

UNIVERSIDAD NACIONAL DE COLOMBIA SEDE MEDELLÍN
FACULTAD DE MINAS

“Análisis de distribución de datos y anomalías geoquímicas de cobre (Cu) y molibdeno (Mo) mediante el uso de Python 3 en el Distrito de Montebello, sur de Antioquia.”

EDIER VICENTE ARISTIZABAL GIRALDO

Profesor Asociado Universidad Nacional de Colombia sede Medellín
evaristizabalg@unal.edu.co

RICARDO TENJO SARMIENTO

Geólogo
Estudiante de Maestría en Ingeniería de Recursos Minerales
jrtensos@unal.edu.co

RESUMEN

En el presente trabajo se muestra el tratamiento dado mediante las herramientas de Python en el ambiente de Google Colab a un conjunto de 351 datos de tenores de cobre y molibdeno, con el fin de identificar su distribución espacial, realizar una interpolación estadística mediante Kriging Ordinario (KO), identificar posibles zonas geoquímicamente anómalas y establecer una correlación espacial entre dichas variables. Los resultados indicaron concentraciones promedio de cobre de 55,4 ppm en el área de estudio, las cuales están por encima del promedio de cobre en la corteza terrestre continental, el cual equivale a 28 ppm. Así mismo, permitieron identificar una zona de concentración anómala de cobre en la vereda Las Mercedes, del municipio de Santa Bárbara, con un tenor de hasta 885 ppm. Por su parte, el molibdeno no se presenta en concentraciones anómalas en el área de estudio, su concentración promedio fue de 1,07 ppm, mientras que el promedio para las rocas de la corteza superior es de 1,1 ppm. Por último, se halló que no existe una dependencia espacial entre las variables cobre y molibdeno.

INTRODUCCIÓN

Desde la década de 1950, el Servicio Geológico Nacional (hoy Servicio Geológico Colombiano) ha venido desarrollando estudios geológicos y mineros para determinar la naturaleza y origen de algunas ocurrencias de oro y elementos metálicos asociados presentes en de los municipios de Montebello, Fredonia, Santa Bárbara, al sur del Departamento de Antioquia. En tales estudios fueron identificados varios prospectos de mineralizaciones de origen hidrotermal, específicamente al sur del municipio de Santa

1. DATOS

2. ÁREA DE ESTUDIO

[illegible]

2

3. METODOLOGÍA

La metodología empleada consistió en una limpieza y cargue de datos en la aplicación Google Colab, un análisis de puntos, un análisis de ráster y una regresión lineal de los datos. A continuación, se describe el procedimiento implementado:

3.1. Limpieza y Cargue de Datos

3.1.1. Preparación de los Datos (Data Cleaning)

En este primer paso se seleccionaron los elementos de interés cobre (Cu) y molibdeno (Mo) para ser objeto del posterior tratamiento. Se preparó una tabla de Excel con los valores de concentraciones para cada elemento en formato Número, se eliminaron los espacios vacíos y se estableció como símbolo decimal la coma (,).

	A	B	C	D	E
1	Coord_X	Coord_Y	Cu_ppm	Mo_ppm	
2	-75,46305031	5,5870477	91,00	0,22	
3	-75,48556546	5,58808818	17,20	4,12	
4	-75,48205784	5,58908193	4,50	0,31	
5	-75,49263627	5,59023026	49,30	0,13	
6	-75,47704568	5,59106463	252,00	0,69	
7	-75,47600822	5,59496341	1,90	0,44	
8	-75,47603076	5,5968618	81,80	0,07	
9	-75,47324272	5,59692266	285,00	0,33	
10	-75,46544512	5,60013232	79,70	0,12	

Figura 2 Primeras 10 líneas de la tabla de datos utilizada.

3.1.2. Montaje de los Datos en Google Colab

Se importó la librería “pandas” de Python en la aplicación Google Colab, la cual permite leer archivos tipo ‘xlsx’, se crearon los objetos “Coord_X” y “Coord_Y” en el dataframe, conteniendo los valores de longitud y latitud, luego se corroboró que la tabla de datos fuese cargada correctamente usando la función “.head()”

3.1.3. Montaje del Polígono de Estudio

Se instaló la librería “geopandas” para leer el archivo shapefile que representa el área de estudio, denominada Distrito Montebello.

3.2. Análisis de Puntos

3.2.1. Cálculo de Estadísticas Básicas

Se usó la función “.describe” para obtener las estadísticas descriptivas de las columnas de datos “Cu_ppm” y “Mo_ppm”. Así mismo se elaboró un histograma para ver la distribución de los datos.

3.2.2. Distribución de los Datos

Se importaron las librerías “seaborn”, “contextily”, “pointpats” y “matplotlib”. Se elaboró un histograma con 90 bins o intervalos para ver la distribución de los datos. Luego se plotearon las muestras en un mapa fijando los límites del campo a representar mediante un “X_min, X_max”, y un “Y_min, Y_max”, se crearon objetos denominados “joint_axes”, se agregó un mapa base Toner Lite proporcionado por Stamen Design y se agregó una flecha que señala el norte mediante la función “ax.arrow”. Finalmente se plotó el gráfico utilizando la función “plt.imshow”. Adicionalmente se construyó el gráfico de conteo de puntos por cuadrantes usando los módulos “Qstatistic y PointPattern”.

3.2.3. Centrografía

Se importó la librería “numpy”, y el módulo “centrography” de “pointpats”. En la gráfica de distribución de los datos se representó el promedio y la mediana, luego se agregó la elipse estándar de los datos utilizando “matplotlib”. Posteriormente, se convirtieron los datos en un arreglo numpy, se seleccionaron 3 grupos para “K” y se aplicó el algoritmo K-means. Este fue repetido con 5 grupos y finalmente se obtuvieron los centroides de los grupos y sus respectivas etiquetas. Se calcularon los centros promedio y medianos, y se generó el gráfico de dispersión con su respectiva leyenda y norte.

3.2.4. Relación de Puntos y Polígonos

Usando las librerías “geopandas”, “matplotlib” y “contextily” se creó un mapa de calor que representa la cantidad de datos (puntos) y las veredas (polígonos) que componen la zona de estudio. Para este análisis se cargó un shapefile de las veredas del sur de Antioquia. Se agregó el recuento de los puntos, se filtraron las veredas con recuento de puntos mayor a cero y se obtuvo el nombre (etiqueta) de la vereda con mayor cantidad de puntos. Posteriormente se realizó el mismo ejercicio pero con base en el valor de concentración de cobre (Cu_ppm) según las veredas.

3.3. Análisis de Ráster

3.3.1. Modelo de Elevación Digital

Se importó la imagen satelital ALOS-PALSAR con código ALPSRP261120100 desde la plataforma ASF Data Search Vertex que a su vez hace parte de los datos abiertos de la Earth Information Center de la NASA. Para esto se instaló la librería “rasterio”. Se cargó y se representó usando mapa de colores Spectral. Posteriormente se instaló la librería EarthPy y se instaló el módulo “spatial”, el cual permitió elaborar un mapa de sombras a partir de la imagen de dem cargada.

3.3.2. Análisis de Imágenes Satelitales

3.3.3. Interpolación Estadística

Inicialmente se instalaron las librerías “geostatspy” y “scikit-gstat”. Se calculó una matriz de distancias entre los puntos usando la librería “numpy”. Así mismo, se convirtieron los datos de coordenadas X-Y y las concentraciones de cobre (Cu) y molibdeno (Mo) a arreglos matriciales. Se creó y calculó un objeto denominado variograma a partir de semivarianzas, usando la función “variogram.fit()”, seguidamente se obtuvieron los parámetros de dicho variograma y se creó un modelo regresor del proceso gaussiano de los datos. Luego se definió una malla de interpolación, se realizó kriging ordinario (KO), se calculó el error de la estimación y se generó un mapa de isovalores para las concentraciones de cobre y molibdeno.

3.4. Regresión Lineal

En este apartado se intentó determinar la relación entre las variables de cobre (Cu) y molibdeno (Mo). Para esto se importó la librería “statsmodels.api”. Se designó la variable cobre (Cu) como independiente y la variable molibdeno como (dependiente) y se procedió a ejecutar un modelo de regresión lineal de mínimos cuadrados ordinarios (OLS). Se imprimieron los resultados del modelo, se calcularon los residuos y por último fueron graficados.

4. RESULTADOS

4.1. Análisis de Puntos

Se identificó que los datos tienen una distribución de cola pesada, caracterizada altas frecuencias en los valores iniciales, que disminuyen gradualmente a medida que los valores de cobre (Cu) aumenta.

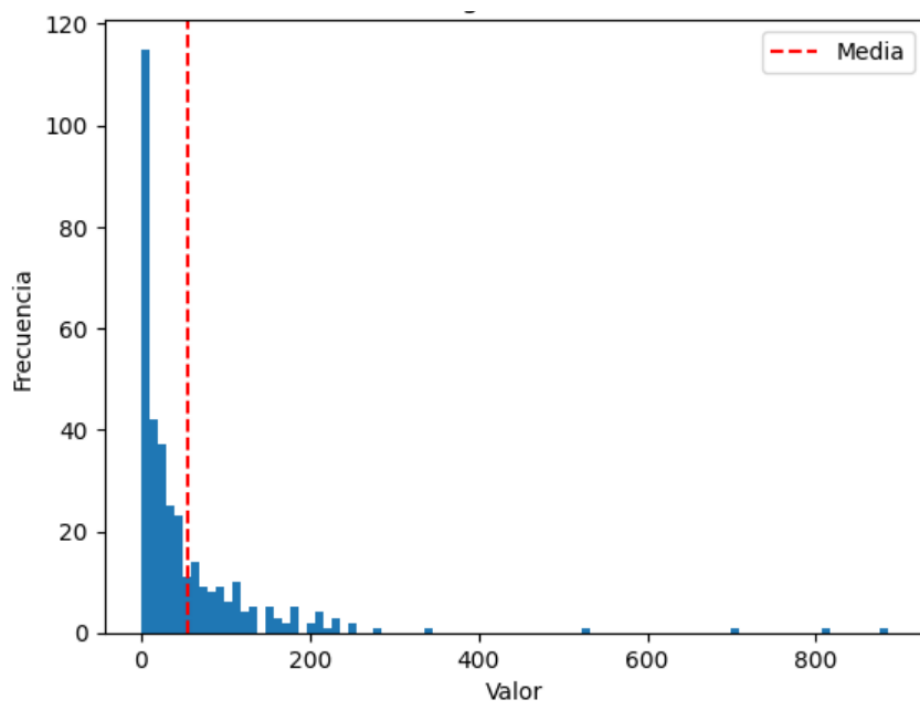


Figura 3 Histograma de los valores de cobre (Cu_ppm).

La concentración promedio de cobre en el Distrito de Montebello (54.5 ppm) se encuentra enriquecida dos veces (54,54) respecto al promedio de cobre en la corteza continental terrestre (28 ppm) ^[1]. Por su parte, el molibdeno se encuentra en un rango de concentración promedio respecto a este mismo parámetro (1,1 ppm).

Tabla 1 Estadísticas Descriptivas de las variables cobre (Cu) y molibdeno (Mo).

Parámetro	Cu ppm	Mo ppm
Count	351	351
Mean	54,54	1,07
Std Dev	94,43	2,72
Min	0,43	0,03
25%	5,24	0,19
50%	24,80	0,41
75%	65,92	0,83
max	885,80	35,70

La visualización de los datos permite identificar que hacia la región central del área de estudio se localiza el centro promedio y el centro mediano.

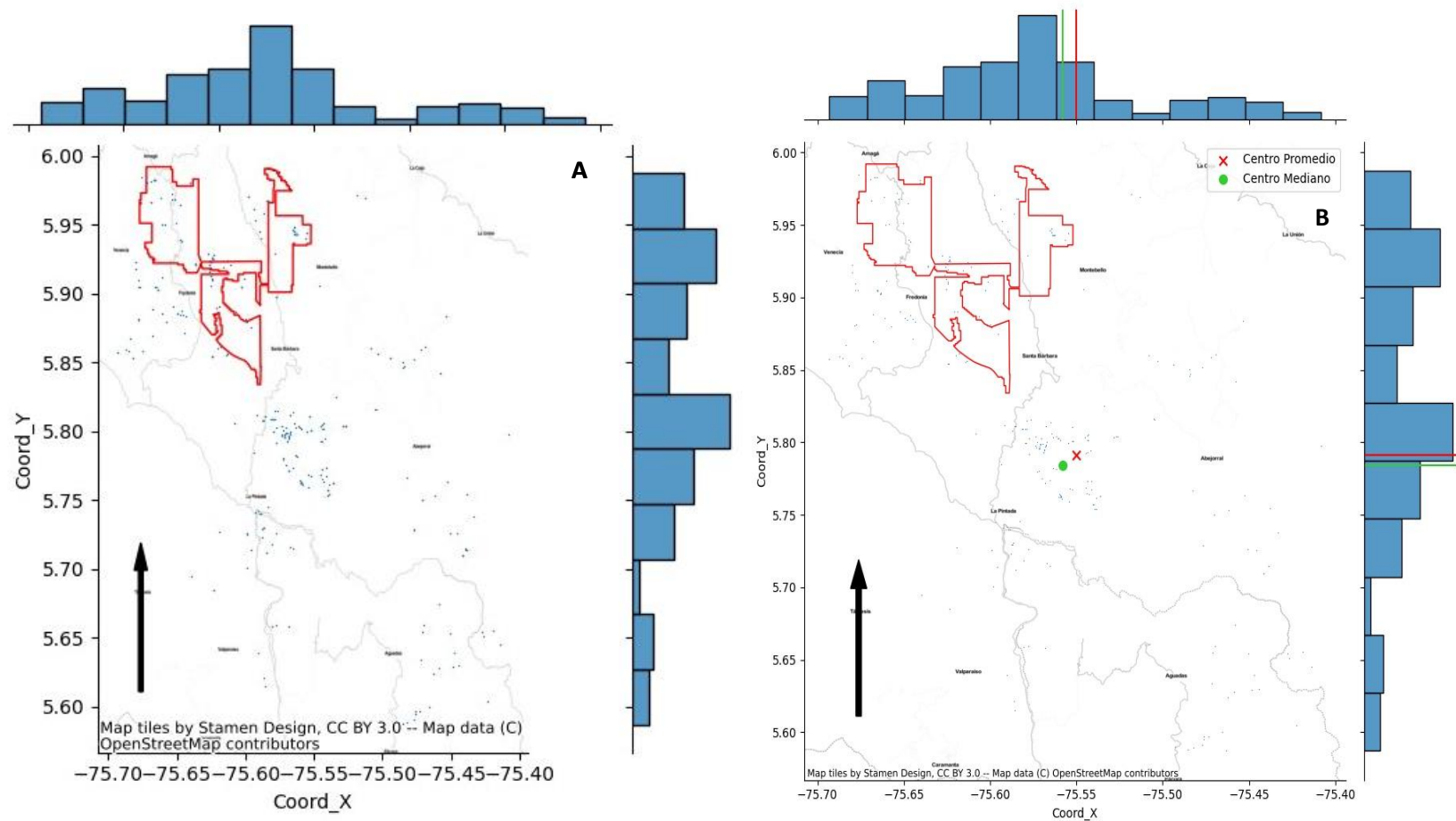


Figura 4 Visualización (A) y tendencias (B) en la distribución de las muestras de roca analizadas.

Mediante el análisis de cuatro (4) cuadrantes se evidenció que los datos se encuentran mayormente localizados en el cuadrante 1, noroccidental. Mientras que la elipse estándar indica una tendencia de distribución central con dirección noreste.

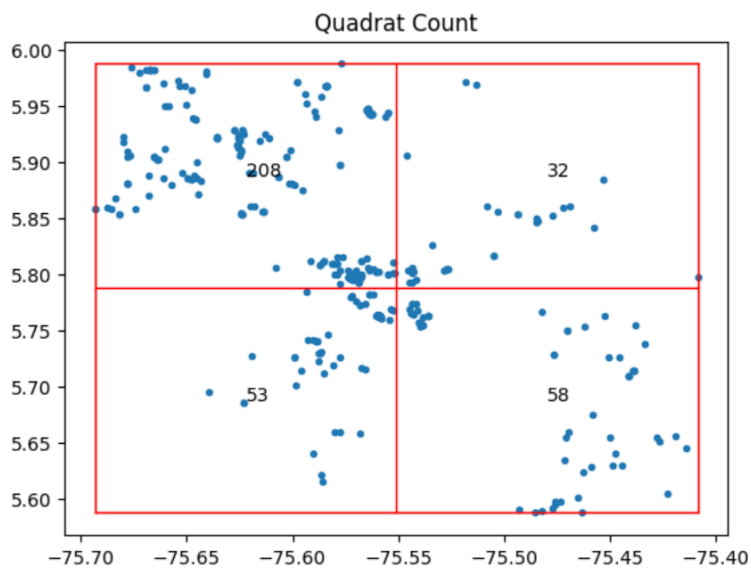


Figura 5 Conteo de muestras por cuadrantes.

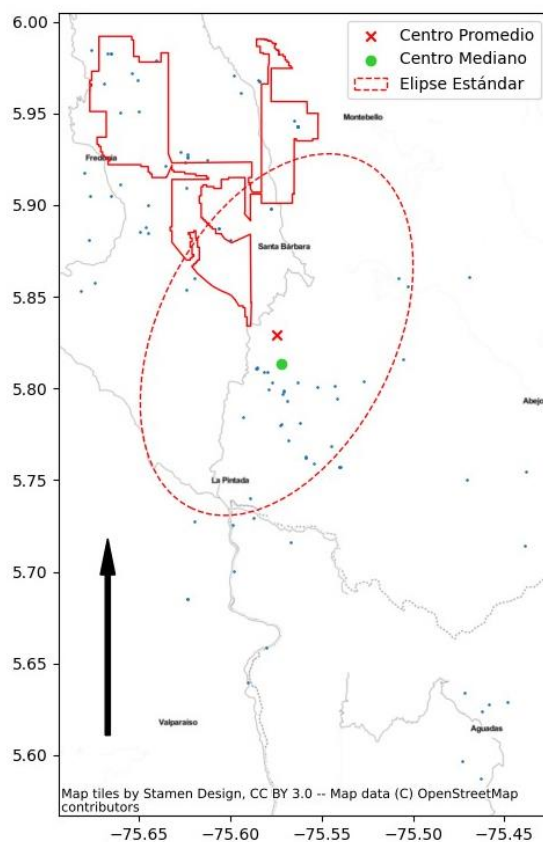


Figura 6 Análisis de la dispersión de las muestras mediante una elipse estándar.

El análisis de Clusterización permitió agrupar 3 grupos generales y 5 grupos más específicos con distintas elipses de distribución, que revelan que la tendencia general de los datos puede ser mejor descrita en dirección noroeste-sureste a norte-sur.

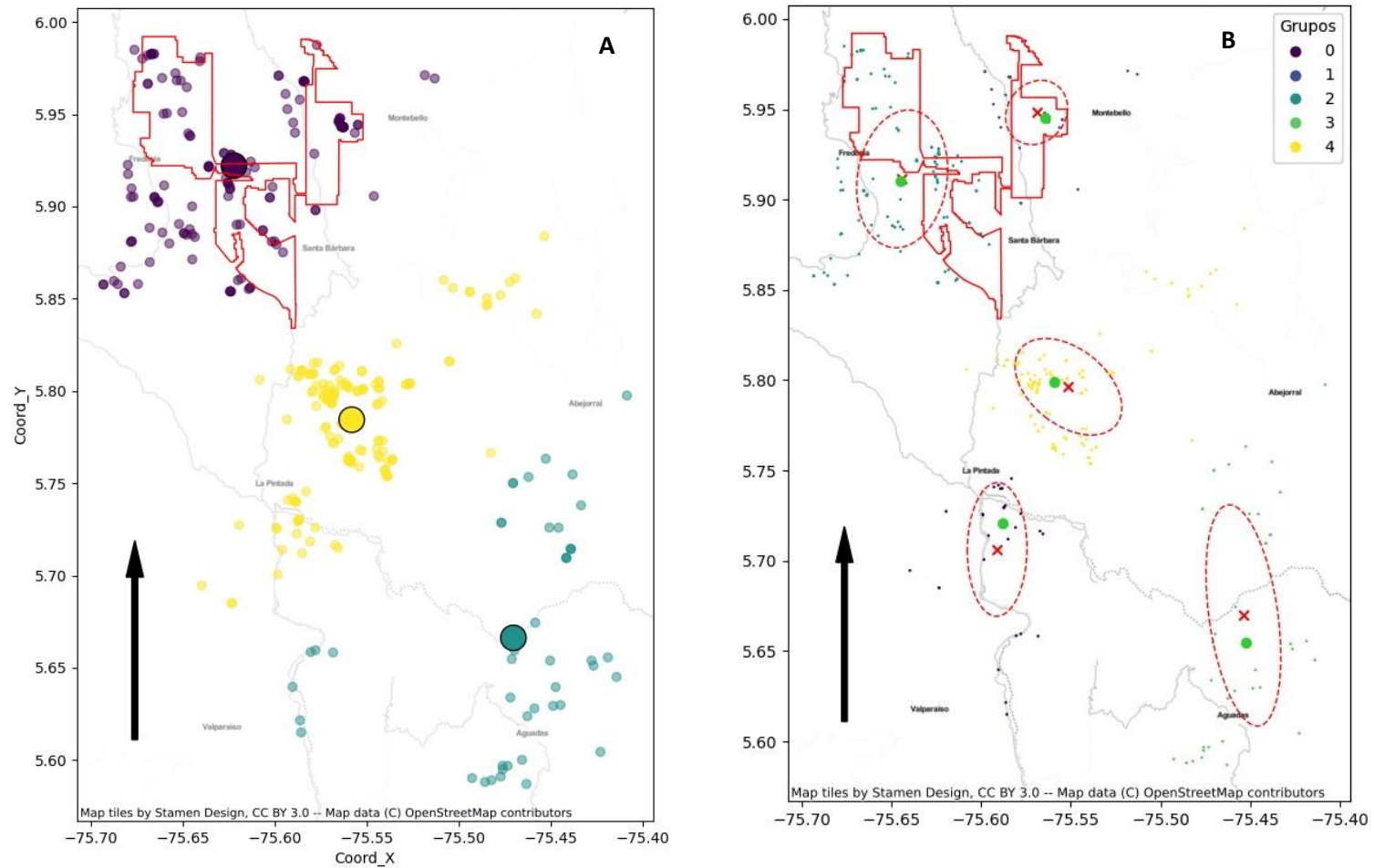


Figura 7 Análisis de centrografía de las muestras colectadas para 3 grupos (A) y 5 grupos (B).

Al contrastar las veredas (polígono) que componen el área con la distribución de las muestras (puntos) y sus respectivos valores de cobre (Cu_ppm) se determinó que la Vereda Damasco presenta el mayor número de muestras, con un total de 49 registros, mientras que en la Vereda Las Mercedes, del mismo municipio, se presentan las mayores concentraciones de cobre (885 ppm).

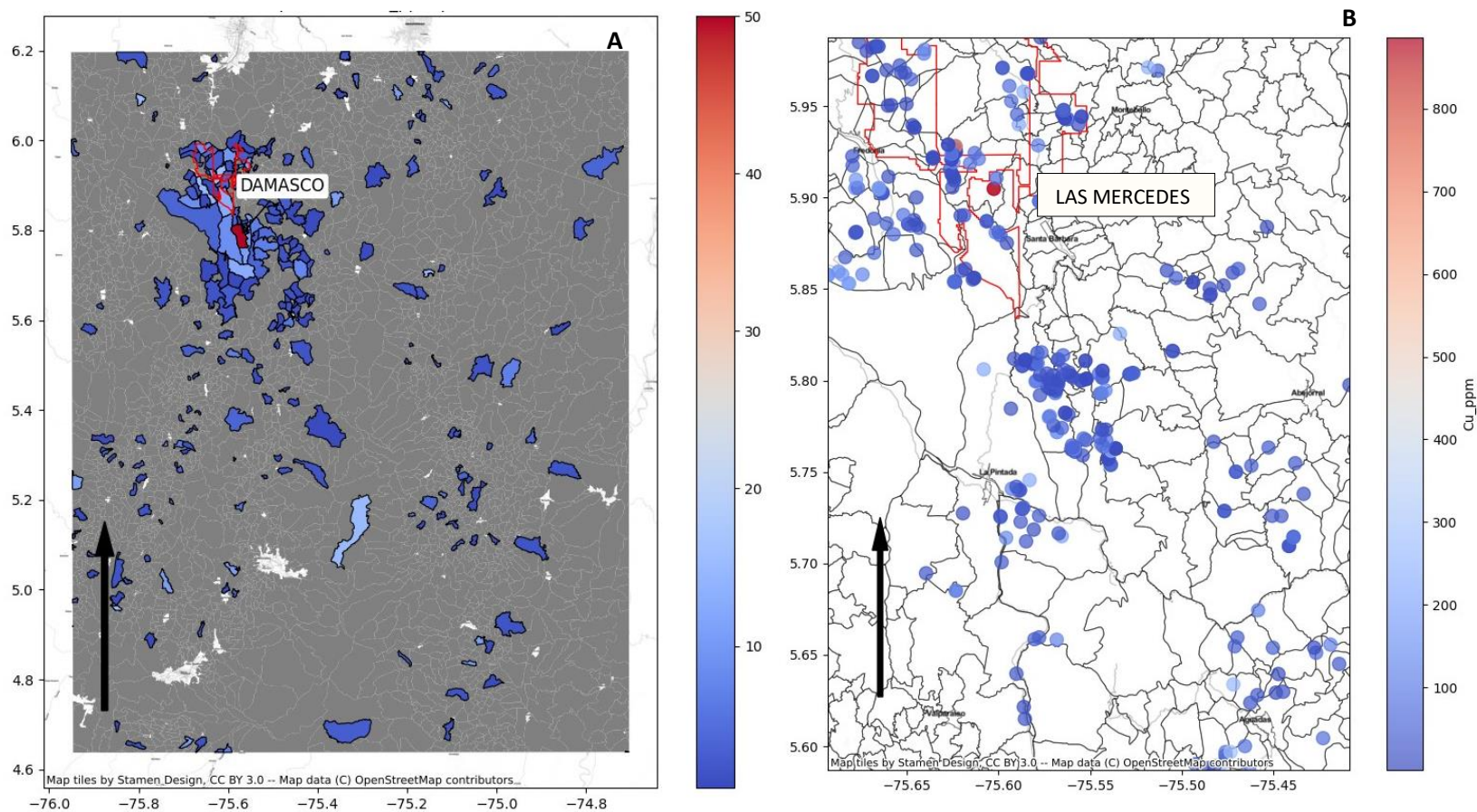


Figura 8 Mapa de calor de densidad de muestras según (A): las veredas del área de estudio y (B): la concentración de cobre (Cu).

4.2. Análisis Raster

El análisis ráster permitió obtener un mapa espectral del modelo de elevación digital (DEM) del área de estudio y de esta manera generar un mapa de sombras que resalta las estructuras geológicas del área que permite ver estructuras en dirección sureste-noroeste, así como suroeste-noreste.

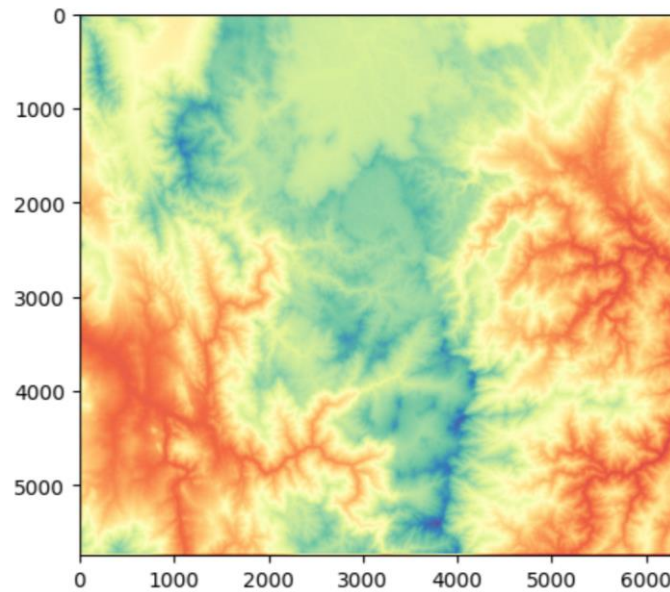


Figura 9 Modelo de elevación digital (DEM) de la zona de estudio con representación espectral.

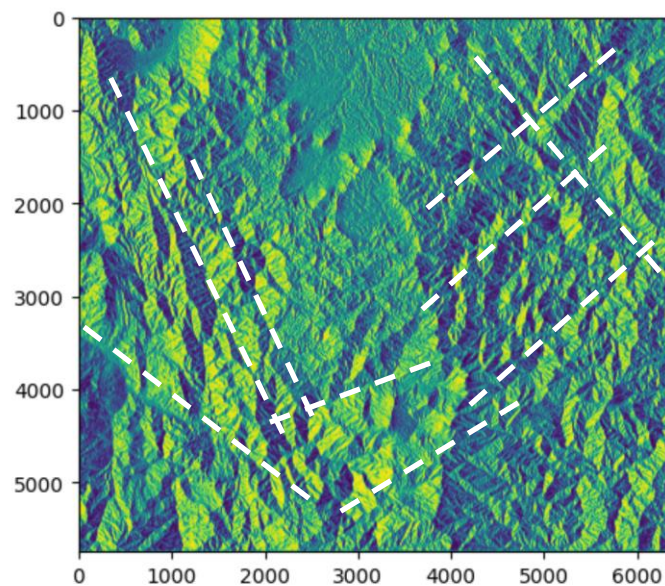


Figura 10 Hillshade obtenido a partir de la imagen DEM.

Para la interpolación estadística se creó el variograma experimental, a partir de la semivarianza en la diferencia de las distancias de los puntos. Para el caso del cobre (Cu) se ajustó su meseta a una semivarianza de 8000, lo cual indica una alta dispersión de los datos.

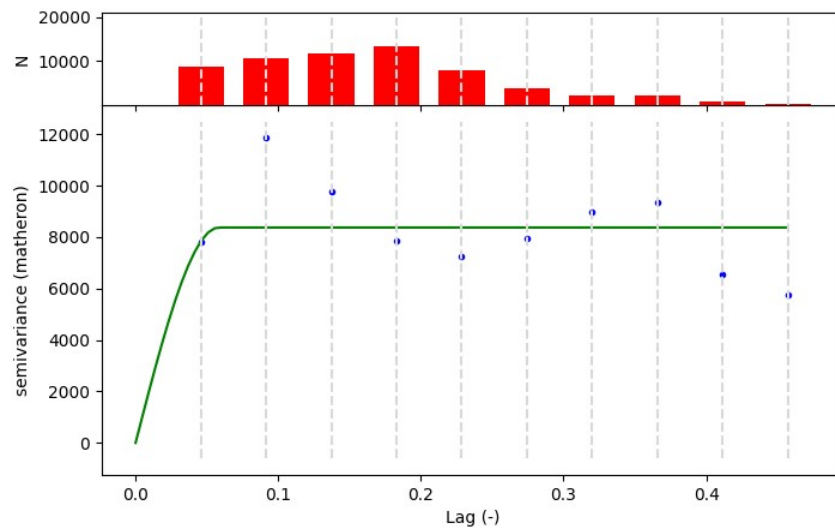


Figura 11 Variograma geométrico de la variable cobre (Cu).

Seguidamente se obtuvo la respectiva interpolación mediante Kriging Ordinario (KO):

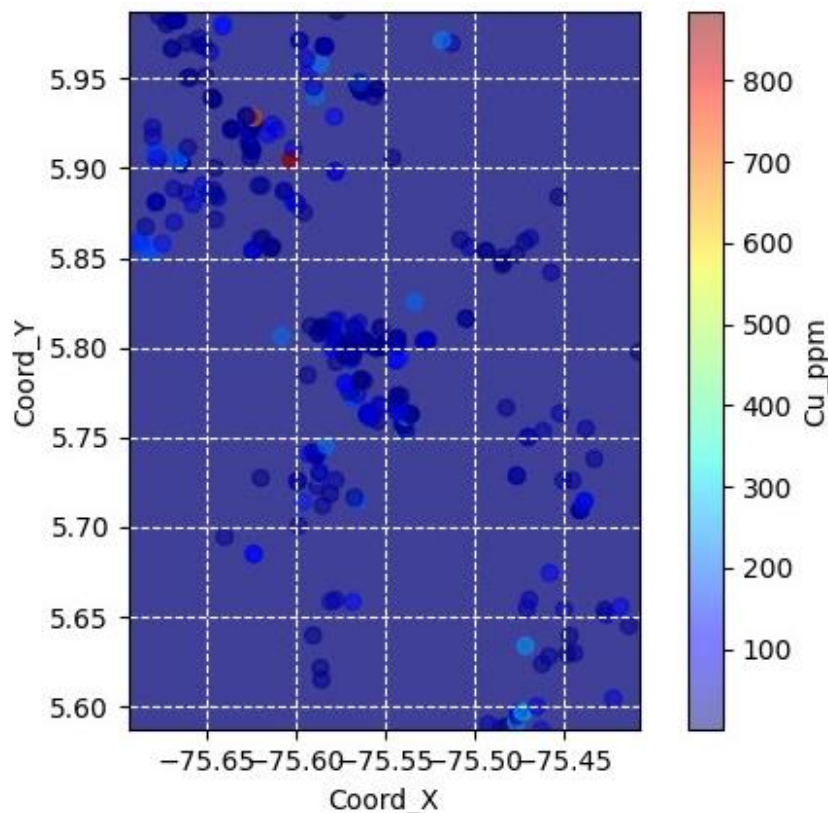


Figura 12 Interpolación mediante Kriging Ordinario (KO) de la variable cobre (Cu)

A partir de la interpolación generada se construyó el mapa de isovalores de cobre el cuál muestra una zona de altas concentraciones de cobre en inmediaciones del sector central del Distrito de Montebello, en el municipio Santa Bárbara, donde los valores estimados de cobre alcanzan hasta 885 ppm.

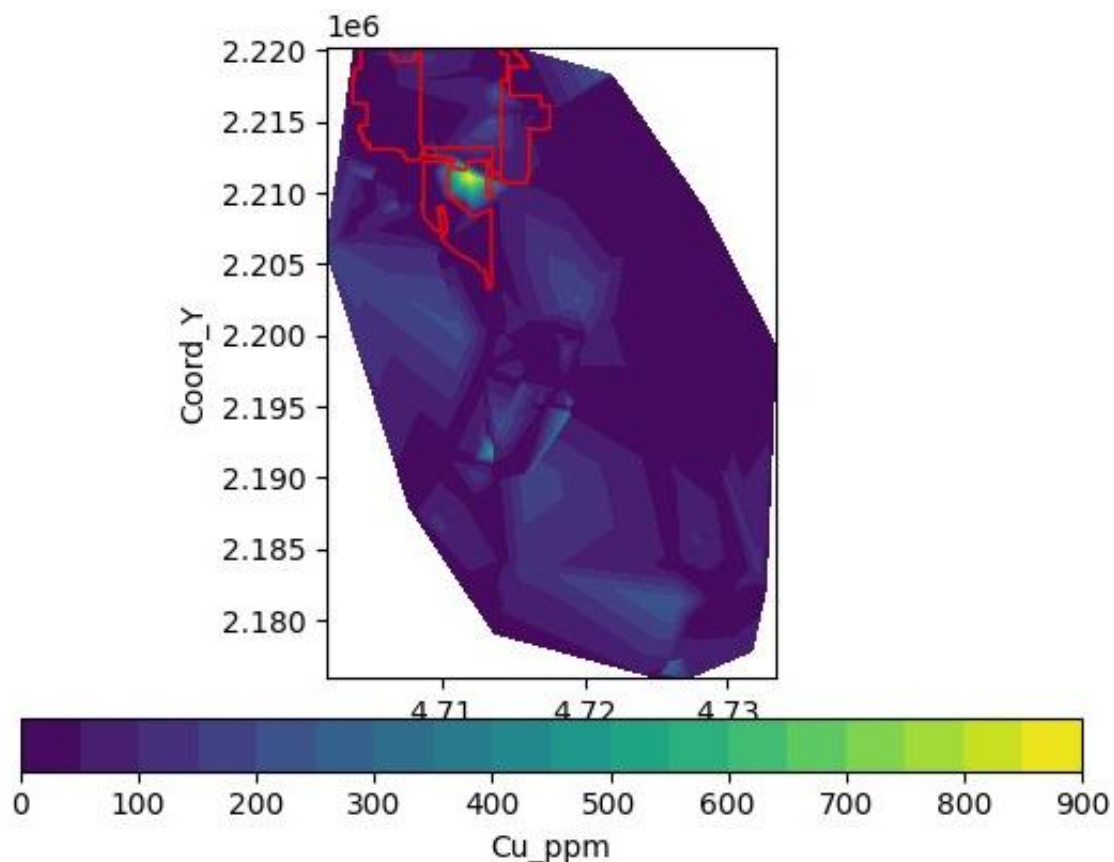


Figura 13 Mapa de isovalores interpolados de la variable cobre (Cu).

Este mismo ejercicio se repitió para la variable molibdeno (Mo), obteniéndose un variograma cuya meseta fue de 7,8, indicando una menor dispersión de los datos. Los resultados de Kriging Ordinario (KO) indican que las mayores concentraciones estimadas de este elemento alcanzan los 36 ppm y se ubican en la parte suroriental del área de estudio.

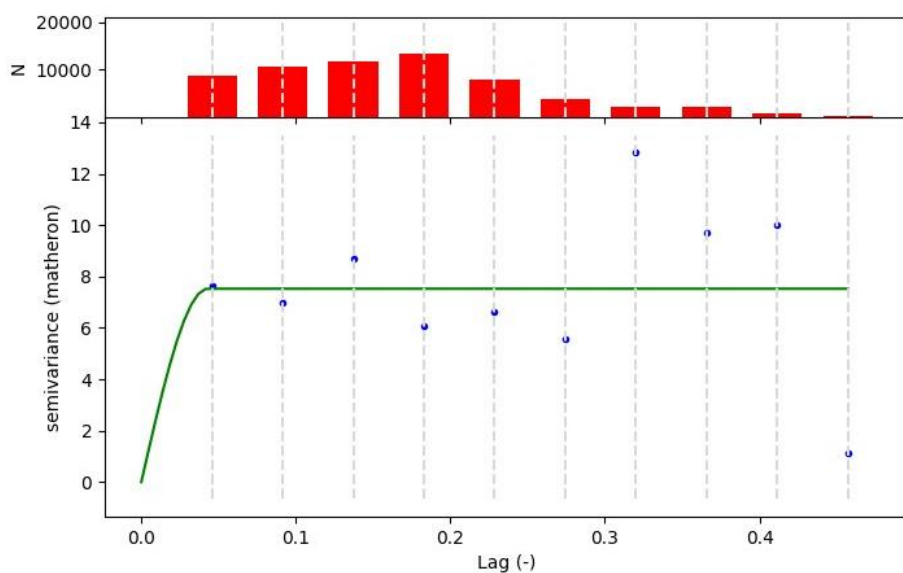


Figura 14 Variograma geométrico de la variable molibdeno (Mo).

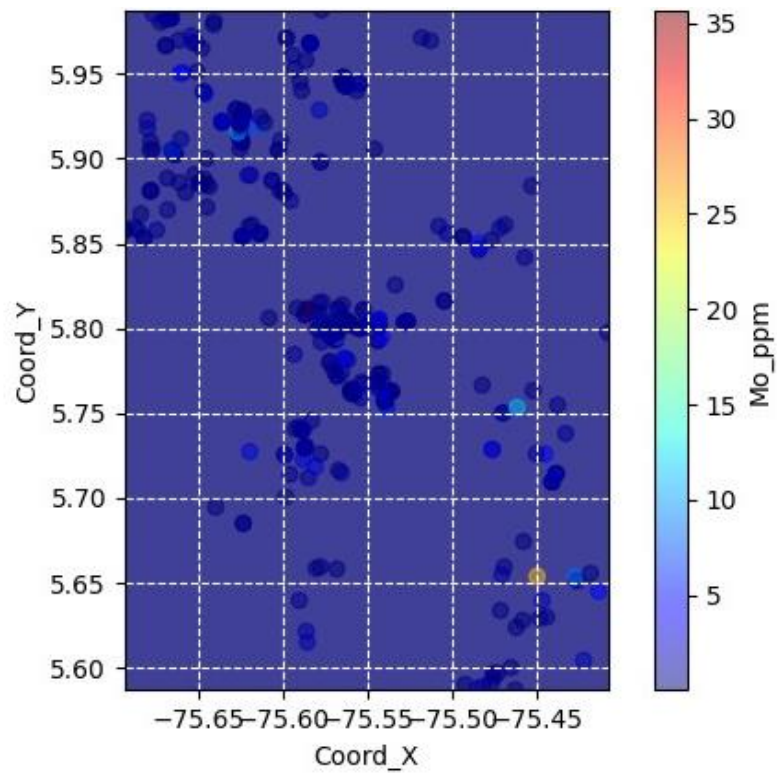


Figura 15 Interpolación mediante Kriging Ordinario (KO) de la variable molibdeno (Mo)

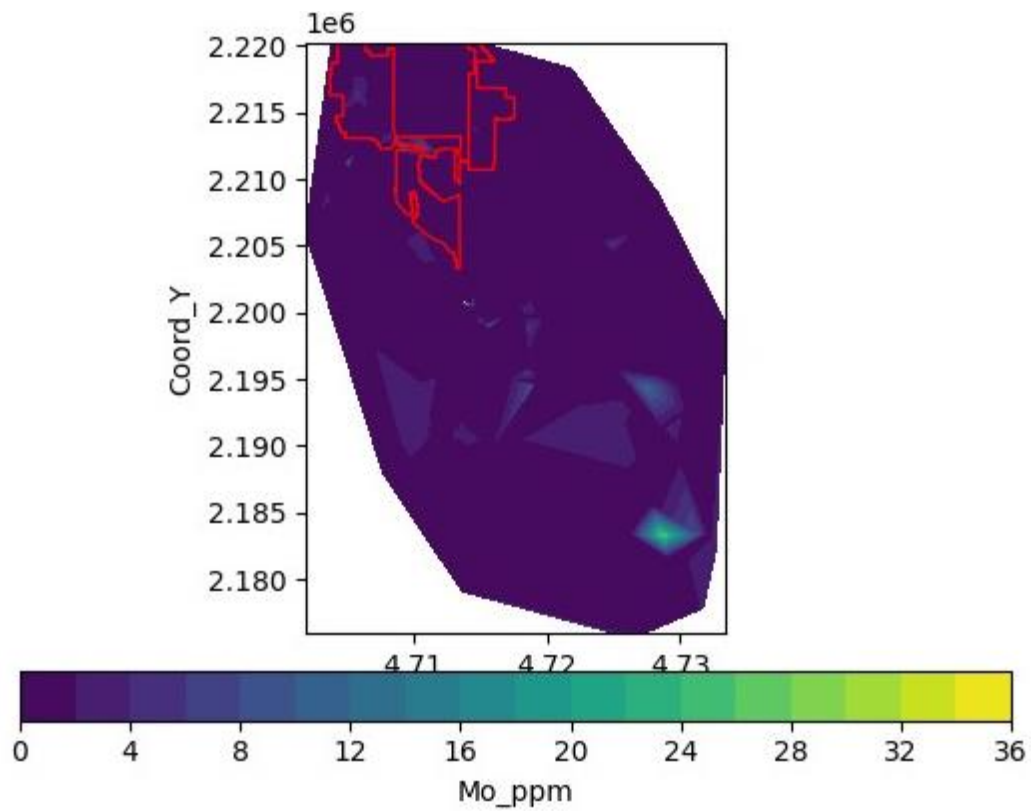


Figura 16 Mapa de isovalores interpolados de la variable molibdeno (Mo).

4.3. Regresión Lineal

Los resultados de la regresión lineal de mínimos cuadrados ordinarios (OLS) implementada, teniendo como variable dependiente el cobre (Cu) y como variable independiente el molibdeno, indican que:

- El coeficiente para la constante es 55.9100.
- El coeficiente para molibdeno (Mo) es -1.2768.
- El valor p para molibdeno (Mo) 0.493, lo cual indica que no hay evidencia suficiente para rechazar la hipótesis nula de que el coeficiente sea igual a cero.
- El valor R-cuadrado es 0.001, lo que significa que el 0.1% de la variabilidad en cobre (Cu) puede ser explicada por el molibdeno (Mo).
- El valor F-estadístico es 0.4712 y el valor p asociado es 0.493, lo que indica que el modelo en general no es estadísticamente significativo.

En resumen, no se encontró una relación espacial significativa entre cobre (Cu) y molibdeno (Mo) con base en estos resultados.

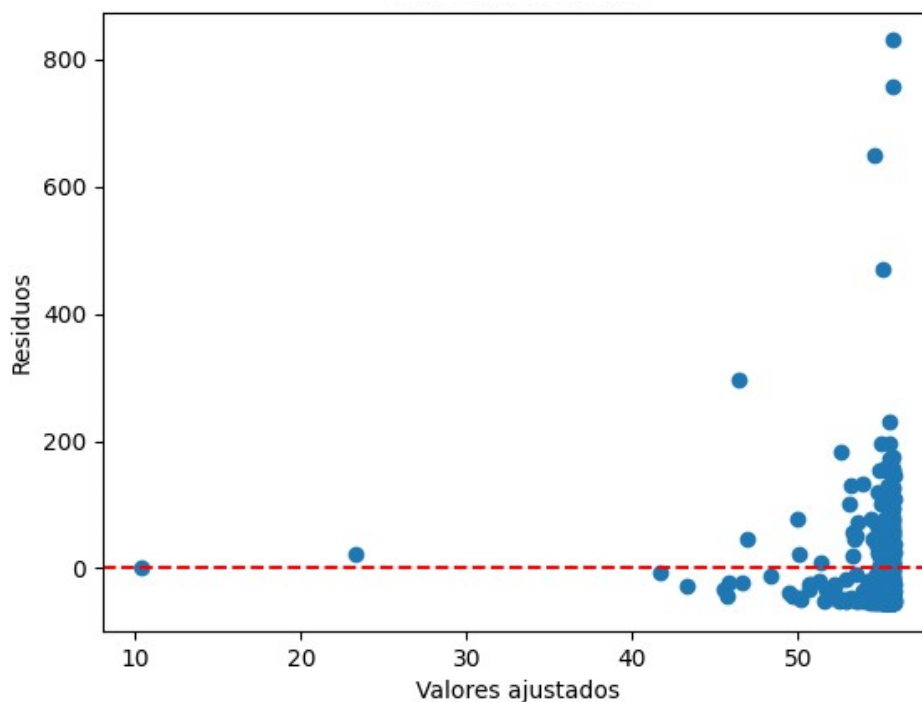


Figura 17 Grafico de Residuos de la regresión lineal entre las variables cobre (Cu) y Molibdeno (Mo)

```

OLS Regression Results
=====
Dep. Variable:          Cu_ppm    R-squared:                0.001
Model:                  OLS       Adj. R-squared:           -0.002
Method:                 Least Squares   F-statistic:              0.4712
Date:                   Fri, 30 Jun 2023   Prob (F-statistic):       0.493
Time:                   00:18:08    Log-Likelihood:           -2093.6
No. Observations:       351         AIC:                      4191.
Df Residuals:           349         BIC:                      4199.
Df Model:                1
Covariance Type:        nonrobust
=====
               coef      std err          t      P>|t|      [0.025      0.975]
-----
const          55.9100      5.425      10.306      0.000      45.240      66.580
Mo_ppm         -1.2768      1.860      -0.686      0.493      -4.935      2.381
=====
Omnibus:                 396.317    Durbin-Watson:              1.419
Prob(Omnibus):            0.000    Jarque-Bera (JB):           19527.874
Skew:                     5.074    Prob(JB):                   0.00
Kurtosis:                 38.104    Cond. No.                   3.19
=====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```

Figura 18 Resultados de la regresión lineal entre las variables cobre (Cu) y Molibdeno (Mo)

5. CONCLUSIONES

- Mediante un mapa de calor se determinó que la Vereda Damasco, ubicada en el municipio de Santa Barbara, presenta el mayor número de muestras por vereda, con un total de 49 registros. Mientras que en la Vereda Las Mercedes, del mismo municipio, se presentan las mayores concentraciones de cobre (885 ppm).
- La concentración promedio de cobre en el área de estudio equivale a 54, 5ppm, la cual se encuentra dos veces por encima del promedio de cobre para la corteza continental (28 ppm), mientras que la concentración promedio de molibdeno en el área de estudio no presenta anomalías geoquímicas respecto a la concentración promedio de este elemento en la corteza continental (1,1 ppm).
- Mediate Kriging Ordinario (KO) pudo identificarse una zona de concentración anómala de cobre en la vereda de Las Mercedes, ubicada en el municipio de Santa Bárbara, con un tenor máximo de 885 ppm.
- La interpolación realizada para molibdeno evidencia que las mayores concentraciones para este elemento se localizan hacia el sureste del área de estudio donde alcanzan valores de hasta 36 ppm.
- Se determinó mediante que no existe una dependencia espacial entre las variables cobre y molibdeno en el área de estudio.

6. REFERENCIAS BIBLIOGRÁFICAS

Servicio Geológico Colombiano. (2020). *Atlas Geoquímico de Colombia, Concentración de Cobre (Cu)*.

Servicio Geológico Colombiano. (2020). *Atlas Geoquímico de Colombia, Concentración de Molibdeno (Mo)*.

Índice de Figuras

Figura 1 Mapa geológico y localización del Distrito de Montebello, Sur de Antioquia.	2
Figura 2 Primeras 10 líneas de la tabla de datos utilizada.	3
Figura 3 Histograma de los valores de cobre (Cu_ppm).	6
Figura 5 Visualización (A) y tendencias (B) en la distribución de las muestras de roca analizadas.	7
Figura 9 Conteo de muestras por cuadrantes.	8
Figura 7 Análisis de la dispersión de las muestras mediante una elipse estándar.	8
Figura 10 Análisis de centrografía de las muestras colectadas para 3 grupos (A) y 5 grupos (B).	9
Figura 14 Mapa de calor de densidad de muestras según (A): las veredas del área de estudio y (B): la concentración de cobre (Cu).	10
Figura 15 Modelo de elevación digital (DEM) de la zona de estudio con representación espectral.	11
Figura 16 Hillshade obtenido a partir de la imagen DEM.	11
Figura 17 Variograma geométrico de la variable cobre (Cu).	12
Figura 18 Interpolación mediante Kriging Ordinario (KO) de la variable cobre (Cu)	12
Figura 19 Mapa de isovalores interpolados de la variable cobre (Cu).	13
Figura 20 Variograma geométrico de la variable molibdeno (Mo).	13
Figura 21 Interpolación mediante Kriging Ordinario (KO) de la variable molibdeno (Mo)	14
Figura 22 Mapa de isovalores interpolados de la variable molibdeno (Mo).	14
Figura 23 Grafico de Residuos de la regresión lineal entre las variables cobre (Cu) y Molibdeno (Mo)	15
Figura 21 Resultados de la regresión lineal entre las variables cobre (Cu) y Molibdeno (Mo)	16

Índice de Tablas

Tabla 1 Estadísticas Descriptivas de las variables cobre (Cu) y molibdeno (Mo).	6
Tabla 2 Estadísticas básicas	Error! Bookmark not defined.