International Journal of Computer Science and Applications, 2008, Vol. 5, No. 3a, pp 67 - 79

METHODOLOGY WHICH APPLIES GEOSTATISTICS TECHNIQUES TO THE TOPOGRAPHICAL SURVEY

R. ORTIZ-MARIN, G. DEL RIO-CIDONCHA, J. MARTINEZ-PALACIOS,

AND L. GARCIA-RUESGAS

Department of Engineering Graphic, University of Seville
Avda. de los Descubrimientos s/n, 41092, Seville (Spain)
e-mail: rafortiz@esi.us.es

Abstract

Until today, the topographical surveys in a specific point had a problem: if a point was measured several times, it would be impossible to know exactly which measure would be the nearest to the real value. The methodology suggested in this article allows to reject those points which are further from the real value and to keep the points that are nearer to this value.

Keywords
Geostatistic, matriz of cofactors, regionalized variables, spatial autocorrelation, semivariogram

1. Introduction

Geostatistics Techniques will be used to distinguish among the points which are nearer or further to the real value. These techniques are based on the theory of regionalized variable. This theory reports us that in nature there are values which have a spatial change and it is possible to quantify this change. In the following epigraphs, a quick view of the Geostatistics and the theory of regionalized variable will be described.

2. Variability

Most of the natural phenomena which are studied are variables in the space and time. If the topographical surface or water pollution is considered, it will be possible to observe the high variability in small distances. The conditions under which it occurs are not known completely, it's not possible to describe them totally with physical and chemical laws.

3. Geostatistics

The analysis of time series is one of the first fields where the variability has been considered and has been described with stochastical methods. These methods have been extended to analyse spatial variability. The spatial methods are part of the discipline called Geostatistics.

It is used with two different meanings:

1. As a collection of all statistical and probabilistic methods applied to sciences of Earth.

2. As another name for the theory of regionalized variable.

In the next pages, it will be talked about the theory of regionalized variables which appeared at the beginning of the fifties. During this period, D. Krige and his colleagues started to apply sta tistics techniques to estimate mineral reserve.

In the sixties, G. Matheron [1] (a French mathematician) gave theoretical fundamentals to the geostatistics methods. Mining used the geostatistics in first place because the high costs of bore holes made data analysis extremely important. Books and publications of Geostatistics are targeted fundamentally at problems of mine.

4. Regionalized variables

The theory of regionalized variable is a geostatistic method which is used to interpolate in space. According to this theory, the interpolation of points in space shouldn't be based on a continuous object. It must be based on a stochastical model which must consider trends in the original system of points. The theory considers three types of connections within any data group:

1. A structural part, which is also called the trend.
2. Correlated change.
3. Change without correlation or noise.

In the theory of regionalized variables, the concept of random functions plays a crucial role [2]. A random function is a system of random variables which belong to the points of *D* domain under study. This means that for each u point in *D* there's a random variable *Z (u)*.

A regionalized variable is the achievement of a random function. This means that for each u point in the dimensional space of *D, z (u)* is an achievement of random function *Z (u)*. This interpretation of natural phenomena recognizes the fact that it's not possible to describe them completely using only deterministic methods [3].

It is possible to describe a random function using multidimensional distribution functions. This means that for each system of points $u_1 ... u_n$ in *D* domain, it is assigned an accumulative distribution function $F\ u_1 ... u_n$. It is possible to find *P* probability using these functions for each system of possible values $w_1 ... w_n$:

$$P(Z(u_1) < w_1 \cdots Z(u_2) < w_n) = F_{u_1 \cdots u_n}(w_1 \cdots w_n)$$

This will mean that the conditional probabilities may be used to value the local or global averages etc.

Unfortunately, there are many finite subsets in *D* domain, and generally for each point in *D* there's only one value which is available for obtaining the distribution functions based on the data. Even if the parameters were measured in several occasions (e.g. groundwater quality) there wouldn't be enough measures to determine the distribution functions.

A general hypothesis that reduces the complexity of the problem is the supposed strong immobility:

The *Z (u)* random function is stationary if it is carried out the next equation for each system of points in *D* domain $u_1 ... u_n$ en el dominio *D*, and for each system of possible values $w_1 ... w_n$, and for each *h* vector:

$$P(Z(u_1) < w_1, \cdots, Z(u_n) < w_n) = P(Z(u_1 + h) < w_1, \cdots, Z(u_n + h) < w_n)$$

This equation means that the distribution of the random function depends on the configuration of points, but it doesn't depend on their locations.

5. The variogram

The $2g(x, y)$ theoretical variogram is a function which describes the degree of spatial dependence of a spatial random field or a *Z (x)* stochastical process. It is defined as the hope of increasing square of the values between *x y* locations [4]:

$$2g(x, y) = E((Z(x) - Z(y))^2$$

A $g(x, y)$ it is called semivariogram.

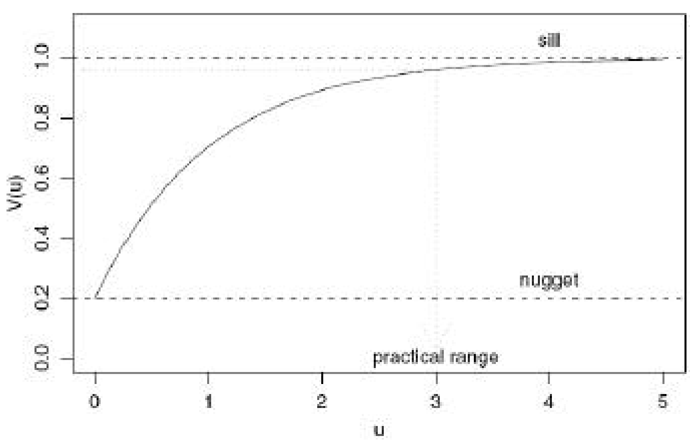

Figure 1. Semivariogram

5.1 The experimental semivariogram

For $z_i, i = 1, \ldots, N$ observations which are situated in $x_1, \ldots, x_N$ locations, it is defined the empirical semivariogram as [5]:

$$\hat{g} = \frac{1}{2|N(h)|} \sum_{(i,j) \in N(h)} |z_i - z_j|^2$$

Where *N* (*h)* is the system of pair observations i, j placed at a distance of h.

Normally, the semivariogram is an increasing monotonous function which can reach a limit value, known as Plateau, equivalent to the sampling variance [6]. The plateau is reached for a known value of h as range or scope. The range determines the influence zone around a point, beyond this point the autocorrelation is nil. However, all variograms don't reach a plateau. It is possible that a variogram doesn't reach the variance asymptotically and it reaches infinity [7].

The variogram represents the average rate of change for a property with the distance. If two observations are situated next one to another, they will be more similar than if they are very separate. This is the concept of the variogram. The spatial dependence is reduced when the h distance is increased and it finishes at a certain distance, known as range. Beyond the range, the average rate of change is independent of the separation between observations.

The empirical variogram is used in Geostatistics as an initial estimation of the variogram (theoretical) which is needed for spatial interpolation by cokriging.

6. Cokriging

It is common to have the measure of several variables. These measures can belong to points of the sample or to different points.

It will be considered the case in which one of the variables is identified as priority (Z main variable) without losing generality and the others are considered as secondary variables. To simplify, it will be supposed that there's only one secondary variable (Y).

How can we use the information provided by the secondary variable to improve the estimation of main variable?

This can be done through the estimation called ordinary cokriging.

6.1 Ordinary Cokriging

It is possible to establish a linear estimation of the main variable from the observations of main and secondary variable [8]:

$$Z_0^* = \sum_{i=1}^{nz} l_i Z_i + \sum_{i=1}^{ny} a_i Y_i$$

The estimator must be 'insesgado' [9] this can be insured by imposing:

$$\sum_{i=1}^{nz} l_i = 1$$

$$\sum_{i=1}^{ny} a_i = 0$$

The variance of estimation is written:

$$Var(Z_0 - Z_0^*) = Var(Z_0) + \sum_{i=1}^{nz}\sum_{j=1}^{ny} l_i l_j Cov(Z_i, Z_j) + \sum_{i=1}^{nz}\sum_{j=1}^{ny} a_i a_j Cov(Y_i, Y_j) +$$
$$2\sum_{i=1}^{nz}\sum_{j=1}^{ny} l_i a_j Cov(Z_i, Y_j) - 2\sum_{i=1}^{nz} l_i Cov(Z_0, Z_i) - 2\sum_{i=1}^{ny} a_i Cov(Z_0, Y_i)$$

Lagrangiana is established and it is derived in relation to unknown weights and two Lagrange multipliers which have been introduced to take into account the restrictions of insesgado estimator. It is obtained the system of ordinary cokriging which can be expressed in a matrix way like that:

$$K\,l = k$$

$$K = \begin{pmatrix} Cov(Z_1,Z_1) & Cov(Z_1,Z_2) & \cdots & Cov(Z_1,Z_{nz}) & Cov(Z_1,Y_1) & Cov(Z_1,Y_2) & \cdots & Cov(Z_1,Y_{ny}) & 1 & 0 \\ Cov(Z_2,Z_1) & Cov(Z_2,Z_2) & \cdots & Cov(Z_2,Z_{nz}) & Cov(Z_2,Y_1) & Cov(Z_2,Y_2) & \cdots & Cov(Z_2,Y_{ny}) & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ Cov(Z_{nz},Z_1) & Cov(Z_{nz},Z_2) & \cdots & Cov(Z_{nz},Z_{nz}) & Cov(Z_{nz},Y_1) & Cov(Z_{nz},Y_2) & \cdots & Cov(Z_{nz},Y_{ny}) & 1 & 0 \\ Cov(Y_1,Z_1) & Cov(Y_1,Z_2) & \cdots & Cov(Y_1,Z_{nz}) & Cov(Y_1,Y_1) & Cov(Y_1,Y_2) & \cdots & Cov(Y_1,Y_{ny}) & 0 & 1 \\ Cov(Y_2,Z_1) & Cov(Y_2,Z_2) & \cdots & Cov(Y_2,Z_{nz}) & Cov(Y_2,Y_1) & Cov(Y_2,Y_2) & \cdots & Cov(Y_2,Y_{ny}) & 0 & 1 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ Cov(Y_{ny},Z_1) & Cov(Y_{ny},Z_2) & \cdots & Cov(Y_{ny},Z_{nz}) & Cov(Y_{ny},Y_1) & Cov(Y_{ny},Y_2) & \cdots & Cov(Y_{ny},Y_{ny}) & 0 & 1 \\ 1 & 1 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 & 0 & 0 \end{pmatrix}$$

$$l = \begin{pmatrix} l_1 \\ l_2 \\ \vdots \\ l_{nz} \\ a_1 \\ a_2 \\ \vdots \\ a_{ny} \\ m_z \\ m_y \end{pmatrix} \qquad k = \begin{pmatrix} Cov(Z_0,Z_1) \\ Cov(Z_0,Z_2) \\ \vdots \\ Cov(Z_0,Z_{nz}) \\ Cov(Z_0,Y_1) \\ Cov(Z_0,Y_2) \\ \vdots \\ Cov(Z_0,Y_{ny}) \\ 1 \\ 0 \end{pmatrix}$$

K matrix which have $(nz+ny+2)\cdot(nz+ny+2)$ dimension contains all the covariances, the k vector $(nz+ny+2)$ contains all the covariances between the point we want to estimate and the points of data for the two variables and the l matrix contains the unknown quantities (weights and Lagrange multipliers).

7. Theoretical approach and problem analysis

From the data obtained in measurements, the emq[1], the q_{xxi} [2] and q_{yyi} [3] are defined regionalized variables (main and secondary variable).

Z main variable is defined as:

$$Z_i(x_i) = emq^2 \cdot q_{xxi} \qquad i = 1 \ldots nz$$

$$Y_i(x_i) = emq^2 \cdot (q_{xxi} + q_{yyi}) \qquad i = 1 \ldots ny$$

8. Methodology

The diagram of figure 2 shows the steps to get the value of the coordinate nearest to real value (X coordinate in this case).

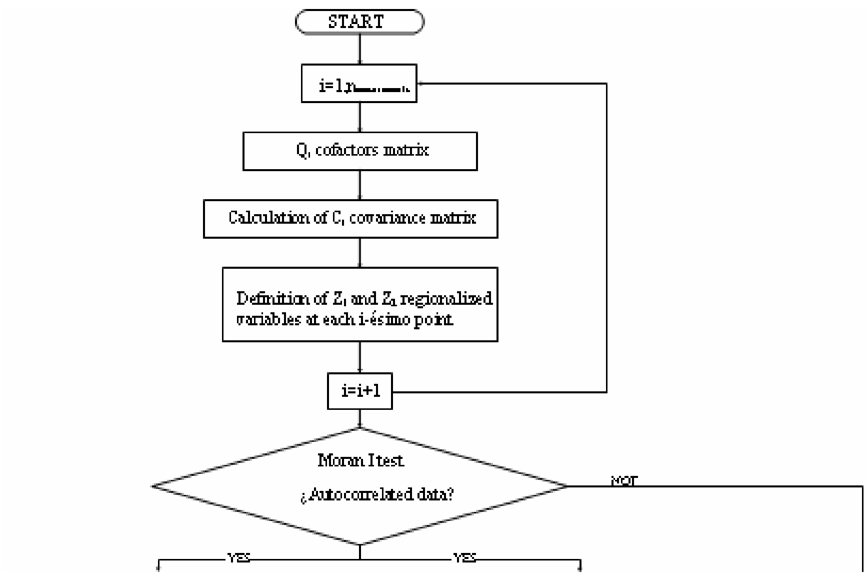

Figure 2a. Steps to get the value of the coordinate (continued into figure 2b)

[1] Average quadratic error

[2] Cofactor matrix element which corresponds to X coordinate

[3] [3] Cofactor matrix element which corresponds to Y coordinate

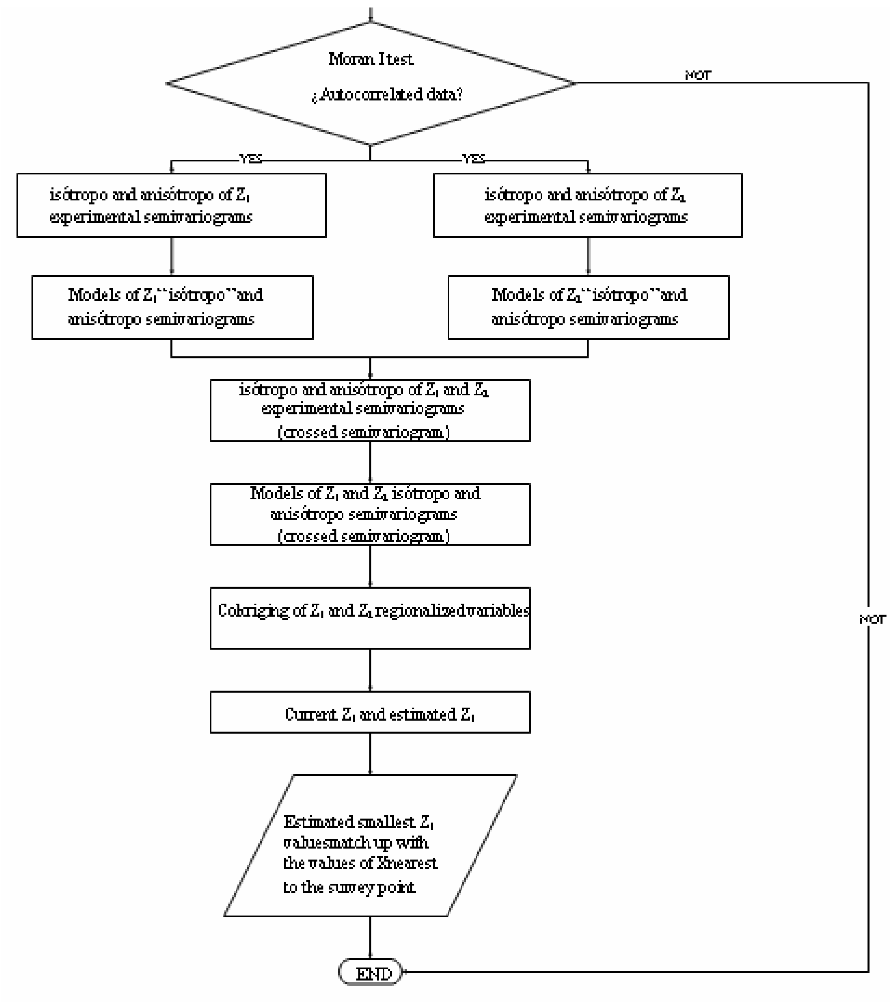

Figure 2b. Steps to get the value of the coordinate

8.1 Implementation of methodology

The starting data are shown in table 1

X(m)	Y(m)	*z*	*y*	X(m)	Y(m)	*z*	*y*
243,886.22	4,164,265.55	0.020	0.029	243,886.31	4,164,265.09	0.024	0.035
243,886.19	4,164,265.41	0.045	0.065	243,886.18	4,164,265.04	0.017	0.025
243,886.11	4,164,265.28	0.067	0.097	243,886.15	4,164,265.05	0.011	0.016
243,886.07	4,164,265.03	0.0467	0.067	243,886.28	4,164,264.97	0.016	0.024
243,886.20	4,164,264.81	0.024	0.035	243,886.26	4,164,265.01	0.013	0.019
243,886.23	4,164,265.01	0.030	0.043	243,886.30	4,164,264.95	0.020	0.029
243,886.03	4,164,265.09	0.022	0.033	243,886.49	4,164,265.02	0.014	0.020
243,886.17	4,164,264.95	0.022	0.032	243,886.71	4,164,265.06	0.015	0.023
243,886.14	4,164,265.07	0.014	0.021	243,886.80	4,164,265.21	0.010	0.015
243,886.15	4,164,265.12	0.021	0.031	243,887.10	4,164,265.29	0.008	0.012
243,886.31	4,164,265.25	0.017	0.024	243,887.27	4,164,265.16	0.013	0.020
243,886.29	4,164,265.26	0.016	0.023	243,887.26	4,164,265.36	0.013	0.020
243,886.38	4,164,265.15	0.010	0.015	243,887.19	4,164,265.40	0.009	0.013
243,886.63	4,164,265.28	0.008	0.012	243,887.10	4,164,265.37	0.011	0.016
243,886.75	4,164,265.34	0.008	0.011	243,887.15	4,164,265.40	0.010	0.014
243,886.55	4,164,265.32	0.009	0.014	243,887.19	4,164,265.26	0.012	0.018
243,886.68	4,164,265.50	0.007	0.010	243,887.03	4,164,265.23	0.008	0.011
243,886.54	4,164,265.61	0.013	0.019	243,887.18	4,164,265.11	0.012	0.018
243,886.30	4,164,265.60	0.011	0.016	243,886.91	4,164,265.07	0.010	0.015
243,886.22	4,164,265.58	0.015	0.022	243,886.85	4,164,264.93	0.013	0.019
243,886.24	4,164,265.46	0.016	0.023	243,886.84	4,164,265.13	0.018	0.026
243,886.29	4,164,265.41	0.012	0.017	243,886.75	4,164,265.16	0.024	0.036
243,886.42	4,164,265.34	0.010	0.015	243,886.75	4,164,265.16	0.032	0.048
243,886.56	4,164,265.13	0.015	0.023	243,886.39	4,164,265.10	0.033	0.048
243,886.42	4,164,265.09	0.019	0.028	243,886.56	4,164,265.23	0.068	0.100

Table 1. Pelagatos, values of main and secondary (Z and Y) regionalized variables

The directional semivariograms are shown in figures 3, 4 and 5.

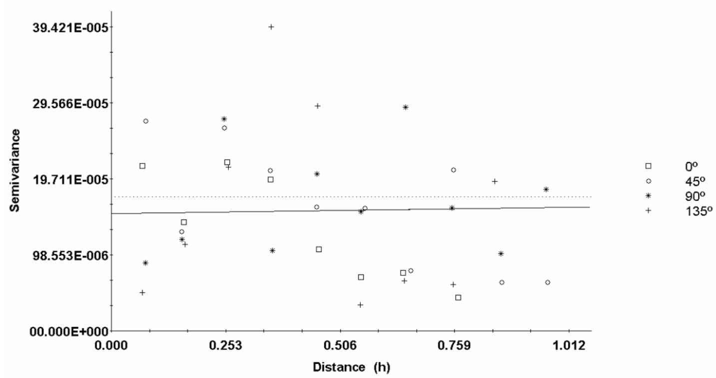

Figure 3. Anisótropo semivariogram of Z variable

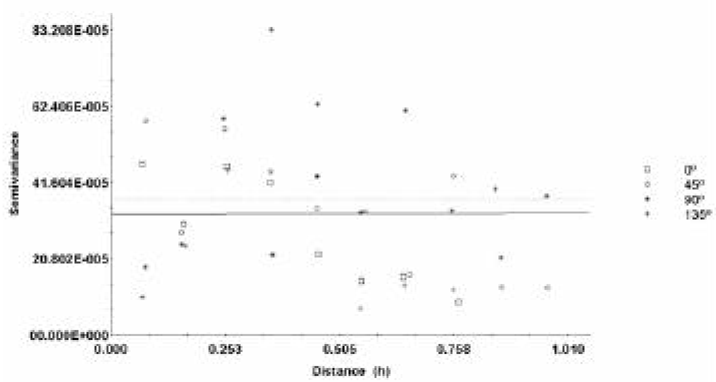

Figure 4. Anisótropo semivariogram of Y variable

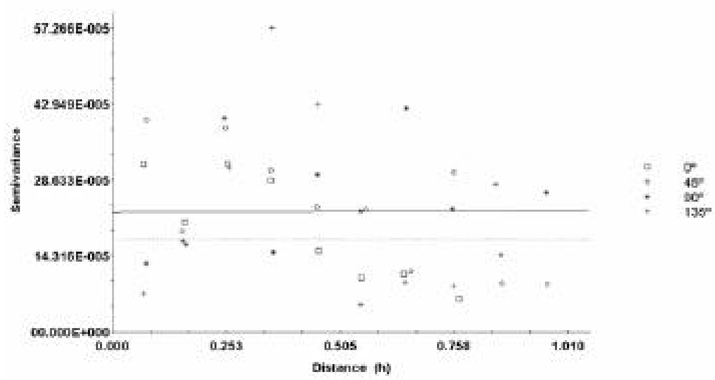

Figure 5. Crossed anisótropo semivariogram of Z and Y variables

Once the cokriging has been done in each (X, Y) sample point, we can have a current Z variable (which has been defined initially) and a Z estimated variable, see Table 2.

Measurement	Z current	Z estimated	Measurement	Z current	Z estimated
1	0.020	0.0183	26	0.024	0.018
2	0.045	0.017	27	0.017	0.018
3	0.067	0.017	28	0.011	0.018
4	0.046	0.017	29	0.016	0.018
5	0.024	0.018	30	0.013	0.018
6	0.029	0.018	31	0.020	0.018
7	0.022	0.017	32	0.014	0.018
8	0.022	0.018	33	0.015	0.019
9	0.014	0.018	34	0.010	0.020
10	0.021	0.017	35	0.008	0.020
11	0.017	0.018	36	0.013	0.020
12	0.016	0.018	37	0.013	0.020
13	0.010	0.018	38	0.009	0.020
14	0.008	0.019	39	0.011	0.020
15	0.008	0.019	40	0.010	0.020
16	0.009	0.019	41	0.012	0.020
17	0.007	0.019	42	0.007	0.020
18	0.013	0.019	43	0.012	0.020
19	0.011	0.018	44	0.010	0.020
20	0.015	0.018	45	0.013	0.019
21	0.016	0.018	46	0.018	0.019
22	0.012	0.018	47	0.024	0.019
23	0.010	0.018	48	0.032	0.019
24	0.015	0.019	49	0.033	0.018
25	0.019	0.018	50	0.068	0.018

Table 2. Values of current Z variable and estimated Z variable

Table 2 can be depicted by lines where positions of sample points and the estimated Z value are shown. Figure 6**Error! Reference source not found.**.

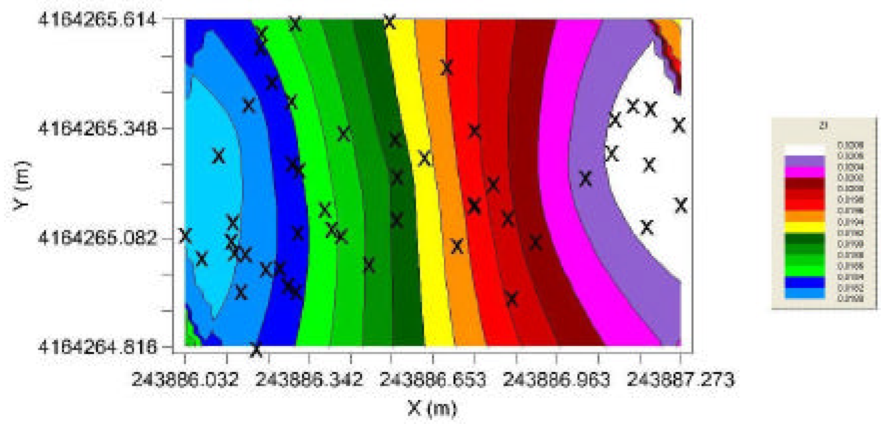

Figure 6. Map with the Z variable distribution

If we replace measurement column by (X, Y) coordinate in Table 2 (with the only presence of estimated Z value), we will obtain Table 3.

X (m)	Y (m)	Z estimated	X (m)	Y (m)	Z estimated
243886.220	4164265.550	0.018	243886.313	4164265.099	0.018
243886.191	4164265.410	0.017	243886.183	4164265.047	0.018
243886.115	4164265.287	0.017	243886.154	4164265.050	0.018
243886.073	4164265.037	0.017	243886.288	4164264.970	0.018
243886.209	4164264.816	0.018	243886.269	4164265.013	0.018
243886.234	4164265.010	0.018	243886.309	4164264.955	0.018
243886.032	4164265.093	0.017	243886.491	4164265.021	0.018
243886.171	4164264.955	0.018	243886.713	4164265.068	0.019
243886.147	4164265.079	0.018	243886.804	4164265.216	0.020
243886.151	4164265.124	0.017	243887.100	4164265.292	0.020
243886.317	4164265.251	0.018	243887.273	4164265.166	0.020
243886.299	4164265.266	0.018	243887.268	4164265.362	0.020
243886.381	4164265.155	0.018	243887.197	4164265.401	0.020
243886.630	4164265.281	0.019	243887.108	4164265.375	0.020
243886.757	4164265.347	0.019	243887.153	4164265.408	0.020
243886.55	4164265.32	0.019	243887.19	4164265.26	0.020

243886.68	4164265.50	0.019	243887.03	4164265.23	0.020
243886.54	4164265.61	0.019	243887.18	4164265.11	0.020
243886.30	4164265.60	0.018	243886.91	4164265.07	0.020
243886.22	4164265.58	0.018	243886.85	4164264.93	0.019
243886.24	4164265.46	0.018	243886.84	4164265.13	0.019
243886.29	4164265.41	0.018	243886.75	4164265.16	0.019
243886.42	4164265.34	0.018	243886.75	4164265.16	0.019
243886.56	4164265.13	0.019	243886.39	4164265.10	0.018
243886.42	4164265.09	0.018	243886.56	4164265.23	0.018

Table 3. (X,Y) coordinates and estimated values of Z variable

In Figure 6 and Table 3, we can see that smallest values of estimated Z are those which are smaller or equal to 0.018. If we extract these values of Table 3, we will obtain Table 4.

X (m)	Y (m)	$z_1$estimated
243,886.19	4,164,265.41	0.017
243,886.11	4,164,265.28	0.017
243,886.07	4,164,265.03	0.017
243,886.23	4,164265.01	0.018
243,886.03	4,164,265.09	0.017
243,886.17	4,164,264.95	0.018

Table 4. Minimum values of the estimated Z variable

The X value in bold colour corresponds with the nearest point to X coordinate of survey point. Pelagatos, this value and the rest of values have been obtained applying the criteria of minimum value of estimated Z variable.

The average of X values is compared to X value of Pelagatos survey point and average X of all observations in Table 5

$X_{\text{average of all observations}}$ (m)	X_{average} (m)	$X_{\text{survey point}}$ (m)
243,886.54	243,886.15	243,886.00

Table 1. Average of X and X value of Pelagatos survey point

As we can see in Table 5, the calculus of the "best X" using the method of regionalized variable is more exact than the average of all observed X coordinates.

9. Conclusions

Measurements in survey points which belong to the new REGENT (National Geodetic Network for Space Techniques) it has been checked that the values of the coordinates in these survey points are very accurate.

It is possible to see in tables that the methodology which has been used allows us to extract a subgroup of data from a group of data which have been obtained from the repeated measurements of a point.

This methodology has been applied in several survey points of the REGENT communications network and it allows us to obtain a group points which are the nearest to the real value of the survey point.

References

[1] Matheron, G. (1971): *The Theory of Regionalized Variables and its Applications*. Les Cahiers du Centre de Morphologie Mathématique, Fasc. 5

[2] Matheron, G. (1982): *Pour une Analyse Krigeante des Données Regionalisées*. Les Cahiers du CGMM, N. 732, Fontaineblau, 21 pp.

[3] Journel, A. G. (1983): *Non parametric estimation of spatial distributions*. Mathematical Geology, 15, 445-468

[4] Wackernagel, H. (1988): Geostatistical Techniques for Interpreting Multivariate Spatial Information. *Quantitative Analysis of Mineral and Energy Resources*, NATO ASI Series, C 223, 393-409, Reidel, Dordrecht, pp. 393-409

[5] Cressie, N. (1993). Fitting variogram models by weighted least squares. Mathematical Geology 17:563-586

[6] Journel, A. G. and Zhu, H. (1990): Integrating Soft Seismic Data: MarkovBayes updating, an alternative to cokriging and traditional regression. In Stanford Center for Reservoir Forecasting, Report 3, Stanford, 1- 62

[7] Journel, A., Huijbregts, Ch. J. (1978): *Mining Geostatistics*. Academic Press, London, 600 pp

[8] Wackernagel, H., Petitgas, P., Touffait, Y. (1989): Overview of methods for coregionalization analysis. M. Armstrong (ed.), *Geostatistics*, Kluwer Academic Publishers, Dordrecht, pp. 409-420

[9] Sousa, A. J. (1988): Factorial Kriging as a Method to include Spatial Structure into Classification. A Case Study on a Sulphide Orebody. *Quantitative Analysis of Mineral and Energy Resources*, NATO ASI, Series C 223, 385-392, Reidel, Dordrecht.