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

ARTÍCULO ORIGINAL

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

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

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

[1]Universidad de Moa, Holguín, Cuba.

[2]Centro de Investigaciones para la Industria Minero Metalúrgica, La Habana, Cuba.

[3]Sociedad Minera de Lunhinga, Lucapa, Lunda Norte, Angola.

[*]Autor para la correspondencia: alegra@ismm.edu.cu

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.

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.

1. INTRODUCCIÓN

Durante el estudio de yacimientos minerales se realizan análisis muestrales geo-localizados de sus propiedades que permiten obtener datos que, con el auxilio de estimadores matemáticos puntuales, conducen a modelos que explican y pronostican el comportamiento de las propiedades que se estudian (Ding *et al.*, 2018; Legrá-Lobaina y Terrero-Matos, 2019; Tomás-Antonio *et al*., 2020). Sin perder generalidad se puede plantear la siguiente situación:

1.1. Datos disponibles, problema técnico general y problema matemático específico

Supongamos que a partir de muestreos preliminares se tienen dos bases de datos:

a. La primera (denominada BD_1) está referida a un estudio de muestreo de seis propiedades de un yacimiento en k geolocalizaciones espaciales (dadas por sus coordenadas). En cada punto, para algunas propiedades se determinan sus valores numéricos y otras propiedades toman valores clasificatorios. Este estudio, espacialmente denso genera k elementos de BD_1 y por su costo es relativamente económico. Para cada uno estos elementos se tienen los valores de los campos:

   - Referencias particulares geológicas de cada muestreo.
   - Coordenadas planas: X (este-oeste); Y (sur-norte); Z (cota o altura del muestreo, sobre el nivel del mar). A la triada P=(X;Y;Z) se le denomina coordenadas espaciales o localización geográfica de cada punto de muestreo;
   - Datos de las seis variables en estudio: Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, medidos o determinados en cada punto P de muestreo. En el presente trabajo se asume que los valores de las variables Z_1, Z_2, Z_3 cuantifican

resultados de laboratorio y mediciones; por otra parte, las variables Z_4, Z_5, Z_6 informan sobre clasificaciones o tipologías de las muestras;

b. La segunda base de datos (BD_2) está referida a otro estudio de 10 variables donde el muestreo de las nuevas variables es más complejo y caro, por ello es espacialmente menos denso con $n<k$ elementos con sus correspondientes valores de campo:

- Referencias particulares geológicas de cada muestreo.
- Coordenadas espaciales: P=(X; Y; Z) de cada muestreo;
- Datos medidos o determinados en cada muestreo: Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, Z_7, Z_8, Z_9, Z_{10}, donde los valores de las variables Z_7, Z_8, Z_9 y Z_{10} cuantifican otros resultados de mediciones y pruebas de laboratorio.

Se presenta el problema técnico general de modelar en el yacimiento mineral el comportamiento de la propiedad descrita por la variable Z_7 mediante una malla (Legrá-Lobaina, 2017) usando estimadores puntuales y los datos disponibles.

El problema matemático específico que se deriva es encontrar un estimador matemático puntual que, teniendo en cuenta los datos de BD_1 y BD_2, permita estimar los valores de la variable Z_7 en todos los puntos de BD_1 así como en otros puntos del espacio que ocupa el yacimiento.

1.2. Estimador eficaz de Z_7

En este contexto un estimador de Z_7 será denominado eficaz si:

1. usa la información disponible en los datos de entrada garantizando la eficiencia computacional de los cálculos y las exigencias para los rangos de los valores estimados;
2. asegura adecuadas capacidades explicativa y pronosticadora, valoradas monitoreadas mediante técnicas de validación cruzada aplicada a los datos de entrada de la BD_1 y BD_2;
3. determina cuáles datos de entrada son admisibles para el estimador;
4. asegura las condiciones para que el estimador sea capaz de obtener el valor de Z_{7e} para cada localización P_e de la BD_1;
5. calcula los valores de la cota del error para cada estimación puntual Z_{7e} realizada sobre la localización P_e de la BD_1, y

6. minimiza los valores de la cota del error de Z_{7e} realizada sobre P_e de la BD_1.

El objetivo del presente artículo es presentar un estimador matemático puntual eficaz que, teniendo en cuenta los datos de BD_1 y BD_2, permita estimar los valores de la variable Z_7 en todos los puntos de BD_1, así como en otros puntos del subespacio 3D que ocupa el yacimiento.

2. ESTIMADOR PUNTUAL (A,U,Θ) CON FUNCIÓN NÚCLEO *UPD* PARA OBTENER Z_{7E}

Se necesita un estimador capaz de determinar el valor Z_{7e} en cada localización $P_e = (X_e;Y_e;Z_e)$ de BD_1. En P_e se conocen los valores Z_{1e}, Z_{2e}, Z_{3e}, Z_{4e}, Z_{5e} y Z_{6e}

Para estimar Z_{7e} solo se usarán m datos de entrada (m≤n) de BD_2 lo cual se conceptualiza expresando que la estimación tiene un Soporte Compacto. Se propone que el estimador cumpla seis condiciones:

1. El resultado Z_{7e} depende de la relación espacial entre P_e y de m datos $(P_i;\ Z_{7i})$ pertenecientes a la BD_1.

2. Z_{7e} depende de los valores Z_{1e}, Z_{2e} y Z_{3e}.

Bajo estas dos condiciones se propone que el estimador sea lineal y tenga la forma (1):

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

Donde deberá definirse la función núcleo Θ de un estimador (A,U,Θ), según define Legrá-Lobaina en su trabajo del 2017.

Si se utiliza la notación matemática vectorial, la ecuación (1) se puede escribir como la suma de dos productos escalares (2):

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

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

Se propone la tercera condición (3):

3. La Función Núcleo Θ es del tipo *PD* (potencia de distancia euclidiana de parámetros ***p*** y ***s***):

$\Theta_{ei} = d_{ei}^{p}$ con: p>0 y diferente de 2, 4, 6,…; s≥0 (3)

Donde se calcula d_{ei} como la distancia euclidiana entre los dos puntos P_e y P_i incluyendo el factor de suavización s (4):

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

Para los puntos P_j y P_i se puede definir (5):

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

A continuación, se explica una técnica para encontrar los vectores *[L]* y *[b]* que, sustituidos en la expresión (2), determinan el valor de Z_{7e}. Para ello se tienen otras dos condiciones:

4. El estimador que se propone tiene la capacidad de estimar en cualquier punto P_i de la BD_2 y el resultado Z_{7e} coincide con Z_{7i}

Eso quiere decir que se cumple para *j=1,…,m*:

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

Y de esta manera se puede definir el sistema de ecuaciones lineales *(SEL)* (6)*:*

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

Este SEL tiene *m+4* incógnitas: *L_1, …,L_m, b_1, …,b_4* y solo *m* ecuaciones (o sea, no es cuadrado) lo cual es un impedimento esencial para que tenga solución única para cualquier punto de la BD_2. Es por ello que, según Legrá-Lobaina (2017), se agregan al SEL las siguientes ecuaciones (7):

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

Y queda constituido el *SEL* cuadrado y simétrico (8):

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

Que se puede escribir matricialmente como la expresión (9):

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

Y denotando las matrices:

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

Entonces (9) se escribe en forma compacta (10):

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

Si existe la inversa de la matriz $\begin{bmatrix} A & \Theta_{m4} \\ \Theta_{4m} & 0 \end{bmatrix}$ entonces se dice que existe la solución del sistema (10) y se obtiene como (11):

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

Con los resultados [L] y [b] se puede realizar la estimación puntual según la expresión (2).

Dado que para obtener cada valor estimado Z_{7e} es necesario resolver un *SEL* cuadrado de orden *m+4* y, además, el comportamiento de la variable Z_7 es propio de las diversas regiones del yacimiento, entonces es conveniente imponer reglas respecto al soporte de los datos que se utilizan en cada estimación y se enuncia la condición:

5. Para realizar la estimación puntual Z_{7e} en P_e se utilizará un Soporte Compacto de la BD_2, cuyos m puntos *($m \leq n$)* cumplan al menos una de las condiciones:

a. Tengan la misma clasificación Z_{4e} que tiene P_e;

b. Tengan la misma clasificación Z_{5e} que tiene P_e;

c. Tengan la misma clasificación Z_{6e} que tiene P_e;

d. Debe fijarse el número máximo de *m* puntos a emplear de manera que (por razones de eficiencia computacional) m no sobrepase los *100*, y estos *m* puntos, deberán ser los más cercanos a P_e desde el punto de vista de la distancia euclidiana.

El valor de *m* puede ser prefijado de antemano o puede ser determinado bajo ciertas condiciones. En este trabajo se propone fijar heurísticamente el valor máximo de *m* para cada estimación.

Finalmente, se enuncia una importante condición que garantiza que la estimación está en un rango cercano y conveniente al rango definido por los datos de la BD_2:

6. Se categoriza la estimación Z_{7e} como estrictamente admisible si existe Z_{7e} y:

a. Z_{7e} es menor o igual que el máximo de Z_7 en la BD_2

b. Z_{7e} es mayor o igual que el mínimo de Z_7 en la BD_2

Sin embargo, esta condición puede y debe suavizarse cuando es necesario aumentar la viabilidad del estimador. Para ello se define la estimación como admisible si ella puede calcularse y su resultado está en los rangos ajustados, es decir se cumple:

a. Z_{7e} es menor o igual que el *1,05* del máximo de Z_7 en la BD_2.

b. Z_{7e} es mayor o igual que el *0,95* del mínimo de Z_7 en la BD_2.

Debe observarse que estos valores de ajuste de rango *(1,05* y *0,95)* pueden redefinirse según convenga al procedimiento donde se implemente al estimador.

3. ESTIMADOR DUAL Y ACOTACIÓN DEL ERROR DE ESTIMACIÓN

Como se ha demostrado (Legrá-Lobaina, 2017), el estimador descrito tiene una forma dual que, usando la notación del soporte compacto de m puntos, se representa de la siguiente manera:

- El SEL (10) se redefine usando la matriz transpuesta de *A* (denotada A_T), (12):

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

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

Nótese que al resolver el SEL (12) ahora se calcula el vector *[λ],* en lugar del vector *[L].*

- La nueva expresión para obtener el valor estimado es (13) y (14):

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

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

- La expresión para acotar el error de estimación en este caso Legrá-Lobaina, 2018) es (15):

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

4. VALIDACIÓN CRUZADA PARA EVALUAR ESTIMADORES

Existen diversas técnicas de validación cruzada (Arlot y Celisse, 2010; Tibshirani, 2013; Zhang & Yang, 2015; Legrá-Lobaina, 2020). En el escenario de la presente investigación se tomará como procedimiento para evaluar la calidad de los datos y del estimador seleccionado, la técnica *leave one out* ("omita uno"). Esta técnica consiste en estimar el valor de Z_7 en cada punto de los datos P_i de la BD_2, pero sin incluir ese punto. El resultado de la estimación, denominado Z_{7ie}, se compara con el valor muestreado Z_{7i} mediante algún criterio que indique la calidad de la estimación. Esta calificación es mayor en la medida en que las estimaciones sean más cercanas a sus correspondientes valores muestreados.

Como medida porcentual de cercanía entre cada valor muestreado Z_{7i} y su correspondiente valor estimado con el resto de los datos (Z_{7ie}) se propone tomar la diferencia relativa absoluta porcentual (16):

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

Ahora puede evaluarse la calidad del par (Datos, Estimador) a partir de la distribución estadística de los valores dZ_{7i} admisibles determinados para los *m* valores de BD_2. Por ejemplo, son significativos los valores de las frecuencias con que aparecen resultados de dZ_{7i} en [0%;10%], en (10%;20%], en (20%;30%] y mayores que 30%.

También es productivo considerar al coeficiente de correlación lineal *C* entre los valores Z_{7i} y los valores Z_{7ie} como medida de la correspondencia entre los valores medidos y los valores estimados. La fórmula para calcular *C* (Miller *et al.*, 2005) es dada en (17) :

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

Donde σ_{Z_7} *y* $\sigma_{Z_{7e}}$ son las correspondientes desviaciones estándar de las variables Z_7 y Z_{7e}. Asimismo $\sigma_{Z_7 Z_{7e}}$ es la covarianza entre las variables.

Entonces, para evaluar cuantitativamente la calidad del estimador:

- Se debe saber cuántas y cuáles estimaciones no fueron admisibles (porque no pudieron realizarse o están fuera de los rangos ajustados).

Si todas las estimaciones fueron realizadas el estimador se dice que es un Estimador Total; y si todas las estimaciones están en los rangos ajustados entonces se dice que es un Estimador en Rango;

- Se debe determinar cuántos valores dZ_i están por debajo del valor 10%. También es útil conocer cuántos valores de los restantes están por debajo de *20%*; y del resto, cuántos están por debajo del *30%*;
- También se determinará el valor de coeficiente de correlación *C,* que preferiblemente debe ser cercano a *1*.

Estos criterios cuantitativos permiten determinar si el estimador es satisfactorio para los datos conocidos y permiten tomar la decisión de revisar los datos o cambiar la configuración del estimador que en este caso será tomar nuevos valores de *p, s* y *m*.

5. PROCEDIMIENTO PARA SELECCIONAR ESTIMADORES ÓPTIMOS

Un problema de optimización matemática es aquel donde se busca el valor extremo (máximo o mínimo) absolutos o relativos de una función objetivo que depende de los datos del problema y de variables (denominadas intermedias y de decisión) vinculadas mediante ecuaciones e inecuaciones, relaciones de enlace que definen al conjunto de soluciones factibles del problema (Legrá-Lobaina, 2022). En síntesis, resolver un problema de optimización es encontrar para cuales de los valores de las variables de decisión que cumplen las relaciones de enlace, la función objetivo alcanza un valor extremo, o sea, mínimo o máximo.

Desde un punto de vista práctico, como vía para resolver problemas científico-tecnológicos muy complejos, se ha flexibilizado el concepto y el proceso de la optimización matemática y este también se entiende como la búsqueda de un conjunto de soluciones factibles, para entre estas seleccionar aquellas que satisfagan ciertos criterios de optimización (Arzola-Ruiz, 2000; Blum & Oli, 2003).

En el presente trabajo, el primer paso será definir formalmente que las variables de decisión son los parámetros *p, s* y *m* y que cada una de las combinaciones de sus valores define un estimador *UPD,* el cual será considerado factible si es un estimador total y en rangos ajustados.

El último paso es seleccionar el mejor entre los estimadores factibles utilizando uno de los siguientes criterios:

1. Tener menor cantidad de valores de $dMV_i>E=10\%$;
2. Tener mayor coeficiente de correlación lineal C.

En la práctica, la primera tarea es delimitar heurísticamente los rangos donde toman sus valores los parámetros *p, s* y *m*. Ellos son denominados p_{min}, p_{max}, s_{min}, s_{max}, m_{min} y m_{max}.

Un aspecto importante a considerar está relacionado con el carácter discreto del parámetro *m* y el carácter continuo de los parámetros *p* y *s*, por lo cual se clasifica este problema de optimización como del tipo continuo-discreto. La decisión que se propone ante esta situación es convertir al problema de optimización al tipo discreto mediante la "*discretización*" de las variables *p* y *s.*

La segunda tarea práctica es elegir los pasos *kp, ks* y *km* que permiten determinar los valores posibles de *p, s* y *m*:

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

La elección de los valores de *kp, ks* y *km* es un aspecto muy sensible para la efectividad de la búsqueda de un conjunto de estimadores factibles y determinar entre ellos el óptimo. Debe tenerse en cuenta que si *kp, ks* y *km* son muy grandes entonces se tendrá que hacer una gran cantidad de cálculos pudiendo ser inviable la tarea; por el contrario, si *kp, ks* y *km* son muy pequeños entonces será difícil encontrar suficientes soluciones admisibles para determinar correctamente un estimador óptimo. La práctica indica que hay que tomar estos valores tan grandes como los recursos de cálculo lo permitan.

Bajo estas condiciones se propone el procedimiento de Búsqueda Exhaustiva (Rivera, 2004), que corresponde a la denominada Optimización Combinatoria (Peng *et al.*, 2003; Vidal *et al.*, 2012). Nótese que para cada combinación de los *kp* valores de *p*, de los *ks* valores de *s* y de los *km* valores de *m* se puede obtener un estimador particular al cual se le aplica la prueba de validación cruzada explicada en el punto 4 y se determina si es Total y si está en Rango (o sea, si es factible), la distribución estadística de dZ_i y el valor del coeficiente de correlación *C*.

Aplicando el procedimiento de validación cruzada a las $kp\times ks\times km$ estimadores posibles se obtiene el conjunto *F* de *T* soluciones factibles denominadas F_1, F_2,…, F_T y la tarea que sigue es aplicar el criterio de optimización.

Como se ha dicho, los posibles procedimientos para determinar cuál de los *T* estimadores factibles es óptimo pueden ser los siguientes:

1. Procedimiento *Min[>10%]:* A cada estimador factible F_q, $q=1,\ldots,T$, se le calcula cuantos de sus valores cumplen que: $(dMV_i)_q>10\%$. Se ordenan las soluciones factibles de menor valor a mayor valor según $(dMV_i)_q$ resultando el conjunto $[E_1,\ldots,E_T]$ y, finalmente, se toma como estimador óptimo, $E=[E_1]$, el primero de la lista ordenada, es decir, el de menor valor;

2. Procedimiento *Min[C]:* A cada solución factible F_q, $q=1,\ldots,T$, se les calcula el coeficiente de correlación lineal $(C)_q$. Se ordenan las soluciones factibles de mayor valor a menor valor según $(C)_k$ resultando el conjunto $[E_1,\ldots,E_T]$ y, por último, se toma como estimador óptimo, $E=[E_1]$, el primero de la lista ordenada, es decir, el de mayor valor.

En este paso se propone como novedad que, en lugar de obtener un único estimador óptimo por cualquiera de los dos procedimientos Min*[>10%] o* Min*[C]* descritos, mejor se obtenga un conjunto de estimadores óptimos tomando los $t \leq T$ primeros de la lista ordenada en cada caso. Entonces los elementos del conjunto $S = [E_1,\ldots,E_t]$ son estimadores factibles E_r $(r=1,\ldots,t)$ de calidad aceptables con parámetros m_r, p_r, s_r. En este caso se dice que el conjunto *S* es un estimador múltiple.

Al seleccionar los mejores *t* estimadores, ordenados según el criterio de optimación seleccionado, entonces se dispone de un amplio conjunto $S=[E_1,\ldots,E_t]$ de estimadores puntuales que van a permitir determinar cuáles son los puntos de BD_2 donde todos (o cierta mayoría) de los estimadores probados tienen valores demasiado grandes de dZ_i. Si estos puntos fuesen rectificados o eliminados en la BD_2, entonces se obtiene una nueva tabla de datos que puede denominarse BD_{21}.

Con los datos de BD_{21} y parámetros más convenientes de la "discretización", se puede repetir el proceso descrito con lo cual se pudiera refinar la optimización. Estos refinamientos pueden sucederse hasta que el investigador considere que se obtuvo un conjunto *S* de estimadores adecuados para los datos que queden en BD_{21}.

6. ESTIMACIÓN 3D DE LOS VALORES DE Z_7 EN LA BD_1

Puesto que en BD_{21} ya se tienen *n* datos aceptables y en *S* están los *t* mejores estimadores que cumplen las condiciones descritas en el epígrafe *3*, ahora es el momento de estimar para cada localización P_j $(j=1,\ldots,k)$ de la BD_1 el valor de Z_7.

Para cualquier estimador E_r *de S* $(r=1,\ldots,t)$, el nuevo criterio de evaluación de la calidad de cada estimación puntual Z_{7e} será el coeficiente de variación (18):

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

Donde:

α_e - Error de estimación definida por la ecuación (15).

Z_{7e} - Valor estimado por cualquiera de las expresiones (2) y (14).

De la ecuación (12) se conoce que la suma de los ponderadores es $\sum_{i=1}^{m} \lambda_i = 1$ reconociéndose que los valores de λ_i pueden ser negativos, nulos o positivos. Dado que al calcular α_e se usan los valores de $|\lambda_i|$ entonces, cuando estos valores absolutos son convenientemente pequeños se obtendrán valores pequeños de α_e. Ahora surgen las siguientes preguntas:

1. Para cada punto P_e de la BD_1. *A priori*, ¿cuál de los estimadores de *S* minimiza CV_e?

   Respuesta: A priori no se sabe cuál estimador de S minimiza CV_e.

2. Si ninguno de los estimadores de *S* genera una estimación Z_{7e} en P_e, tal que CV_e sea satisfactoriamente pequeño, ¿cuáles nuevos estimadores se pueden probar para tratar de obtener menores valores de CV_e?

   Respuesta: Pueden probarse nuevos estimadores obtenidos mediante las variaciones positivas y negativas admisibles del valor de m_r que define el tamaño del soporte compacto de cada estimador de S.

Entonces, se propone el siguiente:

Procedimiento de Estimación de Z_{7e}

A. Tomar el parámetro $d_m=0$

B. Estimar Z_{7e} en P_e con cada uno de los $r=1,…,t$ estimadores de *S* tomando $m=m_r+d_m$ y guardar cada vez $(Z_{7er}\ CV_{er}\ m_r\ p_r\ s_r)$.

C. De los resultados del paso B, seleccionar el resultado que tenga el menor valor de CV_e *al que se denomina* $(Z_{eM}\ CV_{eM}\ m_M\ p_M\ s_M)$. *Si el valor* obtenido $\boldsymbol{CV_{eM}}$ es satisfactoriamente pequeño entonces finaliza el procedimiento y la respuesta es $(Z_{eM}\ CV_{eM}\ m_M\ p_M\ s_M)$.

D. Se repiten pasos B y C tomando d_m:=-1. El paso D se repite disminuyendo cada vez d_m hasta que sea $m=m_r+d_m=3$ en cuyo caso se inicializa $d_m=0$ y se ejecuta el paso E.

E. Se repiten pasos B y C tomando $d_m:=d_m+1$ pero solo hasta que $m=m_r+d_m=30$.

7. CONCLUSIONES

- Para estimar el valor Z_{7e} en cualquier punto de P_e de la BD_1 ahora se dispone de un Estimador Multivariado Múltiple (conjunto *S*) que, aplicado exhaustivamente en las variaciones posibles de sus parámetros *p, s y m*, permite obtener resultados óptimos respecto al coeficiente de variación CV_e que potencialmente califican como eficaz al estimador múltiple.

- A partir de los resultados explicados en este trabajo, se ha considerado conveniente estudiar en el futuro la utilidad del uso de la función núcleo del estimador UPD-L introducida en Legrá-Lobaina (2015) como una forma de estimador multivariado.

8. REFERENCIAS

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-Ruiz, J. (2000). *Sistemas de Ingeniería*. La Habana: Félix Varela. 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. Consultado: 24/10/2024. Disponible en: http://kurs info.himolde.no/forskningsgrupper/optimering/phdkurs/Metaheuristics%20in%20Combinatorial%20Optimization.pdf

Ding, J., Tarokh, V. & Yang, Y. (2018). Model Selection Techniques - An Overview. *IEEE Signal Processing Magazine*, 1-21. DOI: 10.1109/MSP.2018.2867638. Consultado: 23/09/2024. 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). *HOLOS*, *33*(4), 88-110. DOI: 10.15628/holos. 2017.5351.

Legrá-Lobaina, A. A. (2018). Evaluación del error en estimaciones (A,U,Θ). *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). *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. La Habana: Félix Varela. ISBN: 978-959-07-2494-7. Consultado: 12/08/2024. 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. Consultado: 17/10/2024. Disponible en: http://dx.doi.org/10.1109/TNN.2003.809395

Rivera, S. (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. Consultado: 14/07/2024. 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, R. (2013). Model selection and validation 1: Cross-validation. *Data Mining*, *36*, 1-26. Consultado: 24/02/2024. Disponible en: https://www.stat.cmu.edu/~ryantibs/datamining/lectures/18-val1.pdf

Tomás-Antonio, J., Polanco-Almanza, R. 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. Consultado: 10/10/2024. 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. Consultado: 19/09/2024. Disponible en: http://users.stat.umn.edu/~yangx374/papers/AC V v30.pdf

Información adicional

Conflicto de intereses

No se declaran conflictos de intereses.

Contribución de autores

AALL: Idea original, diseño de la investigación, Redacción del borrador, revisión crítica de su contenido y aprobación de la versión final. **REPA**: diseño de la investigación, revisión y aprobación de la versión final. **ETM, RGPA, JTA**: revisión y aprobación de la versión final.

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

Recibido: 03/12/2024

Aceptado: 12/02/2025