Minería y Geología / v.41 n.2/ abril-junio / 2025/ p. 172-187 ISSN 1993 8012

ORIGINAL ARTICLE

Modeling mining deposit properties using multivariate multiple (A,U,Θ) estimators

Modelación de propiedades de yacimientos minerales usando estimadores multivariables múltiples (A,U,Θ)

Arístides Alejandro Legrá-Lobaina[1*], Ramón Eddie Peña-Abreu[2], Eduardo Terrero-Matos[1], Ramón Gilberto Polanco-Almanza[1], Justino Tomás-António[3]

[1]University of Moa, Holguín, Cuba.

[2]Mining-Metallurgical Industry Research Center, Havana, Cuba.

[3]Lunhinga Mining Society, Lucapa, Lunda Norte, Angola.

[*]Corresponding author: alegra@ismm.edu.cu

Abstract

This paper describes a procedure to estimate at a point of coordinates P_e, the value Z_{7e} of a dependent variable Z_7 that quantifies the spatial behavior of a property of a mineral deposit (for example, the volumetric mass), simultaneously using information from two databases named respectively BD_1 and BD_2. The procedure is based on the use of a Multiple Multivariate Estimator that, applied exhaustively to the combinations that are established for the possible values of the parameters (p, s and m) of a UPD kernel function, allows obtaining optimal results with respect to a coefficient of variation CV_e which evaluates the relationship between the estimation error and the estimated value. The form in which the multiple multivariate estimator presented has been defined allows it to be classified as viable and effective.

Keywords: mathematical modeling, mining deposits, estimator (A,U,Θ), multivariate estimator, effective estimator, feasible estimator, multiple estimator, estimation error

Resumen

Este artículo describe un procedimiento para estimar en un punto de coordenadas P_e, el valor Z_{7e} de una variable dependiente Z_7 que cuantifica el comportamiento espacial de una propiedad de un yacimiento mineral (por ejemplo; la masa volumétrica), utilizando simultáneamente información de dos bases de datos denominadas respectivamente BD_1 y BD_2. El procedimiento se basa en el uso de un Estimador Multivariado Múltiple que aplicado exhaustivamente a las combinaciones que sean establecidas para los posibles valores de los parámetros (p, s y m) de una función núcleo UPD, permite obtener resultados óptimos respecto a un coeficiente de variación CV_e que evalúa la relación entre el error de estimación y el valor estimado. La forma en que ha sido definido el estimador multivariado múltiple que se presenta, permite calificarlo como viable y eficaz.

Palabras clave: modelación matemática, yacimientos mineros, estimador (A,U,Θ), estimador multivariable, estimador eficaz, estimador factible, estimador múltiple, error de estimación

1. INTRODUCTION

During the study of mineral deposits, geo-referenced sample studies of their properties are conducted to obtain data. With the aid of mathematical point estimators, these data lead to models that explain and predict the behavior of the properties under investigation (Ding et al., 2018; Legrá-Lobaina and Terrero-Matos, 2019; Tomás, 2020). Without loss of generality, the following scenario can be presented:

1.1. Available Data, General Technical Problem, and Specific Mathematical Problem

Assume that, based on preliminary sampling, two databases are available:

a. The first (denoted DB1) refers to a sampling study of six properties of a deposit at *k* spatial geo-locations (given by their coordinates). At each point, numerical values are determined for some properties, while others take classificatory values. This spatially dense study generates *k* elements in DB1 and is relatively inexpensive. For each of these elements, the values of the following fields are available:

- Specific geological references for each sample.
- Planar coordinates: X (east-west); Y (south-north); Z (elevation or sample height above sea level). The triple P=(X; Y; Z) is termed the spatial coordinates or geographical location of each sampling point.
- Data for the six variables under study: Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, measured or determined at each sampling point P. In the present work, it is assumed that the values of variables Z_1, Z_2, Z_3 quantify laboratory results and measurements; whereas variables Z_4, Z_5, Z_6 report on sample classifications or typologies.

b. The second database (DB2) refers to another study of 10 variables where sampling for the new variables is more complex and expensive. Consequently, it is spatially less dense, with *n* < *k* elements, each with its corresponding field values:

- Specific geological references for each sample.
- Spatial coordinates: P=(X; Y; Z) for each sample.
- Data measured or determined in each sample: Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, Z_7, Z_8, Z_9, Z_{10}, where the values of variables Z_7, Z_8, Z_9, and Z_{10} quantify other measurement results and laboratory tests.

The general technical problem arises: to model the behavior of the property described by variable Z_7 within the mineral deposit using a grid (Legrá-Lobaina, 2017) by employing point estimators and the available data.

The specific mathematical problem derived is to find a mathematical point estimator that, considering the data from DB1 and DB2, allows for the estimation of the values of variable Z_7 at all points in DB1, as well as at other points within the space occupied by the deposit.

1.2. Effective Estimator for Z_7

In this context, an estimator for Z_7 will be termed *effective* if it:

1. Uses the information available in the input data while guaranteeing computational efficiency of the calculations and meeting the requirements for the ranges of the estimated values.
2. Ensures adequate explanatory and predictive capabilities, assessed and monitored through cross-validation techniques applied to the input data from DB1 and DB2.
3. Determines which input data are admissible for the estimator.
4. Ensures the conditions for the estimator to be capable of obtaining the value Z_{7_e} for each location P_e in DB1.
5. Calculates the error bounds for each point estimate Z_{7_e} performed for location P_e in DB1.
6. Minimizes the values of the error bounds for Z_{7_e} performed for P_e in DB1.

The objective of the present article is to present an effective mathematical point estimator that, taking into account the data from DB1 and DB2, allows for the estimation of the values of variable Z_7 at all points in DB1, as well as at other points within the 3D subspace occupied by the deposit.

2. POINT ESTIMATOR (A,U,Θ) WITH UPD KERNEL FUNCTION TO OBTAIN Z_{7E}

An estimator capable of determining the value Z_{7e} at each location $P_e = (X_e; Y_e; Z_e)$ in DB1 is required. At P_e the values Z_{1e}, Z_{2e}, Z_{3e}, Z_{4e}, Z_{5e}, and Z_{6e} are known.

To estimate Z_{7e} only m input data points ($m \leq n$) from DB2 will be used, which is conceptualized by stating that the estimation has a *Compact Support*. It is proposed that the estimator satisfies six conditions:

1. The result Z_{7e} depends on the spatial relationship between P_e and the ***m*** data points (P_i; Z_{7i}) belonging to DB2.
2. Z_{7e} depends on the values Z_{1e}, Z_{2e}, and Z_{3e}.

Under these two conditions, it is proposed that the estimator be linear and take the form:

$$Z_{7e} = L_1\Theta_{e1} + L_2\Theta_{e2} + \ldots + L_m\Theta_{em} + b_1 + b_2 Z_{1e} + b_3 Z_{2e} + b_4 Z_{3e} \quad \textbf{(1)}$$

Where the kernel function Θ of an estimator (A, U, Θ), as defined by Legrá-Lobaina in their 2017 work, must be specified.

Using mathematical vector notation, equation (1) can be written as the sum of two scalar products:

$$Z_{7e} = [L] \bullet [\Theta_e] + [b] \bullet [c] \quad \textbf{(2)}$$

Where:

$$[\boldsymbol{L}] = \begin{bmatrix} \boldsymbol{L}_1 \\ \ldots \\ \boldsymbol{L}_m \end{bmatrix} \qquad [\theta_e] = \begin{bmatrix} \Theta_{e1} \\ \ldots \\ \Theta_{em} \end{bmatrix} \qquad [\boldsymbol{b}] = \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \boldsymbol{b}_3 \\ \boldsymbol{b}_4 \end{bmatrix} \qquad [c] = \begin{bmatrix} 1 \\ Z_{1e} \\ Z_{2e} \\ Z_{3e} \end{bmatrix}$$

The third condition is proposed:

3. The Kernel Function Θ is of the PD type (Euclidean distance power with parameters ***p*** and ***s***):

$$\Theta_{ei} = d_{ei}^{-p} \text{ with: } p > 0 \text{ and } p \neq 2,4,6,\ldots; s \geq 0 \quad \textbf{(3)}$$

Where d_{ei} is calculated as the Euclidean distance between the two points P_e and P_i, including the smoothing factor s:

$$d_{ei} = \sqrt{(X_e - X_i)^2 + (Y_e - Y_i)^2 + (Z_e - Z_i)^2 + s^2} \qquad \textbf{(4)}$$

For points P_j and P_i, one can define:

$$d_{ji} = \sqrt{(X_j - X_i)^2 + (Y_j - Y_i)^2 + (Z_j - Z_i)^2 + s^2} \qquad \textbf{(5)}$$

Subsequently, a technique for finding the vectors *[L]* and *[b]* which, when substituted into expression (2), determine the value of Z_{7e}, is explained. For this, two further conditions are given:

4. The proposed estimator has the capability to estimate at any point P_i in DB2, and the result Z_{7e} coincides with Z_{7i}.

This means that for *j=1, …, m*:

$$Z_{7j} = L_1\Theta_{j1} + L_2\Theta_{j2} + \ldots + L_m\Theta_{jm} + b_1 + b_2 Z_{1j} + b_3 Z_{2j} + b_4 Z_{3j}$$

And thus, the following system of linear equations (SLE) can be defined:

$$\begin{cases} L_1\Theta_{11} + L_2\Theta_{12} + \ldots + L_m\Theta_{1m} + b_1 + b_2 Z_{11} + b_3 Z_{21} + b_4 Z_{31} = Z_{71} \\ \ldots \\ L_1\Theta_{m1} + L_2\Theta_{m2} + \ldots + L_m\Theta_{mm} + b_1 + b_2 Z_{1m} + b_3 Z_{2m} + b_4 Z_{3m} = Z_{7m} \end{cases} \qquad \textbf{(6)}$$

This system of linear equations (SLE) has *m + 4* unknowns: $L_1, \ldots, L_m, b_1, \ldots, b_4$ and only *m* equations (i.e., it is not square), which is an essential impediment to having a unique solution for any point in DB2. It is for this reason that, according to Legrá-Lobaina (2017), the following equations are added to the SLE:

$$\begin{cases} L_1 + L_2 + \ldots + L_m = 0 \\ L_1 Z_{11} + L_2 Z_{12} + \ldots + L_m Z_{1m} = 0 \\ L_1 Z_{21} + L_2 Z_{22} + \ldots + L_m Z_{2m} = 0 \\ L_1 Z_{31} + L_2 Z_{32} + \ldots + L_m Z_{3m} = 0 \end{cases} \qquad (7)$$

And thus, the square and symmetric system of linear equations (SLE) is constituted as (8):

$$\begin{cases} L_1\Theta_{11} + L_2\Theta_{12} + ... + L_m\Theta_{1m} + b_1 + b_2Z_{11} + b_3Z_{21} + b_4Z_{31} = Z_{71} \\ L_1\Theta_{21} + L_2\Theta_{22} + ... + L_m\Theta_{2m} + b_1 + b_2Z_{12} + b_3Z_{22} + b_4Z_{32} = Z_{72} \\ ... \\ L_1\Theta_{m1} + L_2\Theta_{m2} + ... + L_m\Theta_{mm} + b_1 + b_2Z_{1m} + b_3Z_{2m} + b_4Z_{3m} = Z_{7m} \\ L_1 + L_2 + ... + L_m = 0 \\ L_1Z_{11} + L_2Z_{12} + ... + L_mZ_{1m} = 0 \\ L_1Z_{21} + L_2Z_{22} + ... + L_mZ_{2m} = 0 \\ L_1Z_{31} + L_2Z_{32} + ... + L_mZ_{3m} = 0 \end{cases} \quad (8)$$

Which can be written in matrix form as expression (9):

$$\begin{bmatrix} \Theta_{11} & ... & \Theta_{1m} & 1 & Z_{11} & Z_{21} & Z_{31} \\ ... & ... & ... & ... & ... & ... & ... \\ \Theta_{m1} & ... & \Theta_{mm} & 1 & Z_{1m} & Z_{2m} & Z_{3m} \\ 1 & ... & 1 & 0 & 0 & 0 & 0 \\ Z_{11} & ... & Z_{1m} & 0 & 0 & 0 & 0 \\ Z_{21} & ... & Z_{2m} & 0 & 0 & 0 & 0 \\ Z_{31} & ... & Z_{3m} & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} L_1 \\ ... \\ L_m \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} Z_{71} \\ ... \\ Z_{7m} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (9)$$

And by denoting the matrices:

$$\Theta_{m4} = \begin{bmatrix} 1 & Z_{11} & Z_{21} & Z_{31} \\ ... & ... & ... & ... \\ 1 & Z_{1m} & Z_{2m} & Z_{3m} \end{bmatrix} \quad \Theta_{4m} = \begin{bmatrix} 1 & ... & 1 \\ Z_{11} & ... & Z_{1m} \\ Z_{21} & ... & Z_{2m} \\ Z_{31} & ... & Z_{3m} \end{bmatrix} \quad Z_{d7} = \begin{bmatrix} Z_{71} \\ Z_{72} \\ ... \\ Z_{7m} \end{bmatrix}$$

Then (9) can be written in compact form as:

$$\begin{bmatrix} A & \Theta_{m4} \\ \Theta_{4m} & 0 \end{bmatrix} \begin{bmatrix} L \\ b \end{bmatrix} = \begin{bmatrix} Z_{d7} \\ 0 \end{bmatrix} \quad (10)$$

If the inverse of the matrix $\begin{bmatrix} A & \Theta_{m4} \\ \Theta_{4m} & 0 \end{bmatrix}$ exists, then the solution to the system (10) exists and is obtained as:

$$\begin{bmatrix} L \\ b \end{bmatrix} = \begin{bmatrix} A & \Theta_{m4} \\ \Theta_{4m} & 0 \end{bmatrix}^{-1} \begin{bmatrix} Z_{d7} \\ 0 \end{bmatrix} \quad (11)$$

 Minería y Geología / v.41 n.2/ abril-junio / 2025/ p. 172-187 ISSN 1993 8012

With the results [L] y [b], the point estimation can be performed according to expression (2).

Given that obtaining each estimated value Z_{7e} requires solving a square system of linear equations of order *m+4*, and considering that the behavior of the variable Z_7 is specific to the various regions of the deposit, it is convenient to impose rules regarding the data support used in each estimation. Thus, the following condition is stated:

5. To perform the point estimation Z_{7e} at P_e, a **Compact Support** from DB_2 will be used, comprising *m* points ($m \leq n$) that satisfy at least one of the following conditions:

a. Have the same classification Z_{4e} as P_e;

b. Have the same classification Z_{5e} as P_e;

c. Have the same classification Z_{6e} as P_e;

d. The maximum number of m points to be used should be fixed so that (for computational efficiency reasons) m does not exceed 100, and these m points should be the closest to P_e in terms of Euclidean distance.

The value of *m* can be predetermined or determined under certain conditions. In this work, it is proposed to heuristically set the maximum value of *m* for each estimation.

Finally, an important condition is stated to guarantee that the estimation lies within a range close and convenient to the range defined by the data in DB2:

6. The estimation Z_{7e} is categorized as **strictly admissible** if Z_{7e} exists and:

   a. Z_{7e} is less than or equal to the maximum value of Z_7 in DB2.
   b. Z_{7e} is greater than or equal to the minimum value of Z_7 in DB2.

However, this condition can and should be relaxed when it is necessary to increase the estimator's feasibility. For this purpose, the estimation is defined as **admissible** if it can be calculated and its result lies within the adjusted ranges, meaning the following holds:

   a. Z_{7e} is less than or equal to 1.05 times the maximum value of Z_7 in DB_2.
   b. Z_{7e} is greater than or equal to 0.95 times the minimum value of Z_7 in DB_2.

It should be noted that these range adjustment values (1.05 and 0.95) can be redefined as suitable for the procedure in which the estimator is implemented.

3. DUAL ESTIMATOR AND ERROR BOUNDING

As has been demonstrated (Legrá-Lobaina, 2017), the described estimator has a dual form which, using the notation of the compact support of ***m*** points, is described as follows:

- The SLE (10) is redefined using the transpose of matrix A (denoted A_T):

$$\begin{bmatrix} A_T & \Theta_{mt} \\ \Theta_{tm} & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ b \end{bmatrix} = \begin{bmatrix} \Theta_e \\ c \end{bmatrix} \quad (12)$$

$$\text{Donde}: \begin{bmatrix} \Theta_e \\ c \end{bmatrix} = \begin{bmatrix} \Theta_{e1} \\ \ldots \\ \Theta_{em} \\ 1 \\ Z_{1e} \\ Z_{2e} \\ Z_{3e} \end{bmatrix}$$

Note that by solving the SLE (12), the vector [**λ**] is now calculated instead of the vector *[L]*.

- The new expression for obtaining the estimated value is:

$$Z_{7e} = \begin{bmatrix} \lambda \\ b \end{bmatrix} \bullet \begin{bmatrix} Z_{d7} \\ 0 \end{bmatrix} = [\lambda] \bullet [Z_{d7}] = \sum_{i=1}^{m} \lambda_i Z_{7i} \quad (13)$$

$$Z_{7e} = \lambda_1 Z_{71} + \lambda_2 Z_{72} + \ldots + \lambda_m Z_{7m} \quad (14)$$

- The expression for bounding the estimation error in this case (Legrá-Lobaina, 2018) is:

$$\alpha_e = \sum_{i=1}^{m} |\lambda_i| \, |Z_{7i} - Z_{7e}| \quad (15)$$

4. CROSS-VALIDATION FOR EVALUATING ESTIMATORS

Various cross-validation techniques exist (Arlot and Celisse, 2010; Tibshirani, 2013; Zhang & Yang, 2015; Legrá-Lobaina, 2020). In the scenario of the present investigation, the *leave-one-out* procedure will be adopted as the method to evaluate the quality of the data and the selected estimator. This technique involves estimating the value of Z_7 at each data point P_i in DB_2 but excluding that specific point from the estimation process. The result of this estimation, denoted Z_{7ie}, is compared with the sampled value Z_{7i} using a criterion that indicates the quality of the estimation. This rating is higher the closer the estimates are to their corresponding sampled values.

As a percentage measure of closeness between each sampled value Z_{7i} and its corresponding value estimated using the remaining data (Z_{7ie}), the absolute percentage relative difference is proposed (16):

$$dZ_{7i} = 100\frac{\left|Z_{7i} - Z_{7ie}\right|}{\left|Z_{7i}\right|} \quad (16)$$

The quality of the (Data, Estimator) pair can now be evaluated based on the statistical distribution of the admissible dZ_{7i} values determined for the *m* values in DB_2. For example, the frequencies of dZ_{7i} results falling within [0%; 10%], (10%; 20%], (20%; 30%], and greater than 30% are significant.

It is also productive to consider the linear correlation coefficient C between the Z_{7i} values and the Z_{7ie} values as a measure of the correspondence between the measured and estimated values. The formula for calculating C (Miller et al., 2005) is given in (17):

$$C = \left|\frac{\sigma_{Z_7} \times \sigma_{Z_{7e}}}{\sigma_{Z_7 Z_{7e}}}\right| \quad (17)$$

Where:

σ_{Z_7} *y* $\sigma_{Z_{7e}}$ are the standard deviations of the variables Z_7 and Z_{7e}, respectively. Likewise, $\sigma_{Z_7 Z_{7e}}$ is the covariance between the variables.

Therefore, to quantitatively evaluate the quality of the estimator:

- It must be known how many and which estimations were not admissible (because they could not be performed or were outside the adjusted ranges). If all estimations were performed, the estimator is termed a **Total Estimator**; and if all estimations are within the adjusted ranges, it is termed an **In-Range Estimator**.

- The number of dZ_i values below 10% must be determined. It is also useful to know how many of the remaining values are below 20%; and of the rest, how many are below 30%.
- The value of the correlation coefficient C will also be determined, which preferably should be close to 1.

These quantitative criteria allow for determining whether the estimator is satisfactory for the known data and enable the decision to either review the data or change the estimator's configuration—in this case, selecting new values for *p*, *s*, and *m*.

5. PROCEDURE FOR SELECTING OPTIMAL ESTIMATORS

A mathematical optimization problem is one where the absolute or relative extreme value (maximum or minimum) of an objective function is sought. This function depends on the problem's data and on variables (termed intermediate and decision variables) linked through equations and inequalities—constraint relations that define the problem's feasible solution set (Legrá-Lobaina, 2022). In summary, solving an optimization problem involves finding for which values of the decision variables that satisfy the constraints the objective function reaches an extreme value, i.e., a minimum or maximum.

From a practical standpoint, as a means to solve highly complex scientific-technological problems, the concept and process of mathematical optimization have been made more flexible. It is also understood as the search for a set of feasible solutions, from which to select those that satisfy certain optimization criteria (Arzola-Ruiz, 2000; Blum & Oli, 2003).

In the present work, the first step will be to formally define the decision variables as the parameters *p*, *s*, and *m*. Each combination of their values defines a UPD estimator, which will be considered feasible if it is a total estimator and operates within the adjusted ranges.

The final step is to select the best among the feasible estimators using one of the following criteria:

1. Having the smallest number of dZ_{7i} values > E=10%.
2. Having the highest linear correlation coefficient C.

In practice, the first task is to heuristically delimit the ranges within which the parameters *p*, *s*, and *m* take their values. These are denoted as p_{min}, p_{max}, s_{min}, s_{max}, m_{min}, and m_{max}.

An important aspect to consider is related to the discrete nature of the parameter ***m*** and the continuous nature of the parameters *p* and *s*, which classifies this optimization problem as a mixed continuous-discrete type. The decision proposed to address this situation is to convert the optimization problem into a discrete type by "discretizing" the variables *p* and *s*.

The second practical task is to choose the step sizes k_p, k_s, and k_m that determine the possible values of *p*, *s*, and *m*:

- $p = p_{min} = p_1; p_2; p_3; \ldots; p_{kp} = p_{max}$
- $s = s_{min} = s_1; s_2; s_3; \ldots; s_{ks} = s_{max}$
- $m = m_{min} = m_1; m_2; m_3; \ldots; m_{km} = m_{max}$

The choice of the values for k_p, k_s, and k_m is a very sensitive aspect for the effectiveness of the search for a set of feasible estimators and for determining the optimal one among them. It must be considered that if k_p, k_s, and k_m are very large, then a vast number of calculations will be required, potentially making the task unfeasible; conversely, if k_p, k_s, and k_m are very small, it will be difficult to find a sufficient number of admissible solutions to correctly determine an optimal estimator. Practice indicates that these values should be taken as large as the computational resources allow.

Under these conditions, an **Exhaustive Search** procedure is proposed (Rivera, 2004), which corresponds to so-called **Combinatorial Optimization** (Peng et al., 2003; Vidal et al., 2012). Note that for each combination of the k_p values of *p*, the k_s values of *s*, and the k_m values of *m*, a particular estimator can be obtained. The cross-validation test explained in point 4 is applied to each, determining if it is Total and if it is in Range (i.e., if it is feasible), the statistical distribution of dZ_i, and the value of the correlation coefficient C.

By applying the cross-validation procedure to the $k_p \times k_s \times k_m$ possible estimators, the set **F** of **T** feasible solutions, denoted F_1, F_2, …, F_t, is obtained. The subsequent task is to apply the optimization criterion.

As stated, the possible procedures for determining which of the *T* feasible estimators is optimal can be the following:

1. **Procedure Min[>10%]:** For each feasible estimator F_q, $q=1,\ldots,T$,, calculate how many of its values satisfy: $(dMV_i)_q>10\%$. The feasible solutions are ordered from the smallest to the largest value according to $(dMV_i)_q$, resulting in the set $[E_1,\ldots,E_T]$. The optimal estimator, E =

[E_1], is then taken as the first one in the ordered list, i.e., the one with the smallest count.

2. **Procedure Max[C]:** For each feasible solution F_q, *q=1,…,T*, calculate the linear correlation coefficient $(C)_q$. The feasible solutions are ordered from the largest to the smallest value according to $(C)_k$, resulting in the set *[E_1, …, E_t]*. The optimal estimator, *E = [E_1]*, is then taken as the first one in the ordered list, i.e., the one with the highest value.

In this step, a novel proposal is made: instead of obtaining a single optimal estimator via either the *Min[>10%]* or *MIN[C]* procedure described, it is better to obtain a **set of optimal estimators** by taking the first *t≤T* elements from each respective ordered list. Thus, the elements of the set *S = [E_1, …, E_t]* are feasible estimators E_r (r=1,…,t) with acceptable quality and parameters m_r, p_r, s_r. In this case, the set S is termed a **multiple estimator**.

By selecting the top *t* estimators, ordered according to the chosen optimization criterion, a broad set *S = [E_1, …, E_t]* of point estimators is made available. This set allows for the identification of which points in DB_2 exhibit excessively large dZ_i values across all (or a certain majority) of the tested estimators. If these points were rectified or removed from DB_2, a new data table, which could be termed DB_{21}, would be obtained.

Using the data from DB_{21} and potentially more suitable parameters for the "discretization," the described process can be repeated, potentially refining the optimization. These refinements can be iterated until the researcher deems that a suitable set S of estimators has been obtained for the data remaining in DB_{21}.

6. 3D ESTIMATION OF Z_7 VALUES IN DB1

Given that DB_{21} now contains *n* acceptable data points and *S* holds the *t* best estimators meeting the conditions described in section 3, the next step is to estimate the value of Z_7 for each location P_j *(j=1,…,k)* in DB*1*.

For any estimator E_r in *S (r=1,…,t)*, the new criterion for evaluating the quality of each point estimate Z_{7_e} will be the coefficient of variation:

$$CV_e = \frac{\alpha_e}{Z_{7e}} \quad (18)$$

Where:

α_e - Estimation error defined by equation (15).

Z_{7e} - Value estimated by either expression (2) or (14).

From equation (12) it is known that the sum of the weights is $\sum_{i=1}^{m} \lambda_i = 1$, acknowledging that the values of λ_i can be negative, zero, or positive. Since α_e is calculated using the absolute values $|\lambda_i|$, when these absolute values are conveniently small, small values of α_e will be obtained. The following questions now arise:

1. For each point P_e in DB_1. A priori, which estimator in S minimizes CV_e? **Answer:** A priori, it is not known which estimator in S minimizes CV_e.
2. If none of the estimators in S produce an estimation Z_{7e} at P_e such that CV_e is satisfactorily small, which new estimators can be tested to try to obtain lower values of CV_e?

   Answer: New estimators obtained through admissible positive and negative variations of the value m_r, which defines the size of the compact support for each estimator in S, can be tested.

Procedure for Estimating Z_{7e}

A. Set the parameter dm = 0.
B. Estimate Z_{7e} at P_e using each of the *r=1,…,t* estimators in S, setting $m = m_r + dm$, and each time store $(Z_{7er}\ CV_{er}\ m_r\ p_r\ s_r)$.
C. From the results of step B, select the result with the smallest value of CV_e, denoted as $(Z_{eM}\ CV_{eM}\ m_M\ p_M\ s_M)$. If the obtained value CV_{em} is satisfactorily small, then the procedure terminates, and the answer is $(Z_{eM}\ CV_{eM}\ m_M\ p_M\ s_M)$.
D. Repeat steps B and C, setting d_m:= -1. Step D is repeated, decreasing dm each time, until $m = m_r + d_m = 3$, at which point d_m is reset to 0 and step E is executed.
E. Repeat steps B and C, setting $d_m := d_m + 1$, but only until $m = m_r + d_m = 30$.

7. CONCLUSIONS

- For estimating the value Z_{7e} at any point P_e in DB1, a **Multiple Multivariate Estimator** (set S) is now available. Exhaustively applied across the possible variations of its parameters *p*, *s*, and *m*, it allows for obtaining results that are optimal with respect to the coefficient of variation CV_e, potentially qualifying the multiple estimator as effective.

- Based on the results explained in this work, it has been deemed appropriate to future study the utility of employing the kernel function of the UPD-L estimator introduced in Legrá-Lobaina (2015) as a form of multivariate estimator.
-

REFERENCES

Arlot, S. y Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4, 40-79. DOI: 10.1214/09-SS054

Arzola, José (2000). *Sistemas de Ingeniería*. Editorial Félix Varela, La Habana, Cuba. ISBN: 978-959-07-1762-8, DOI: 10.13140/RG.2.1. 2349.4249

Blum, A. & Oli, R. A. (2003). Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys,* 35(3), 268–308. Disponible en: http://kurs info.himolde.no/forskningsgrupper/optimering/phdkurs/Metaheuristics%20in%20Combinatorial%20Optimization.pdf

Ding, Jie; Tarokh, Vahid & Yang, Yuhong (2018). Model Selection Techniques - An Overview. *IEEE Signal Processing Magazine*, 1-21. DOI: 10.1109/MSP.2018.2867638. Disponible en: https://arxiv.org /pdf/1810.09583.pdf

Legrá-Lobaina, A. A. (2015). Método UPD-L para estimar valores de una variable geominera medidos en un conjunto de puntos de R^n. *Minería y Geología,* 31(1), 1-12.

Legrá-Lobaina, A. A. (2017). Modelos de malla basados en estimadores (A,U,Θ). *Revista HOLOS*, 33(4), 88-110. DOI: 10.15628/holos. 2017.5351.

Legrá-Lobaina, A. A. (2018). Evaluación del error en estimaciones (A,U,Θ). *Revista HOLOS*, 34(3), 1-23. DOI: 10.15628/holos. 2018.6193.

Legrá-Lobaina, A. A. y Terrero-Matos, E. (2019). Modelación de variables eólicas mediante estimadores (A,U,Θ) multivariados. *Minería y Geología*, 35(1), 84-99. ISSN 1993 8012.

Legrá-Lobaina, A. A. (2020). Sensibilidad de los estimadores (A,U,Θ). *Revista HOLOS*, 36(1), 1-18. DOI: 10.15628/holos.2020.7282.

Legrá-Lobaina, A. A. (2022). *Elementos teóricos y prácticos de la investigación científico-tecnológica*. Primera Edición Digital, Editorial Félix

Varela, ISBN: 978-959-07-2494-7. Disponible en: http://bibliografía.eduniv.cu:8083/

Miller, I.; Freund, J. y Johnson, R. (2005). *Probabilidades y Estadísticas para ingenieros.* Vol. I y II. 4ta ed. México: Prentice-Hall Hispanoamericana S. A. 624 p. ISBN: 0-13-712-761-8.

Peng, H.; Ozaqui, T.; Haggan-Ozaqui, V. y Toyoda, Y. (2003). A parameter optimization method for radial basis function type models. *IEEE Trans Neural Netw.,* 14(2), 432-440. Disponible en: http://dx.doi.org/10.1109/TNN.2003.809395

Rivera, Sergio. (2004). Estado del Arte en la Ubicación Óptima de Capacitores y Estudio de Optimalidad de la Solución mediante Búsqueda Exhaustiva. *Revista Ingeniería e Investigación,* 56, 67–72. Disponible en: https://www.researchgate.net/publication/259074530_Estado_del_Arte_en_la_Ubicacion_Optima_de_Capacitores_y_Estudio_de_Optimalidad_de_la_Solucion_mediante_Busqueda_Exhaustiva

Tibshirani, Ryan (2013). Model selection and validation 1: Cross-validation. *Data Mining*, 36, 1-26. Disponible en: https://www.stat.cmu.edu/~ryantibs/datamining/lectures/18-val1.pdf

Tomás Antonio, Justino; Polanco Almanza, Ramón G. y Legrá-Lobaina, A. A. (2020). Modelación 3D de la masa volumétrica mediante estimadores (A,U,Θ). *Minería y Geología*, 36(3), 300-315.

Vidal, V; Wolf, C & Dupont, F. (2012). Combinatorial mesh optimization. *The Visual Computer,* 28(5), 511-525. Disponible en: http://liris.cnrs.fr/Documents/Liris-5258.pdf

Zhang, Y. & Yang, Y. (2015). Cross-Validation for Selecting a Model Selection Procedure. *Journal of Econometrics,* 187, 95-112. Disponible en: http://users.stat.umn.edu/~yangx374/papers/ACV_v30.pdf

Additional Information

Conflict of Interest

No conflicts of interest are declared.

Author Contributions

AALL: Original idea, research design, drafting of the manuscript, critical revision of its content, and approval of the final version. **REPA:** Research design, review, and approval of the final version.**ETM, RGPA, JTA:** Review and approval of the final version.

ORCID
AALL: https://orcid.org/0000-0002-4793-4754
REPA: https://orcid.org/0000-0002-7873-9716
ETM: https://orcid.org/0000-0002-0686-3566
RGPA: https://orcid.org/0000-0001-8229-7044
JTA: https://orcid.org/0000-0001-7757-3251

Received: 03/12/2024
Accepted: 12/02/2025