# Geostatistics

## Kriging with drifts
## Likelihood-based approaches

Nicolas Desassis

freely adapted from Thomas Romary's course

nicolas.desassis@minesparis.psl.eu
Mines Paris
PSL University
Centre for geosciences and geoengineering

ATHENS Geostatistics 2024

MINES PARIS | PSL

# Reminder on second order stationary RF

## Definition: second order stationarity

A RF is second order stationary if its first two moments exist, are finite and are translantion invariant

- Expectation independent from $x$: $\mathbb{E}(Z(x)) = m$
- Covariance independent from $x$:

$$\mathsf{Cov}(Z(x), Z(x+h)) = C(h)$$

$\rightarrow$ $C$ is called the **covariance function**

- Consequently, the variance is independent from $x$

$$\mathbb{V}(Z(x)) = \mathsf{Cov}(Z(x), Z(x)) = C(0)$$

NB: we also talk about weak stationarity

# Reminder on second order RF

## Second order random function

A RF $Z$ is a second order RF if its first two moments exist and are finite

For a second order random function $Z$, $x, y \in \mathcal{X}$, we can define

- the expectation $m(x) = \mathbb{E}(Z(x))$
- the variance $\mathbb{V}(Z(x)) = \mathbb{E}((Z(x) - m(x))^2)$
- the covariance $\mathrm{Cov}(Z(x), Z(y)) = \mathbb{E}((Z(x) - m(x))(Z(y) - m(y)))$

# Reminder on second order RF

## Second order random function

A RF $Z$ is a second order RF if its first two moments exist and are finite

For a second order random function $Z$, $x, y \in \mathcal{X}$, we can define

- the expectation $m(x) = \mathbb{E}(Z(x))$
- the variance $\mathbb{V}(Z(x)) = \mathbb{E}((Z(x) - m(x))^2)$
- the covariance $\mathsf{Cov}(Z(x), Z(y)) = \mathbb{E}((Z(x) - m(x))(Z(y) - m(y)))$

# Several approaches to deal with non-stationarity

- relax the stationary assumption on the expectation
  $\Longrightarrow$ Universal Kriging or kriging with external drifts
- relax the stationary assumption on the covariance
  $\Longrightarrow$ non-stationary covariance functions

# Universal kriging

The model is given by

$$Z(x) = m(x) + Y(x)$$

where the *drift* $m(x)$ admits a functional representation

$$m(x) = \sum_{l=0}^{L} \beta_l f_l(x)$$

$Y(x)$ is centered and generally $f_0 = 1$

The universal kriging system can be obtained with the standard recipe

# Universal kriging

Predictor $Z^*(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) = \Lambda^t Z$

Unbiasedness condition

$$\begin{aligned} &\mathbb{E}(Z^*(x_0)) = \mathbb{E}(Z(x_0)) \\ \iff &\sum_{l=0}^{L} \sum_{i=1}^{n} \lambda_i \beta_l f_l(x_i) = \sum_{l=0}^{L} \beta_l f_l(x_0) \\ \iff &\sum_{l=0}^{L} \beta_l \left( \sum_{i=1}^{n} \lambda_i f_l(x_i) - f_l(x_0) \right) = 0 \end{aligned}$$

whatever the values of the unknown $\beta_l$. Hence the *universality conditions*

$$f_l(x_0) = \sum_{i=1}^{n} \lambda_i f_l(x_i), \ \forall \ l \in \{0, \dots, L\}$$

# Universal kriging

Predictor $Z^*(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) = \Lambda^t Z$

Unbiasedness condition

$$\begin{aligned}
&\mathbb{E}(Z^*(x_0)) = \mathbb{E}(Z(x_0)) \\
\iff &\sum_{l=0}^{L} \sum_{i=1}^{n} \lambda_i \beta_l f_l(x_i) = \sum_{l=0}^{L} \beta_l f_l(x_0) \\
\iff &\sum_{l=0}^{L} \beta_l \left( \sum_{i=1}^{n} \lambda_i f_l(x_i) - f_l(x_0) \right) = 0
\end{aligned}$$

whatever the values of the unknown $\beta_l$. Hence the *universality conditions*

$$f_l(x_0) = \sum_{i=1}^{n} \lambda_i f_l(x_i), \ \forall\ l \in \{0, \ldots, L\}$$

# Universal kriging

Prediction variance $\sigma_E^2 = \mathbb{V}(Z^*(x_0) - Z(x_0))$

The universality conditions being met $\sigma_E^2$ do not depend on $m(x)$ anymore, therefore

$$\sigma_E^2 = \mathrm{C}(0) - 2\sum_{i=1}^{n} \lambda_i C_{i0} + \sum_{i,j} \lambda_i \lambda_j C_{ij}$$

We minimise

$$Q = \sigma_E^2 + 2\sum_{l=0}^{L} \mu_l \left( \sum_{i=1}^{n} \lambda_i f_l(x_i) - f_l(x_0) \right)$$

# Universal kriging

Prediction variance $\sigma_E^2 = \mathbb{V}(Z^*(x_0) - Z(x_0))$

The universality conditions being met $\sigma_E^2$ do not depend on $m(x)$ anymore, therefore

$$\sigma_E^2 = \mathrm{C}(0) - 2\sum_{i=1}^{n} \lambda_i C_{i0} + \sum_{i,j} \lambda_i \lambda_j C_{ij}$$

We minimise

$$Q = \sigma_E^2 + 2\sum_{l=0}^{L} \mu_l \left( \sum_{i=1}^{n} \lambda_i f_l(x_i) - f_l(x_0) \right)$$

# Universal kriging

The universal kriging system writes

$$\begin{cases} \displaystyle\sum_{j=1}^{n} \lambda_j \mathrm{C}_{ij} + \sum_{l=0}^{L} \mu_l f_l(x_i) = \mathrm{C}_{i0} & \forall\ i \in \{1, \dots, n\} \\ \displaystyle\sum_{i=1}^{n} \lambda_i f_l(x_i) = f_l(x_0) & \forall\ l \in \{0, \dots, L\} \end{cases}$$

and the universal kriging variance is

$$\sigma_{UK}^2 = \mathrm{C}(0) - \sum_{i=1}^{n} \lambda_i \mathrm{C}_{i0} - \sum_{l=1}^{L} \mu_l f_l(x_0)$$

# Universal kriging

## Matrix representation

$$f_0 = \begin{pmatrix} f_{00} \\ \vdots \\ f_{0L} \end{pmatrix}, \ \mathrm{F} = \begin{pmatrix} f_{10} & \cdots & f_{1L} \\ \vdots & \ddots & \vdots \\ f_{n0} & \cdots & f_{nL} \end{pmatrix}, \ \mu = \begin{pmatrix} \mu_0 \\ \vdots \\ \mu_L \end{pmatrix}$$

and

$$Z^*(x_0) = \Lambda^t \mathrm{z}$$

The kriging system and variance writes

$$\begin{pmatrix} C & \mathrm{F} \\ \mathrm{F}^t & 0 \end{pmatrix} \begin{pmatrix} \Lambda \\ \mu \end{pmatrix} = \begin{pmatrix} C_0 \\ f_0 \end{pmatrix} \text{ et } \sigma^2_{UK} = \mathrm{C}(0) - \Lambda^t \mathrm{C}_0 - \mu^t f_0$$

# Universal kriging
## Solution

We solve the universal kriging system through the block matrix inversion formula

### Block matrix inversion

Noting $S = (D - CA^{-1}B)$ (Schur complement of block A, assumed inversible), we have

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}BS^{-1}CA^{-1} & -A^{-1}BS^{-1} \\ -S^{-1}CA^{-1} & S^{-1} \end{pmatrix}$$

Hence

$$\Lambda = C^{-1}C_0 + C^{-1}\mathrm{F}(\mathrm{F}^t C^{-1}\mathrm{F})^{-1}(\mathrm{F}^t C^{-1}C_0 - f_0)$$
$$\mu = (\mathrm{F}^t C^{-1}\mathrm{F})^{-1}(f_0 - \mathrm{F}^t C^{-1}C_0)$$

# Exercise

Propose an estimator of the drift coefficients $\beta_l$ and give the variance of the estimator.

We consider linear estimator of the form

$$\beta_l^\star = \sum_{i=1}^{n} \lambda_i^{(l)} Z(x_i) = \Lambda_l^t Z$$

The unbiasedness condition writes

$$\sum_{i=1}^{n} \lambda_i^{(l)} \left( \sum_{k=0}^{L} \beta_k f_k(x_i) \right) = \beta_l$$

$$\sum_{k=0}^{L} \beta_k \left( \sum_{i=1}^{n} \lambda_i^{(l)} f_k(x_i) - \delta_{lk} \right) = 0$$

And we deduce the $L$ unbiasedness conditions:

$$\forall l \in \{0, \ldots, L\}, \qquad \sum_{i=1}^{n} \lambda_i^{(l)} f_k(x_i) = \delta_{lk}$$

# Exercise

Propose an estimator of the drift coefficients $\beta_l$ and give the variance of the estimator.

We consider linear estimator of the form

$$\beta_l^\star = \sum_{i=1}^{n} \lambda_i^{(l)} Z(x_i) = \Lambda_l^t Z$$

The unbiasedness condition writes

$$\sum_{i=1}^{n} \lambda_i^{(l)} \left( \sum_{k=0}^{L} \beta_k f_k(x_i) \right) = \beta_l$$

$$\sum_{k=0}^{L} \beta_k \left( \sum_{i=1}^{n} \lambda_i^{(l)} f_k(x_i) - \delta_{lk} \right) = 0$$

And we deduce the $L$ unbiasedness conditions:

$$\forall l \in \{0, \ldots, L\}, \qquad \sum_{i=1}^{n} \lambda_i^{(l)} f_k(x_i) = \delta_{lk}$$

The kriging system for the $l^{\mathrm{th}}$ coefficient is:

$$\begin{pmatrix} C & F \\ F^t & 0 \end{pmatrix} \begin{pmatrix} \Lambda_l \\ \mu \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ e_l \end{pmatrix}$$

where the $k^{\mathrm{th}}$ term of $e_l$ is $\delta_{lk}$

## Block matrix inversion

Noting $S = (D - CA^{-1}B)$ (Schur complement of block A, assumed inversible), we have

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}BS^{-1}CA^{-1} & -A^{-1}BS^{-1} \\ -S^{-1}CA^{-1} & S^{-1} \end{pmatrix}$$

$$\beta_l^\star = \Lambda_l^t Z = e_l^t \left(\mathrm{F}^t C^{-1} \mathrm{F}\right)^{-1} \mathrm{F}^t C^{-1} Z$$

Since the LHS are the same for all the coefficients, we can obtain a formula for the vector

$$\beta^\star = \left(\mathrm{F}^t C^{-1} \mathrm{F}\right)^{-1} \mathrm{F}^t C^{-1} Z$$

$$\mathbb{V}(\beta^\star) = \left(\mathrm{F}^t C^{-1} \mathrm{F}\right)^{-1}$$

The kriging system for the $l^{\mathrm{th}}$ coefficient is:

$$\begin{pmatrix} C & F \\ F^t & 0 \end{pmatrix} \begin{pmatrix} \Lambda_l \\ \mu \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ e_l \end{pmatrix}$$

where the $k^{\mathrm{th}}$ term of $e_l$ is $\delta_{lk}$

## Block matrix inversion

Noting $S = (D - CA^{-1}B)$ (Schur complement of block A, assumed inversible), we have

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}BS^{-1}CA^{-1} & -A^{-1}BS^{-1} \\ -S^{-1}CA^{-1} & S^{-1} \end{pmatrix}$$

$$\beta_l^\star = \Lambda_l^t Z = e_l^t \left(\mathrm{F}^t C^{-1} \mathrm{F}\right)^{-1} \mathrm{F}^t C^{-1} Z$$

Since the LHS are the same for all the coefficients, we can obtain a formula for the vector

$$\beta^\star = \left(\mathrm{F}^t C^{-1} \mathrm{F}\right)^{-1} \mathrm{F}^t C^{-1} Z$$

$$\mathbb{V}(\beta^\star) = \left(\mathrm{F}^t C^{-1} \mathrm{F}\right)^{-1}$$

The kriging system for the $l^{\text{th}}$ coefficient is:

$$\begin{pmatrix} C & F \\ F^t & 0 \end{pmatrix} \begin{pmatrix} \Lambda_l \\ \mu \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ e_l \end{pmatrix}$$

where the $k^{\text{th}}$ term of $e_l$ is $\delta_{lk}$

### Block matrix inversion

Noting $S = (D - CA^{-1}B)$ (Schur complement of block A, assumed inversible), we have

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}BS^{-1}CA^{-1} & -A^{-1}BS^{-1} \\ -S^{-1}CA^{-1} & S^{-1} \end{pmatrix}$$

$$\beta_l^\star = \Lambda_l^t Z = e_l^t \left(\mathrm{F}^t C^{-1} \mathrm{F}\right)^{-1} \mathrm{F}^t C^{-1} Z$$

Since the LHS are the same for all the coefficients, we can obtain a formula for the vector

$$\beta^\star = \left(\mathrm{F}^t C^{-1} \mathrm{F}\right)^{-1} \mathrm{F}^t C^{-1} Z$$

$$\mathbb{V}(\beta^\star) = \left(\mathrm{F}^t C^{-1} \mathrm{F}\right)^{-1}$$

The kriging system for the $l^{\text{th}}$ coefficient is:

$$\begin{pmatrix} C & F \\ F^t & 0 \end{pmatrix} \begin{pmatrix} \Lambda_l \\ \mu \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ e_l \end{pmatrix}$$

where the $k^{\text{th}}$ term of $e_l$ is $\delta_{lk}$

## Block matrix inversion

Noting $S = (D - CA^{-1}B)$ (Schur complement of block A, assumed inversible), we have

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}BS^{-1}CA^{-1} & -A^{-1}BS^{-1} \\ -S^{-1}CA^{-1} & S^{-1} \end{pmatrix}$$

$$\beta_l^\star = \Lambda_l^t Z = e_l^t \left(\mathrm{F}^t C^{-1} \mathrm{F}\right)^{-1} \mathrm{F}^t C^{-1} Z$$

Since the LHS are the same for all the coefficients, we can obtain a formula for the vector

$$\beta^\star = \left(\mathrm{F}^t C^{-1} \mathrm{F}\right)^{-1} \mathrm{F}^t C^{-1} Z$$

$$\mathbb{V}(\beta^\star) = \left(\mathrm{F}^t C^{-1} \mathrm{F}\right)^{-1}$$

# Universal kriging

## Additivity relationship

We can decompose the UK predictor into

$$Z^*(x_0) = m^\star(x_0) + R^\star(x_0)$$

where

$$m^\star(x_0) = \sum_{l=0}^{L} \beta_l^\star f_l(x_0)$$

is the drift estimated at the target location
and $R^\star(x_0)$ is the simple kriging (with mean 0) of the residuals

$$R(x_i) = Z(x_i) - m^\star(x_i)$$

This property is also valid for the variance

$$\sigma_{UK}^2 = \mathbb{V}(m^\star(x_0)) + \sigma_{SK}^2$$

# Universal kriging

## Additivity relationship

We can decompose the UK predictor into

$$Z^{*}(x_0) = m^{\star}(x_0) + R^{\star}(x_0)$$

where

$$m^{\star}(x_0) = \sum_{l=0}^{L} \beta_l^{\star} f_l(x_0)$$

is the drift estimated at the target location
and $R^{\star}(x_0)$ is the simple kriging (with mean 0) of the residuals

$$R(x_i) = Z(x_i) - m^{\star}(x_i)$$

This property is also valid for the variance

$$\sigma^2_{UK} = \mathbb{V}(m^{\star}(x_0)) + \sigma^2_{SK}$$

# Variogram fitting in the presence of a drift

1. Compute the OLS estimator of $\beta$

$$\widehat{\beta}_{OLS} = (F^t F)^{-1} F^t \mathrm{Z}$$

2. get the estimated residuals

$$\widehat{y}_i = (\mathrm{Z} - F\widehat{\beta}_{OLS})_i$$

3. fit the variogram of $\widehat{y}_i$
4. Compute the GLS estimator of $\beta$

$$\widehat{\beta}_{GLS} = (F^t C_{\widehat{\theta}}^{-1} F)^{-1} F^t C_{\widehat{\theta}}^{-1} \mathrm{Z}$$

5. get the estimated residuals

$$\widehat{y}_i = (\mathrm{Z} - F\widehat{\beta}_{GLS})_i$$

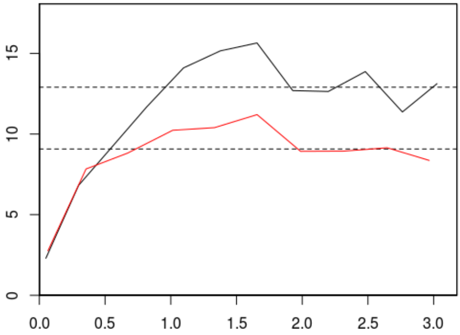
6. Back to 3 or stop

# Drawbacks

- Inherent drawbacks from the empirical variogram approach
- The variogram of the estimated residuals is biased (even when computed with the true parameters value!)

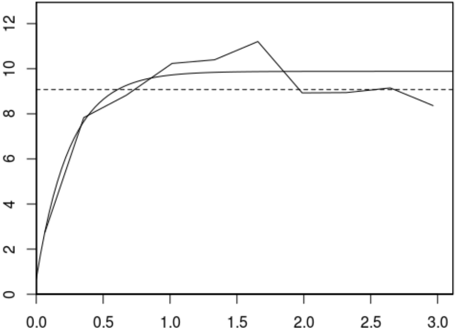
$$\mathbb{E}\left[\frac{1}{2}(Y^*(x_i) - Y^*(x_j))^2\right] = \gamma(x_i - x_j) - \frac{1}{2}(F_i - F_j)^t(F^tC_\theta^{-1}F)^{-1}(F_i - F_j)$$

# Universal Kriging of the Cobalt concentration: residuals variogram



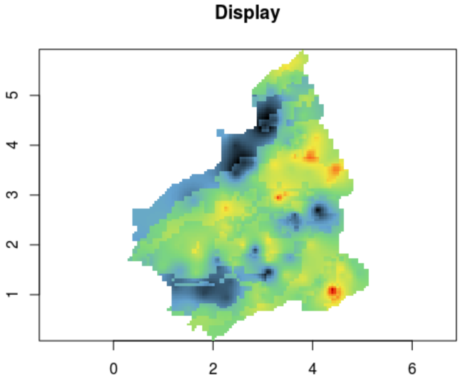

auxiliary variable : rocktype

# Universal Kriging of the Cobalt concentration: variography



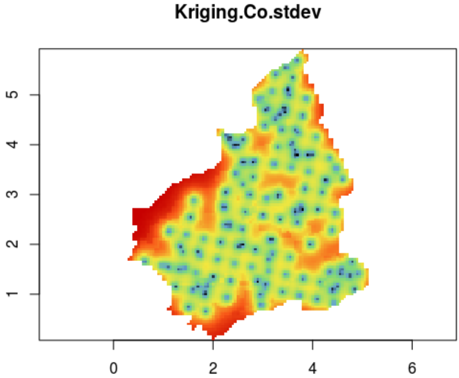

structure 0.67nugget + 9.2exponential(0.75)

# Universal Kriging of the Cobalt concentration




RMSE on the validation set: 3.89

# Standard error of the Universal Kriging of the Cobalt concentration




RMSE on the validation set: 3.89

# Principle of Maximum likelihood estimation

Let $(z_1, \ldots, z_n)$ be an i.i.d. sample from a parametric pdf $f_\theta$, $\theta \in \Theta$

The likelihood of $(z_1, \ldots, z_n)$ is

$$l(z_1, \ldots, z_n; \theta) = \prod_{i=1}^{n} f_\theta(z_i)$$

Generally more convenient to work with the log-likelihood

$$L(z_1, \ldots, z_n; \theta) = \log l(z_1, \ldots, z_n; \theta) = \sum_{i=1}^{n} \log(f_\theta(z_i))$$

The Maximum likelihood estimator (MLE) of $\theta$ is

$$\widehat{\theta} = \operatorname*{argmax}_{\theta \in \Theta} l(z_1, \ldots, z_n; \theta) = \operatorname*{argmax}_{\theta \in \Theta} L(z_1, \ldots, z_n; \theta)$$

# Exercise

Let $(z_1, \ldots, z_n)$ be an i.i.d. sample from $\mathcal{N}(\mu, \sigma^2)$

1. compute the log-likelihood
2. compute the MLE of $\mu$
3. compute the MLE of $\sigma^2$

# Some (asymptotic) properties of MLE

- Consistency (i.e. $\widehat{\theta} \xrightarrow{\mathbb{P}} \theta$)
- Asymptotic normality
- Efficiency (achieves the Cramer-Rao bound)

# The Gaussian spatial model

We consider the model

$$Z(x) = F(x)^t \beta + Y(x),\ x \in \mathcal{X} \subset \mathbb{R}^d$$

where

- $F(x) \in \mathbb{R}^p$ is a vector of covariates
- $\beta \in \mathbb{R}^p$ is a vector of unknown parameters
- $Y(x)$ is a zero-mean Gaussian process with covariance

$$\text{Cov}(Y(x_1), Y(x_2)) = \mathrm{C}(x_1, x_2; \theta)$$

where $\theta \in \Theta$ is a set of parameters

In the isotropic stationary case, generally, $\theta = (\sigma, \kappa, a)$

$$\mathrm{C}(x_1, x_2; \theta) = \sigma^2 \mathrm{R}_\kappa \left( \frac{\|x_1 - x_2\|}{a} \right)$$

with $\mathrm{R}$ a correlation function

# The Gaussian spatial model

$Z$ is sampled at $(x_1, \ldots, x_n) \in \mathcal{X}^{n \times d}$

Let $\mathrm{Z} = (Z(x_1), \ldots, Z(x_n)) \in \mathbb{R}^n$

We have

$$\mathrm{Z} \sim \mathcal{N}\left(F\beta, \sigma^2 C(\kappa, a)\right)$$

where

- $F_{ik} = F_k(x_i)$, $1 \leq k \leq p$, $1 \leq i \leq n$
- $C_{ij}(\kappa, a) = \mathrm{R}_\kappa\left(\frac{\|x_i - x_j\|}{a}\right)$, $1 \leq i, j \leq n$

# Classical approach

1. compute the OLS estimate of $\beta$

$$\widehat{\beta}_{OLS} = (F^t F)^{-1} F^t \mathrm{Z}$$

2. compute the estimated residuals

$$\widehat{y}_i = (\mathrm{Z} - F\widehat{\beta}_{OLS})_i$$

3. fit the variogram of $\widehat{y}_i$
4. compute the GLS estimate of $\beta$

$$\widehat{\beta}_{GLS} = (F^t C_{\widehat{\theta}}^{-1} F)^{-1} F^t C_{\widehat{\theta}}^{-1} \mathrm{Z}$$

5. compute the estimated residuals

$$\widehat{y}_i = (\mathrm{Z} - F\widehat{\beta}_{GLS})_i$$

6. go back to step 3

# Maximum likelihood estimation

Under the Gaussian hypothesis, the likelihood of $\mathrm{Z}$ is

$$l(\beta, \theta; \mathrm{Z}) = (2\pi)^{-n/2} |C_\theta|^{-1/2} \exp\left( - \frac{\|\mathrm{Z} - F\beta\|^2_{C_\theta^{-1}}}{2} \right)$$

where $\|A\|^2_B = A^t B A$

MLE of $\beta$ and $\theta$ are defined as

$$(\widehat{\beta}_{ML}, \widehat{\theta}_{ML}) = \underset{\beta \in \mathbb{R}^p, \theta \in \Theta}{\operatorname{argmax}}\ l(\beta, \theta; \mathrm{Z}) = \underset{\beta \in \mathbb{R}^p, \theta \in \Theta}{\operatorname{argmax}}\ L(\beta, \theta; \mathrm{Z})$$

where

$$\begin{aligned} L(\beta, \theta; \mathrm{Z}) &= \log(l(\beta, \theta; \mathrm{Z})) \\ &= -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |C_\theta| - \frac{\|\mathrm{Z} - F\beta\|^2_{C_\theta^{-1}}}{2} \end{aligned}$$

# Profile log-likelihood

It is easy to see (exercise) that for fixed $\theta$

$$\widehat{\beta}_{ML} = (F^t C_\theta^{-1} F)^{-1} F^t C_\theta^{-1} \mathrm{Z}$$

Then, plugging $\widehat{\beta}_{ML}$ into $L$

$$L(\theta; \mathrm{Z}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |C_\theta| - \frac{\|\mathrm{Z}\|^2_{P(\theta)}}{2}$$

where

$$P(\theta) = C_\theta^{-1} - C_\theta^{-1} F (F^t C_\theta^{-1} F)^{-1} F^t C_\theta^{-1}$$

# Optimizing the log-likelihood

- Non-linear, constrained (over $\Theta$) optimization problem
- Grid search
- Gradient descent algorithms

$$\theta^{(k+1)} = \theta^{(k)} + M^{(k)} g^{(k)}$$

  where
  - $M^{(k)}$ is a $m \times m$ matrix
  - $g^{(k)} = \frac{\partial L(\theta; \mathrm{Z})}{\partial \theta}|_{\theta=\theta^{(k)}}$
- 2 examples
  - Newton-Raphson: $M^{(k)} = (B^{(k)})^{-1}$, $B_{ij}^{(k)} = -\frac{\partial^2 L(\theta; \mathrm{Z})}{\partial \theta_i \partial \theta_j}|_{\theta=\theta^{(k)}}$
  - Fisher scoring: $M^{(k)} = (B^{(k)})^{-1}$, $B_{ij}^{(k)} = \mathbb{E}\left[-\frac{\partial^2 L(\theta; \mathrm{Z})}{\partial \theta_i \partial \theta_j}|_{\theta=\theta^{(k)}}\right]$

# Some asymptotic results

Fisher information matrix

$$\mathcal{I}(\theta) = \mathbb{E}_\theta \left( \frac{\partial}{\partial \theta} L(\beta, \theta) \left( \frac{\partial}{\partial \theta} L(\beta, \theta) \right)^t \right)$$

More precisely

$$\mathcal{I}_{j,k}(\theta) = \frac{1}{2} \mathrm{Tr} \left( \mathrm{C}_\theta^{-1} \mathrm{C}_\theta^j \mathrm{C}_\theta^{-1} \mathrm{C}_\theta^k \right)$$

where $\mathrm{C}_\theta^j = \frac{\partial \mathrm{C}_\theta}{\partial \theta_j}$

provides an estimate of the inverse covariance matrix of the estimated covariance parameters (under particular assumptions)

# Computational issues and approximations

When $n$ is large, the evaluation of $L$ becomes computationally demanding

$$L(\beta, \theta; \mathrm{Z}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log |C_\theta| - \frac{\|\mathrm{Z} - F\beta\|^2_{C_\theta^{-1}}}{2}$$

storage scales with $O(n^2)$

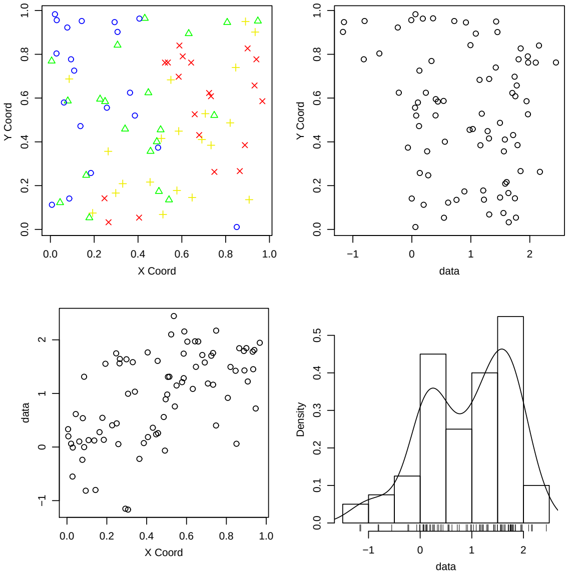
computation time scales with $O(n^3)$

# Computational issues and approximations

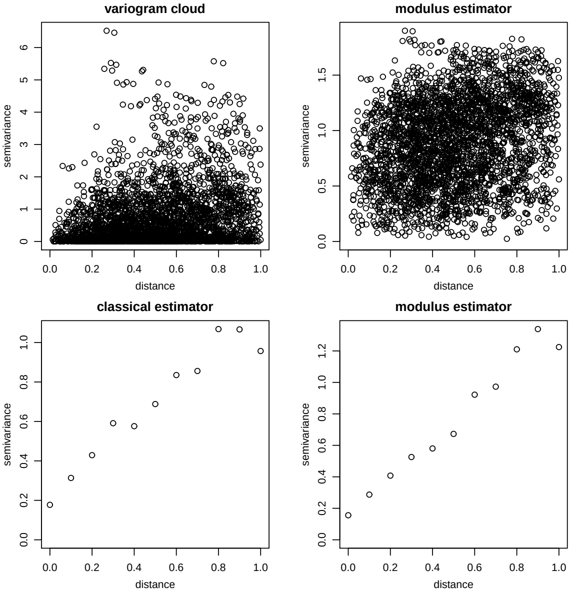
Several solutions

- profile with respect to the scale parameter
- assume/enforce sparsity of the covariance matrix (small range, compact support)
- approximate likelihood
- composite likelihood

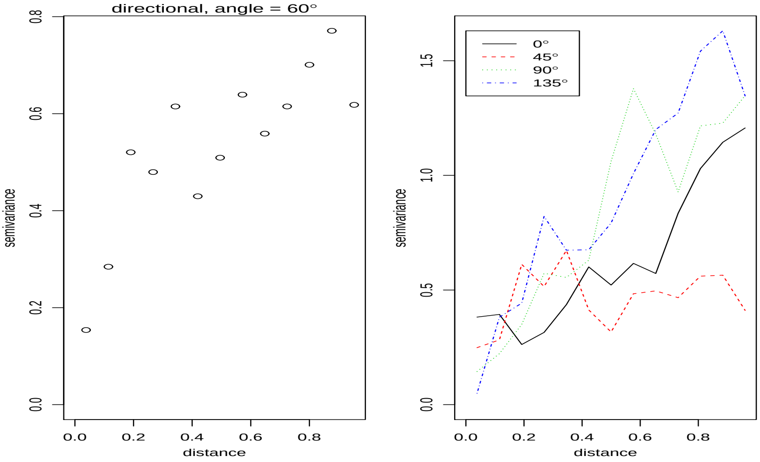
# Example – dataset



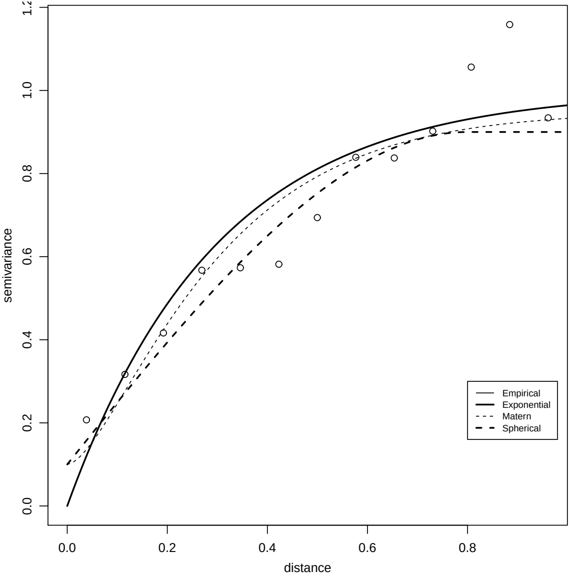
# Example – variogram cloud



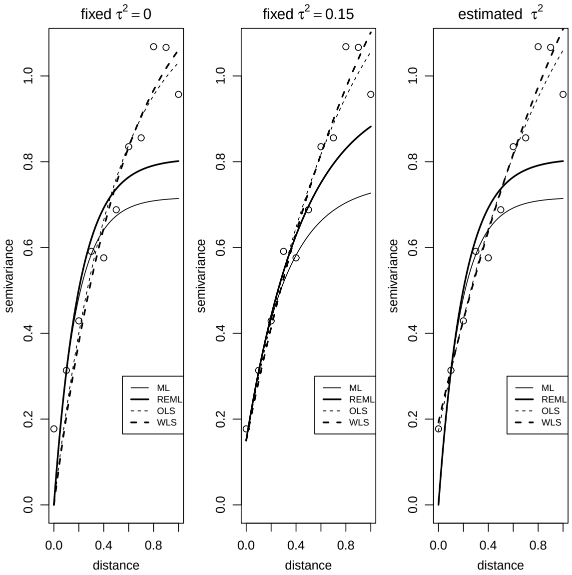
# Example – several directions



# Example – variogram fitting (isotropic)



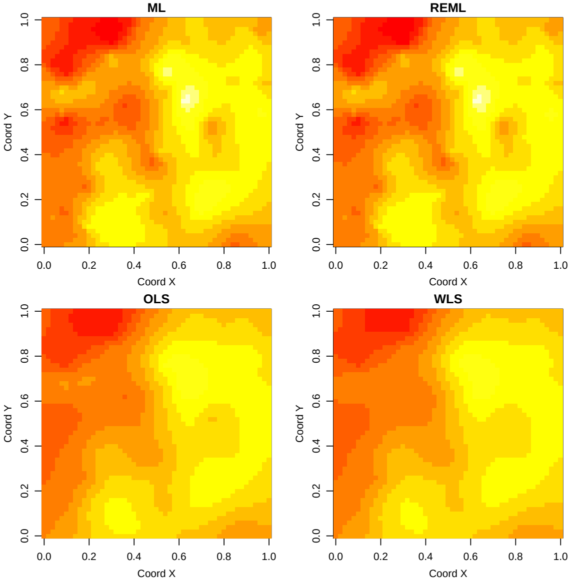
# Example – Comparison



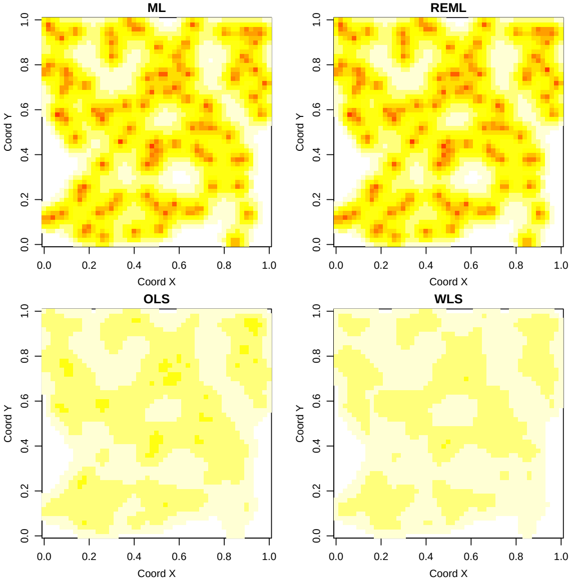
# Example – Summary

| | **ML** | **REML** | **OLS** | **WLS** | **True** |
|---|---|---|---|---|---|
| $\beta$ | 0.78 | 0.75 | | | 0 |
| $\tau^2$ | 0 | 0 | 0.16 | 0.19 | 0 |
| $\sigma^2$ | 0.75 | 0.85 | 0.99 | 1.27 | 1 |
| **range** | 0.55 | 0.63 | 1.60 | 2.49 | 0.9 |

# Example – Prediction



# Example – Prediction standard deviation



# Example – Prediction MSE

We have separated the sample into a learning dataset of size 80 and a validation dataset of size 20
The prediction MSE is computed at the validation location

| **ML** | **REML** | **OLS** | **WLS** | **True** |
|---|---|---|---|---|
| 0.23 | 0.25 | 0.47 | 0.54 | 0.28 |

# Pros and cons for the MLE

Pros

- No tedious parameterization for the fitting procedure
- Statistical properties of the estimators

Cons

- Gaussian assumption
- Computationally intensive

# Beyond Gaussian

## Transformed Gaussian random field

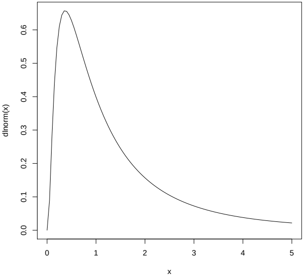
- Striclty monotonic transforms, $Z(x) = \phi(Y(x))$, $Y$ is a GRF

- Lognormal model

$$\log(Z(x)) = \mu + \sigma Y(x), \; x \in \mathcal{X}$$

- Box - Cox transform

$$\frac{Z(x)^\lambda - 1}{\lambda} = Y(x)$$




- Truncated Gaussian process $Z(x) = \phi(Y(x))\mathbb{1}_{Y(x)>v}$

- Latent gaussian model
  example logit for (0,1) data

$$\mathbb{P}(Z = 0) = \frac{1}{1 + \exp\left(-F(x)^t\beta - Y(x)\right)}$$

# References


- Matheron, G. (1970). La théorie des variables régionalisées, et ses applications: Fascicule 5. Les Cahiers du Centre de Morphologie Mathématique, Ecole des Mines de Paris.
- Mardia, K. V., & Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. Biometrika, 71(1), 135-146.
- Kitanidis, P. K., & Lane, R. W. (1985). Maximum likelihood parameter estimation of hydrologic spatial processes by the Gauss-Newton method. Journal of Hydrology, 79(1), 53-71.
- Stein, M. L. (2012). Interpolation of spatial data: some theory for kriging. Springer Science & Business Media.
- Gelfand, A. E., Diggle, P., Guttorp, P., & Fuentes, M. (Eds.). (2010). Handbook of spatial statistics. CRC press.