

Article

Influence of the Sampling Density in the Coestimation Error of a Regionalized Locally Stationary Variable

Heber Hernandez Guerra ^{1,†} , Elisabete Alberdi ^{2,*,†} and Aitor Goti ^{3,†}¹ Department of Mining, University of Aconcagua UAC, 1700000 La Serena, Chile; heber.hernandez@uac.cl² Department of Applied Mathematics, University of the Basque Country UPV/EHU, 48013 Bilbao, Bizkaia, Spain³ Deusto Digital Industry Chair, University of Deusto, 48007 Bilbao, Bizkaia, Spain; aitor.goti@deusto.es

* Correspondence: elisabete.alberdi@ehu.eus; Tel.: +34-946-017-790

† These authors contributed equally to this work.

Received: 9 December 2019; Accepted: 18 January 2020; Published: 21 January 2020



Abstract: In the present study, the influence of the sampling density on the coestimation error of a regionalized, locally stationary and geo-mining nature variable is analyzed. The case study is two-dimensional (2D) and synthetic-type, and it has been generated using a non-conditional Sequential Gaussian Simulation (SGS), with subsequent transformation to Gaussian distribution, seeking to emulate the structural behavior of the aforementioned variable. A primary and an auxiliary variable with different spatial and statistical properties are constructed using the same methodology. The collocated ordinary cokriging method has been applied, in which the auxiliary variable is spatially correlated with the primary one and it is known exhaustively. Fifteen sampling densities are extracted from the target population of the primary variable, which are compared with the simulated values after performing coestimation. The obtained results follow a potential function that indicates the mean global error (MGE) based on the sampling density percentage (SDP) ($MGE = 1.2366 \cdot SDP^{-0.224}$).

Keywords: collocated ordinary cokriging; sampling density; regionalized; local stationary variables

1. Introduction

Mining is a business that bases its productive parameters and those of an economic–financial nature on the basis of fragmentary information, which must be used to limit the various existing uncertainties about the mineral resource, guaranteeing profitability [1]. Accurate evaluation of viable mineral resources is crucial in optimal sustainable development and mine planning procedures [2]. In mineral resource estimation, it is crucial the identification of the geological domains to be used for modeling [3]. Several types of variables, such as regionalized variables or spatial variables, are relevant [2,4,5] and they must be taken into account for business profitability maximization. Often, it is difficult to get reliable information of this type.

According to the Pan European Reserves and Resources Reporting Committee (PERC) Reporting Standard [6], a mineral resource is a concentration or occurrence of material of intrinsic economic interest in or on the earth's crust in the form and quantity in which there are reasonable probabilities of eventual economic extraction. These resources are studied through variables that by nature have spatial structure. This means that they are not randomly arranged, but rather have a certain order. This feature is distinctive of data from the field of geology, and they are called regionalized variables in geostatistics [7].

For Isaaks and Srivastava [8], geostatistics offers a way to describe spatial continuity, an essential feature of many natural phenomena that takes advantage from the classical regression techniques. When the limits of spatial dependence of the variable of interest are known, a linear, unbiased and optimal estimation technique called kriging is used [8,9]. This is a generic term applied to a variety of estimation methods that depend on minimizing the estimation error by a least squares procedure [10]. The name of the technique was introduced by Matheron [11], in honor of the pioneering works of D. G. Krige when estimating gold concentrations in South Africa [12]. When this technique is used in the mining industry, in most of the cases, the process only depends on a primary variable, and auxiliary variables (such as sub-elements, pollutants or others) are not considered. Pan et al. [13] state that cokriging is theoretically more attractive than univariate kriging when estimating simultaneously a set of correlated variables. Wackernagel [14] defines cokriging as the extension of kriging, where an auxiliary variable is used to improve the accuracy of the estimate. Due to its complexity when modeling joint spatial continuity, this coestimation technique has not received sufficient credit. Almeida and Journel [15] emphasize the use of more easily applicable but still precise techniques, and they introduce a simplified cokriging method called collocated ordinary cokriging.

The Sequential Gaussian Simulation (SGS) algorithm is also widely used because it is fast and efficient when building conditional cumulative distribution functions [16]. Techniques like fractal and multifractal modeling are also used along with SGS to separate various geological processes [17,18].

In this study, the collocated ordinary cokriging and the SGS are applied. When mapping spatial variables two important stages must be taken into account: the way in which sampling is done, and the techniques used in prediction. Both processes determine the accuracy of the results. Sampling aspects are analyzed in diverse works. It is known that purposive sampling is in general more efficient for model-based mapping [19]. In [20] the universal kriging variance is used as a quality measure to design samples. In [21] the size of the sample is studied, and they conclude that it is an important factor to obtain accurate empirical variograms, resulting crucial to sample sufficiently and without bias.

However, costs associated with obtaining information from the earth's crust are high, especially in the sector of hard rock mining in which the only way to extract samples is through mechanized methods. This is the reason this study analyzes the influence of the sampling density on the coestimation error in an area defined by the target population. The aim of this work is to calculate an approximation of the appropriate sampling size to obtain good results when applying the collocated ordinary cokriging technique.

2. Theoretical Framework

The nature of geological data allows examination of some characteristics of the whole population only in some specific cases. This is the reason in most of the cases part of the population is analyzed, from which inferences about the total population are made [22]. The objective is to estimate a parameter of the whole population using information of a sample.

Geostatistics is an applied science that focuses on modeling the spatial continuity of natural phenomena. This information is used to estimate values in not sampled locations using an unbiased optimal linear method. This rigorous mathematical formulation was born in France in the 1960s [11]. An essential aspect of geostatistical modeling is the establishment of quantitative measures of spatial variability or continuity, which are subsequently used as input data in the estimation. Modeling the spatial variability is a standard process of mineral resource analysts, where for the past 25 years, the experimental semivariogram has been the most applied tool in mining [23].

Geostatistics studies regionalized phenomena, i.e., phenomena that in the geographical space follow a certain structure. Sinclair and Blackwell [10] define regionalized variables as variables distributed in a partially structured manner in the space, such that there is some degree of spatial self-correlation. These variables can be interpreted as a natural extension of random variables when they depend on the position. A regionalized variable, at each point x belonging to a domain $D \subset \mathbb{R}^d$

in space, being d the dimension of the space ($d = 2$ for 2D and $d = 3$ for 3D), makes it correspond a variable $Z(x)$ that depends on the position of the point x in space. The regionalized variable is stationary if its joint distribution function is invariant by translation in space.

The semivariogram is the basic geostatistical tool to measure the spatial self-correlation of a regionalized variable [24]. It is used in different directions of the space to study the continuity of the variable. In almost all cases of reservoir modeling, there is a need to model the joint distribution of multiple variables [25]. In this context, cross semivariograms are defined.

Interpolation algorithms are used to map a primary variable (variable of interest). Some interpolation algorithms are based on regression of the unknown value on the sample data. Kriging is one of these regression algorithms. It is a geostatistical estimation technique that gives the best unbiased linear estimator of a variable. Known certain values of a variable of interest Z_1 , $Z_1(x_1)$, $Z_1(x_2)$, $\dots$, $Z_1(x_n)$, the unknown value at point x_0 , $Z_1^*(x_0)$, is estimated by a linear combination:

$$Z_1^*(x_0) = \sum_{i=1}^n \lambda_i Z_1(x_i), \quad (1)$$

being λ_i the estimation weights calculated in a way that they minimize the variance.

Sometimes, the sampling of the primary variable is poor but there are other variables (auxiliary variables) more densely sampled. Assume that $Z_1(x)$ and $Z_2(x)$ are two regionalized variables. Variable Z_1 will be the primary variable and $Z_2(x)$ will work as the auxiliary one. Known certain values of the first variable $Z_1(x_1)$, $Z_1(x_2)$, $\dots$, $Z_1(x_n)$, and the auxiliary variable $Z_2(x_1)$, $Z_2(x_2)$, $\dots$, $Z_2(x_m)$, ordinary cokriging consists of calculating $Z_1^*(x_0)$ as a linear combination of data from both variables located at sample points in the vicinity of the point:

$$Z_1^*(x_0) = \sum_{i=1}^n \lambda_i^{(1)} Z_1(x_i) + \sum_{j=1}^m \lambda_j^{(2)} Z_2(x_j). \quad (2)$$

Each variable is defined in different number of samples. Cokriging is the extension of kriging to the multivariate case, i.e., it considers not only the variable to be estimated in space, but also the information of one or several additional variables in nearby sites [26]. Although cokriging has the same characteristics as kriging, considering additional variables improves the precision of the estimation in some cases, obtaining also more consistent results than when each variable is estimated independently using kriging in a multivariate study. However, it should be noted that it makes estimation more difficult.

Cokriging has several variants and collocated ordinary cokriging is one of them [27]. It can be used when primary data are available in sparsely distributed points while secondary data are present in all points of the mesh. According to Xu et al. [28] a full cokriging approach may cause matrix instabilities due to high autocorrelation between closer secondary data, as opposed to large separation and low autocorrelation in primary data. This problem can be overcome by retaining at each location to be estimated the collocated secondary data. The collocated ordinary cokriging estimator is given by [28]:

$$Z_1^*(x_0) - m_1 = \sum_{i=1}^n \left(\lambda_i^{(1)} Z_1(x_i) - m_1 \right) + \lambda^{(2)} (Z_2(x_0) - m_2), \quad (3)$$

being m_1 and m_2 the means of the primary and secondary variables, $Z_2(x_0)$ is the secondary variable sampled at the same location in which the primary variable will be estimated. The weights $\lambda_i^{(1)}$ for $i = 1, 2, \dots, n$ and $\lambda^{(2)}$ are calculated from the system:

$$\begin{aligned} \sum_{j=1}^n \lambda_j^{(1)} C_{12}(x_j - x_0) + \lambda^{(2)} C_2(0) &= C_{12}(0), \\ \sum_{j=1}^n \lambda_j^{(1)} C_1(x_j - x_i) + \lambda^{(2)} C_{21}(x_0 - x_i) &= C_1(x_0 - x_i), \quad i = 1, 2, \dots, n, \end{aligned} \quad (4)$$

where C_1 and C_2 are the covariances of the primary and secondary variables, respectively; and C_{12} and C_{21} are the cross-covariances (being $C_{12}(h) = C_{21}(h)$). These can be obtained by calculating the variograms and cross-variograms. Because of the complexity of this process, Markov-type screening hypothesis are considered [29]. In Markov's model the influence of any primary data $Z_1(x+h)$ on the secondary collocated variable $Z_2(x)$ is screened by the primary data $Z_1(x)$. In this context, the cross-covariance can be calculated using:

$$C_{12}(h) = \frac{C_{12}(0)}{C_{11}(0)} C_{11}(h), \forall h, \quad (5)$$

or equivalently:

$$\rho_{12}(h) = \rho_{12}(h) \rho_1(h), \forall h, \quad (6)$$

being $\rho_1(h)$ the correlogram function of the primary variable and $\rho_{12}(h)$ the cross-correlogram function between primary and secondary variables.

Under the Markov model, the collocated cokriging equations (3) and (4) can be written as:

$$\frac{Z_1^*(x_0) - m_1}{\sigma_1} = \sum_{i=1}^n \frac{(\lambda_i^{(1)} Z_1(x_i) - m_1)}{\sigma_1} + \lambda^{(2)} \frac{(Z_2(x_0) - m_2)}{\sigma_2}, \quad (7)$$

and:

$$\begin{aligned} \sum_{j=1}^n \lambda_j^{(1)} \rho_{12}(0) \rho_1(x_j - x_0) + \lambda^{(2)} &= \rho_{12}(0), \\ \sum_{j=1}^n \lambda_j^{(1)} \rho_1(x_j - x_i) + \lambda^{(2)} \rho_{12}(0) \rho_1(x_0 - x_i) &= \rho_1(x_0 - x_i), i = 1, 2, \dots, n, \end{aligned} \quad (8)$$

where σ_1 and σ_2 are the two stationary standard deviations of the variables Z_1 and Z_2 respectively.

3. Methodology

This section summarizes the most relevant stages of the study, which are implemented using the GSLIB algorithms, acronym for "Geostatistical Software Library", developed in the Petroleum Engineering Department of Stanford University. It is considered one of the most relevant computer developments in the history of geostatistics by some researchers [30]. The whole process is computed in Python by the authors. The sequence of the main stages of this study is represented in Figure 1.

3.1. Synthetic Case Study Creation

First, a synthetic case study has been created. This option has been chosen because it enables having data at the population level (universe), making possible to compare this reality with estimations that depend on different sampling densities. In this way, the real error can be calculated, and the influence of the sampling density can be measured. Thus, an area of 1 km² is defined as spatial domain with coordinates ranging from the origin to 1 km in the East (X) and in the North (Y). A 200 × 200 squared geometry is considered, thus, 40,000 cells of 25 m² are considered to be target population.

In each cell a primary variable (Z_1) and an auxiliary one (Z_2) are created following the Sequential Gaussian Simulation, and using a random number generator (random seed) and spatial structure parameters. In this first stage, both variables are measured in the whole population and they are spatially correlated. Hence, conditions required by the collocated ordinary cokriging technique are satisfied. On the other hand, real values of Z_1 are also known, making possible the comparison with the estimations obtained for different sampling densities.

3.2. Semivariographic Modeling of Variables

The spatial structure of Z_1 and Z_2 is described using the experimental semivariogram. The spatial continuity is modeled and parameters such as the anisotropy direction, features of the ellipse and plateau (maximum semivariance) are obtained. These parameters are required in the next stage.

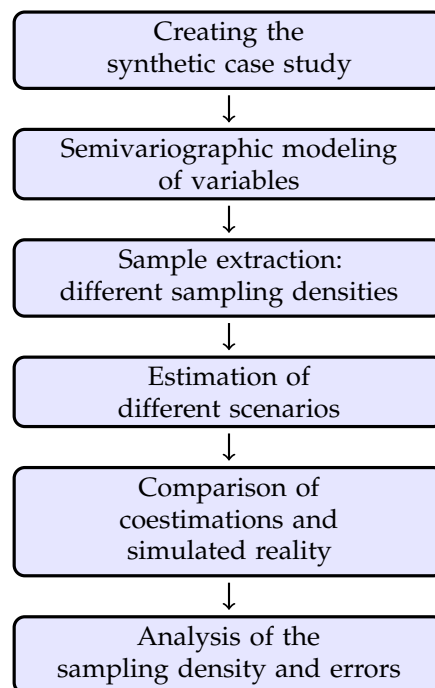


Figure 1. Sequence of the main stages of the study.

3.3. Sample Extraction at Different Densities

From the exhaustive variable Z_1 , 15 different sampling densities are extracted. The minimum is equivalent to 0.06% of the population and the maximum is the 5% of the population. These scenarios have been used to develop the study. The spatial arrangement of the samplings is regular, it does not have clusters.

3.4. Estimation of Different Scenarios

Results of 15 scenarios are estimated using the collocated ordinary cokriging method. The technique uses parameters that have been calculated in the structural modeling, the spatial correlation between the variables and the scopes defined for sample search.

3.5. Comparison of Estimations and Simulated Reality

With the purpose of evaluating the representativeness of the sample and the accuracy of the technique, the estimated results are compared with the simulated reality of the variable Z_1 , using the mean squared error (MSE).

3.6. Analysis of the Sampling Density and Errors

At this stage, the relationship between sampling density and the coestimation errors is analyzed. They result inversely proportional, and it can be expressed by a potential function restricted to a maximum sample size equivalent to 5% of the target population. It must be taken into account that it is difficult to have bigger samples than the indicated above in the evaluation of metallic mineral deposits.

4. Application

The first stage is the creation of the synthetic case study. Two regionalized and spatially correlated variables in a two-dimensional space are created. Both variables, the primary Z_1 and auxiliary Z_2 have a location in the georeferenced space in East (X) and North (Y). The unconditional SGS method has been used, taking into account a spatial structure and a random seed. This process is developed using the SGSIM algorithm (Sequential Gaussian simulation program) of GSLIB reimplemented in Python

code, with data presented in Tables 1 and 2. Table 2 contains the structural anisotropic-type parameters for both variables. The contribution refers to the plateau of the semivariogram (the value at which the semivariogram stops changing).

Table 1. Parameters to create the area of study.

Parameter	Value	Unit of Measurement
X coordinate in the origin	0	meters (m)
Y coordinate in the origin	0	meters (m)
Last coordinate in X	1000	meters (m)
Last coordinate in Y	1000	meters (m)
Cell width in X	5	meters (m)
Cell width in Y	5	meters (m)
Number of cells in X	200	unities
Number of cells in Y	200	unities
Total number of cells	40,000	unities

Table 2. Structural parameters of Z_1 and Z_2 .

Parameter	Value	Units of Measurement
Anisotropy azimuth Z_1	45	degrees
Anisotropy azimuth Z_2	45	degrees
Maximum value of Z_1	500	meters (m)
Minimum value of Z_1	200	meters (m)
Maximum value of Z_2	300	meters (m)
Minimum value of Z_2	80	meters (m)

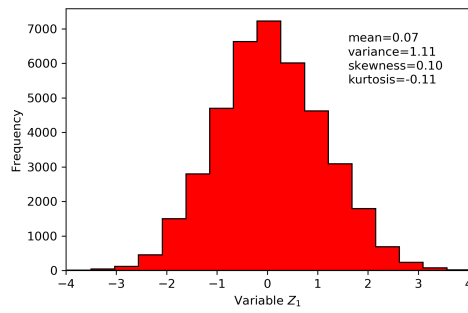
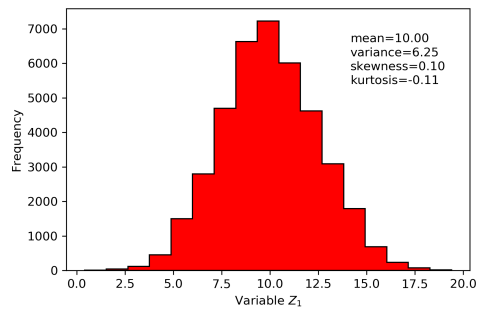
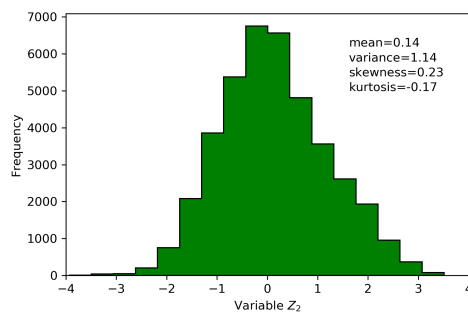
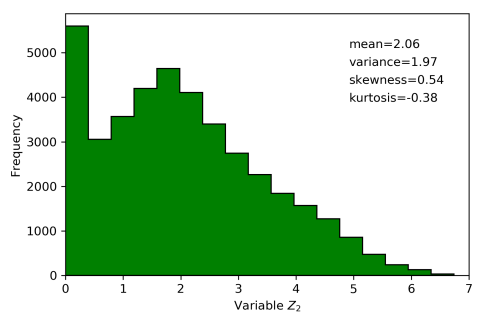
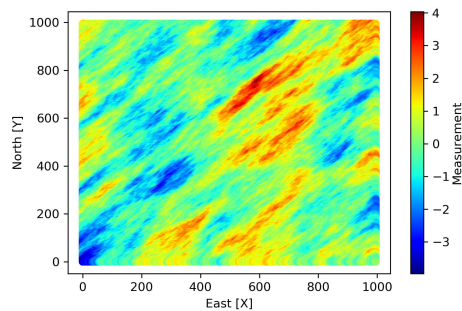
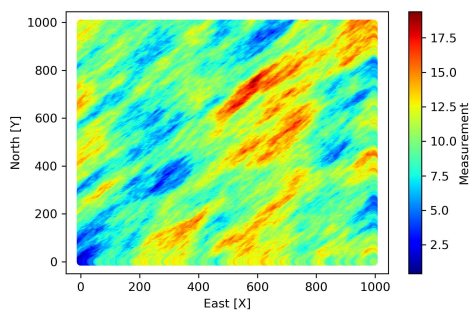
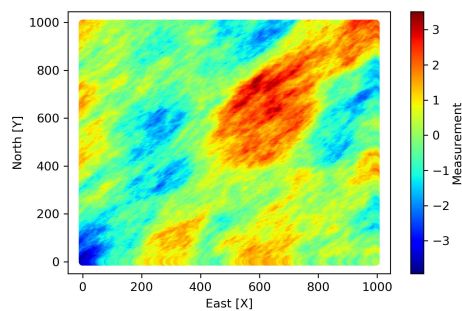
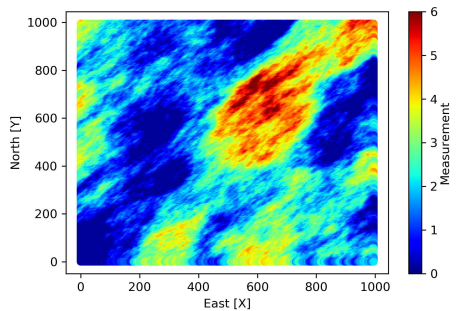
Subsequently, a corrective affine transformation is applied to the distribution of the variable correcting the mean and the standard deviation. Table 3 contains the statistics of form and dispersion applied in the correction of the Gaussian distributions of the simulation.

Table 3. Corrective affine distribution of Z_1 and Z_2 .

Parameter	Value
Mean of Z_1	10
Mean of Z_2	2
Standard deviation of Z_1	2.5
Standard deviation of Z_2	1.5

For Z_2 , the 47% of the corrected data were negative and they were replaced by a positive small value. Thus, the histogram has a positive asymmetric form, see Figure 2. The simulations for both variables are shown in Figure 3.

With the aim of analyzing the behavior of the primary variable, a diagonal cut is performed from southwest to northeast on the map of this variable (45° azimuth), coinciding with the direction of anisotropy. We observe the existence of local stationarity between 500 and 900 meters approximately, see Figure 4. Armstrong [31] states that this assumption of quasi-stationarity is essential between the homogeneity scale of the phenomenon and the sampling density.

(a) Histogram, Gaussian Z_1 .(b) Histogram, corrected Z_1 .(c) Histogram, Gaussian Z_2 .(d) Histogram, corrected Z_2 .**Figure 2.** Histograms, Gaussian variables and transformed variables.(a) Gaussian Z_1 map.(b) Corrected Z_1 map.(c) Gaussian Z_2 map.(d) Corrected Z_2 map.**Figure 3.** Maps of Z_1 and Z_2 with Gaussian sequential simulation and corrective affine transformation.

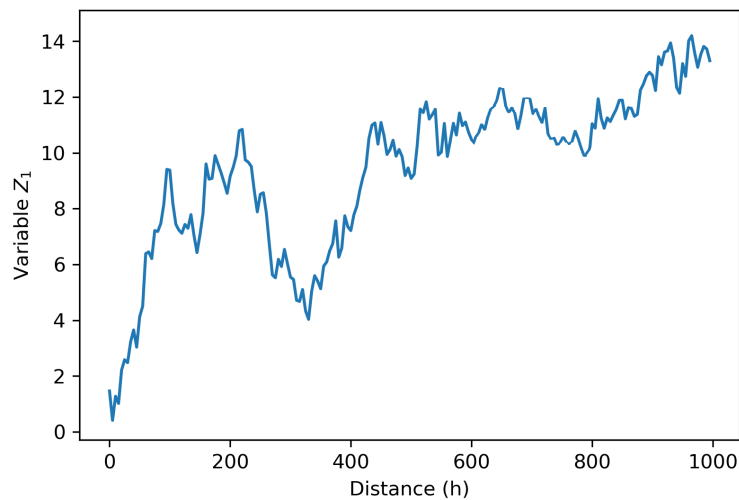


Figure 4. Local stationarity behavior of Z_1 .

In the second stage, the data required for the estimation processes is prepared. The collocated ordinary cokriging method with a Markov model is used. The required data are the following:

- Semivariographic fitting model for the primary variable (Z_1).
- Semivariographic fitting model for the exhaustive auxiliary variable (Z_2).
- Linear correlation coefficient between Z_1 and Z_2 .

At this point both Z_1 and Z_2 are exhaustive. The auxiliary variable that contributes in the estimation process of the primary variable is static. The semivariographic fitting models are described according to the target population of each variable and they will remain constant during the estimation processes. In Figure 5 the semivariograms of the variables are presented.

From Figure 5, a zonal anisotropy is concluded in 45° azimuth direction as the main axis, for both variables Z_1 and Z_2 . In both variables, there is a slower variability in the direction of 45° azimuth than in its orthogonal direction (135° azimuth), where the variables present a faster variability. The experimental data of Z_1 and Z_2 fit two nested spherical structures: $\gamma_1(h) = 6.0 \cdot Sph(300, 250) + 0.8 \cdot Sph(\infty, 250)$ for Z_1 ; and $\gamma_2(h) = 2.0 \cdot Sph(500, 300) + 0.75 \cdot Sph(\infty, 300)$ for Z_2 . The linear correlation coefficient between Z_1 and Z_2 will vary depending on the sampling density.

The third stage consists of extracting 15 different sampling densities (SD) from the exhaustive variable Z_1 . These values will be combined with the values of Z_2 to create the scenarios of estimation. Table 4 shows the sampling densities for Z_1 . It can be observed that at higher sampling densities the error in statistical characterization is lower. The sampling density is constant for Z_2 , i.e., all the spatial area of 1 km^2 is considered for Z_2 . A maximum sample size of 5% of the population has been considered, associated with high costs when retrieving information of metallic elements from the earth's crust. It is typical to have case studies with smaller sample size than the 1% of the target population [17,32]. The limit scenarios (lowest and highest sampling densities) are presented in Figure 6.

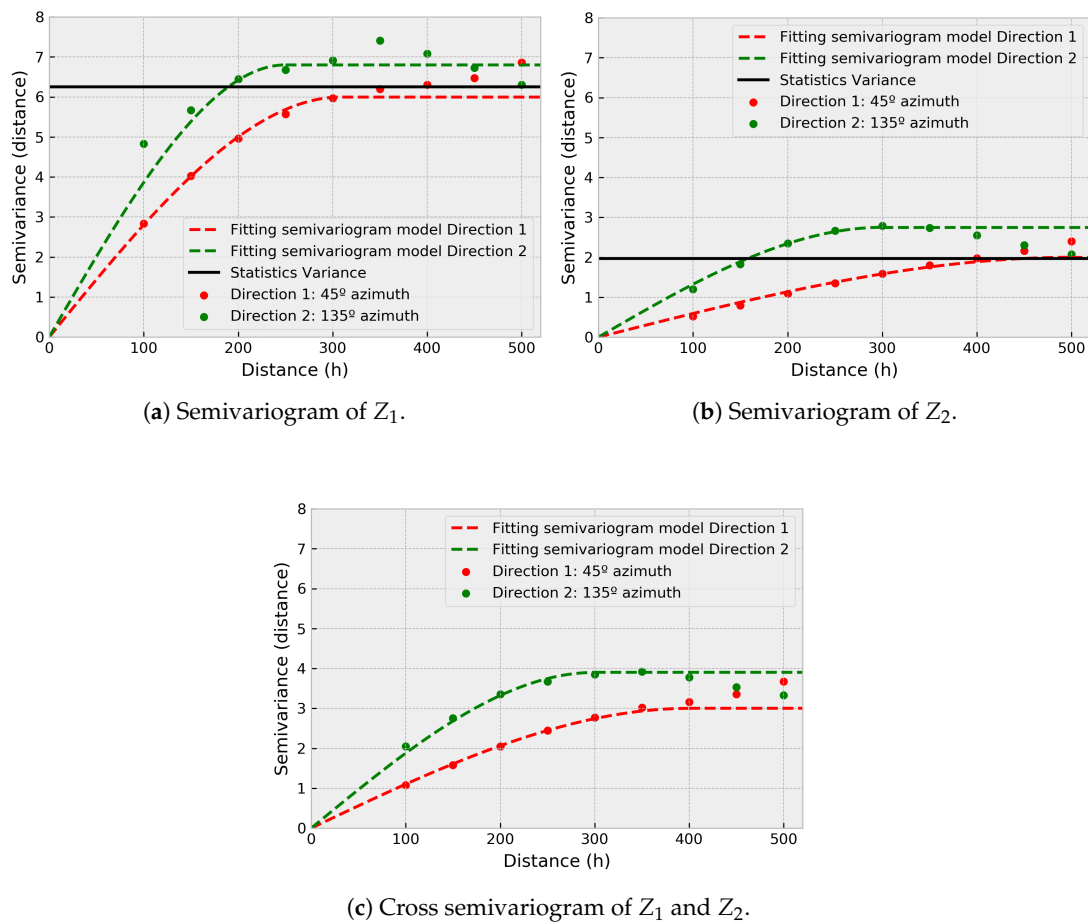


Figure 5. Semivariograms of the variables.

Table 4. Different scenarios depending on sampling densities.

Scenario	Number of Samples	Sampling Density Z_1 (m^2)	Sampling Size Z_1 (%)	Mean	Variance	Deviation
1	25	625	0.06	10.16	2.97	1.72
2	40	1000	0.10	10.16	2.43	1.56
3	60	1500	0.15	9.86	2.55	1.60
4	70	1750	0.18	9.87	2.88	1.70
5	80	2000	0.20	10.11	3.32	1.82
6	120	3000	0.30	10.10	3.32	1.82
7	160	4000	0.40	10.15	3.60	1.90
8	200	5000	0.50	9.91	4.67	2.16
9	240	6000	0.60	9.98	4.51	2.12
10	280	7000	0.70	9.95	4.77	2.18
11	320	8000	0.80	9.95	4.95	2.23
12	360	9000	0.90	10.01	5.14	2.27
13	400	10,000	1.00	10.02	5.25	2.29
14	1600	40,000	4.00	10.02	5.41	2.33
15	2000	50,000	5.00	10.01	5.55	2.36

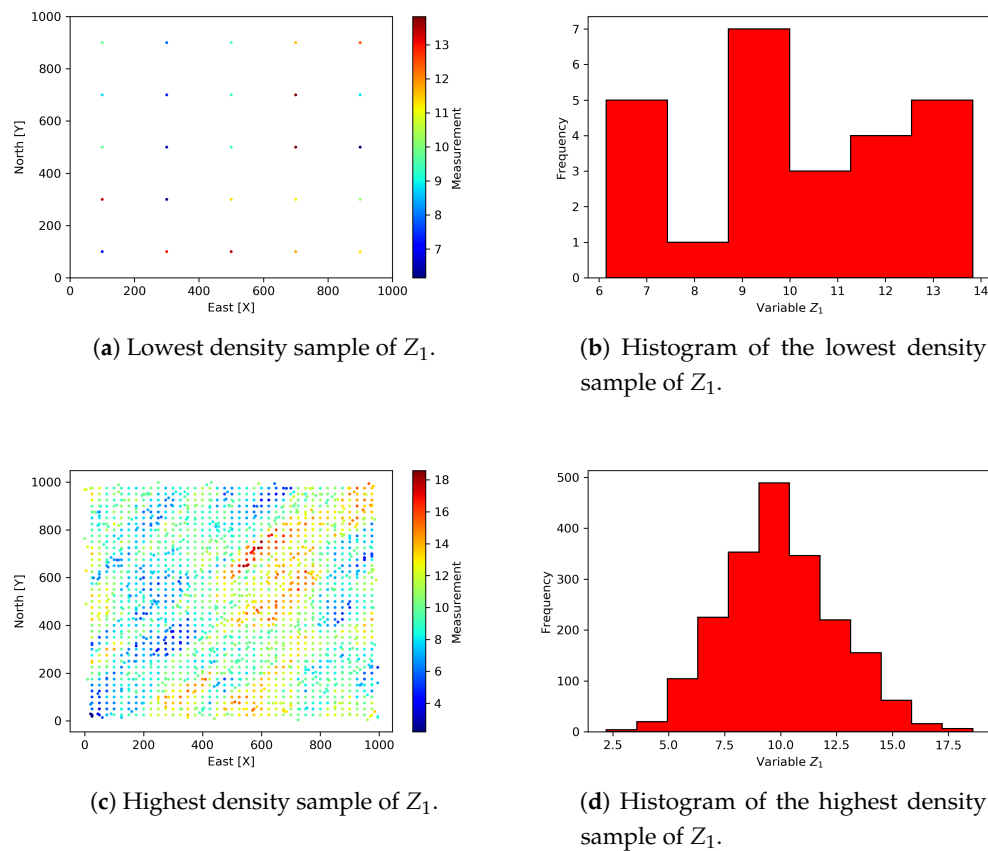


Figure 6. Maps and histogram of samples of Z_1 , limit scenarios.

In the fourth stage, the collocated ordinary cokriging process for each scenario is developed (Table 5), varying the linear correlation coefficient depending on the sampling density. Before performing the coestimation process, it is necessary to define the search neighborhood, determining the spatial limits that restrict the number of observations that interact in the coestimation. In a study about the kriging method, Rivoirard [33] points out that the higher the search limits, the greater the accuracy in the estimation process. However, the latter is limited to the number of existing observations and their distances over the estimation grid. For example, it would not be adequate to estimate at point (0,0) for the lowest density sample of Z_1 (in an univariate assumption) in the mesh of samples in Figure 6 using the existing 25 observations, given that the influence of those farther observations will hardly describe the reality at that point, unless there exists a complete stationarity in the study area. In this case, the information obtained from the semivariogram is used to define the search limits, reducing the size of the ellipse given the increase of the sampling density. The anisotropy direction for the searched ellipse is 45° azimuth and it varies based on the amount of information available in each scenario. The number of samples used for estimating Z_1 varies also based on the amount of information available. In Table 5 the minimum and the maximum number of samples for coestimation are presented. The ellipse will be larger in lower sampling density scenarios because of the absence of samples close to the sites to be estimated.

Table 5. Different scenarios depending on sampling densities.

Scenario	Sampling Density	Linear correlation Coefficient	Ellipse	Minimum Number of Samples Z_1	Maximum Number of Samples Z_1
1	625	0.91	300/150	1	4
2	1000	0.96	300/150	1	12
3	1500	0.92	300/150	1	20
4	1750	0.87	300/150	1	20
5	2000	0.88	300/150	1	20
6	3000	0.88	300/150	1	20
7	4000	0.84	300/80	1	20
8	5000	0.88	300/80	1	20
9	6000	0.86	300/80	1	20
10	7000	0.85	300/80	1	20
11	8000	0.85	200/80	1	20
12	9000	0.91	150/80	1	20
13	10,000	0.89	150/80	1	20
14	40,000	0.90	150/40	1	20
15	50,000	0.90	150/40	1	30

5. Results

The estimation scenarios are compared with the simulated real map from which the samples of Z_1 are extracted. It can be observed that smoothing is more noticeable in those scenarios with lower sampling density, see Figure 7. The coestimation is perfect when the linear correlation coefficient between the real scenario and the estimated scenario is 1. The linear correlation coefficient will depend on the number of observations and the semivariographic model that can be fitted according to these observations. Not very dense scenarios will poorly describe the spatial correlation, obtaining less accurate results. In Figure 8 the dispersion diagrams between the real variable Z_1 and the 15 scenarios are represented, indicating the linear correlation coefficient. The linear correlation coefficient of scenario 11 (320 observations) is 0.88 and the map of the variable Z_1 respects the real spatial structure. Scenarios with higher sampling density present also good features.

In Figure 9 the direct semivariograms of Z_1 are shown. It can be observed that they are similar to the real one for densities equivalent to 0.8% or greater, which is associated with the fact that in these cases the mathematical adjustment should also be similar to the real with independence of the modelization technique. When the sampling density decreases, the direct semivariogram of Z_1 becomes more erratic, decreasing its spatial correlation.

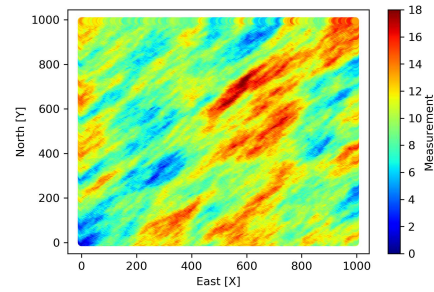
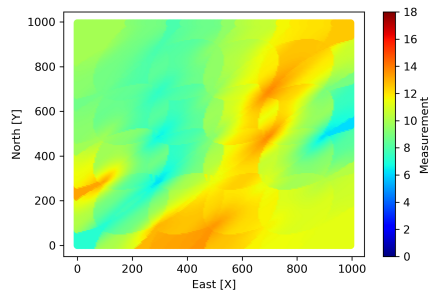
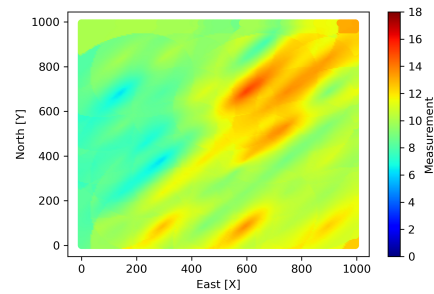
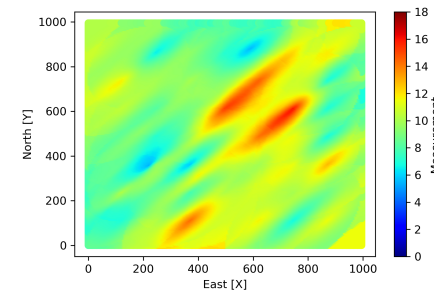
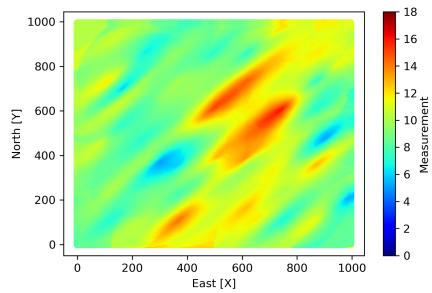
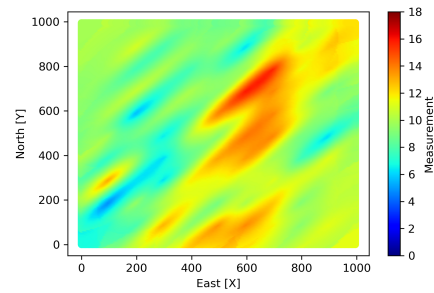
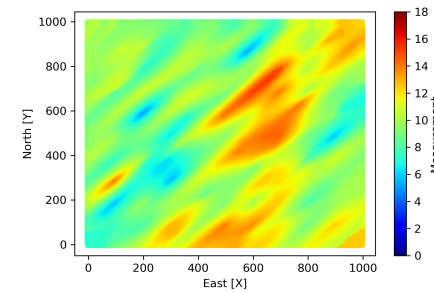
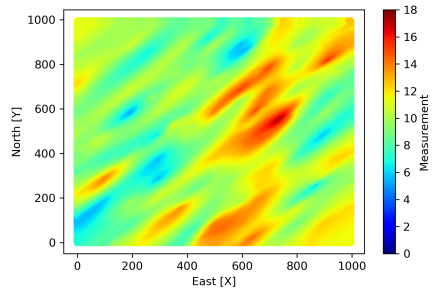
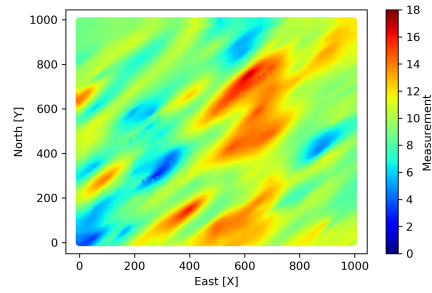
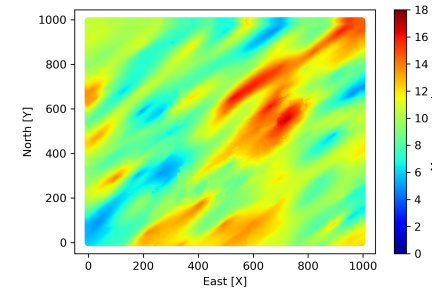
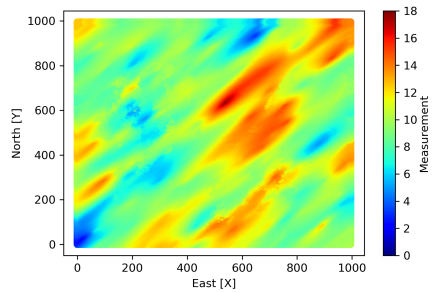
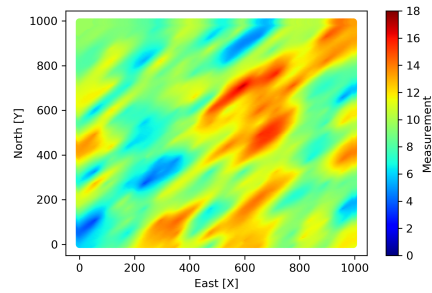
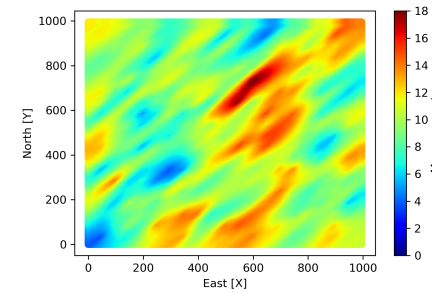
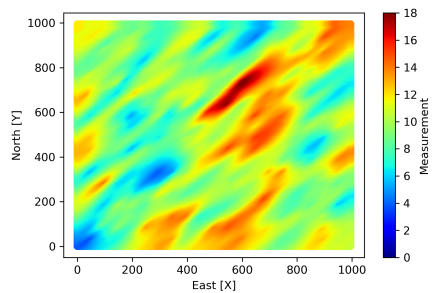
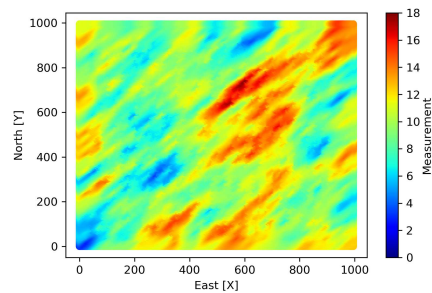
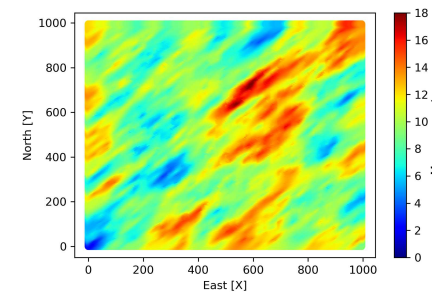
(a) Real map Z_1 .(b) Scenario 1, $SD=625 \text{ m}^2$.(c) Scenario 2, $SD=1000 \text{ m}^2$.(d) Scenario 3, $SD=1500 \text{ m}^2$.(e) Scenario 4, $SD=1750 \text{ m}^2$.(f) Scenario 5, $SD=2000 \text{ m}^2$.(g) Scenario 6, $SD=3000 \text{ m}^2$.

Figure 7. Cont.

(h) Scenario 7, $SD=4000 \text{ m}^2$.(i) Scenario 8, $SD=5000 \text{ m}^2$.(j) Scenario 9, $SD=6000 \text{ m}^2$.(k) Scenario 10, $SD=7000 \text{ m}^2$.(l) Scenario 11, $SD=8000 \text{ m}^2$.(m) Scenario 12, $SD=9000 \text{ m}^2$.(n) Scenario 13, $SD=10,000 \text{ m}^2$.(o) Scenario 14, $SD=40,000 \text{ m}^2$.(p) Scenario 15, $SD=50,000 \text{ m}^2$.Figure 7. Coestimation maps versus simulated real Z_1 .

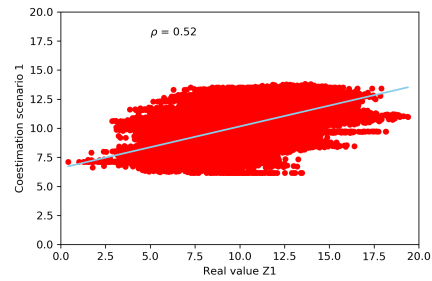
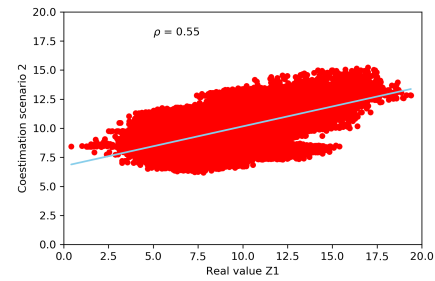
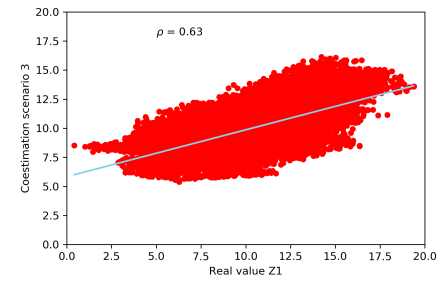
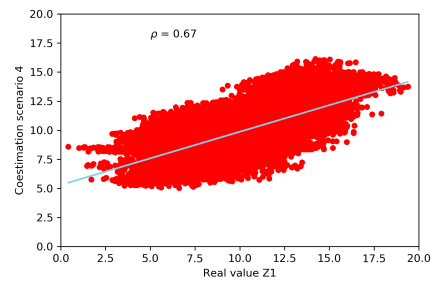
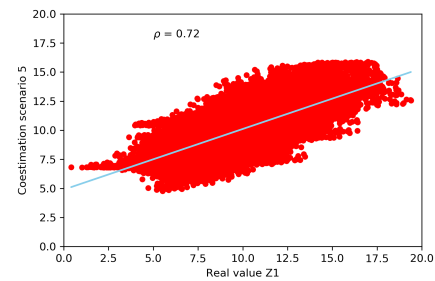
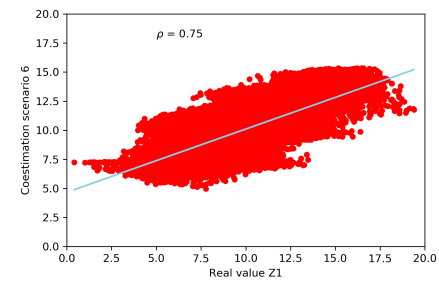
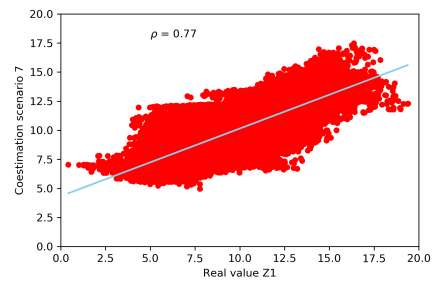
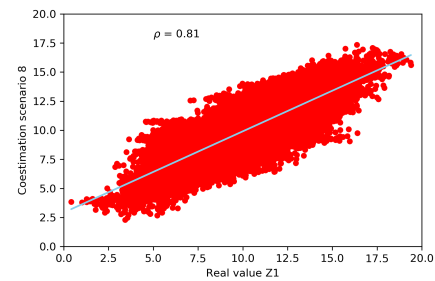
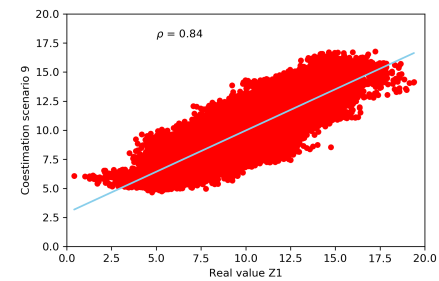
(a) Scenario 1, SD=625 m².(b) Scenario 2, SD=1000 m².(c) Scenario 3, SD=1500 m².(d) Scenario 4, SD=1750 m².(e) Scenario 5, SD=2000 m².(f) Scenario 6, SD=3000 m².(g) Scenario 7, SD=4000 m².(h) Scenario 8, SD=5000 m².(i) Scenario 9, SD=6000 m².

Figure 8. Cont.

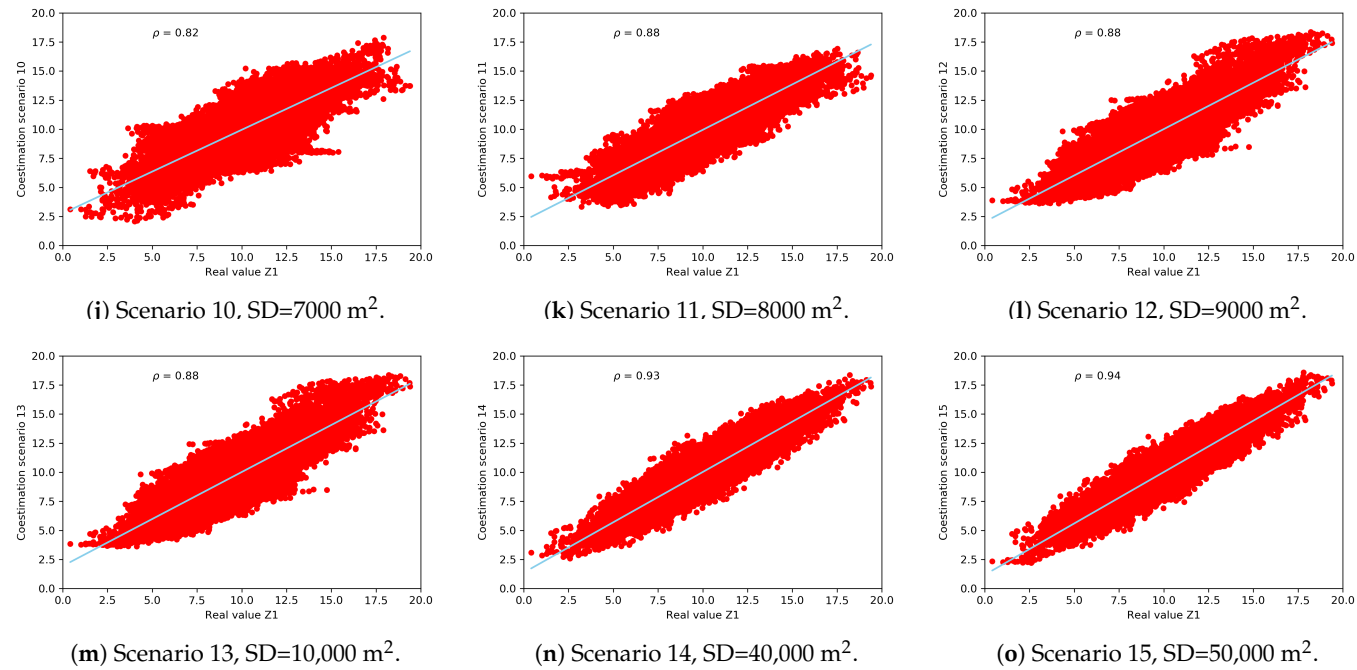


Figure 8. Dispersion diagrams between the real variable Z_1 and the 15 scenarios.

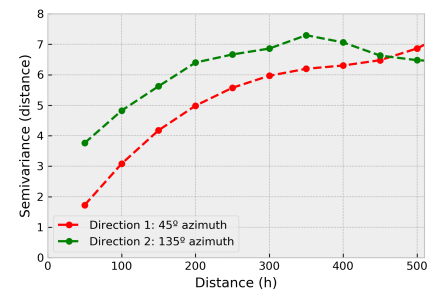
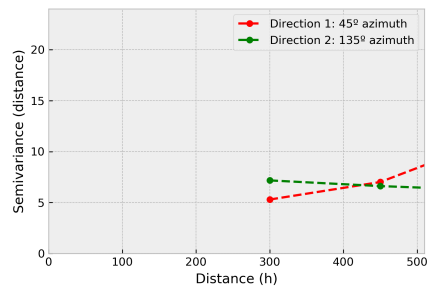
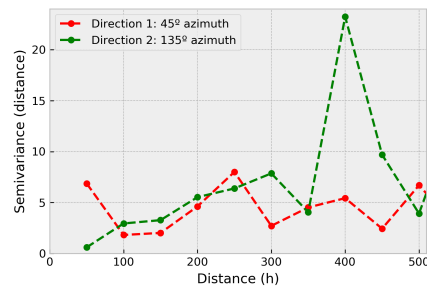
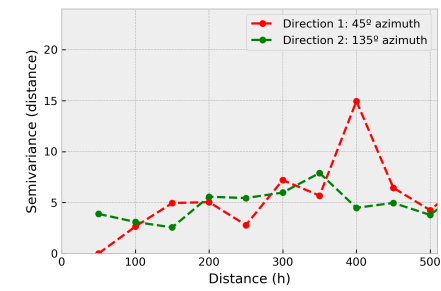
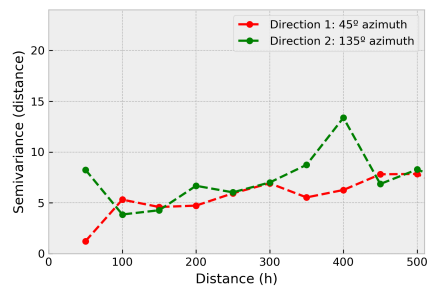
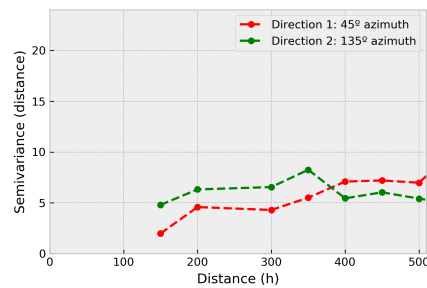
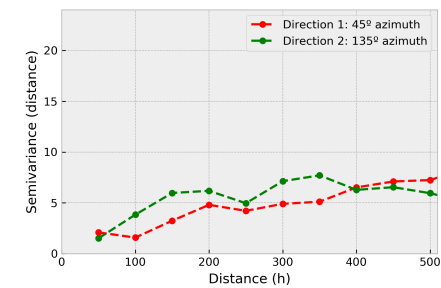
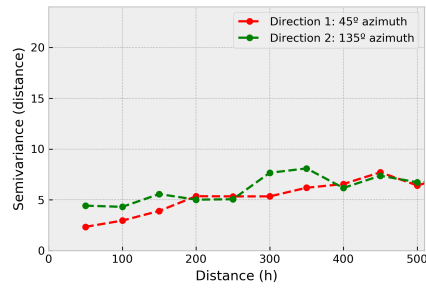
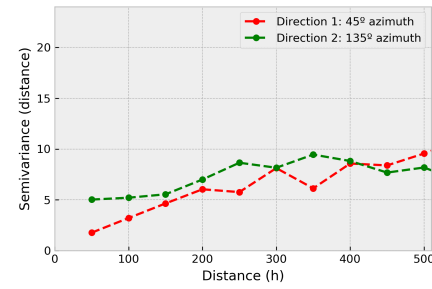
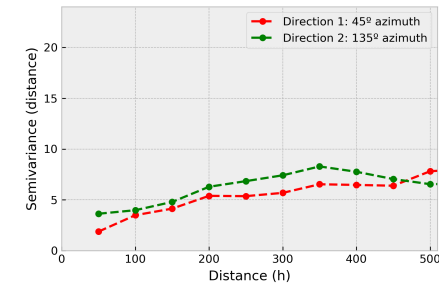
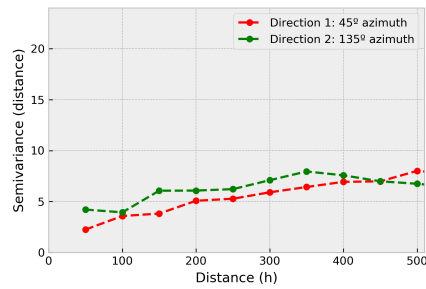
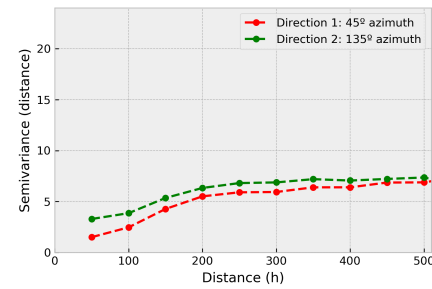
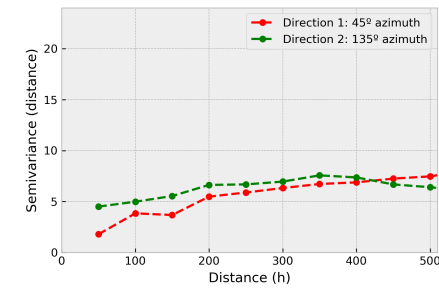
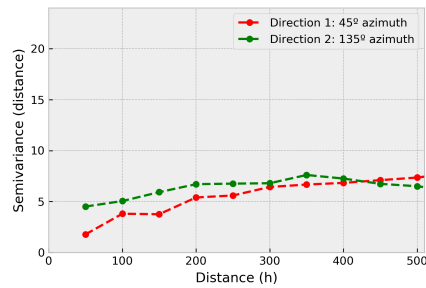
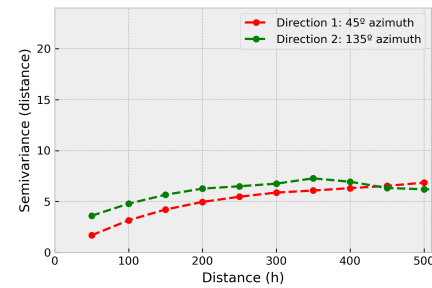
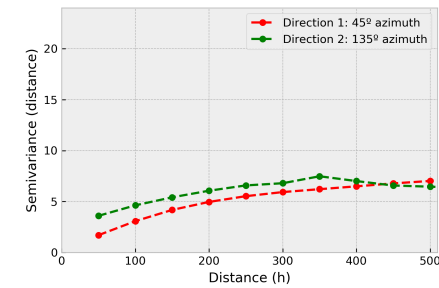
(a) Real semivariogram Z_1 .(b) Scenario 1, $SD = 625 \text{ m}^2$.(c) Scenario 2, $SD=1000 \text{ m}^2$.(d) Scenario 3, $SD=1500 \text{ m}^2$.(e) Scenario 4, $SD=1750 \text{ m}^2$.(f) Scenario 5, $SD=2000 \text{ m}^2$.(g) Scenario 6, $SD=3000 \text{ m}^2$.

Figure 9. Cont.

(h) Scenario 7, SD=4000 m².(i) Scenario 8, SD=5000 m².(j) Scenario 9, SD=6000 m².(k) Scenario 10, SD=7000 m².(l) Scenario 11, SD=8000 m².(m) Scenario 12, SD=9000 m².(n) Scenario 13, SD=10,000 m².(o) Scenario 14, SD=40,000 m².(p) Scenario 15, SD=50,000 m².Figure 9. Coestimation direct semivariograms versus simulated real Z_1 .

As expected at higher sampling densities, the mean squared error is smaller (MSE), see Figure 10. The relationship between the sampling density in percentage measure (SDP) and the mean global error (MGE) can be expressed using this potential function:

$$MGE = 1.2366 \cdot SDP^{-0.224}. \quad (9)$$

In Figure 11 the mean global error versus the equivalent sampling density is shown.

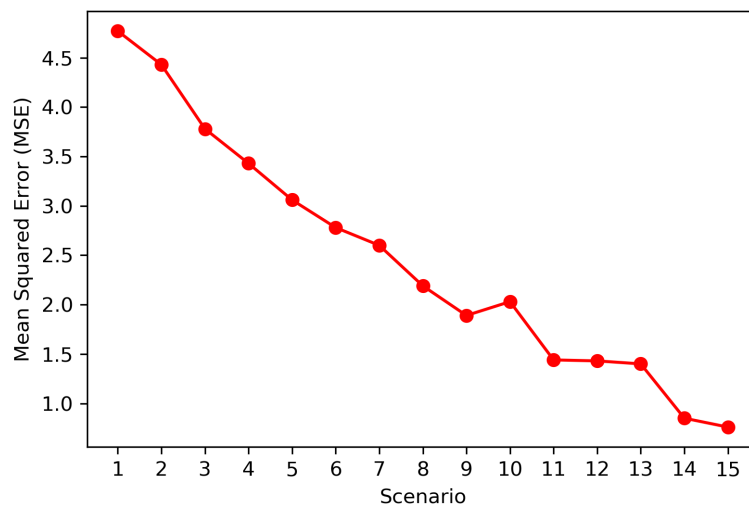


Figure 10. Mean Squared Error of the coestimation in each scenario.

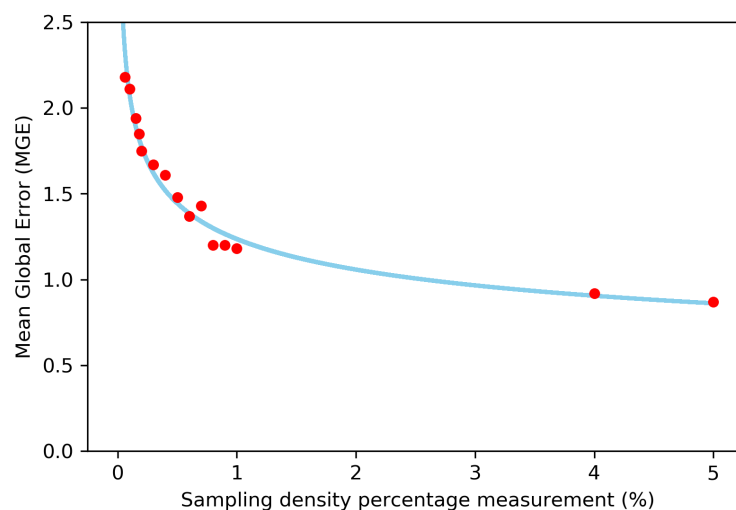


Figure 11. Mean global error versus the equivalent sampling density.

The deviation of the coestimation is directly proportional to the mean global error. The correlation coefficient is 0.99, which indicates that the application of the collocated ordinary cokriging method was consistent and reliable, see Figure 12. Applying linear regression to find the relationship between the mean global error and the coestimated deviation, the corrected coestimated deviation $\hat{\sigma}_{COCK}$ is obtained:

$$\hat{\sigma}_{COCK} = 0.8781 \cdot \sigma_{COCK} + 0.0833. \quad (10)$$

There is no optimal scenario, since the overall process depends on factors such as the sensitivity of the variable of interest, the amount of resources to invest to obtain the sample, the error tolerance,

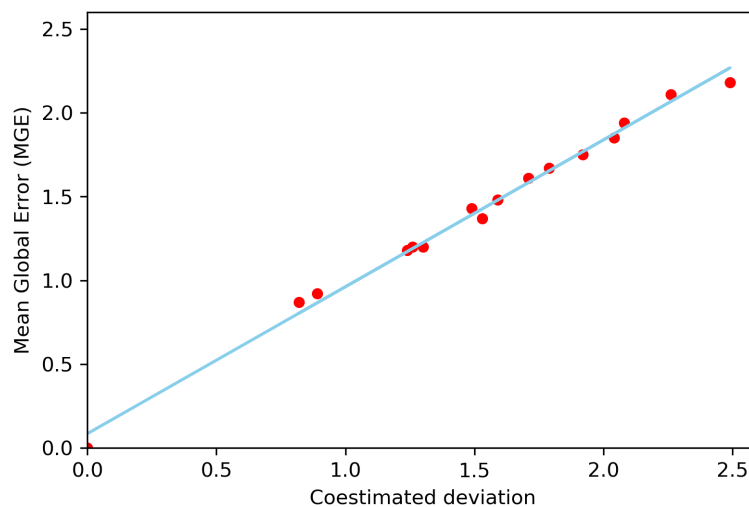


Figure 12. Relationship between the mean global error and the collocated ordinary cokriging deviation.

and so on. For example, in the 13th scenario with a sampling density equivalent to 1% of the target population, the estimated mean is 10.02 and the deviation 1.24: $\mu_{COCK} = 10.02 \pm 1.24$. This means that the estimated value could vary between 8.78 and 11.26. A similar value $\mu_{COCK} = 10.02 \pm 1.18$ is obtained after correcting the deviation with the proposed function (10). Different parameters of the 15 scenarios are presented in Table 6.

Table 6. Coestimation results.

Scenario	Sampling Density	SDP	MSE	MGE	μ_{COCK}	σ_{COCK}^2	σ_{COCK}	σ_{COCK}/μ_{COCK}	Inferior Limit	Superior Limit
1	25	0.06	4.77	2.18	10.16	6.20	2.49	0.25	7.67	12.65
2	40	0.10	4.43	2.11	10.16	5.11	2.26	0.22	7.90	12.42
3	60	0.15	3.78	1.94	9.86	4.32	2.08	0.21	7.79	11.94
4	72	0.18	3.43	1.85	9.87	4.16	2.04	0.21	7.83	11.91
5	80	0.20	3.06	1.75	10.11	3.67	1.92	0.19	8.19	12.03
6	120	0.30	2.78	1.67	10.10	3.20	1.79	0.18	8.31	11.89
7	160	0.40	2.60	1.61	10.15	2.93	1.71	0.17	8.43	11.86
8	200	0.50	2.19	1.48	9.91	2.54	1.59	0.16	8.31	11.50
9	240	0.60	1.89	1.37	9.98	2.34	1.53	0.15	8.45	11.51
10	280	0.70	2.03	1.43	9.95	2.23	1.49	0.15	8.46	11.45
11	320	0.80	1.44	1.20	9.95	1.68	1.30	0.13	8.65	11.24
12	360	0.90	1.43	1.20	10.01	1.58	1.26	0.13	8.75	11.27
13	400	1.00	1.40	1.18	10.02	1.53	1.24	0.12	8.78	11.26
14	1600	4.00	0.85	0.92	10.02	0.80	0.89	0.09	9.13	10.91
15	2000	5.00	0.76	0.87	10.01	0.68	0.82	0.08	9.19	10.84

6. Conclusions

Collocated ordinary cokriging is reliable and easier to implement than other multivariable geostatistical methods when an auxiliary variable is spatially correlated with the primary one, and when the auxiliary data is much more extensive than the primary data. In advance, the primary variable must be estimated in some domains checking delimitations where homogeneity by translation exists.

The sampling density is inversely proportional to the coestimation error; the larger the sampling density the lower the uncertainty of the estimated population. However, there is not a quantification of the sampling density that guarantees obtaining perfect results, since the coestimation error is subject

to an intrinsic error. It is crucial to manage the degree of sparsity of the primary data. This must be related to the sensitivity of the coestimation error in commercial terms. It is assumed that an error of 1 unit of measure will affect a variable with a low level of sparsity, which has a greater variability. For decision-making, the global mean deviation of the coestimation is a relevant parameter, since it determines the mean global error with a very high correlation coefficient. This parameter defines the worst ($\mu_{COCK} - \sigma_{COCK}$) and best ($\mu_{COCK} + \sigma_{COCK}$) of the cases, deriving in the choice of a sample size with an assumed and controlled risk.

Scenario 11 (320 observations) is presented as one of the best options balancing the number of samples versus results. It has a linear correlation coefficient of 0.88, a mean squared error of 1.44 and its linear correlation coefficient is similar to higher density sampling scenarios.

Author Contributions: H.H.G., conceptualization and investigation; E.A., writing, conceptualization and investigation; and A.G. writing and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the HAZITEK call of the Basque Government, project acronym HORDAGO.

Acknowledgments: We would like to thank the partners of the Deusto Digital Industry Chair (Etxe-Tar, General Electric, Idom, Accenture, Fundación Telefónica, Fundación BBK) the interest and support shown during this research.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

PERC	Pan European Reserves and Resources Reporting Committee
SGS	Sequential Gaussian Simulation
GSLIB	Geostatistical Software Library
MSE	Mean Squared Error
SGSIM	Sequential Gaussian Simulation Program
SD	Sampling Density
SDP	Sampling Density Percentage
MGE	Mean Global Error

References

1. Tulcanaza, E. *Evaluación de Recursos y Negocios Mineros: Incertidumbres, Riesgos y Modelos Numéricos*; Instituto de Ingenieros de Minas de Chile: Santiago de Chile, Chile, 1999.
2. Battalgazy, N.; Madani, N. Stochastic Modeling of Chemical Compounds in a Limestone Deposit by Unlocking the Complexity in Bivariate Relationships. *Minerals* **2019**, *9*, 683.
3. Kasmaee, S.; Raspa, G.; de Fouquet, C.; Tinti, F.; Bonduà, S.; Bruno, R. Geostatistical Estimation of Multi-Domain Deposits with Transitional Boundaries: A Sensitivity Study for the Sechahun Iron Mine. *Minerals* **2019**, *9*, 115.
4. Madani, N.; Yagiz, S.; Coffi Adoko, A. Spatial Mapping of the Rock Quality Designation Using Multi-Gaussian Kriging Method. *Minerals* **2018**, *8*, 530.
5. Nevskaya, M. A.; Seleznev, S. G.; Masloboev, V. A.; Klyuchnikova, E. M.; Makarov, D. V. Environmental and Business Challenges Presented by Mining and Mineral Processing Waste in the Russian Federation. *Minerals* **2019**, *9*, 445.
6. PERC Reporting Standard. *Pan-European Standard for Reporting of Exploration Results, Mineral Resources and Reserves*; The Pan-European Reserves and Resources Reporting Committee: Brussels, Belgium, 2017.
7. Matheron, G. The theory of regionalized variables and its applications. In *Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau*; École Nationale Supérieure des Mines de Paris: Paris, France, 1971.
8. Isaaks, E.H.; Srivastava, M.R. *Applied Geostatistics*; Oxford University Press: New York, NY, USA, 1989.
9. Journel, A.G.; Huijbregts, C.J. *Mining Geostatistics*; Academic Press: New York, NY, USA, 1978.
10. Sinclair, A.J.; Blackwell, G.H. *Applied Mineral Inventory Estimation*; Cambridge University Press: Cambridge, UK, 2002.

11. Matheron, G. Principles of geostatistics. *Econ. Geol.* **1963**, *58*, 1246–1266.
12. Oliver, M.A.; Webster, R. *Basic steps in Geostatistics: The Variogram and Kriging*; Springer: New York, NY, USA, 2015.
13. Pan, G.; Gaard, D.; Moss, K.; Heiner, T. A comparison between cokriging and ordinary kriging: Case study with a polymetallic deposit. *Math. Geol.* **1993**, *25*, 377–398.
14. Wackernagel, H. *Multivariate Geostatistics. An Introduction with Applications*; Springer: Berlin, Germany, 1995.
15. Almeida, A.S.; Journel, A.G. Joint simulation of multiple variables with a Markov-type coregionalization model. *Math. Geol.* **1994**, *26*, 565–588.
16. Chen, F.; Chen, Sh.; Peng, G. Using Sequential Gaussian Simulation to assess geochemical anomaly areas of lead element, computer and computing technologies in agriculture VI IFIP advances. *Inf. Commun. Technol.* **2013**, *393*, 69–76.
17. Soltani, F.; Afzal, P.; Asghari, O. Delineation of alteration zones based on sequential gaussian simulation and concentration-volume fractal modeling in the hypogene zone of Sungun copper deposit, NW Iran. *J. Geochem. Explor.* **2014**, *140*, 64–76.
18. Paravarzar, S.; Maarefvand, P.; Maghsoudi, A.; Afzal, P. Correlation between geological units and mineralized zones using fractal modeling in Zarshuran gold deposit (NW Iran). *Arab. J. Geosci.* **2015**, *8*, 3845–3854.
19. Brus, D.; De Gruijter, J. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). *Geoderma* **1997**, *80*, 1–44.
20. Brus, D.J.; Heuvelink, G. Optimization of sample patterns for universal kriging of environmental variables. *Geoderma* **2007**, *138*, 86–95.
21. Oliver, M.; Webster, R. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena* **2014**, *113*, 56–69.
22. Alperin, M. *Introducción al Análisis Estadístico de Datos Geológicos*; Editorial de la Universidad de La Plata: La Plata, Argentina, 2013.
23. Rossi, M.E.; Deutsch, C.V. *Mineral Resource Estimation*; Springer: Dordrecht, The Netherlands, 2014.
24. Hohn, M.E. *Geostatistics and Petroleum Geology*; Springer: Boston, MA, USA, 1988.
25. Pyrcz, M.J.; Deutsch, C.V. *Geostatistical Reservoir Modeling*; Oxford University Press: New York, NY, USA, 2014.
26. Goovaerts, P. *Geostatistics for Natural Resources Evaluation*; Oxford University Press: Nueva York, NY, USA, 1997.
27. Rocha, M.M.; Yamamoto, J.K.; Watabane, J.; Fonseca, P.P. Studying the influence of a secondary variable in collocated cokriging estimates. *An. Acad. Bras. Cienc.* **2012**, *84*, 335–346.
28. Xu, W.; Tran, T.T.; Srivastava, R.M.; Journel, A.G. Integrating seismic data in reservoir modeling: The collocated cokriging alternative. In Proceedings of the 67th Annual Technical Conference and Exhibition of the Society of Petroleum Engineers, Washington, DC, USA, 4–7 October 1992; Society of Petroleum Engineers Inc.: London, UK, 1992; pp. 833–842.
29. Journel, A.G. Markov models for cross-covariances. *Math. Geol.* **1999**, *31*, 955–964.
30. Remy, N.; Boucher, A.; Jianbing, W. *Applied Geostatistics with SGeMS; A User's Guide*; Cambridge University Press: New York, NY, USA, 2009.
31. Armstrong, M. *Basic Linear Geostatistics*; Springer: Berlin, Germany, 1998.
32. Ortiz, J.M.; Emery, X. Geostatistical estimation of mineral resources with soft geological boundaries: A comparative study. *J. S. Afr. Inst. Min. Metall.* **2006**, *106*, 577–584.
33. Rivoirard, J. Two key parameters when choosing the kriging neighborhood. *Math. Geol.* **1987**, *19*, 851–856.

