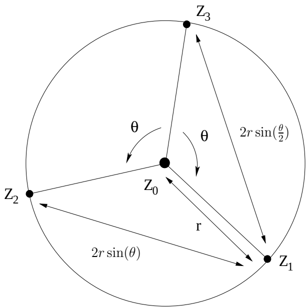
Consider the ordinary Kriging of Z_0 as a function of Z_1, Z_2, Z_3, with $\gamma(h) = |h|$



Solve the kriging system and compute the kriging weights when $\theta = \pi$, $2\pi/3$, $\pi/2$ and when θ goes to 0

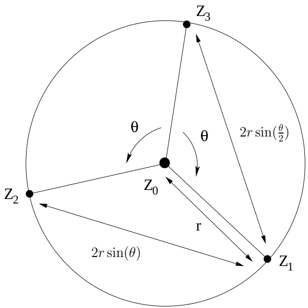
■ EXERCISE

Solution



EXERCISE

Solution



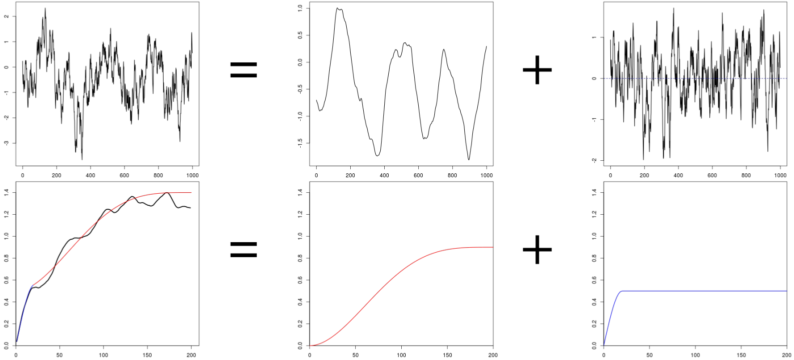
$$\lambda_1 = \lambda_2 = \frac{1}{4 - \cos(\theta/2)}, \; \lambda_3 = 1 - 2\lambda_1$$

weight / angle	π	$2\pi/3$	$\pi/2$	$\to 0$
λ_1	1/4	1/3	0.39	1/2
λ_2	1/4	1/3	0.39	1/2
λ_3	1/2	1/3	0.22	0

THE FILTERING PROBLEM

Motivation: We observe a field Z defined as a superposition of 2 unccorrelated fields Z_S and Z_N.

We seek to "extract" the component Z_S.



Observed field Z= Sum of uncorrelated fields Z_S and Z_N

$\Rightarrow$ **Covariance function of Z = Covariance function of Z_S + Covariance function of Z_N**

FACTORIAL KRIGING

Setting:

- We observe $Z = m + Z_S + Z_N$, $m \in \mathbb{R}$ and Z_S, Z_N are *zero-mean* uncorrelated random fields
  $\Rightarrow$ Covariance function: $C_Z = C_{Z_S} + C_{Z_N}$
- We want to predict the value of Z_S at $x_0 \in \mathcal{D}$ from observations of Z at $x_1, \dots, x_n \in \mathcal{D}$

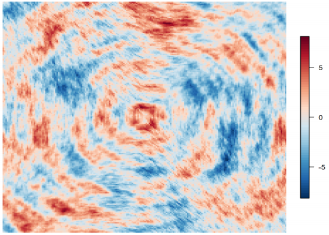
$\rightarrow$ **Factorial kriging predictor** = BLUP of $Z_S(x_0)$ given $\boldsymbol{Z} = (Z(x_1), \dots, Z(x_n))^T$:

$$Z_S^{\text{FK}}(x_0) = \sum_{i=1}^{n} \lambda_i^{\text{FK}} Z(x_i)$$

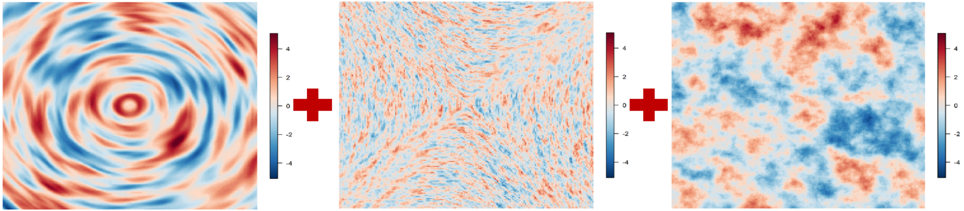
The same computations as before yield the system

$$\begin{pmatrix} \boldsymbol{C}^Z & \boldsymbol{1}_n \\ \boldsymbol{1}_n^T & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda}^{\text{FK}} \\ \mu^{\text{FK}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{C}^S + \boldsymbol{C}^N & \boldsymbol{1}_n \\ \boldsymbol{1}_n^T & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda}^{\text{FK}} \\ \mu^{\text{FK}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{c}_0^S \\ 0 \end{pmatrix}, \quad \begin{cases} \boldsymbol{C}^* = [C_*(x_i - x_j)]_{1 \le i,j \le n} \\ \boldsymbol{c}_0^* = [C_*(x_i - x_0)]_{1 \le i \le n} \end{cases}$$

EXAMPLE



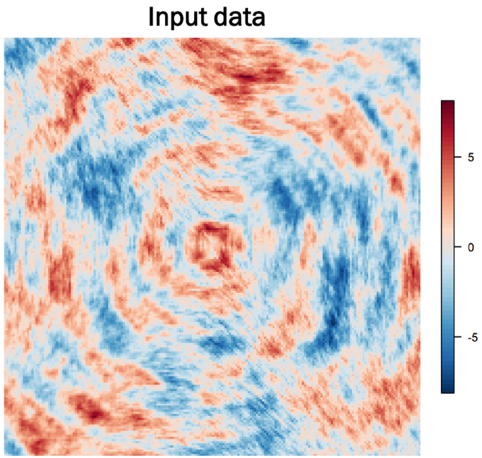
Noisy signal

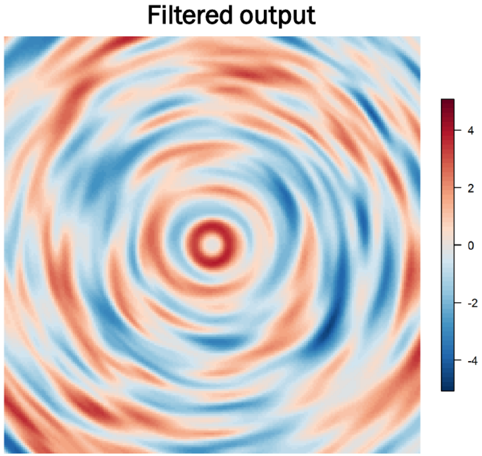


True signal

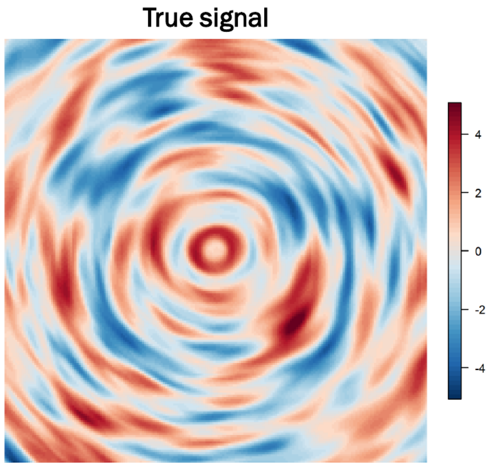
Noise components

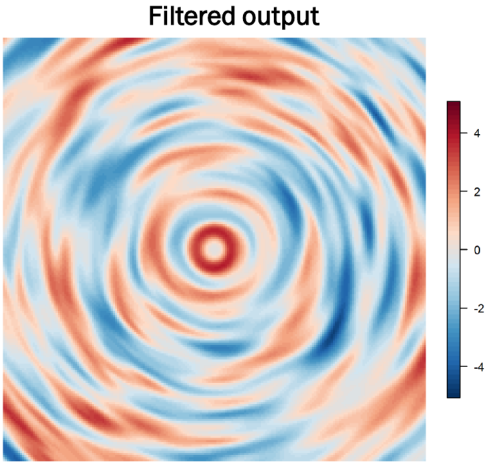
■ EXAMPLE





EXAMPLE





OUTLINE

I. (Simple and Ordinary) Kriging

II. Extensions

III. Multivariate setting (from H. Wackernagel's lectures)

Multivariate Geostatistics

Cross-covariance function and cross-variogram

TOPICS

1. Cross-covariance function
2. Cross variogram
3. Pseudo Cross-variogram

Measures of the joint spatial variation of a pair of variables

Cross-Covariance Function

$\mathbf{x}, \mathbf{x}' \in \mathcal{D}$	two points in a spatial region
$\mathbf{h} \;=\; \mathbf{x} - \mathbf{x}'$	separation vector between $\mathbf{x}$ and $\mathbf{x}'$
$Z_i(\mathbf{x}),\;\; i = 1, \ldots, N$	set of random functions

We assume second-order stationarity:

- $\mathrm{E}\big[Z_i(\mathbf{x})\big] = m_i$

  i.e. the means do not depend on $\mathbf{x}$.

- $\mathrm{E}\Big[\Big(Z_i(\mathbf{x}) - m_i\Big) \cdot \Big(Z_j(\mathbf{x}+\mathbf{h}) - m_j\Big)\Big] = C_{ij}(\mathbf{h})$

  i.e. the cross-covariance depends only on $\mathbf{h}$.

Asymmetry

In general:

$$C_{ij}(\mathbf{h}) \neq C_{ji}(\mathbf{h})$$

$$C_{ij}(\mathbf{h}) \neq C_{ij}(-\mathbf{h})$$

but,

$$C_{ij}(\mathbf{h}) = C_{ji}(-\mathbf{h})$$

Asymmetry

In general:

$$C_{ij}(\mathbf{h}) \neq C_{ji}(\mathbf{h})$$

$$C_{ij}(\mathbf{h}) \neq C_{ij}(-\mathbf{h})$$

but,

$$C_{ij}(\mathbf{h}) = C_{ji}(-\mathbf{h})$$

Asymmetry

In general:

$$C_{ij}(\mathbf{h}) \neq C_{ji}(\mathbf{h})$$

$$C_{ij}(\mathbf{h}) \neq C_{ij}(-\mathbf{h})$$

but,

$$C_{ij}(\mathbf{h}) = C_{ji}(-\mathbf{h})$$

Even and Odd Part

The cross-covariance function can be split into an even and an odd term:

$$\begin{aligned} C_{ij}(\mathbf{h}) \quad &= \quad \underbrace{\frac{1}{2}\Big(C_{ij}(\mathbf{h}) + C_{ij}(-\mathbf{h})\Big)}_{\text{even term}} \\ &\qquad + \underbrace{\frac{1}{2}\Big(C_{ij}(\mathbf{h}) - C_{ij}(-\mathbf{h})\Big)}_{\text{odd term}} \end{aligned}$$

Stationary Increments

We assume second-order stationary increments:

- $\mathrm{E}\Big[\, Z_i(\mathbf{x}+\mathbf{h}) - Z_i(\mathbf{x}) \,\Big] = 0$

  i.e. the expectation of increments
  does not depend on location $\mathbf{x}$.

- $\mathrm{cov}\Big[\, Z_i(\mathbf{x}+\mathbf{h}) - Z_i(\mathbf{x})\,,\; Z_j(\mathbf{x}+\mathbf{h}) - Z_j(\mathbf{x}) \,\Big]$

  $$= 2\,\gamma_{ij}(\mathbf{h})$$

  i.e. the cross-covariance of increments
  depends only on separation $\mathbf{h}$.

Cross-variogram

The cross-variogram:

$$\gamma_{ij}(\mathbf{h}) = \frac{1}{2}\,\mathrm{E}\Big[\Big(Z_i(\mathbf{x}+\mathbf{h}) - Z_i(\mathbf{x})\Big)\cdot\Big(Z_j(\mathbf{x}+\mathbf{h}) - Z_j(\mathbf{x})\Big)\Big]$$

is an even function.

It does not depend on:

- the sign of $\mathbf{h}$: $\gamma_{ij}(\mathbf{h}) = \gamma_{ij}(-\mathbf{h})$
- the ordering of the variables: $\gamma_{ij}(\mathbf{h}) = \gamma_{ji}(\mathbf{h})$

Cross-Variogram *vs* CCF

Assuming second-order stationarity:

$$\gamma_{ij}(\mathbf{h}) = C_{ij}(0) - \frac{1}{2}\left(C_{ij}(\mathbf{h}) + C_{ij}(-\mathbf{h})\right)$$

we see that the cross-variogram only retains the even term of the cross-covariance function (plus a constant).

Multivariate variogram modeling

Linear Model of Coregionalization

Spatial and multivariate representation of $Z_i(\mathbf{x})$ using uncorrelated factors $Y_u^p(\mathbf{x})$ with coefficients a_{ip}^u:

$$\boxed{Z_i(\mathbf{x}) = \sum_{u=0}^{S} \sum_{p=1}^{N} a_{ip}^u \, Y_u^p(\mathbf{x})}$$

Given u, all factors $Y_u^p(\mathbf{x})$ have the same variogram $g_u(\mathbf{h})$.

This implies a **multivariate nested variogram**:

$$\boldsymbol{\Gamma}(\mathbf{h}) = \sum_{u=0}^{S} \mathbf{B}_u \, g_u(\mathbf{h})$$

MULTIVARIATE LINEAR PREDICTIONS

Motivation: We observe two fields Y and Z. **We seek to estimate the value of Z at some locations using the information provided by both Y and Z**

- We seek to exploit both spatial correlation of Y and Z, **considered separately and jointly**
  $\rightarrow$ Tools: Covariance and Cross covariance functions (or variograms)

- Allows to complete the information provided by the observations of Z by extraction information from observations of Y also

Idea: Once again, look for a BLUP!

SIMPLE COKRIGING

Setting:

- We observe two zero-mean second-order stationary fields Y and Z at locations $x_1, \ldots, x_n \in \mathcal{D}$
  $\rightarrow$ Covariance functions C_Y, C_Z and Cross covariance functions C_{ZY}, C_{YZ} defined by

$$C_{ZY}(h) = \mathrm{Cov}(Z(x+h), Y(x)) = C_{YZ}(-h), \quad h \in \mathbb{R}^d$$

- We want to predict the value of Z at $x_0 \in \mathcal{D}$ from observations of Z and Y at $x_1, \ldots, x_n \in \mathcal{D}$

SIMPLE COKRIGING

Setting:

- We observe two zero-mean second-order stationary fields Y and Z at locations $x_1, \dots, x_n \in \mathcal{D}$
  $\rightarrow$ Covariance functions C_Y, C_Z and Cross covariance functions C_{ZY}, C_{YZ} defined by

$$C_{ZY}(h) = \mathrm{Cov}(Z(x+h), Y(x)) = C_{YZ}(-h), \quad h \in \mathbb{R}^d$$

- We want to predict the value of Z at $x_0 \in \mathcal{D}$ from observations of Z and Y at $x_1, \dots, x_n \in \mathcal{D}$

Cokriging predictor = BLUP of $Z(x_0)$ given $\boldsymbol{Z} = (Z(x_1), \dots, Z(x_n))^T$ and $\boldsymbol{Y} = (Y(x_1), \dots, Y(x_n))^T$:

$$Z^{\mathrm{CoK}}(x_0) = \sum_{i=1}^{n} \lambda_i^{\mathrm{CoK}} Z(x_i) + \sum_{i=1}^{n} w_i^{\mathrm{CoK}} Y(x_i)$$

The same computations as before yield the system

$$\begin{pmatrix} \boldsymbol{C}^Z & \boldsymbol{C}^{ZY} \\ \boldsymbol{C}^{YZ} & \boldsymbol{C}^Y \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda}^{\mathrm{CoK}} \\ \boldsymbol{w}^{\mathrm{CoK}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{c}_0^Z \\ \boldsymbol{c}_0^{ZY} \end{pmatrix}, \quad \begin{cases} \boldsymbol{C}^* = [C_*(x_i - x_j)]_{1 \le i,j \le n} \\ \boldsymbol{c}_0^* = [C_*(x_i - x_0)]_{1 \le i \le n} \end{cases}$$

and a cokrging variance $\sigma^2_{\mathrm{CoK}}(x_0) = \mathrm{C}^Z(0) - (\boldsymbol{\lambda}^{\mathrm{CoK}})^T \boldsymbol{c}_0^Z - (\boldsymbol{w}^{\mathrm{CoK}})^T \boldsymbol{c}_0^{ZY}$

REMARKS ON COKRIGING

- Same properties on spatial dependence and interpolation as kriging
- Takes cross correlations into account
- Not directly usable under intrinsic hypothesis
- Cokriging requires to infer the variagroms of two variables, and their cross variogram $\rightarrow$ Hard in practice
- Possible extensions to unknown mean case and data from different variables in different locations

REFERENCES AND FURTHER READING

- Emery, X. (2001). Géostatistique Linéaire. Notes de cours à l'École Nationale Supérieure des Mines de Paris.
- Chauvet, P. (2008). Aide-Mémoire de Géostatistique Linéaire. Presses des MINES.
- Wackernagel, H. (2003). Multivariate Geostatistics: an introduction with applications. Springer Science & Business Media.