Spatial statistics

Lucia CLAROTTO

MIA Paris-Saclay, AgroParisTech
lucia.clarotto@agroparistech.fr

M2 MSV, February 2026

AgroParisTech
Talents d'une planète soutenable

Overview of topics

- Introduction to spatial statistics
    - Spatial data : geostatistical data, lattice data and point processes
- Spatial dependency in the continuous case (geostatistics)
    - Covariance and variogram functions
    - Variogram fitting
- Parametric prediction of geostatistical data : Kriging

Introduction

Spatial Statistics : provides statistical methods for processing data relating to phenomena observed on a spatial set S

$\hookrightarrow$ Geographic information is attached to each element of S.

Spatial data : any observation for which both the observed value and the observation location are available.

⚠ Locations are an important piece of information to consider in statistical analysis

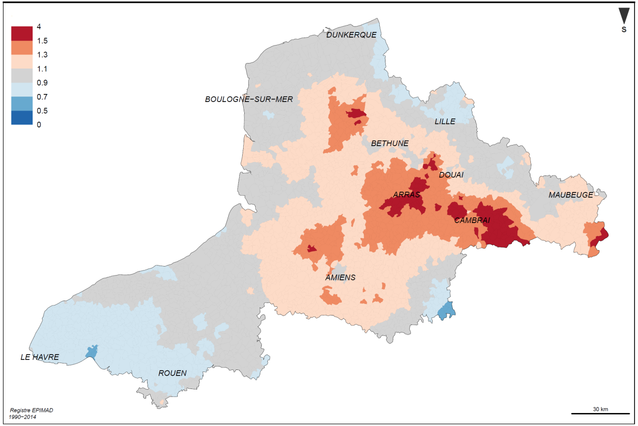
Spatial dependence : the first law of geography (Tobler, 1979) "*Everything interacts with everything, but two objects close together are more likely to do so than two objects far apart*"

Domains of applications

Domains of application

Epidemiology : Disease mapping, cluster detection, etc.

Example : The smoothed standardised incidence ratio of Crohn's disease in 2014 in the North of France adjusted for age and sex.

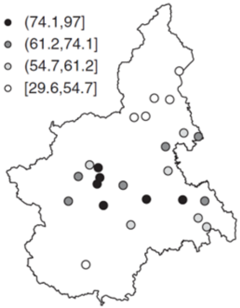


Domains of application

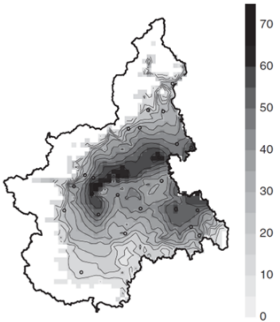
Environment : provide forecast maps of air quality, concentrations of heavy metals, etc.

Example : Prediction of daily PM10 concentration throughout the Piedmont region. Identification of locations where PM10 concentrations are likely to exceed the standard.

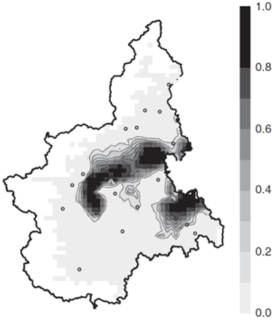
PM10 at 24 stations



Map of PM10



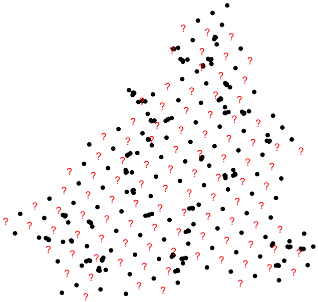
$Prob(PM10 > 50\mu g/m^3)$



Source : Blangiardo, Marta, and Michela Cameletti. Spatial and spatio-temporal Bayesian models with R-INLA. John Wiley and Sons, 2015.

Domains of application

Agriculture : Determine contaminated soils in order to avoid farming in some areas.
Example : Jura Swiss Data



- region of 14,5 km^2
- concentration on Cadmium,...
- 259 sites used as observations
- 100 sites to predict

Data collected by Ecole Polytechnique Fédérale de Lausanne (Switzerland).

Source : Goovaerts, P. Geostatistics for natural resources evaluation. Applied Geostatistics, 1997

Domains of application

Domains of application :

- Geology/Mining : predict the potential quantity of oil at a location given a sample taken at certain points spread over an area to optimise drilling locations.
- Geomarketing : predict customer flows from a given geographical area to a given shop.
- Hydrology : define catchment areas, study precipitation measured at weather stations.
- Economy : Unemployment rates, property price modelling, etc.
- ...

Types of spatial data

Spatial data

Mathematically, a random phenomenon observed on a spatial ensemble S is modelled by a spatial process $\{Z_s \in \mathbb{R}^d\ ,\ s \in S\}$.

Definition

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $S \subset \mathbb{R}^N$, $(N > 1)$. A spatial process on S with values in $\mathbb{R}^p$ $(p \geq 1)$ is a function of two variables, denoted $Z(s, \omega)$, such that for $s \in S$, $Z(s, \omega)$ is a random variable on $(\Omega, \mathcal{A}, \mathbb{P})$ and for each $\omega \in \Omega$, $Z(s, \omega)$ is a function of $S \to \mathbb{R}^p$, denoted by $\{Z_s, s \in S\}$.

⚠ The S elements (the locations of the observation sites) can be fixed or random.

There are three types of spatial data :[1]

- Continuous (geostatistical) data
- Lattice data
- Point data

1. Cressie, Noel A.C. (1993b). Statistics for spatial data. Wiley series in probability and statistics. Wiley-Interscience, New York 15, p. 105–209.

Continuous data (Geostatistics)

Definition

- S is a continuous subset of $\mathbb{R}^d$.
- $\{Z_s, s \in S\}$ is continuously observable in S.
- $\{Z_s, s \in S\}$ is measured only at n sites $\{s_1, \dots, s_n\}$ of S.

Objective of spatial analysis of continuous data (geostatistics)

- Prediction of $\{Z_s, s \in S\}$ at a location where it has not been measured
- Reconstruction of $\{Z_s, s \in S\}$ everywhere on S.

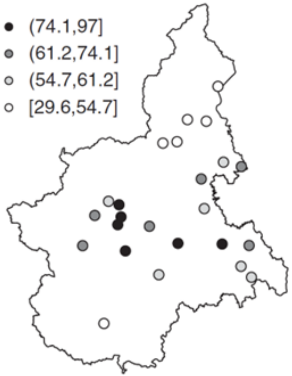
Examples of continuous data

- Mining industry : chemical composition of soil
- Pollution study : water or air quality
- Weather forecast : precipitation, temperature, etc.

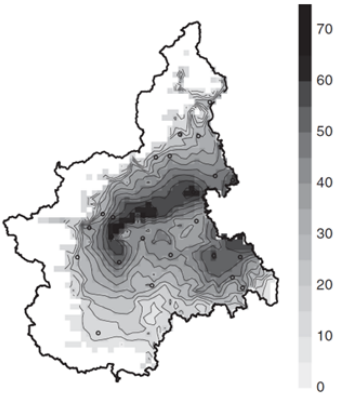
Continuous data : example

- Prediction of PM10 concentration in the Piedmont region.
- Identification of locations where PM10 concentrations are likely to exceed the standard.

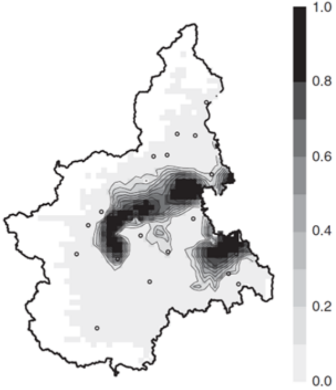
PM10 at 24 stations



Map of PM10



$Prob(PM10 > 50\mu g/m^3)$



Source : Blangiardo, Marta, and Michela Cameletti. Spatial and spatio-temporal Bayesian models with R-INLA. John

Lattice data

Definition

- S is a deterministic fixed discrete set of $\mathbb{R}^d$.
- S elements (generally) represent geographical units : stations or administrative areas (towns, departments, etc.).
- $\{Z_s, s \in S\}$ is observed (generally) everywhere on S.

Objective of spatial analysis of lattice data

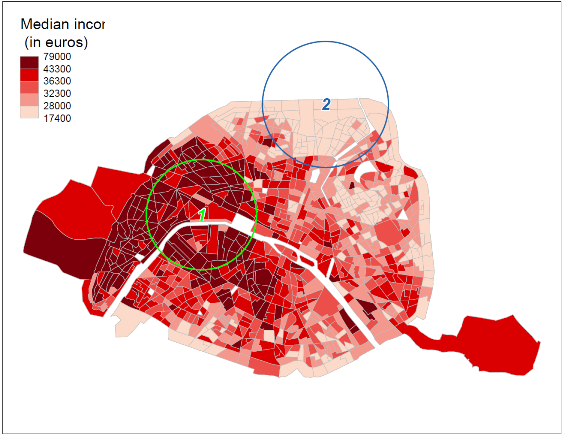
- Characterization of the spatial structure of the spatial process $\{Z_s, s \in S\}$.
- Construction and analysis of explanatory models.

Examples of lattice data

- Economics : modelling prices
- Health : modelling number of patients by municipality

Lattice data : example

- Describe the portion of median household income related to spatial dependence.
- Identify clusters of rich/poor people (spatial zone) and adjust the cluster detection to spatial dependence.



Point data

Definition

- S is a continuous subset of $\mathbb{R}^d$.
- $\mathbf{z} = \{z_1, \ldots, z_n\}$, $z_i \in S$, the set of locations where the observations took place.
- $\mathbf{z}$ and the number of observations $n = n(\mathbf{z})$ are random.
- The process Z is said *spatial point process* (SPP) and $\mathbf{z}$ a realization of this SPP.

Objective of spatial analysis of point data

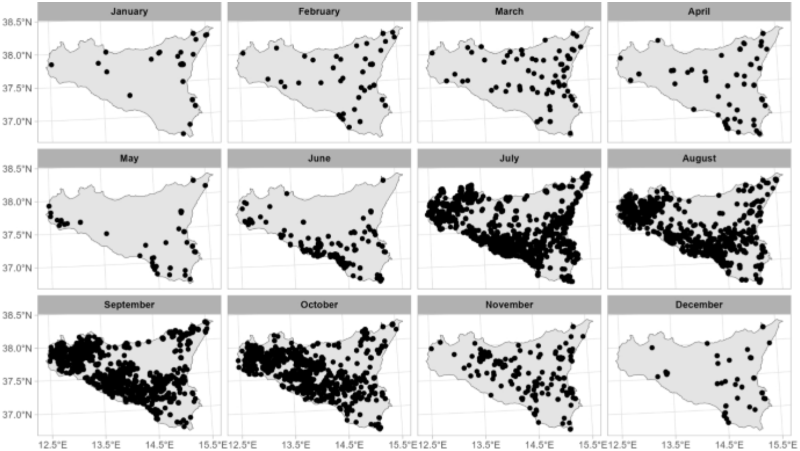
- Measure of the difference between the spatial distribution of observations and a completely random distribution in space. How are observations grouped together in space ?

Examples of point data

- Ecology : distribution of plant species
- Epidemiology : individual patient positions

Point data : example

- Point data of fire occurrences in Sicily.
- Are there zones where fires are most probable ? Do fires occur close to other fires ?



Geostatistics

Geostatistics

Geostatistics

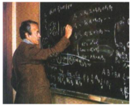
- A set of statistical methods for studying phenomena that can be observed in space (continuously) and that present a form of continuity (spatial dependence)
- The central objective is to provide forecast maps of a phenomenon over an area of interest based on a finite number of observations
- It is useful in a number of fields : geology, agronomy, environment, epidemiology, etc.

Geostatistics : history

'50 Initially developed by Danie Gerhardus Krige (1919-2013) in South Africa for the purpose of estimating deposits exploited in mines.

Danie Krige learning the trade, 1939.

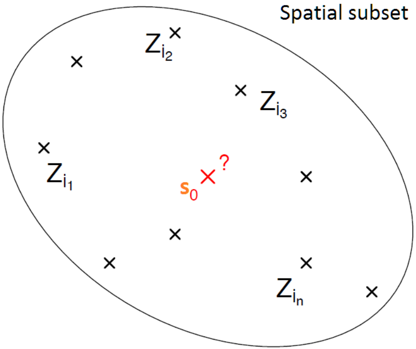
'60 Mathematical Foundation by G. Matheron (1930-2000) and his colleagues at Mines Paris.



- Firstly mining and petroleum applications.
- Nowadays almost all data (especially in environmental science) are spatial or spatio-temporal.

Context

- S is a continuous subset of $\mathbb{R}^d$
- The spatial process $\{Z_s, s \in S\}$ (or $\{Z(s), s \in S\}$) is continuously observable in S.
- $\{Z_s, s \in S\}$ is only observed in n sites $\{s_1, \ldots, s_n\}$ of S.
- We want to predict the process $\{Z_s, s \in S\}$ at any site $s_0 \in S$.



Geostatistics notation

- $S \subset R^d$ is the **spatial domain** contained within :
  - line ($d = 1$),
  - plane ($d = 2$),
  - cube ($d = 3$),
  - hypercube ($d > 3$)
- s is generally a **vector** because there are usually 2 or 3 elements defining location : latitude/longitude, northing/easting, etc.
- We assume that the outcome variable

$$\mathbf{Z} = \left(Z(s_1), \ldots, Z(s_n)\right)^{\mathsf{T}}$$

has a **Multivariate Normal distribution** :

$$\mathbf{Z} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

- $\mathbf{Z}$ is sometimes called a **Gaussian process**.

Multivariate Normal Structure

$$\mathbf{Z}_{n\times 1} \sim \mathcal{N}_n\left(\boldsymbol{\mu}_{n\times 1}, \boldsymbol{\Sigma}_{n\times n}\right)$$

- Mean vector :

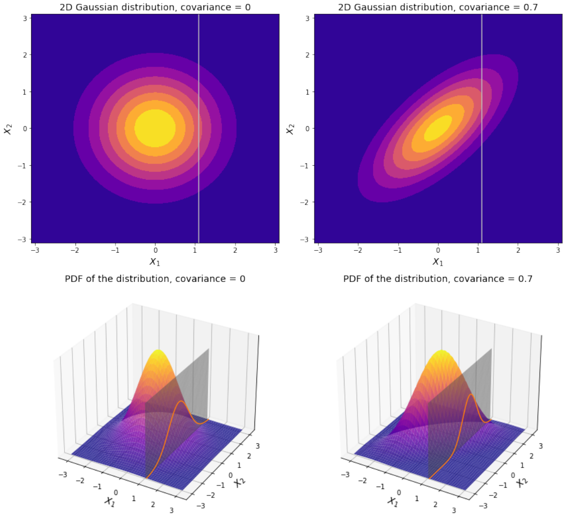
$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$$

- Covariance matrix :

$$\boldsymbol{\Sigma} = \sigma^2 \mathbf{R}$$

- $\mathbf{R}$ is an $n \times n$ **correlation matrix**
- Properties :
  - symmetric
  - positive definite
  - entries defined by spatial correlation (depending on distance)

Bivariate Gaussian distribution



Single realization

- Spatial statistics routinely works with a **single realization** of an n-dimensional multivariate Gaussian field.
  $\hookrightarrow$ Makes estimation hard...
- Luckily, nearby observations are likely similar :

$$\text{spatial correlation} \;\Rightarrow\; \text{structure imposed on data}$$

Spatial variability

Covariance

Spatial data are not independent as in classical statistics

- Similar (dissimilar) data at similar (dissimilar) locations
- Can we quantify similarity ?
  $\hookrightarrow$ Similarity in data is measured by (positive) **correlation** or **covariance**
  $\hookrightarrow$ Similarity in locaton is measured by a **distance metric**

Covariance function : allows to take into account the relationships between a set of pairs of points.

$$c(s_1, s_2) = \mathrm{Cov}(Z(s_1), Z(s_2)) \qquad \forall s_1,\ s_2 \in S$$

Spatial stationarity

Idea : the phenomenon is sufficiently homogeneous in space such that we can replace multiple realizations of the field by a repetition in space

Definition : Strict stationarity

The spatial process $\{Z(s), s \in S\}$ is said to be strictly stationary if $\forall k \in \mathbb{N}^*$, $\forall s_1, \ldots, s_k \in S$ and $\forall h \in S$ the law of $(Z(s_1 + h), \ldots, Z(s_k + h))$ does not depend on h. The process $\{Z(s), s \in S\}$ is said to be strictly stationary if the joint distribution of $Z(s_i)$ is the same as that of $Z(s_i + h)$, h indicating a translation with respect to the initial position s_i.

⚠ Strict stationarity is not operational and is very restrictive.

Spatial stationarity

Idea : the phenomenon is sufficiently homogeneous in space such that we can replace multiple realizations of the field by a repetition in space

Definition : Strict stationarity

The spatial process $\{Z(s), s \in S\}$ is said to be strictly stationary if $\forall k \in \mathbb{N}^*$, $\forall s_1, \ldots, s_k \in S$ and $\forall h \in S$ the law of $(Z(s_1 + h), \ldots, Z(s_k + h))$ does not depend on h. The process $\{Z(s), s \in S\}$ is said to be strictly stationary if the joint distribution of $Z(s_i)$ is the same as that of $Z(s_i + h)$, h indicating a translation with respect to the initial position s_i.

⚠ Strict stationarity is not operational and is very restrictive.

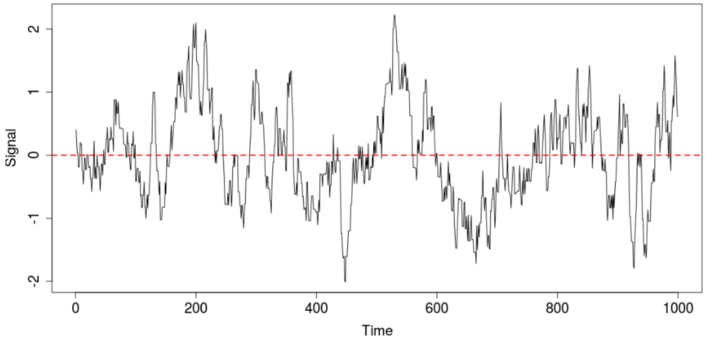
Definition : Second-order stationarity (weak)

The spatial process $\{Z(s), s \in S\}$ is said to be second-order stationary on S if the mean of Z is constant and if the covariance function of Z is invariant by translation :

$$\forall s_1, s_2 \in S \qquad \mathbb{E}(Z(s_1)) = \mathbb{E}(Z(s_2)) = m \in \mathbb{R} \text{ and } \mathrm{Cov}(Z(s_1), Z(s_2)) = C(s_1 - s_2) = C(h)$$

Then also the variance is constant : $\mathrm{Var}(Z(s_1)) = \mathrm{Var}(Z(s_2)) = C(0)$

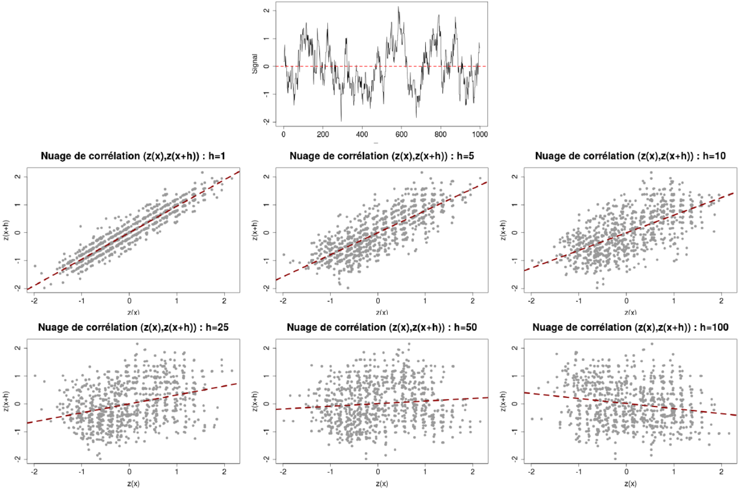
Illustration of a second order stationary 1D spatial process



One realization of a second order stationary random field (mean in red)

- fluctuations around a constant value in space
- the dispersion of the fluctuations are similar everywhere

Correlation clouds

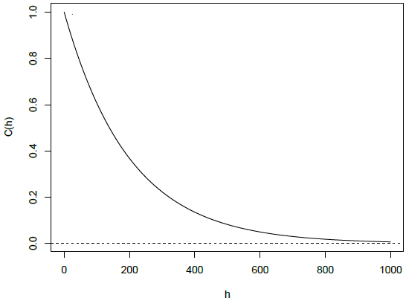


Given a lag h, we represent the scatter plot $(z(s), z(s+h))$ for every $s \in S$ such that $s + h \in S$

Isotropic covariance function

Definition : Isotropy

Let Z be a second-order stationary process. We say that Z is isotropic if, and only if, its covariance depends only on the (Euclidean) distance separating the sites : $C(h) = C_0(\|h\|) \quad \forall h = s_1 - s_2 \in S$



⚠ The covariance function expresses how the covariance between two observations evolves as their distance (in the isotropic case) increases.

Covariance : properties

Properties

If Z is a second-order stationary process, then :

1. $C(h) = C(-h), \qquad \forall h \in S$ (symmetry)
2. $|C(h)| \leq C(0) = \mathrm{Var}(Z), \qquad \forall h \in S$ (boundedness)
3. $C(\cdot)$ is positive semidefinite
   $\forall n \geq 1, \quad s_1, \ldots, s_n \in S \quad$ and $\quad \alpha_1, \ldots, \alpha_n \in \mathbb{R}$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \, C(s_i - s_j) \geq 0$$

Intrinsic stationarity and variogram

⚠ In practice, second-order stationarity is often too strong :

- the average may change over the area of interest
- the variance may not be bounded when this area grows

Definition : Intrinsic stationarity

The spatial process $\{Z(s), s \in S\}$ is an intrinsically stationary process, or is an intrinsic process, if for any $h \in S$, the process $\Delta Z^{(h)} = \{\Delta Z^{(h)}(s) = Z(s+h) - Z(s),\ s \in S\}$ is second-order stationary :

$$\mathbb{E}(Z(s+h) - Z(s)) = m \in \mathbb{R}$$

$$\mathrm{Var}(Z(s+h) - Z(s)) = \mathbb{E}((Z(s+h) - Z(s))^2) = 2\gamma(h)$$

The function $\gamma : \ S \to \mathbb{R}^+$ defined by

$$\gamma(h) = \frac{1}{2}\mathrm{Var}\left(Z(s+h) - Z(s)\right)$$

is called *semi-variogram* (or sometimes *variogram*)

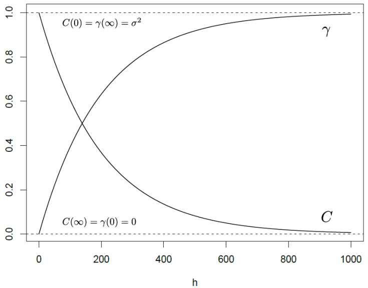
Relationship covariance/variogram

The covariance function can be used to analyse spatial variability provided that the assumption of second-order stationarity is satisfied.

Theorem : Fundamental covariance/variogram relationship

For a second-order stationary process Z, for any distance h, the variogram and the covariance are related by the relation

$$\gamma(h) = C(0) - C(h), \qquad \forall h \in S$$



Variogram properties

Properties

Let Z be an intrinsic process. Then

1. $\gamma(h) = \gamma(-h), \qquad \forall h \in S$
2. $\gamma(h) \geq 0, \qquad \forall h \in S$
3. $\gamma(0) = 0$
4. $\gamma(\cdot)$ is conditionally negative definite
   $\forall n \geq 1, \quad s_1, \ldots, s_n \in S \quad$ and $\quad \alpha_1, \ldots, \alpha_n \in \mathbb{R}$ such that $\sum_{i=1}^n \alpha_i = 0$, then

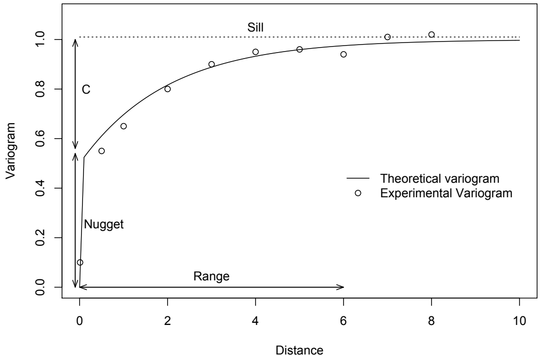
$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \gamma(s_i - s_j) \leq 0$$

General appearance of variogram : sill

Variogram sill (palier) : the limit of the variogram to infinity (when it exists)

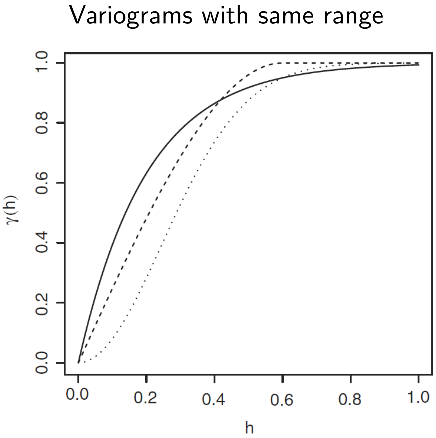
$$C = \lim_{||h|| \to \infty} \gamma(h)$$

- If the sill is finite, the sill equals its variance.



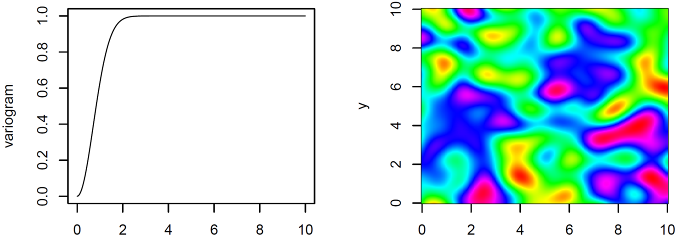
General appearance of the variogram : range

Variogram range (portée) : the distance h from which the variogram reaches its plateau, i.e., the distance beyond which observations are no longer spatially correlated.

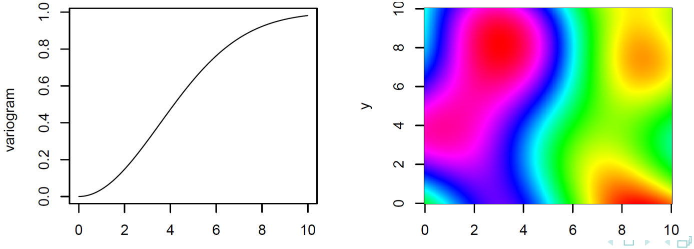


General appearance of variogram : effect of range

Range equal to 1



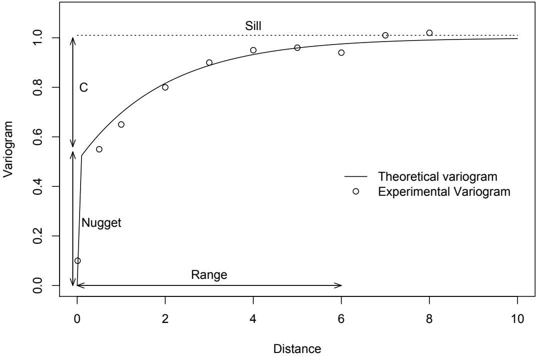
Range equal to 8



General appearance of variogram : nugget

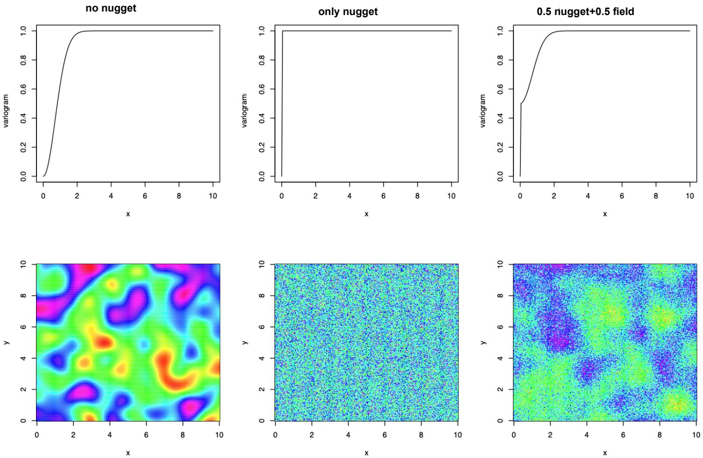
Nugget (pépite) : the limit of the variogram to 0 (possibly zero)

$$\tau^2 = \lim_{||h|| \to 0^+} \gamma(h)$$



General appearance of the variogram : effect of nugget

A nugget represents noise in the observation process.
↪ The statistical error of the measurement instrument.



Anisotropy

Anisotropy is the opposite of isotropy. It happens when the variogram function depends on the direction.

Definition

- *Isotropy* : the variogram between $Z_(s_1)$ and $Z(s_2)$ depends only on the *distance* $||s_1 - s_2||$ between s_1 and s_2.
- *A field Z is anisotropic if at least two directional variograms differ.*
  If $\overrightarrow{e}$ is a direction in $\mathbb{R}^d$, $||\overrightarrow{e}|| = 1$, then the *directional variogram of a stationary field in the direction* $\overrightarrow{e}$ is defined as :
  $$\gamma(h) = \frac{1}{2}\mathrm{Var}(Z(s + h\overrightarrow{e}) - Z(s)) \qquad \forall h \in S$$

Geometric Anisotropy

We distinguish two types of anisotropy :

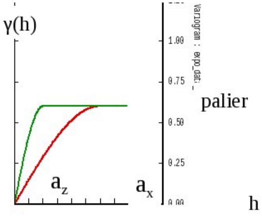
- Geometric anisotropy : it is associated with the *A-linear deformation* of an isotropic model, meaning that : $\gamma(h) = \gamma_0(||Ah||)$.

Such a variogram retains the *same sill in all directions*, but *the ranges differ* depending on the directions.

- Zonal anisotropy : if the *variogram* $h \mapsto \gamma(h)$, after a possible change of coordinates, *depends only on certain coordinates of h*.

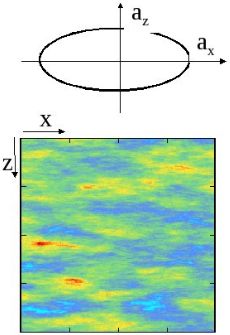
The *sill* of γ *depends on the direction*.

Example : Geometric anisotropy



Variogrammes suivant les 2 directions principales

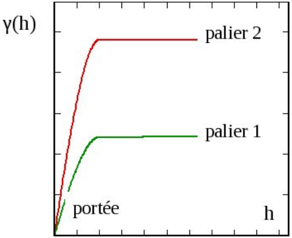
$$\gamma(h) = \gamma_0 \sqrt{\frac{h_x^2}{a_x^2} + \frac{h_z^2}{a_z^2}}$$



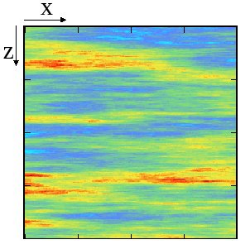
« Lentilles minéralisées »

Groundwater contamination in an alluvial aquifer flowing in one specific direction.
↪ Same overall variance in all directions, but long spatial correlation in the flow direction.

Example : Zonal anisotropy



Variogrammes suivant 2 directions



Exemple $\gamma(h) = \gamma_1(h_x) + \gamma_2(h_x, h_z)$

Rainfall fields associated with meteorological fronts or rain bands.
$\hookrightarrow$ Along the front, rainfall intensity is relatively uniform. Across the front, rainfall alternates rapidly between high and low values.

Variogram models

Usual isotropic variograms

Nugget model :

$$\gamma(h) = \begin{cases} 0, & \text{if} \quad \|h\| = 0 \\ \sigma^2, & \text{if} \quad \|h\| > 0 \end{cases}$$

Gaussian model :

$$\gamma(h) = \sigma^2(1 - \exp(-\|h\|^2/a)), \quad a > 0$$

Exponential model :

$$\gamma(h) = \sigma^2(1 - \exp(-\|h\|/a)), \quad a > 0$$

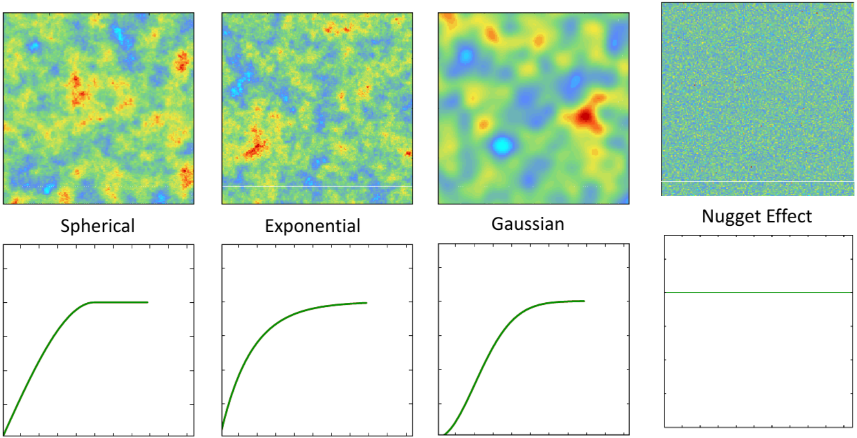
Spherical model :

$$\gamma(h) = \begin{cases} \sigma^2\left(\frac{3}{2}\frac{\|h\|}{a} - \frac{1}{2}\frac{\|h\|^3}{a^3}\right), & \text{if} \quad 0 \leq \|h\| \leq a \\ \sigma^2, & \text{if} \quad \|h\| > a \end{cases}$$

Power Model :

$$\gamma(h) = \sigma^2\|h\|^\alpha, \quad \alpha < 2$$

Usual isotropic variograms



Usual isotropic variograms

Matérn Model :

$$\gamma(h) = \sigma^2\Big[1 - \frac{1}{2^{\nu-1}\Gamma(\nu)}\Big(\frac{2\nu^{1/2}\|h\|}{\rho}\Big)\mathcal{K}_\nu\Big(\frac{2\nu^{1/2}\|h\|}{\rho}\Big)\Big]$$

where

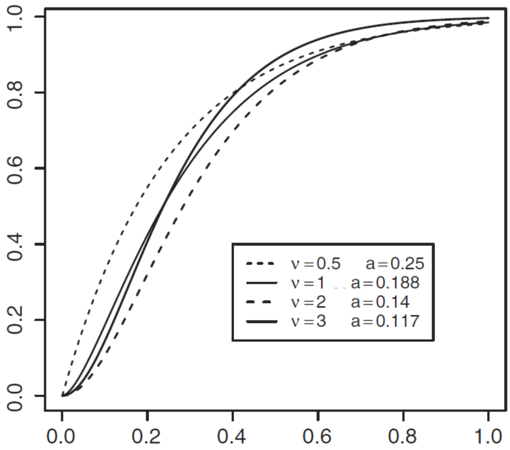
- $\Gamma(\cdot)$ is the gamma function.
- $\mathcal{K}_\nu(\cdot)$ is the modified Bessel of second kind with parameter ν.

ν is the regularity parameter at the origin :

- $\nu = 1/2$: exponential model
- $\nu \to +\infty$: Gaussian model

Matérn variogram

Class of Matérn variograms



Variogram estimation

Estimation of experimental variogram

The experimental variogram is estimated empirically by

$$\widehat{\gamma}(h) = \frac{1}{2n(h)} \sum_{(s_i,s_j)\in S_h}^{n(h)} (z(s_i) - z(s_j))^2,$$

where

- S_h is an approximate class of pairs (s_i, s_j) at distance h $(h \in S)$ for a tolerance Δ.
- $n(h) = \#S(h)$, the cardinal of $S(h)$.

Istropic case :

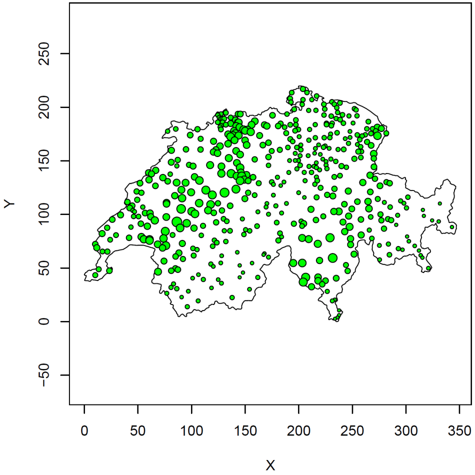
$$S_h = \left\{(s_i, s_j) \in S^2 : \|h\| - \Delta \leq \|s_i - s_j\| \leq \|h\| + \Delta\right\}$$

In practice :

- We estimate $\gamma(\cdot)$ at a finite number k of locations $h_1, h_2, \cdots, h_k$ ensuring that each interval contains at least 30 pairs of points.
- We estimate $\gamma(\cdot)$ in different directions to highlight any anisotropy in the phenomenon under study.

Application (Swiss rainfall)

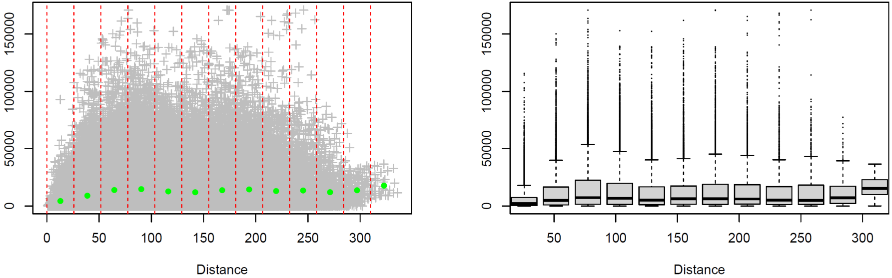
Rainfall records from 467 weather stations in Switzerland



Application (Swiss rainfall)

Cloud variogram is used to express variability according to interdistances

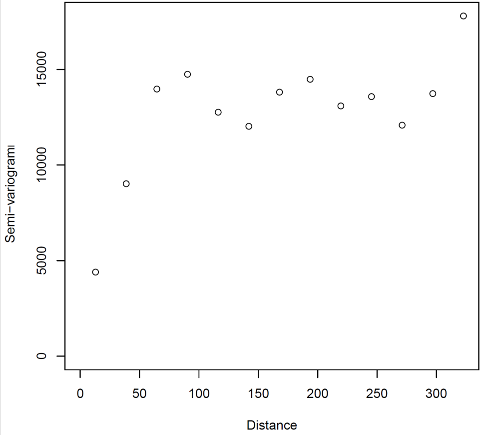
$$\text{The point cloud } \left(\|s_i - s_j\|, (z(s_i) - z(s_j))^2/2\right), \qquad i, j = 1 \ldots, n$$



The green dots indicate the experimental variogram

Application (Swiss rainfall)

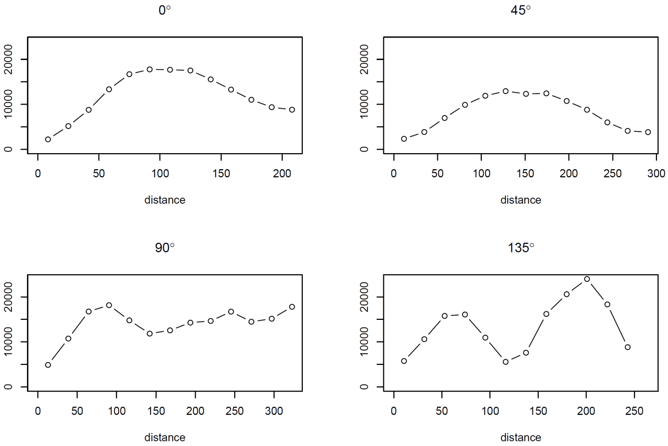
Experimental variogram



Application (Swiss rainfall)

Directional variogram : represent the variogram in several directions.

↪ Assess the possible anisotropy of the phenomenon



Variogram fitting : parametric models

The experimental variogram can't be directly used in geostatistical models.
$\hookrightarrow$ It does not check the variogram properties.

An adjustment to theoretical models with well-defined analytical forms is necessary.

The choice of the theoretical model (variogram) must take into account both empirical knowledge of the phenomenon and the shape of the experimental variogram obtained.

Estimation procedure :

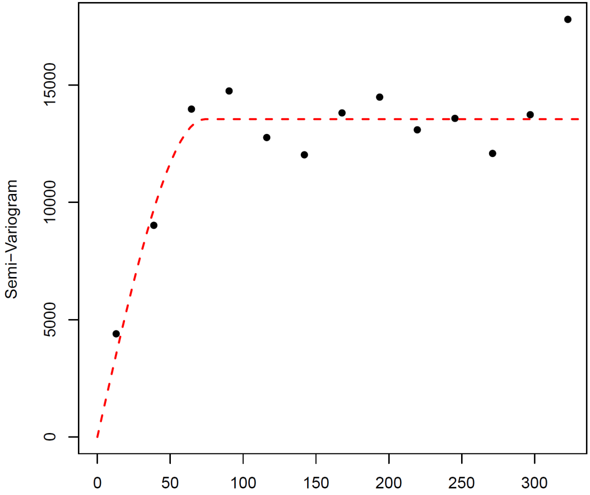
1. Select a parametric model $\gamma_\theta(\cdot)$ whose parameters θ are to be estimated (θ includes all the parameters of the variogram model).
2. Use a parametric estimation method : least squares, maximum likelihood, etc.
   Ordinary Least Squares (OLS) method :

$$\widehat{\theta} = \arg\min_{\theta} \sum_{k=1}^{K} (\gamma_\theta(h_k) - \hat{\gamma}(h_k))^2,$$

where $h_1, \dots, h_K$ are the distances at which the experimental variogram was calculated.

Application (Swiss rainfall) : variogram fitting

Adjusted spherical model



Drawbacks of the empirical variogram

- Sensitive to outliers
- Bad characterization of the behaviour at the origin
- Arbitrary choice of a parametric family

Alternative approaches

- Maximum likelihood (under Gaussian hypothesis) and variants (case where $\mathbb{E}(Z)(s) = 0$)

$$L(\theta) = \frac{1}{(2\pi)^{n/2}\sqrt{|\mathbf{C}_\theta|}} \exp\left(-\frac{1}{2}\mathbf{z}^\top \mathbf{C}_\theta^{-1}\mathbf{z}\right)$$

  where $\mathbf{C}_\theta$ is the (n, n) covariance matrix such that $C_{i,j} = C(s_i - s_j)$

  Log-likelihood

$$l(\theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log\left(\det(\mathbf{C}_\theta)\right) - \frac{1}{2}\mathbf{z}^\top \mathbf{C}_\theta^{-1}\mathbf{z}$$

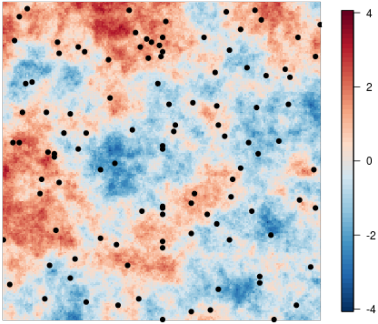
- Bayesian approach (hierarchical model based on the Gaussian model)

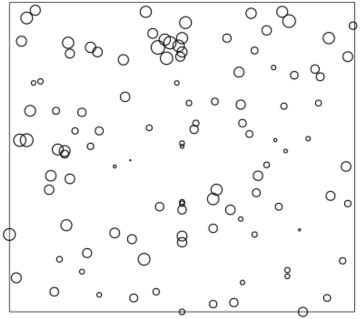
Predictions by kriging

Motivation

- Data : Observations $\{Z(s_i)\}_{i=1}^n$ of a variable Z at n locations $s_1, \dots, s_n$ of a domain $S \subset \mathbb{R}^d$
- Goal : Predict Z at an unobserved site $s_0 \in S$

Start simple : **Linear predictor** $\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i)$



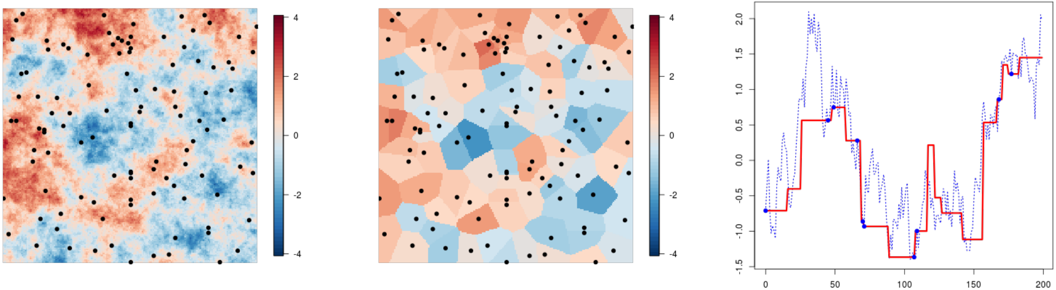


Left : Variable Z. Right : Observations

Motivation

- Data : Observations $\{Z(s_i)\}_{i=1}^{n}$ of a variable Z at n locations $s_1, \ldots, s_n$ of a domain $S \subset \mathbb{R}^d$
- Goal : Predict Z at an unobserved site $s_0 \in S$

Nearest neighbor : $\hat{Z}(s_0) = Z(s_{\mathrm{NN}})$, where s_{NN} = nearest neighbor of s_0

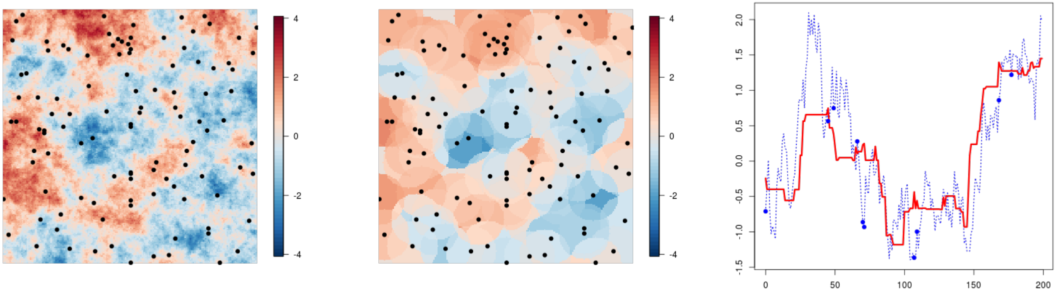


Left : Variable Z. Center : Estimation. Right : Estimation (Red) and True values (Blue) on a section.

Motivation

- Data : Observations $\{Z(s_i)\}_{i=1}^n$ of a variable Z at n locations $s_1, \ldots, s_n$ of a domain $S \subset \mathbb{R}^d$
- Goal : Predict Z at an unobserved site $s_0 \in S$

Moving average : $\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i)$, such that $\lambda_i = 0$ if $\mathrm{dist}(s_i, s_0) > R$, $\sum_{i=1}^n \lambda_i = 1$

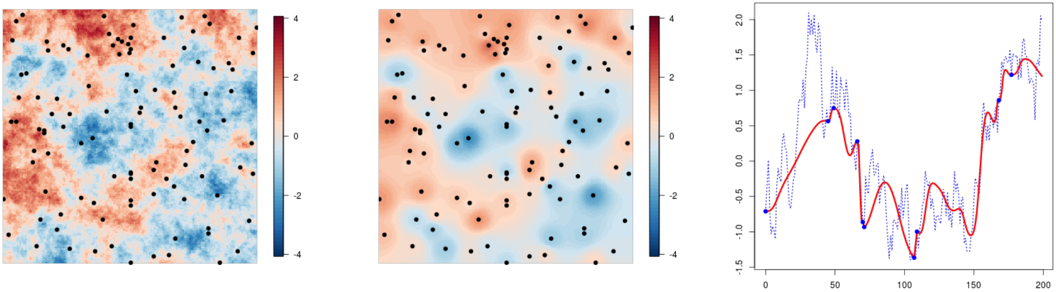


Left : Variable Z. Center : Estimation. Right : Estimation (Red) and True values (Blue) on a section.

Motivation

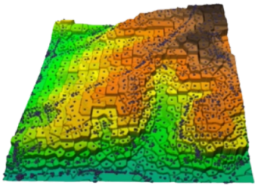
- Data : Observations $\{Z(s_i)\}_{i=1}^n$ of a variable Z at n locations $s_1, \ldots, s_n$ of a domain $S \subset \mathbb{R}^d$
- Goal : Predict Z at an unobserved site $s_0 \in S$

Inverse distance : $\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i), \quad \lambda_i = \mathrm{dist}(s_0, s_i)^{-1} / \big(\sum_{1 \le j \le n} \mathrm{dist}(s_0, s_j)^{-1} \big)$

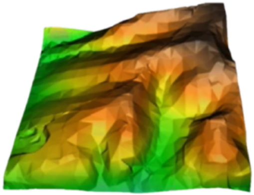


Left : Variable Z. Center : Estimation. Right : Estimation (Red) and True values (Blue) on a section.

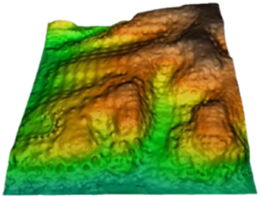
How predictions change ?



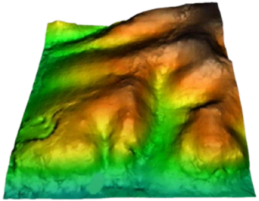
voronoi polygons



linear interpolation



inverse distance weighting



kriging

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

$\hookrightarrow$ **Kriging** : geostatistical method of local estimation, developed by Danie G. Krige (1951) and theorised by Georges Matheron (1962)

Formulation of the problem

Let Z be a spatial process observed at n sites $s_1, \ldots, s_n$ in S.

Kriging consists of predicting Z at an unobserved site s_0 such that the proposed prediction $\widehat{Z}(s_0)$ satisfies the following constraints :

- Linear

$$\widehat{Z}(s_0) = a + \sum_{i=1}^{n} \lambda_i Z(s_i)$$

- Unbiased

$$\mathbb{E}(\widehat{Z}(s_0) - Z(s_0)) = 0$$

- Best

$$\mathrm{Var}(\widehat{Z}(s_0) - Z(s_0)) \quad \text{is minimal}$$

$\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ are the weights to be estimated.

$\widehat{Z}(s_0)$ is called the *Linear Unbiased Predictor* of $Z(s_0)$ (BLUP)

Simple Kriging

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

- **Simple kriging predictor** $Z^{SK}(s_0)$ of $Z(s_0)$ given $Z(s_1), \ldots, Z(s_n)$ = BLUP of $Z(s_0)$:
  - Linear : $Z^{SK}(s_0) = \sum_{i=1}^{n} \lambda_i^{SK} Z(s_i)$, for some weights $\{\lambda_i^{SK}\}_{i=1}^{n}$
  - Unbiased : $\mathbb{E}[Z^{SK}(s_0)] = \mathbb{E}[Z(s_0)] = m = 0$
  - Best : Error variance (aka kriging variance) $\mathrm{Var}[Z^{SK}(s_0) - Z(s_0)]$ is minimal

Simple Kriging

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

- **Simple kriging predictor** $Z^{SK}(s_0)$ of $Z(s_0)$ given $Z(s_1),\dots,Z(s_n)$ = BLUP of $Z(s_0)$:
  - Linear : $Z^{SK}(s_0) = \sum_{i=1}^{n} \lambda_i^{SK} Z(s_i)$, for some weights $\{\lambda_i^{SK}\}_{i=1}^n$
  - Unbiased : $\mathbb{E}[Z^{SK}(s_0)] = \mathbb{E}[Z(s_0)] = m = 0$
  - Best : Error variance (aka kriging variance) $\mathrm{Var}[Z^{SK}(s_0) - Z(s_0)]$ is minimal
- Take $C_{ij} := \mathrm{Cov}(Z(s_i), Z(s_j)) = C(s_i - s_j)$. The conditions above imply :

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

The corresponding kriging variance is $\sigma_{SK}^2(s_0) = \mathrm{Var}[Z^{SK}(s_0) - Z(s_0)] = C(0) - \sum_{i=1}^n \lambda_i^{SK} C_{i0}$

Simple Kriging

- In vectorized form : recall that $C_{ij} = C(s_i - s_j)$ and introduce

$$\boldsymbol{Z} = \begin{pmatrix} Z(s_1) \\ \vdots \\ Z(s_n) \end{pmatrix}, \ \boldsymbol{\lambda}^{SK} = \begin{pmatrix} \lambda_1^{SK} \\ \vdots \\ \lambda_n^{SK} \end{pmatrix}, \ \mathbf{c}_0 = \begin{pmatrix} C_{10} \\ \vdots \\ C_{n0} \end{pmatrix}, \mathbf{C} = \begin{pmatrix} C_{11} & \dots & C_{1n} \\ \vdots & \ddots & \vdots \\ C_{n1} & \dots & C_{nn} \end{pmatrix}$$

Then,

$$Z^{SK}(s_0) = (\boldsymbol{\lambda}^{SK})^T \mathbf{Z}$$

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

$$\mathbf{C}\boldsymbol{\lambda}^{SK} = \mathbf{c}_0 \quad \text{and} \quad \sigma_{SK}^2(s_0) = C(0) - (\boldsymbol{\lambda}^{SK})^T \mathbf{c}_0$$

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

Simple Kriging

- The simple kriging predictor and its variance at the location s_0 can be rewritten as

$$Z^{SK}(s_0) = \mathbf{c}_0^T \mathbf{C}^{-1} \mathbf{Z}, \quad \sigma^2_{SK}(s_0) = C(0) - \mathbf{c}_0^T \mathbf{C}^{-1} \mathbf{c}_0$$

Simple Kriging

- The simple kriging predictor and its variance at the location s_0 can be rewritten as

$$Z^{SK}(s_0) = \mathbf{c}_0^T \mathbf{C}^{-1}\mathbf{Z}, \quad \sigma^2_{SK}(s_0) = C(0) - \mathbf{c}_0^T \mathbf{C}^{-1}\mathbf{c}_0$$

- For multiple target locations $y_1, \ldots, y_p \in S$ we get

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

where $\mathbf{C}_{\mathrm{TD}} = \left[C(y_i - s_j)\right]_{\substack{1\le i\le p \\ 1\le j\le n}} \begin{pmatrix} C(s_1 - y_1) & \ldots & C(s_n - y_1) \\ \vdots & \ddots & \vdots \\ C(s_1 - y_p) & \ldots & C(s_n - y_p) \end{pmatrix} \in \mathbb{R}^{p\times n}$

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

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 $Z - m$

$$Z^{SK}(s_0) - m = \mathbf{c}_0^T \mathbf{C}^{-1}(\mathbf{Z} - m\mathbf{1}_n), \quad \sigma^2_{SK}(s_0) = C(0) - \mathbf{c}_0^T \mathbf{C}^{-1}\mathbf{c}_0$$

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

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

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

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

What if the mean is unknown ?

Goal : Find a linear predictor $\boxed{\widehat{Z}(s_0) = \beta + \sum_{i=1}^{n} \lambda_i Z(s_i)}$ that is unbiased and with minimal error,

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

What if the mean is unknown ?

Goal : Find a linear predictor $\boxed{\widehat{Z}(s_0) = \beta + \sum_{i=1}^{n} \lambda_i Z(s_i)}$ that is unbiased and with minimal error,

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

- Unbiasedness gives

$$0 = \mathbb{E}[\widehat{Z}(s_0) - Z(s_0)] = \beta + \sum_{i=1}^{n} \lambda_i m - m = \beta + m(\sum_{i=1}^{n} \lambda_i - 1)$$

$\rightarrow$ 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 $\boxed{\widehat{Z}(s_0) = \beta + \sum_{i=1}^{n} \lambda_i Z(s_i)}$ that is unbiased and with minimal error,

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

- Unbiasedness gives

$$0 = \mathbb{E}[\widehat{Z}(s_0) - Z(s_0)] = \beta + \sum_{i=1}^{n} \lambda_i m - m = \beta + m(\sum_{i=1}^{n} \lambda_i - 1)$$

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

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

- Consequently, the error variance is given by :

$$\mathrm{Var}[\widehat{Z}(s_0) - Z(s_0)] = C(0) - 2\boldsymbol{\lambda}^T \mathbf{c}_0 + \boldsymbol{\lambda}^T \mathbf{C} \boldsymbol{\lambda}$$

What if the mean is unknown ?

Goal : Find a linear predictor $\boxed{\widehat{Z}(s_0) = \beta + \sum_{i=1}^{n} \lambda_i Z(s_i)}$ that is unbiased and with minimal error,

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

- The resulting predictor is called the **ordinary kriging predictor**
- **Ordinary kriging predictor** $Z^{OK}(s_0)$ of $Z(s_0)$ given $Z(s_1), \ldots, Z(s_n)$ = BLUP of $Z(s_0)$ when the mean m of Z is unknown :

$$\boxed{Z^{OK}(s_0) = \sum_{i=1}^{n} \lambda_i^{OK} Z(s_i)} = (\boldsymbol{\lambda}^{OK})^T \mathbf{Z}$$

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

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

Ordinary kriging

- The weights $\boldsymbol{\lambda}^{OK}$ of the ordinary kriging predictor $Z^{OK}(s_0) = (\boldsymbol{\lambda}^{OK})^T \mathbf{Z}$ are defined as

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

- The Lagrangian of the system writes

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

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

$$\begin{pmatrix} \mathbf{C} & \mathbf{1}_n \\ \mathbf{1}_n^T & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda}^{OK} \\ \mu^{OK} \end{pmatrix} = \begin{pmatrix} \mathbf{c}_0 \\ 1 \end{pmatrix}, \quad \begin{cases} \mathbf{C} = [C(s_i - s_j)]_{1 \le i,j \le n} \\ \mathbf{c}_0 = [C(s_i - s_0)]_{1 \le i \le n} \end{cases}$$

Ordinary kriging

- The weights $\boldsymbol{\lambda}^{OK}$ of the ordinary kriging predictor $Z^{OK}(s_0) = (\boldsymbol{\lambda}^{OK})^T \mathbf{Z}$ are defined as

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

- The Lagrangian of the system writes

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

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

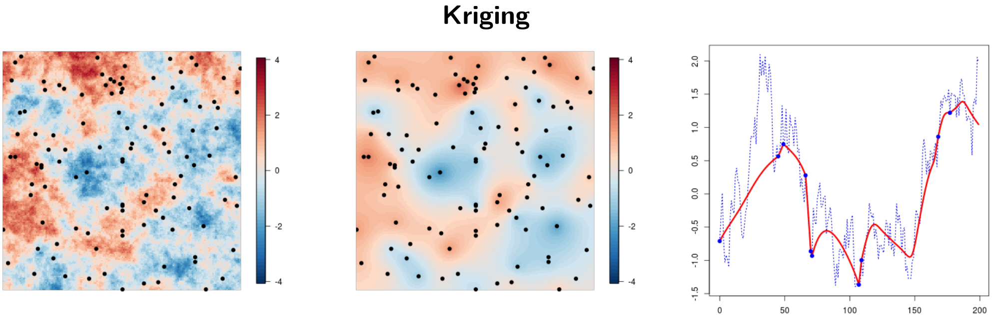
$$\begin{pmatrix} \mathbf{C} & \mathbf{1}_n \\ \mathbf{1}_n^T & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda}^{OK} \\ \mu^{OK} \end{pmatrix} = \begin{pmatrix} \mathbf{c}_0 \\ 1 \end{pmatrix}, \quad \begin{cases} \mathbf{C} = [C(s_i - s_j)]_{1 \leq i,j \leq n} \\ \mathbf{c}_0 = [C(s_i - s_0)]_{1 \leq i \leq n} \end{cases}$$

- The associated kriging variance is given by

$$\sigma^2_{OK}(s_0) = C(0) - (\boldsymbol{\lambda}^{OK})^T \mathbf{c}_0 - \mu^{OK}$$

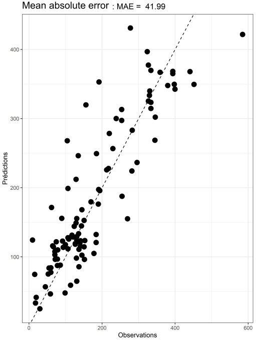
- Once again, the kriging weights variance do not depend on the locations of the observations

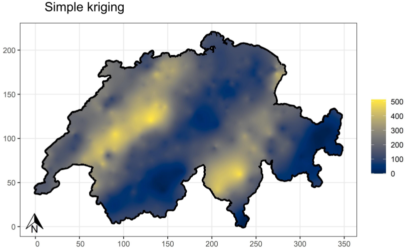
Ordinary kriging

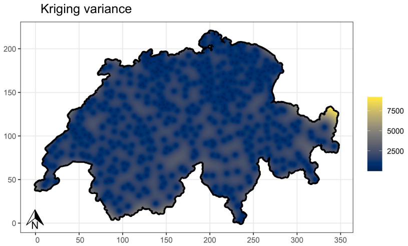


Left : Variable Z. Center : Estimation. Right : Estimation (Red) and True values (Blue) on a section.

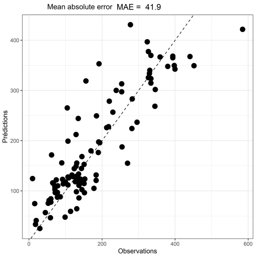
Simple Kriging : application (Swiss rainfall)

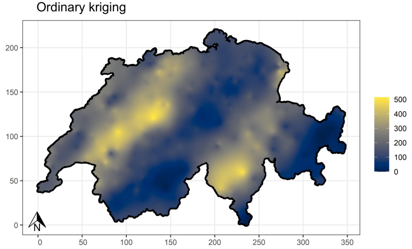


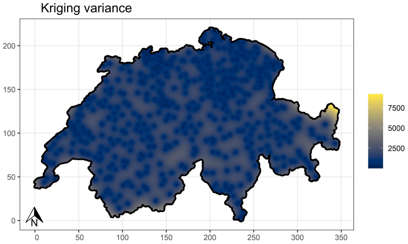




Ordinary Kriging : application (Swiss rainfall)







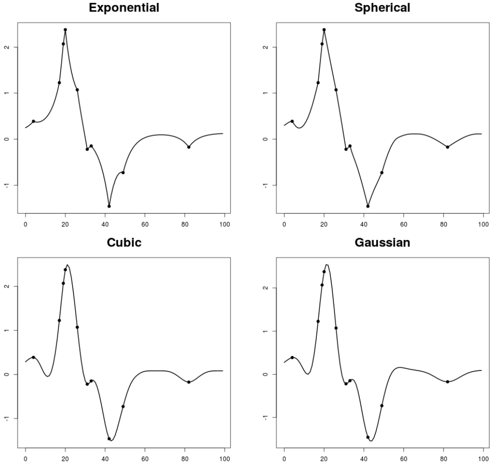
Properties of SK and OK predictors

Let $Z^K(s_0) = Z^{SK}(s_0)$ or $Z^{OK}(s_0)$ using observations $Z(s_1), \ldots, Z(s_n)$ 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. $\forall i \in \{1, \ldots, n\}$, $Z^K(s_i) = Z(s_i)$

  Eg. Simple kriging : If $s_0 = s_1$ then $\boldsymbol{c}_0 =$ First column of $\boldsymbol{C}$ and therefore $\boldsymbol{\lambda}^{SK} = (1, 0, \ldots, 0) \Rightarrow$ $Z^{SK}(s_1) = Z(s_1)$
- Kriging weights **can be negative** $\rightarrow$ Allows to predict outside the range of the data

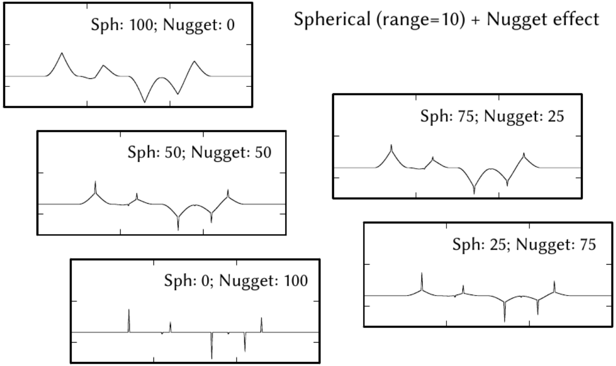
Properties : Influence of covariance choice



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

- Regularity of the covariance affects the regularity of the estimate

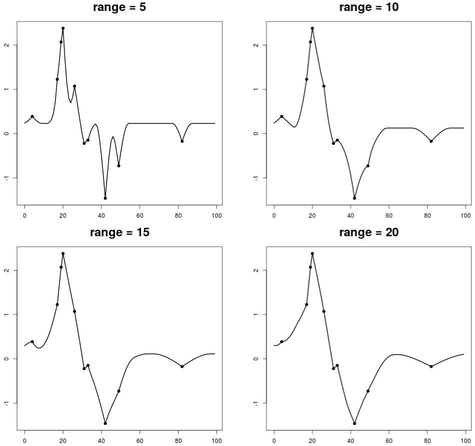
Properties : Influence of nugget effect



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 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
      $\hookrightarrow$ The mean m has an important role !

Influence of variogram parameters

Nugget

- If the variogram is only nugget, then $\lambda_i = 1/n$
  $\hookrightarrow$ Kriging is constant (except at observation sites), as well as Kriging variance.
- Presence of a nugget in variogram function smooths the prediction.
- Absence of a nugget implies that the regularity of the variogram at 0 is the regularity of the process itself.

Range

- $\lambda_i = 0$ if $||s_i - s_0|| > \text{range}$.
- Screen effect.

Sill

- Almost no influence.

Conditional simulations

Conditional Simulations

Principle : simulate random fields that follow a given law such that values at measurement locations are equal to observations

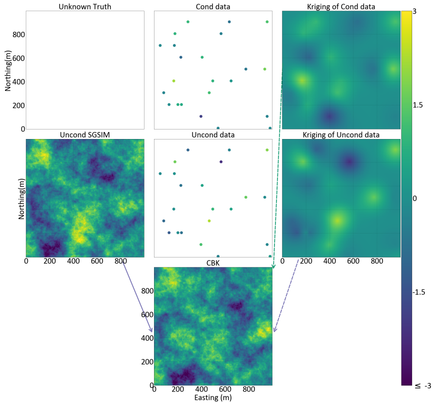
- Estimate the covariance function of the field $C(h)$
- Compute Kriging from observations on a grid

$$\hat{Z} = \sum_{i=1}^{n} \lambda_i Z(s_i)$$

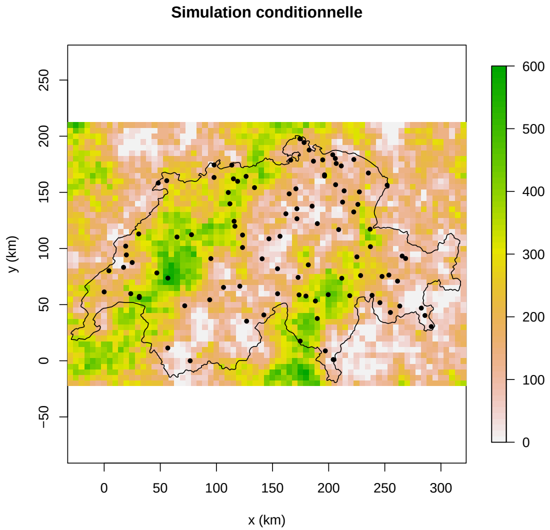
- Simulate a field Y on S with covariance function $C(h)$
- Obtain Kriging from simulated values at measurement locations on the same grid

$$\hat{Y} = \sum_{i=1}^{n} \lambda_i Y(s_i)$$

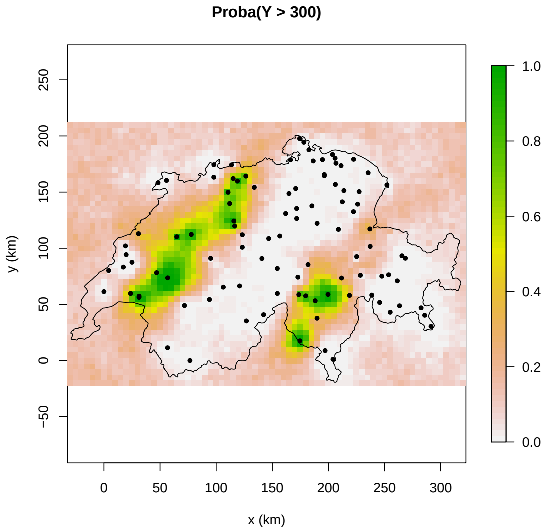
- Obtain a conditional simulation as $\hat{Z} + Y - \hat{Y}$



Swiss rainfall



Swiss rainfall



From Geostatistics to Spatio-Temporal Modeling

Spatio-Temporal Extension

- Spatio-temporal process : $Z(s,t)$, $(s,t) \in \mathbb{R}^d \times \mathbb{R}$
- Covariance : $C(h,u) = \mathrm{Cov}(Z(s,t), Z(s+h, t+u))$
- Separable models : $C(h,u) = C_S(h)\, C_T(u)$
- Non-separable models : space–time interaction
- Dynamic view : state-space and evolution equations

Key Questions

- How does spatial dependence evolve over time ?
- Are space and time separable ?
- Can temporal dynamics improve spatial prediction ?

Computational Challenges in Spatio-Temporal Modeling

Main Computational Bottlenecks

- Large datasets : $n = n_s \times n_t$ observations
- Dense covariance matrices of size $n \times n$
- $\mathcal{O}(n^3)$ cost for likelihood evaluation
- Storage cost $\mathcal{O}(n^2)$
- Non-separable covariance models increase complexity

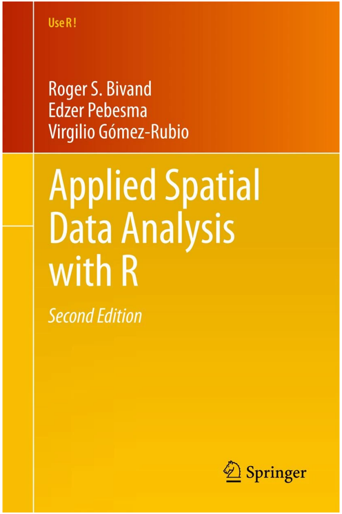
Statistical and Numerical Extensions

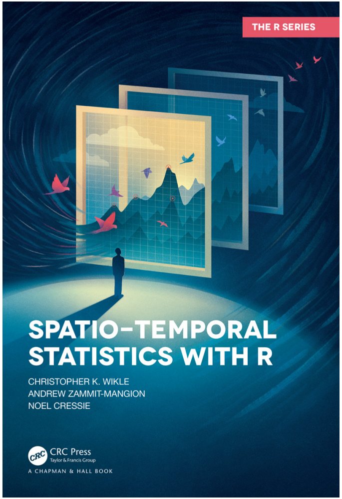
- Separable structures : exploit Kronecker products
- Sparse precision models : Gaussian Markov random fields (GMRFs)
- Approximate inference : composite likelihood, Vecchia approximations

Key Idea

- Replace dense covariance matrices with sparse or structured representations
- Trade exactness for scalability when necessary

Spatio-(temporal) data science books with R





Spatial data science resources

- Books :
  Spatial analysis manual (Theory and practical implementation with R)
  `https://www.insee.fr/fr/information/3635442`
  Spatial modelling and statistics
  `https://link.springer.com/book/10.1007/978-3-540-79226-0`
- Packages R :
  RandomFields, GeoR, fields, spdep,...
- Extensive number of resources : published papers, scientific events, journals, societies, ...
- More than 100 free R, Python, Matlab,..., softwares