

Geophysicists and other geoscientists are among the potential readers of this book. The purpose of this chapter is to summarize the basics of geostatistics, focusing on the set of methods used in stochastic simulations (Chap. 3) and seismic inversion (Chaps. 4 and 5). For those seeking more detailed information on geostatistics, we recommend Journel and Huijbreghts (1978), David (1977), Isaaks and Srivastava (1989), Goovaerts (1997) and Chilès and Delfiner (1999). As for the broad range of applications of geostatistics the following compilation of papers of different international geostatistics congresses are recommended: Verly et al. (1984), Armstrong et al. (1989), Soares (1993), Baafi and Schofield (1997), Kleingeld and Krige (2001), Leuangthong and Deutsch (2004), Ortiz and Emery (2008), Abrahamsen et al. (2012).

Geostatistics began playing a central role in the modeling and characterization workflows of hydrocarbon reservoir characterization in recent years (Dubrule 2003). By definition, geostatistics is a set of statistical tools that seek to describe the spatial and/or temporal distribution of a given property of interest, of which one only knows its value at sparse and discrete locations (Goovaerts 1997). Despite its potential, the use of geostatistical tools as part of the reservoir geo-modeling workflow is still traditionally restricted to the three-dimensional interpolation of the reservoirs' internal properties of interest (e.g. porosity, velocities) in between the sparse well locations.

The importance of geostatistical techniques has grown considerably largely as a result of their ability to integrate, within the same framework—the reservoir grid—geophysical and well-log data of a different nature and support. For example, the integration of seismic reflection data during the geo-modeling procedure allows for more detailed, heterogenic and reliable reservoir models when compared with those based exclusively on well-log data. This is due to the much higher spatial coverage the seismic data provides compared with the well data (Doyen 2007). Among the most well-known geostatistical algorithms are the Kriging methods (Sect. 2.3: Deutsch and Journel 1992) and conditional simulations [Chap. 3: e.g. sequential Gaussian simulation, direct sequential simulation; (Deutsch and Journel 1992; Gomez-Hernandez and Journel 1993; Verly 1993; Soares 2001)].

Deterministic models estimate the value of the property of interest,  $z(x_0)^*$  at location  $x_0$  by using a linear combination of the observed values, i.e. the experimental data. Within this framework, the inferred value is believed to correspond to the true value for that unknown location,  $z(x_0) = z(x_0)^*$ . The interpolated values are interpreted as having no associated error and the underlying assumption is that the physical system being modelled is fully known (Goovaerts 1997). By assuming no uncertainty in the inferred parameter, deterministic models are hardly suitable for describing

complex and heterogeneous systems like hydrocarbon reservoirs. In these environments, the lack of knowledge of the physical system being modelled is large, and uncertainty should be assessed during the modeling process (Caers 2011).

Unlike deterministic models, the use of a probabilistic framework reflects the lack of knowledge we have about the natural Earth system being modelled. At location  $x_0$ , the probabilistic framework provides a distribution of possible values for the property of interest along with its probability of occurrence, allowing the assessment of the spatial uncertainty of the property being modelled at a particular location of interest (Caers 2011). Finally, it is worth noting that any model resulting from a probabilistic approach is constrained by a set of assumptions about prior probability distributions that are estimated from available experimental data and the spatial continuity model imposed by, for instance, a variogram model or training image (Goovaerts 1997; Strebelle 2002).

This section introduces the main geostatistical stochastic sequential simulation approaches, due to their importance in assessing spatial uncertainty in recent modeling workflows and the lack of a real understanding of this family of algorithms within the oil and gas industry. Chapter 3 deals exclusively with stochastic sequential simulation algorithms. These algorithms are the basis of the geostatistical modeling techniques presented in Chaps. 4, 5 and 6.

## 2.1 Spatial Continuity Patterns Analysis and Modeling

Much of the success related with geostatistical models in Earth sciences relates to the ability to reproduce subsurface three-dimensional numerical models with the relevant statistics of the variables retrieved from available experimental data. This reproduction is particularly effective for the spatial continuity and variability of the physical property under investigation.

The purpose of the geostatistical methodologies for estimation, simulation and inversion introduced here is to generate numerical subsurface models that reproduce the main statistics and spatial distribution as they are quantified (or estimated) as a result of available information and experimental data. Therefore, in this section we deal with the geostatistical tools that allow inference of the spatial continuity patterns of a natural resource for a given property measured at sparse locations within the study area.

Modeling the spatial behavior of a given property plays a key role in geostatistical methodologies, fulfilling two objectives: first, the characterization and quantification of the spatial pattern of a reservoir property, commonly designated in geostatistics as spatial continuity analysis, i.e. the quantification of the spatial continuity for the property of interest and the way how it varies in different spatial directions; second it is also the basis for the spatial inference/estimation, simulation and geostatistical inversion methodologies presented in the following chapters.

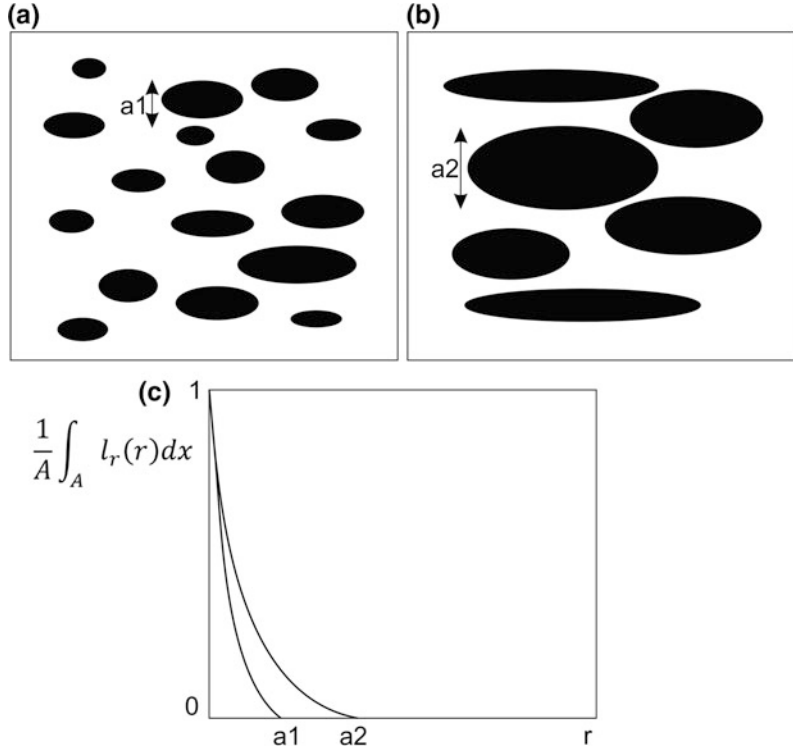
### 2.1.1 Bi-point Statistics

Let us start with a simple example of an image, or two-dimensional model, with a biphasic phenomenon: a body  $X$  within a given area  $A$ , which is composed by  $X$  and its complementary  $X^c$  ( $A = X \cup X^c$ ) in two distinct cases (Fig. 2.1), for which we intend to quantify the degree of spatial continuity.

To calculate the proportion of  $X$  in  $A$ , we can use a point that visits all possible locations within  $A$  and takes the value of '1' if it intersects body  $X$ , and the '0' if it intersects its complementary  $X^c$ . The ratio between the absolute frequency of intersections '1' and the total number of positions within  $A$  is an estimator of the proportion of  $X$  in  $A$ .

Similarly, to measure the spatial continuity or dispersion of  $X$ , we can consider a circle with radius  $r$  and count the number of times it is

**Fig. 2.1** Based on the circle as the basic structural element, the two images (a) and (b) have a spatial continuity of  $X$  as illustrated in (c)



wholly contained in  $X$  while visiting all the locations within  $A$ . Increasing the radius  $r$  allows us to assess the continuity of  $X$  for both situations, assuming an indicator variable  $I_r(x) = 1$  if the circle of radius  $r$ , centered in  $x$ , is wholly contained within  $X$ , and  $I_r(x) = 0$  if the circle is not wholly contained in  $X$ . This relative measurement of the spatial continuity, which varies inversely with the size of  $r$ , is given by the following integral calculation (Eq. 2.1):

$$\frac{1}{A} \int_A l_r(x) dx. \quad (2.1)$$

In the case previously described, the circle of radius  $r$  is considered, in image processing and particularly for mathematical morphology, to be a structural element that allows the inference of the spatial continuity of  $X$  (Serra 1982). However, there are many other structural elements that may be used for the same purpose: a line segment  $l$  or the bi-point—pairs of points separated by a vector distance  $h$ —that are richer than the circle in a

morphological point of view. These tools, when compared with the circle as structural element, allow the measurement of other parameters, such as the anisotropy degree of  $X$ : i.e. the way the continuity of  $X$  varies in different spatial directions.

Although less rich than the line segment, the bi-point acting as structural element for measuring the spatial continuity of a natural resource is the privileged structural element in geostatistics. Notice that, unlike the line segment, the bi-point does not include the notion of connected sets (two points may simultaneously belong to  $X$  while not being connected). However, the knowledge we have from a given resource is not normally acquired from a two-dimensional representation, as in Fig. 2.1. In practice, we normally have access to a sparse limited discrete group of samples located within the study area (e.g. well data, core samples, soil samples). For this reason, the inference of the spatial continuity of a given property is frequently performed by returning to the bi-point as structural element (Sect. 2.1.4).

### 2.1.2 Complex Morphologic Patterns: Auxiliary and Reference Images

There are cases in which the connectivity of bodies is of outmost importance for the characterization of a resource (e.g. meandering sand channels in braided river sedimentary environments associated with some hydrocarbon reservoirs). In these cases, given the limitations of the bi-point in characterizing the connectivity of two distinct bodies (see above), we may have to return to proxy images of co-variables for the successful reproduction of these complex spatial patterns. The spatial distribution of these complex structures may be inferred from 2D and/or 3D images, conceptual models or interpreted from available geophysical data acquired from the hydrocarbon reservoir of interest, for example. Although they are auxiliary variables, which may be directly or indirectly related to the properties being studied, that information (e.g. seismic amplitudes, resistivity) may be used as auxiliary variables for joint simulation or as target images for geostatistical inversion, which then allows the reproduction of those complex spatial patterns within the geo-modeling workflow.

When geophysical data is scarce or unavailable, for example during early exploratory phases, another reliable alternative is to use feasible geological models of a given sedimentary environment. These representations of reality are built by gathering all the information about the system being studied (e.g. information from analogous and neighboring fields, expert opinion from geologists), and are often referred to as reference images. These reference images are then used to quantify the continuity and connectivity of the features from multi-point structural elements (Strebelle 2002; Arpat and Caers 2007; Mariethoz et al. 2010; Renard and Allard 2013; Mariethoz and Caers 2014), as in the example of the line segment.

Although these multi-point statistics methodologies are not the focus of this book, it is important to stress their use in complex geological environments. The choice of multi-point statistics as tools to quantify the spatial continuity patterns of a given property should be

exclusively directed by the trust and knowledge of the degree of similitude between reality and the reference image.

### 2.1.3 Spatial Random Fields

A random variable (RV) is defined as one that can assume all the values contained within a probability distribution function. It can be continuous if the possible range of outcomes is continuous, or discrete if the outcomes are finite and without any specific order. By using the concept of RVs, the value of a property (e.g. porosity, acoustic impedance) at a given location within a study area (e.g. a reservoir grid) is interpreted as a single realization,  $z(x_1)$  of the RV  $Z(x_1)$ . The group of these dependent RVs, located for example along a reservoir grid, is defined as a random field (RF) (Ventsel 1973).

To properly model a stochastic process there is no need to explicitly characterize the entire number of associated RVs and their corresponding multivariate distributions. Instead, and under some a priori assumptions, all that is required to spatially characterize a given property is to describe a certain number of parameters, such as the mean (Eq. 2.2) and variance (Eq. 2.3; Isaaks and Srivastava 1989):

$$E\{Z(x_i)\} = m(x_i) = \int_{-\infty}^{+\infty} z dF_{x_i}(z), \quad (2.2)$$

$$var\{Z(x_i)\} = \int_{-\infty}^{+\infty} [z - m(x_i)]^2 dF_{x_i}(z), \quad (2.3)$$

where  $F_{x_i}(z)$  is the probability distribution function of the RV  $Z(x_i)$ .

If we consider two RVs, such as  $Z(x_1)$  and  $Z(x_2)$ , the covariance between both variables is given by:

$$C(Z(x_1), Z(x_2)) = E\{Z(x_1)Z(x_2)\} - m(x_1)m(x_2), \quad (2.4)$$

with

$$E\{Z(x_1), Z(x_2)\} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy d^2 F_{x_1, x_2}(x, y), \quad (2.5)$$

where  $F_{x_1, x_2}(x, y)$  is the bivariate probability distribution function

$$F_{x_1, x_2}(x, y) = \text{prob}\{Z(x_1) < x \text{ and } Z(x_2) < y\}. \quad (2.6)$$

The way the two RVs are spatially correlated is frequently described by a variogram model. The variogram between two RVs can be expressed as:

$$\gamma(Z(x_1), Z(x_2)) = E\{[Z(x_1) - Z(x_2)]^2\}. \quad (2.7)$$

Under the spatial RF assumption, the available experimental data is interpreted as being a single realization,  $z(x_i)$ ,  $i = 1, \dots, N$  (with  $N$  equal to the total number of samples of the available experimental dataset), of a random function that comprises a set of spatially-correlated RVs. Therefore, by definition it is impossible to sample more than a single realization,  $z(x_1)$ , for a given RF. Even if the same location is sampled twice, each sample set will correspond to two different realizations of  $z(x_i)$  in the same RF (Goovaerts 1997; Soares 2006).

However, a single realization of a random function is not enough to completely describe its statistical moments. The inference of these moments is only possible if we are somehow able to repeatedly sample a given location within the study area. The first and second statistical moments of a given RF can only be calculated by assuming different levels of stationarity within a specified study area: the stationarity of the mean and the stationarity of the spatial covariance (Goovaerts 1997; Soares 2006).

Within a pre-defined area of interest,  $A$ , the decision about the stationarity refers to a constant mean and spatial continuity pattern, as estimated from the available experimental data. By considering stationarity we can assume all RVs have the same mean within a limited area. With this assumption, the mean is not dependent on the

location,  $x_0$ , since it remains constant for the entire field. In this framework, the mean can then be estimated as the arithmetic mean of all the realizations of RF (i.e. the experimental dataset) composed by  $N$  samples,  $Z(x_\alpha)$ ,  $\alpha = 1, \dots, N$ :

$$m = \frac{1}{N} \sum_{i=1}^N Z(x_\alpha). \quad (2.8)$$

The second order of stationarity is defined if the correlation between two RVs depends exclusively on the distance between the two variables—the vector  $\mathbf{h}$ —and not on the specific location,  $x_0$ , of each variable. For example, the variogram between  $Z(x_1)$  and  $Z(x_2)$ :

$$\gamma(Z(x_1), Z(x_2)) = \gamma(Z(x_1), Z(x_1 + \mathbf{h})) = \gamma(\mathbf{h}). \quad (2.9)$$

By definition, the decision on stationarity can never be proved or refuted since we only know a single realization of the random function. Note that the stationarity is a property of the random function, or geostatistical model, needed to spatially infer the value of a given property far from the location of experimental data. This decision does not assume that the Earth's physical system we are trying to model is itself stationary. We should also test the available experimental data for its homogeneity across the entire study area. If this hypothesis cannot be assumed for the entire study area, then we may have to divide the field in smaller areas in which the decision about stationarity is more suitable (Goovaerts 1997).

#### 2.1.4 Variograms and Spatial Covariances

Given a quantitative property,  $Z(x)$ , the diagrams representing the pairs of points,  $Z(x)$ , versus  $Z(x + \mathbf{h})$  calculated for different values of  $\mathbf{h}$  are the statistics parameters that contain more, and richer, information about the spatial continuity of  $Z(x)$ .

Figure 2.2 shows an example of well-log data with samples located along the well path. For each well we sample each pair of points with

distance  $h$ ,  $Z(x)$  and  $Z(x + h)$ . Figure 2.3 shows the cross-plots between pairs of points  $Z(x)$  and  $Z(x + h)$  for different values of  $h$  in the vertical direction:  $h = 1, 2, 3, 10$ . For  $h = 1$  we may infer a good linear correlation between the values of samples  $Z(x)$ . This means there is a good correlation between the values of samples located in  $x$  and the values of the samples located immediately below. As soon as the values of  $h$  increase, the clouds of points start to scatter and the spatial correlation of the samples decreases. We may interpret from Fig. 2.3 that there is no correlation between samples separated by a distance of  $h = 10$ .

A group of diagrams constructed from different steps,  $h$ , comprises almost all the information related with the degree of dispersion/continuity for the variable,  $Z(x)$ , at that well location that we may retrieve from bi-point statistics. However, for a better interpretation and further use one must synthesize the bi-plots shown in Fig. 2.3 into a single tool. Summarizing the dispersion between pairs of points allows for a better visualization of the behavior of the property with increasing  $h$ . One way, for example, would be to represent the correlation coefficients (Pearson's correlation) in function of  $h$ ,

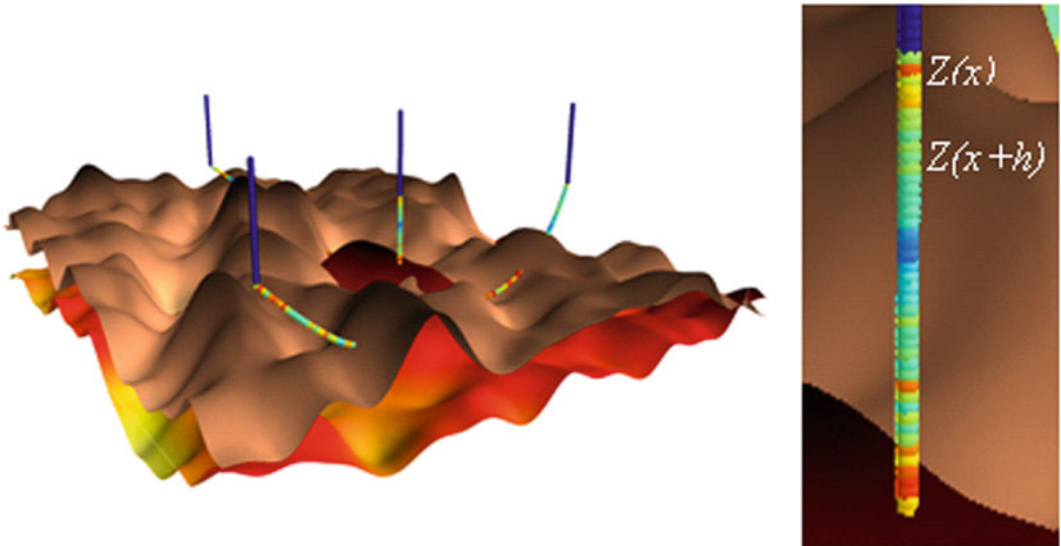
resulting in what is commonly called a correlogram (Fig. 2.4).

In addition to the correlogram (Fig. 2.4), there are other measurements that synthesize the dispersion of different clouds of point ( $Z(x)$ ,  $Z(x + h)$ ), and which may result in a series of statistics quantifying the continuity of  $Z(x)$ . For example, each cross-plot from may be summarized by the mean of the least squares between  $Z(x)$  and  $Z(x + h)$ , which is commonly called a variogram (or semi-variogram: Eq. 2.10):

$$\gamma(h) = \frac{1}{2N(h)} \sum_{x=1}^{N(h)} [Z(x_x) - Z(x_x + h)]^2, \quad (2.10)$$

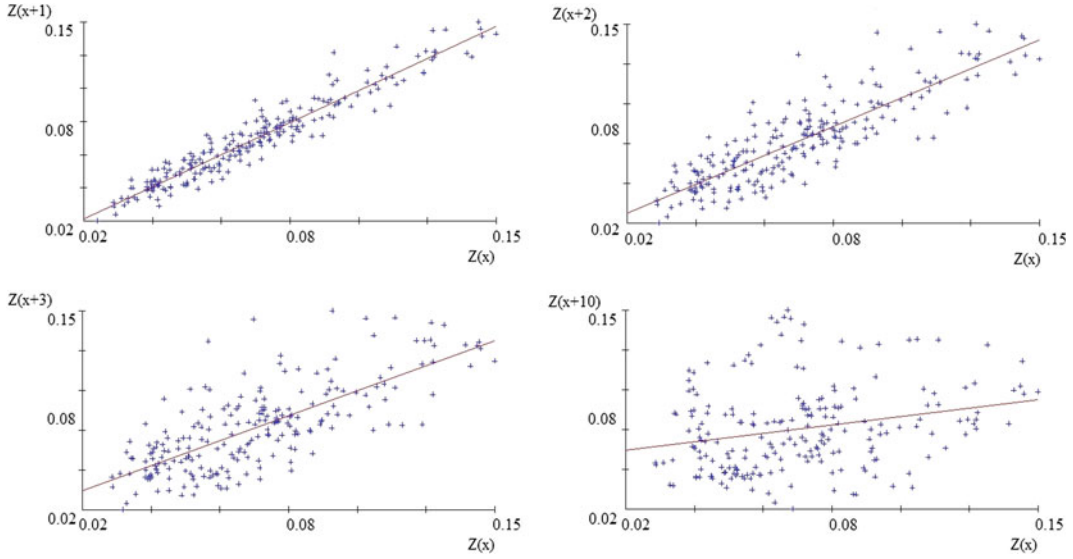
where  $N(h)$  is the number of pairs of points for each value of  $h$ .

Note that all these statistics represent the spatial continuity of the variable  $Z(x)$  in a given experimental location. In other words, the spatial continuity and its representativeness are limited to the region around the well. If instead of a single well we simultaneously consider all the wells with the same direction located within a reservoir and intersecting different geological layers, the resulting variogram (Eq. 2.10), or

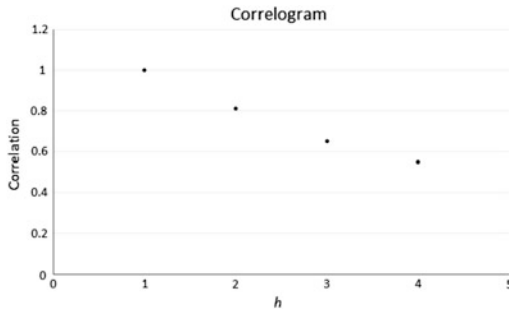


**Fig. 2.2** *Left* Spatial representation of four wells and samples of a given subsurface property of interest measured along the well path. *Right* Detail of a well with the samples measured vertically along the well path





**Fig. 2.3** Bi-plots of  $Z(x)$  versus  $Z(x + h)$  for different distances (values of  $h$ ) in the vertical direction:  $h = 1, 2, 3$  e  $10$



**Fig. 2.4** Example of a correlogram for  $h$  between 1 and 4

correlogram, would represent the space covered by the group of available wells.

Let us now consider an example of a sedimentary environment related with large sinuous channels. This kind of geological setting is normally described as anisotropic, i.e. the spatial continuity/variability is different depending on the direction of space. For example, petrophysical properties, such as porosity and permeability, are frequently more continuous and consequently less variable along a given geological formation and are more variable between formations. The same exercise can be carried out using the pairs of values  $(Z(x), Z(x + h))$  for different directions

in space. This exercise allows the assessment of spatial continuity for the property  $Z(x)$  along the entire domain of spatial analysis.

A different measurement of spatial continuity is given by the average of the product  $Z(x)Z(x + h)$ , for a given distance  $h$ . This results in a non-centered covariance estimate (Eq. 2.11):

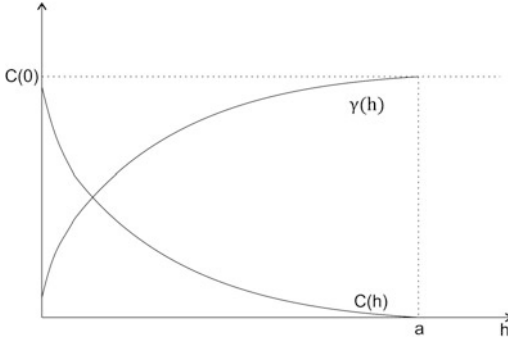
$$C(h) = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} [Z(x_{\alpha}) \cdot Z(x_{\alpha} + h)]. \quad (2.11)$$

This covariance estimator may be centered by:

$$C(h) = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} [Z(x_{\alpha}) \cdot Z(x_{\alpha} + h)] - m(x_{\alpha}), m(x_{\alpha} + h), \quad (2.12)$$

given  $m(x_{\alpha}) = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} Z(x_{\alpha})$  and  $m(x_{\alpha} + h) = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} Z(x_{\alpha} + h)$  as the arithmetical averages for all the points at locations  $x_{\alpha}$  and  $x_{\alpha} + h$ ,  $\alpha = 1, \dots, N(h)$ .

The covariance estimator (Eq. 2.12) may be expressed in terms of a correlogram (or normalized covariance):



**Fig. 2.5** Relationship between covariance,  $C(h)$ , and variogram,  $\gamma(h)$ , functions with the increment of step  $h$

$$\rho(h) = \frac{C(h)}{\sqrt{\sigma_{(x_x)}^2 \cdot \sigma_{(x_x+h)}^2}}, \quad (2.13)$$

where

$$\sigma_{(x_x)}^2 = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} [Z(x_\alpha) - m(x_\alpha)]^2,$$

and

$$\sigma_{(x_x+h)}^2 = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} [Z(x_\alpha + h) - m(x_\alpha + h)]^2.$$

By assuming the stationarity of increments  $h$  (Sect. 2.1.3), the mean of the least squares and the mean of the products are estimates of the second moments: the variogram (Eq. 2.14) and the centred covariance (Eq. 2.15):

$$\gamma(h) = \frac{1}{2} E\{[Z(x) - Z(x+h)]^2\}, \quad (2.14)$$

$$C(h) = E\{Z(x)Z(x+h)\} - E\{Z(x)\}E\{Z(x+h)\}, \quad (2.15)$$

we reach the relationship between a variogram and the covariance (Eq. 2.16):

$$\gamma(h) = C(0) - C(h). \quad (2.16)$$

We then may express the variogram in terms of a correlogram (Eq. 2.17):

$$\rho(h) = \frac{C(h)}{C(0)}. \quad (2.17)$$

The relationship between the variogram and the covariance functions (Eq. 2.16) is synthesized in Fig. 2.5.

### 2.1.5 Spatial Representativeness of the Variogram

As noted in Sect. 2.1.3. above, it is important to relate the stationary assumption of the probabilistic model to the notion of the representativeness and homogeneity of the experimental samples, which are the basis for calculating the stationary statistics (mean and variance).

For the first statistical moment, the mean of the  $N$  samples within an area  $A$ ,  $m_z = \frac{1}{N} \sum_{\alpha=1}^N z(x_\alpha)$  is an estimator of the expected value of the random function  $Z(x) - E\{Z(x)\}$ , under the assumption of stationary about the mean.

Given that we only have access to a single realization of the random function  $Z(x)$ , i.e. the set of values  $z(x_\alpha)$  in  $A$ , we ensure that  $m_z$  is a good estimate of the spatial integral (Eq. 2.18):

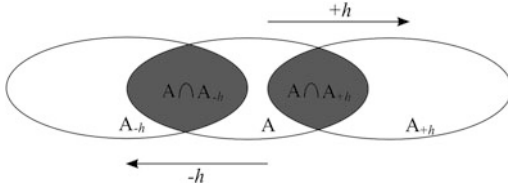
$$m_z = \frac{1}{A} \int_A z(x) dx. \quad (2.18)$$

Assuming a stationarity mean for the random function  $Z(x)$  is equivalent to considering the set of available experimental data—the only known realization of the random function  $Z(x)$ —as homogenous and representative for the entire area  $A$ . The same principle can easily be extrapolated for the second statistical moments: the variograms (Eq. 2.14) and covariance (Eq. 2.15).

Let us consider a given area,  $A$ , as a closed body and a distance vector,  $h$ , as schematically represented in Fig. 2.6. A new area,  $A_{+h}$ , can then be defined by the samples  $x_{+h}$  with distance  $+h$  from the samples  $x$  in  $A$  (Eq. 2.19):

$$A_{+h} : \{x+h | x \in A\} \text{ or } \{x | x-h \in A\}. \quad (2.19)$$





**Fig. 2.6** Schematic representation of the spatial representativeness for covariance and variance estimates

In the same way we may mathematically define  $A_{-h}$  (Eq. 2.20):

$$A_{-h} := \{x - h | x \in A\} \text{ or } \{x | x + h \in A\}. \quad (2.20)$$

The variogram is by definition the variance of the deviations  $(Z(x) - Z(x + h))$  when both samples  $x$  and  $x + h$  belong to  $A$ . Hence, its representativeness refers to the union of both grey areas as shown in Fig. 2.6:  $(A \cap A_{+h}) \cup (A \cap A_{-h})$ .

In terms of spatial integrals of  $A$ , the variogram may be described as follows:

$$2\gamma(h) = \left[ \frac{1}{A \cap A_{+h}} \int_{A \cap A_{+h}} [Z(x) - Z(x + h)]^2 dx + \frac{1}{A \cap A_{-h}} \int_{A \cap A_{-h}} [Z(x) - Z(x - h)]^2 dx \right]. \quad (2.21)$$

When the distance vector  $h$  is small—for example smaller than half the dimension of  $A$ —then the representativeness of the variogram is similar to  $A$ :

$$(A \cap A_{+h}) \cup (A \cap A_{-h}) \approx A.$$

Thus, when computing the variogram estimate we must take into account that its representativeness in  $A$  is given by the dimension of the distance vector  $h$ . For practical applications, the representativeness of the variogram ( $\gamma(h)$ ) should be questioned for distances,  $h$ , larger than half the size of  $A$  along the direction of  $h$ .

Besides the spatial representativeness, each value of the variogram must be related to a homogeneous cloud of pairs of points in the bi-plot  $(Z(x), Z(x + h))$  for each step  $h$ . For example, one single anomalous value  $Z(x_i)$  may result, along with its neighbor samples  $Z(x_i + h)$ , in large values of  $[Z(x_i) - Z(x_i + h)]^2$ , and consequently a large mean  $\gamma(h)$  translated in practice as a weak spatial correlation for that given value of  $h$ . We may consider this sample anomalous and as having a restrict representativeness. If the computed experimental variogram, without taking into account the anomalous value  $Z(x_i)$ , has a more regular behavior (in a way that it is more similar to the rest of the values of  $\gamma(h)$ ), it may and should be adopted as representative of the remaining samples for the whole area. Note that in these cases, during the process of local estimation the areas surrounding  $Z(x_i)$  should be considered with special care to ensure the area of influence of the sample does not have a large impact on the estimate. By definition,  $\gamma(h)$  does not translate the behaviour of the large spatial variability between  $Z(x_i)$  and the neighbor samples.

### 2.1.6 Spatial Continuity for Multivariate Systems

Consider those cases in which, at a given spatial location for a single sample, we measure more than one attribute,  $Z_1(x_i)$ ,  $Z_2(x_i)$ , ...  $Z_N(x_i)$ : for example, P-wave velocity, S-wave velocity and density measured at the same locations along a well path.

The correlation between each pair of these attributes  $Z_1(x)$ ,  $Z_2(x)$ , ...  $Z_N(x)$  may be measured through the correlation coefficient for the set of  $N$  samples (Eq. 2.22):

$$\rho(Z_1, Z_2) = \frac{1}{N\sigma_1\sigma_2} \sum_{i=1}^N [(Z_1(x_i) - m_1)(Z_2(x_i) - m_2)], \quad (2.22)$$

where  $m_1, m_2, \sigma_1^2 \in \sigma_2^2$  are the mean and variance of  $Z_1(x)$  and  $Z_2(x)$  respectively.

We can now generalize the correlation between the different variables and calculate the correlation of variable  $Z_1(x)$  located in  $x$  and the variable  $Z_2(x + \mathbf{h})$  located in  $x + \mathbf{h}$ . The spatial dependency between each pair of variables with distance  $\mathbf{h}$  may be characterized by cross-variograms, cross-covariance and cross-correlation (Goovaerts 1997).

Assessing the spatial dependency between variables is important as we often have to use an auxiliary variable (frequently more abundant) to estimate a primary variable (less abundant) assuming a spatial correlation between both. This is frequently the case when estimating porosity models with known values at the well locations from models of acoustic impedance retrieved, for example, from seismic inversion in the entire area  $A$ .

The random function model used so far (Sect. 2.1.3) may be generalized for multivariate cases. The set of RVs defining  $N_v$  random function,  $I = 1, \dots, N_v$ , can also be designated as multivariate random function:

$$Z_i(x), \quad i = 1, \dots, N_v, \quad \forall x \in A. \quad (2.23)$$

The joint distribution of two variables  $Z_i(x)$  and  $Z_j(x)$  depends on the distance vector  $\mathbf{h}$  (Eq. 2.24):

$$F_{ij}(\mathbf{h}, z_i, z_j) = \text{prob}\{Z_i(x) \leq z_i, Z_j(x + \mathbf{h}) \leq z_j\}, \quad \forall ij. \quad (2.24)$$

The spatial dependency between two variables  $Z_i(x)$  and  $Z_j(x)$  may be measured by the cross-covariance function:

$$C_{ij}(\mathbf{h}) = E\{[Z_i(x) - m_i] \cdot [Z_j(x + \mathbf{h}) - m_j]\}, \quad \forall ij, \quad (2.25)$$

or by the cross-variogram:

$$\gamma_{ij}(\mathbf{h}) = \frac{1}{2} E\{[Z_i(x) - Z_i(x + \mathbf{h})] \cdot [Z_j(x) - Z_j(x + \mathbf{h})]\}, \quad \forall ij. \quad (2.26)$$

Note that  $\gamma_{ij}(\mathbf{h}) = \gamma_{ji}(\mathbf{h})$ , but  $C_{ij}(\mathbf{h})$  may not equal  $C_{ji}(\mathbf{h})$ , meaning the function is not symmetrical with  $\mathbf{h}$ .

The relationship between cross-variogram and cross-covariance may be described as:

$$\gamma_{ij}(\mathbf{h}) = C_{ij}(0) - \frac{1}{2} [C_{ij}(\mathbf{h}) + C_{ij}(-\mathbf{h})]. \quad (2.27)$$

If the cross-covariance in Eq. 2.25 is rewritten in terms of the sum between two terms dependent on  $\mathbf{h}$  (Eq. 2.27), it can easily be understood that the cross-variogram (Eq. 2.27) comprises only the first term. This is the reason for the symmetry around  $\mathbf{h}$ :

$$C_{ij}(\mathbf{h}) = \frac{1}{2} [C_{ij}(\mathbf{h}) + C_{ij}(-\mathbf{h})] + \frac{1}{2} [C_{ij}(\mathbf{h}) - C_{ij}(-\mathbf{h})]. \quad (2.28)$$

In practice, the asymmetry component of the cross-covariance is usually ignored for two main reasons (Journel and Huijbregts 1978):

- The amount of available experimental data rarely allows comprehension and consequent validation across the whole study area of the physical phenomenon that results in the asymmetry in the cross-covariance;
- Modeling the asymmetric cross-covariance is extremely complex.

For this reason, the geostatistical tools that quantify the spatial continuity of a multivariate system are frequently the cross-variograms and the symmetrical cross-covariance:  $C_{ij}(\mathbf{h}) = C_{ji}(\mathbf{h})$ , or the mean of  $C_{ij}(\mathbf{h})$  and  $C_{ji}(\mathbf{h})$ .

Finally, the cross-correlogram may be synthesized by (Eq. 2.29):

$$\rho_{ij}(\mathbf{h}) = \frac{C_{ij}(\mathbf{h})}{\sqrt{C_{ii}(0) \cdot C_{jj}(0)}} \in [-1, 1]. \quad (2.29)$$

### 2.1.7 Variogram Modeling Workflow

#### *Modeling by a Mean Representative Function*

Figure 2.7a represents a given reality, i.e. a model with all the values of  $Z(x)$  within an area  $A$ . From that reference data  $Z(x)$ , a limited set of experimental data was randomly sampled (black circles on Fig. 2.7a) and an experimental variogram calculated (Fig. 2.7b). As a reference dataset from which reality is known, we may also calculate the experimental variogram for all points within the study area, the entire  $\gamma(h)$  (Fig. 2.7c).

The example illustrated in Fig. 2.7 synthesizes the main objective during the variogram modeling stage: the estimation of the real variogram using a discrete and limited set of samples  $z(x)$  and the corresponding experimental variogram.

Once the values of the variograms for different distances,  $h$ , are calculated for a given area  $A$ , it is necessary to model them using a function describing the spatial behavior of the property of interest for the entire study area. In practice, we adjust a smooth function of a reduced number of parameters that describe the spatial continuity of  $Z(x)$ .

This step is of utmost importance within the geostatistical framework, since it allows the synthesis of the structural characteristics of the spatial phenomena, e.g. degree of dispersion/continuity and anisotropies, into a single and coherent variogram model.

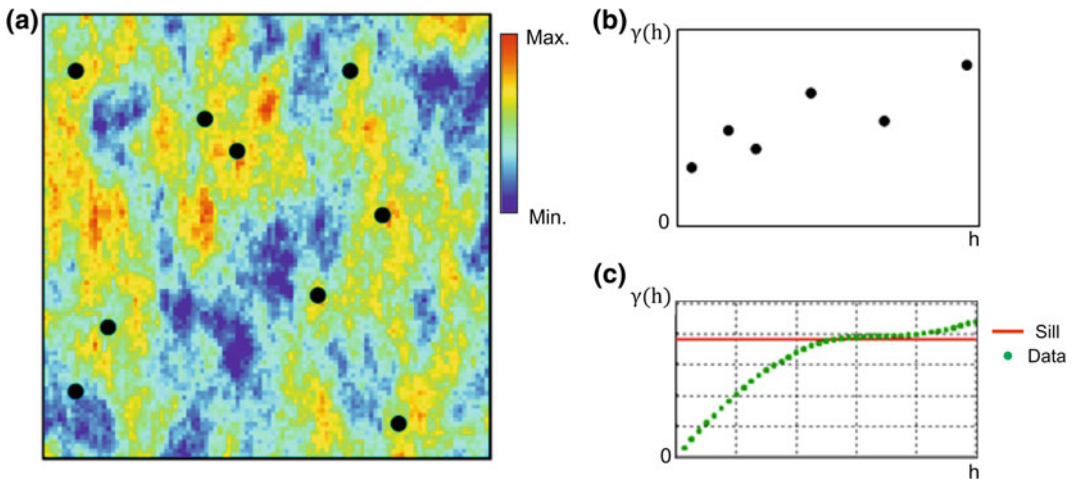
It is also common practice to adjust a model to the experimental variogram by conditioning it from expert knowledge about the phenomena being modelled.

#### *Positive Definite Models*

From the many functions that may be used to interpolate the points of an experimental variogram we need to constrain our options to those allowing stable solutions when calculating linear estimates (Sect. 2.2). To meet this condition, the variogram and covariance must be positive definite. The necessary condition for the positive definite of a covariance matrix is:

$$\sum_i \sum_j \lambda_i \lambda_j C(i, j) \geq 0. \quad (2.30)$$

Any linear combination of covariance between pairs of points within an area  $A$  is always positive definite. If we consider a given



**Fig. 2.7** **a** Known given data from an area  $A$  and set of samples retrieved from this data (black filled circles). **b** Experimental variogram computed from the set of

available samples. **c** Experimental variogram computed from the known given data (**a**)

variable  $Z(x_0)$  resulting from a linear combination of RVs  $Z(x_1), Z(x_2), \dots, Z(x_N)$  (Eq. 2.31):

$$Z(x_0) = \sum_i \lambda_i Z(x_i), \quad (2.31)$$

then a definite positive covariance ensures that the variance of  $Z(x_0)$  is always positive:

$$\begin{aligned} \text{var}\{Z(x_0)\} &= E\left\{\sum_i \sum_j \lambda_i \lambda_j Z(x_i) Z(x_j) - m^2\right\} \\ &= \sum_i \sum_j \lambda_i \lambda_j E\{Z(x_i) Z(x_j) - m^2\} \\ &= \sum_i \sum_j \lambda_i \lambda_j C(i, j) \geq 0. \end{aligned} \quad (2.32)$$

By replacing Eq. 2.16 in Eq. 2.31 the variance may be written in function of the variogram:

$$\begin{aligned} \text{var}\{Z(x_0)\} &= C(0) \sum_i \sum_j \lambda_i \lambda_j \\ &\quad - \sum_i \sum_j \lambda_i \lambda_j \gamma(i, j) \geq 0. \end{aligned} \quad (2.33)$$

In the cases in which  $C(0)$  does not exist (i.e. non-stationary random functions) the variance  $Z(x_0)$  exists if  $\sum_i \lambda_i = 0$ . Thus, the necessary condition of positive variance is ensured if  $\sum_i \sum_j \lambda_i \lambda_j \gamma(i, j) \leq 0$  conditioned to the sum of the weights being zero.

### 2.1.8 Theoretical Variogram Models

The positive definite condition limits the number of models that can be used for interpolating experimental variograms. In practice, within a geostatistical framework a limited range of positive definite interpolating functions are used. The following models are presented: spherical, exponential, Gaussian and power.

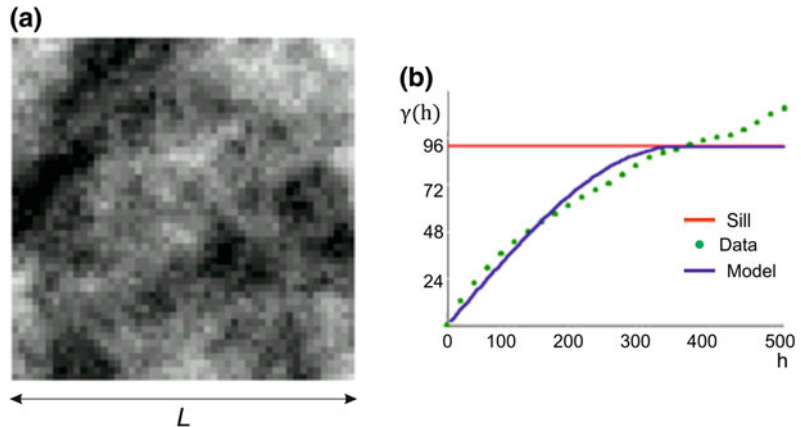
#### Spherical Model

The spherical model is one of the most common in geostatistics and it is a function of two parameters (Eq. 2.34): the sill ( $C$ ), upper limit to which the values of the variogram tend when  $h$  is increased; and the range,  $a$ , distance from where the values of  $\gamma(h)$  stop increasing and are approximately equal to the sill, normally the total variance of the experimental data  $Z(x)$ . The range of a variogram measures the distance from where the data  $Z(x)$  is no longer correlated:

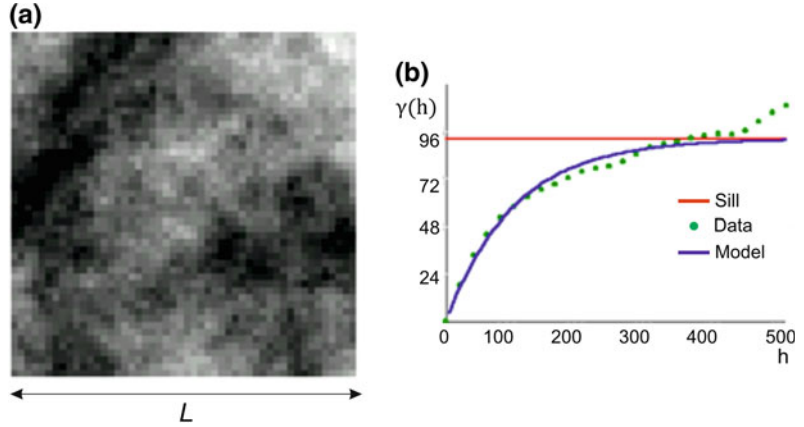
$$\gamma(h) = \begin{cases} C \left[ 1.5 \frac{h}{a} - 0.5 \left( \frac{h}{a} \right)^3 \right] & \text{for } h \leq a \\ C & \text{for } h > a. \end{cases} \quad (2.34)$$

Figure 2.8a shows a map with the spatial distribution of a variable  $Z(x)$  modelled with a spherical model (Fig. 2.8b) in which the amplitude is one-third of the dimension map length.

**Fig. 2.8** **a** Map with the spatial distribution of the variable  $Z(x)$  according to **b** a spherical model with amplitude equal to  $1/3 L$



**Fig. 2.9** **a** Map with the spatial distribution of the variable  $Z(x)$  according to **b** an exponential model with amplitude equal to 0.3  $L$



### Exponential Model

The exponential model is a function of the same parameters of the spherical model—sill and range (Eq. 2.35)—while the variogram tends asymptotic to the sill value:

$$\gamma(h) = C \left[ 1 - \exp\left(-\frac{3h}{a}\right) \right]. \quad (2.35)$$

In this model, the range value is the distance in which the model reaches 95% of the sill:  $\gamma(h) = 0.95 C$ .

Figure 2.9a shows the spatial distribution of a variable modelled with an exponential variogram (Fig. 2.9b) with range equal to 0.3  $L$  of the map length.

Comparing both Figs. 2.8 and 2.9, besides the rapid growth of the spherical model near the

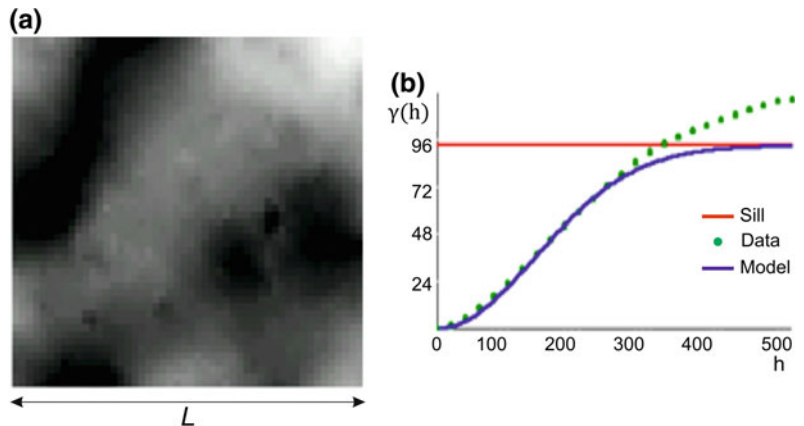
origin, it shows structures with larger spatial continuities resulting from larger spatial correlation for larger distances  $h$ .

### Gaussian Model

The two variogram models previously presented—the spherical and exponential—have a relatively fast increase near the origin, translating into a typical behavior of irregular natural phenomena. Other phenomena, more regular and continuous, are translated by a slow increase of the variogram values near the origin, for example a parabolic behavior. This is the case of Gaussian models (Eq. 2.36):

$$\gamma(h) = C \left[ 1 - \exp\left(\frac{-3h^2}{a^2}\right) \right]. \quad (2.36)$$

**Fig. 2.10** **a** Map with the spatial distribution of the variable  $Z(x)$  according to **b** a Gaussian model with amplitude equal to 0.2  $L$



As in the exponential model, range,  $a$ , is the distance from where the model reaches 95% of the sill:  $\gamma(h) = 0.95 C$ .

Figure 2.10 shows the behavior of a Gaussian variable model in which the range is 0.2  $L$  of the length of the map. Note the much larger and smooth spatial continuity compared with those obtained from the spherical and exponential models (Figs. 2.8 and 2.9).

### Power Models

So far all the variogram models described have a sill as upper limit for where the variogram values of  $\gamma(h)$  tend when  $h$  increases infinitely. These models are adequate for transition phenomena characterized by a distance—the range—from where the spatial correlation between samples no longer exists. In these transitional phenomena there is always a relationship between covariance and variogram as described in Eq. 2.16.

However, there are other natural phenomena in which the growth of  $\gamma(h)$  is continuous with  $h$  and does not tend to the sill. These are non-stationary phenomena in which there is no finite variance or notion of covariance—the variance grows with the dimension of the dispersion field  $Z(x)$ .

The most common variogram model applied in these situations is the power model (Eq. 2.37):

$$\gamma(h) = Ch^\alpha, \quad (2.37)$$

with  $\alpha$  between 0 and 2. Depending on  $\alpha$ , the variogram may be linear ( $\alpha = 1$ ), logarithmic ( $0 < \alpha < 1$ ) or parabolic ( $1 < \alpha < 2$ ) (Fig. 2.11).

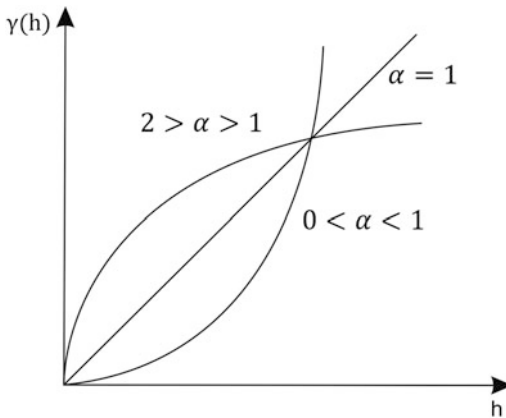
### 2.1.9 Linear Combinations of Variogram Models: Imbricated Structures

For most natural phenomena, spatial continuity patterns are rarely simple and modelled by a single variogram model. Normally different structures coexist simultaneously with distinct spatial continuities and distinct characteristics. A simple example is the one illustrated by Fig. 2.12, which shows a binary process with structures at two different scales: the first structure is composed of bodies with average size  $a_1$ ; the grouping of these structures results in bodies with dimension  $a_2$  (second structure).

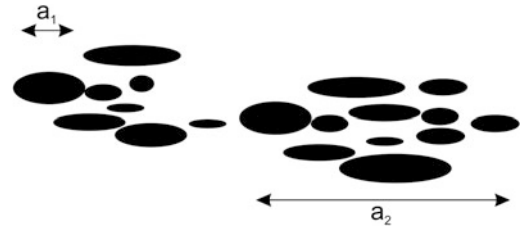
In these cases, the experimental variogram reveals the simultaneous effect of the imbricated structures through discontinuities in the growing behavior of the variogram values with the distance  $h$ .

For comparing the effect of modeling a property with different variogram models, Fig. 2.13 shows four 2D models (with size  $1000 \times 1000$  m) with distinct spatial variogram models, respectively: (a) a spherical model with range equal to 150 m; (b) a spherical model with range equal to 750 m. Figure 2.13c, d show two imbricated structures with different contributions:

- Figure 2.13c has a spatial distribution modelled by a weighted mean of 70% for the first structure ( $a = 150$  m) and 30% for the second



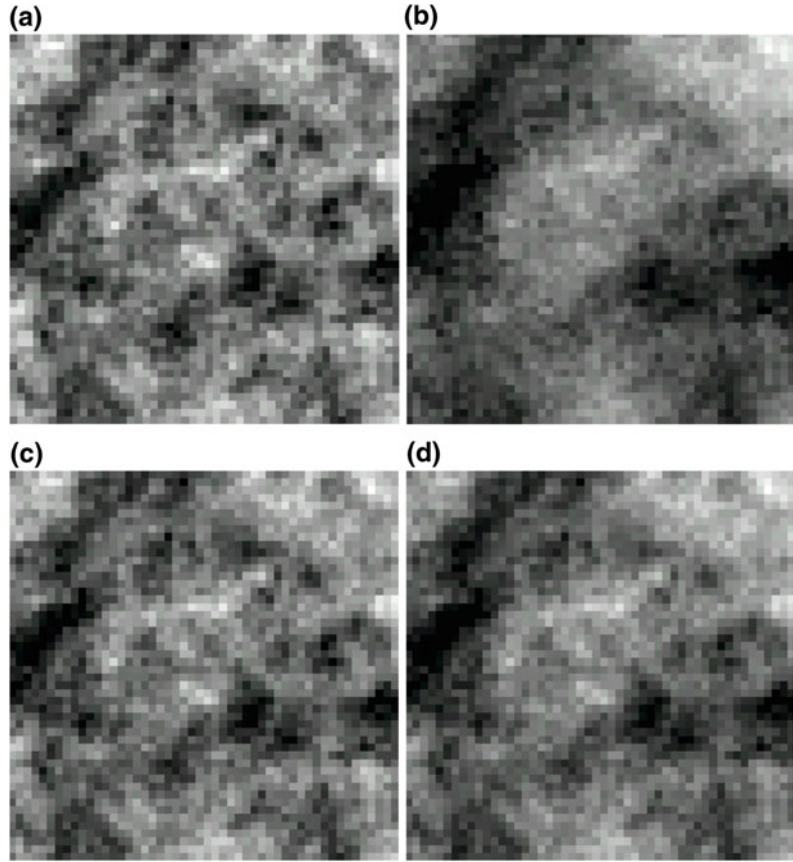
**Fig. 2.11** Schematic representation of power models



**Fig. 2.12** Example of a binary process with structures at two different scales,  $a_1$  and  $a_2$



**Fig. 2.13** Images with the same attribute modelled with different variogram models: **a** spherical model with zero nugget effect for a single structure with range 150 m; **b** spherical model with zero nugget effect for a single structure with range 750 m; **c** imbricated structure quantified by a weighted mean of 70% for a first structure with  $a = 150$  m and 30% for the second structure with  $a = 750$  m; **d** imbricated structure quantified by a weighted mean of 30% for a first structure with  $a = 150$  m and 70% for the second structure with  $a = 750$  m



( $a = 750$  m). The greater influence of the smallest structure is clear. The resulting variogram model may be expressed by the following linear combination:

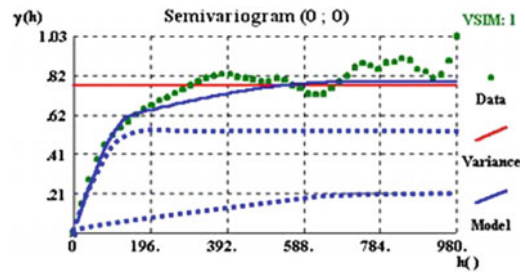
$$\gamma(h) = 0.7 \text{ Sph}(a = 150 \text{ m}) + 0.3 \text{ Sph}(a = 750 \text{ m}).$$

- Figure 2.13d has a spatial distribution modeled by a weighted mean of 30% for the first structure and 70% for the second. In this image the greater influence of the largest scale structure is clear. The resulting variogram model may be synthesized by the following linear combination:

$$\gamma(h) = 0.3 \text{ Sph}(a = 150 \text{ m}) + 0.7 \text{ Sph}(a = 750 \text{ m}).$$

The imbricated structures are modelled through a linear combination of the variogram models presented above (Fig. 2.14). They benefit from the propriety of the positive definite models:

in any linear combination of positive coefficients of models definite positive is positive. Any linear combination of variogram models (Eq. 2.38) is an imbricated group of variogram  $\gamma_i(h)$  where the weights  $C_i(0)$  are the sills—total variance—for each single structure. The sum of the different sills is equal to the global sill:



**Fig. 2.14** Variogram model resulting from the sum of two distinct structures:  $\gamma(h) = 0.48 \text{ Sph}(a = 150 \text{ m}) + 0.24 \text{ Sph}(a = 750 \text{ m})$

$$\gamma(\mathbf{h}) = \sum_i C_i(0) \gamma_i(\mathbf{h}). \quad (2.38)$$

### Nugget Effect

Theoretically, the value of the variogram is zero for  $h = 0$  ( $\gamma(\mathbf{h}) = 0$ , for  $h = 0$ ). In practice, between consecutive samples there is a minimum value of  $h$  for which the value of  $\gamma(\mathbf{h})$  may be calculated. When this minimum value ( $\gamma(h_{min})$ ) is high, it means there is a high variability in the natural phenomenon at the small-scale, i.e. for distances smaller than the distances between samples or observations:  $\gamma(\mathbf{h})$  may not tend to zero while  $h$  tends to zero. In these cases, there is an inflexion or discontinuity in the growth of the variogram at a scale not sampled by the available experimental data, i.e.  $h = 0$  and  $h_{min}$ . For these cases, the variogram is modelled by a constant,  $C_0$ , which is called the nugget effect (Eq. 2.39). The nugget effect is the first structure summed to the linear combination of the remainder of the structures:

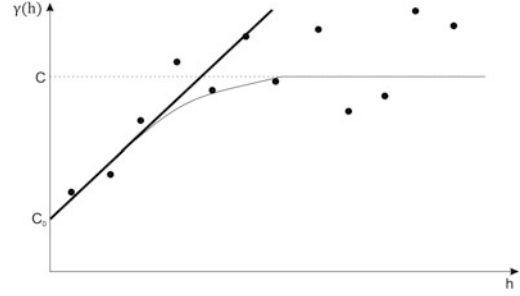
$$\gamma(\mathbf{h}) = C_0 + \sum_i C_i(0) \gamma_i(\mathbf{h}), \quad (2.39)$$

while the total variance is described by:

$$C(0) = C_0 + \sum_i C_i(0). \quad (2.40)$$

The nugget effect summarizes the effect of two distinct parts of the total variability in the natural phenomenon being modelled: (i) the small-scale variability not comprised in the sampling grid; (ii) the variability at the scale support sample introduced by non-systematic sampling errors that add to the structure of the natural phenomenon random noise component.

Modeling the nugget effect by a constant different from zero translates the lack of knowledge about the system at the small-scale by increasing the uncertainty during the estimation procedure (Sect. 2.2 and Chap. 3). The nugget effect can be interpreted as the intersect value at the ordinate axis. It is usually inferred by the intersection of a line approximating the first



**Fig. 2.15** Nugget effect inference based on the linear regression (thick black line) of the first points of  $\gamma(\mathbf{h})$ . Variogram model is represented by the thin black line

points of the variogram with the ordinate axis (Fig. 2.15).

### Anisotropy Models

The spatial continuity of a natural resource frequently varies as a function of the direction of the space, e.g. greater continuity for porosity values along a channelized structure versus lower continuity across the channel direction. A given attribute of any natural resource (e.g. porosity in hydrocarbon reservoirs) has an isotropic spatial continuity if the variogram (or covariance) has the same behavior in all directions (that is,  $\gamma(\mathbf{h})$  depends exclusively on the modulus of the distance vector  $\mathbf{h}$ ). There are cases, however, in which the attribute being studied is more continuous along a preferential direction resulting in structural anisotropy. The latter may be seen as the variability of a given attribute depending on the directions of the space we consider when inferring the spatial behavior of a specific property.

Modeling anisotropic structures seeks to reduce structures of continuity depending on the direction of a single variogram model. The anisotropy is normally processed in terms of geometric transforms of the coordinate system in such a way as that the several variograms along different directions are equivalent to a single model, transforming them into isotropic structures.

Two of the most common anisotropy models are the geometric and the zonal.

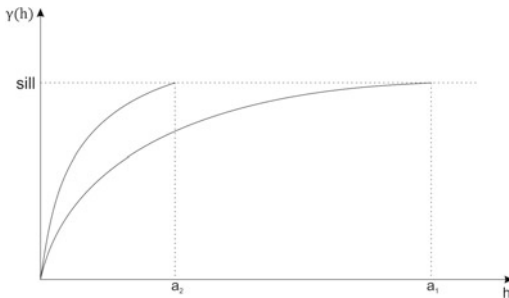
### Geometric Anisotropy

Geometric anisotropy is a model in which the spatial continuity, revealed by the variogram amplitudes, varies gradually from the direction of larger continuity/range through the direction of smallest amplitude, perpendicular to the first following the equation of an ellipse defined by those directions.

A geometric anisotropy is characterized by variograms with the same model and the same sill in all directions, but different ranges, while the minimum and maximum amplitudes are in perpendicular directions (Fig. 2.16).

Geometric anisotropy means the rose diagram that describes the ranges of the variogram along the different directions of the space may be modelled by an ellipse in 2D or an ellipsoid in 3D. In fact, when representing the different variogram ranges as a function of the angle of the variogram direction in a diagram, the ellipse is the geometric figure that best describes a geometric anisotropy.

An ellipse may be seen as a linear transformation of a circle (or sphere in three-dimensions), which corresponds to an isotropic condition: that is to say, the variogram range does not change with the variogram direction. A simple method for combining a group of variograms, with ranges  $a_x$ ,  $a_y$  and  $a_z$  in the three directions of space, respectively, into a single model with range  $a = 1$  is given by the following geometric transformation:



**Fig. 2.16** Schematic representation of geometrical anisotropy: variograms with the same sill, but different ranges

$$\begin{aligned} \gamma_{a=1}(\mathbf{h}) &= \gamma_x(\mathbf{h}_x) & \text{with } h &= h_x/a_x, \\ \gamma_{a=1}(\mathbf{h}) &= \gamma_y(\mathbf{h}_y) & \text{with } h &= h_y/a_y, \\ \gamma_{a=1}(\mathbf{h}) &= \gamma_z(\mathbf{h}_z) & \text{with } h &= h_z/a_z. \end{aligned} \quad (2.41)$$

This corresponds to the normalization of the distances within the Cartesian space as follows:

$$h = \sqrt{\left(\frac{h_x}{a_x}\right)^2 + \left(\frac{h_y}{a_y}\right)^2 + \left(\frac{h_z}{a_z}\right)^2} \quad (2.42)$$

$$\gamma_{a=1}(h) = \gamma_x(hx) \text{ if } h = h_x/a_x$$

If we choose to transform the anisotropy into a reference variogram (for example, the variogram with the largest range instead of a variogram with range equal to 1), the normalized distance  $h$  can be described as:

$$h = \sqrt{h_x \left(\frac{a_x}{a_x}\right)^2 + h_y \left(\frac{a_y}{a_y}\right)^2 + h_z \left(\frac{a_z}{a_z}\right)^2}, \quad (2.43)$$

where  $a_x$  is the reference variogram range and  $r_x = a_x/a_x$ ,  $r_y = a_x/a_y$  e  $r_z = a_x/a_z$  are the three anisotropic ranges in the three main ranges.

This methodology of transforming coordinates may be applied equally to non-stationary models that do not reach the sill.

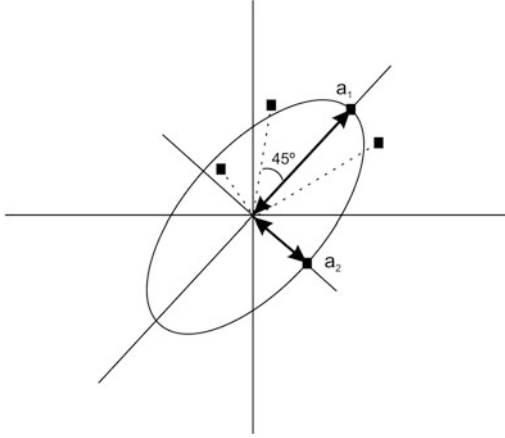
If the direction of largest range does not match the axis of the reference coordinate system, it must be rotated so the axis  $xx'$  matches  $a_x$ ,  $yy'$  matches  $a_y$  and  $zz'$  matches  $a_z$  before applying the geometry transform (Eq. 2.43).

Figure 2.17 shows a two-dimensional example in which the maximum range is observed in the direction of  $45^\circ$  and the smallest amplitude perpendicular to this ( $135^\circ$ ). Therefore, any vector  $h$  must be first rotated  $45^\circ$  (Eq. 2.44) before any normalization operation:

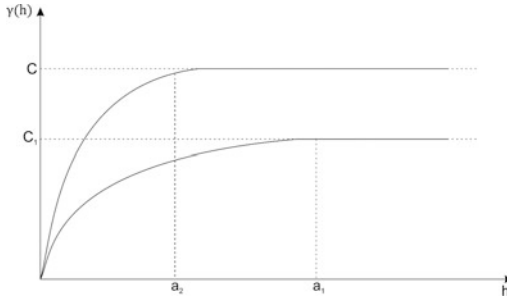
$$\begin{bmatrix} h_{x_{45^\circ}} \\ h_{y_{45^\circ}} \end{bmatrix} = \begin{bmatrix} \cos 45^\circ & -\sin 45^\circ \\ \sin 45^\circ & \cos 45^\circ \end{bmatrix} \cdot \begin{bmatrix} h_x \\ h_y \end{bmatrix}. \quad (2.44)$$

### Zonal Anisotropy

Zonal anisotropy is common in stratified phenomena in which the spatial continuity along a stratum is in contrast with the variability between



**Fig. 2.17** Schematic representation of the rose diagram for the group of ranges and geometric anisotropic model



**Fig. 2.18** Example of zonal anisotropy

a stratum and, consequently, the variogram along the stratum does not reach the sill of the variogram (Fig. 2.18).

The zonal anisotropy may be modelled by a linear combination of two structures:

$$\gamma(\mathbf{h}) = C_1\gamma_1(\mathbf{h}_1) + (C_2 - C_1)\gamma_2(\mathbf{h}_2), \quad (2.45)$$

where the second structure  $\gamma_2(\mathbf{h}_2)$ , here the vector  $\mathbf{h}_2$ , is related only to the direction of larger variability—that is, between stratum. The example of Fig. 2.18 may be modelled by a first structure with a sill  $C_1$  and a geometric anisotropy with an anisotropic ratio and a second structure with variance equal to  $C_2 - C_1$  and range  $a_2$  that only exists in direction  $\mathbf{h}_2$ :

$$\gamma(\mathbf{h}) = C_1\gamma_1(\mathbf{h}_1) + (C_2 - C_1)\gamma_2(\mathbf{h}_2),$$

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

$$\text{and } h_2 = \frac{h_z}{a_z}.$$

The two structures are defined as follows:

$\gamma_1(\mathbf{h})$ —an isotropic model with a sill equal to  $C_1$ , and with ranges  $a_1$  in direction  $\mathbf{h}_1$  and  $a_2$  in direction  $\mathbf{h}_2$ .

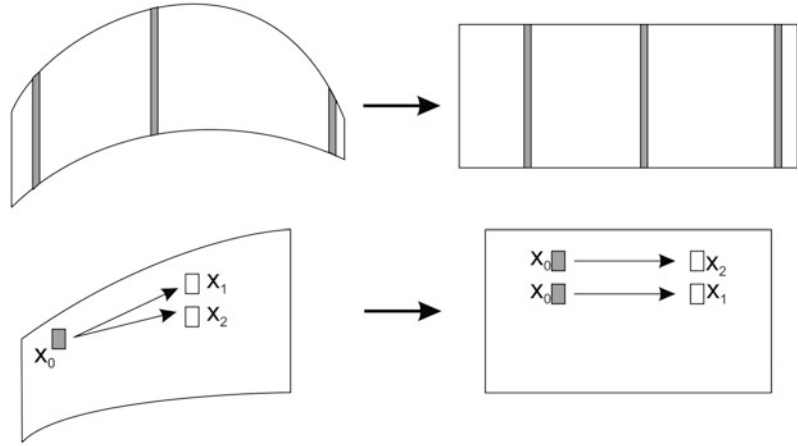
$\gamma_2(\mathbf{h})$ —anisotropic model with a sill equal to  $C_2 - C_1$ , with range  $a_2$  in direction  $\mathbf{h}_2$  and with ‘infinite’ range along  $\mathbf{h}_1$ . Note that choosing a very large range in direction  $\mathbf{h}_1$  (for example, 10 times the dimension of the field) implies a small contribution of  $\gamma_1(\mathbf{h})$  for values of  $\mathbf{h}_1$  near the dimension of the field.

### Structural Transforms

There are cases in which the anisotropic phenomena do not regularly vary from the maximum to the minimum range: for example, when the spatial dispersion is conditioned by external factors to the genesis of this resource. Structural geology, such as folds and faults, are typical events in which these cases occur.

Each of these situations requires specific geometric transforms in order to achieve simple and generalized variogram models for the entire study area. For example, the transform associated with a folded geological formation, or dome, into a regular shape in order to better identify the spatial continuity between samples within a very non-isotropic structure (Fig. 2.19) (Mallet 2002, 2004). The conceptual basis of this transform is the following: the value of a sample in the thin part of the formation is correlated with more than one value in the thickest part of the folded layer. For the example in Fig. 2.19, in order to calculate the mean variogram for the pairs of points  $(x_0, x_1)$  and  $(x_0, x_2)$ , it is the same as doubling the sample  $x_0$ . Note that this type of transformation is possible in geological formations with a large horizontal continuity and vertical heterogeneity, which is the case in some oil and gas reservoirs.

**Fig. 2.19** Structural geometric transform of a folded geological layer



### 2.1.10 Co-regionalized Models of Multivariate Systems

So far we introduced the spatial continuity models for univariate systems. This section deals with a co-regionalization model for multivariate systems (Sect. 2.1.6). In a multivariate domain, the variogram models and cross-covariance, as in the univariate case, must ensure a positive variance of any linear combination between variables. There are several co-regionalization models, however the most common is the linear model. In this mode the single and cross-variograms are the result of the linear combination of basic models.

The basic models (nugget effect, first structure, second structure and so on): with  $L$  as the total number of structure,  $\gamma_0(\mathbf{h}), \gamma_1(\mathbf{h}), \dots, \gamma_L(\mathbf{h})$ , must all belong to the simple and cross-variograms:

$$\begin{aligned}\gamma_{ii}(\mathbf{h}) &= b_{ii}^0 \gamma_0(\mathbf{h}) + b_{ii}^1 \gamma_1(\mathbf{h}) + \dots + b_{ii}^L \gamma_L(\mathbf{h}). \\ \gamma_{jj}(\mathbf{h}) &= b_{jj}^0 \gamma_0(\mathbf{h}) + b_{jj}^1 \gamma_1(\mathbf{h}) + \dots + b_{jj}^L \gamma_L(\mathbf{h}), \\ \gamma_{ij}(\mathbf{h}) &= b_{ij}^0 \gamma_0(\mathbf{h}) + b_{ij}^1 \gamma_1(\mathbf{h}) + \dots + b_{ij}^L \gamma_L(\mathbf{h}).\end{aligned}\quad (2.46)$$

The co-regionalization model of the group of  $N_v \times N_v$  simples and cross-covariance may be defined as

$$\gamma_{ij}(\mathbf{h}) = \sum_{l=0}^L b_{ij}^l \gamma_l(\mathbf{h}), \quad \forall_{i,j},$$

or

$$C_{ij}(\mathbf{h}) = \sum_{l=0}^L b_{ij}^l C_l(\mathbf{h}), \quad \forall_{i,j}, \quad (2.47)$$

where  $b_{ij}^l$  are the sill of the simple model  $C_{ij}(\mathbf{h})$ .

In order to ensure that the groups of covariances  $C_{ij}(\mathbf{h})$  are allowed (i.e. we need to guarantee that the variance of any linear combination between variables is positive) we need the following:

- (1) The functions  $C_l(\mathbf{h})$  are positive definite;
- (2) The matrices  $b_{ij}^l \forall_{i,j,l=0,L}$  are positive define.

This condition implies that:

$$b_{ii}^l > 0 \text{ e } b_{jj}^l > 0 \quad \text{and} \quad b_{ii}^l \cdot b_{jj}^l > b_{ij}^l b_{ij}^l \quad l = 0, \dots, L.$$

These relationships suggest two basic rules:

- (i) one structure that exists in the simple covariance may not exist in the cross-covariance:

$$b_{ii}^l \neq 0 \text{ does not imply that } b_{ij}^l \neq 0;$$

- (ii) one structure that exists in the cross-covariance needs to necessary exist in the corresponding simple covariances:

$$b_{ij}^l \neq 0 \Rightarrow b_{ii}^l \neq 0 \text{ and } b_{jj}^l \neq 0;$$

In practice the co-regionalization models of multivariate systems by a linear model may be summarized in the following sequence of steps (Goovaerts 1997):

- (1) Selection of a group of structures  $\gamma_l(\mathbf{h})$ ,  $l = 0, \dots, L$ , that reproduce the behavior of the different simple variogram models  $\gamma_{ii}(\mathbf{h})$ ,  $\gamma_{jj}(\mathbf{h})$ ,. Following the rules already stated in the previous paragraphs (i.e., one structure in the cross-covariance needs to exist necessarily in the corresponding simple covariances). There is no need, at this stage, to perform an analysis of the cross-variograms;
- (2) Estimation of the sills  $b_{ij}^l$  for the different structures of each cross-variogram;
- (3) The matrices  $b_{ij}^l$  need to be positive:

$$b_{ii}^l > 0 \text{ and } b_{jj}^l > 0 \quad l = 0, \dots, L.$$

Repeat (2) and (3) until the adjustment of the group of simple and cross-variograms is satisfactory.

Note that is due to the last two steps of this sequence that in practice, make the adjustment of co-regionalization not simple.

Nevertheless, in the cases where there is the need to model multivariate systems and there is also the need to approximate the adjustments of the simple and cross-variograms, the priority should be focused on the simple variograms, and in particular, the variograms of the main variable.

## 2.2 Estimation Models

The reason geostatistics has grown in many research fields of Earth and Environmental Sciences relates to the efficiency and simplicity of its interpolation methods from known sparse experimental data to unknown locations within a study area.

Here we present a group of geostatistical methods for the characterization of spatial phenomena by

integrating different types of experimental data with different scale support, resolution and uncertainty. The spatial inference methodologies presented for continuous and indicator variables are the basis for the stochastic sequential simulation algorithms presented in Chap. 3, and one of the main foundations of the methodologies for integrating geophysical data into reservoir modeling (Chaps. 4, 5 and 6).

### 2.2.1 Linear Estimation of Local Statistics

Generally speaking, it is possible to define the spatial inference, or estimation, of a variable at any given scale support (point, area or volume) for a location not sampled  $Z(x_0)$ , located in  $x_0$  as a linear combination of the known value,  $N$ , for that variable  $Z(x_\alpha)$  located at other different spatial positions,  $x_\alpha = 1, \dots, N$ :

$$Z(x_0)^* = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha Z(x_\alpha). \quad (2.48)$$

The weights,  $\lambda_\alpha$ , (Eq. 2.48) should summarize two extremely important characteristics in spatial inference procedures: first, they should be sensitive to the distance between the known samples  $Z(x_\alpha)$  and the point to be estimated  $Z(x_0)$ ; while they should also be able to disaggregate clusters of experimental data in order to avoid biasing the estimate at the unknown location,  $x_0$ , by these groups of clustered samples.

### 2.2.2 Probabilistic Model of the Geostatistical Linear Estimator

In the probabilistic model described in Sect. 2.1, the unknown value  $Z(x_0)$ , and the neighboring experimental samples  $Z(x_\alpha)$ ,  $x_\alpha = 1, \dots, N$ , are interpreted as a RV located in  $x_0$  and  $x_\alpha$  and respectively.

If we assume the stationary hypothesis for the statistical moments related to the structural element bi-point (Sect. 2.1.1), the first moment of



each of these RVs is individually defined as (Eq. 2.49):

$$E\{Z(x_\alpha)\} = E\{Z(x_0)\} = m. \quad (2.49)$$

On the other hand, it is possible to assume each pair of RVs separated by the same distance vector  $\mathbf{h}$  has the same joint distribution laws:

$$\begin{aligned} F_{ZZ'}[Z(x_1); Z(x_1 + \mathbf{h})] &= \text{prob}\{Z(x_1) < z, Z(x_1 + \mathbf{h}) < z'\} \\ &= F_{ZZ'}[Z(x_2); Z(x_2 + \mathbf{h})] \\ &= F_{ZZ'}[Z(x); Z(x + \mathbf{h})]. \end{aligned} \quad (2.50)$$

This means any bivariate law depends exclusively on the distance vector  $\mathbf{h}$ —the distance between  $Z(x)$  and  $Z(x + \mathbf{h})$ —and not on the location  $x_0$ . The second-order stationarity implies the variogram and covariance,  $\gamma(\mathbf{h})$  and  $C(\mathbf{h})$ , are functions depending exclusively on the vector  $\mathbf{h}$ .

The linear estimator described by Eq. 2.48 is interpreted as a RV located in  $x_0$  and resulting from the linear combination of the variables  $Z(x_\alpha)$ ,  $x_\alpha = 1, \dots, N$ . Let  $\varepsilon(x_0)$  be the difference between the estimated value,  $Z(x_0)^*$ , and the real value,  $Z(x_0)$ , the error associated with estimating the value of  $Z(x)$  in  $x_0$  (Eq. 2.51):

$$\varepsilon(x_0) = Z(x_0)^* - Z(x_0) = \sum_{\alpha} \lambda_{\alpha} Z(x_{\alpha}) - Z(x_0). \quad (2.51)$$

The two quality criteria mentioned above may be expressed in terms of the mean and the variance of the new RV  $\varepsilon(x_0)$ :

$$(i) \text{ Unbiased condition: } E\{\varepsilon(x_0)\} = 0$$

The first quality criteria of the estimate,  $Z(x_\alpha)^*$ , is related to its expected value:

$$E\{Z(x_\alpha)^*\} = E\{Z(x_0)\} = m. \quad (2.52)$$

Ensuring, within the probabilistic formalism of this estimation model, there is no bias in the estimation  $E\{\varepsilon(x_0)\} = 0$  results in the following:

$$\begin{aligned} E\{\varepsilon(x_0)\} &= E\left\{\sum_{\alpha} \lambda_{\alpha} Z(x_{\alpha})\right\} - E\left\{\sum_{\alpha} Z(x_0)\right\} = 0 \\ \sum_{\alpha} \lambda_{\alpha} E\{Z(x_{\alpha})\} &= E\{Z(x_0)\}. \end{aligned} \quad (2.53)$$

Since the random function is stationary,  $E\{Z(x_{\alpha})\} = E\{Z(x_0)\} = m$ , the equality from Eq. 2.53 is ensured if the sum of the weights is equal to 1:

$$\sum_{\alpha} \lambda_{\alpha} = 1. \quad (2.54)$$

$$(ii) \text{ Variance minimization: } E\{[\varepsilon(x_0)]^2\}$$

The second quality factor of this estimator is related to the variance of the error  $\varepsilon(x_0)$ . Two estimators may have a null mean of  $\varepsilon(x_0)$ , but the minimum dispersion around the mean states the difference in terms of the quality of the estimators:

$$\begin{aligned} \text{var}\{\varepsilon(x_0)\} &= \text{var}\{Z(x_{\alpha})^* - Z(x_0)\} \\ &= E\left\{\left[\sum_{\alpha} \lambda_{\alpha} Z(x_{\alpha}) - Z(x_0)\right]^2\right\}. \end{aligned} \quad (2.55)$$

Decomposing the squared terms, we have:

$$\begin{aligned} &= E\left\{\sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} E\{Z(x_{\alpha}) \cdot Z(x_{\beta})\}\right\} + E\{Z(x_0)^2\} \\ &\quad - 2E\left\{\sum_{\alpha} \lambda_{\alpha} Z(x_{\alpha}) \cdot Z(x_0)\right\} \\ &= \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} E\{Z(x_{\alpha}) \cdot Z(x_{\beta})\} + E\{Z(x_0)^2\} \\ &\quad - 2 \sum_{\alpha} \lambda_{\alpha} E\{Z(x_{\alpha}) \cdot Z(x_0)\}. \end{aligned} \quad (2.56)$$

Once the covariance or variance ( $C(\mathbf{h})$  or  $\gamma(\mathbf{h})$ ), model is defined and validated for the entire study area,  $Z(x_0)$ , the variance of estimation of any estimator may be expressed as a

function of the covariance between the samples and the unknown location at which the estimation is performed:

$$\begin{aligned} \text{var}\{\varepsilon(x_0)\} &= C(0) + \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} C(x_{\alpha} x_{\beta}) \\ &\quad - 2 \sum_{\alpha} \lambda_{\alpha} C(x_{\alpha} x_0). \end{aligned} \quad (2.57)$$

### 2.3 Kriging Estimate<sup>1</sup>

Ordinary Kriging is the most common Kriging algorithm from a family of algorithms that comprise the following non-stationary estimators: simple Kriging, universal Kriging (also known as Kriging with trend), Kriging with external drift, co-Kriging, the estimator of probability distribution functions—indicator Kriging for categorical indicator variables, and the non-linear estimates—multi-Gaussian Kriging and disjunctive Kriging (Matheron 1965).

The linear geostatistical estimator (Eq. 2.48) named by ordinary Kriging is defined as a linear combination of  $N$  neighbors of  $x_0$ — $Z(x_{\alpha})$ ,  $\alpha = 1, \dots, N$ —that verifies the two criteria related to the estimation error  $\varepsilon(x_0) = Z(x_0)^* - Z(x_0)$ , unbiasedness  $E\{\varepsilon(x_0)\} = 0$  and minimum estimation variance (Eq. 2.58):

$$\min\{\text{var}(\varepsilon(x_0))\}. \quad (2.58)$$

The first criterion is reached by imposing a condition on the weights (Eq. 2.54).

$N$  partial derivatives to zero in order to  $\lambda_{\alpha}$  ( $\alpha = 1, \dots, N$ ) and solving the system of  $N$  equations with  $N$  unknowns by any mathematical method. However, since the solution of the  $N$  unknowns is conditioned by Eq. 2.54, the minimization of Eq. 2.57 may be solved by resorting to Lagrange formalism, which implies adding an extra equation to Eq. 2.57 and, consequently, an extra unknown (the Lagrange parameter  $\mu$ ) to Eq. 2.57:

$$\begin{aligned} \text{var}\{\varepsilon(x_0)\} &= C(0) + \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} C(x_{\alpha} x_{\beta}) \\ &\quad - 2 \sum_{\alpha} \lambda_{\alpha} C(x_{\alpha} x_0) + 2\mu \left[ \sum_{\alpha} \lambda_{\alpha} - 1 \right], \end{aligned} \quad (2.59)$$

the last term is null.

The minimization of Eq. 2.59 consists of calculating the  $N + 1$  partial derivatives to achieve  $\lambda_{\alpha}$  and  $\mu$ , using an equality to zero, obtaining the system of  $N + 1$  equations with  $N + 1$  unknowns in this way. The resulting solution is the  $N$  weights  $\lambda_{\alpha}$  that fulfil the unbiasedness condition (Eq. 2.54) while at the same time minimizing the estimation variance:

$$\begin{aligned} \frac{\partial \left[ E\{[Z(x_0)^* - Z(x_0)]^2\} + 2\mu \left[ \sum_{\alpha} \lambda_{\alpha} - 1 \right] \right]}{\partial \lambda_{\alpha}} &= 0, \alpha = 1, \dots, N \\ \frac{\partial \left[ E\{[Z(x_0)^* - Z(x_0)]^2\} + 2\mu \left[ \sum_{\alpha} \lambda_{\alpha} - 1 \right] \right]}{\partial \mu} &= 0. \end{aligned} \quad (2.60)$$

The development of the  $N$  first equations results in:

$$\begin{aligned} \frac{\partial \left[ C(0) + \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} C(x_{\alpha} x_{\beta}) - 2 \sum_{\alpha} \lambda_{\alpha} C(x_{\alpha} x_0) + 2\mu \left[ \sum_{\alpha} \lambda_{\alpha} - 1 \right] \right]}{\partial \lambda_{\alpha}} &= 0, \alpha = 1, \dots, N \\ 2 \sum_{\beta} \lambda_{\beta} C(x_{\alpha} x_{\beta}) - 2C(x_{\alpha} x_0) + 2\mu &= 0, \alpha = 1, \dots, N. \end{aligned} \quad (2.61)$$

Minimizing the estimation variance (Eq. 2.58) is ensured by the classic method of equating the

The last partial derivative in order to  $\mu$  results in the following equation:

$$\sum_{\alpha} \lambda_{\alpha} = 1. \quad (2.62)$$

<sup>1</sup>The geostatistical estimator was named Kriging by Georges Matheron (1965) as a tribute to the pioneering work of Danie G. Krige (1951).

Finally, the Kriging system of  $N + 1$  equations that allows calculation of the  $N$  weights  $\lambda_\alpha$  is the following:

$$\begin{cases} \sum_{\beta} \lambda_{\beta} C(x_{\alpha} x_{\beta}) + \mu = C(x_{\alpha} x_0), \alpha = 1, \dots, N \\ \sum_{\alpha} \lambda_{\alpha} = 1. \end{cases} \quad (2.63)$$

The minimum estimation variance is obtained by replacing Eq. 2.63 into Eq. 2.57:

$$\begin{aligned} \sigma_E^2 &= C(0) + \sum_{\alpha} \lambda_{\alpha} C(x_{\alpha} x_0) - \mu - 2 \sum_{\alpha} \lambda_{\alpha} C(x_{\alpha} x_0) \\ \sigma_E^2 &= C(0) - \sum_{\alpha} \lambda_{\alpha} C(x_{\alpha} x_0) - \mu. \end{aligned} \quad (2.64)$$

The Kriging system may also be described in terms of the variogram  $\gamma(\mathbf{h})$ , knowing that  $\gamma(\mathbf{h}) = C(0) - C(\mathbf{h})$ :

$$\begin{cases} \sum_{\beta} \lambda_{\beta} \gamma(x_{\alpha} x_{\beta}) - \mu = \gamma(x_{\alpha} x_0), \alpha = 1, \dots, N \\ \sum_{\alpha} \lambda_{\alpha} = 1, \end{cases} \quad (2.65)$$

with the estimation variance defined as:

$$\sigma_E^2 = \sum_{\alpha} \lambda_{\alpha} \gamma(x_{\alpha} x_0) + \mu. \quad (2.66)$$

### 2.3.1 Kriging System Resolution

In practice, the system of  $N + 1$  equations may be written in a matrix notation (Eq. 2.67). Considering  $K$  the covariance matrix between samples,  $M$  the second member matrix—the covariance between samples and the unknown location—and  $\lambda$  the weighting matrix:

$$[K] = \begin{pmatrix} C(x_1, x_1) & \cdots & C(x_1, x_N) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C(x_N, x_1) & \cdots & C(x_N, x_N) & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix},$$

$$[M] = \begin{bmatrix} C(x_1, x_0) \\ \vdots \\ C(x_N, x_0) \\ 1 \end{bmatrix} \quad [\lambda] = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_N \\ \mu \end{bmatrix}, \quad (2.67)$$

The Kriging system may be written as follows:

$$[K] \cdot [\lambda] = [M], \quad (2.68)$$

where the solution is achieved by inverting the matrix  $K$ :

$$[\lambda] = [K]^{-1} \cdot [M], \quad (2.69)$$

and

$$\sigma_E^2(x_0) = C(0) - [\lambda]^T \cdot [M]. \quad (2.70)$$

By defining  $[Z]$  as the vector of the values  $Z(x_{\alpha})$ ,  $[Z] = [z(x_1), \dots, z(x_N)]$ , the Kriging estimator  $Z(x_0)^*$  is given by Eq. 2.71:

$$Z(x_0)^* = [\lambda]^T \cdot [Z] = [M]^T \cdot [K]^{-1} \cdot [Z]. \quad (2.71)$$

Note that the Kriging variance ( $\sigma_E^2(x_0)$ , Eq. 2.66) depends exclusively on the location of the experimental data ( $x_{\alpha}$ ) against the location where the estimation is being performed ( $x_0$ ), and not on the values of the experimental data. In other words, the Kriging variance is not dependent on the property that is being modelled, but depends exclusively on the configuration of the experimental data against the location  $x_0$  (Goovaerts 1997).

All the different available Kriging-based techniques share some important properties:

- (1) Kriging is an exact interpolator, the values of the experimental data are honored in the interpolated model;
- (2) The interpolation is constrained by a spatial continuity model, represented by a variogram model in two-point geostatistics;
- (3) Kriging techniques are able to weigh differently isolated samples from clusters of samples (declustering);

- (4) Models interpolated with Kriging tend to reproduce the mean value of the experimental data in areas far from the experimental data location (Deutsch and Journel 1992).

## 2.4 Linear Estimation of Non-stationary Phenomena: Simple Kriging

The ordinary Kriging estimator assumes the mean of the variable  $Z(x)$  within the study area  $A$  is not known but constant. However, there are natural phenomena in which the values of a given property we aim to estimate are not homogeneous across the entire study area. In such cases, it can be said that a drift in the values of  $Z(x)$  exists and that the stationarity hypothesis of Eq. 2.49 is not verified for the entire  $A$ .

There are some linear estimation methodologies of  $Z(x)$  that take into account the way the values of  $Z(x)$  drift spatially its local means ( $m(x)$ ) vary: simple Kriging; Kriging with a trend model (universal Kriging) and Kriging with an external drift. Due to its importance in the following chapters, we will only refer to the simple Kriging estimate. For discussions of the other non-stationary Kriging estimates, the reader is referred to Journel and Huijbreghts (1978), Govaerts (1997), and Deutsch and Journel (1992).

The simple Kriging estimate is the most general Kriging algorithm in its non-stationary version. It assumes the mean of the set of RVs from the available experimental data and the locations not sampled within the study area is known.

In practice, this algorithm is applied where the theoretical formalism of the probabilistic model imposes the knowledge of the mean of the random function (as in the multi-Gaussian Kriging of a random function with null mean) or when there is good knowledge about the trend, or drift, of the natural phenomenon. In these cases, the values of the drift (if known for the entire field) as the local mean of the RV of the sampled and non-sampled values within the study area can be assumed.

If we consider the Kriging estimator  $Z(x_0)^*$  in its most general form—as a linear combination of the  $N$  data  $Z(x_\alpha)$ , then:

$$Z(x_0)^* = \lambda_0 \cdot 1 + \sum_{\alpha=1}^N \lambda_\alpha Z(x_\alpha). \quad (2.72)$$

For those non-stationary cases in terms of the first statistical moment of the RVs are known, but not constant, the unbiasedness condition is defined as:

$$E\{Z(x_0)\} = E\{Z(x_0)^*\} = \lambda_0 + \sum_{\alpha=1}^N \lambda_\alpha E\{Z(x_\alpha)\}, \quad (2.73)$$

which implies:

$$\lambda_0 = m_{x_0} - \sum_{\alpha=1}^N \lambda_\alpha m_{x_\alpha}. \quad (2.74)$$

By plugging Eq. 2.74 into Eq. 2.72 we obtain the simple Kriging estimate:

$$Z(x_0)^* - m(x_0) = \sum_{\alpha=1}^N \lambda_\alpha [Z(x_\alpha) - m(x_\alpha)], \quad (2.75)$$

in which the residual  $Z(x_0) - m(x_0)$  is estimated based on the residual between samples  $Z(x_\alpha) - m(x_\alpha)$ .

Note that in those situations, in which uncertainty about the knowledge of the drift phenomena allows matching the drift of the set of RVs, simple Kriging is a method for estimating the residuals. The variance of the error  $\varepsilon(x_0) = Z(x_0)^* - Z(x_0)$  may be written in function of the covariances (Eq. 2.57). For simple Kriging, the  $N$  weights are calculated by minimizing this variance, obtained by the  $N$  partial derivatives in order to  $\lambda_\alpha$ , resulting in the following system of  $N$  equations:

$$\sum_{\beta} \lambda_\beta C(x_\alpha x_\beta) = C(x_\alpha x_0), \alpha = 1, \dots, N, \quad (2.76)$$

resulting in the solution of  $N$  unknowns  $\lambda_\alpha$  and the consequent calculus of the estimation variance associated with simple Kriging:

$$\sigma_E^2(x_0) = C(0) - \sum_{\beta} \lambda_{\alpha} C(x_{\alpha} x_0). \quad (2.77)$$

$$E\{[Z_1(x_0)]_{CK}^* - Z_1(x_0)\} = 0, \quad (2.79)$$

given the stationarity of the first statistical moment of both variables:

$$E\left\{\sum_i a_i Z_1(x_i) + \sum_j b_j Z_2(x_j) - Z_1(x_0)\right\}, \quad (2.80)$$

$$\left[\sum_i a_i - 1\right] m_1 + \sum_j b_j m_2 = 0. \quad (2.81)$$

The following conditions imposed over the weights ensure the unbiasedness of the estimate:

$$\sum_i a_i = 1 \text{ and } \sum_j b_j = 0. \quad (2.82)$$

The estimation variance is given by:

$$\begin{aligned} \text{var}\{\varepsilon\} &= \text{var}\{[Z_1(x_0)]_{CK}^* - Z_1(x_0)\} = \text{var}\left\{\sum_{i=1}^{N_1} a_i Z_1(x_i) + \sum_{j=1}^{N_2} b_j Z_2(x_j) - Z_1(x_0)\right\} \\ &= \sum_{i=1}^{N_1} \sum_{j=1}^{N_1} a_i a_j C_{Z_1}(x_i, x_j) + \sum_{i=1}^{N_2} \sum_{j=1}^{N_2} b_i b_j C_{Z_2}(x_i, x_j) + \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} a_i b_j C_{Z_1 Z_2}(x_i, x_j) \\ &\quad + 2 \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} a_i b_j C_{Z_1 Z_2}(x_i, x_j) - 2 \sum_{j=1}^{N_1} a_j C_{Z_1}(x_j, x_0) \\ &\quad - 2 \sum_{j=1}^{N_2} b_j C_{Z_2}(x_j, x_0) + C_{Z_1}(x_0, x_0), \end{aligned} \quad (2.83)$$

sampled in  $N_2$  points. The linear estimate,  $Z_1(x_0)$ , in an unknown location,  $x_0$ , may be described by the following linear combination of the neighbor samples of both variables  $Z_1(x_i)$  and  $Z_2(x_j)$ :

$$[Z_1(x_0)]_{CK}^* = \sum_{i=1}^{N_1} a_i Z_1(x_i) + \sum_{j=1}^{N_2} b_j Z_2(x_j). \quad (2.78)$$

Equation 2.78 describes the co-Kriging estimator, which like Kriging should be non-biased and with minimum variance error. The unbiasedness condition implies that the expected value for the error is null:

where  $C_{Z_1}(\mathbf{h})$  and  $C_{Z_2}(\mathbf{h})$  are the covariances of  $Z_1(x)$  and  $Z_2(x)$ , respectively, and  $C_{Z_1 Z_2}(\mathbf{h})$  the cross-covariance between  $Z_1(x)$  and  $Z_2(x)$ . The weights  $a_i$  and  $b_j$  are calculated by minimizing the estimation variance (Eq. 2.83) with constraints from Eq. 2.82.

The Lagrange formalism may be applied to Eq. 2.83 as it was in Eq. 2.57:

$$\text{var}\{\varepsilon\} = 2\mu_1 \left(\sum_{i=1}^{N_1} a_i - 1\right) + 2\mu_2 \left(\sum_{j=1}^{N_2} b_j\right), \quad (2.84)$$

where:

$$\varepsilon = [Z(x_0)]_{CK}^* - Z(x_0). \quad (2.85)$$

The minimization of Eq. 2.83 is achieved by resorting to the  $N_1 + N_2$  partial derivatives with respect to the weights  $a_i$  and  $b_j$  equal to zero:

$$\begin{aligned} \frac{\partial(\text{var}\{\varepsilon\})}{\partial a_j} &= 0 \text{ for } j = 1, \dots, N_1 \\ \frac{\partial(\text{var}\{\varepsilon\})}{\partial b_j} &= 0 \text{ for } j = 1, \dots, N_2 \\ \frac{\partial(\text{var}\{\varepsilon\})}{\partial \mu_1} &= 2 \sum_{i=1}^{N_1} a_i - 1 \\ \frac{\partial(\text{var}\{\varepsilon\})}{\partial \mu_2} &= 2 \sum_{i=1}^{N_2} b_i. \end{aligned}$$

Resulting in the  $N_1 + N_2 + 2$  equations:

$$\begin{aligned} \sum_{i=1}^{N_1} a_i C_{Z_1}(x_i, x_j) + \sum_{i=1}^{N_2} b_i C_{Z_1 Z_2}(x_i, x_j) + \mu_1 &= C_{Z_1}(x_0, x_j) \quad j = 1, \dots, N_1 \\ \sum_{i=1}^{N_1} a_i C_{Z_1 Z_2}(x_i, x_j) + \sum_{i=1}^{N_2} b_i C_{Z_2}(x_i, x_j) + \mu_2 &= C_{Z_1 Z_2}(x_0, x_j) \quad j = 1, \dots, N_2 \\ \sum_{i=1}^{N_1} a_i &= 1 \text{ and } \sum_{i=1}^{N_2} b_i = 0. \end{aligned} \quad (2.86)$$

In sum, here are some practical notes regarding the co-Kriging estimate:

In practice, using an auxiliary variable through co-Kriging is only advantageous compared to the ordinary Kriging of the primary variable, if the primary variable is sub-sampled in comparison to the secondary variable and if both variables are correlated. For more on this topic, please see Sousa (1989), Wackernagel (1995), Bourgault and Marcotte (1991), Marcotte (1991), Myers (1982, 1984), Samper and Carrera (1990) and Goovaerts (1997).

However, there are a few points that should be noted about the co-Kriging estimate:

- (1) The co-Kriging estimate is the natural extension of the Kriging estimate to incorporate a secondary variable. As in ordinary Kriging, the first member matrix must be definite positive. This condition is satisfied if the individual covariance and cross-covariance are the ones described in Sect. 2.1.8;
- (2) The resolution of the system of equations for co-Kriging may have numerical instability problems if there are large differences in the variances of primary and secondary variables. In such cases, simple or cross-correlograms should be used instead of their variogram and covariance equivalents;
- (3) Theoretically, the co-Kriging estimator introduces estimation errors that are smaller than those introduced by ordinary Kriging. However, this advantage must take into

---

account the costs associated with modeling the cross-variograms, which are often just rough approximations and forced adjustments;

- (4) The co-Kriging estimate may be generalized for a group of auxiliary variables:

$$[Z_1(x_0)]_{CK}^* = \sum_{\alpha_1=1}^{N_1} \lambda_{\alpha_1} Z_1(x_{\alpha_1}) + \sum_{i=2}^{N_v} \sum_{i=2}^{N_2} \lambda_{\alpha_i} Z_i(x_{\alpha_i}). \quad (2.87)$$

The resolution of this estimate implies the knowledge of the  $(N_v + 1)^2 / 2$  variogram models.



## 2.6 Co-estimation with a Secondary Variable in a Much Denser Sample Grid: Collocated Co-kriging

In some cases, the secondary variable is much more abundant compared to the available number of samples or observations of the primary variable. This is so when the secondary variable is known for the entire area  $A$ , resulting in a 2D or 3D model that is intended to be used as conditioning data to estimate the primary variable. Note that the term model here refers to knowledge of the variable for the entire study area. When this happens, the co-Kriging system (Eq. 2.86) becomes unstable due to difference in the sampling density between the primary and secondary variables.

Moreover, the value of the secondary variable at the location at which the estimation is performed tends to minimize the effect from the primary variable samples located at great distances. In these cases, one possible solution is to retain the values of the secondary variable at the location being studied. This Kriging version is known as collocated co-Kriging (Xu et al. 1992; Almeida and Journel 1994).

By considering  $Z_1(x)$  to be the primary variable the values of which are known at  $N_1$  locations, and  $Z_2(x)$  the secondary variable known for the entire study area, the collocated co-Kriging estimate  $Z_1(x_0)$  is defined by:

$$[Z_1(x_0)]_{CK}^* = \sum_{\alpha_1=1}^{N_1} a_i Z_1(x_i) + b_0 Z_2(x_0). \quad (2.88)$$

Resulting in a system of equation of  $N_1 + 2$  equations:

$$\begin{aligned} \sum_{i=1}^{N_1} a_i C_{Z_1}(x_i, x_j) + b_0 C_{Z_1 Z_2}(x_i, x_0) + \mu_1 &= C_{Z_1}(x_0, x_j) \quad j = 1, \dots, N_1 \\ \sum_{i=1}^{N_1} a_i C_{Z_1 Z_2}(x_i, x_j) + b_0 C_{Z_2}(0) + \mu_2 &= C_{Z_1 Z_2}(0) \quad j = 1, \dots, N_1 \\ \sum_{i=1}^{N_1} a_i + b_0 &= 1.. \end{aligned} \quad (2.89)$$

When the variances of the primary and secondary variables are very different, the system of equations (Eq. 2.89) should be expressed in terms of correlograms.

## 2.7 Estimation of Local Probability Distribution Functions

The Kriging estimate presented above is an optimal solution for inferring mean global or local characteristics of a quantitative property. The resulting models are interpolated models of mean values that are traditionally suitable for characterizing variables homogeneously spatially distributed, i.e. variables with low variability in which the mean value is enough to represent it within a study area.

For heterogeneous variables, such as the internal petro-elastic properties of a hydrocarbon reservoir, since Kriging results in smooth models, the Kriging estimate is not enough to characterize their spatial distribution.

For such complex variables there are geostatistical models that aim to locally estimate the probability distribution function of a given property. These probability distribution functions are the basis for the sequential stochastic simulation methodologies introduced in the Chap. 3.

These can be used in the context of this book to map extreme values or to assess local uncertainty (Goovaerts 1997). But the indicator formalism as a method for estimating local probability distribution functions is based in the work developed by Switzer (1977). However, this method is difficult to implement and was replaced by alternative stochastic simulation processes for continuous variables, such as the

Gaussian formalism. Readers interested in the indicator formalism should read Goovaerts (1997) and Deutsch and Journel (1992).

### Multi-Gaussian formalism

This approach for estimating the local probability functions of a given variable consists in using a single model known for the distribution function of the group of RVs. It assumes that a group of RVs  $\{Y(x), x \in A\}$  follows a joint multi-Gaussian function. This is an easy way to estimate the local probability distribution function when compared to the indicator formalism, but this strong assumption may lead to some consistency problems with the available experimental data.

The probability distribution function at any location  $x_0$  is perfectly described by the conditional expected value and variance:

$$\begin{aligned} E\{Y(x_0)|Y(x_1)\dots Y(x_N)\}, \\ \text{var}\{Y(x_0)|Y(x_1)\dots Y(x_N)\}. \end{aligned} \quad (2.90)$$

Resulting in the Gaussian probability function at  $x_0$  as:

$$G(x_0; y) = G\left[\frac{y - E\{Y(x_0)|Y(x_\alpha), \alpha = 1, \dots, N\}}{\sqrt{\text{var}\{Y(x_0)|Y(x_\alpha), \alpha = 1, \dots, N\}}}\right]. \quad (2.91)$$

Under the multi-Gaussian assumption both first statistical moments (Eq. 2.90) are equal to the linear simple Kriging estimate (Eq. 2.75) and the corresponding Kriging variance (Eq. 2.77; Journel and Huijbregts 1978):

$$\begin{aligned} E\{Y(x_0)|Y(x_\alpha), \alpha = 1, \dots, N\} &= [Y(x_0)^*] \\ &= m(x_0) + \sum_{\alpha=1}^n \lambda_\alpha [Y(x_\alpha) - m(x_\alpha)] = \sum_{\alpha=1}^N \lambda_\alpha Y(x_\alpha), \end{aligned} \quad (2.92)$$

given that the means are known and constant:

$$\begin{aligned} m(x_0) &= m(x_\alpha) = 0, \\ \text{var}\{Y(x_0)|Y(x_\alpha), \alpha = 1, \dots, N\} &= \sigma_E^2(x_0). \end{aligned} \quad (2.93)$$

where the weights  $\lambda_\alpha$  are computed by the simple Kriging system (Eq. 2.76).

The probability distribution function in  $x_0$  is defined by the two parameters estimated by simple Kriging—mean and variance:

$$G(x_0; y) = G\left[\frac{y - [Y(x_0)^*]}{\sigma_E^2(x_0)}\right]. \quad (2.94)$$

### 2.7.1 Gaussian Transform of the Experimental Data

One of the greatest advantages of this approach concerns the simplicity of its implementation: the probability distribution function is defined for every location  $x_0$  with the simple Kriging estimate of  $Y(x_0)$ . However, we need a Gaussian transformation of the experimental data  $Z(x_\alpha)$ ,  $\alpha = 1, \dots, N$  to ensure the Gaussian marginal distribution at least:

$$Y(x_\alpha) = \Phi(Z(x_\alpha)), \alpha = 1, \dots, N, \quad (2.95)$$

where  $Y(x_\alpha)$  follows a Gaussian function with zero mean and variance one.

The Gaussian transform ( $\Phi$ ) may be calculated using a polynomial approximation—Hermite's polynomial (Matheron 1974; Muge 1982)—or by a simple graphical transform, which due to its simplicity is more suitable for this operation.

Given two distribution function of variables  $Z(x)$  and  $Y(x)$ :

$$\begin{aligned} F(x) &= \text{prob}\{Z(x) < z\}, \\ G(y) &= \text{prob}\{Y(x) < y\}, \end{aligned}$$

the value  $z$  corresponding to the Gaussian value  $y$  satisfies  $F(z) = G(y)$ .

Generalizing:

$$\begin{aligned} Y(x_\alpha) &= \Phi(Z(x_\alpha)) = G^{-1}(F(Z(x_\alpha))) \\ \alpha &= 1, \dots, N, \end{aligned} \quad (2.96)$$

with the Gaussian transformation of the experimental data  $Z(x)$ , with a probability distribution

function  $F(z)$  into Gaussian  $Y(x)$ , and assuming that these follow a joint multi-Gaussian probability function, all the formalisms previously described may be applied following this sequential approach:

- (1) The experimental data is transformed into Gaussian (Eq. 2.96);
- (2) After calculating the variograms of the transformed values  $Y(x_z)$  for each single point  $x_0$ , a local probability distribution function is calculated:

$$G(x_0; y) = \text{prob}\{Y(x_0) < y\}; \quad (2.97)$$

- (3) The values of the probability distribution  $F(z')$  for any threshold value  $z'$  are obtained by the inverse transform  $\phi$ : first the value of  $y'$  corresponding to  $z'$  is calculated:

$$y' = \phi(z') = G^{-1}[F(z')]. \quad (2.98)$$

Then, we may calculate  $F(x_0, z')$  from  $G(x_0; y')$  estimated by Eq. 2.94:

$$F(x_0, z') = G(x_0; y'). \quad (2.99)$$

If the threshold  $z'$  is not coincident with the values of experimental data  $Z(x_z)$ , and because  $F(x, z)$  is monotonous crescent, the inverse transformation may be calculated by a linear or power interpolation (Goovaerts 1997).

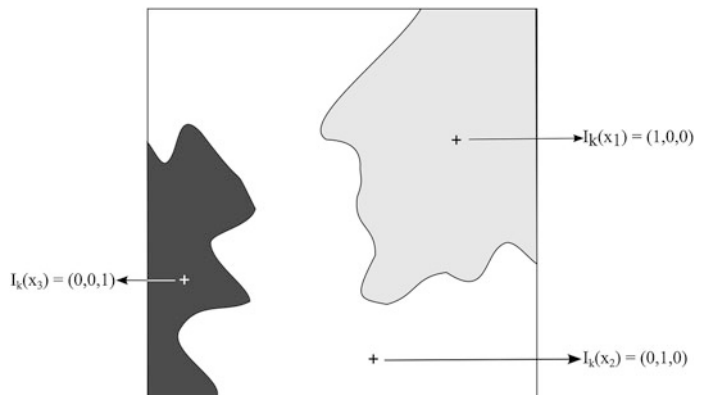
## 2.8 Estimation of Categorical Variables

When the petrophysical properties of a given hydrocarbon reservoir have some degree of homogenization, one can frequently classify the group of petrophysical properties through the concept of lithofacies. Lithofacies may be defined as geological bodies that share an identical behavior in terms of petrophysical and/or elastic response; this concept does not strictly refer to distinct types of lithologies (or sedimentary facies). In reservoir characterization, these lithofacies are frequently modelled using categorical variables, which are further modelled in terms of their internal properties and to continuous variables, such as porosity and permeability.

Within the conceptual geostatistical framework for categorical variables, the unit element consists in the probability of a point located within the study area that belongs to a group of complementary and disjunctive bodies. The shapes of the different bodies (or lithofacies) result from the classification of these elements with the greater probability of belonging to each body (or lithofacies). Let's assume a group  $K$  of disjunctive lithofacies,  $X_k = 1, \dots, K$ . For each point located in  $x$  within the study area  $A$  we may define a binary vector  $I_k(x)$  as follows (Fig. 2.20):

$$I_k(x) = \begin{cases} 1, & \text{if } x \in X_k \\ 0, & \text{if } x \in X_j \text{ and } j \neq k. \end{cases} \quad (2.100)$$

**Fig. 2.20** Schematic representation of a group of three lithofacies in a 2D model



From the multi-phase group, we may define the individual statistics for each single phase:

$$\begin{aligned} m_k &= E\{I_k(x)\}, \\ \sigma_k^2 &= \text{var}\{I_k(x)\}. \end{aligned} \quad (2.101)$$

For the group of different lithofacies we may define a measurement of average continuity of global structure— $C(\mathbf{h})$  (Eq. 2.102)—with the probability of two points  $x$  and  $x + \mathbf{h}$  with distance  $\mathbf{h}$ , belonging to the same lithofacies  $X_k$  for all  $k = 1, \dots, K$  (Soares 1992):

$$C(\mathbf{h}) = E\left\{\sum_{k=1}^K I_k(x) \cdot I_k(x + \mathbf{h})\right\}. \quad (2.102)$$

The multi-phase covariance (Eq. 2.102) may be decomposed by the sum of the single-phase covariance:

$$C(\mathbf{h}) = \sum_{k=1}^K E\{I_k(x) \cdot I_k(x + \mathbf{h})\} = \sum_{k=1}^K C_k(\mathbf{h}), \quad (2.103)$$

which may be written in the form of a variogram.

From the structural point of view, the covariance  $C(\mathbf{h})$  and the variogram  $\gamma(\mathbf{h})$  quantify the average morphological variability of the set of all lithofacies:

$$\gamma(\mathbf{h}) = \frac{1}{2} E\left\{\sum_{k=1}^K [I_k(x) - I_k(x + \mathbf{h})]^2\right\}. \quad (2.104)$$

Note that for most cases, when dealing with multi-phase structures there are not enough samples to estimate reliable individual variograms or covariances. This gets worse as the number of lithofacies increases. In this way the multi-phase variogram or covariance may be directly estimated from Eq. 2.105:

$$\begin{aligned} \gamma(\mathbf{h}) &= \frac{1}{KN(\mathbf{h})} \sum_{k=1}^K \sum_{\alpha=1}^{N(\mathbf{h})} [I_k(x_\alpha) - I_k(x_\alpha + \mathbf{h})]^2, \\ C(\mathbf{h}) &= \sum_{i=1}^K \sigma_k^2 - \gamma(\mathbf{h}). \end{aligned} \quad (2.105)$$

A multi-phase group may be composed of subgroups with distinct spatial behavior. This can be seen as having a single variable with a non-stationary behavior for the entire study area. By adopting an average global spatial continuity model there will be regional areas that are not modelled correctly. In such cases, each subgroup should be modelled independently, avoiding as much as possible, the integration within the same model subgroups with very distinct spatial behaviors.

With these structural tools for categorical variables, we are now ready to introduce a morphological estimation methodology for multi-phase structures (Soares 1992). With the geostatistical estimation of multi-phase structures, the aim is to calculate, for each point  $x_0$  within the study area  $A$ , the joint probability of  $x_0$  belonging to lithofacies  $X_k$  for all  $k = 1, \dots, K$  based on the  $I_i(x_\alpha)$  from the experimental samples  $x_\alpha = 1, \dots, N$ :

$$\begin{aligned} \text{prob}\{x_0 \in X_1\} &= [I_1(x_0)]^* = \sum_{\alpha} \lambda_{\alpha,1} I_1(x_\alpha), \\ \text{prob}\{x_0 \in X_2\} &= [I_2(x_0)]^* = \sum_{\alpha} \lambda_{\alpha,2} I_2(x_\alpha), \\ \text{prob}\{x_0 \in X_K\} &= [I_K(x_0)]^* = \sum_{\alpha} \lambda_{\alpha,K} I_K(x_\alpha). \end{aligned} \quad (2.106)$$

If when constructing these estimators, we use the same covariance model—multi-phase covariance—then the weights  $\lambda_\alpha$  are the same for all phases:

$$\lambda_{\alpha,1} = \lambda_{\alpha,2} = \dots = \lambda_{\alpha,K} = \lambda_\alpha. \quad (2.107)$$

The estimator  $[I_K(x_0)]^* = \sum \lambda_\alpha I_K(x_\alpha)$ ,  $k = 1, \dots, K$ , is calculated in such a way that it is not biased and the estimation variance is minimized:

Unbiasedness condition

$$\begin{aligned} E\{[I_k(x_0)]^*\} &= E\{I_k(x_0)\}, \\ \sum_{\alpha} \lambda_\alpha E\{I_k(x_\alpha)\} &= E\{I_k(x_0)\}, \end{aligned} \quad (2.108)$$

implies that

$$\sum_{\alpha} \lambda_{\alpha} = 1. \quad (2.109)$$

Consequently, from Eq. 2.106 the result is that the sum of the estimated probabilities of belonging to each one of the phases is 1:

$$\sum_{k=1}^K [I_k(x_0)]^* = \sum_{\alpha} \lambda_{\alpha} = 1. \quad (2.110)$$

In addition to the unbiasedness condition, Eq. 2.108 calculated the weights of the minimization in the sum of the estimation variances that is also imposed on each lithofacies individually:

$$\begin{aligned} & \min \left\{ E[[I_1(x_0)]^* - I_1(x_0)]^2 + E[[I_2(x_0)]^* - I_2(x_0)]^2 + \dots + E[[I_K(x_0)]^* - I_K(x_0)]^2 \right\} \\ & = \min \left\{ E \left[ \sum_k E[[I_k(x_0)]^* - I_k(x_0)]^2 \right] \right\} \end{aligned} \quad (2.111)$$

which we also may express in terms of multi-phase covariance  $C(\mathbf{h})$ :

$$\begin{aligned} & \sum_k E \left\{ [[I_k(x_0)]^* - I_k(x_0)]^2 \right\} \\ & = C(0) + \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} C(x_{\alpha} x_{\beta}) + \sum_{\alpha} \lambda_{\alpha} C(x_{\alpha} x_0). \end{aligned} \quad (2.112)$$

By minimizing this equation under the constraining from Eq. 2.109, we obtain the classical Kriging system with multi-phase covariance:

$$\begin{cases} \sum_{\beta} \lambda_{\beta} C(x_{\alpha} x_{\beta}) + \mu = C(x_{\alpha} x_0) & \forall \alpha = 1, \dots, N \\ \sum_{\alpha} \lambda_{\alpha} = 1. \end{cases} \quad (2.113)$$

The minimization of the sum of the variances does not directly imply the minimization of the variances for each class individually. This means that the best estimate of the multi-phase group may not be the best estimate of each class/lithofacies individually: this is only possible if each class is

estimated individually with independent covariance for each class.

When we can calculate individual variograms for each phase, they can be different, and as the weights are phase-dependent, the sum of probabilities estimated at a given point may not be 1. However, there are methods to overcome this limitation (Suro-Perez and Journel 1990).

Considering:

$$S_i = \sum_{\alpha} [I_{ki}(x)]^* \neq 1, \quad (2.114)$$

then the estimated probability for each phase is reconverted by factor

$$[I_K(x_0)]^{**} = \frac{[I_K(x_0)]^*}{S_i} \text{ being } \sum_k [I_k(x_0)]^{**} = 1.$$

Thus, both the multi-phase estimation and the individual estimation of each phase are valid methods for reaching the same goal—the spatial characterization of a multi-phase structure—but in different situations:

- The individual estimation of each phase has the advantage of taking into consideration the structural differences quantified by individual variogram or covariance models. However, in practice this is not often used since while the number of phases increase, the estimation of the individual variograms becomes more difficult as the number of samples per phase decreases.
- Using multi-phase variograms does not mean we need to use a single variogram model. Heterogeneous multi-phase groups may and should be modelled by different multi-phase variograms within the same estimation procedure.

Geostatistical Methods for Reservoir Geophysics

Azevedo, L.; Soares, A.

2017, XXVII, 141 p. 114 illus., 84 illus. in color.,

Hardcover

ISBN: 978-3-319-53200-4