

Estrategia para el pretratamiento masivo de datos geológicos en yacimientos lateredieíticos

Strategy for the mass treatment of geological data in lateritic deposits



<http://opn.to/a/P59ey>

Ramón Eddie Peña Abreu ^{1*}, Lázaro Fernández Martínez ², Alfonso Puentes Martín ³

RESUMEN : En las labores geológicas se trabaja con un gran número de datos, generado por etapas en las que intervienen numerosos técnicos y especialistas, esto produce una gama de errores. Pueden corregirse por simple inspección, pero en una tabla voluminosa resulta complicado. Se conocen diversos métodos automáticos de limpieza y corrección de datos para el caso de los errores estandarizados, a cada tipo se le aplica un algoritmo específico, pero no existe un método que sea capaz de limpiar o identificar todos los errores. La especificidad es la mejor herramienta de trabajo para lograr resultados satisfactorios, cuando los datos son multivariados es más complejo porque se necesitan criterios consistentes. Se planteó como objetivo crear un sistema de reglas que permitiera identificar los errores y tipificarlos de forma automática. Se presentan un conjunto de criterios, métodos de detección y corrección de errores propios de la Geología que deja abierta una ventana metodológica para dar un tratamiento similar a los datos de los procesos industriales.

Palabras clave: detección de errores, bases de datos, algoritmos, clasificación, componentes principales.

ABSTRACT : In the geology, when you arrive at the end of the exploration and the exploitation tasks and mining planning are begun, it works with a big number of data. These they are generated in different stages and for different people, that produces a range of shapes in the errors that adds to the common ones, taken place during the collection, the preparation of the sample and in the analytic laboratory. Most can be detected by simple inspection, but in a voluminous chart it can be very slow their elimination or correction. Diverse automatic methods of cleaning and correction of data are known for several types of standardized errors, inclusive the errors have been classified to apply to each type a specific algorithm of correction and cleaning, but it doesn't exist a method that is able to clean or to identify errors or possible errors in a global way, the specificity it is the best work tool to achieve satisfactory results, in the case of multivariate data the task are complicated and consistent approaches are needed that can be applied in simple algorithms, It thought about as objective to create a system of rules that allow to identify the errors and automatically correct them, with simple algorithms. It is presented a group of approaches and methods for detection and correction of errors characteristic of the geology, but it leaves open a methodological window to give a similar treatment to the data of the industrial processes.

Key words: errors detection, databases, algorithms, classification, main components.

Recibido: 22/04/2019

Aprobado en su forma original: 28/05/2019

¹Centro de Investigaciones del Níquel, Cuba

²Empresa del Níquel Ernesto Guevara de la Cerna, Grupo Cubaniquel, Ministerio de Energía y Minas (MINEM), Cuba

³Universidad de Holguín, Centro Universitario de Banes, Holguín, Cuba

*Autor para correspondencia: Ramón Eddie Peña Abreu. E-mail: ramoneddiepena@gmail.com

INTRODUCCIÓN

Las bases de datos geológicas (BDG) son estáticas desde el punto de vista de la variabilidad en el tiempo. Su conformación es un proceso dinámico, en el que se conoce aproximadamente la cantidad de datos que alcanzarán. Esto se define en el diseño del muestreo de exploración y es igual a la cantidad de sondeos multiplicada por las muestras por pozos, estas últimas se estiman por analogía con otros yacimientos y en ocasiones se obtienen con métodos geofísicos ([Vera & Legrá, 2001](#); [Vera, 2001](#); [Acosta & Gentoiu, 2007](#)). Las coordenadas de los puntos del sondeo se generan en ese proceso, previo al sondeo se conforman las tablas con las coordenadas diseñadas, lo que se corrige posteriormente pues suceden cambios entre la posición diseñada y la real. Las muestras obtenidas se entregan a los laboratorios, donde se realizan los análisis químicos y físicos, los resultados se añaden a la BDG, muestra a muestra de cada intervalo perforado. Se genera información textual con las descripciones geológicas de cada testigo que pueden tener varios tipos, categoriales, nominales, cardinales y textuales, se generan otras, entre ellas las mediciones geofísicas. En resumen una BDG es una Base de Datos (BD) estática con un gran volumen de datos de diferentes tipos.

El enlace de las tablas de la BD es generalmente por las coordenadas de cada punto al que le corresponde un registro. El proceso de construir las BD transcurre en largos períodos de tiempo, donde intervienen muchas personas, por esto suceden errores de anotación que pueden solaparse como datos reales e incoherencias de diversos tipos y formatos. La correcta confección de una BDG depende en gran medida de la solidez del conocimiento en la institución que la genere, fundamentalmente en la disciplina lograda, los aspectos informáticos y de organización. Debe contar con un sistema de trabajo y de control de la calidad que logre normalizar el producto final, que es la información del yacimiento.

Se ha propuesto como objetivo crear un sistema de reglas que permita identificar los errores y anomalías, tipificarlos de forma automática con algoritmos sencillos y generar

escenarios para tomar decisiones en el procesamiento estadístico.

Se tiene como hipótesis que el conocimiento de las propiedades de los datos que permitirá establecer reglas y algoritmos para la detección de errores y anomalías. Entre las propiedades se encuentran las cotas lógicas de los contenidos de las variables y la clasificación de los tipos de errores en: anomalías informacionales y del comportamiento geoquímico. La primera evalúa la cantidad de datos ausentes, sea en cada campo de la tabla que es una variable o en cada registro de una muestra evaluada en las variables que se le denominará individuo o muestra. Se evalúa la cantidad de información que aporta cada variable.

Para cada tipo de anomalía se le adiciona un campo en la BD, ellas se tipifican en categorías que incluyen los estados normales o deseados de las variables, de tal forma que se tenga un conjunto completo de los sucesos posibles. De igual forma se generan varios registros que no se incluyen en las tablas pero en ellos se realizan las mediciones de cada variable. El final de este proceso genera una gama de escenarios a partir de las combinaciones de las categorías de ambas anomalías. En estos escenarios se aplican los métodos estadísticos y se evalúa la información que pueden aportar cada uno de ellos. El análisis por escenarios genera nuevas perspectivas de interpretación que dejan al descubierto la información invisible.

MATERIALES Y MÉTODOS

Las BDG no están exentas de los errores típicos de las tablas de datos, los registros duplicados, los valores vacíos que no representan el cero (valores perdidos), la falta de sincronización entre las tablas y otros para los que existen herramientas desarrolladas en el software Microsoft Access 2013, el Lenguaje de Consulta Estructurado (SQL) y en Microsoft Excel 2013. Entre las herramientas estadísticas empleadas se encuentran los sistemas Statistica, Statgraphics Centurion XV y Spad 5.6.

El análisis de la tabla de datos se enfrenta en dos direcciones, en la búsqueda de *anomalías informacionales* y del *comportamiento geoquímico*. El principio fundamental es no eliminar ningún registro, excepto los duplicados o sin datos. Se crean dos variables categoriales,

Tabla 1. Errores más frecuentes en las bases de datos geológicas

Error	Anomalía que provoca	Solución
Espacio oculto	Si es una celda numérica, no se reconoce como tal y provoca errores en la aplicación de los cálculos	En Excel se marca el (los) campos completos y se utiliza la función de remplazo. En “Buscar” se escribe un espacio (“ ”) y en “Remplazar con” no se escribe nada y Excel asume vacío.
Cambio del separador decimal	Ídem al anterior	Ídem. En “Buscar” se escribe el separador incorrecto y en “Remplazar con”, el correcto.
Cadena alfanumérica donde debe estar un valor numérico	Ídem al anterior	En Excel se puede emplear la función “ESNUMERO” combinada con funciones lógicas, aunque se recomienda programar una macro que sea capaz de definir si hay un valor numérico en la celda aunque esté mezclado con otros caracteres.
En una variable categorial, una misma categoría se escribe de dos o más formas.	Se eleva la cantidad de categorías de la variable. Afecta los análisis estadísticos y desborda la capacidad de los softwares para esos análisis.	En Excel, se debe filtrar la tabla por la variable categorial y sustituir los valores incorrectos.
Celda en blanco	Genera un valor perdido	En algunos softwares es necesario rellenar estas celdas con un valor que identifique que se trata de un valor perdido.

una que califica los registros desde el punto de vista informacional y otra desde el geoquímico, para formar por filtrado, combinaciones que se procesan y se someten a los análisis estadísticos.

En la primera variable “Anomalía Informacional” se tipificaron como categoría, la escasa o dudosa información y otras en función de la cantidad de variables con datos por registro. La segunda variable “Anomalía Geoquímica” se analiza desde una perspectiva univariante, para determinar los intervalos que contienen valores atípicos y contrastarlos con la opinión de los especialistas expertos en los yacimientos estudiados.

Las variables se analizan en todos sus registros, la normalidad se analiza además en cada intervalo de profundidad, en 16 sectores angulares de un círculo virtual creado en el centro de cada zona del yacimiento, este procedimiento de validación aplicado a la BD que se muestra como ejemplo, se puede encontrar en (Peña & Perdomo 2014; Tomás *et al.*, 2018).

Las dos variables categoriales creadas para las anomalías informacionales y geoquímicas generan escenarios diferentes de procesamiento estadístico, le corresponden informaciones diferentes que pueden confrontarse. Esto es posible porque no todos los escenarios de

anomalías informacionales son excluibles de los análisis estadísticos, algunos no completan un conjunto determinado de variables. En el caso de las anomalías geoquímicas generalmente se excluyen, pero no se descarta su análisis por separado, en el caso de estudio que se presentan se analizan varias anomalías geoquímicas, lo que permitió dilucidar entre errores en la anotación y rarezas propias de los procesos de formación de los yacimientos.

Búsqueda de Anomalías Informacionales

Se comienza con la eliminación automática de pequeños errores de anotación que pueden ser corregidos de forma sencilla en el M. S. Excel y se relacionan a continuación:

Una vez se corrigen los valores de las celdas, se procede al análisis de la integridad de las tablas de datos (TD) de la BD. Se calcula el por ciento de celdas vacías y la integridad se calcula por la fórmula siguiente:

$$I_{TD} = \left(1 - \frac{n_v}{n_T}\right) \cdot 100 \quad (1)$$

Dónde: I_{TD} – integridad de la TD; n_v – celdas vacías; n_T – celdas totales a llenar.

Durante la etapa de análisis y validación de la TD se pueden completar los datos perdidos o reducir en registros y campos innecesarios de

acuerdo a la información que se logre obtener en el proceso de muestreo del yacimiento, por lo que puede variar la integridad en el tiempo, este último parámetro es discreto ya que se calcula en balances hasta que se garantice un valor representativo o se decida trabajar con los valores alcanzados.

El principio para clasificar los registros es el mismo, se valora su integridad por la [fórmula \(2\)](#), se denota $I_r^j(t)$ la integridad del registro en el momento t , donde $j = \overline{1, m}$; m es la cantidad de registros. $n_v^j(t)$ es el número de celdas vacías en el registro j en el balance t y n_T^j es el número de celdas a llenar en el registro j . Los valores de $I_r^j(t)$ genera un vector columna, cada elemento corresponde a un registro de la TD que se denotará $\widehat{\zeta}(t) = (I_r^1, I_r^2, I_r^3 \dots I_r^m)^T$ ¹, su norma es $C_{ex} = \widehat{\zeta}$, donde $\widehat{\zeta} = \max_j(I_r^j)$, de tal forma el criterio para excluir o marcar según un grado de información, será proporcional a la norma C_{ex} . Las categorías se establecen de acuerdo a $I_r^j < K \cdot C_{ex}$, $K \in \mathbb{R}^+$.

$$I_r^j(t) = \left(1 - \frac{n_v^j(t)}{n_T^j}\right) \cdot 100 \quad (2)$$

El análisis de los campos de la TD se evalúa la integridad de las variables y luego se analiza el grado de información que aporta cada una de ellas. Para validar la integridad se emplea la misma [fórmula \(2\)](#) transformada, en este caso $j = \overline{1, n}$, donde n es la cantidad de campos. El criterio de decantación es proporcional a la norma del vector de integridad, el valor de $K \in \mathbb{R}^+$ lo determina el investigador de acuerdo a las necesidades del trabajo, tanto para las filas como en las columnas. Las variables en los campos deben cumplir una función informacional, por ello se debe medir la capacidad de aporte de información que según [Shannon \(1948\)](#) se expresa en [\(4\)](#).

$$C = \lim_{t \rightarrow \infty} \frac{\log(n(t))}{t} \quad (3)$$

Donde $n(t)$ es la ley de variación de la probabilidad y t el parámetro de variación, C es la variación de la entropía con respecto a t , que en [Shannon \(1948\)](#) es el tiempo. Note que si $n(t)$ es una distribución continua de tipo exponencial o asintótica C tiende al infinito, entonces es más sencillo probar el ajuste de cada variable continua a una distribución. Cuando la cantidad de datos es grande, según los teoremas del límite ([Projorov & Rozanov, 1973](#); [Kolmogorov et al., 1982](#)), se asume que todas las distribuciones tienden a la normal, entonces basta con probar la normalidad de los datos². De no cumplirse la normalidad, el cálculo de la probabilidad se realiza con métodos no paramétricos, donde la distribución binomial sirve de apoyo. Para obtener C en el caso de las variables discretas, la probabilidad se calcula de acuerdo a la magnitud del cambio de sus modalidades.

La variable categorial que tipifica las anomalías informacionales contiene una codificación del estado en que se encuentra cada registro de la BD, expuesta en la [tabla 2](#). Los niveles se fijan de acuerdo con el tipo de yacimiento y las variables que le sean más importantes, en el caso de algunos procesos metalúrgicos se tienen como imprescindibles los elementos útiles y los nocivos al proceso, en otros solo interesan los elementos útiles. De esta forma puede que un grupo de variables deban estar siempre presentes, por ello los códigos de la [tabla 2](#) pueden variar en cada caso, sin perder el sentido que en ella se describe.

Búsqueda de anomalías del comportamiento geoquímico

El comportamiento geoquímico se valida de acuerdo a las relaciones entre las variables geoquímicas y los intervalos de valores en que ellas se enmarcan en cada zona, de acuerdo a las características geológicas reportadas en la bibliografía especializada o por especialistas expertos, ([Lavaut, 1987, 2004, 2005](#); [Rojas et al.,](#)

¹ $\widehat{\zeta}$ la marca encima de ζ indica que el vector se transpuso a una fila.

² Aunque es común en la literatura matemática y técnica usar los términos "Normal" y "No Normal" para referirse al ajuste o no de los datos a la distribución de Gauss, preferimos usar los términos en español "normal" y "anormal" según corresponde, de igual forma "normalidad" y "anormalidad".

Tabla 2. Códigos de las anomalías informacionales

Descripción	Código	Descripción	Código
Suma de los elementos químicos >100	3	Tercer nivel de completamiento de los datos. Se completan sólo con las variables imprescindibles.	7
Primer nivel de completamiento de los datos. Llenas todas las variables	4	Primer nivel de valor crítico en el completamiento de los datos. Con valores en dos variables.	8
Segundo nivel de completamiento de los datos. Ausente una variable.	5	Segundo nivel de valor crítico en el completamiento de los datos. Con valores en una sola variable.	9
Tercer nivel de completamiento de los datos. Se completan las variables fundamentales para el proceso metalúrgico.	6	Último nivel de valor crítico en el completamiento de los datos. Sin datos.	10

1998, 2001, 2005; Muñoz, 2007, 2009; Peña & Vera, 2013). Se pueden emplear criterios geológicos que permiten dilucidar anomalías reales de falsas anotaciones. Un criterio poco empleado es la propiedad composicional de los datos geológicos, en un individuo la suma de todos sus elementos es igual al 100 si se mide en por ciento, en otros casos será la base del sistema en que se mida la composición. Esta propiedad permite medir la parte del todo a la que no se le han realizado análisis químicos. Los valores elementales conforman un vector $(x_1, x_2, x_3 \dots x_n)$ para el cual se cumple (4) (En Mateu *et al.*, 2003).

$$\sum_{i=1}^n x_i = 100 \quad (4)$$

En la práctica empresarial es común que no se complete el vector con todos los elementos, entonces se ha realizado el siguiente análisis para identificar el nivel de información. Se tiene una parte conocida del vector y otra incógnita $(x_1, x_2, x_3 \dots x_k, x_{k+1} \dots, x_n)$. Se asume que de 1 a k están los elementos conocidos en el vector y de $k+1$ a n los elementos incógnitas, lo que se expresa:

$$\sum_{i=1}^k x_i + \sum_{i=k+1}^n x_i = 100 \quad (5)$$

Si se denota:

$$\sum_{i=1}^k x_i = d$$

y con:

$$\sum_{i=k+1}^n x_i = D$$

Se despeja en la ecuación (5), entonces:

$$D = 100 - d \quad (6)$$

Se deduce que para cada registro se puede calcular D_j y d_j con los que se pueden construir índices que validan la información geoquímica que contiene cada registro. El índice de completamiento o aporte de información se expresa en (7).

$$I_j^c = \frac{d_j}{D_j} \quad (7)$$

El por ciento de información obtenida e incógnita se exponen respectivamente en (8) y (9).

$$I_j^{obt.} = \frac{d_j}{100} \quad (8)$$

$$I_j^{inc.} = \frac{D_j}{100} \quad (9)$$

El índice $I_j^c > 1$ identifica los individuos que logran un valor de información de la muestra superior al desconocimiento, en caso contrario $I_j^c < 1$, el caso $I_j^c = 1$ es trivial, pero ocurre muy poco. Se construye una variable categorial para generar escenarios que permitan el procesamiento de los individuos con diferentes niveles de información. Los índices $I_j^{obt.}$ e $I_j^{inc.}$ aportan una información similar y pueden ser empleados con el mismo objetivo.

La clasificación por niveles de información definida por los coeficientes 7, 8 ó 9, permite el procesar datos del mismo nivel, esto disminuye el efecto que tiene el simplex (5) sobre clasificación

multivariada con la distancia euclidiana, fundamentalmente en la comparación de los individuos ([Mateu, et al., 2003](#); [Bear & Billheimer, 2016](#); [Egozcue & Pawlowsky, 2016](#)).

RESULTADOS Y DISCUSIÓN

Análisis de las anomalías informacionales

La búsqueda de anomalías informacionales se resume en la [tabla 6](#), se puede observar que aproximadamente el 90 % de las muestras del sondeo contienen datos válidos para el análisis estadístico, que en lo adelante le denominamos *efectivos*, sin embargo de ellos cerca del 60 % sólo cuenta con tres análisis químicos y menos del 30 % cuenta con seis. Los mayores problemas en la información se encuentran en el yacimiento PTG, un número grande de pozos se perforaron solo para medir el Ni, por lo que no son efectivos para un análisis multivariado. Estos escenarios de las variables serán los que se empleen para el procesamiento estadístico.

Análisis de las anomalías en el comportamiento geoquímico

Para definir con mayor precisión las anomalías que constituyen errores, se emplearon los métodos explicados en los puntos anteriores. Los valores fuera de los rangos o que no concuerdan con las relaciones conocidas se tipifican como anómalos ([Lavaut, 2004, 2005](#)). Se encontraron

ocho comportamientos anómalos que no superan el 3 % de los datos efectivos, los intervalos de valores en los que oscilan se exponen en la [tabla 3](#). En los grupos 2, 3, 6 y 7 la alúmina oscila por encima del 20 %, lo que difiere de los rangos establecidos en [Lavaut \(2004, 2005\)](#).

La génesis de estos yacimientos se encuentran en dos tipos fundamentales de rocas, las básicas y las ultrabásicas, en el primer caso puede originar tipos específicos de menas bauxíticas que se han detectado en algunas zonas de estos yacimientos ([Bergues, 2006](#)). Para definir estas menas a partir del análisis de datos, se utilizó el coeficiente α propuesto por [Bergues \(2006\)](#), los intervalos de α para determinar el tipo de mena se exponen en la [tabla 4](#).

$$\alpha = 0,85 \cdot \frac{\text{Silíce} - \text{Alúmina}}{\text{Alúmina}} \quad (10)$$

El empleo de los criterios de [Lavaut \(2004, 2005\)](#) y [Bergues \(2006\)](#) explican las anomalías, en la [tabla 5](#), tipificadas en los códigos del 1 al 10.

Los comportamientos anómalos no representan un peso importante en la totalidad de los datos ([tabla 6](#)), su distribución espacial es en extremo dispersa. Esas muestras se excluyen de los análisis multivariados porque sus valores anómalos en las principales variables distorsionan las tendencias del agrupamiento y la interpretación de los fenómenos geológicos. Las

Tabla 3. Intervalo de valores de las anomalías en las variables

Cód.	Nº de muestras	Fe		Ni		Co		SiO ₂		MgO		Al ₂ O ₃	
		Min	Máx	Min	Máx	Min	Máx	Min	Máx	Min	Máx	Min	Máx
1	5245	1,5	9,9	0,02	3,36	0,002	0,085	26,13	57,91	20,01	43,12	0,28	18,4
2	2431	1,1	21,4	0,01	1,52	0,001	0,180	20,56	67,47	0,04	23,42	20	47,29
3	52	20,4	29,5	0,03	0,4	0,005	0,076	0,32	5,43	0,1	1,56	33,33	44,66
4	74	1	11,6	0,01	0,36	0,001	0,041	22,21	70,89	1,16	16,67	12,93	19,96
5	4	4,7	11,6	0,13	0,45	0,010	0,022	69,00	77,1	0,72	11,96	0,56	8,78
6	119	17,3	39,9	0,1	1,27	0,101	0,614	0,87	14,8	0,06	4,82	25,02	44,04
7	47	12,6	27,8	0,22	0,79	0,102	0,253	16,30	35,08	0,16	4,48	25,08	35,72
8	6	41,2	72,2	0,29	0,72	0,012	0,086	2,33	38,96	1,62	70,81	11,64	19,83
Total	7978												
10	301812	0,02	59,3	0,01	4,85	0,001	1,5	0,15	66,06	0,01	38,07	0,09	54,36

Tabla 4. Clasificación de las anomalías de aluminio y silicatos según [Bergues \(2006\)](#)

Materiales	α	Materiales	α
Bauxitas	[-1; -0,75)	Arcillas poco bauxíticas	[-0,24; 0)
Bauxitas arcillosas	[-0,74; -0,50)	Materiales arcillosos	[0,0; 0,25]
Arcillas bauxíticas	[-0,49; 0,25)	Materiales detríticos	≥0,26

Tabla 5. Descripción de las anomalías en el comportamiento geoquímico de cuatro yacimientos.

Código	Descripción del comportamiento anómalo
1	Valores de sílice y MgO significativamente altos que concuerdan con valores bajos de Fe, la alúmina presenta valores muy bajos. Según Lavaut (2005) corresponde a una <i>serpentina hipergénica</i> . En la sílice y la alúmina los valores exceden lo común. El MgO y el Fe significativamente bajos. Provenientes de <i>cortezas por mafitas</i> según Lavaut (2005) , donde predominan los aluminosilicatos (Hidromicas, Montmorillonita, Caolinita, Halloysita y Gibbsita). Según la metodología propuesta por Bergues (2006) sólo el 7 % de las muestras en este grupo se estiman que sean bauxíticas, el resto se consideran materiales arcillosos y detríticos.
2	La alúmina muy alta excede lo común, el Fe en valores que no corresponden a las menas limoníticas, una sílice muy baja al igual que el MgO. Este comportamiento no se asocia a ninguno de los descritos en Lavaut (2005) . Según los valores de α calculados para las 52 muestras de este grupo, se clasifican como bauxitas lateríticas.
3	Existe una alta correlación del Al_2O_3 con el MgO y son independientes $MgO \perp SiO_2 \perp Fe$, esto no concuerda con los comportamientos conocidos de las lateritas (Lavaut, 2005 ; Muñoz, Orozco, Rojas et al., 2005 ; Muñoz, Rodríguez y Vera, 2009 ; Muñoz, Rodríguez y Barea, 2015) y confirmada en la actividad práctica de explotación (Peña y Sam, 2013 ; Peña y Vera, 2013). Según los valores de α no corresponden a materiales bauxíticos.
4	Este grupo de sólo cuatro muestras presenta la sílice muy alta, valores de MgO, Al_2O_3 y Fe muy bajos, tiene alteradas las correlaciones del $Fe \leftrightarrow MgO$ y $MgO \uparrow SiO_2$, mantiene la correlación $Fe \leftrightarrow SiO_2$. Se le denominó <i>producto atípico del intemperismo</i> .
5	Valores muy altos de la alúmina y el cobalto, no hay la correlación $MgO \perp Fe$, según el coeficiente α es <i>bauxita de alto cobalto</i> .
6	La sílice, la alúmina y el cobalto son altos, según el coeficiente α cerca del 66 % tienen carácter bauxítico, fundamentalmente <i>arcilla bauxítica de alto cobalto</i> .
7	Este grupo se compone de valores dudosos o escasos de información.
8	Código reservado para anomalías aún no detectadas.
9.	Sin anomalías notables.
10.	

Símbolos: $\perp$ - sin correlación; $\leftrightarrow$ correlación inversa; $\uparrow$ correlación directa.

anomalías se estudian como casos raros por separado. Se particulariza su análisis ya que no se conoce la influencia de estos materiales en el proceso metalúrgico ([Peña et al., 2007](#)).

La forma de tipificar las anomalías informacionales ([tabla 2](#)) se adaptó al tipo de yacimiento y se expone en la [tabla 6](#). Se observa que se logró un alto grado de información primaria de acuerdo a los recursos de la época, segunda mitad de la pasada centuria, sin medios de cómputos, ni recursos para los análisis avanzados que se fueron incorporando posteriormente.

En los dos primeros campos de la [tabla 6](#) se definieron varios escenarios acorde al nivel de datos en los registros. Se particularizó el estudio estadístico, la clasificación multivariada se realizó en el sexto escenario que contempla seis elementos químicos y los cuatro yacimientos. No se aplica al cuarto escenario porque se enmarca

en un sólo yacimiento. El software Spad 5.6 asocia los individuos inactivos a las clases, luego se corrige con un análisis discriminante realizado en los softwares Statgraphics Centurion y Statistica V. 8. Estos resultados no forman parte del objetivo en esta contribución, pero se encuentran en [Peña & Vera \(2013\)](#).

Se excluyeron y corrigieron errores de anotación como resultado de un profundo análisis de la zona, el pozo, la profundidad y otros parámetros que aportan elementos para decidir si existe una lógica desde el conocimiento mineralógico de las menas o se trata de errores, se empleó el conocimiento expuesto en la literatura especializada y se consultaron especialistas de dos empresas productoras de Ni laterítico ([Rojas & Orozco, 1998](#); [Rojas, 2001](#); [Rojas et al., 2005](#); [Muñoz et al., 2005](#); [Muñoz et al., 2009](#); [Muñoz et al., 2015](#); [Peña et al., 2017](#)).

Tabla 6. Resumen de las anomalías informacionales en las bases de datos de cuatro yacimientos

Anomalías		Yacimientos			Total	
Código	Descripción	CE	PTG	YN	YS	
	Total de Registros en la BD	51207 (100,00 %) (16,53 %)	141696 (100,00 %) (45,74 %)	55574 (100,00 %) (17,94 %)	61313 (100,00 %) (19,79 %)	309790 (100,00 %) (100,00 %)
4	Ocho análisis químicos y la susceptibilidad magnética	6888 (13,45 %) (2,22 %)				6888 (2,22 %) (2,22 %)
5	Ocho análisis químicos	9297 (18,16 %) (3,00 %)				9297 (3,00 %) (3,00 %)
6	Seis análisis químicos	13675 (26,71 %) (4,41 %)	(0,30 %) (0,14 %)	38483 (69,25 %) (12,42 %)	27867 (45,45 %) (9,00 %)	80457 (25,97 %) (25,97 %)
7	Tres análisis químicos	21345 (41,68 %) (6,89 %)	109358 (77,18 %) (35,30 %)	17035 (30,65 %) (5,50 %)	33438 (54,54 %) (10,79 %)	181176 (58,48 %) (58,48 %)
	Total de registros efectivos	51205 (100,00 %) (16,53 %)	106066 (74,85 %) (34,24 %)	55315 (99,53 %) (17,86 %)	61305 (99,99 %) (19,79 %)	273891 (88,41 %) (88,41 %)
Registros excluidos del análisis estadístico						
3	La suma del % de elementos químicos supera el 100 %				(0,00 %) (0,00 %)	(0,00 %) (0,00 %)
8	Sólo dos análisis químicos		(0,00 %) (0,00 %)			(0,00 %) (0,00 %)
9	Un Solo análisis Químico		31526 (22,25 %) (10,18 %)			31526 (10,18 %) (10,18 %)
	Sin Datos	(0,00 %) (0,00 %)	(0,27 %) (0,12 %)	(0,10 %) (0,02 %)	(0,01 %) (0,00 %)	(0,14 %) (0,14 %)
10	Pozos sin coordenadas		3724 (2,63 %) (1,20 %)	(0,37 %) (0,07 %)		3927 (1,27 %) (1,27 %)
	Tota excluidos	(0,00 %) (0,00 %)	35630 (25,15 %) (11,50 %)	(0,47 %) (0,08 %)	(0,01 %) (0,00 %)	35899 (11,59 %) (11,59 %)
Resumen por pozos						
	Total de pozos	5901 (100,00 %) (23,79 %)	9844 (100,00 %) (39,69 %)	3045 (100,00 %) (12,28 %)	6015 (100,00 %) (24,25 %)	24805 (100,00 %) (100,00 %)
	Pozos sin datos, excluidos		3064 (31,13 %) (12,35 %)			3064 (12,35 %) (12,35 %)
	Pozos sin coordenadas, excluidos		(2,3 %) (0,91 %)	(0,4 %) (0,05 %)		(1,0 %) (0,96 %)
	Total de pozos efectivos	51205 (100,00 %) (16,53 %)	106066 (74,85 %) (34,24 %)	55315 (99,53 %) (17,86 %)	61305 (99,99 %) (19,79 %)	273891 (88,41 %) (88,41 %)

Nota: El primer porcentaje se calcula respecto al total de registros del yacimiento y el segundo respecto al total de los cuatro yacimientos.

La tipificación de las anomalías geoquímicas se describe en la [tabla 5](#). Por ejemplo:

- se detectaron dos muestras con Fe>60 %, se marcaron para excluirlas de cualquier análisis estadístico.
- en una de ellas la suma de sus elementos superan el 100 %, el Fe, SiO₂ y Al₂O₃ exceden los límites conocidos, lo que impide definir el tipo de error que inutilizó el registro.

• en otra muestra la composición química es incoherente, la anomalía en el Fe coincidió con valores altos de SiO₂, Al₂O₃, Ni y Co. Se dedujo que corresponde a un material serpentinitico, que por error en la anotación se corrió un lugar decimal del Fe de 7,2 a 72, esta deducción no implica que se incluye la muestra en el procesamiento estadístico, esto es imposible sin que se certifique objetivamente su contenido, el análisis sirve

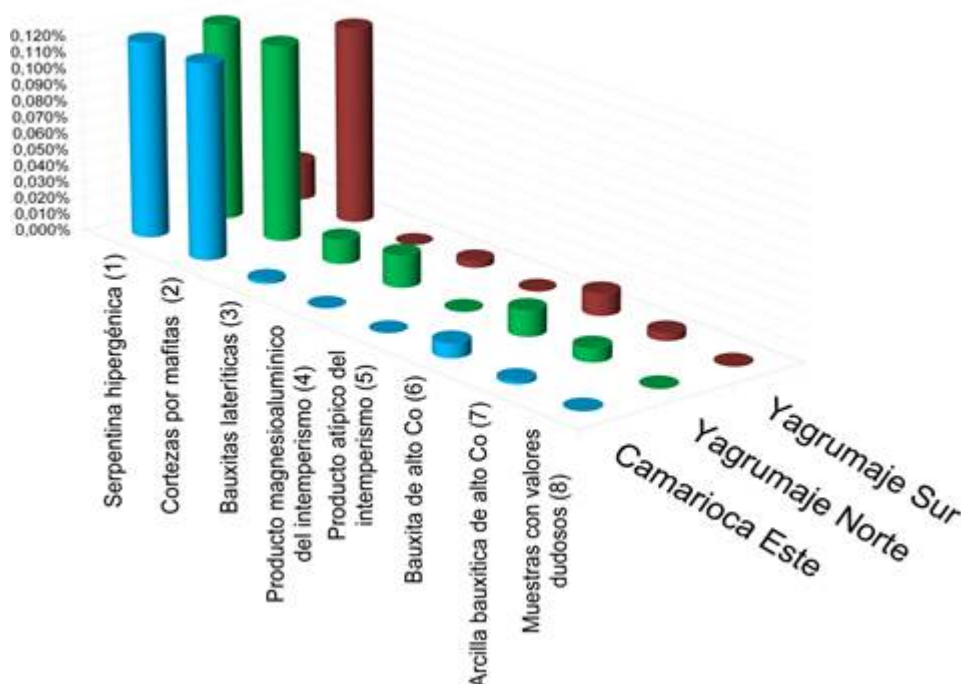


Figura 1. Resumen de las anomalías en el comportamiento geoquímico de los cuatro yacimientos. para otras decisiones puntuales en caso de que esto fuese necesario.

Las otras muestras atípicas se clasificaron como anomalías propias de los procesos geológicos. Las muestras con valores atípicos de SiO_2 y Al_2O_3 se analizan por la [fórmula \(10\)](#) y se clasifican según los intervalos de α (ver [tabla 4](#)) y se incluyen en la [tabla 5](#). Las muestras marcadas como bauxíticas se excluyen del análisis multivariado, se analizan por separado y se ubican espacialmente para definir las zonas de su posible aparición. En la [figura 1](#) se exponen las anomalías geoquímicas, se observa que las bauxitas no son importantes en ninguno de los yacimientos, nótese que no se incluye Punta Gorda, debido a que sus bases de datos no contemplan la información necesaria para definir anomalías geoquímicas. En total las anomalías que se exponen no llegan a alcanzar un 3 % de la BD, por lo que no son significativas, en general se descartan del análisis cerca de un 13 % de los registros lo que indica que la mayor parte de los problemas son de carácter informacional.

CONCLUSIONES

1. Se definieron dos tipos de anomalías, las informacionales y geoquímicas para los yacimientos lateríticos.
2. Se establecieron los criterios para determinar, en los yacimientos lateríticos, las anomalías en los registros y los campos.

3. Se logró establecer una estrategia de tratamiento para un gran volumen de datos geológicos, con el empleo secuencial de los criterios que identifican los dos tipos básicos de anomalías, las informacionales y las geoquímicas. Las primeras definen los escenarios de validez de los datos a partir de los que se construirá la información después del correspondiente procesamiento estadístico. La segunda define los tipos de individuos que contienen la base de datos, aísla aquellos que no responden de forma lógica a los procesos geológicos que dieron formación al yacimiento, para analizarlos sin valores extremos que afecten la interpretación y estudiar con detalle por separado los datos aislados. Lo anterior se logró mostrar a partir de un ejemplo, se aplicó la estrategia expuesta a las bases de datos de cuatro yacimientos lateríticos.

RECOMENDACIONES

Para el empleo de la metodología expuesta se debe trabajar con un equipo de expertos en la geología de los yacimientos que se analicen y posean las habilidades en el uso de los softwares que le permitan realizar el proceso expuesto.

Se trabaja en el desarrollo de una herramienta de software para preprocesar bases de datos geológicas con esta metodología

BIBLIOGRAFÍA

- Acosta Breal, J.; Gentoiu M. 2007. "Prospección con Georradar en los Yacimientos Lateríticos de Yagrumaje Sur". II Congreso de Minería (MINERIA'2007). III Simposio Geología, Exploración y Explotación de las Lateritas Niquelíferas, II Convención Cubana de Ciencias de la Tierra, La Habana, ISBN: 978-959-7117-16-2.
- Bear J. & Billheimer D. 2016. "A Logistic Normal Mixture Model for Compositional Data Allowing Essential Zeros." *Austrian Journal of Statistics*, 45(4):3-23, DOI:10.17713/ajs.v45i4.117.
- Bergues Garrido P. S. 2006. "Presencia de perfiles lateríticos con intercalaciones atípicas y bauxitas en el sector septentrional del río Yagrumaje." *Minería y Geología*, 22(1):1-8, ISSN 0258-5979.
- Egozcue J. J. & Pawlowsky Glahn V. 2016. "Changing the Reference Measure in the Simplex and Its Weighting Effects." *Austrian Journal of Statistics*, 45(4):25-44, DOI:10.17713/ajs.v45i4.126.
- Kolmogorov A. N.; Zhupbenkov I. G. & Projorov A. B. 1982. *Vidienie v teorii veroiatnostiey*, Moskva: Ed. Nauka.
- Lavaut Copa W. 1987. Control litológico-mineralógico de la mineralización en la corteza de intemperismo de ultramafitas del campo mineral yacimientos Punta Gorda. Tesis Doctoral, Camariocas y Piloto: Archivo E.G.S. Moscú.
- Lavaut Copa W. 2004. "Patrones de Meteorización de Rocas Ofiolíticas de Cuba Oriental: su Importancia para la Minería." *Minería y Geología* : 3-4, ISSN: 0258-5979, Available: <http://www.ismm.edu.cu/revistamg>, [Consulted: October 21, 2018].
- Lavaut Copa W. 2005. Clasificador para Modelación Litológica de Lateritas. Available: http://www.ismm.edu.cu/R_Geociencias%20y%20Quimica/volumen1.htm, [Consulted: March, 2007].
- Mateu Figueras G.; Martín Fernández J. A.; Pawlowsky Glahn V. & Barceló Vidal C. 2003. "El problema del análisis estadístico de datos composicionales". 27 Congreso Nacional de Estadística e Investigación Operativa, Lleida, abril del 8 al 11, Available: http://ima.udg.edu/~jamf/paper_INTRO_MMPB.pdf, [Consulted: February, 2019].
- Muñoz Gómez J. N.; Orozco Melgar G.; Rojas Purón A. & Cruz Orosa I. 2005. "Mineralogía de las menas lateríticas del yacimiento Punta Gorda: implicaciones técnica - operativas durante la explotación", II Simposio Geología, Exploración y Explotación de las Lateritas Niquelíferas, I Congreso de Minería (MINERIA'2005), I convención Cubana de Ciencias de la Tierra, La Habana, ISBN 959-7117-03-7.
- Muñoz Gómez J. N.; Rodríguez Infante A.; Vera Sardiñas L.O. & Martínez Vargas A. 2009. "Modelo Genético Teórico de los Yacimientos Lateríticos Asociados al Macizo Ofiolítico de Moa - Baracoa: Implicaciones de la Exploración de Explotación, III Simposio Geología, Exploración y Explotación de las Lateritas Niquelíferas, III Congreso de Minería (MINERIA'2009), III Convención Cubana de Ciencias de la Tierra, La Habana, ISBN: 978-959-7117-19-3.
- Muñoz Gómez J. N.; Rodríguez Infante A. & Barea Pérez I. 2015. "Variabilidad de los perfiles lateríticos en los yacimientos ferroniquelíferos de Moa". *Minería y Geología*, 31(2): 21-33, Available: <http://www.ismm.edu.cu/revistamg>, [Consulted: February, 2019].
- Peña Abreu R. E.; Sam Palanco C. M.; Majendíe Cemitier R.; Rivero Piña B. & Laborí Leyva N. 2007. "Determinación de las causas de degradación del óxido de níquel. Diagnóstico estadístico multivariado". *Tecnología Química Vol. XXVII (1):1-11*. Available: <https://www.uo.edu.cu/ojs/index.php/tq/article/viewPDFInterstitial/396/284>, [Consulted: May, 2009].
- Peña Abreu, R. E. C. & Sam Palanco. 2013. "Optimización del Muestreo Tecnológico en Yacimientos Lateríticos Cubanos." *Ciencias de la Tierra y Espacio*, Vol.14 (1). Available:

- Available: <http://www.iga.cu/Publicaciones/revista/index.htm> , [Consulted: October 14, 2018].
- Peña Abreu R.E., Vera Sardiñas L. O. 2013. Identificación de clases patrones para la modelación matemática en yacimientos de cubanos níquel, Ciencias de la Tierra y el Espacio, Vol. 14, No. 2, ISSN: 1729-3790. Available: <http://www.iga.cu/Publicaciones/revista/index.htm> , [Consulted: January 9, 2019].
- Peña Abreu, R. E. & Perdomo Orellana E. A. 2014. “Comportamiento Probabilístico del Hierro y el Níquel en Tres Yacimientos Lateríticos Cubanos”. Ciencia de la Tierra y el Espacio. Vol. 15(2), ISSN 1729-3790.
- Peña Abreu, R. E ; Cala Cardero, Y.; Capote Flores N.; Casanova Gómez, A; Toledo Sánchez, C. A.; Milia González, I. & Lisabet Sarracen E. 2017. “Caracterización mineralógica de tipologías del comportamiento geoquímico de las lateritas del oriente cubano”. Ciencias de la Tierra y el Espacio, Vol. 18(1): 12-23, ISSN 1729-3790. Available: <http://www.iga.cu/Publicaciones/revista/index.htm> , [Consulted: February 8, 2019].
- Projorov Y. B. & Rozanov. Y.A. 1973. Teoria Veroiatnostey. Asnovnie Poniatia, Predielnie Teoremi y Sluchainie Protsessin. 2^{da} ed. Masckva: Nauka. 494 p.
- Rojas Purón A . & Orozco G. 1998. “Iniciación al estudio de las fases minerales portadoras de Ni en el horizonte limonítico del yacimiento Moa”. Minería y Geología , Vol. XI (2):11-15.
- Rojas Purón A. L . 2001. “Evidencias a Favor de que la Goethita es la Principal Portadora de Níquel en los Horizontes Lateríticos de las Cortezas Ferroniquelíferas.” Minería y Geología , Vol. 18 (3-4): 21-31. Available: <http://www.ismm.edu.cu/revistamg> , [Consulted: February 8, 2019].
- Rojas Purón A .; Orozco Melgar G .; Vera Sardinias O.; Muñoz Gómez N. & Arderí A. 2005. “Caracterización mineralógica de los perfiles lateríticos del yacimiento Punta Gorda”. II Simposio Geología, Exploración y Explotación de las Lateritas Niquelíferas, I Congreso de Minera (MINERIA'2005), I Convención Cubana de Ciencias de la Tierra, La Habana, Cuba, ISBN: 959-7117-03-7.
- Tomás Antonio J., Peña Abreu, R. E . & Polanco Almanza R. G. 2018. “Comportamiento probabilístico espacial de la composición sustancial en yacimientos lateríticos del oriente cubano.” Vol. 34 (3): 251-264, ISSN: 1993-8012. <http://www.ismm.edu.cu/revistamg> , [Consulted: February 8, 2019].
- Shannon C. E. 1948. “A Mathematical Theory of Communication. Reprinted with corrections from The Bell System.” Technical Journal, July Vol. 27: 379-423, 623-656.
- Vera Sardiñas L. O.; Legrá Lobaina A. A. & Medina Arce M. 2001. “Principios básicos para la obtención de redes racionales en la exploración de los yacimientos lateríticos cubanos.” Revista Minería y Geología , Vol. XVIII (1):1-7, ISSN: 0258-5979. <http://www.ismm.edu.cu/revistamg> , [Consulted: February 8, 2019].
- Vera Sardiñas L. O . 2001. Procedimiento para la determinación de las redes racionales de exploración de los yacimientos lateríticos de níquel y cobalto en la región de Moa. Tesis Doctoral, Instituto Superior Minero Metalúrgico de Moa, 120 p.

Los autores de este trabajo declaran no presentar conflicto de intereses.

Este artículo se encuentra bajo licencia [Creative Commons Reconocimiento-NoComercial 4.0 Internacional \(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)