

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

MAT283

ANÁLISIS Y PREDICCIÓN UTILIZANDO TÉCNICAS  
DE KRIGING PARA ARIZONA'S COPPER DEPOSIT  
(KD)

---

## Laboratorio de Modelación II

---

*Autores:*

Bastián Aceitón  
Francine Palacios

2024-1

# Índice

<b>Descripción</b>	<b>3</b>
Introducción al tema de trabajo, motivación y relevancia . . . . .	3
Estado del arte del problema . . . . .	3
Objetivos generales y específicos desarrollados durante el semestre . . . . .	4
Herramientas matemáticas necesarias para cumplir los objetivos anteriores . . . . .	5
Declaración contribución y uso de herramientas de Inteligencia Artificial . . . . .	6
<b>Metodología de trabajo</b>	<b>6</b>
Fechas y duración de reuniones con especialista . . . . .	6
Planificación de reuniones con especialista, tareas realizadas en el semestre por cada integrante del grupo y los períodos de cada tarea . . . . .	7
<b>RESULTADOS</b>	<b>8</b>
Background Matemático . . . . .	8
Transformación Box-Cox . . . . .	8
Campo Aleatorio . . . . .	9
Semi-variograma . . . . .	10
Semi-Variograma empírico . . . . .	11
Kriging . . . . .	12
K-Folds Cross-Validation . . . . .	13
Herramientas de programación . . . . .	14
Procedimiento y resultados principales . . . . .	14
<b>Conclusiones, preguntas abiertas y potencial trabajo a futuro</b>	<b>22</b>
<b>Bibliografía</b>	<b>24</b>
<b>Anexo</b>	<b>25</b>

# Descripción

## Introducción al tema de trabajo, motivación y relevancia

El trabajo que se nos encomendó, fue el de implementar (mediante código) distintas herramientas provenientes de la estadística para hacer una predicción espacial de una variable de interés, para nuestro caso específico, la variable de interés fue la concentración de cobre y el espacio ambiente del problema es un subconjunto de  $\mathbb{R}^3$ . Concretamente, para un dataset dado, nos enfocamos en extraer los datos  $(x, y, z, \%cu)$ , con lo cual, la formulación del problema es:

predecir  $\%cu$  en un punto  $(\bar{x}, \bar{y}, \bar{z})$  dado.

Cabe mencionar que esta manera de proceder puede ser perfectamente aplicada para predecir la concentración de otros minerales, como lo puede ser el oro.

La motivación y relevancia de nuestro problema, no debería ser ajena a ningún ciudadano o ciudadana Chileno, pues, nuestro país es considerado internacionalmente como un país minero y gran parte de nuestro PIB es debido al sector minero, en consecuencia, es crucial trabajar y resolver problemas mineros como el nuestro, el cual, dado un trabajo de previo de recopilación datos o muestras, nos permite dar indicios de los lugares del espacio en donde hay mayor concentración del mineral de interés, y por lo tanto, ayuda a la planificación minera en la toma de mejores decisiones a la hora de extraer el mineral.

## Estado del arte del problema

A modo de ejemplo, en una de las publicaciones que revisamos durante este trabajo [4] se afirma que hay problemas de minería a rajo abierto con restricciones de capacidad que son NP-difícil, con lo cual, constantemente se están implementando nuevos algoritmos y aproximaciones al problema que hacen del proceso general para hallar las predicciones más eficiente, de hecho, sin temor, se puede decir que este problema está tan avanzado como la optimización estocástica y la geoestadística lo estén. Sin embargo, las librerías y software especializados más avanzadas por lo general son closed-source, por lo tanto, es también de interés poder aplicar librerías de libre acceso para poder abordar este tipo de problemas. Adicionalmente, algunos de los métodos vigentes son (que pueden ser vistos en [7]):

1. (Estimación por Kriging). Una vez escogido un modelo de variograma se procede a definir los parámetros para la correcta aplicación del kriging. El kriging es una herramienta de interpolación geoestadística que depende de la distancia de separación existente entre el sitio a estimar y las observaciones conocidas y de la distribución geométrica de las observaciones conocidas. De forma adicional, incorpora la continuidad espacial de los valores gracias a la función de variograma de la variable aleatoria.

2. (Simulación). Dado el conjunto de observaciones conocidas  $\{Z(x_i), i = 1, \dots, n\}$ , existen técnicas de simulación no condicional que solo tienen por objetivo reproducir el mismo comportamiento espacial de la variable regionalizada. Razón por la que no consideran la relación entre los valores simulados y los datos experimentales. También se dispone de métodos de simulación exacta o condicional que además, buscan que la simulación resultante sea igual a las muestras en las ubicaciones de datos reales, tal que:

$$Z_s(x_i) = Z(x_i), \forall x \in D, i = 1, \dots, n$$

El método más simple para producir la simulación exacta de una variable aleatoria es el método secuencial, seguido del método de descomposición de Cholesky de la matriz de covarianza  $C$  o  $LU$ , llamado así por la descomposición de la matriz de covarianzas  $C$ .

## Objetivos generales y específicos desarrollados durante el semestre

Se nos dio la tarea de trabajar con el dataset KD (Arizona's Copper Deposit), el cual no corresponde a los datos reales de la empresa, pero, tienen el mismo formato, y por lo tanto, es sencillo adaptar el código para los datos reales de la empresa. En el contexto del Depósito de Cobre de Arizona (KD), la precedencia se calcula utilizando un ángulo de 45 grados y 8 niveles. Esto significa que la extracción de un bloque depende de la extracción previa de otros bloques que están directamente encima de él, formando una especie de “escalera” o “rampa” que respeta el ángulo de inclinación establecido. Este ángulo es importante para mantener la estabilidad de las paredes del pozo y prevenir deslizamientos o derrumbes. Por ejemplo, si un bloque está en un nivel inferior, no se puede extraer hasta que se hayan extraído todos los bloques que están en los niveles superiores y que caen dentro del cono de influencia definido por el ángulo de 45 grados. Esto asegura que no se comprometa la integridad estructural de la mina durante la operación (cada bloque tiene dimensiones de  $(20 \times 20 \times 15)$ [m], lo que representa el volumen físico de cada unidad extraída). El dataset puede ser hallado en este link. Para ser más precisos, el dataset consiste de 14153 observaciones y de nueve columnas, las cuales son:

1. **id**: almacena un identificador único para el bloque, donde los identificadores del bloque están numerados, comenzando desde cero.
2. **x**: representa una de las coordenadas del bloque y el eje  $x$  apunta hacia la izquierda del observador.
3. **y**: representa una de las coordenadas del bloque y el eje  $y$  apunta directamente hacia el observador.
4. **z**: representa una de las coordenadas del bloque, donde la coordenada  $z$  igual a cero corresponde a la capa más baja en el yacimiento y el eje  $z$  apunta hacia arriba.
5. **tonn**: peso en toneladas del bloque.

6. **blockvalue**: valor del bloque, calculado en base a los costos y ganancias.
7. **destination**: destino del bloque, mineral (1) o desecho (2).
8. **CU %**: porcentaje de cobre en el bloque.
9. **process\_profit**: ganancia por procesamiento del bloque.

Posteriormente, extrajimos un sub-dataset del dataset original el cual solamente contiene las columnas x, y, z y CU% y la cantidad total de observaciones es la misma. En primera instancia, el objetivo consistía en entender el dataset y hacer un análisis exploratorio de datos. Posteriormente, el objetivo principal durante el resto del semestre fue implementar en código distintos métodos de interpolación a este dataset e informar sobre las hipótesis que éstos métodos deben satisfacer, como lo son: el kriging simple, kriging ordinario y kriging universal para los cuales a su vez se probaron distintos variogramas, como lo son el esférico, exponencial, gaussiano y lineal. En la figura 1 puede ver como lucen los datos.

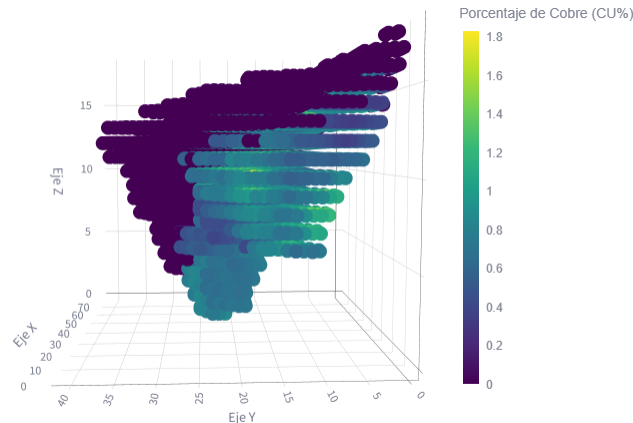


Figura 1: Porcentaje de Cobre en cada coordenada

## Herramientas matemáticas necesarias para cumplir los objetivos anteriores

Las herramientas matemáticas son las siguientes:

1. (**Box – Cox**). La transformación de Box-Cox es una familia de transformaciones de potencia que tienen como objetivo estabilizar la varianza y hacer que los datos sigan una distribución normal. Es particularmente útil para transformar variables dependientes no normales en una forma normal, lo cual es una suposición de muchas técnicas estadísticas.
2. (**Kriging – Simple**). El Kriging Simple es un método de interpolación geoestadística utilizado para predecir valores desconocidos en ubicaciones no muestreadas basándose en valores conocidos en ubicaciones muestreadas. Es un tipo de kriging que asume una media conocida y constante sobre toda el área de estudio.

3. (**Kriging – Ordinario**). El Kriging Ordinario es un método de interpolación geoestadística ampliamente utilizado que, a diferencia del Kriging Simple, no asume una media conocida para el proceso aleatorio subyacente. En cambio, el Kriging Ordinario asume una media constante desconocida sobre una vecindad local de la ubicación de predicción. Esto lo hace más flexible y generalmente más aplicable a situaciones del mundo real donde la media no es conocida.
4. (**Kriging – Universal**). El Kriging Universal, también conocido como Kriging con tendencia, es un método geoestadístico avanzado que extiende el Kriging Ordinario al permitir un componente de tendencia determinista en los datos. Este método es particularmente útil cuando el proceso espacial subyacente muestra una tendencia que puede ser modelada por una función conocida.
5. (**Variograma**). Un variograma es una herramienta fundamental en la geoestadística utilizada para cuantificar la dependencia espacial de un campo aleatorio o proceso estocástico. Mide cómo la similitud de los datos disminuye con el aumento de la distancia entre los puntos de datos. Comprender el variograma es crucial para métodos de interpolación espacial como el kriging. Y así como hay distintos métodos de Kriging, también hay distintos variogramas, entre los cuales se encuentran el esférico, exponencial, gaussiano y lineal que serán detallados en la última sección.

A pesar de que no son herramientas matemáticas como tal, vale la pena mencionar que se usó python y R para implementar las herramientas matemáticas antes mencionadas.

## Declaración contribución y uso de herramientas de Inteligencia Artificial

Ambos integrantes del grupo trabajamos tanto en la elaboración de código al igual que en la confección del informe. El uso de herramientas de inteligencia artificial fue equivalente al del uso de un motor de búsqueda tradicional.

## Metodología de trabajo

### Fechas y duración de reuniones con especialista

Las reuniones fueron todos jueves de cada semana a las 12:00 a partir del 28 de marzo y con termine el 13 junio, mediante la plataforma Zoom. Con una duración promedio de 40 minutos.

## **Planificación de reuniones con especialista, tareas realizadas en el semestre por cada integrante del grupo y los períodos de cada tarea**

Los trabajos asignados en las 11 semanas entre el 28 de marzo y el 13 de junio fueron las siguientes:

- Semana 1: Análisis exploratorio de datos e implementación de página web: Francine Palacios.  
Estudio de librerías e implementación de código en Python: Bastián Aceitón.
- Semana 2: Análisis exploratorio de datos e implementación de página web: Francine Palacios.  
Estudio de librerías e implementación de código en Python: Bastián Aceitón.
- Semana 3: Implementación del kriging simple en la página web: Francine Palacios.  
Código en python para el kriging simple: Bastián Aceitón.
- Semana 4: Implementación del kriging simple en la página web: Francine Palacios.  
Código en python para el kriging simple: Bastián Aceitón.
- Semana 5: Implementación del kriging ordinario en la página web: Francine Palacios.  
Código en python para el kriging ordinario: Bastián Aceitón.
- Semana 6: Implementación del kriging ordinario en la página web: Francine Palacios.  
Código en python para el kriging ordinario: Bastián Aceitón.
- Semana 7: Implementación del kriging universal en la página web: Francine Palacios.  
Código en python para el kriging universal: Bastián Aceitón.
- Semana 8: Implementación del kriging universal en la página web: Francine Palacios.  
Código en python para el kriging universal: Bastián Aceitón.
- Semana 9: Implementación del mejor modelo de kriging en la página web: Francine Palacios.  
Código en python para el mejor modelo de kriging: Bastián Aceitón.
- Semana 10: Conversión del código en python a R en la pagina web: Francine Palacios.  
Confección informe: Bastián Aceitón.
- Semana 11: Confección informe: Francine Palacios.  
Confección informe: Bastián Aceitón.

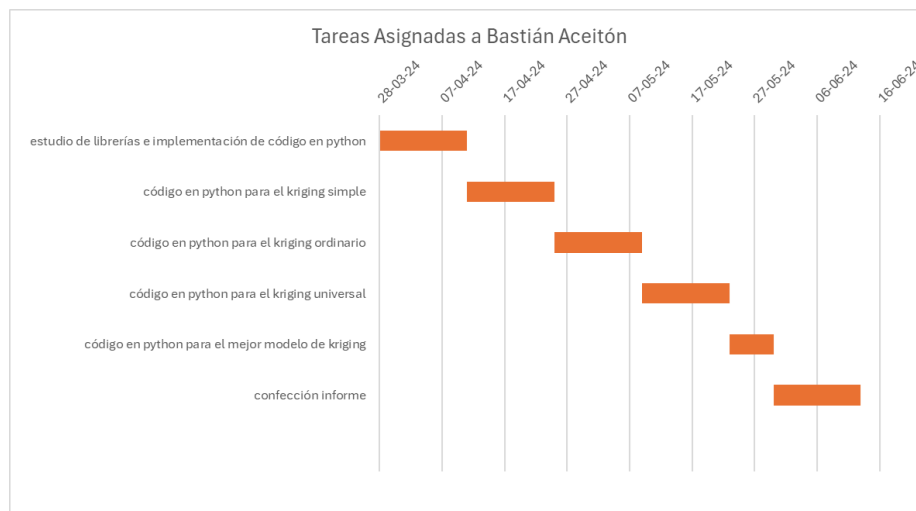


Figura 2: Carta Gantt para integrante Bastián Aceitón

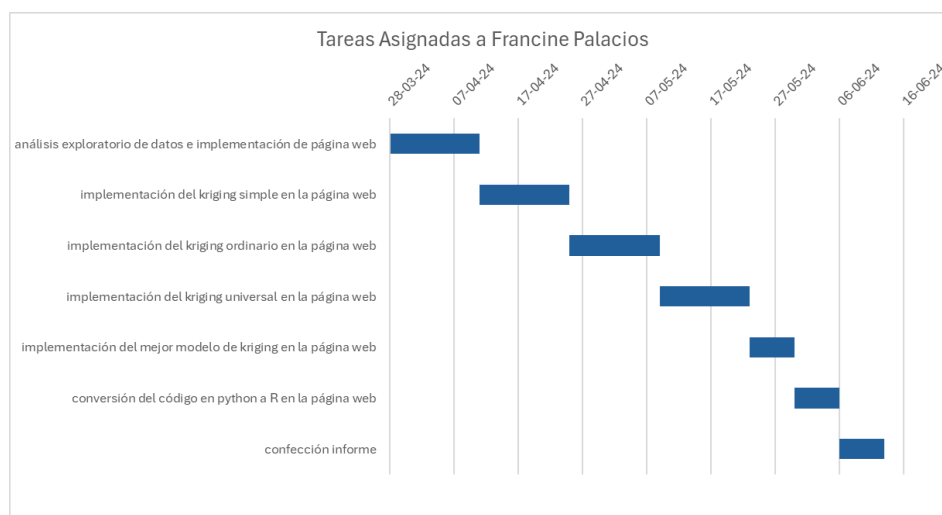


Figura 3: Carta Gantt para integrante Francine Palacios

## RESULTADOS

### Background Matemático

#### Transformación Box-Cox

La transformación de Box-Cox es una técnica ampliamente utilizada en la estadística para estabilizar la varianza y aproximar los datos a una distribución normal. Esta transformación fue introducida por los estadísticos George Box y David Cox en su artículo seminal en 1964 (ver [1]) pero la revisión de Sakia en 1992 (ver [6]) ofrece una perspectiva más completa y moderna



**Definición:** Bajo la hipótesis de que los datos deben ser estrictamente positivos, independientes entre sí, y misma distribución subyacente. La transformación de Box-Cox se define de la siguiente manera:

$$T_{\lambda}(x) = \begin{cases} \frac{x^{\lambda}-1}{\lambda} & , \lambda \neq 0 \\ \ln(x) & , \lambda = 0 \end{cases}$$

donde  $x$  es el dato a transformar y  $\lambda$  es el parámetro de transformación.

Se utiliza un criterio de evaluación, como el métrico  $d_{\lambda}$ , definido como:

$$d_{\lambda} = \frac{\bar{x}_{\lambda} - q_{0,5,\lambda}}{\text{IQR}_{\lambda}}$$

donde  $\bar{x}_{\lambda}$  es la media de los datos transformados,  $q_{0,5,\lambda}$  es la mediana de los datos transformados y  $\text{IQR}_{\lambda}$  es el rango intercuartílico de los datos transformados. Se selecciona el valor de  $\lambda$  que minimice  $d_{\lambda}$  o cualquier otro criterio de evaluación utilizado, indicando que este valor de  $\lambda$  proporciona la mejor transformación de los datos.

Las definiciones y métodos enunciados a continuación fueron extraídos de [5].

## Campo Aleatorio

**Definición:** Un campo aleatorio es una familia  $X = (X_t)_{t \in T}$  de variables aleatorias  $X_t$  que están definidas en el mismo espacio de probabilidad y están indexadas por  $t$  en un subconjunto  $T$  de  $\mathbb{R}^d$ .

**Notación:** Las distribuciones de dimensión finita involucran dos parámetros, el vector de medias y la matriz de covarianza. Los elementos de esta última son  $\text{Cov}(X_{t_i}, X_{t_j})$ ,  $i, j = 1, \dots, n$ . Definimos las funciones

$$m : T \rightarrow \mathbb{R}; \quad m(t) = \mathbb{E}X_t$$

y

$$\rho : T \times T \rightarrow \mathbb{R}; \quad \rho(s, t) = \text{Cov}(X_s, X_t).$$

Las cuales llamaremos la función de media y la función de covarianza de  $X$  respectivamente.

**Definición** Un campo aleatorio  $X = (X_t)_{t \in \mathbb{R}^d}$  es débilmente estacionario si

- $\mathbb{E}X_t^2 < \infty$  para todo  $t \in \mathbb{R}^d$ ;
- $\mathbb{E}X_t \equiv m$  es constante;
- $\text{Cov}(X_{t_1}, X_{t_2}) = \rho(t_2 - t_1)$  para alguna  $\rho : \mathbb{R}^d \rightarrow \mathbb{R}$ .

**Definición:** Un campo aleatorio  $X = (X_t)_{t \in \mathbb{R}^d}$  es intrínsecamente estacionario si cumple con las siguientes condiciones:

- $\mathbb{E}X_t^2 < \infty$  para todo  $t \in \mathbb{R}^d$ ;
- $\mathbb{E}X_t \equiv m$  es constante;
- $\text{Var}(X_{t_2} - X_{t_1}) = f(t_2 - t_1)$  para alguna función  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ .

## Semi-variograma

En geoestadística, se prefiere a menudo el semi-variograma en lugar de la función de covarianza.

**Definición:** Sea  $X = (X_t)_{t \in \mathbb{R}^d}$  intrínsecamente estacionario. Entonces, el semi-variograma  $\gamma : \mathbb{R}^d \rightarrow \mathbb{R}$  se define como

$$\gamma(t) = \frac{1}{2} \text{Var}(X_t - X_0), \quad t \in \mathbb{R}^d.$$

Notemos que

$$\gamma(t) = \rho(0) - \rho(t)$$

para campos aleatorios débilmente estacionarios. En particular,  $\gamma(0) = \rho(0) - \rho(0) = 0$ . Sin embargo, la definición de un semi-variograma solo requiere la suposición más débil de estacionariedad intrínseca.

A continuación, listamos los semi-variogramas correspondientes a algunas de las funciones de covarianza:

- Para la función de covarianza exponencial  $\rho(t) = \exp[-\beta\|t\|]$ ,

$$\gamma(t) = 1 - \exp[-\beta\|t\|]$$

- Para la función de covarianza Gaussiana  $\rho(t) = \exp[-\beta\|t\|^2]$ ,

$$\gamma(t) = 1 - \exp[-\beta\|t\|^2]$$

- Para la función de variograma Nugget:

$$\gamma(t) = \begin{cases} 0 & \text{si } \|t\| = 0 \\ 1 & \text{si } \|t\| > 0 \end{cases}$$

- Para la función de variograma esférica:

$$\gamma(t) = \begin{cases} \frac{3\|t\|}{2a} - \frac{1}{2} \left( \frac{\|t\|}{a} \right)^3 & \text{si } 0 \leq \|t\| \leq a \\ 1 & \text{si } \|t\| > a \end{cases}$$

- Para la función de variograma lineal:

$$\gamma(t) = \|t\|, \quad \|t\| \geq 0$$

- Para la función de variograma lineal con umbral:

$$\gamma(t) = \begin{cases} \frac{\|t\|}{a} & \text{si } 0 \leq \|t\| \leq a \\ 1 & \text{si } \|t\| > a \end{cases}$$

## Semi-Variograma empírico

En muchas aplicaciones, solo se dispone de una única muestra finita  $X_{t_1}, \dots, X_{t_n}$ ,  $n \in \mathbb{N}$ , del campo aleatorio  $X$ . Para poder llevar a cabo una inferencia estadística, se debe obtener una réplica artificial asumiendo al menos la estacionariedad intrínseca. La idea es, para un desfase  $t$ , considerar todos los pares de observaciones que están 'aproximadamente' a una distancia  $t$  y promediar.

**Definición** *Estimador clásico de variograma de Matheron:* Usando la idea anteriormente comentada se obtiene el estimador suavizado de Matheron

$$\hat{\gamma}(t) = \frac{1}{2|N(t)|} \sum_{(t_i, t_j) \in N(t)} (X_{t_j} - X_{t_i})^2,$$

donde la vecindad  $t$ -vecindad  $N(t)$  se define por

$$N(t) = \{(t_i, t_j) : t_j - t_i \in B(t, \varepsilon)\},$$

$B(t, \varepsilon)$  es la bola cerrada de radio  $\varepsilon$  centrada en  $t$  y  $|\cdot|$  denota la cardinalidad.

La elección de  $\varepsilon$  es un arte. Debe ser lo suficientemente pequeña para que  $\gamma(t_j - t_i) \approx \gamma(t)$  para  $t_j - t_i$  en la bola  $\varepsilon$  alrededor de  $t$  y, por otro lado, lo suficientemente grande para tener un número razonable de puntos en  $N(t)$  para que el promedio sea estable. En otras palabras, hay un equilibrio entre sesgo y varianza.

Este estimador es muy afectado por la presencia de valores atípicos, e incluso un solo dato discordante puede distorsionar las estimaciones finales del variograma.

**Definición** *Estimador robusto de variograma de Cressie-Hawkins:* El estimador propuesto por Cressie y Hawkins (1980) (ver [2]) se define como

$$2\hat{\gamma}_{CH}(h) = \frac{1}{C_h} \left[ \left( \frac{1}{N_h} \sum_{N(h)} |Z(x_i) - Z(x_j)|^{\frac{1}{2}} \right)^4 \right]$$

donde  $C_h = \left(0,457 + \frac{0,494}{N_h} + \frac{0,494}{N_h^2}\right)$  es un factor de corrección para el sesgo cuando la variable  $Z(\cdot)$  es normal. Este estimador fue diseñado para ser robusto frente a valores atípicos y no normalidad en distribuciones que son similares a la normal en la región central pero más pesadas que la normal en las colas.

**Observación:** Notar que, aunque el estimador de Matheron es no paramétrico, se puede ajustar una familia específica  $\gamma_\theta$  minimizando el contraste

$$\sum_j w_j (\hat{\gamma}(h_j) - \gamma_\theta(h_j))^2$$

con, por ejemplo,  $w_j = |N(h_j)|$  o  $w_j = |N(h_j)|/\gamma_\theta(h_j)^2$  y la suma abarca una familia finita de desfases  $h_j$ . Para la segunda elección, cuanto menor sea el valor del semivariograma teórico, mayor será el peso dado a un par de observaciones en aproximadamente ese desfase para compensar su esperada rara ocurrencia. Misma idea se puede realizar para el estimador de Cressie-Hawkins.

## Kriging

**Definición** *Kriging Simple*. Supongamos que se observan los valores  $X_{t_1} = x_{t_1}, \dots, X_{t_n} = x_{t_n}$  de un campo aleatorio  $X = (X_t)_{t \in \mathbb{R}^d}$  en  $n$  ubicaciones  $t_i \in \mathbb{R}^d, i = 1, \dots, n$ . Basándonos en estas observaciones, el objetivo es predecir el valor en alguna ubicación  $t_0$  de interés en la que no se ha realizado ninguna medición. Necesitaremos la función de media  $m$  y la función de covarianza  $\rho$  de  $X$ . Restrinjámonos a predictores lineales de la forma

$$\hat{X}_{t_0} = c(t_0) + \sum_{i=1}^n c_i X_{t_i}$$

**Teorema:** Sean  $X_{t_1}, \dots, X_{t_n}$  muestras de un campo aleatorio  $(X_t)_{t \in \mathbb{R}^d}$  en  $n$  ubicaciones  $t_i \in \mathbb{R}^d, i = 1, \dots, n$ , y júntelos en el  $n$ -vector  $Z$ . Denotemos por  $\Sigma$  la matriz de covarianza de  $Z$  y asumamos que  $\Sigma$  existe y es no singular. Adicionalmente, sea  $K = (K_i)_{i=1}^n$  el vector  $n \times 1$  con entradas  $K_i = \rho(t_i, t_0)$ . Entonces

$$\hat{X}_{t_0} = m(t_0) + K' \Sigma^{-1} (Z - \mathbb{E}Z)$$

es el mejor predictor lineal de  $X_{t_0}, t_0 \in \mathbb{R}^d$ , en términos del error cuadrático medio. El error cuadrático medio de la predicción está dado por

$$\rho(t_0, t_0) - K' \Sigma^{-1} K$$

**Definición:** *Kriging Ordinario*. Consideremos el modelo

$$X_t = \mu + E_t, \quad t \in \mathbb{R}^d,$$

donde  $\mu \in \mathbb{R}$  es la media global desconocida y  $E_t$  es un campo aleatorio de media cero con función de covarianza  $\text{Cov}(E_t, E_s) = \rho(t, s)$ . Basándonos en muestras de  $X$  en  $t_1, \dots, t_n, n \in \mathbb{N}$ , busquemos un predictor lineal no sesgado

$$\hat{X}_{t_0} = c(t_0) + \sum_{i=1}^n c_i X_{t_i}$$

en alguna otra ubicación  $t_0 \in \mathbb{R}^d$  que optimice el error cuadrático medio.

**Teorema.** Sean  $X_{t_1}, \dots, X_{t_n}$  muestras de un campo aleatorio  $(X_t)_{t \in \mathbb{R}^d}$  con media constante desconocida en  $n$  ubicaciones  $t_i \in \mathbb{R}^d, i = 1, \dots, n$ , y júntelos en el  $n$ -vector  $Z$ . Denotemos por  $\Sigma$  la matriz de covarianza de  $Z$  y asumamos que  $\Sigma$  existe y es no singular. Adicionalmente, sea  $K = (K_i)_{i=1}^n$  el vector  $n \times 1$  con entradas  $K_i = \rho(t_i, t_0)$ . Entonces

$$\hat{X}_{t_0} = K' \Sigma^{-1} Z + \frac{1 - \mathbf{1}' \Sigma^{-1} K}{\mathbf{1}' \Sigma^{-1} \mathbf{1}} \mathbf{1}' \Sigma^{-1} Z$$

es el mejor predictor lineal en términos del error cuadrático medio. El error cuadrático medio de la predicción está dado por

$$\rho(t_0, t_0) - K' \Sigma^{-1} K + \frac{(1 - \mathbf{1}' \Sigma^{-1} K)^2}{\mathbf{1}' \Sigma^{-1} \mathbf{1}}.$$

**Definición** *Kriging Universal*. El kriging universal relaja la suposición de media constante del kriging ordinario a la suposición más general de que

$$\mathbb{E}X_t = m(t)'\beta$$

para alguna función conocida  $m : \mathbb{R}^d \rightarrow \mathbb{R}^p$  y un vector de parámetros desconocido  $\beta \in \mathbb{R}^p$ . Tal modelo sería apropiado en un contexto de regresión espacial donde se considera que los valores muestreados dependen linealmente de  $p$  variables explicativas  $m(t)_i, i = 1, \dots, p$ . Un estimador lineal  $\hat{X}_{t_0} = c(t_0) + \sum_{i=1}^n c_i X_{t_i}$  es insesgado siempre que

$$m(t_0)'\beta = c(t_0) + \sum_{i=1}^n c_i m(t_i)'\beta$$

para todo  $\beta$ . Nótese que ambos lados son polinomiales en  $\beta$ . Por lo tanto, todos los coeficientes deben ser iguales. En otras palabras,  $c(t_0) = 0$  y

$$m(t_0) = \sum_{i=1}^n c_i m(t_i).$$

El kriging universal optimiza el error cuadrático medio

$$\mathbb{E} \left[ \left( \sum_{i=1}^n c_i X_{t_i} - X_{t_0} \right)^2 \right]$$

bajo la restricción anterior. Siempre que  $M'\Sigma^{-1}M$  sea no singular para la matriz  $n \times p$   $M$  cuyas filas están dadas por  $m(t_i)'$ , se puede demostrar que el vector óptimo de coeficientes lineales es

$$c = \Sigma^{-1} \left[ K + M (M'\Sigma^{-1}M)^{-1} (m(t_0) - M'\Sigma^{-1}K) \right].$$

El error cuadrático medio de la predicción correspondiente es

$$\rho(t_0, t_0) - K'\Sigma^{-1}K + (m(t_0) - M'\Sigma^{-1}K)' (M'\Sigma^{-1}M)^{-1} (m(t_0) - M'\Sigma^{-1}K).$$

## K-Folds Cross-Validation

La validación cruzada es una técnica estadística utilizada para evaluar el rendimiento predictivo de un modelo en datos no vistos. Es especialmente útil en el aprendizaje automático y la minería de datos para estimar la precisión de un modelo en términos de su capacidad para predecir datos no utilizados para el entrenamiento.

**Definición:** *K-Folds Cross-Validation* Es una variante común de la validación cruzada que divide el conjunto de datos en  $K$  pliegues (folds) aproximadamente iguales. Se entrena el modelo  $K$  veces, utilizando cada uno de los pliegues como conjunto de prueba una vez y los restantes como conjunto de entrenamiento. Esto permite obtener  $K$  medidas de rendimiento, que luego se promedian para obtener una estimación global del rendimiento del modelo (ver [3]).

## Herramientas de programación

Se emplearon dos lenguajes principales para el desarrollo del proyecto: R y Python. R fue utilizado para el análisis, predicción e implementación del modelo matemático, incluyendo la técnica de transformación Box-Cox. Por otro lado, Python se utilizó para crear una interfaz que permitiera visualizar los resultados de manera clara y explicativa. Esta interfaz se integró en una página web accesible al público mediante Streamlit.

En R, se utilizaron los siguientes paquetes:

- **MASS:** Este paquete proporciona funciones y conjuntos de datos para soporte de los métodos modernos de estadística aplicada, incluyendo la función `boxcox` para la transformación Box-Cox.
- **sp:** Proporciona clases y métodos para manipular datos espaciales en R.
- **gstat:** Este paquete es utilizado para modelar y predecir datos espaciales mediante kriging y otras técnicas de interpolación espacial.
- **caret:** Utilizado para realizar tareas de aprendizaje automático y validación cruzada de manera eficiente.

En Python, se emplearon los siguientes paquetes:

- **Streamlit:** Utilizado para crear aplicaciones web interactivas con facilidad, permitiendo la visualización de datos de manera dinámica.
- **Plotly:** Proporciona herramientas para crear gráficos interactivos y visualizaciones estadísticas.
- **Pandas:** Esencial para la manipulación y análisis de datos estructurados.
- **NumPy:** Fundamental para operaciones numéricas y cálculos científicos eficientes.

## Procedimiento y resultados principales

Para este procedimiento se emplearon las mismas herramientas y en el mismo orden que en la subsección anterior.

Inicialmente, se realizó una transformación de los datos utilizando Box-Cox. Como se mencionó en el Background, esta transformación normaliza los datos y puede mejorar las predicciones. Sin embargo, como se puede observar en la Figura 4, existe un alto porcentaje de valores 0, lo cual inicialmente dificulta el uso de esta técnica debido a la hipótesis de estricta positividad requerida por los datos. Para abordar esto, se decidió reparametrizar los datos de manera que en lugar de predecir el porcentaje de cobre, se predice el porcentaje de NO cobre, calculado restando el porcentaje original de cobre a 100. Esto asegura que se cumple

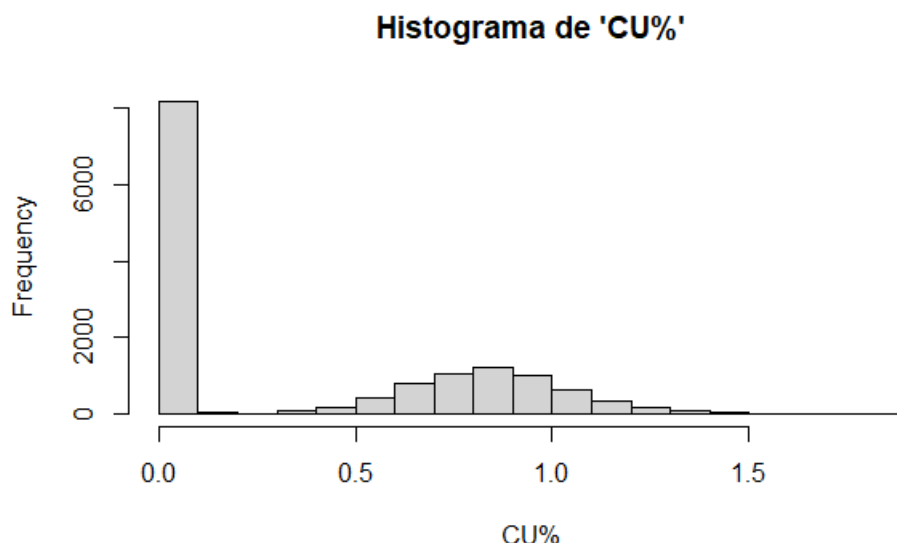


Figura 4: Histograma del porcentaje de Cobre

la hipótesis necesaria. Además, se aplicó un escalado dividiendo por 100, transformando la variable predictora a  $\frac{100-CU\%}{100}$ .

Posteriormente, se aplicó la transformación Box-Cox siguiendo el procedimiento descrito en el Background matemático. La gráfica de  $d_\lambda$  se puede observar en la Figura 6. Utilizando la técnica del codo, se identificó que un mínimo adecuado podría ser alrededor de  $\lambda = 140$ . Además, para validar este resultado, se utilizó la biblioteca MASS, la cual encontró un óptimo  $\lambda$  aproximadamente igual a  $\lambda = 145$ . Puede ver el histograma de los datos luego de la transformación en la figura 5.

Se realizó un análisis del variograma. Siguiendo la propuesta de Cressie y Hawkins, se obtuvo el variograma empírico bajo dos suposiciones: que la media es constante y que existe una tendencia en las tres coordenadas. Ambos variogramas empíricos se muestran en las figuras 7 y 8, respectivamente.

Procedimos a estimar el mejor modelo que se ajustara al variograma empírico. Evaluamos varios modelos como el Gaussiano, lineal, exponencial y esférico. Para comparar su rendimiento, calculamos la suma de errores cuadráticos (SSE) entre los puntos y el ajuste de cada modelo. Los resultados se presentan en las siguientes tablas: la Tabla 1 asumiendo una media constante, y la Tabla 2 considerando una tendencia en la media. En ambas tablas, uno de los mejores modelos fue el Gaussiano. Sus ajustes se pueden observar en las Figuras 9 y 10 para media constante y tendencia en la media, respectivamente (el ajuste de los demás modelos los puede ver en la sección del Anexo, en las figuras 14, 15, 16, 17, 18,19)

Posteriormente realizamos las estimaciones de Kriging utilizando el procedimiento descrito en el Background Matemático.

Para el *kriging simple*, utilizamos el promedio usual como media conocida y el modelo Gaus-

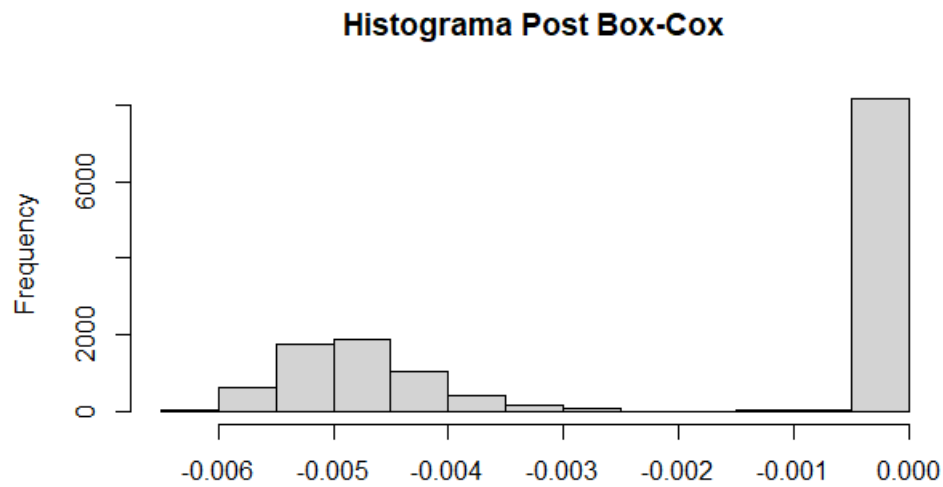


Figura 5: Histograma luego de la reparametrización y transformación Box-Cox

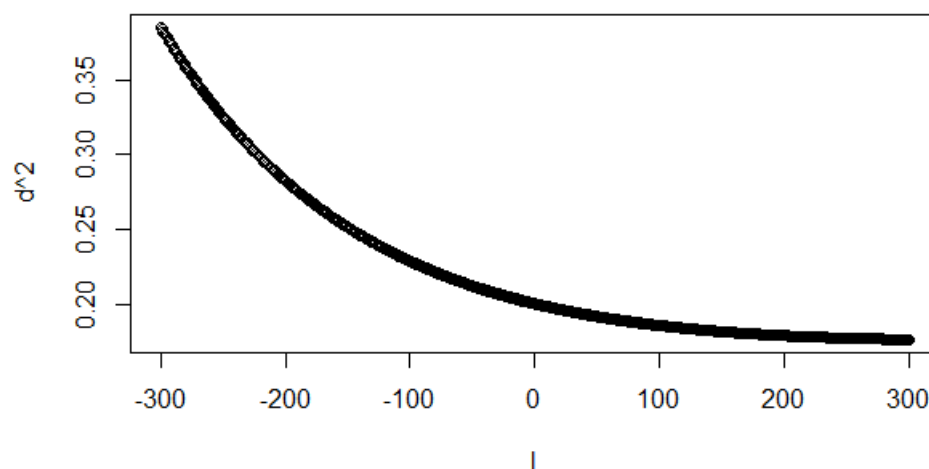


Figura 6: Grafico de  $d_\lambda$  vs  $\lambda$  para encontrar el optimo



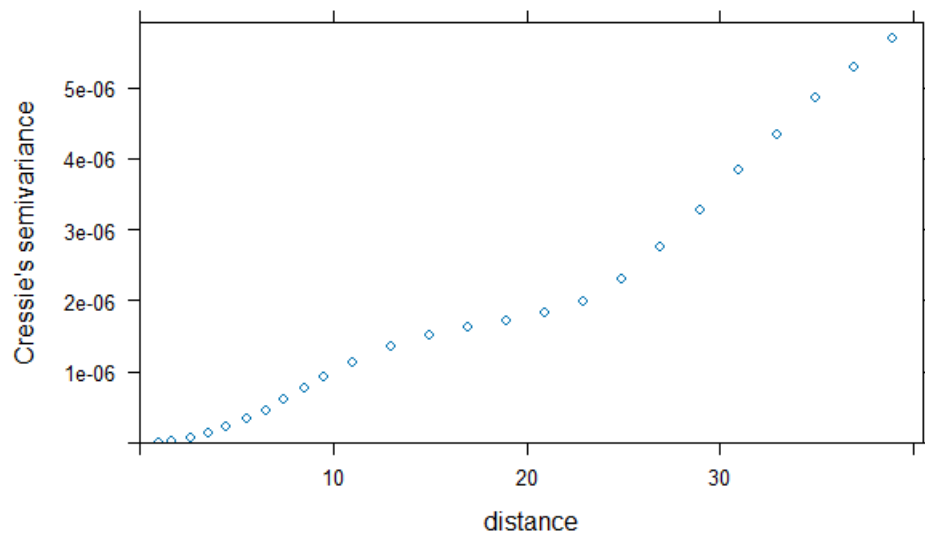


Figura 7: Variograma Empirico con media constante

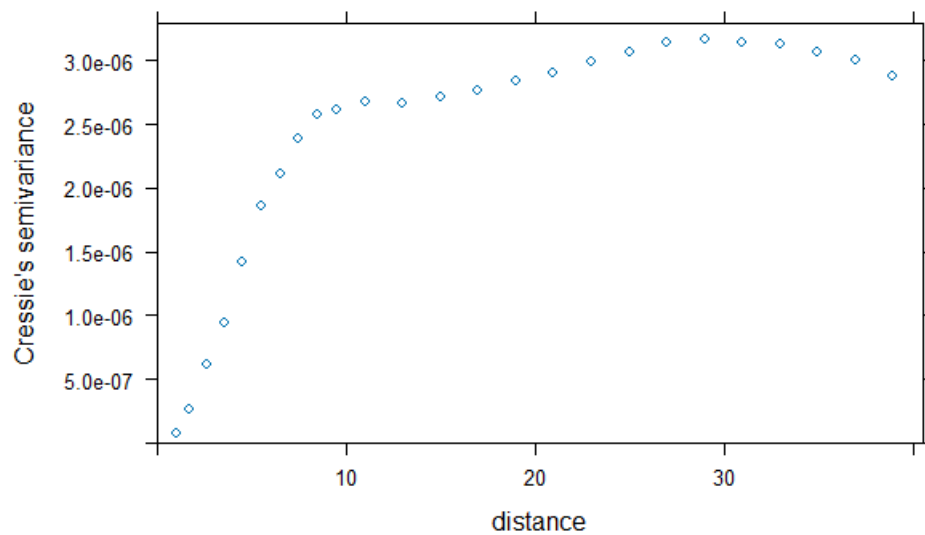


Figura 8: Variograma empirico con tendencia

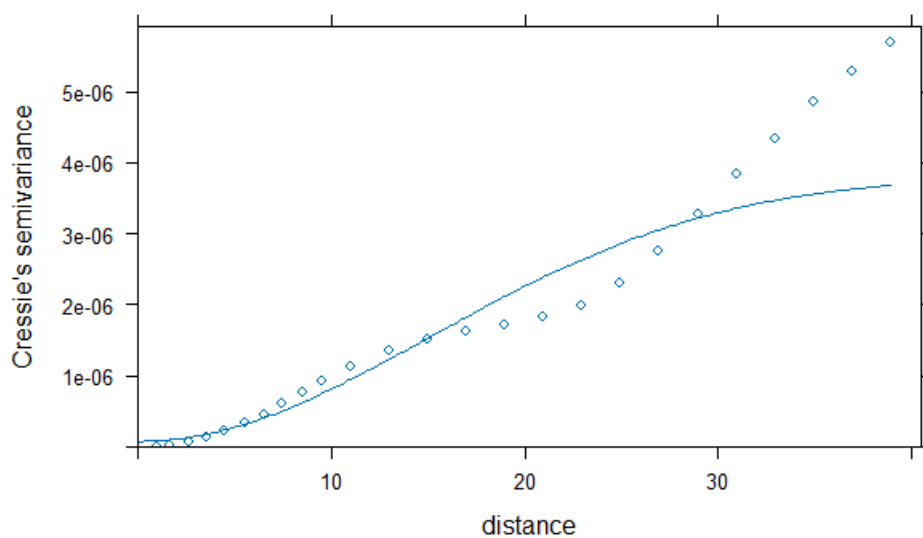


Figura 9: Estimación usando modelo de Variograma Gaussiano con media constante

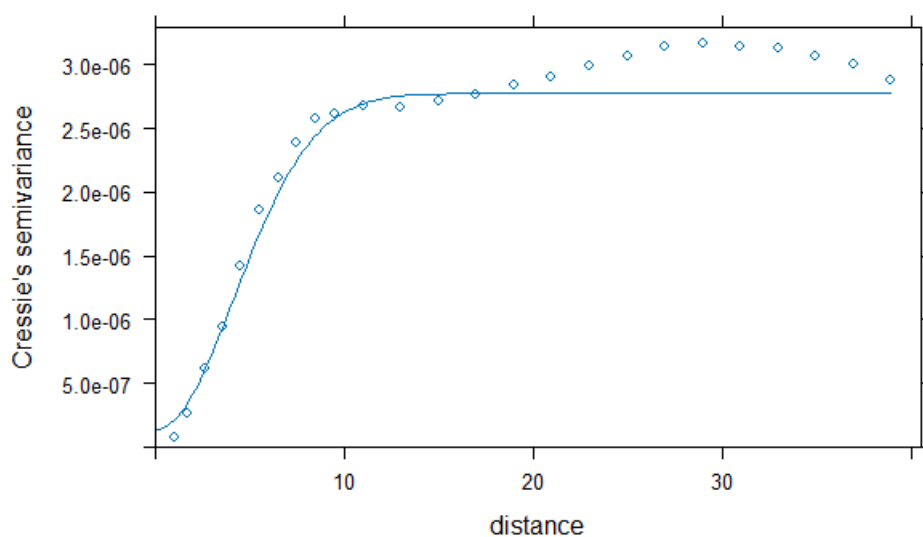


Figura 10: Estimación usando modelo de Variograma Gaussiano con tendencia en la media

Modelo	SSE
Esférico	$3,95533 \times 10^{-8}$
Exponencial	$8,264444 \times 10^{-8}$
Gaussiano	$3,984313 \times 10^{-8}$
Lineal	$3,911204 \times 10^{-8}$

Cuadro 1: Valores de SSE para diferentes modelos asumiendo una media constante

Modelo	SSE
Esférico	$2,235154 \times 10^{-8}$
Exponencial	$4,414818 \times 10^{-8}$
Gaussiano	$9,676946 \times 10^{-9}$
Lineal	$1,75465 \times 10^{-8}$

Cuadro 2: Valores de SSE para diferentes modelos asumiendo una tendencia en la media

siano para el variograma, asumiendo la hipótesis de media constante.

Para el *kriging ordinario*, estimamos la media como se explica en el Background Matemático y también utilizamos el modelo Gaussiano para el variograma, bajo la hipótesis de media constante que caracteriza este método.

Para el *kriging universal*, realizamos la estimación considerando una posible tendencia detectada visualmente en los datos. Esto también se observó mediante regresiones lineales en cada coordenada, cuyos resultados se detallan en las Tablas 3, 4 y 5. Las regresiones indicaron una tendencia significativa, lo que cuestiona la suposición de media constante en los modelos de kriging anteriores. Esto se refleja en el valor muy bajo del p-valor en la última columna de las tablas.

Se compararon las predicciones realizadas por los tres métodos de kriging. En cada caso, se predijo el cubo tridimensional más pequeño que contiene por completo el conjunto de datos de entrenamiento, resultando en más de 60,000 predicciones. Como se observa en la figura 11, las predicciones obtenidas mediante Kriging Simple y Kriging Ordinario son prácticamente iguales y muestran un comportamiento muy similar. Aunque las predicciones del Kriging Universal no se alejan demasiado, tienden a predecir valores más positivos o nulos. Las alternativas de Kriging Simple y Kriging Ordinario tienden a no predecir valores de 0. En las figuras tridimensionales 12 y 13, se puede ver el resultado de las predicciones tanto para Kriging Simple como para Kriging Universal.

Para los tres modelos matemáticos, realizamos validación cruzada utilizando K-fold con K igual a 10. Este número de folds es comúnmente utilizado en la literatura y adecuado dado el tamaño de nuestro conjunto de datos (aproximadamente 14000 observaciones), lo cual permite mantener suficientes datos tanto para el entrenamiento como para la evaluación. En cada iteración del K-fold, se realizaron las estimaciones de variograma y kriging correspondientes, comparándolas con las predicciones reales y calculando los errores. Posteriormente, se calcularon las métricas de Root Mean Squared Error (RMSE) y Mean Absolute Error (MAE) para cada modelo. El RMSE es sensible a los errores grandes debido a la operación

Comparación de CU% entre diferentes métodos de Kriging

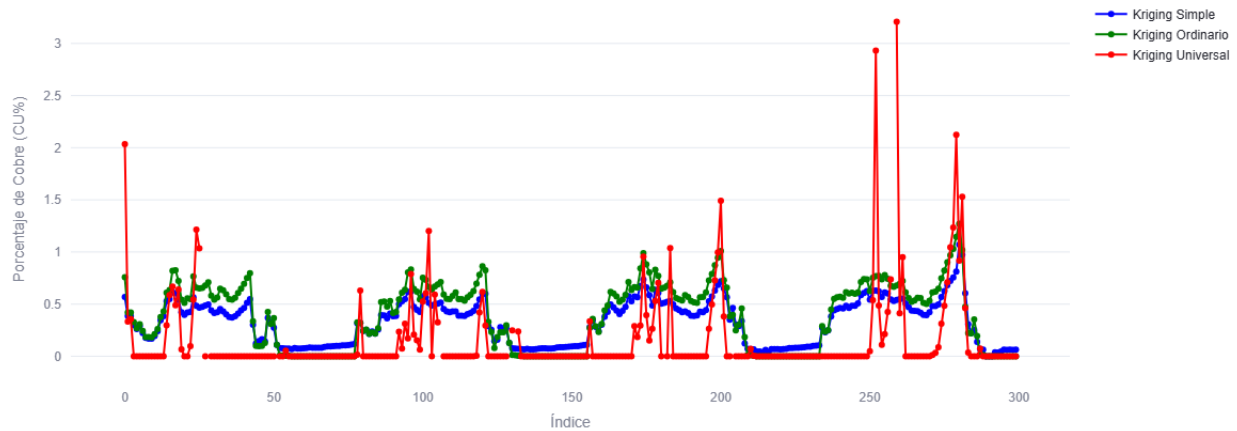


Figura 11: Comparación de las primeras 300 predicciones entre los tres métodos de Kriging.

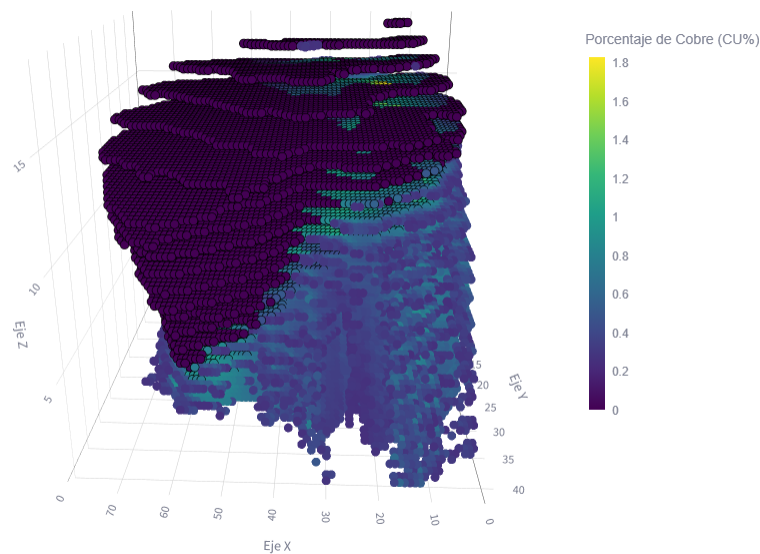


Figura 12: Predicción de Kriging Simple, con umbral 0.25. Aquellos puntos que no aparecen en el plot, es porque la predicción está debajo del umbral

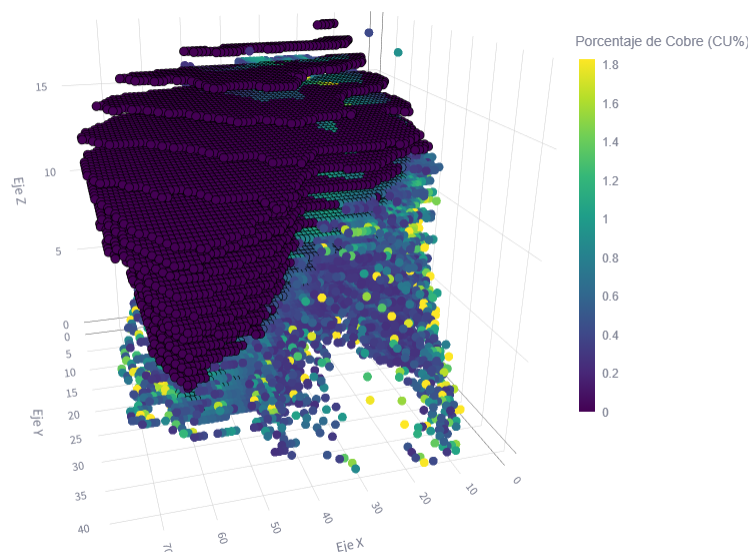


Figura 13: Predicción de Kriging Universal, con umbral 0.25. Aquellos puntos que no aparecen en el plot, es porque la predicción está debajo del umbral

de elevación al cuadrado, mientras que el MAE proporciona una medida lineal de los errores sin dar un peso adicional a los errores más grandes. Estas métricas para cada caso se resumen en la Tabla 6. La cual entrega que *kriging universal* es el mejor modelo para nuestros datos, ya que presenta los menores valores de RMSE y MAE. Esto sugiere que tener en cuenta la tendencia en los datos mejora la precisión de las predicciones.

Todo lo anterior descrito se llevó a cabo en R utilizando diversos paquetes mencionados en secciones anteriores. Los gráficos de los resultados de predicción, así como otros gráficos utilizados, fueron implementados en Python. Puede revisar los códigos en la página de Github en la cual están almacenados. Además, se desarrolló una página web disponible para el público utilizando el paquete Streamlit, que permite interactuar con los resultados. Puede acceder a ella mediante este link [Kd-Analysis](#).

La razón principal para implementar estos gráficos en Python y ofrecer una interfaz interactiva se debe a la naturaleza tridimensional de las predicciones realizadas. Aunque las proyecciones bidimensionales son útiles para visualizar los datos, un gráfico interactivo tridimensional proporciona una perspectiva más completa y detallada. Esto facilita la exploración de las relaciones espaciales y la comprensión de la distribución de las variables en el contexto de las predicciones espaciales realizadas mediante Kriging.

	Estimate	Std. Error	t value	$\Pr(>  t )$
(Intercept)	$-4,641 \times 10^{-3}$	$4,326 \times 10^{-5}$	-107.3	$< 2 \times 10^{-16}$
x	$6,308 \times 10^{-5}$	$9,487 \times 10^{-7}$	66.5	$< 2 \times 10^{-16}$

Cuadro 3: Regresión respecto a la coordenada  $x$ .

	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	$-4,310 \times 10^{-3}$	$5,114 \times 10^{-5}$	-84.28	$< 2 \times 10^{-16}$
y	$1,208 \times 10^{-4}$	$2,504 \times 10^{-6}$	48.26	$< 2 \times 10^{-16}$

Cuadro 4: Regresión respecto a la coordenada  $y$ .

	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	$-4,427 \times 10^{-3}$	$6,255 \times 10^{-5}$	-70.77	$< 2 \times 10^{-16}$
z	$2,264 \times 10^{-4}$	$5,586 \times 10^{-6}$	40.53	$< 2 \times 10^{-16}$

Cuadro 5: Regresión respecto a la coordenada  $z$ .

Modelo	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)
Kriging Simple	0.0006955978	0.0003329818
Kriging Ordinario	0.0006966028	0.0003321576
Kriging Universal	0.0006768341	0.0002693542

Cuadro 6: Comparación de errores RMSE y MAE para diferentes métodos de Kriging.

## Conclusiones, preguntas abiertas y potencial trabajo a futuro

En este estudio, hemos explorado exhaustivamente técnicas estadísticas y herramientas computacionales avanzadas para el análisis y predicción espacial del contenido de cobre en datos geológicos. Nuestro enfoque se centró en la optimización de predicciones precisas mediante técnicas como la transformación Box-Cox, modelado de variogramas y métodos de kriging, respaldados por una sólida fundamentación matemática y el uso de herramientas de programación especializadas.

Inicialmente, se quería realizar una transformación Box-Cox, pero identificamos la necesidad de reparametrizar los datos de manera adecuada para cumplir con los requisitos de hipótesis, especialmente debido a la presencia de ceros en los datos. La transformación Box-Cox demostró ser efectiva, permitiendo la aplicación de kriging con un modelo de variograma bien ajustado, como lo confirmaron los análisis de variogramas empíricos y la validación cruzada mediante K-fold.

Implementamos y evaluamos tres métodos de kriging: Kriging Simple, Kriging Ordinario y Kriging Universal. Observamos que todos estos métodos ofrecieron predicciones altamente precisas, con errores (RMSE y MAE) comparativamente bajos. Es importante destacar que, aunque el error fue pequeño, este resultado también se debe al reescalado de los datos, que normalizó los valores a una escala de 0 a 1.

El modelo de Kriging Universal, que incorpora la tendencia espacial detectada mediante regresión lineal, mostró una ligera mejora en la precisión de las predicciones en comparación con Kriging Simple y Ordinario. Sin embargo, los tres métodos demostraron ser robustos y adecuados para la estimación de variables espaciales en este contexto geológico específico.

Además, desarrollamos herramientas computacionales en R y Python para el análisis de datos y la visualización interactiva, facilitando la replicabilidad de nuestros resultados y la exploración detallada de las relaciones espaciales. Estas herramientas fueron documentadas y compartidas públicamente en un repositorio en GitHub, así como a través de una interfaz web interactiva desarrollada con Streamlit disponible a todo público.

Algunas preguntas abiertas que se plantearon durante este trabajo fueron ¿Cómo afectaría el uso de otras transformaciones de datos, además de Box-Cox, la precisión de las predicciones en el contexto geológico? ¿Existen transformaciones más adecuadas para datos con muchos ceros? , ¿Qué impacto tendrían modelos de variograma alternativos (por ejemplo, matemáticamente más complejos o adaptados específicamente a variabilidades geológicas) en la precisión de las predicciones de kriging? ¿Cómo se podría mejorar la precisión de las estimaciones considerando datos adicionales, como información geofísica o geoquímica, y su integración con datos geológicos para un modelado más completo de la distribución de recursos minerales?

Se propone como un futuro trabajo, un análisis, procesando los datos mineros utilizando el concepto de 'tubos'. Los 'tubos' se refieren a las muestras extraídas verticalmente desde la superficie hacia abajo en coordenadas específicas ( $x$ ,  $y$ ). Estas muestras contienen información valiosa sobre la distribución del mineral a lo largo de la profundidad ( $z$ ). Al entrenar nuestro modelo de Kriging usando la propuesta anterior, seguiremos el orden natural en que se obtuvieron los datos, es decir, respetando la secuencia de las muestras en los 'tubos'. Esto es crucial porque la correlación espacial en la minería a menudo se presenta en la dirección vertical, y nuestro modelo debe capturar esta relación para hacer predicciones precisas.

## Bibliografía

- [1] George EP Box y David R Cox. «An analysis of transformations». En: *Journal of the Royal Statistical Society: Series B (Methodological)* 26.2 (1964), págs. 211-252.
- [2] Noel Cressie y Douglas M. Hawkins. «Robust Estimation of the Variogram: I». En: *Mathematical Geology* 12.2 (1980), págs. 115-125.
- [3] Trevor Hastie, Robert Tibshirani y Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2009.
- [4] Tomás Lagos et al. «A framework for adaptive open-pit mining planning under geological uncertainty». En: *Optimization and Engineering* (2020), págs. 1-36.
- [5] Maria Nicolette Margaretha van Lieshout. *Theory of spatial statistics: a concise introduction*. Chapman y Hall/CRC, 2019.
- [6] R M Sakia. «The Box-Cox transformation technique: A review». En: *The Statistician* 41.2 (1992), págs. 169-178.
- [7] Héctor Daniel Vega Ulloa et al. «Diseño de mallas de muestreo en presencia de anisotropías.» En: (2021).



## Anexo

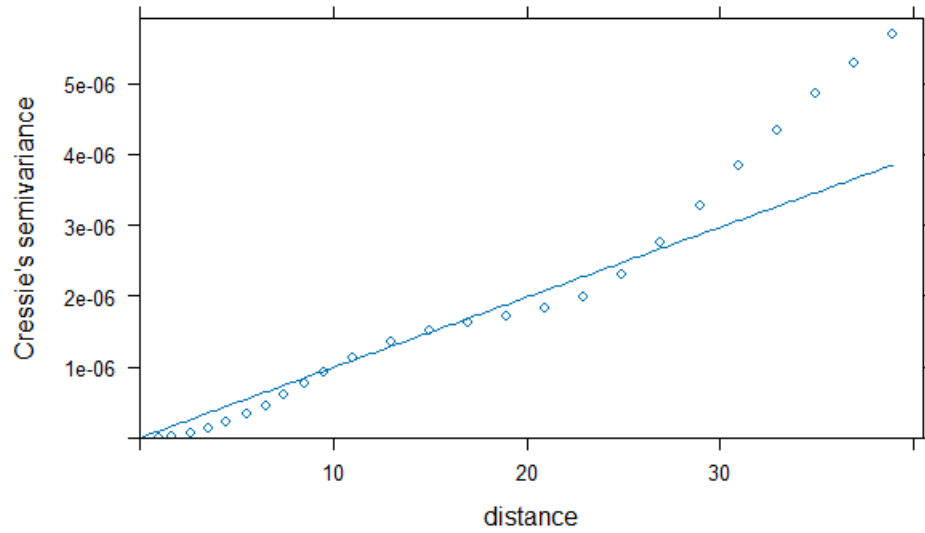


Figura 14: Estimación del Variograma usando un Modelo Esferico, con media constante

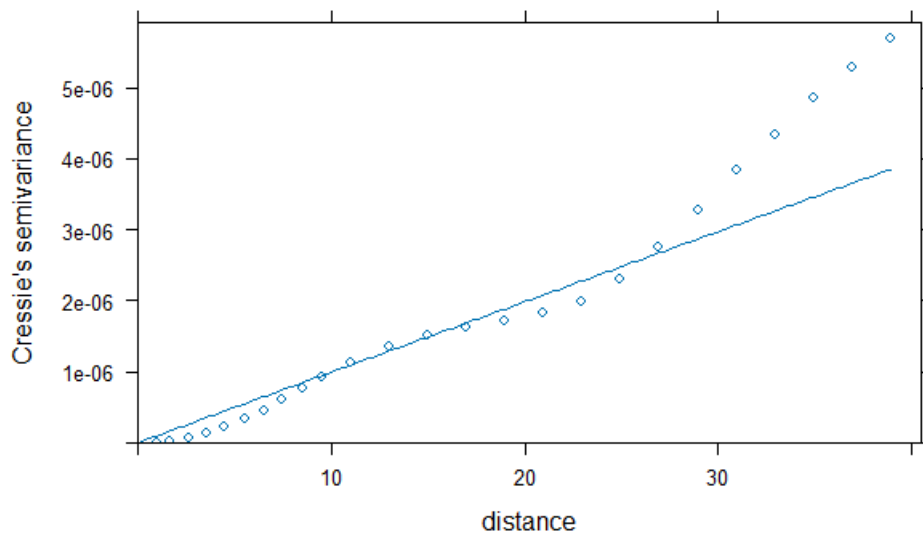


Figura 15: Estimación del Variograma usando un Modelo Esferico, con tendencia en la media

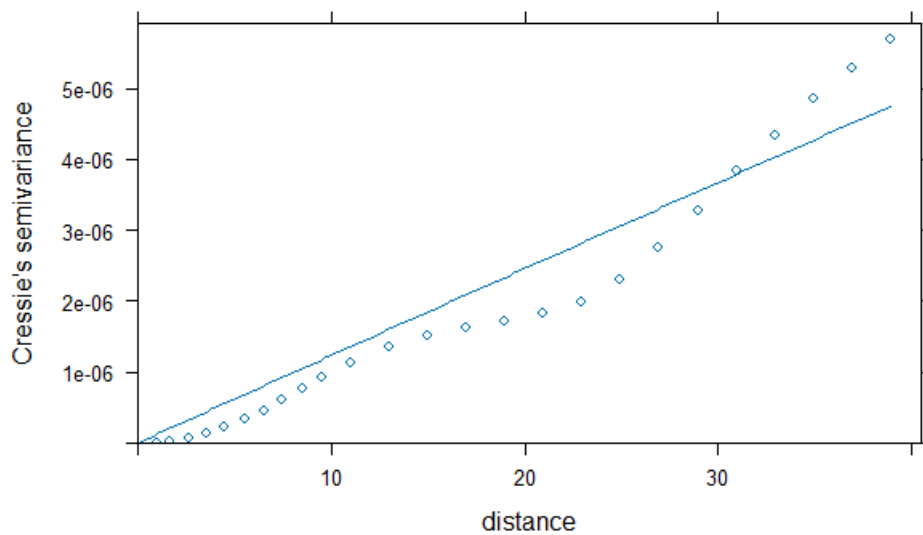


Figura 16: Estimación del Variograma usando un Modelo Exponencial, con media constante

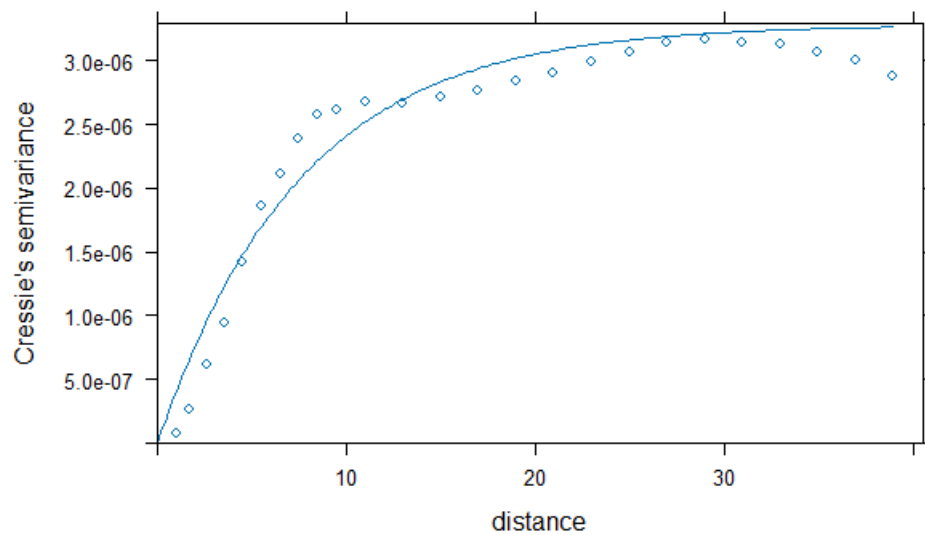


Figura 17: Estimación del Variograma usando un Modelo Exponencial, con tendencia en la media

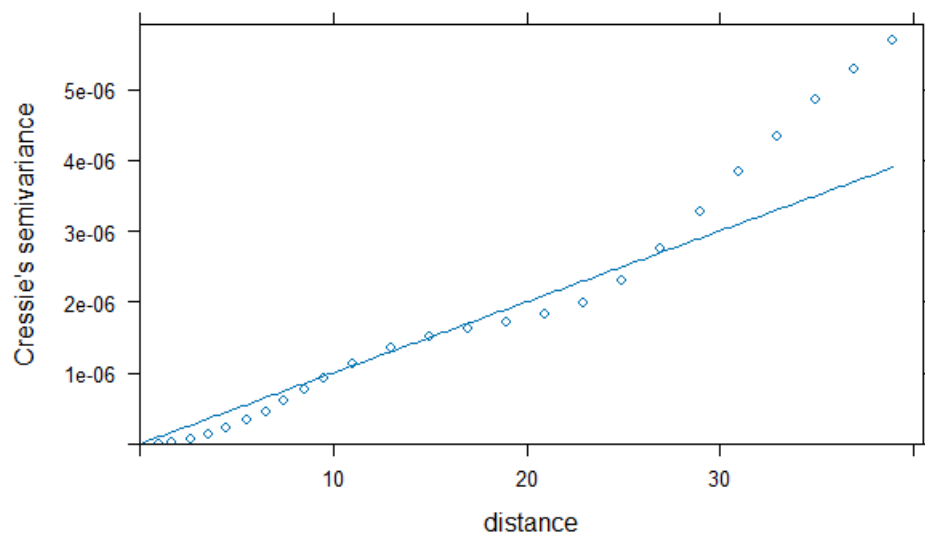


Figura 18: Estimación del Variograma usando un Modelo lineal, con media constante

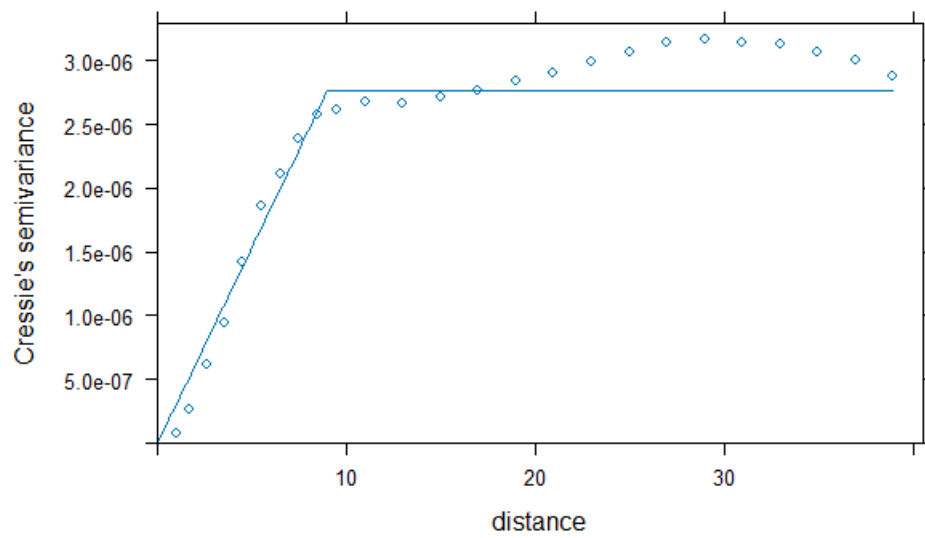


Figura 19: Estimación del Variograma usando un Modelo Lineal, con tendencia en la media