Chapter 1

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 surface 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 kriging and then we proceed at describing the method under different assumptions and finally report a few applications.

1.1 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 $\vec{x}$ denotes 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 $Z(\vec{x}, \xi)$.

1.2 Kriging for Weak or Second Order stationary RF

An RF is said to be second order stationary if:

1. Constant mean:
$$E[Z(\vec{x}, \xi)] = m \tag{1.1}$$
2. the autocovariance (another name for the covariance) is a function of the distance between the reference points $\vec{x}_1$ and $\vec{x}_2$:
$$\text{cov}[\vec{x}_1, \vec{x}_2] = E[(Z(\vec{x}_1, \xi) - m)(Z(\vec{x}_2, \xi) - m)] = C(\vec{h}) \tag{1.2}$$

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 (1.1), the covariance (1.2) can be written as:

$$\begin{aligned} \text{cov}[\vec{x}_1, \vec{x}_2] &= E[(Z(\vec{x}_1, \xi) - m)(Z(\vec{x}_2, \xi) - m)] \\ &= E[Z(\vec{x}_1, \xi)Z(\vec{x}_2, \xi)] - mE[Z(\vec{x}_2, \xi)] - E[Z(\vec{x}_1, \xi)]m + m^2 \\ &= E[Z(\vec{x}_1, \xi)Z(\vec{x}_2, \xi)] - m^2 \end{aligned} \tag{1.3}$$

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 $E[Z(x)] = m$ 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:

$$Y^*(x_0) = \sum_{i=1}^{n} \lambda_i Y_i \tag{1.4}$$

Note that the estimator (1.4) 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 zero expecetd value and minimum variance:

$$\begin{aligned} E[\epsilon(x_0)] &= 0 \\ \text{var}[\epsilon(x_0)] &= E[(Y(x_0) - Y^*(x_0))^2] = \text{minimum} \end{aligned}$$

Substituting eq. (1.4) in the expression of the variance we have:

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

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:

$$C\lambda = b \tag{1.5}$$

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 covariance 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 (1.5) can be solved with either direct or iterative methods. Once the solution vector λ is obtained, equation (1.4) 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 (1.5) for different values of x_0. In this case it is convenient to factorize matrix C using Cholesky 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:

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

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 dispersion 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:

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

1.3 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 (1.2). In this case the "Intrinsic hypothesis" must be used, in which we assume that the first order increments

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

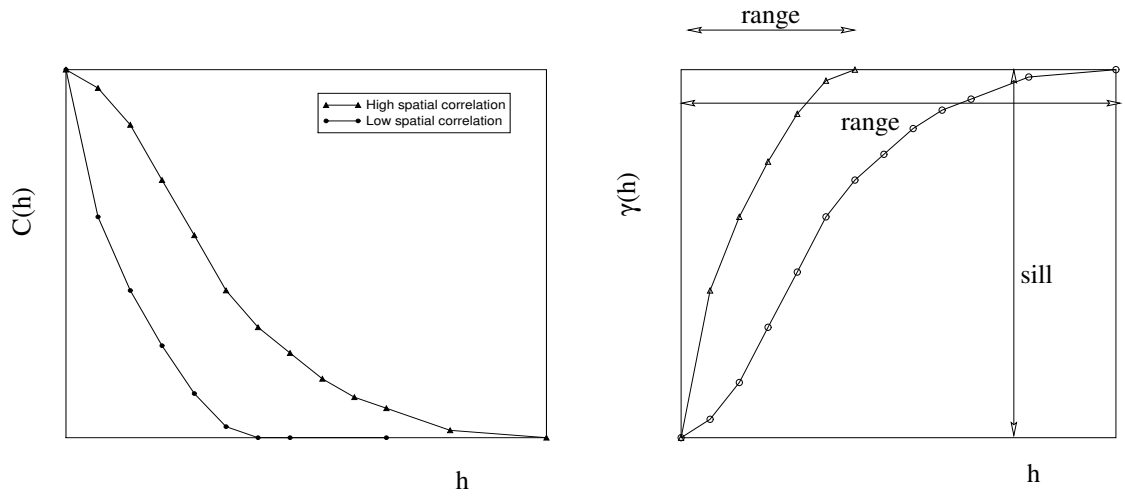


Figure 1.1: Behavior of the covariance as a function of distance (left) and the corresponding variogram (right).

are second order RF:

$$E[Y(x+h) - Y(x)] = m(h) = 0 \tag{1.6}$$
$$\text{var}[Y(x+h) - Y(x)] = 2\gamma(h) \tag{1.7}$$

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:

$$\gamma(h) = \frac{1}{2}\text{var}[Y(x+h) - Y(x)] = \frac{1}{2}E[(Y(x+h) - Y(x))^2] \tag{1.8}$$

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 (1.2), as changed into (1.3), implies (1.8), but not viceversa.

The covariance $C(h)$ has a decreasing behavior as shown in Fig. 1.1. 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. 1.1).

1.3.1 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 given $x_1, \ldots, x_n$ are n points belonging to the domain of interpolation, for all the coefficients $\eta_1, \ldots, \eta_n \in \mathbb{R}$ satisfing:

$$\sum_{i=1}^{n} \eta_i = 0$$

then

$$-\sum_{i=1}^{n}\sum_{j=0}^{n} \eta_i \eta_j \gamma(x_i - x_j) \geq 0$$

i.e. $-\gamma(h)$ is a positive semidefinite function, and

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

i.e. $\gamma(h)$ is tends to infinity slowlier than $|h|^2$ 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. 1.2 and are of the form:

1. polinomial variogram:
$$\gamma(h) = \omega h^\alpha \qquad\qquad 0 < \alpha < 2$$

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

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 of points that fall within the class;

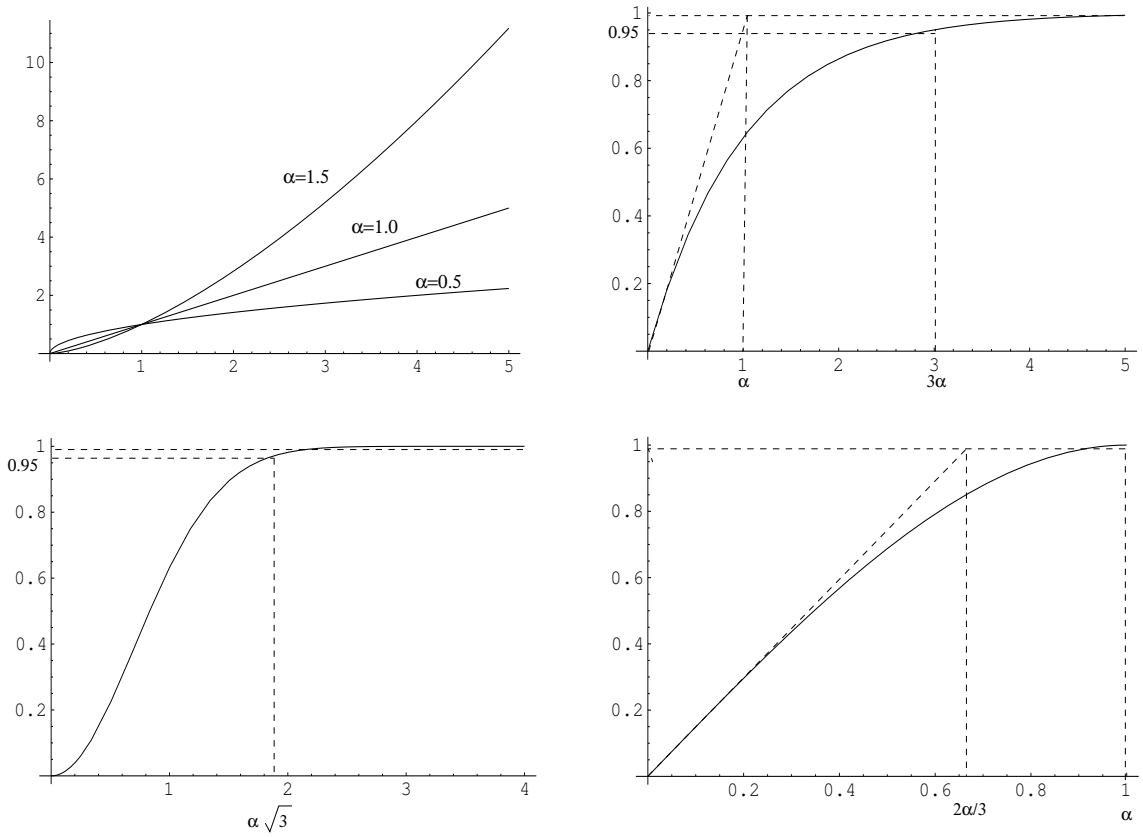


Figure 1.2: Behavior of the most commmonly used variograms: polinomial (top left); exponential (top right); gaussian (bottom left); spherical (bottom right)

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 meaningful 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.

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 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. 1.3 we report a variogram with the nugget effect together with some other special cases that may be encountered.

Remark 3. When we look for an estimator of the type

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

from the intrinsic hypothesis, i.e. eq. (1.6), we see that our random variable $Y(x)$ must have a constant mean m, so that:

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

and, since $E[Y(x_i)] = E[Y_i] = m$, we readily obtain:

$$\sum_{i=1}^{n} \lambda_i = 1$$

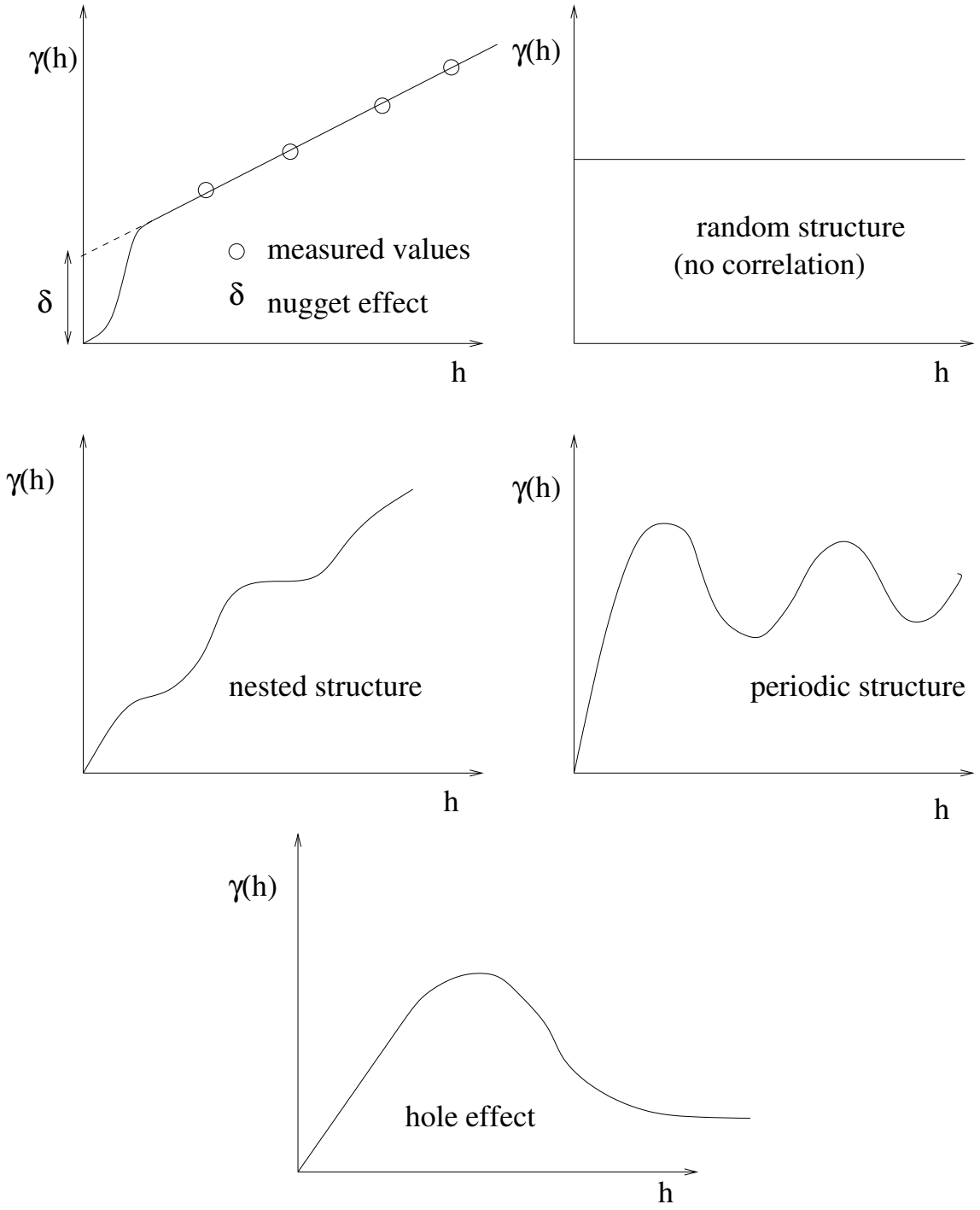


Figure 1.3: Example of typical spatial correletion structures that may be encountered in analyzing measured data

The condition of minimum variance becomes now a constrained minimization problem which can be solve by introducing a lagrange multiplier μ. The system (1.5) must be rewritten as:

$$\begin{bmatrix} 0 & \gamma(x_1 - x_2) & \dots & \gamma(x_1 - x_n) & 1 \\ \vdots & & & & \vdots \\ \gamma(x_n - x_1) & & & & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma(x_1 - x_0) \\ \vdots \\ \gamma(x_n - x_0) \\ 1 \end{bmatrix}$$

1.4 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 (1.5) we obtain immediately $\lambda_i = 1, \lambda_j = 0, j = 1, \dots, n, j \neq i$.

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 kriging estimator (1.4) becomes:

$$Y^*(x_0) = \sum_{i=1}^{n} \lambda_i Y_i \pm 2\sigma_0$$

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.

1.5 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:

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

2. the errors are uncorrelated:

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

3. the errors are not correlated with the RF:

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

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 (1.5) is changed by adding to the main diagonal the quantity $-\sigma_i^2$:

$$C := C - \begin{bmatrix} \sigma_1^2 & \cdot & 0 \\ 0 & \cdot & 0 \\ 0 & \cdot & \sigma_n^2 \end{bmatrix}$$

and everything proceeds as in the standard case.

1.6 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 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:

$$\frac{1}{n} \sum_{i=1}^{n} \left(\frac{Y_i^* - Y_i}{\sigma_i} \right)^2 \approx 1$$

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 theoretical 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 observations and gives rise to a theoretically coherent model, one can choose the variogram with the smallest value of Q.

1.7 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 column of the complete kriging matrix. This can be efficiently accomplished by means of intersections of $n-1$-dimensional lines with appropriate coordinate n-dimensional planes.

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

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

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.

1.8 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 radius R from point x_0 are considered, provided that we are left with an adequate number of data (Fig. 1.4). The radius R is selected so that the lag h will remain within 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 elimination has a computational cost proportional to n^3, where n is the number of equations).

1.9 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 that the trend be recognized not only from the raw data but

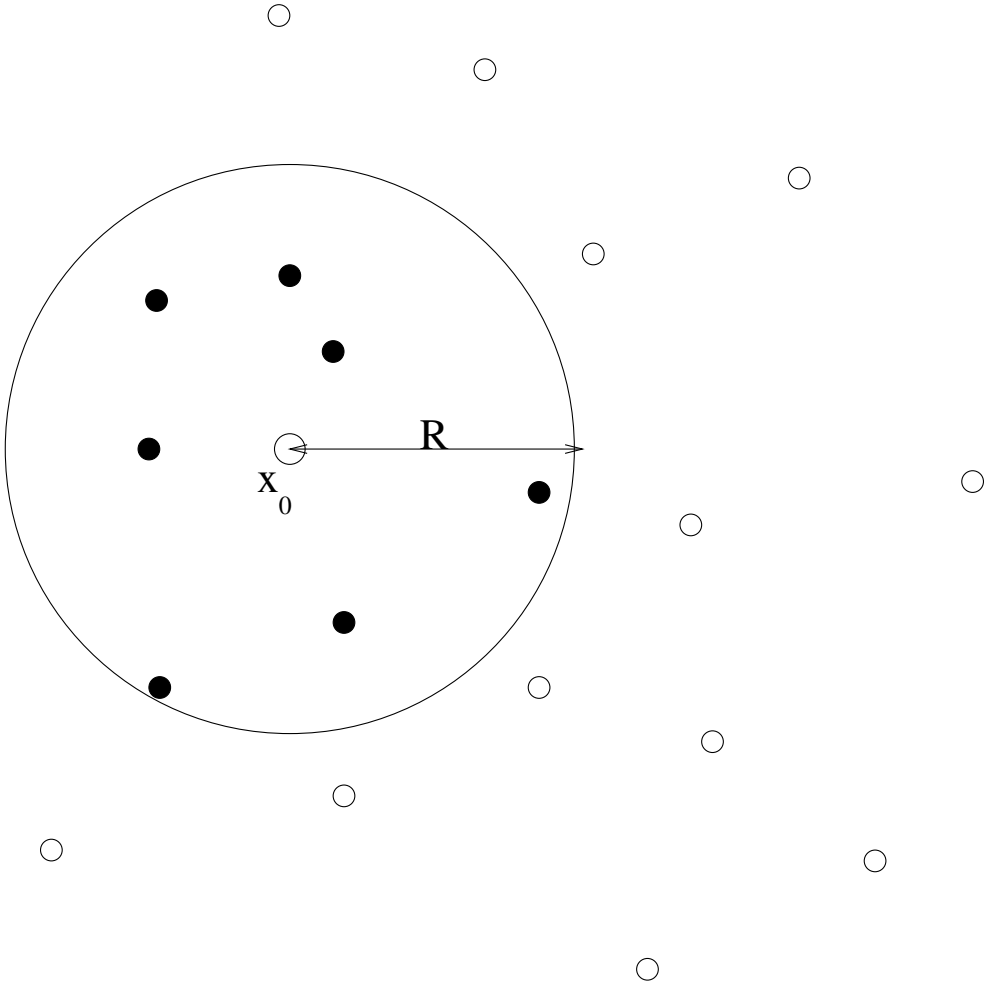
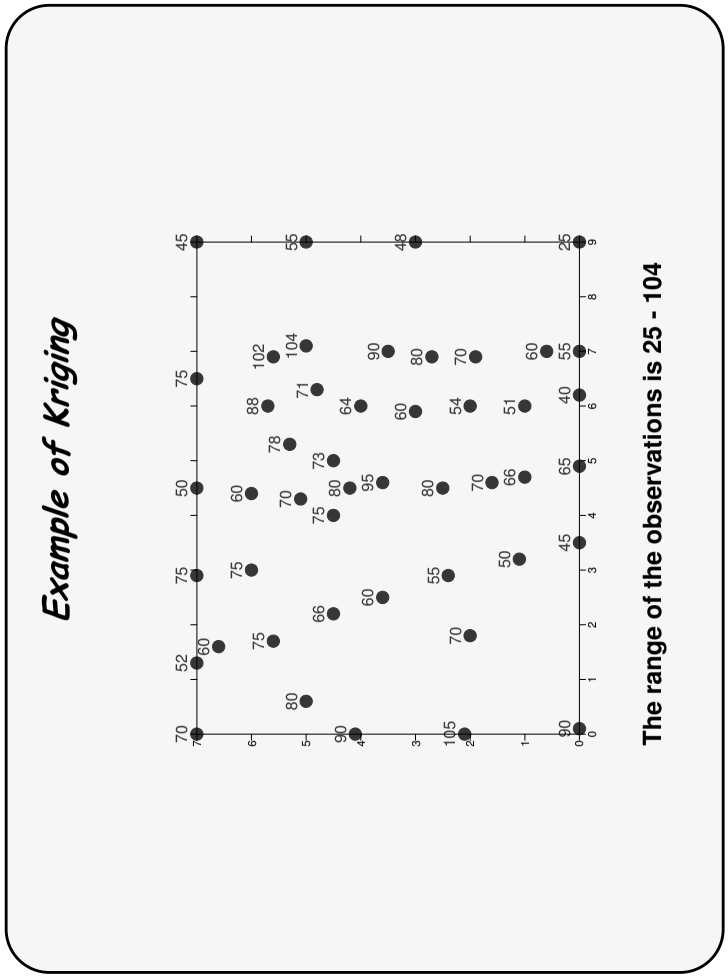
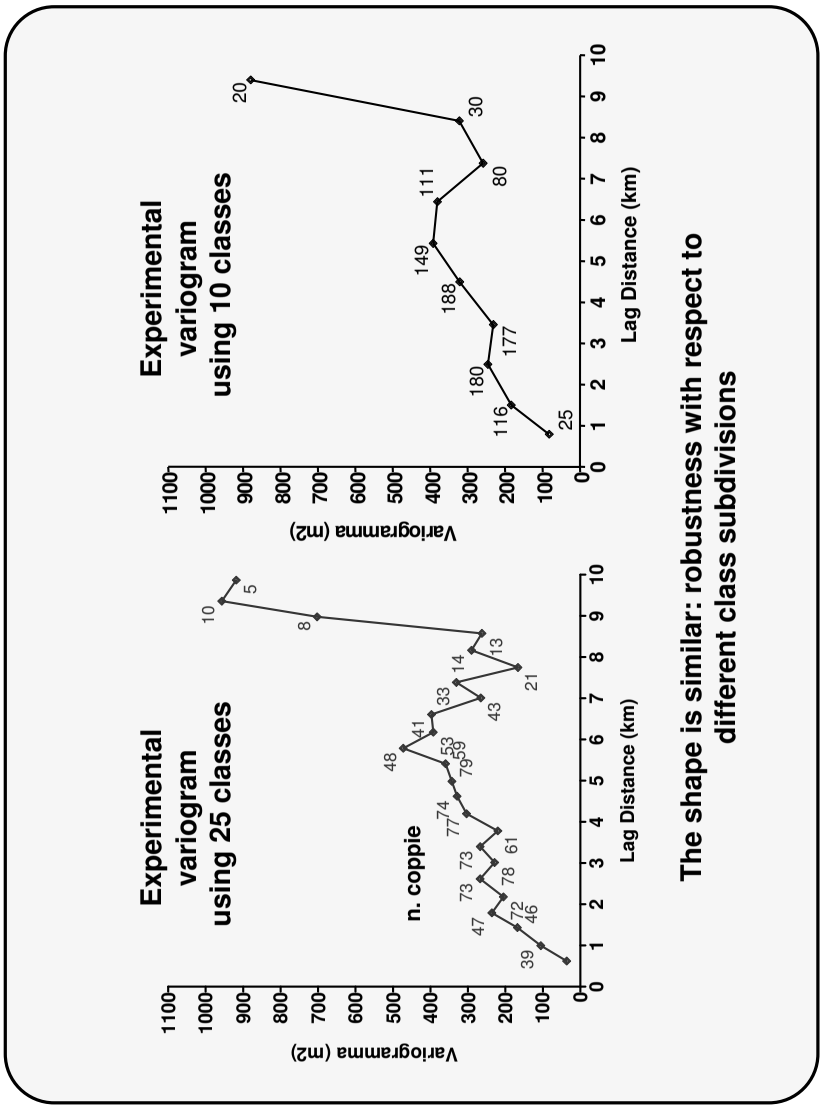


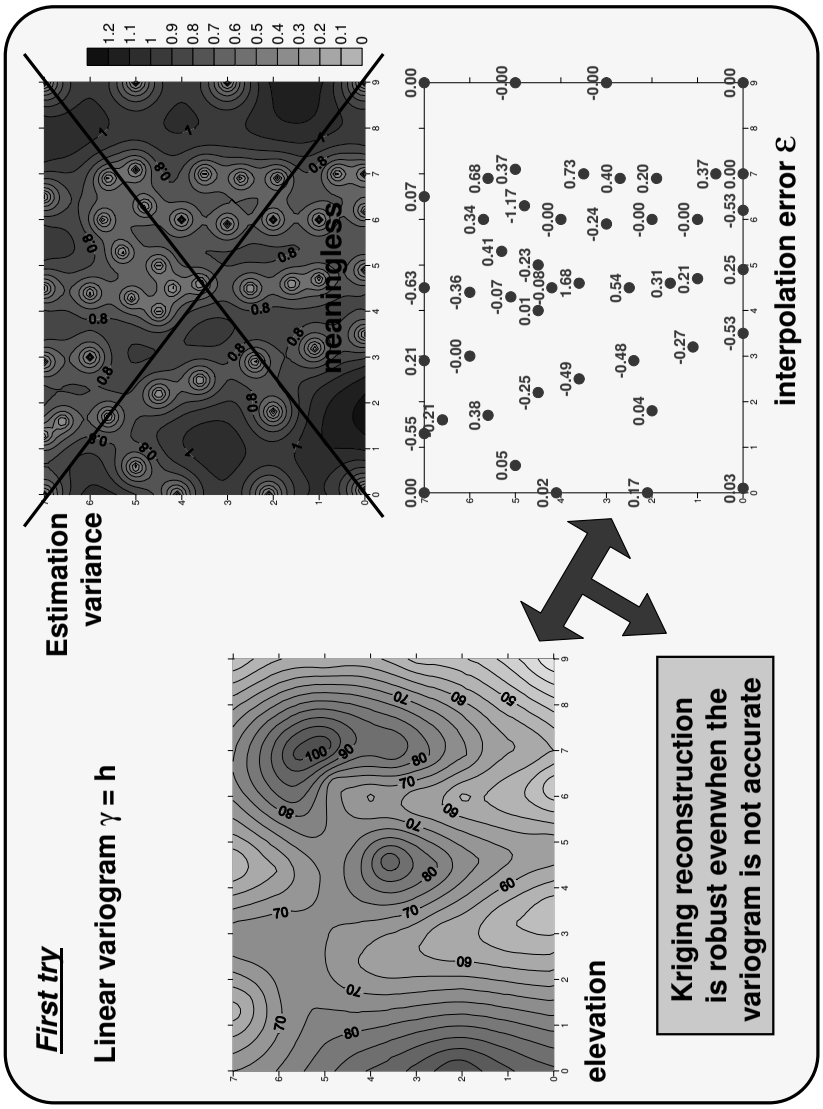
Figure 1.4: Example of interpolation with moving neighborhood

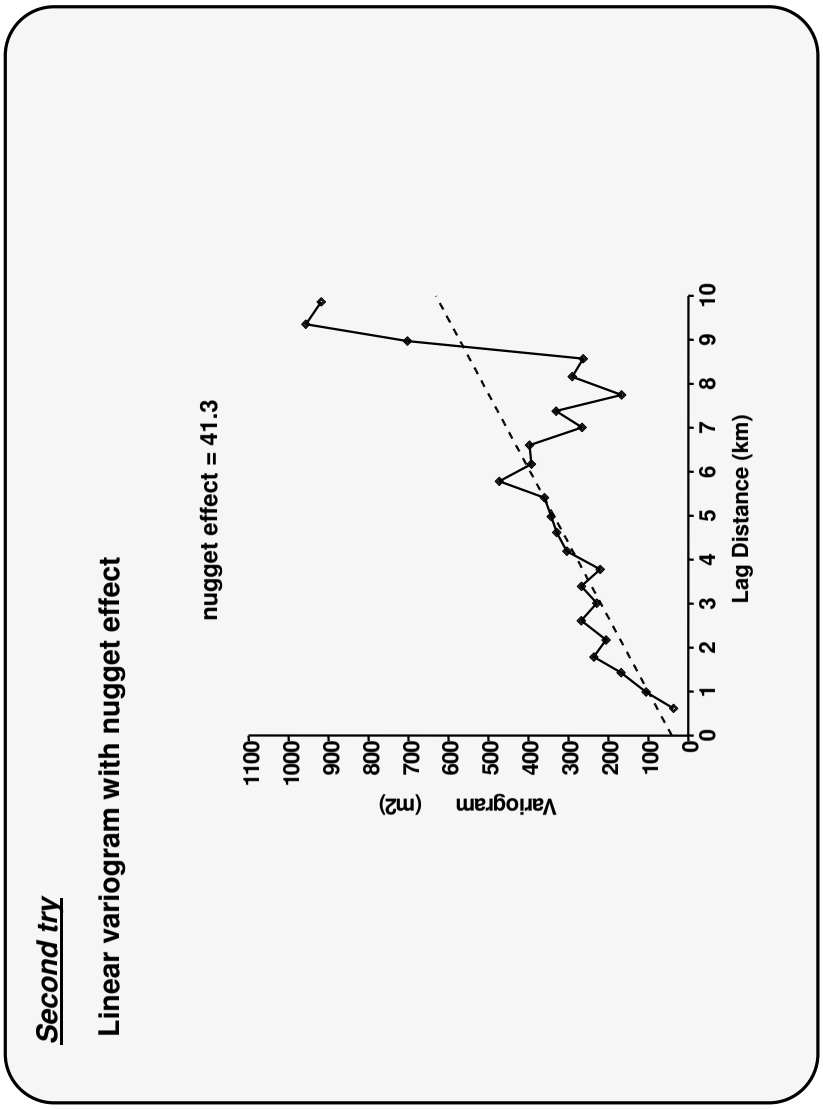
also from the physical behavior of the system from which the data come. If the trend cannot be removed from the observations by simple subtraction, the universal kriging approach [2, 3] can be used.

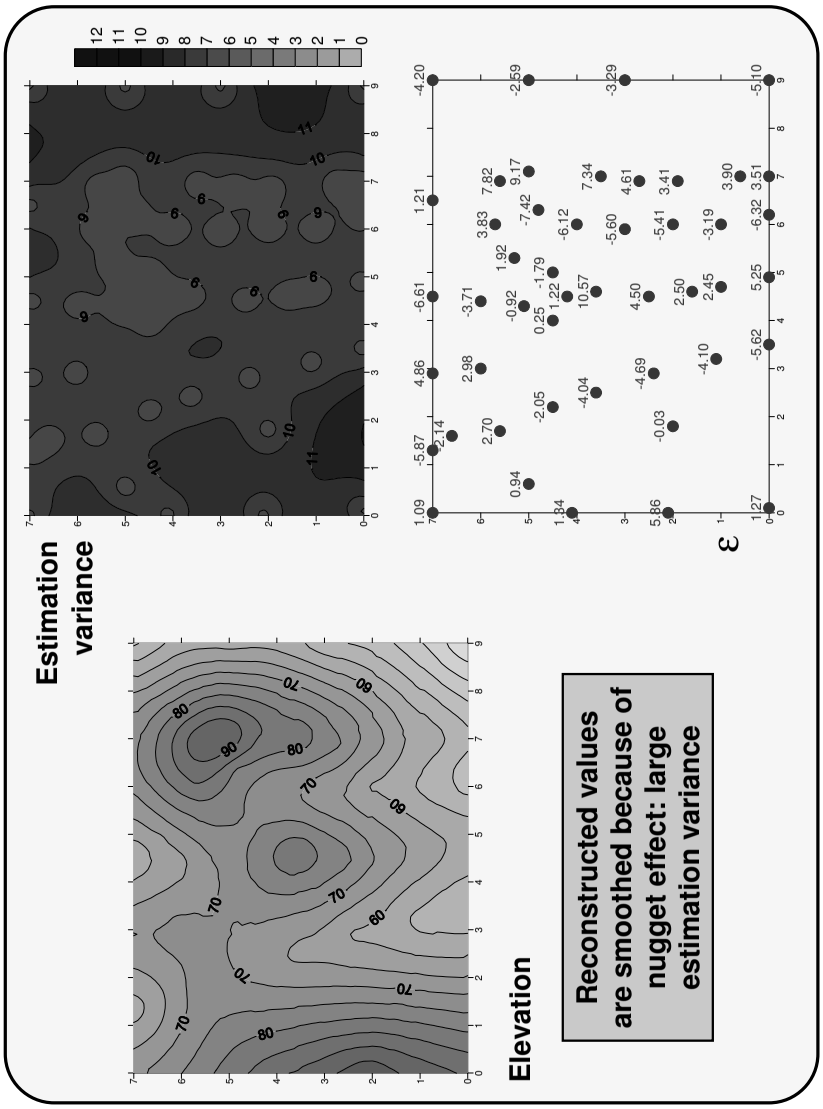
1.10 Example of application of kriging for the reconstruction of an elevation map

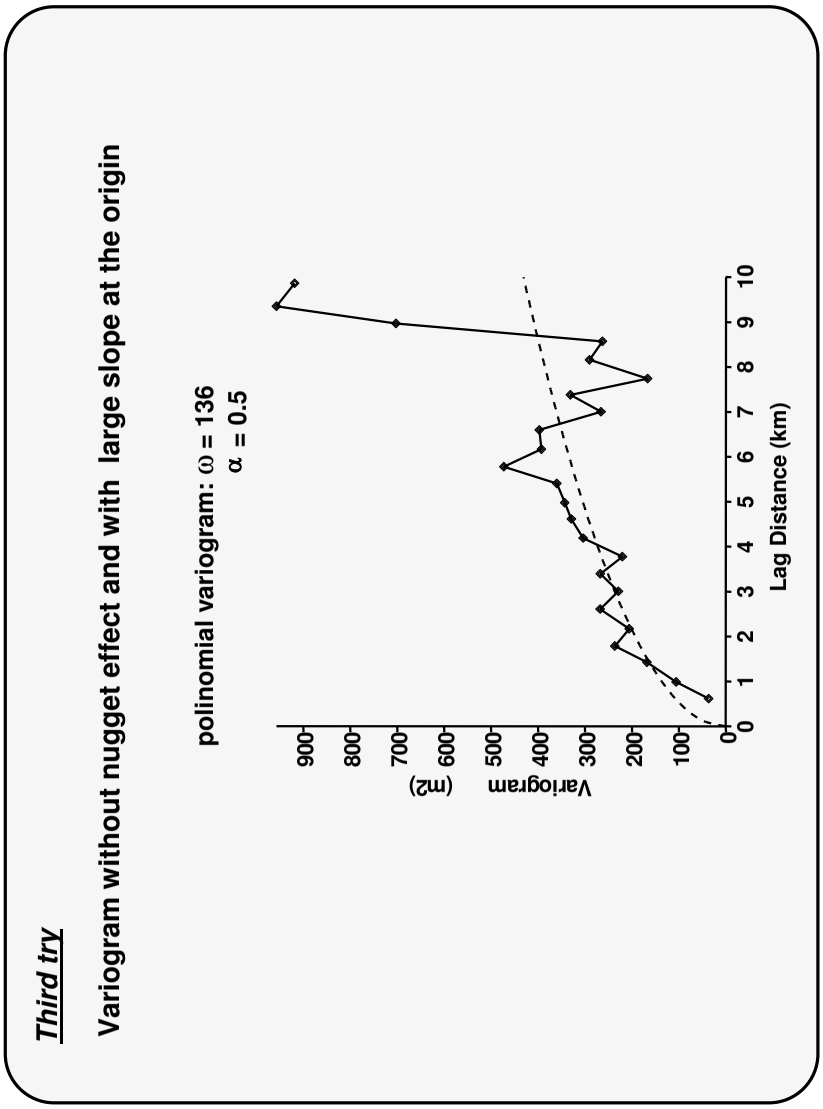


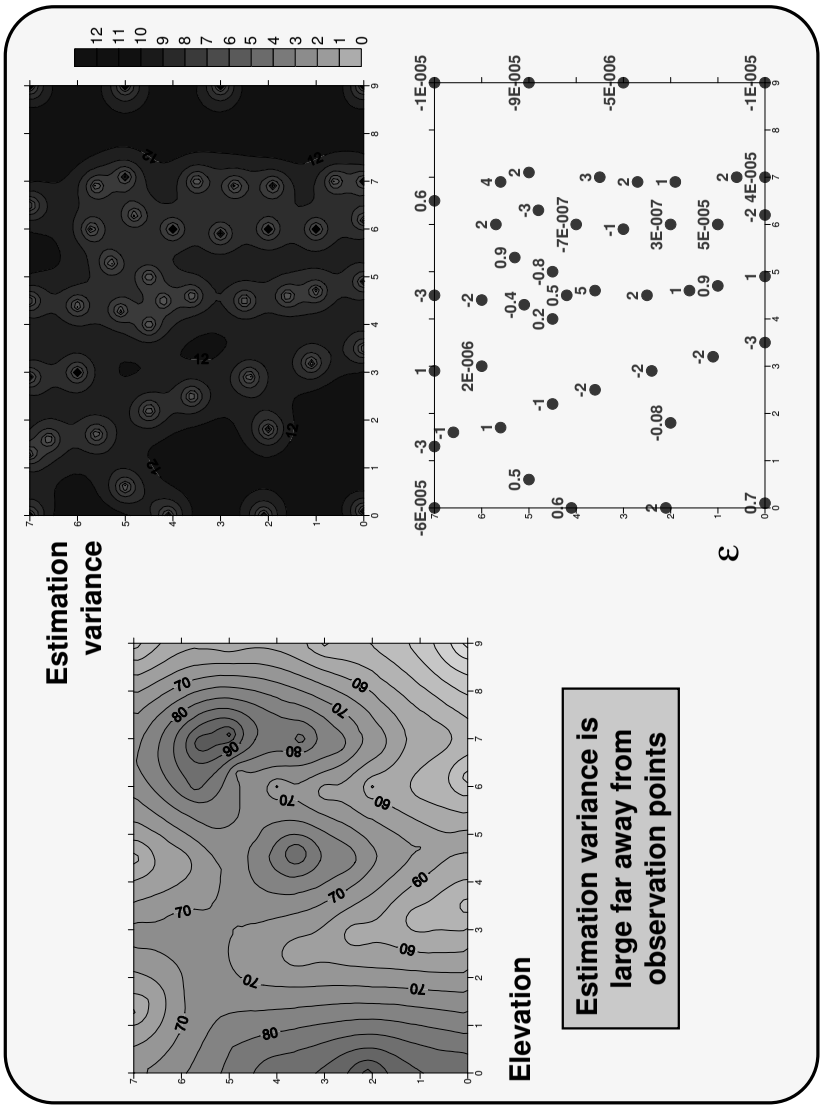

Experimental
variogram
using 10 classes
Variogramma (m2)
Lag Distance (km)
n. coppie
25
116
180
188
149
111
80
30
20
Experimental
variogram
using 25 classes
Variogramma (m2)
Lag Distance (km)
n. coppie
39
46
72
47
73
78
61
73
77
74
79
59
53
48
41
43
33
14
13
21
8
10
5
The shape is similar: robustness with respect to
different class subdivisions

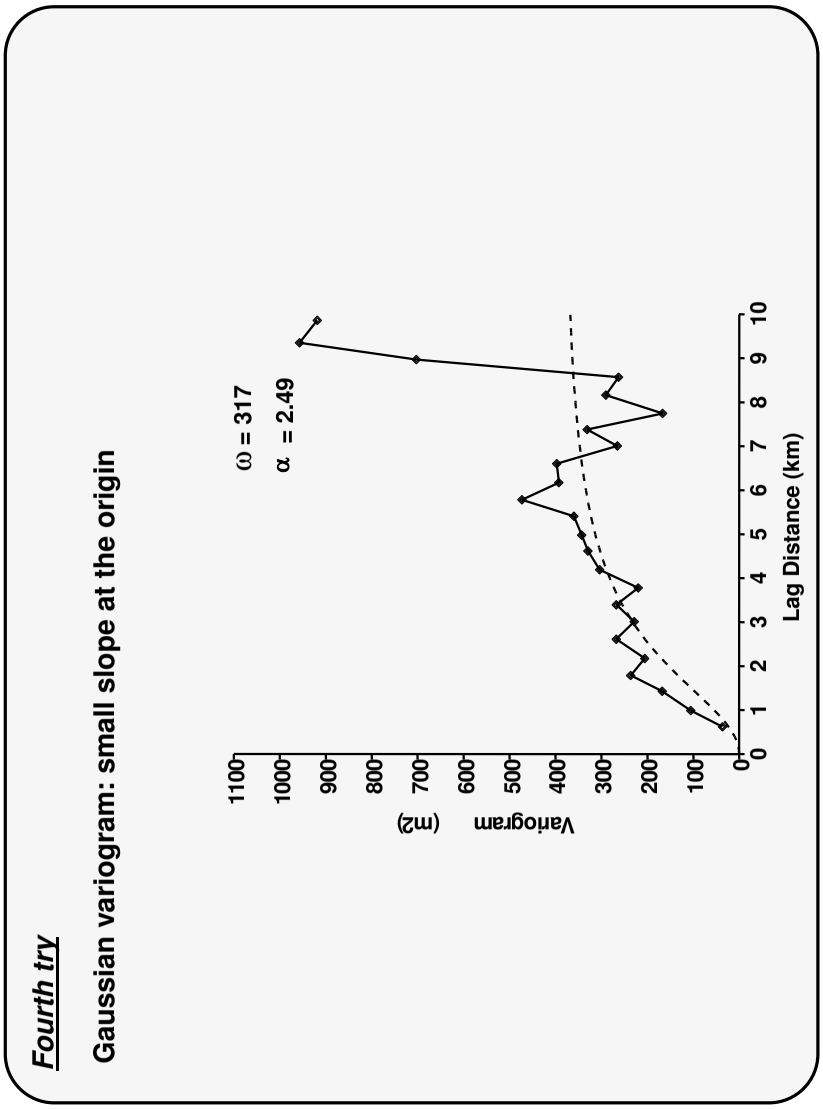

First try
Linear variogram $\gamma = h$
Estimation variance
meaningless
interpolation error ε
elevation
Kriging reconstruction is robust evenwhen the variogram is not accurate

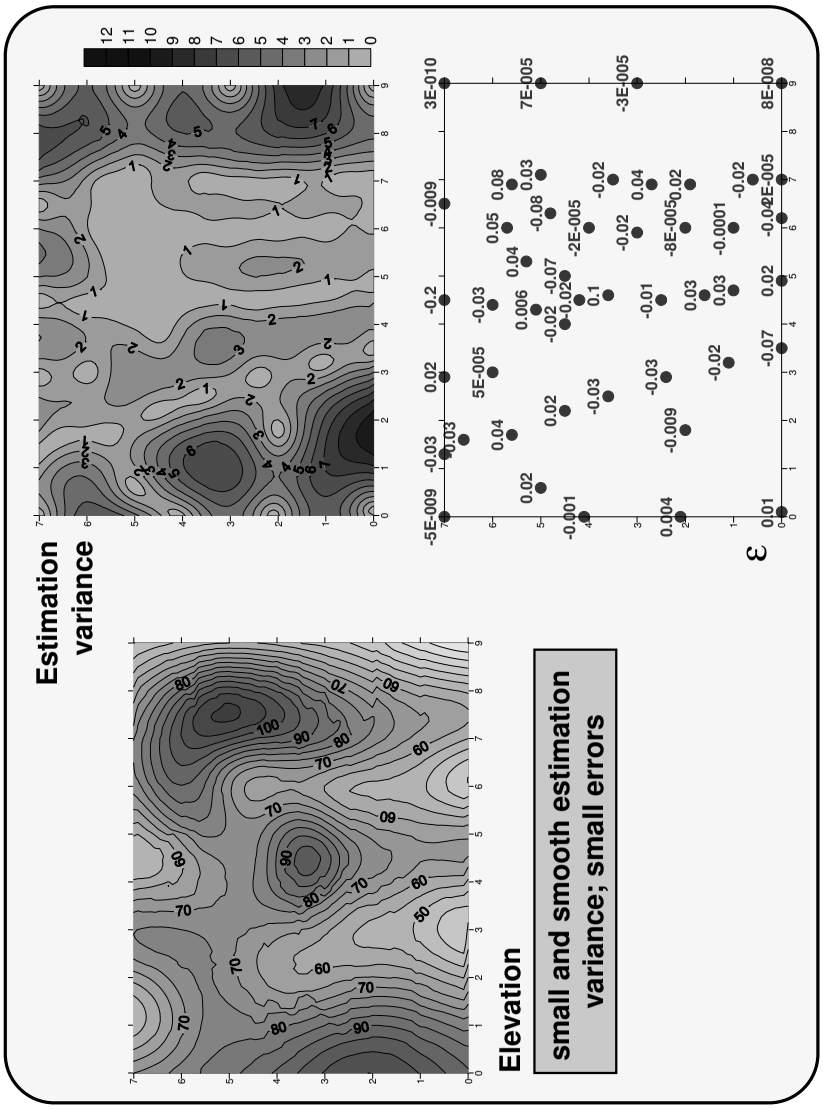

Second try
Linear variogram with nugget effect
nugget effect = 41.3
Variogram (m2)
1100
1000
900
800
700
600
500
400
300
200
100
0
0 1 2 3 4 5 6 7 8 9 10
Lag Distance (km)


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


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
ε
Estimation variance is large far away from observation points


Fourth try
Gaussian variogram: small slope at the origin
ω = 317
α = 2.49
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
ε
small and smooth estimation variance; small errors

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.