

Influencia de la exclusión de outliers en el proceso de interpolación por Kriging ordinario

Livia Maxnelly Caceres Beizaga

livia.caceres@unmsm.edu.pe

Cusco-Perú

RESUMEN

La detección de outliers y su tratamiento es necesario para evitar inexactitudes en los datos de salida de cobre y molibdeno producto de una interpolación por kriging ordinario en un yacimiento mineral. Se realizó un análisis de outliers y distancia de influencia a partir de los cuales se determinó que en el caso del cobre los valores anómalos a partir de los cuales se observa una mejor reconciliación con el valor de referencia varía entre 1.5 hasta 3 % de cobre con una distancia óptima de búsqueda entre 15 y 30 m obteniéndose un error de -0.3% en el mejor de los escenarios. En el caso de molibdeno los valores anómalos a partir de los cuales se observa una mejor reconciliación con el valor de referencia varía entre 0.005% hasta 0.01 % con una distancia óptima de búsqueda entre 15 m obteniéndose un error de 3%. Finalmente se determinó que el valor del outlier a excluir del proceso de interpolación está estrechamente relacionado con la distancia de búsqueda, pero esta relación es variable para cada elemento analizado por lo que con el fin de lograr un mejor ajuste en función al valor de referencia se debe realizar un análisis geoestadístico adecuado.

Palabras clave: outlier; interpolación; kriging ordinario; distancia de búsqueda; valor de referencia

Correspondencia: livia.caceres@unmsm.edu.pe

Artículo recibido: 28 abril 2022. Aceptado para publicación: 15 mayo 2022.

Conflictos de Interés: Ninguna que declarar

Todo el contenido de **Ciencia Latina Revista Científica Multidisciplinar**, publicados en este sitio están disponibles bajo

Licencia [Creative Commons](https://creativecommons.org/licenses/by/4.0/) 

Como citar: Caceres Beizaga, L. M. (2022). Influencia de la exclusión de outliers en el proceso de interpolación por Kriging ordinario. *Ciencia Latina Revista Científica Multidisciplinar*, 6(3), 1906-1919. DOI: https://doi.org/10.37811/cl_rcm.v6i3.2341

Influence of the exclusion of outliers in the interpolation process by ordinary Kriging

ABSTRACT

The detection of outliers and their treatment is necessary to avoid inaccuracies in the output data of copper and molybdenum resulting from interpolation by ordinary kriging in a mineral deposit. An analysis of outliers and influence distance was carried out, from which it was determined that in the case of copper, the anomalous values that present better reconciliation with the reference value varies between 1.5 to 3% of copper with an optimal search distance between 15 and 30 m, obtaining an error of -0.3% in the best case. For molybdenum, the anomalous values that shows better reconciliation with the reference value is between 0.005% and 0.01% with an optimal search distance of 15 m, obtaining an error of 3% . Finally, it will be determined that the value of the outlier to exclude from the interpolation process is closely related to the search distance, but this relationship is variable for each element analyzed, so in order to achieve a better fit based on the reference value we must perform a proper geostatistical analysis.

Keywords: *outlier; interpolation; ordinary kriging; search distance; reference value*

1. INTRODUCCIÓN

Durante el proceso de explotación de un yacimiento, es importante conocer cuál es la naturaleza de la información que se va generando a partir de los ensayos geoquímicos aplicados a las muestras de pozos de perforación de producción. Se sabe que el proceso de planificación de minado y la determinación de la vida útil de las minas se basa netamente en un modelo de recursos y reservas. Por otro lado durante el desarrollo de las operaciones mineras también se implementa modelos de corto plazo que son generados y actualizados constantemente.

El propósito de estos modelos es predecir a corto plazo el valor de cabeza de las leyes la cuales van a ser enviadas a planta de procesamiento, sin embargo frecuentemente existen desviaciones entre los valores predichos y los valores obtenidos, un punto a tomar en cuenta para la elaboración de los modelos de corto plazo son los llamados Outliers. De acuerdo a Hawkins (1980) los outliers son valores que se desvían mucho de otros valores observados y pueden despertar sospechas que fueron generados por mecanismos distintos, a pesar de que dichos valores pueden aparentar ser erróneos, también existe la posibilidad de que sean correctos.

Según Orellana y Cedillo (2020) existen infinidad de metodologías para detectar valores anómalos y plantean una serie de procedimientos para encontrar valores atípicos en una base de datos nominal (conjunto de datos relacionados a temas financieros, esta propuesta se divide en tres fases, la primera es el uso del método estadístico chi-cuadrado para definir una ponderación o peso W que indica la relevancia de la columna del perfil del cliente en comparación con la cantidad de crédito otorgado, en la siguiente fase, se utilizó el algoritmo de agrupación denominado KNN Global Anomaly (KNN) que calcula la puntuación de los valores atípicos y finalmente, el algoritmo K-Means que es utilizado para separar los valores atípicos de los valores falsos-positivos.

Según Knorr (2002) Un outlier es cualquier dato que parece estar fuera de lugar respecto al resto de los datos, es importante realizar el constante análisis de calidad de los datos ya que datos poco confiables podrían llevar la toma de decisiones erróneas, pérdida de dinero, tiempo y credibilidad. Da vila (2019) menciona que existen varios métodos para la detección de valores anómalos en grandes bases de datos como los mecanismos iterativos y otros sin embargo la presente investigación está centrada en la identificación de posibles valores anómalos de forma sencilla y visual mediante ploteos en gráficos de

puntos y diagrama de cajas; y los efectos de la diferenciación de estos durante el proceso de interpolación mediante el método de Kriging ordinario.

En caso de producirse outliers, y no tratarlos adecuadamente a la hora de realizar interpolaciones y/o predicciones se puede llegar a inexactitudes en la predicción puntual (Catalan y Trivez, 2006)

2. ESTRATEGIAS METODOLÓGICAS O MATERIALES Y MÉTODOS

2.1 Análisis estadístico de los datos

Como primer punto se realizó un análisis estadístico de las muestras con datos geoquímicos disponibles a partir de pozos de producción existentes, se utilizó la estadística descriptiva para la revisión de valores máximos y mínimos de Cu% y Mo%.

2.2 Identificación de posibles outliers.

El segundo punto consiste en la identificación de posibles valores que se puede considerar como valores anómalos mediante la utilización de diagrama de cajas o boxplots, Según Castellanos (2012) estos gráficos identifican la distribución global de la población de datos analizada, así como los posibles valores anómalos.

La ventaja de la utilización de estos tipos de gráficos es que muestran un resumen de una cantidad considerable de datos en este caso Cu% y Mo% en cinco medidas descriptivas además de su morfología y simetría.

2.3 Aplicación de la diferenciación de los outliers durante la interpolación por Kriging Ordinario.

Una vez identificados los posibles datos anómalos, se procedió a la aplicación de exclusión del proceso de interpolación mediante Kriging Ordinario, con la finalidad de ver la variación de la media de los datos procedentes de la estimación. El proceso de interpolación se realizó mediante software Minesight.

La exclusión del proceso de interpolación, también está estrechamente relacionado a la distancia a partir de la cual ya no se va a considerar el valor anómalo en cuestión. Una vez realizada la interpolación se procedió a obtener la media y compararla con el dato real de planta para identificar el valor correcto de outlier aplicable durante el proceso de estimación.

3. RESULTADOS Y DISCUSIÓN

3.1 Análisis estadístico general

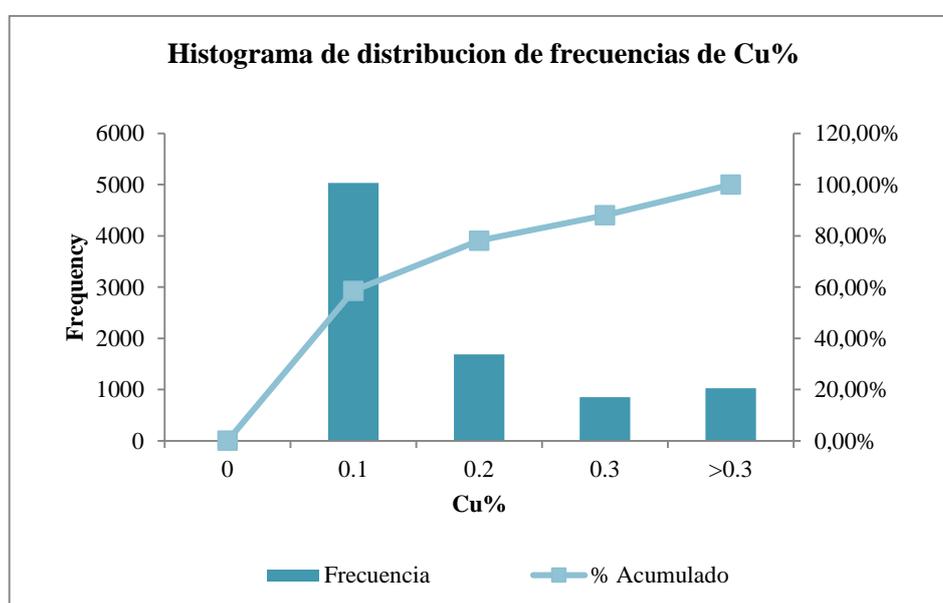
Para la realización del análisis estadístico se utilizaron un total de 8600 muestras de Cobre y 8577 muestras de Mo%, el resultado del análisis se muestra en la tabla 1.

Tabla 1. Estadísticos descriptivos de Cu% y Mo%

<i>Estadística descriptiva</i>	<i>Cu%</i>	<i>Mo%</i>
Media	0.14	0.0053
Error estandar	0.00	0.0002
Mediana	0.07	0.0010
Moda	0.01	0.0005
Desviacion estandar	0.21	0.0226
Varianza de muestras	0.04	0.0005
Kurtosis	19.26	5649.78
Skewness	3.62	68.192
Rango	2.29	1.892
Minimum	0.00	0.00
Maximum	2.29	1.8920
Count	8600	8577

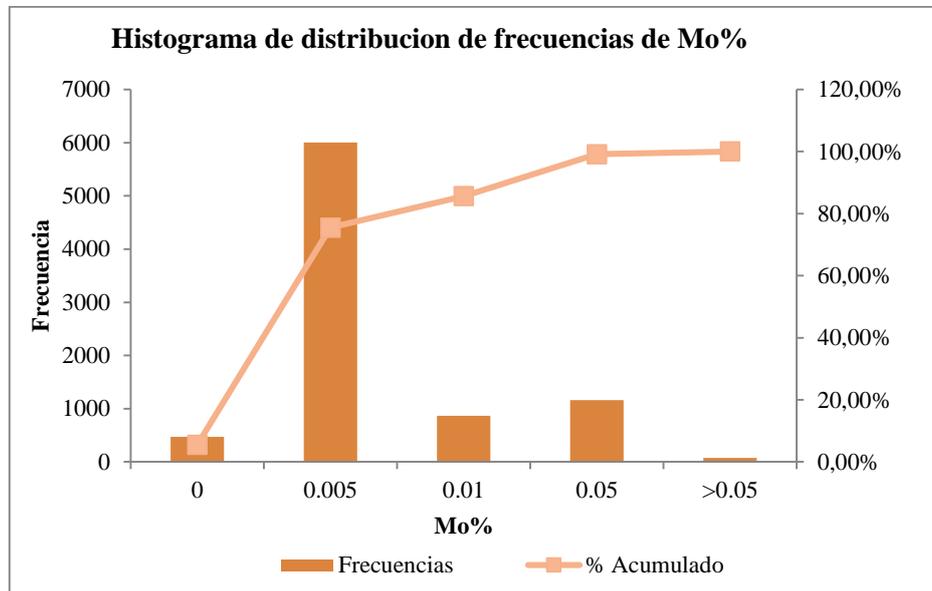
A partir de las 8600 y 8577 muestras se realizó dos histogramas de distribución de Cu% y Mo% que se presentan en las figuras 1 y 2 respectivamente. La distribución de Cu% que se observa en la figura 1 muestra que el 60% de la data procesada supera el 0.1% de cobre a partir del cual se considera como mineral, por lo que la población sobre la cual se hará el análisis de outlier corresponde a zonas económicamente rentables.

Figura 1. Histograma de distribución de Cu% en muestras



Con respecto a los valores de Molibdeno, el 75 % de las muestras analizadas superan 0.005% de molibdeno, por lo que se centrara el análisis de outliers en esa población.

Figura 2. Histograma de distribución de Mo% en muestras

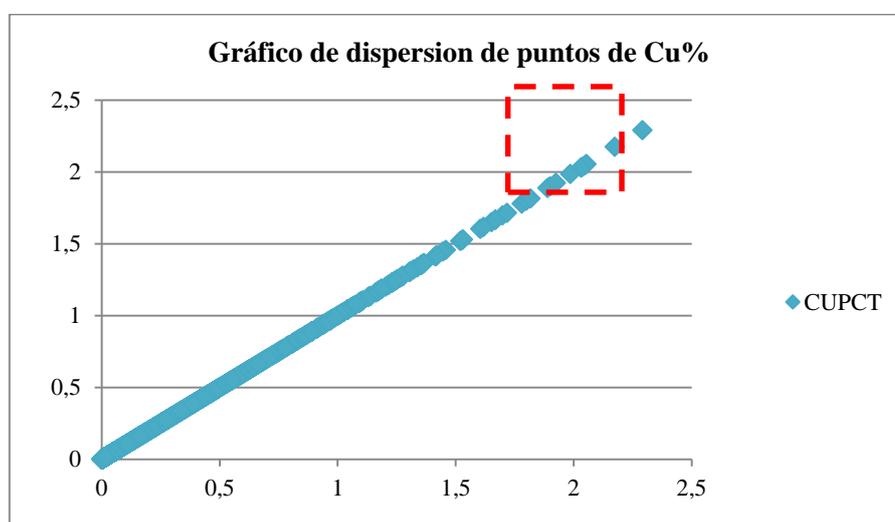


3.2 Identificación de outliers

La identificación de outliers se realizó mediante el gráfico de dispersión de puntos así como el diagrama de cajas, además se realizó el cálculo de los cuartiles correspondientes a cada elemento y la cantidad de outliers detectados.

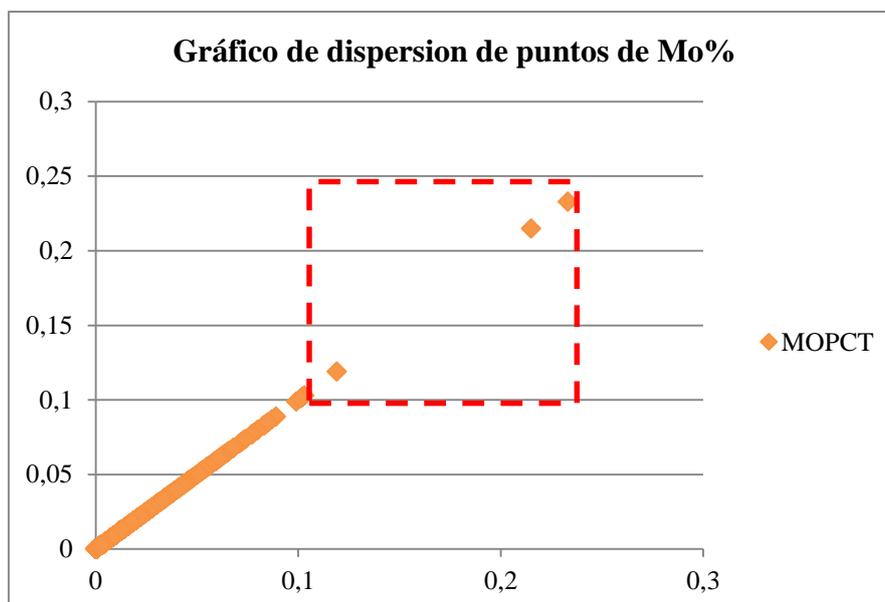
La figura 3 muestra la localización de los posibles outliers máximos encontrados en las muestras de Cu%, se observa que los posibles valores anómalos corresponden a 2.175 ya que a partir de este rango se podría considerar fuera de la distribución normal de la población de muestras.

Figura 3. Histograma de distribución de Zn% en muestras



La figura 4 muestra la localización de los posibles outliers máximos encontrados en las muestras de Mo%, se observa que los posibles valores anómalos corresponden a 0.119 ya que a partir de este rango se podría considerar fuera de la distribución normal de la población de muestras.

Figura 4. Histograma de distribución de Zn% en muestras



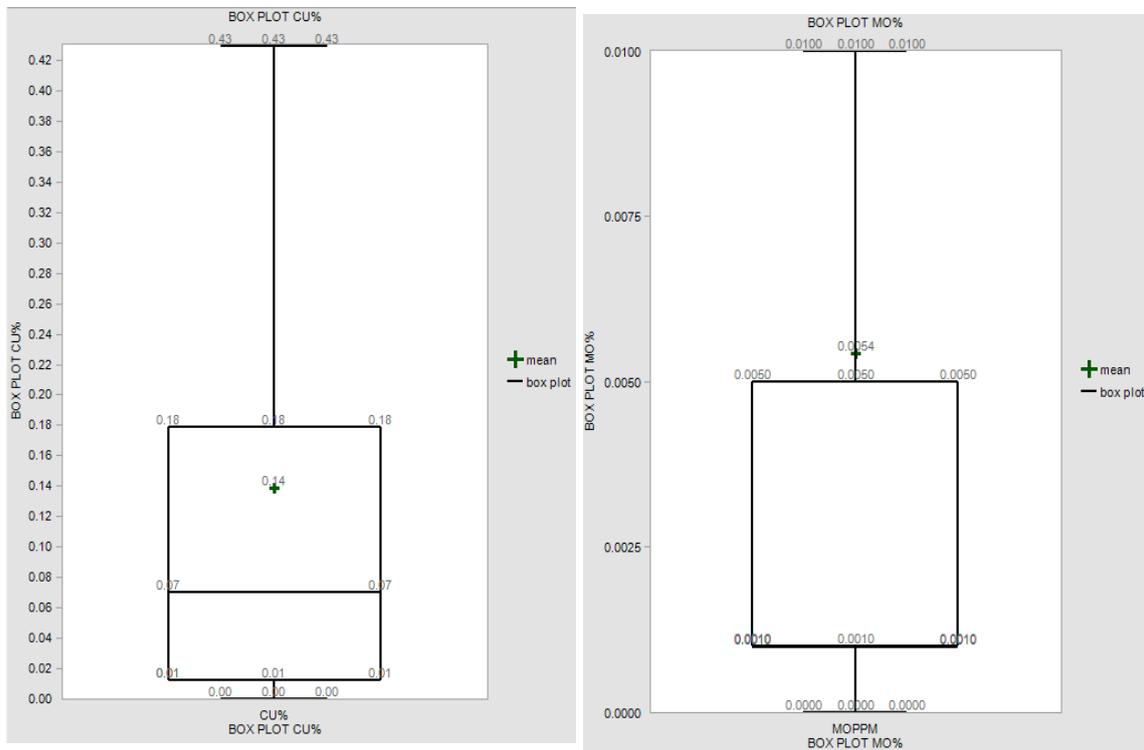
La tabla 2 muestra el análisis estadístico de dispersión y outliers realizado para las muestras de Cu % y Mo% del análisis se observa que el valor máximo sin valor atípico es 0.429 para Cu% y 0.01 para Mo%, también se incluye la cantidad de valores atípicos o llamados outliers leves y extremos.

Tabla 2. Análisis estadístico de outliers para Cu% y Mo%

Estadístico	Cu%	Mo%
Numero de muestras	8600	8577
Minimo	0	0
Maximo	2.291	1.892
Promedio	0.13878	0.0054
Primer cuartil	0.012	0.001
Mediana	0.07	0.001
Tercer cuartil	0.179	0.005
Desviacion estandar	0.20656	0.0226
Varianza	0.04267	0.0005
Valor minimo sin valor atípico	0	0
Valor máximo sin valor atípico	0.429	0.01
Numero de valores atípicos leves	331	544
Numero de valores atípicos extremos	238	689

La figura 5 muestra los diagramas de caja para Cu% y Mo% en los cuales se observa la distribución de las muestras analizadas, los gráficos principalmente destacan la posición del primer cuartil que equivale a 0.012 y 0.001 para Cu% y Mo% respectivamente, la media de los datos es decir el segundo cuartil con valores de 0.07 y 0.001 así como el tercer cuartil que representa que el 75% de los datos son menores o iguales a 0.179 en el caso de Cu% y 0.005 en el caso de Mo%.

Figura 5. Diagrama de cajas de muestras de Cobre y Molibdeno



3.3 Exclusión de outliers en la interpolación mediante Kriging ordinario.

La exclusión de los valores atípicos durante el proceso de interpolación para el cobre se muestra en la tabla 3, la interpolación se realizó considerando dos variables, los valores de outliers y la distancia de búsqueda a partir de la cual se realizara la exclusión de los outliers con la finalidad de ajustar el valor obtenido a un valor de referencia.

Tabla 3. Resultados de la aplicación de exclusión de valores atípicos durante el proceso de interpolación de Cu% por Kriging ordinario.

Valores de Outlier de Cu%	Distancia de Cu%	Valores de Cu% interpolado	Valor de referencia	Porcentaje de error
1.00	20	0.2721	0.277	-1.8%
0.50	20	0.2674	0.277	-3.6%
1.50	20	0.2744	0.277	-0.9%
2.00	20	0.2758	0.277	-0.4%
2.50	20	0.2763	0.277	-0.3%
2.50	15	0.2763	0.277	-0.3%
2.50	30	0.2763	0.277	-0.3%
3.00	30	0.2763	0.277	-0.3%
0.40	10	0.1909	0.277	-45.1%
0.50	10	0.2114	0.277	-31.0%
0.75	15	0.2544	0.277	-8.9%
1.00	15	0.2624	0.277	-5.6%
0.75	20	0.2700	0.277	-2.6%

Las figura 6 muestra la distribución de los valores de cobre obtenidos a partir de la interpolación en función de la distancia de búsqueda de outlier así como una línea de referencia a partir de la cual se obtendrá el porcentaje de error para cada punto de estimación. La figura 7 muestra la distribución de los valores obtenidos a partir de la interpolación en función del valor de outlier también presenta una línea de referencia a partir de la cual se obtendrá el porcentaje de error.

Figura 6. Valores interpolados en función de la distancia de búsqueda de outlier para Cobre

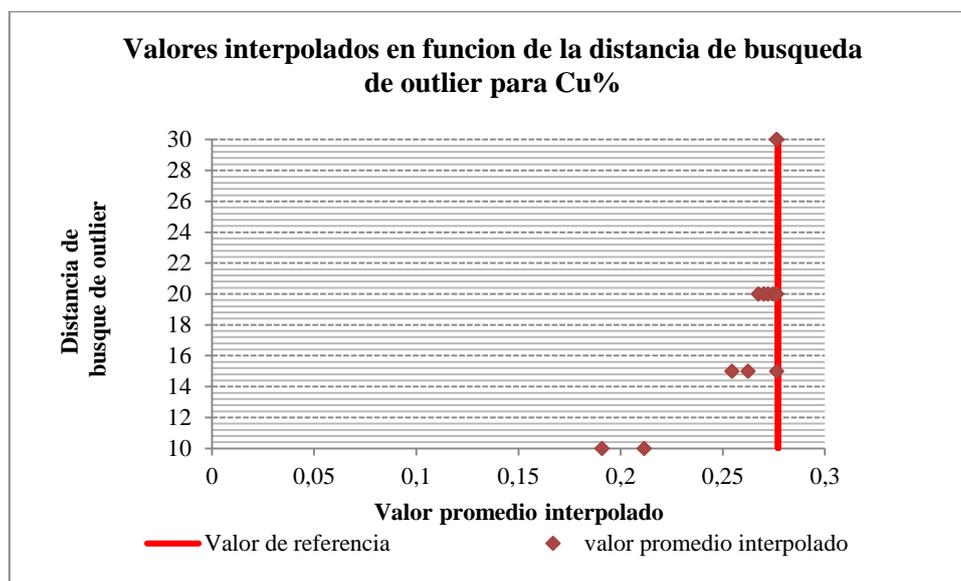
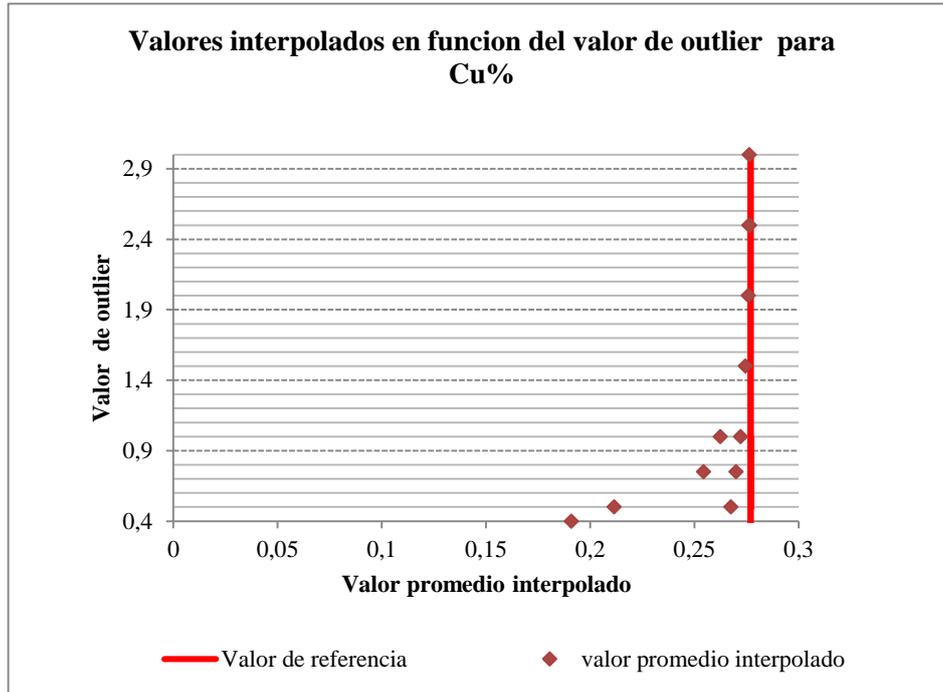


Figura 7. Valores interpolados en función del valor de outlier para cobre.



La exclusión de los valores atípicos durante el proceso de interpolación para el molibdeno se muestra en la tabla 4, la interpolación se realizó considerando dos variables, los valores de outliers y la distancia de búsqueda a partir de la cual se realizara la exclusión de los outliers con la finalidad de ajustar el valor obtenido a un valor de referencia

Tabla 4. Resultados de la aplicación de exclusión de valores atípicos durante el proceso de interpolación de Mo% por Kriging ordinario.

Valores de Outlier de Mo%	Distancia de Mo%	Valores de Mo% interpolado	Valor de referencia	Porcentaje de error
0.030	15	0.0109	0.0095	12.96%
0.040	15	0.0113	0.0095	15.93%
0.025	15	0.1070	0.0095	91.12%
0.010	15	0.0100	0.0095	5.00%
0.005	15	0.0098	0.0095	3.06%
0.003	10	0.0061	0.0095	-55.74%
0.003	20	0.1105	0.0095	91.40%
0.004	20	0.1115	0.0095	91.48%
0.004	10	0.0061	0.0095	-55.74%
0.004	20	0.1115	0.0095	91.48%
0.004	15	0.0098	0.0095	3.06%
0.003	15	0.0099	0.0095	4.04%
0.010	15	0.0099	0.0095	4.04%

Las figura 8 muestra la distribución de los valores de molibdeno obtenidos a partir de la interpolacion en función de la distancia de búsqueda de outlier así como una línea de referencia a partir de la cual se obtendrá el porcentaje de error para cada punto de estimación. La figura 9 muestra la distribución de los valores obtenidos a partir de la interpolacion en función del valor de outlier también presenta una línea de referencia a partir de la cual se obtendrá el porcentaje de error.

Figura 8. Valores interpolados en función de la distancia de búsqueda de outlier para Molibdeno

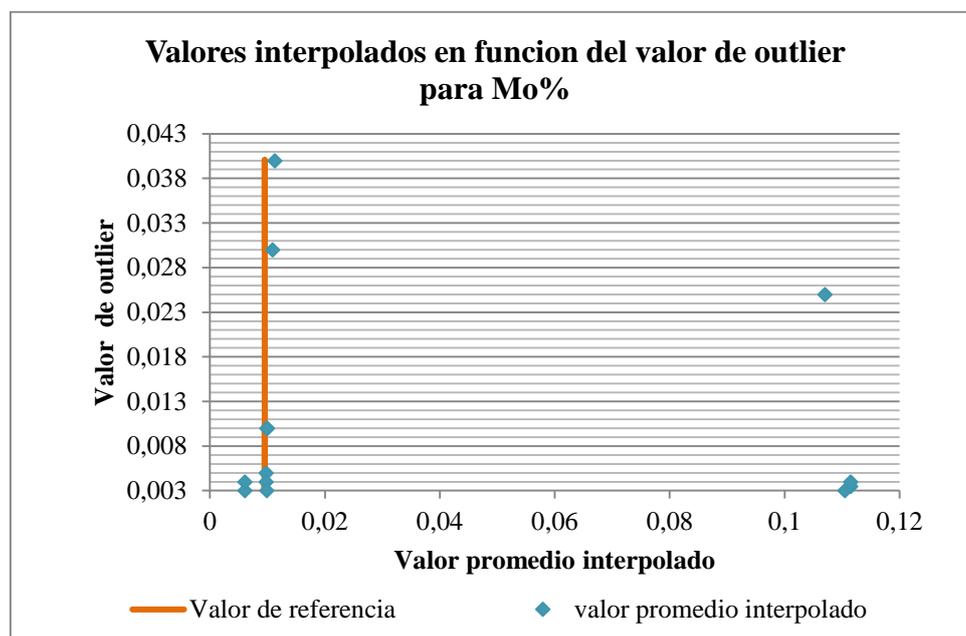
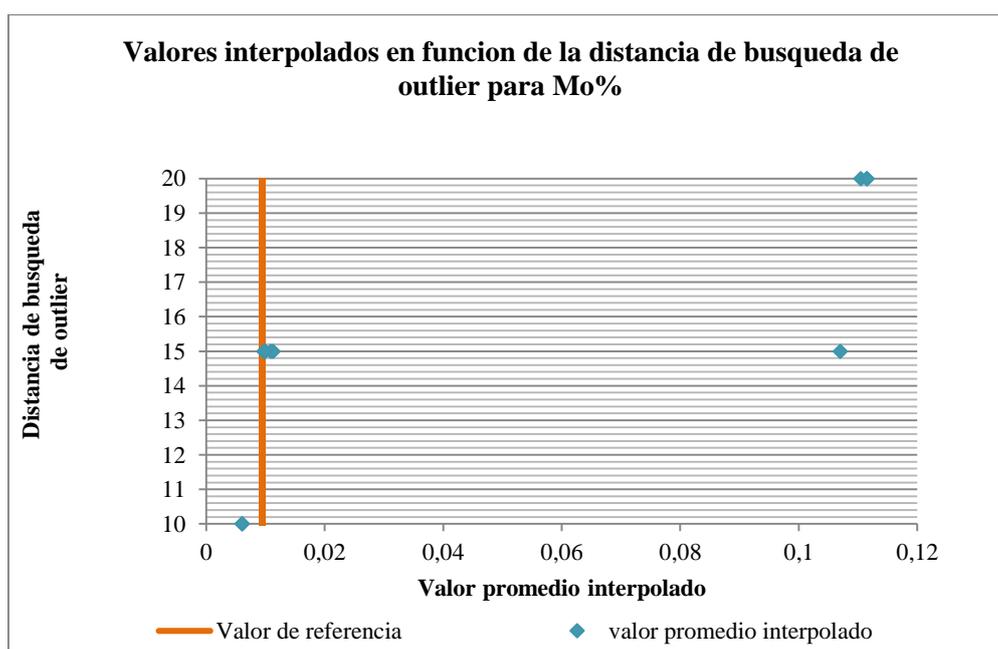


Figura 9. Valores interpolado en función del valor de outlier para Molibdeno



DISCUSIÓN

Del análisis en función de la distancia de los resultados de la interpolación de cobre realizando la exclusión de valores anómalos se observa que existe un menor porcentaje de error para distancias mayores es decir para una distancia de búsqueda de 30 m el porcentaje de error es -0.3% aproximadamente. Por el contrario para distancias de búsqueda de 10 m el porcentaje de error el porcentaje de error es oscila entre -31% y -45% en negativo es decir el valor interpolado se aleja muy por debajo del valor de referencia.

En el caso del análisis en función del valor de outlier el error es mayor es decir entre -31% y 45% para valores menores como 0.4 y 0.5 , a medida que se incrementa el valor de outlier el error disminuye por ejemplo para valores a partir de 1.5 el error oscila entre -0.9% y -0.3% usando 3 como valor extremo.

En el caso de la interpolación de molibdeno el análisis en función de la distancia de búsqueda para exclusión de outliers muestra que existe un mayor porcentaje de error para distancias mayores es decir para una distancia de búsqueda de 20 m el porcentaje de error es 90% aproximadamente. Por el contrario para distancias de búsqueda de 15 m el porcentaje de error disminuye entre 3 y 4%. Sin embargo se observa que para una distancia de búsqueda de 10m el porcentaje de error es 55% en negativo es decir el valor interpolado se aleja muy por debajo del valor de referencia.

Para el caso de los valores de molibdeno usados como outliers, se observa que existe un mayor rango de error para valores muy bajos, es decir entre 0.003 y 0.004 corresponde entre -55 y 91%, lo mismo sucede para los valores anómalos muy altos como 0.025 y 0.04 con errores de 91% y 15% respectivamente. Sin embargo los valores intermedios muestran un mejor comportamiento ya que para valores entre 0.005 y 0.01 corresponde 3% y 5% de error.

4. CONCLUSIÓN O CONSIDERACIONES FINALES

Para el caso de la interpolación de cobre los valores anómalos a partir de los cuales se observa una mejor reconciliación con el valor de referencia varía entre 1.5 hasta 3 % de cobre con una distancia óptima de búsqueda entre 15 y 30 m obteniéndose un error de -0.3% en el mejor de los escenarios.

Para el caso de la interpolación de molibdeno los valores anómalos a partir de los cuales se observa una mejor reconciliación con el valor de referencia varía entre 0.005% hasta

0.01 % con una distancia optima de búsqueda entre 15 m obteniéndose un error de 3% en el mejor de los escenarios.

Se identificó que el valor del outlier a excluir del proceso de interpolación está estrechamente relacionado con la distancia de búsqueda, asimismo esta relación varía de un elemento a otro, por lo que es necesario seleccionar adecuadamente ambos parámetros para lograr un mejor ajuste en función al valor de referencia.

5. LISTA DE REFERENCIAS

- Buenaño López, A. R. (2017). *Mapa estratégico de contaminación como herramienta para el trazado de alternativas de ciclo vías con menor exposición* (Bachelor's thesis, Quito: Universidad de las Américas, 2017).
- Córdoba, M., Paccioretti, P. A., Giannini Kurina, F., Bruno, C. I., & Balzarini, M. G. (2019). *Guía para el análisis de datos espaciales en agricultura*.
- Corea, F. V. G. (2014). *Predicción espacio-temporal de la irradiancia solar global a corto plazo en España mediante geoestadística y redes neuronales artificiales* (Doctoral dissertation, Universidad Politécnica de Madrid).
- Da Vila, S. (2019). *Detección de outliers en grandes bases de datos*. Máster en Técnicas Estadísticas. Universidad da Coruña.
- Debelis, S. P. (2017). *La variabilidad espacial en tierras hidrohalomórficas* (Master's thesis).
- Delgado, A. M. S. *Diseño de Redes de Muestreo Óptimas para el Monitoreo del Carbono Orgánico del Suelo en el CI La Libertad mediante la Aplicación de Algoritmos Genéticos*.
- Duque Gardezabal, N. (2018). *Estimación de campos de precipitación en cuencas hidrográficas colombianas con escasez de datos, combinando datos teledetectados y de estaciones en tierra, utilizando funciones de Kernel*. *Ingeniería Civil*.
- Estrella Salvador, J. R. (2020). *Geoid undulation models for Madrid topographic network through geostatistical techniques designed from free and commercial software*.
- Fernández Palomino, C. A. (2015). *Frecuencia de precipitaciones máximas según los L-momentos en la cuenca hidrográfica del lago Titicaca*.

- Fernández Palomino, C., & Lavado-Casimiro, W. (2014). Análisis regional de precipitaciones máximas en las cuencas Chillón, Rímac, Lurín y parte alta del Mantaro.
- Grisales Vargas, S. C. (2020). Modelos de regresión de usos del suelo para la caracterización espacial de la contaminación del aire por PM_{2.5} en la ciudad de Medellín-Colombia, 2018.
- Gutierrez Corea, F. V. (2014). *Predicción espacio-temporal de la irradiación solar global a corto plazo en España mediante geoestadística y redes neuronales artificiales* (Doctoral dissertation, Topografía).
- Hawkins, D. (1980). *Identification of Outliers*. London, Chapman & Hall.
- Moreno, J. (2012). *Método de detección temprana de outliers*. Pontificia Universidad Javeriana
- Orellana M. y Cedillo, P. (2020). Outlier detection with data mining techniques and statistical methods. *Enfoque UTE*, 11(1), pp. 56 - 67. <https://doi.org/10.29019/enfoque.v11n1.584>
- Salgado, C. A., & Ocampo, J. L. (2018). IMPUTACION DE DATOS FALTANTES DE TEMPERATURA MAXIMA MEDIA MENSUAL MEDIANTE METODOS GEOESTADÍSTICOS EN ESTACIONES CLIMATICAS DEL VALLE DEL CAUCA EN EL PERIODO 2013-2014.
- Santacruz Delgado, A. M. (2010). Diseño de redes de muestreo óptimas para el monitoreo del carbono orgánico del suelo en el CI La Libertad mediante la aplicación de algoritmos genéticos. *Facultad de Agronomía*.
- Soberón Rafael, A. (2017). Estimación de recursos minerales aplicando el paquete de variografía comercial Sage 2001, del proyecto Azod-Cusco, 2017.
- Tolentino Barrios, F. A. (2019). Estimación de recursos y reservas con el uso de software minero para la explotación del proyecto minero-Don Javier.
- Trávez, F. & Catalán, B. (2006). *Efectos de los Outliers aditivos en la predicción de la varianza condicional de un modelo Arch*. *Estudios de Economía Aplicada*, 24(1), 531-543. [fecha de Consulta 24 de Mayo de 2022]. ISSN: 1133-3197. Disponible en: <https://www.redalyc.org/articulo.oa?id=30113179021>