Corso di
Bonifica dei siti inquinati
Modulo di Idrologia Sotterranea: Parte II
Metodi stocastici nell'Idrologia

A.A. 2002/2003

Indice

Elenco delle figure

Chapter 1

Statistical preliminaries

1.1 Random Variables

1.1.1 The concept of Probability

A general model for an experiment could be constructed in the following manner:

1. Take a non-empty set Ω, in such a way that each $\omega \in \Omega$ represents a possible **outcome** of the experiment.

2. To the "greatest possible number" of subsets $A \subset \Omega$, each subset A representing a possible **event**, associate a number $P(A) \in [0, 1]$, called the **probability** of such event.

An event A has been observed when we perform a "realization of the experiment" and an outcome $\omega \in A$ is obtained. The probability of the event can be defined as follows ("frequentist approach"):

Repeat the experiment N times, with N large, and observe the results. If event A has occurred N_A times then

$$f_a = \frac{N_A}{N}$$

is the relative frequency of event A and is an approximation of the probability $P(A)$ of occurrence of A. Then $P(A)$ can be defined as:

$$P(A) = \lim_{N\to\infty} \frac{N_A}{N}$$

In practice the probability is a function relating the elements of the family of events to the interval $[0, 1]$:

$$P : \mathcal{A} \to [0, 1]$$

where $\mathcal{A} \subset \Omega$ and the following properties hold:

$$\begin{aligned} P(\Omega) &= 1 \\ P(\sum_{n=1}^{\infty} A_n) &= \sum_{n=1}^{\infty} P(A_n) \end{aligned}$$

for any family $\mathcal{A} = \{A_n\}$ of events for which $A_n \bigcap A_m = \emptyset$ if $n \neq m$.

In mathematical terms the ordered triple

$$(\Omega, \mathcal{A}, P) \tag{1.1}$$

is called a **probability space** if

- Ω is a non-empty set (the space of **outcomes**),
- $\mathcal{A}$ is a family of subsets of Ω (the **family of events**),
- $P : \mathcal{A} \to [0, 1]$ is a probability.

The structure (1.1) constitutes the basis for a mathematical model of an experiment, keeping in mind the non-reproducibility which is often encountered in empirical sciences. For instance, in a coin toss the outcomes are "heads" (H) or "tails" (T) and we can take

$$\Omega = \{H, T\}$$

All possible events are elements of

$$\mathcal{A} = \{\emptyset, \{H\}, \{T\}, \Omega\}$$

Lastly, if the coin is "fair", intuitively the probability P is given by

$$\begin{aligned} P(\emptyset) = 0 \qquad & P(\Omega) = 1 \\ P(H) = 1/2 \qquad & P(T) = 1/2 \end{aligned}$$

All the properties of probability, without exception, are derived from the mathematical model (1.1).

1.1.2 Random variables and random functions

Let us start with an example related to groundwater hydrology.

- *We wish to study the steady state response of a confined heterogeneous aquifer, under conditions of water withdrawal. We measure the hydraulic conductivity at various points in the aquifer for confidence in the results. Hence it is convenient to consider modeling the aquifer as a medium with random characteristics.*

To this aim, we perform a pumping test: withdraw water from a well at specified rates and observe water drawdown at different wells. By means of pumping test theory, calculate transmissivities at the different well locations (points in space). (Note that the number of observation wells is generally small due to the high drilling costs).

The possible outcomes in our experiment (i.e. the corresponding profiles of the transmissivities) are functions which assume a value at each point in the space which we are working in. Let $S \subset \mathbb{R}^3$ be the mathematical representation of this space (the domain of the aquifer). Then the elements of Ω are the functions $\omega : S \to \mathbb{R}$ where

$$\omega(x) = \text{transmissivity at point } x \ (\omega \in \Omega).$$

and we assume that these functions are continuous ($\Omega = C^1(S)$). A more classical notation for the random transmissivity at point x when the profile is ω would be:

$$T(x\omega) \equiv \omega(x)$$

Ω is made up of continuous functions. Let $h : S \to \mathbb{R}$ be the piezometric head ($h \in C^2(S)$). The aquifer is confined and the steady state mass balance equation together with Neumann (no flow) boundary conditions is:

$$\frac{\partial}{\partial x_i}\left(T_{ij}(\vec{x},\omega)\frac{\partial h}{\partial x_j}\right) = q(x) \text{ in } S \tag{1.2}$$

$$\frac{\partial u}{\partial n} = 0 \text{ in } \partial S \tag{1.3}$$

where $q(x)$ represents the rate of water withdrawal or injection per unit time and unit volume.

The solution to this boundary value problem is a function

$$h : S \times \Omega \to \mathbb{R}$$

For each point $x \in S$ we determine a function $\omega \mapsto h(x,\omega)$ which represents the piezometric level at the point x if the transmissivity rofile in the medium is $\omega \in \Omega$.

The piezometric head is a random function of the random variable ω. Note that a random variable (RV) is also a function from the sample set to the real axis ($\omega : S \to \mathbb{R}$), while the random function $h(x,\omega)$ is a function $S \times \Omega h \to \mathbb{R}$. In other words, $h(x_o,\omega)$ and $h(x_1,\omega)$ are two random variables if $x_o \neq x_1$; $h(x,\omega_o)$ is a realization of $h(x,\omega)$, i.e. the profile (function of x) of T when (among all possible observations) only $\omega_o(x)$ is assumed.

In general we are interested in real functions X defined in the space of outcomes, and in associating a probability with all the events of the type

- The value of X is larger than $b \in \mathbb{R}$.
- The value of X is not larger than $a \in \mathbb{R}$.
- The value of X falls in the interval $[a,b]$, etc.

These events will be denoted by means of the symbols $(X > b)$, $(X \leq a)$, $(a \leq X \leq b)$, etc.

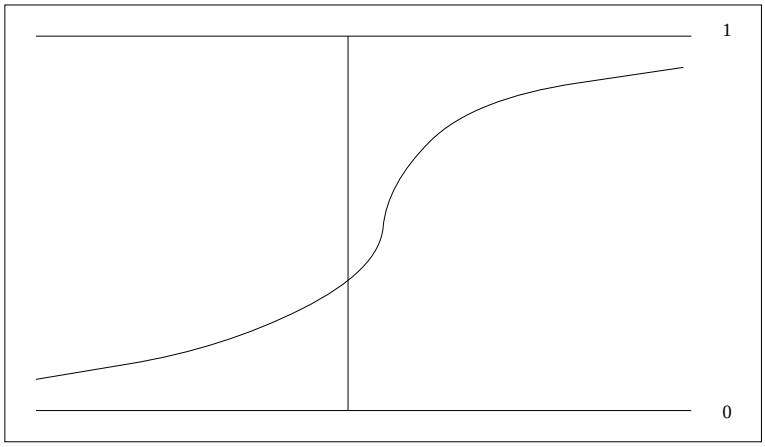


Figure 1.1: A continuous CDF

1.2 Distributions

Let X be a random variable. We can determine a function $F_X : \mathbb{R} \to \mathbb{R}$ (called the **cumulative distribution** – CDF – or simply **distribution function** of X) given by

$$F_X(x) := P(X \leq x). \tag{1.4}$$

Clearly if $x < y$ then $(X \leq x) \subset (X \leq y)$ and hence $F_X(x) \leq F_X(y)$, that is, F_X is monotonically increasing. Furthermore, when $x \to +\infty$ the event $(X \leq x)$ tends to Ω and when $x \to -\infty$ the event $(X \leq x)$ tends to $\emptyset$. Hence it is reasonable that

$$\lim_{x \to -\infty} F_X(x) = 0, \quad \lim_{x \to +\infty} F_X(x) = 1. \tag{1.5}$$

The plots of cumulative distribution functions have the general form shown in figures 1.1 or 1.2.

How can we determine the cumulative distribution function of a given random variable X? We can sample values of X, repeating N times the corresponding experiment, obtaining for each $x \in \mathbb{R}$ the table of results:

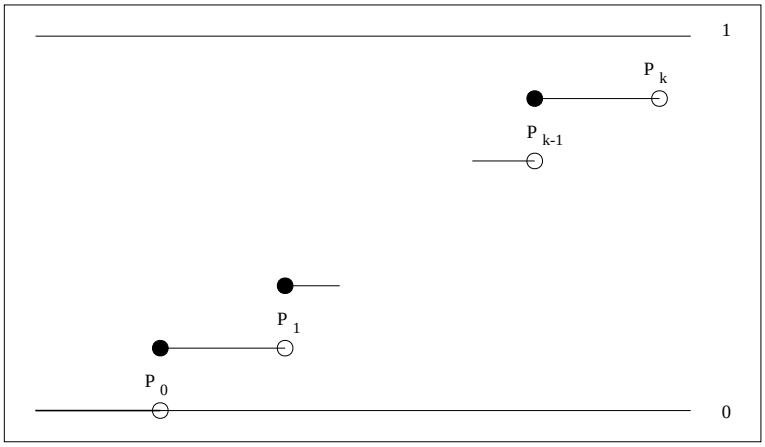


Figure 1.2: A discrete CDF

Repetition	Observation	$x_i \leq x$?
1	x_1	yes
2	x_2	no
$\vdots$	$\vdots$	$\vdots$
N	x_N	yes

Clearly

$$F_X(x) \simeq \frac{\#\{i \leq N : x_i \leq x\}}{N}$$

if N is large, according to the frequency approach. The function F_e given by

$$F_e(x) := \frac{\#\{i \leq N : x_i \leq x\}}{N}, \quad x \in \mathbb{R}$$

is called the **empirical cumulative distribution function** of X. If $x_1 \leq \ldots \leq x_N$, it has a plot of the form represented in figure 1.2.

Def.

A random variable X is said to be **continuous** if its cumulative distribution function has the form

$$F_X(x) = \int_{-\infty}^{x} f_X(\xi) d\xi$$

f_X is called the **probability density function** (briefly **density**) of X.

Clearly

$$f_X(x) \geq 0, \quad \int_{-\infty}^{+\infty} f_X(\xi)d\xi = 1$$

for any density function f_X.

1.2.1 Examples of continuous distributions

A random variable has a cumulative distribution $F_X(x)$ if $P(X \leq x) = F_X(x)$. Examples of cumulative distributions and their relative density functions are given below.

1. Exponential density function:

$$f_X(x) = \begin{cases} 0 & \text{if } x < 0, \\ \lambda e^{-\lambda x} & \text{if } x \geq 0. \end{cases}$$

The random variable X has the cumulative distribution (shown in figure 1.3):

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = (x \geq 0) = \int_0^x \lambda e^{-\lambda t}dt = -\frac{\lambda e^{-\lambda t}}{\lambda}\Big|_0^x = 1-e^{-\lambda x}$$

2. standard Gaussian (Normal) distribution ($N(0,1)$):

$$f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

and

$$F_X(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}dy,$$

called the **standard Gaussian distribution**.

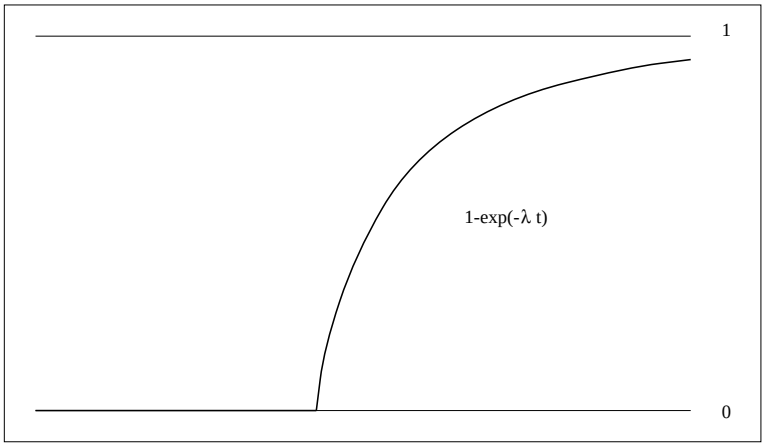


Figure 1.3: Exponential distribution.

If $t = (x - \mu)/\sigma$ then we have the Gaussian distribution with mean μ and variance σ^2 ($N(\mu, \sigma^2)$):

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

and

$$F_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{1}{2}(\frac{y-\mu}{\sigma})^2} dy$$

3. Lognormal distribution:

   let $y = \ln x$ and $y \in N(\mu, \sigma^2)$ is a lognormal random variable. Then, since

$$\frac{dy}{dx} = 1 \qquad f_X(x) = f_Y(y)\frac{dy}{dx}$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{1}{2}(\frac{y-\mu_y}{\sigma_y})^2}$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_y x} e^{-\frac{1}{2}(\frac{\ln x-\mu_y}{\sigma_y})^2}$$

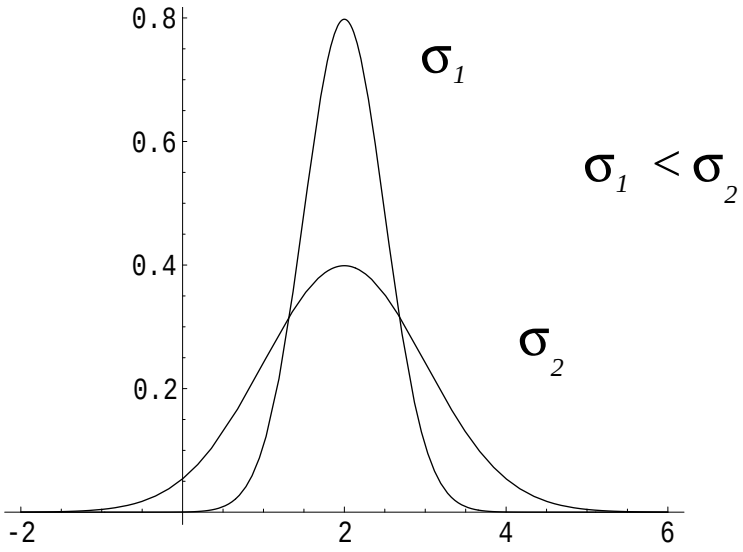


Figure 1.4: Gaussian cumulative distribution

1.2.2 Expected Value

A random variable can take on a large, often infinite, number of values. We are interested in substituting, in place of the possible values of the random variable, a representative value that takes into account the large or small probabilities associated with the RV. This value is called the Expected Value or Expectation (in italiano valore atteso o speranza matematica) and its symbol is $E[\cdot]$. In the discrete case we can define such a value as an average of the observations weighted with the respective probabilities, while in the continuous case we need to substitute sums with integrals:

- Let X be a discrete random variable and let $P(X = x_k) = P_k$. Then

$$E[X] = \sum_{k=1}^{\infty} x_k p_k$$

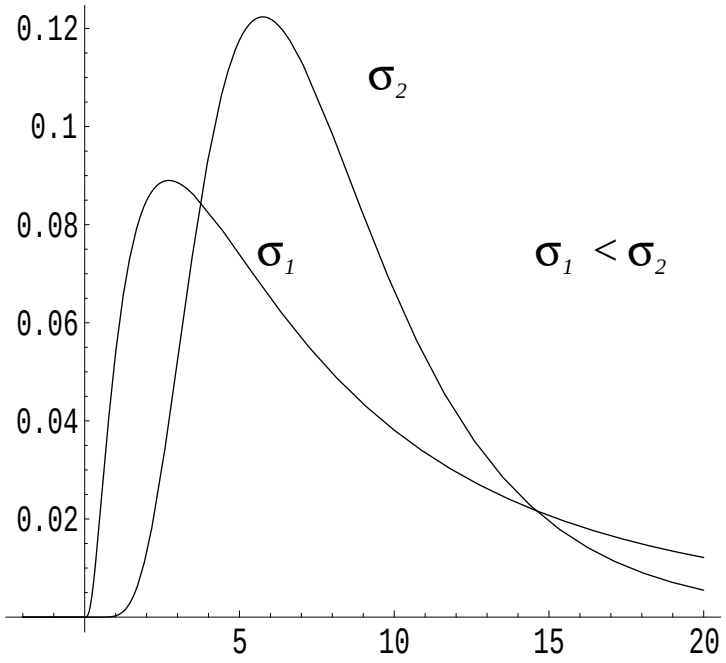


Figure 1.5: Lognormal cumulative distribution

If $\phi(x)$ is a continuous function then:

$$E[\phi(X)] = \sum_{k=1}^{\infty} \phi(x_k) p_k$$

- Let X be a continous random variable with continous density f_X. Then

$$E[X] = \int_{-\infty}^{+\infty} t f_X(t) dt$$

If $\phi(x)$ is a continuous function then:

$$E[\phi(X)] = \int_{-\infty}^{+\infty} \phi(t) f_X(t) dt$$

It can be easily verified that

$$E[\alpha X + \beta Y] = \alpha EX + \beta EY$$

Examples:

- if $X \sim N(0,1)$, then

$$\begin{aligned} E[X] &= \int_{-\infty}^{+\infty} t \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \Big|_{-\infty}^{+\infty} \\ &= 0 \end{aligned}$$

- if $X \sim N(\mu, \sigma^2)$, then

$$\begin{aligned} E[X] &= \int_{-\infty}^{+\infty} \frac{t}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{t-mu}{\sigma})^2} dt \\ &\quad (z = \frac{t-\mu}{\sigma} \Rightarrow dz = \frac{dt}{\sigma} \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} (\sigma z + \mu) e^{-\frac{z^2}{2}} dz \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z e^{-\frac{z^2}{2}} dz + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} dz \\ &= \mu \end{aligned}$$

- if $X \sim Exp(\lambda)$

$$\begin{aligned} E[X] &= \int_0^{+\infty} te^{-\lambda t} dt \\ \text{(p.p.)} &= \lambda e^{-\lambda t}\Big|_0^{+\infty} - \int_0^{+\infty} e^{-\lambda t} dt \\ &= \frac{1}{\lambda} \end{aligned}$$

We are also interested in measuring the discrepancies with respect to the expected value. This is accomplished by the "variance":

$$\text{var}[X] = E[(X - EX)^2] = E[X^2] + E[X]^2 - 2E[X]E[X] = E[X^2] - E[X]^2$$

Examples:

- if $X \sim N(0,1)$, then

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{+\infty} t^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \\ \text{(p.p.)} &= \frac{1}{\sqrt{2\pi}} \left[-te^{-\frac{t^2}{2}}\Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e; -\frac{t^2}{2} dt \right] \\ &= 1 \end{aligned}$$

  and since $E[X] = 0$ we have:

$$\text{var}[X] = 1$$

- if $Y \sim N(\mu, \sigma^2)$, then we can write $Y = \sigma X + \mu$ with $X \sim N(0,1)$. It follows:

$$E[Y] = \mu; E[X^2] = 1$$

$$\text{var}[Y] = E[(y - \mu)^2] = E[(\sigma X + \mu - \mu)^2] = \sigma^2 E[X^2] = \sigma^2$$

- if $X \sim U[0,1]$, we say that a random variable X has uniform distribution in $[0,1]$. Its density is shown in Figure 1.6, and its cumulative density is as represented in figure 1.7. A simple calculation shows that $E[X] = 1/2$ and $\text{var}[X] = 1/12$.

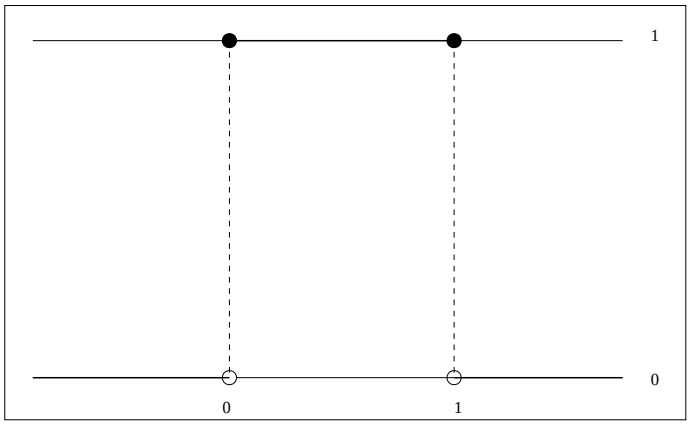


Figure 1.6: Uniform density in $[0, 1]$.

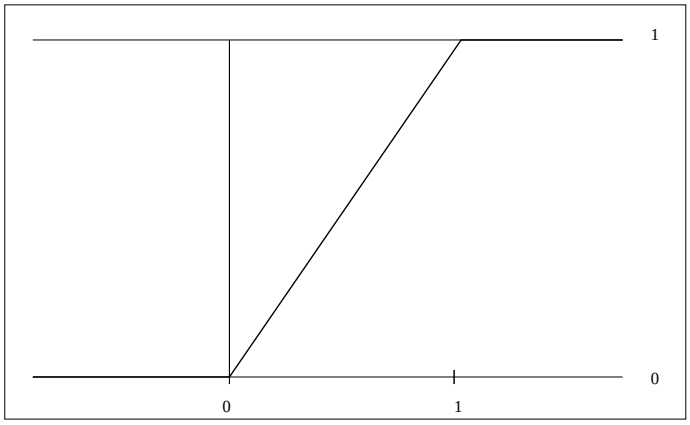


Figure 1.7: Uniform CDF in $[0, 1]$.

1.3 Random Vectors

Consider the example of the confined aquifer in section 1.1.2. An infinite family $\mathcal{F}$ of random variables

$$\{Y_x, x \in S\} \tag{1.6}$$

was constructed there, where

$$Y_x(\omega) := h(x, \omega)$$

is the piezometric head at x corresponding to a transmissivity profile ω.

Suppose a number of points $p_1, \ldots, p_n \in S$ are selected for measurement and let the random variables $X_1, \ldots, X_n$ be defined by

$$X_i := Y_{p_i}, \quad i = 1, \ldots, n.$$

Then, $X := (X_1, \ldots, X_n)$ constitutes a **random vector**: it is a vector valued random quantity.

Just as in the scalar case, the probabilities of the components of a random vector X are embodied in its joint distribution function $F_X : \mathbb{R}^n \to \mathbb{R}$, defined as follows:

$$\begin{aligned} F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) &= P(X_1 \leq x_1, \ldots, X_n \leq x_n) \\ &= P((X_1 \leq x_1) \cap X_2 \leq x_2 \cap \ldots \cap X_n \leq x_n) \end{aligned}$$

For instance if we throw two dice at the same time the outcome of one does not depend on the outcome of the other. Then we can have: zsh: command not found: G if its density is

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\|x\|^2/2}$$

where $\|\cdot\|$ denotes the ordinary Euclidean norm in $\mathbb{R}^n$.

The expectation of a random vector X is defined componentwise:

$$E[X] := (E[X_1], \ldots, E[X_n])^T$$

The **covariance matrix** of X is defined as:

$$\mathrm{Cov}(X) := E[(X - E[X])(X - E[X])^T]$$

The (i, j)-th element of $\mathrm{Cov}(X)$ is

$$c_{ij} := E[(X_i - E[X_i])(X_j - E[X_j])^T]$$

and it is referred to as the **covariance** of X_i and X_j.

A simple computation shows that if $X \sim N(\mathbf{0}, I)$, then X has mean $\mathbf{0} \in \mathbb{R}^n$ and its covariance matrix is the identity matrix $I \in \mathbb{R}^{n \times n}$.

1.3.1 Independence

The **conditional probability of** A **given** $B$is defined as

$$P(A|B) := \frac{P(A \cap B)}{P(B)}$$

provided both A and B are events and $P(B) \neq 0$.

One can say that A and B are **independent** if $P(A|B) = P(A)$ and $P(B|A) = P(B)$. Thus A and B are independent if and only if

$$P(A \cap B) = P(A)P(B) \tag{1.7}$$

Clearly, this last condition makes sense even if either $P(A)$ or $P(B)$ vanishes, hence can be (and usually is) taken as the definition of independence of A and B.

Two random variables X and Y are **independent** if the events $(X \leq x)$ and $(Y \leq y)$ are independent in the sense of (1.7) for each choice of $x, y \in \mathbb{R}$, i.e. if

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y) \tag{1.8}$$

In other words X and Y are independent if and only if

$$F_{X,Y}(x, y) = F_X(x)F_Y(y). \tag{1.9}$$

A random vector $X \in \mathbb{R}^n$ has **independent components** if all its marginals $F_{X_{i1},\ldots,X_{ik}}$ have the multiplicative property

$$F_{X_{i1},\ldots,X_{ik}} = F_{X_{i1}} \cdots F_{X_{ik}},$$

for each ordered k-tuple $(i_1, \ldots, i_k)$ with $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$ with $k \leq n$. N.B. It does **not** suffice to ask that $F_{X_1,\ldots,X_n}$ have the multiplicative property for $k = n$ only.

If X and Y have a joint density f_{XY}, then both X and Y have a density (f_X and f_Y, respectively), namely the marginals

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f_{XY}(x,y)dy \\ f_Y(y) &= \int_{-\infty}^{+\infty} f_{XY}(x,y)dx. \end{aligned}$$

The vectors X and Y are independent with a joint density $f_{X,Y}$, then

$$f_{XY}(x,y) = f_X(x)f_Y(y) \tag{1.10}$$

Conversely, if the joint density factors into the marginal densities, then (1.9) holds. Thus (1.10) is a necessary and sufficient condition for independence.

Let X and Y be independent and let ϕ and ψ be "regular" functions. Then:

$$\begin{aligned} E[\phi(X)\psi(Y)] &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \phi(x)\psi(y)dF_{XY}(x,y) \\ &= \int_{-\infty}^{+\infty} \phi(x)dF_X(x) \int_{-\infty}^{+\infty} \psi(y)dF_Y(y), \end{aligned}$$

i.e. under independence of X and Y,

$$E[\phi(X)\psi(Y)] = E[\phi(X)]E[\psi(Y)] \tag{1.11}$$

In particular, if X, Y are independent, then

$$\mathrm{Cov}[X,Y] = \begin{pmatrix} \mathrm{Var}[X] & 0 \\ 0 & \mathrm{Var}[Y] \end{pmatrix}$$

In fact, by (1.11)

$$E[(X - E[X])(Y - E[Y])] = E[(X - E[X])]E[(Y - E[Y])] = 0$$

Moreover

$$\mathrm{Var}[aX + bY] = a^2\mathrm{Var}[X] + 2abE[(X - E[X])(Y - E[Y])] + b^2\mathrm{Var}[Y]$$

i.e.

$$\mathrm{Var}[aX + bY] = a^2\mathrm{Var}[X] + b^2\mathrm{Var}[Y]$$

if X and Y are independent.

The above results can be generalized for random vectors. If an n-dimensional random vector X has independent components, then

$$\mathrm{Cov}[X] = \mathrm{diag}(\mathrm{Var}[X_1], \ldots, \mathrm{Var}[X_n])$$

and

$$\mathrm{Var}[c^T X] = c_1^2\mathrm{Var}[X_1] + \cdots + c_n^2\mathrm{Var}[X_n]$$

Let us apply the above results to the following particular situation: sampling a given random variable, like when a measurement is repeated a certain number of times.

Suppose X is a given random variable. A **sample** of length n from X is a sequence of independent random variables $X_1, \ldots, X_n$, each of them having the same distribution as X. The components of the sample are said to be **independent and identically distributed** ("iid" for short). The **sample mean**

$$M_n := \frac{X_1 + \ldots + X_n}{n} \tag{1.12}$$

is computed in order to estimate the "value" of X.

If

$$E[X] = m, \quad \mathrm{Var}[X] = \sigma^2$$

Then

$$EM_n = m, \quad \mathrm{Var}M_n = \frac{\sigma^2}{n} \tag{1.13}$$

and the advantage of forming the sample average becomes apparent: While the mean is unaltered, the variance reduces when the number n of

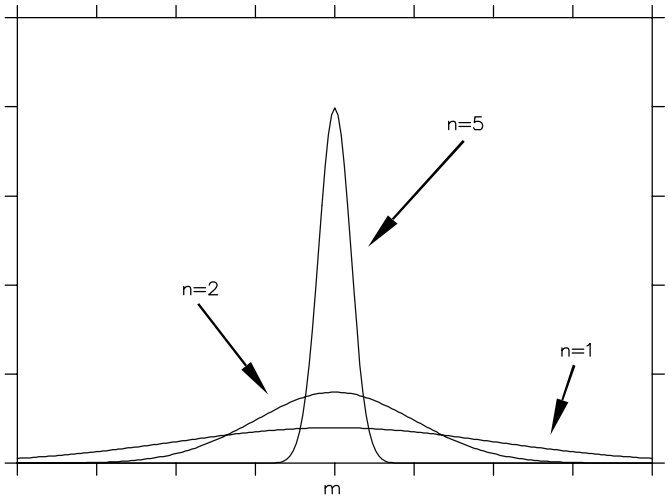


Figure 1.8: Density of the sample mean when the number of observations n increases.

observations increases. Thus, forming the arithmetic mean of a sample of measurements results in a higher precision of the estimate. The situation is as depicted in figure 1.8.

One feels tempted to assert that

$$M_n \to m \text{ as } n \to \infty$$

In fact it is true that

$$P\left(\lim_{n\to\infty} M_n = m\right) = 1 \tag{1.14}$$

if $X_1, X_2, \ldots$ are iid random variables with mean m. This result is known as the **Strong Law of Large Numbers**.

On the other hand, we know that

$$E[M_n] = m, \quad \text{Var}[M_n] = \sigma^2/n$$

but we know nothing about the distribution of M_n. It is true that

$$E[Z_n] = 0, \quad \text{Var}[Z_n] = 1$$

where

$$Z_n := \frac{M_n - m}{\sigma/\sqrt{n}}$$

The **Central Limit Theorem** states that if $X_1, \ldots, X_n$ are iid with mean m and variance σ^2, then

$$\lim_{n\to\infty} P(Z_n \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\xi^2/2} d\xi$$

uniformly in $x \in \mathbb{R}$.

Thus, the sample mean is given by

$$M_n = m + \frac{\sigma}{\sqrt{n}} Z_n,$$

where the normalized errors Z_n are "asymptotically Gaussian".

Chapter 2

Geostatistics in Hydrology: Kriging interpolation

Hydrologic properties, such as rainfall, aquifer characteristics (porosity, hydraulic conductivity, transmissivity, storage coefficient, etc.), effective recharge and so on, are all functions of space (and time) and often display a high spatial variability, also called heterogeneity. This variability is not in general random. It is a general rule that these properties display a so called "scale effect", i.e., if we take measurements at two different points the difference in the measured values dicreases as the two points come closer to each other.

It is convenient in certain cases to consider these properties as random functions having a given spatial structure, or in other word having a given spatial correlation, which can be conveniently described using appropriate statistical quantities. These variables are called "regionalized" variables [5].

The study of regionalized variables starts from the ability to interpolate a given field starting from a limited number of observation, but preserving the theoretical spatial correlation. This is accomplished by means of a technique called "kriging" developed by [4, 5] and largely applied in hydrology and other earth science disciplines [3, 6, 7, 1] for the spatial interpolation of various physical quantities given a number of spatially distributed measurments.

Although theoreticallly kriging cannot be considered superior to other surfave fitting techniques [7] and the use of a few arbitrary parameters may lead absurd results, this method is capable of obtaining "objective" interpolations evaluating at the same time the quality of the results.

In the next sections we discuss first the statistical hypothesis that are needed to develop the theory of krigin and then we proceed at describing the method under different assumptions and finally report a few applications.

2.1 Statistical assumptions

Let $Z(\vec{x}, \xi)$ be a random function (RF) simulating our hydrological quantity of interest. We recall that with this notation we imply that $\vec{x}$ denots a point in space (two or three dimensional space), while ξ denotes the state variable in the space of the realizations. In other words:

- $Z(\vec{x}_0, \xi)$ is a random variable at point $\vec{x}_0$ representing the entire set of realizations of the RF at point $\vec{x}_0$;
- $Z(\vec{x}, \xi_1)$ is a particular realization of the RF Z;
- $Z(\vec{x}_0, \xi_1)$ is a measurement point.

Given a set of sampled values $Z(\vec{x}_i, \xi_1),\ i = 1, 2, \ldots$ we want to reconstruct $Z(\vec{x}, \xi_1)$, i.e. a possible realization of $Z(\vec{x}, \xi)$.

2.2 Kriging for Weak or Second Order stationary RF

An RF is said to be second order stationary if:

1. Constant mean:
$$E[Z(\vec{x}, \xi)] = m \tag{2.1}$$

2. the autocovariance (another name for the covariance) is a function of the distance between the reference points $\vec{x}_1$ and $\vec{x}_2$:

$$\text{cov}[\vec{x}_1, \vec{x}_2] = E[(Z(\vec{x}_1, \xi) - m)(Z(\vec{x}_2, \xi) - m)] = C(\vec{h}) \qquad (2.2)$$

In practice second order stationarity implies that the first two statistical moments (expected value and covariance) be translation invariant. Note that by saying that $E[Z(\vec{x}, \xi)] = m$ we imply that effectively the expected value taken on all the possible realizations ξ does not vary with space. However, for a given realization $Z(\vec{x}, \xi_1)$ is a function of $\vec{x}$.

Because of (2.1), the covariance (2.2) can be written as:

$$\begin{aligned} \text{cov}[\vec{x}_1, \vec{x}_2] &= E[(Z(\vec{x}_1, \xi) - m)(Z(\vec{x}_2, \xi) - m)] \\ &= E[Z(\vec{x}_1, \xi)Z(\vec{x}_2, \xi)] - mE[Z(\vec{x}_2, \xi)] - E[Z(\vec{x}_1, \xi)]m + m^2 \\ &= E[Z(\vec{x}_1, \xi)Z(\vec{x}_2, \xi)] - m^2 \qquad (2.3) \end{aligned}$$

Obviously if $\vec{h} = 0$ we have the definition of variance, also called the "dispersion" variance:

$$C(0) = \text{var}[Z] = \sigma_Z^2$$

For simplicity of notation, from now on, we will denote $x = \vec{x}$, $h = \vec{h}$ and we will drop the variable ξ in the RF.

Case with m and $C(\vec{h})$ known If $[Z(x)] = m$ and $C(h)$ are known, then we can define a new variable $Y(x)$ with zero mean:

$$\begin{aligned} Y(x) &= Z(x) - m \\ E[Y(x)] &= 0 \end{aligned}$$

Given the observed values:

$$\begin{array}{cccc} x_1 & x_2 & \dots & x_n \\ Y_1 & Y_2 & \dots & Y_n \end{array}$$

with $Y_i = Y(x_i)$ being the observation at point x_i. we look for a linear estimator $Y^*(x_0)$ of $Y(x_0)$ at point x_0 using the observed values. The

form of the estimator is:

$$Y^*(x_0) = \sum_{i=1}^{n} \lambda_i Y_i \tag{2.4}$$

Note that the estimator (2.4) is a realization of the RF

$$Y^*(x_0, \xi) = \sum_{i=1}^{n} \lambda_i Y(x_i, \xi)$$

The weights λ_i are calculated by imposing that the statistical error

$$\epsilon(x_0) = Y(x_0) - Y^*(x_0)$$

has minimum variance:

$$\text{var}[\epsilon(x_0)] = E[(Y(x_0) - Y^*(x_0))^2] = \text{minimum}$$

Substituting eq. (2.4) in the expression of the variance we have:

$$\begin{aligned}
E[(Y(x_0) - Y^*(x_0))^2] &= E[(\sum_{i=1}^{n} \lambda_i Y_i - Y_0)^2] \\
&= E[(\sum_{i=1}^{n} \lambda_i Y_i - Y_0)(\sum_{i=1}^{n} \lambda_i Y_i - Y_0)] \\
&= E[(\sum_{i=1}^{n} \lambda_i Y_i)(\sum_{i=1}^{n} \lambda_i Y_i)] - 2E[\sum_{i=1}^{n} \lambda_i Y_i Y_0] + E[Y_0^2] \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j E[Y_i Y_j] - 2 \sum_{i=1}^{n} E[Y_i Y_0] + E[Y_0^2]
\end{aligned}$$

but, since $m = 0$

$$E[Y_i Y_j] = C(x_i - x_j) + m^2 = C(x_i - x_j)$$

and

$$E[Y_0^2] = C(0) = \text{var}[Y]$$

is the dispersion variance of Y. Then:

$$E[(Y(x_0) - Y^*(x_0))^2] = \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j C(x_i - x_j) - 2\sum_{i=1}^{n}\lambda_i C(x_i - x_0) + C(0)$$

The minimum is found by setting to zero the first partial derivatives:

$$\frac{\partial}{\partial\lambda_i}\left(E[(Y(x_0) - Y^*(x_0))^2]\right) = 2\sum_{j=1}^{n}\lambda_j C(x_i - x_j) - 2C(x_i - x_0) = 0 \qquad j = 1, \ldots, n$$

This yields a linear system of equations:

$$C\lambda = b \tag{2.5}$$

where matrix C is given by:

$$C = \begin{bmatrix} C(0) & C(x_1 - x_2) & \ldots & c(x_1 - x_n) \\ \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \\ C(x_n - x_1) & & & C(0) \end{bmatrix}$$

and the right hand side vector b is given by:

$$b = \begin{bmatrix} C(x_1 - x_0) \\ \cdot \\ \cdot \\ \cdot \\ C(x_n - x_0) \end{bmatrix}$$

Matrix C is the spatial covarianve matrix and does not depend upon x_0. It can be shown that if all the x_j's are distinct then C is positive definite, and thus the linear system (2.5) can be solved with either direct or iterative methods. Once the solution vector λ is obtained, equation (2.4) yields the estimation of our regionalized variable at point x_0. Thus the

calculated value for λ is actually function of the estimation point x_0. If we want to change the estimation point x_0, for example if we need to obtain a spatial distribution of our regionalized variable, we need to solve the linear system (2.5) for different values of x_0. In this case it is convenient to factorize matrix C using Cholseky decomposition and then proceed to the solution for the different right hand side vectors.

Evaluation of the estimation variance The estimation variance is defined as the variance of the error:

$$\epsilon = Y_0^* - Y_0$$

Hence:

$$\text{var}[Y_0^* - Y_0] = E[(Y_0^* - Y_0)^2] - E[(Y_0^* - Y_0)]^2$$

but:

$$E[(Y_0^* - Y_0)] = E[Y_0^*] - E[Y_0] = \sum_i \lambda_i E[Y_i] - E[Y_0] = 0$$

$$(E[Y] = m = 0)$$

and since

$$\sum_j \lambda_j C(x_i - x_j) = C(x_i - x_0)$$

we finally obtain:

$$\begin{aligned} \text{var}[Y_0^* - Y_0] &= E[(Y_0^* - Y_0)^2] \\ &= \sum_i \sum_j \lambda_i \lambda_j C(x_i - x_j) - 2 \sum_i \lambda_i C(x_i - x_0) + C(0) \\ &= -\sum_i \lambda_i C(x_i - x_0) + C(0) \\ &= \text{var}[Y] - \sum_i \lambda_i C(x_i - x_0) \end{aligned}$$

which, since $\sum_i \lambda_i C(x_i - x_0) > 0$, shows that the estimation variance of Y_0 is smaller than the dispersion variance of Y (the real variance of

the RF). In statistical terms, we can interpret this result by saying that since we have observed Y at some points x_i then the uncertainty on Y decreases.

It is important to remark the difference between estimation and dispertion variance. The latter is representative of the variation interval of the RF Y within the interpolation domain, while the estimation variance represents the residual uncertainty in the estimation of the realization Y_0^* of Z when n observations are available. The dispersion variance is a constant, while the estimation variance varies from point to point and is zero at the observation points.

Our original variable was $Z = Y + m$ and its estimate is thus:

$$\begin{aligned} Z_0^* &= m + \sum_i \lambda_i (Z_i - m) \\ \text{var}[Z_0^* - Z_0] &= \text{var}[Z_0] - \sum_i \lambda_i C(x_i - x_0) \end{aligned}$$

2.3 Kriging with the intrinsic hypothesis

The hypothesis of second order stationarity of the RF is not always satisfied, for example $C(0)$ increases with the distance, violating hypothesis (2.2). In this case the "Intrinsic hypothesis" must be used, in which we assume that the first order increments

$$\delta = Y(x+h) - Y(x)$$

are second order RF:

$$E[Y(x+h) - Y(x)] = m(h) = 0 \tag{2.6}$$

$$\text{var}[Y(x+h) - Y(x)] = 2\gamma(h) \tag{2.7}$$

where the function $\gamma(h)$ is called the variogram. If the mean $m(h)$ is not zero an obvious change of variable is required. The variogram is defined as the mean quadratic increment of $Y(x)$ (divided by 2) for any two points x_i and x_j separated by a distance h:

$$\gamma(h) = \frac{1}{2}\text{var}[Y(x+h) - Y(x)] = \frac{1}{2}E[(Y(x+h) - Y(x))^2] \tag{2.8}$$

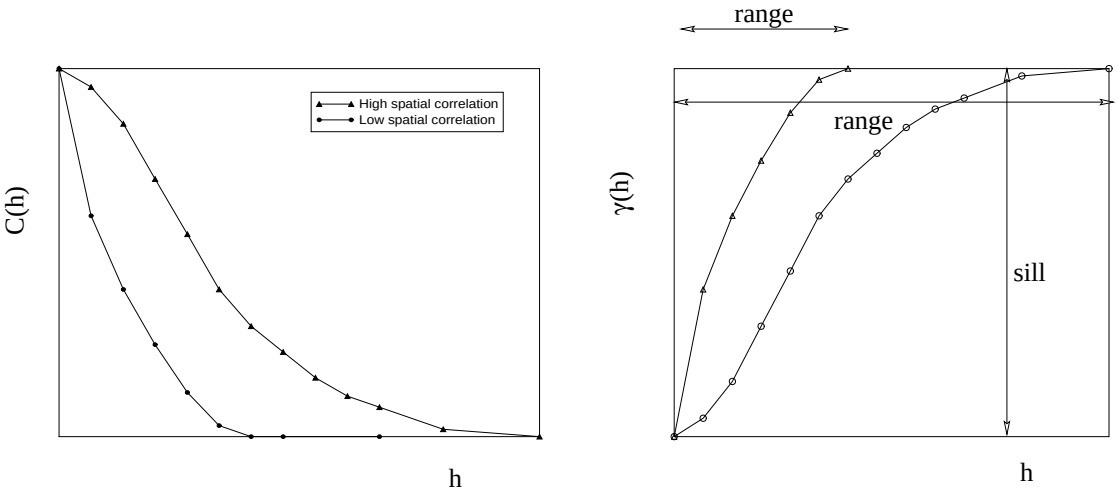


Figure 2.1: Behavior of the covariance as a function of distance (left) and the corresponding variogram (right).

and is related to the covariance function by:

$$\begin{aligned}\gamma(h) &= \frac{1}{2}E[(Y(x+h)-Y(x))^2] \\ &= \frac{1}{2}E[Y^2(x+h)] - E[Y(x+h)Y(x)] + \frac{1}{2}E[Y^2(x)] \\ &= C(0) - C(h)\end{aligned}$$

The intrinsic hypothesis requires a finite value for the mean of $Y(x)$ but not for its variance. In fact, hypothesis (2.2), as changed into (2.3), implies (2.8), but not viceversa.

The covariance $C(h)$ has a decreasing behavior as shown in Fig. 2.1. When $C(h)$ is known then the variogram can be directly calculated. When $C(0)$ is finite, the variogram $\gamma(h)$ is bounded asymptotically by this value. The value of h at which the asymptot can be considered achieved is called the "range", while $C(0)$ is called the "sill" (see Fig. 2.1).

2.3.1 The variogram

The variogram is usually calculated from the experimental observations, and describes the spatial structure of the RF. It can be shown [5] that if $x_0, x_1, \ldots, x_n$ are $n+1$ points belonging to the domain of interpolation and the coefficients $\lambda_0, \lambda_1, \ldots, \lambda_n$ satisfy:

$$\sum_{i=0}^{n} \lambda_i = 0$$

then $\gamma(h)$ has to satisfy:

$$-\sum_{i=0}^{n}\sum_{j=0}^{n} \lambda_i \lambda_j \gamma(x_i - x_j) \geq 0$$

and

$$\lim_{|h|\to\infty} \frac{\gamma(h)}{h^2} = 0$$

i.e. $\gamma(h)$ is tends to infinity slowlier thatn $|h|^2$ as $|h| \to \infty$.

In principle there $\gamma(h)$ could assume different behaviors also with the direction of vector h ("anisotropy"), but this is in general not easily verifiable due to the limited number of data points usually available for hydrologic variables. If the experimental variogram displays anisotropy, then the intrinsic hypothesis is not verified and one has to use the so called "universal" kriging [2, 3].

The most commonly used isotropic variograms are shown in Fig. 2.2 and are of the form:

1. polinomial variogram:

$$\gamma(h) = \omega h^{\alpha} \qquad\qquad 0 < \alpha < 2$$

2. exponential variogram:

$$\gamma(h) = \omega \left[1 - e^{-\alpha h}\right]$$

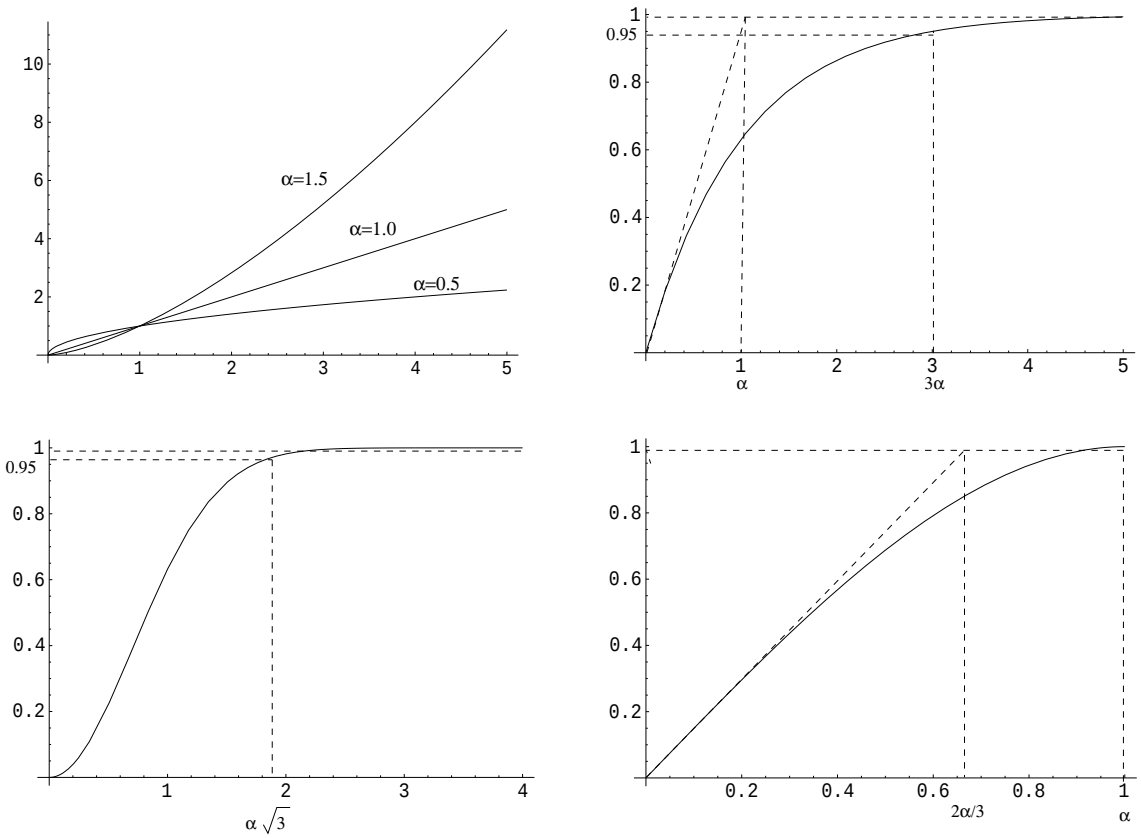


Figure 2.2: Behavior of the most commmonly used variograms: polinomial (top left); exponential (top right); gaussian (bottom left); spherical (bottom right)

3. gaussian variogram:

$$\gamma(h) = \omega \left[1 - e^{-(\alpha h)^2}\right]$$

4. spherical variogram:

$$\gamma(h) = \begin{cases} \frac{1}{2}\omega \left[\frac{3h}{\alpha} - \left(\frac{h}{\alpha}\right)^3\right] & h \leq \alpha \\ \omega & h > \alpha \end{cases}$$

where ω and α are real constants.

The variogram is estimated from the available observations in the following manner. The data points are subdivided into a prefixed number of classes based on the distances between the measurement locations. For each pair i and j of points and for each class calculate:

1. the number M tha fall within tha class;

2. the average distance of the class;

3. the half of the mean quadratic increment

$$\frac{1}{2}\sum (Y_i - Y_j)^2 / M$$

In general the pairs are not uniformly distributed among the different classes as usually there are more pairs for the smaller distances. Thus the experimental variogram will be less meaninful as h increases. A best fit procedure together with visual inspection is then used to select the most appropriate variogram and evaluate its optimal parameters.

Remark 1. An experimental variogram that is not bounded above (for example the polinomial variogram with $\alpha \geq 1$) implies an infinite variance, and thus the covariance does not exist. Only the intrinsic hypothesis is acceptable. If the variogram achieves a "sill" then the phenomenon has a finite variance and the covariance exists.

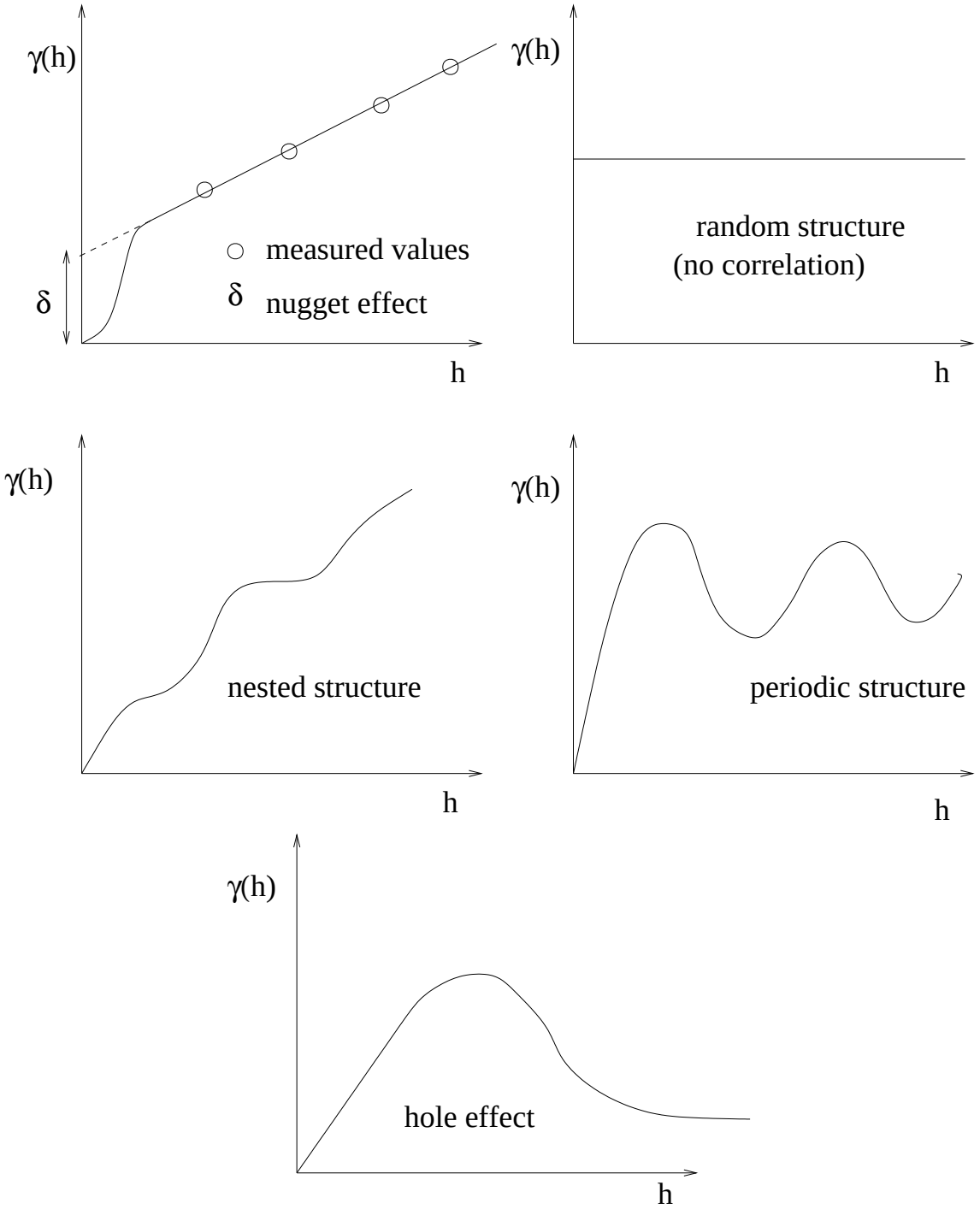


Figure 2.3: Example of typical spatial correletion structures that may be encountered in analyzing measured data

Remark 2. For every variogram, $\gamma(0) = 0$. However sometimes the data may display a jump at the origin. This apparent discontinuity is called the "nugget effect" and can be due to measurement errors or to microregionalization effects that are not evidenced at the scale of the actual data points. If the nugget effect is present the the variogram needs to be changed into:

$$\gamma(h) = \delta + \gamma_0(h)$$

where δ is the jump at the origin and $\gamma_0(h)$ the variogram without the jump. In Fig. 2.3 we report a variogram with the nugget effect together with some other special cases that may be encountered.

2.4 Remarks about Kriging

a. Kriging is a BLUE (Best Linear Unbiased Estimator) interpolator. In other word it is a Linear estimator that matches the correct expected value of the population (Unbiased) and that minimizes the variance of the observations (Best).

b. Kriging is an exact interpolator if no errors are present. In fact, if we set $x_0 = x_i$ in (2.5) we obtain immediately $\lambda_i = 1, \lambda_j = 0, j = 1, \ldots, n, j \neq i$.

c. If we assume that the the error ϵ is Gaussian, then we can associate to the estimate $Y^*(x_0)$ a confidence interval. For example, the 95% confidence interval is $\pm 2\sigma_0$ where:

$$\sigma_0 = \sqrt{\text{var}[Y^*(x_0) - Y(X_0)]}$$

Then the krigin estimator (2.4 becomes:

$$Y^*(x_0) = \sum_{i=1}^{n} \lambda_i Y_i$$

d. The solution of the linear system does not depend on the observed value but only on x_i and x_0.

e. A map of the estimated regionalized variable, and possibly its confidence intervals, can be obtained be defining a grid and solving the linear system for each point in the grid.

2.5 Kriging with uncertainties

We now assume that the observations Y_i are affected by measurement errors ϵ_i, and that:

1. the errors ϵ_i have zero mean:

$$E[\epsilon_i] = 0 \qquad\qquad i = 1, \ldots, n$$

2. the errors are uncorrelated:

$$\text{cov}[\epsilon_i, \epsilon_j] = 0 \qquad\qquad i \neq j$$

3. the errors are not correlated with the RF:

$$\text{cov}[\epsilon_i, Y_i] = 0$$

4. the variance σ_i^2 of the errors is a known quantity and can vary from point to point.

The new coefficient matrix C of the linear system (2.5) is changed by adding to the main diagonal the quantity $-\sigma^2)i$:

$$C + \begin{bmatrix} \sigma_1^2 & \cdot & 0 \\ 0 & \cdot & 0 \\ 0 & \cdot & \sigma_n^2 \end{bmatrix}$$

and everything proceeds as in the standard case.

2.6 Validation of the interpolation model

The chosen model (in practice the variogram) can be validated by interpolating observed values. If n observations $Y(x_i), i = 1, \ldots, n$ are available, the validation process proceeds as follows:

For each j, $j = 1, \ldots, n$:

- discard point $(x_j, Y(x_j))$;
- estimate the $Y^*(x_j)$ by solving the kriging system having set $x_0 = x_j$ and using the remaining points $x_i, i \neq j$ for the interpolation;
- evaluate the estimation error $\epsilon_j = Y_j^* - Y_j$,

The chosen model can be considered theoretically valid if the error distribution is approximately gaussian with zero mean and unit variance ($N(0, 1)$, i.e. satisfies the following:

1. there is no bias:
$$\frac{1}{n} \sum_{i=1}^{n} \epsilon_i \approx 0$$
2. the estimation variance σ_i is coherent with the error standard deviation:
$$\frac{1}{n} \sum_{i=1}^{n} \left(\frac{Y_i^* - Y_i}{\sigma_i} \right)^2 = 1$$

One can also look at the behavior of the interpolation error at each point looking at the mean square error of the vector ϵ:

$$Q = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \epsilon_i^2}$$

The uncertainties connected to the choice of the theoritcal variogram from the experimental data can be minimized by anaylizing the validation

test. In fact, among all the possible variograms $\gamma(h)$, that close to the origin display a slope compatible with the obesrvations and gives rise to a theoretically coherent model, one can choose the variogram with the smallest value of Q.

2.7 Computational aspects

In the validation phase n linear systems of dimension $n-1$ need to be solved. The system matrices are obtained by dropping one row and one complumn of the complete kriging matrix. This can be efficiently accomplished by means of intersections of $n-1$-dimensional lines with apporpriate coordinate n-dimensional planes.

Note that the krigin matrix C is symmetric, and thus its eigenvalues λ_i are real. However, since

$$\sum_{i=1}^{n} \lambda_i = \mathrm{Tr}(C) = \sum_{i=1}^{n} c_{ii} = 0$$

where $\mathrm{Tr}(C)$ is the trace of matrix C, it follows that some of the eigenvalues must be negative and thus C is not positive definite. For this reason, the solution of the linear systems is usually obtained by means of direct methods, such as Gaussian elimination or Choleski decomposition. Full Pivoting is often necessary to maintain stability of the algorithm.

2.8 Kriging with moving neighborhoods

Generally the experimental variogram is most accurate for small values of h, with uncertainties growing rapidly when h is large. The influence of this problem may be decreased by using moving neighborhoods. With this variant, only the points that lie within a prefixed radios R from point x_0 are considered, provided that we are left with an adequate number of data (Fig. 2.4). The redius R is selected so that the lag h will remain with the range of maximum certainty for $\gamma(h)$.

This approach leads also to high saving in the computational cost of the procedure because each linear system is now much smaller (Gaussian

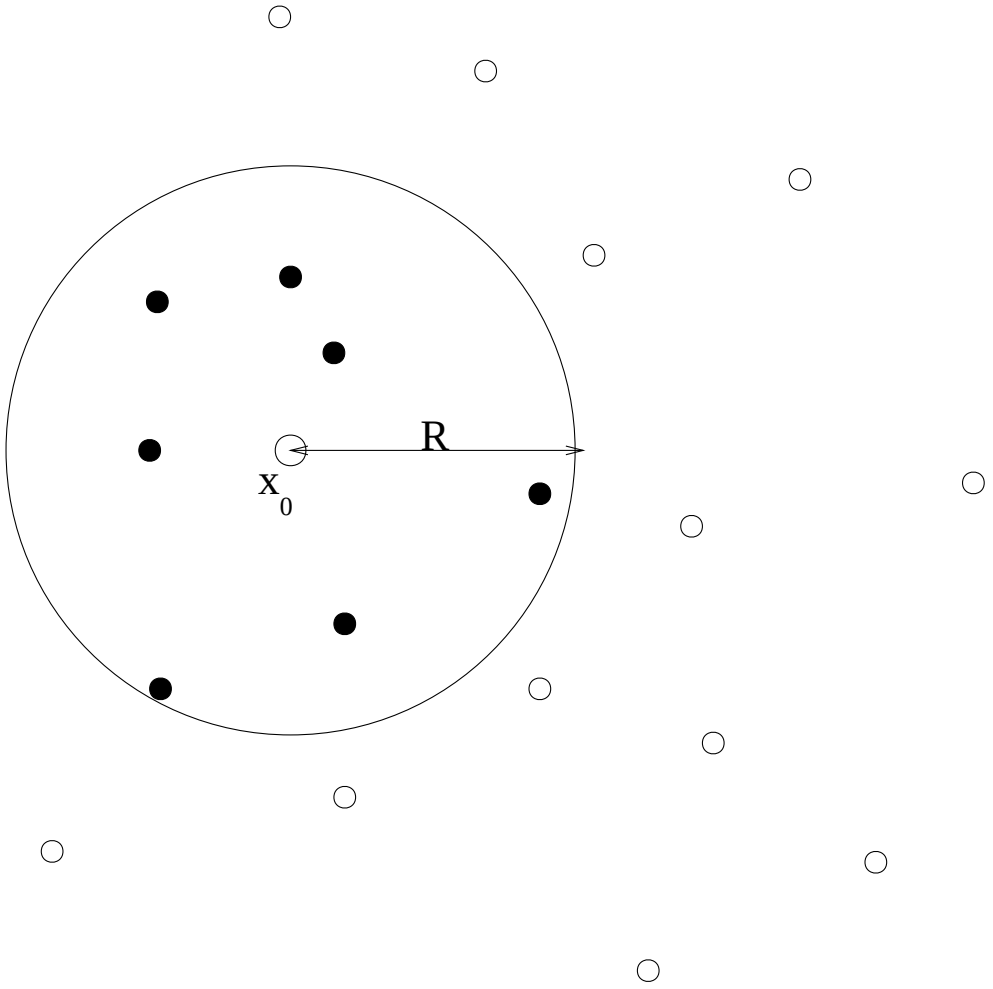


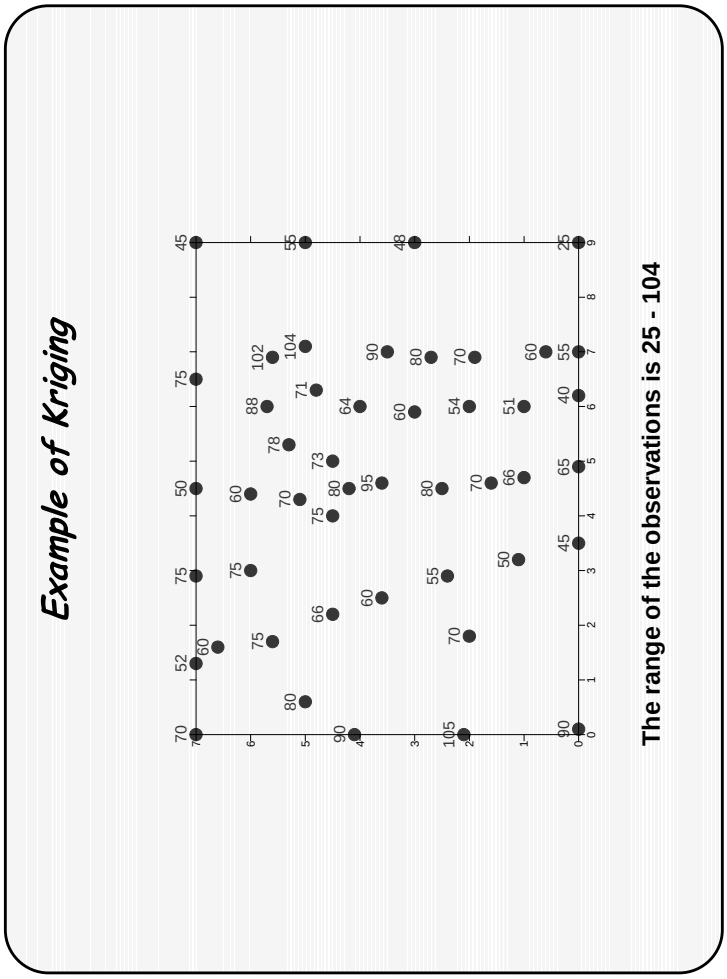
Figure 2.4: Example of interpolation with moving neighborhood

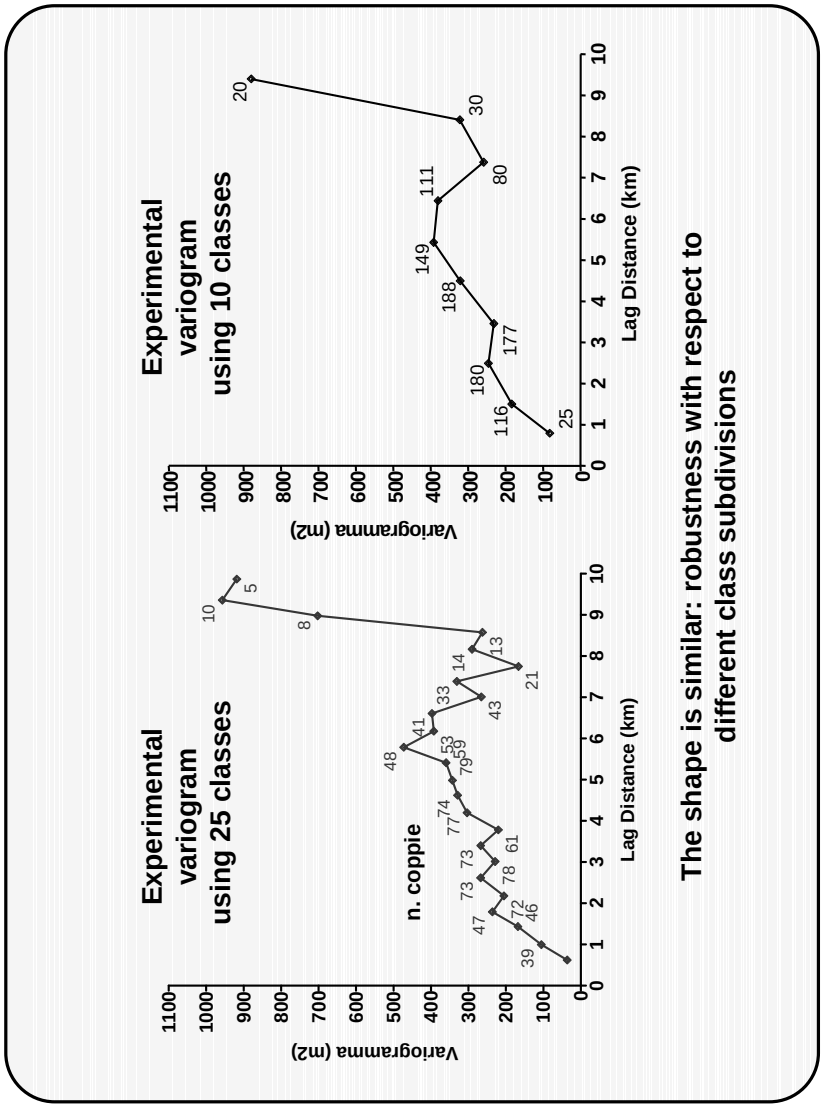
elimination has a computational cost proportional to n^3, where n is the number of equations).

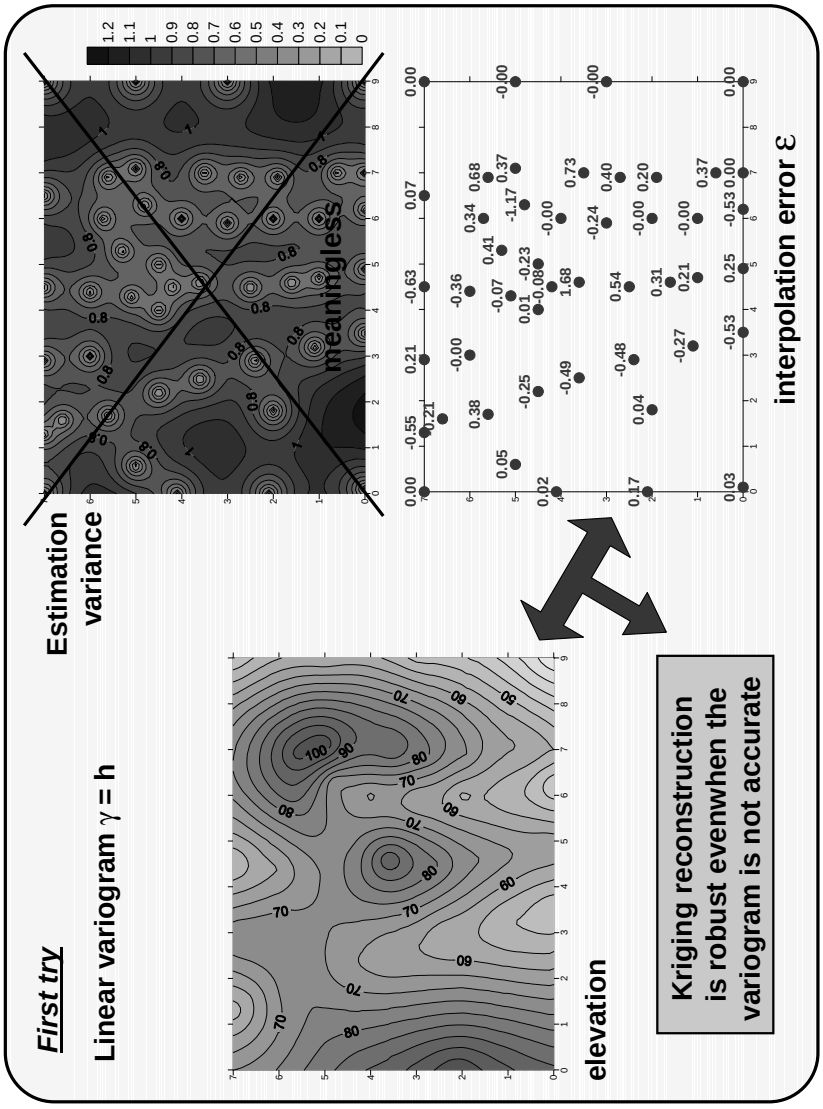
2.9 Detrending

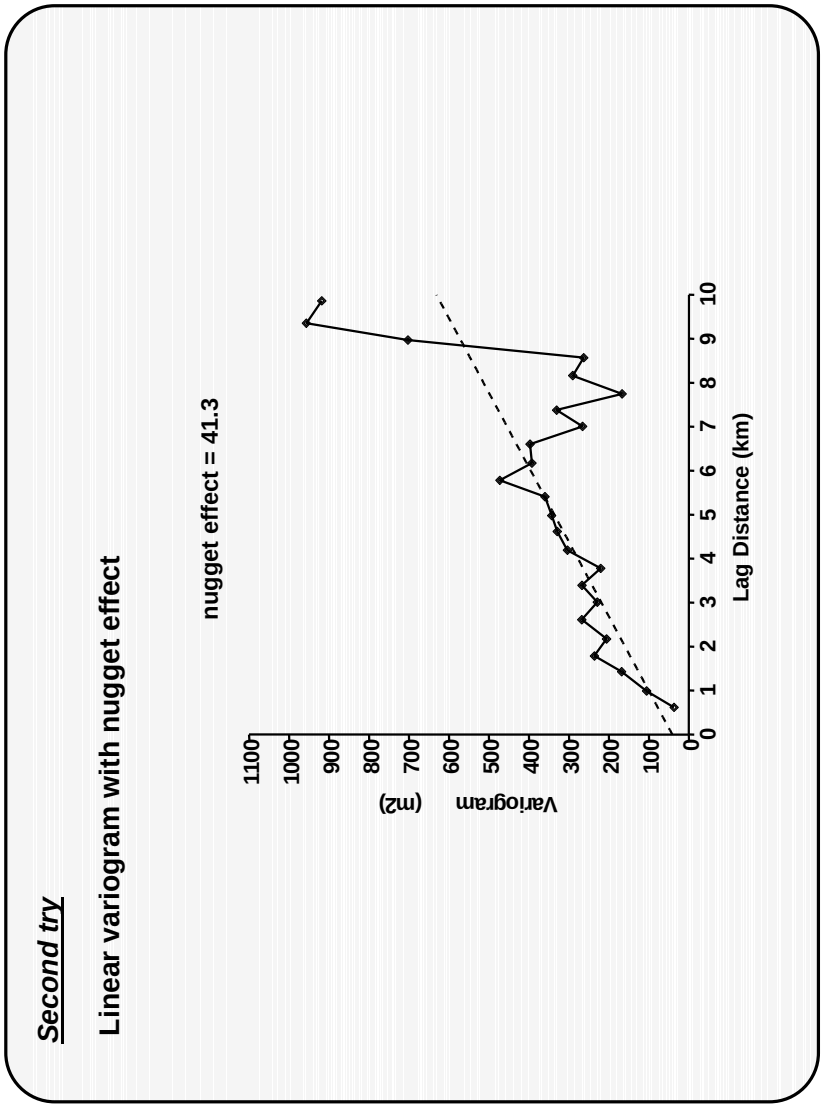
In certain cases the observations display a definite trend that needs to be taken into account. This is the case, for example, when one needs to interpolate piezometric heads observed from wells in an aquifer system where a regional gradient is present. To remove the trend from the data it is possible to work with residuals (= measurements - trend) that have a constant mean. However this detrending procedure may be dangerous as it may introduce a bias in the results. For this reason it is important tha the trend be recognized not only from the raw data but also from the physical behavior of the system from which the data come. If the trend cannot be removed from the observations by simple subtraction, the universal kriging approach [2, 3] can be used.

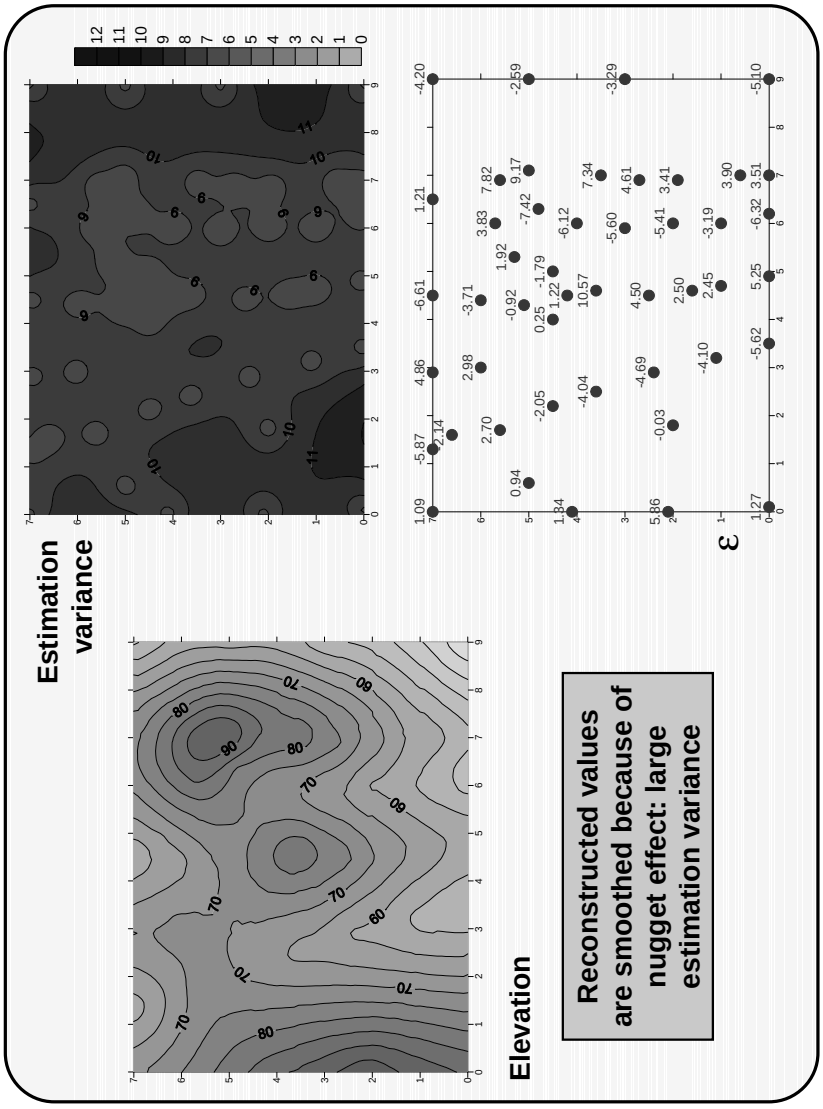
2.10 Example of application of kriging for the reconstruction of an elevation map

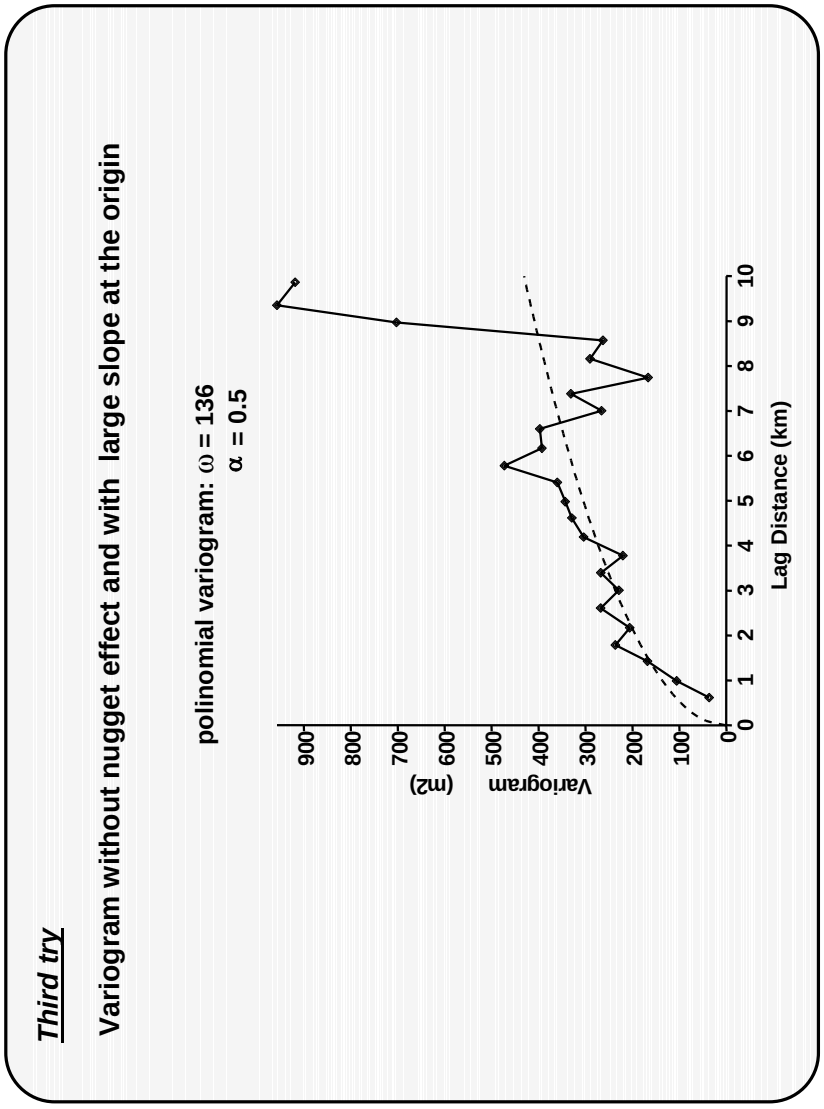


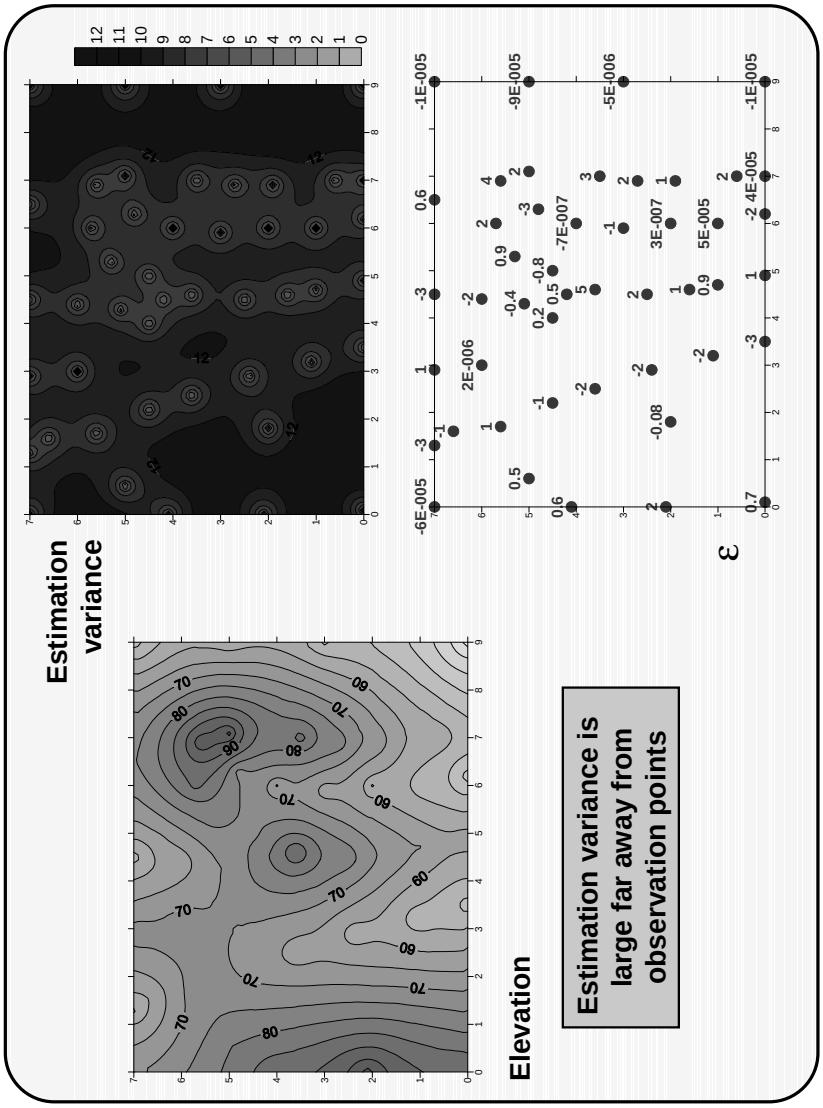

Experimental variogram using 25 classes
n. coppie
Variogramma (m2)
Lag Distance (km)
Experimental variogram using 10 classes
Variogramma (m2)
Lag Distance (km)
The shape is similar: robustness with respect to different class subdivisions

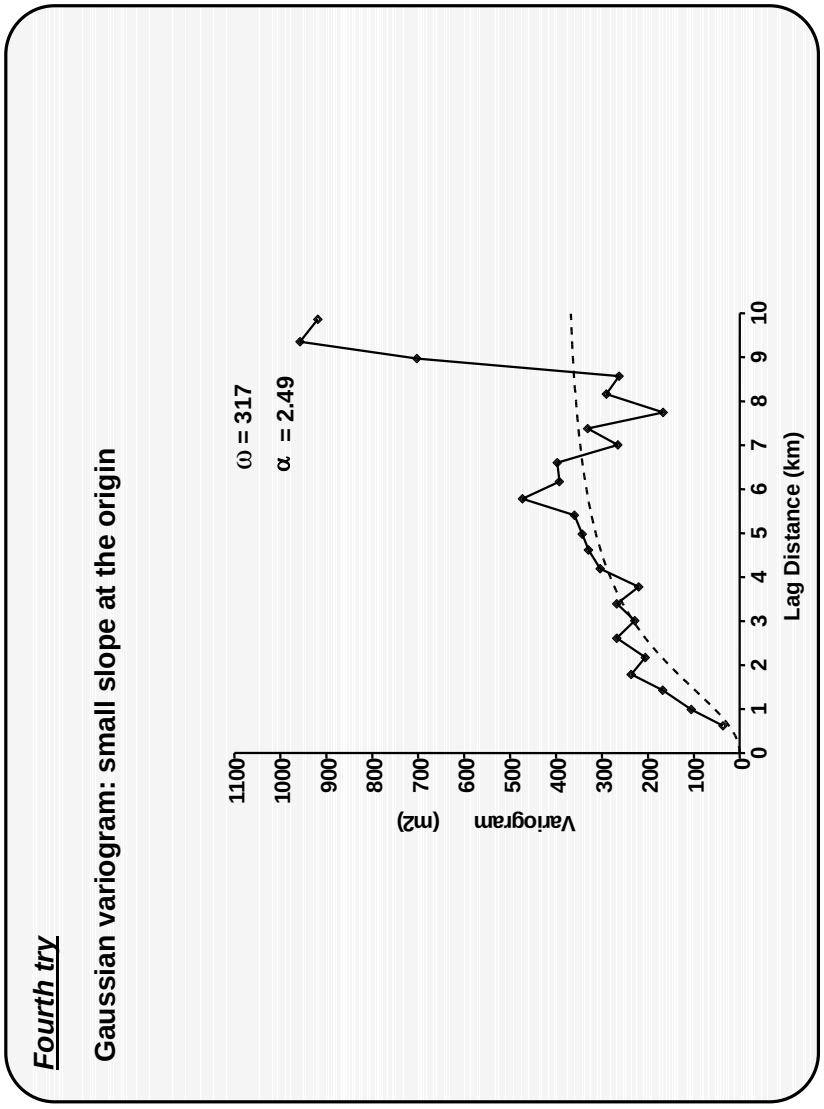

First try
Linear variogram γ = h
Estimation variance
meaningless
elevation
interpolation error ε
Kriging reconstruction is robust evenwhen the variogram is not accurate

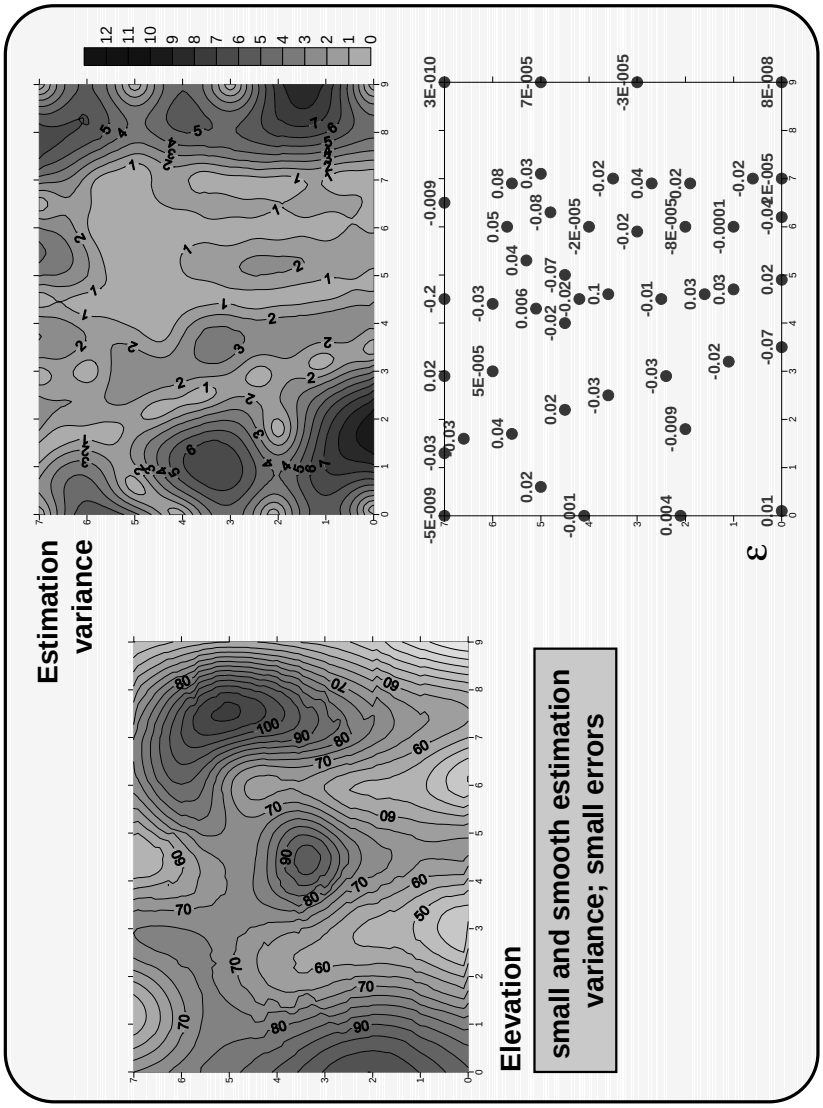

Second try
Linear variogram with nugget effect
nugget effect = 41.3
Variogram (m2)
1100
1000
900
800
700
600
500
400
300
200
100
0
0 1 2 3 4 5 6 7 8 9 10
Lag Distance (km)


Estimation variance
Elevation
Reconstructed values are smoothed because of nugget effect: large estimation variance
ε


Third try
Variogram without nugget effect and with large slope at the origin
polinomial variogram: ω = 136
α = 0.5
Variogram (m2)
900
800
700
600
500
400
300
200
100
0
0 1 2 3 4 5 6 7 8 9 10
Lag Distance (km)


Estimation variance
Elevation
Estimation variance is large far away from observation points
ε


Fourth try
Gaussian variogram: small slope at the origin
ω = 317
α = 2.49
Variogram (m2)
Lag Distance (km)
0
100
200
300
400
500
600
700
800
900
1000
1100
0
1
2
3
4
5
6
7
8
9
10


Estimation variance
Elevation
small and smooth estimation variance; small errors
ε

Bibliography

[1] de Marsily, G. (1984). Spatial variability of properties in porous media: A stochastic approach. In J. Bear and M. Corapcioglu, editors, *Fundemantals of Transport Phenomena in Porous Media*, pages 721–769, Dordrecht. Martinus Nijhoff Publ.

[2] Delhomme, J. P. (1976). *Applications de la thèorie des varaibles règionalisèes dans les sciences de l'eau*. Ph.D. thesis, Ecoles des Mines, Fontainebleaus (France).

[3] Delhomme, J. P. (1978). Kriging in the hydrosciences. *Adv. Water Resources*, **1**, 251–266.

[4] Matheron, G. (1969). Le krigeage universel. Technical Report 1, Paris School of Mines. Cah. Cent. Morphol. Math., Fontainbleau.

[5] Matheron, G. (1971). The theory of regionalized variables and its applications. Technical Report 5, Paris School of Mines. Cah. Cent. Morphol. Math., Fontainbleau.

[6] Volpi, G. and Gambolati, G. (1978). On the use of a main trend for the kriging technique in hydrology. *Adv. Water Resources*, **1**, 345–349.

[7] Volpi, G., Gambolati, G., Carbognin, L., Gatto, P., and Mozzi, G. (1979). Groundwater contour mapping in venice by stochastic interpolators. 2. results. *Water Resour. Res.*, **15**, 291–297.