



Maestría Profesional en Estadística

Curso: SP-1649 Tópicos de Estadística Espacial Aplicada

Profesora: Dra. Marcela Alfaro Córdoba

## ANÁLISIS DE COMPONENTE GEOESPACIAL, EN LA CONCENTRACIÓN DE ZINC EN EL SUELO, SUROESTE DE ALASKA

Elaborado por:

María Elena Zúñiga Brenes

Alejandro Zúñiga Madrigal

- 07 de noviembre de 2019 -



## CONTENIDO

---

1	INTRODUCCIÓN	3
2	ANTECEDENTES Y JUSTIFICACIÓN	3
2.1	Objetivos de la investigación	4
3	ASPECTOS METODOLÓGICOS	4
3.2	Fórmulas utilizadas	4
4	RESULTADOS Y DISCUSIÓN	7
5	CONCLUSIÓN	12
6	BIBLIOGRAFÍA	13

## 1 INTRODUCCIÓN

---

En este trabajo se realiza un análisis estadístico con enfoque geoespacial, tomando como base de datos la concentración de metales pesados Zinc en el suroeste de Alaska.

En colaboración con otros organismos federales y estatales, la industria y el mundo académico, el Servicio Geológico de los Estados Unidos (USGS), como parte de su Programa de Estudios de Recursos Minerales, llevó a cabo el Estudio Geoquímico Nacional (NGS) donde se recopiló la información entre 2003 y 2008.

La intención del trabajo es explorar y analizar un aspecto geoespacial asociado a la concentración de los metales pesados en el suelo, especialmente del Zinc, planteando las siguientes preguntas de interés:

- ¿Existe estructura de correlación espacial en la concentración del Zinc en el Suroeste de Alaska?
- ¿Cuál modelo lineal se puede ajustar para estudiar la concentración de Zinc en el Suroeste de Alaska?

Este trabajo establece como propuesta la pregunta de investigación para lograr con ello identificar si existe distribución espacial y cual modelo lineal puede estudiar mejor la concentración de Zinc.

## 2 ANTECEDENTES Y JUSTIFICACIÓN

---

La mayoría de los fenómenos naturales que se estudian se pueden describir mediante variables regionalizadas, tanto en el espacio como en el tiempo. La variabilidad es el resultado de procesos naturales, por lo tanto es determinista. Pero la mayoría de estos procesos son muy sensibles y las condiciones en las que tienen lugar no se conocen, basándose en leyes físicas y químicas no es posible describirlos por completo (Bárdossy 2001).

La geoestadística primero fue utilizada por la industria minera, en la cual, dado que los costes de las perforaciones eran altos, el análisis de los datos fue de suma importancia. El modelamiento de variables medidas en diferentes sitios de una región con continuidad espacial y que presentan alguna estructura de correlación espacial, ha sido desarrollada desde los años sesenta, con el desarrollo de los análisis geoestadísticos. (Melo, 2012).



Lo anterior permitió el avance en las ciencias del suelo que estudia este como recurso natural en la superficie de la Tierra incluyendo su formación (pedogénesis), su clasificación y cartografía, sus características, su composición química, su biología, su fertilidad y todo lo relacionado con el uso y a la gestión del suelo. Esta ciencia agrupa a la pedología (formación, química, morfología y clasificación del suelo) y la edafología (influencia del suelo en organismos, especialmente las plantas) (Buol, 1973).

La extracción de los numerosos recursos naturales de Alaska representa un sector importante para la economía del estado, especialmente lo que respecta a la minería. Históricamente ha sido una industria clave, y la mayoría de infraestructura fue construida originalmente con la intención de mejorar el acceso a puntos de extracción. Incluye la exploración, desarrollo y producción de oro, plata, carbón, cobre, plomo y zinc. Los minerales representan la segunda partida de exportación más alta (36% del total de exportaciones), por detrás del petróleo (OEC, 2018).

### 3 OBJETIVOS DE LA INVESTIGACIÓN

---

Para este estudio se establecieron los siguientes objetivos de investigación:

- Identificar si existe distribución espacial en la concentración de Zinc en el Suroeste de Alaska.
- Identificar el modelo lineal espacial que se ajuste mejor a la concentración de Zinc en el Suroeste de Alaska.

### 4 ASPECTOS METODOLÓGICOS

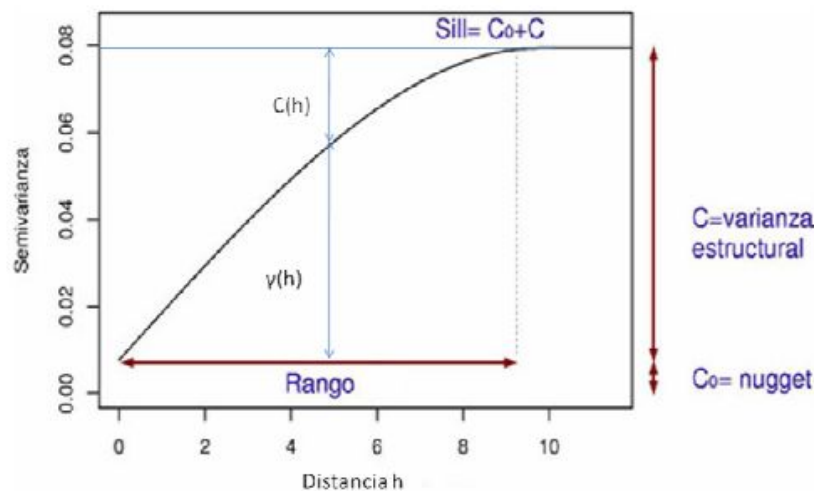
---

#### 4.1 FÓRMULAS UTILIZADAS

##### **Semivariograma**

La manera más usual de representar la dependencia espacial de una variable es a través de su función de covarianza o mediante la semivarianza ya que las dos contienen la misma información. En las aplicaciones geoestadísticas es más común trabajar con la semivarianza, para ello hay diversas razones. De manera general, la ventaja de utilizar semivarianzas frente a covarianzas reside en el hecho de que, para evaluar la semivarianza

no es necesario conocer el valor de la media, sin embargo, si lo es para la covarianza, lo que introduce un factor de sesgo. Además si la función es no estacionaria, la varianza no es finita, por lo que la covarianza no está definida en el origen, mientras que la semivarianza, por definición, debe ser nula en el origen. El semivariograma, además de darnos a conocer la estructura de dependencia espacial de la variable aleatoria, nos proporciona información acerca del comportamiento de la misma. Para una variable aleatoria estacionaria de segundo orden, el semivariograma tiene una forma característica como la mostrada en la siguiente figura.



**Figura 1.** Representación gráfica de la semivarianza o semivariograma

### Kriging ordinario

La componente de la tendencia puede fluctuar suavemente en el vecindario de los datos que intervienen en cada estimación. De manera que supone un valor de la media constante pero desconocido. El modelo lineal planteado es similar al usado para kriging simple pero el valor del parámetro  $\beta_0$  es desconocido  $Z(x) = \beta_0 + \varepsilon(x)$ . En consecuencia se hacen pocas suposiciones, por lo que el método es muy flexible en este aspecto pero menos poderoso que otros.

Si tenemos una variable estacionaria con covarianza conocida el proceso de obtención de las ecuaciones es similar al anterior. Partiendo del estimador:

$$Z^* = \sum_{i=1}^N \lambda_{(i)} Z_{(i)}$$

el hecho de no conocer la media obliga a utilizar el requisito de sesgo nulo

$E(Z^*) = E(Z)$  que da lugar a  $E\left(\sum_i \lambda_i Z_i\right) = E(Z) = \mu$ , o bien,  $\sum_i \lambda_i E(Z_i) = \sum_i \lambda_i \mu = \mu$  donde se obtiene  $\sum_i \lambda_i = 1$  condición necesaria para asegurar que el estimador es insesgado.

Haciendo mínima la varianza de estimación se obtienen el sistema de ecuaciones para los coeficientes de ponderación y con estos, el estimador y su varianza de estimación para kriging ordinario.

$$Z^* = \sum_{i=1}^N \lambda_{(i)} Z_{(i)}$$

$$Var(Z^* - Z) = Var(Z) - \sum_i \lambda_i C(x_i - x) + \mu$$

Es la formulación más común del método kriging, válida tanto para variables estacionarias como intrínsecas, de hecho mediante la ecuación:

$$\gamma(h) = Var(Z) - C(h) = C(0) - C(h)$$

que expresa la relación entre semivarianza y covarianza las dos ecuaciones de varianzas son iguales. (Delgado, 2017)

## Software

Para el procesamiento de la información se utiliza el software estadístico R Studio versión 3.6. Además se utilizan las librerías “sp”, “sf”, “gstat” y “raster”.

## Datos utilizados

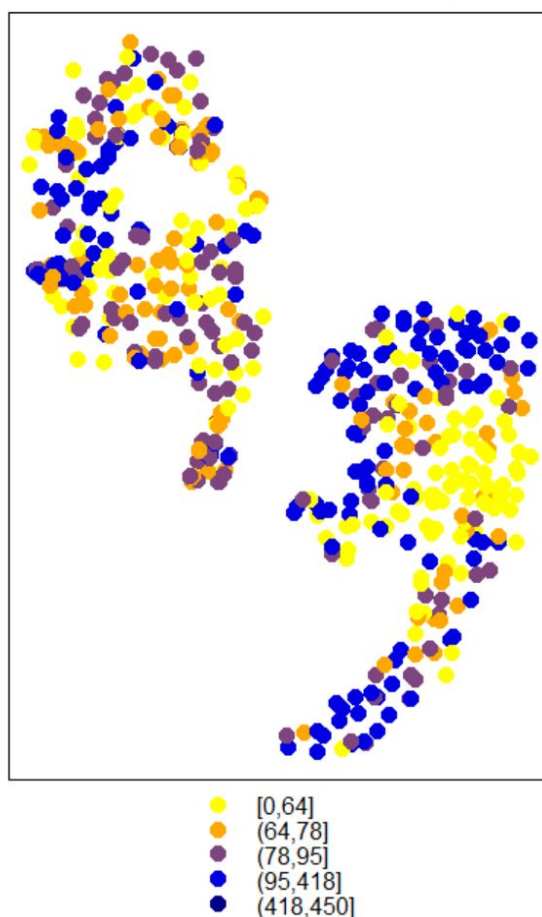
Para este estudio se tomó un subconjunto de 652 muestras de la base de datos de la Encuesta Nacional de Geoquímica (NGS) de EE.UU, obtenidas durante el período comprendido entre 2003 y 2008, estos puntos corresponden al Suroeste de Alaska. Esta base contiene ocho metales pesados, sin embargo, para este estudio se utilizó la información del Zinc (zn).

## Análisis

Primero se lleva a cabo un análisis descriptivo de la concentración de Zinc en el Suroeste de Alaska, posteriormente se elabora un semivariograma de los datos y se ajustan cuatro distribuciones teóricas (exponencial, Gaussiana, Mat y Sph). También se elaboró un kriging ordinario sobre una grilla de 200 a 300.

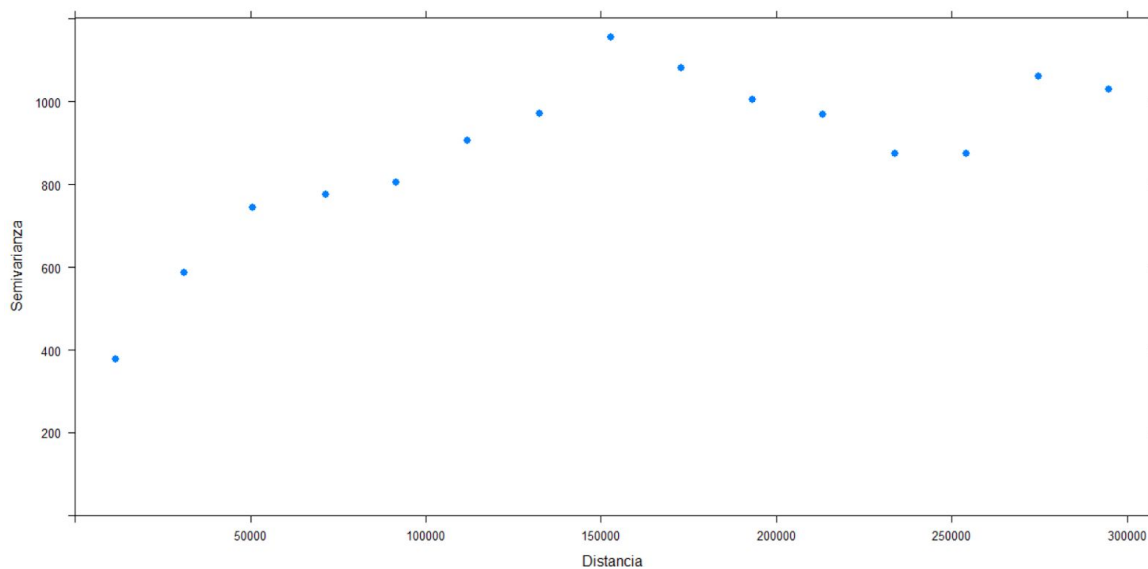
## 5 RESULTADOS Y DISCUSIÓN

La figura 2 permite observar las muestras recolectadas en el suroeste de Alaska y la distribución del zinc de las mismas de acuerdo a los quintiles. Se aprecia que existen dos agrupaciones en los datos, una al noroeste de la región y la otra al sureste, y se puede observar grupos de acuerdo a los diferentes niveles de zinc medidos.



**Figura 2.** Suroeste de Alaska: Concentración de Zinc por Quintiles

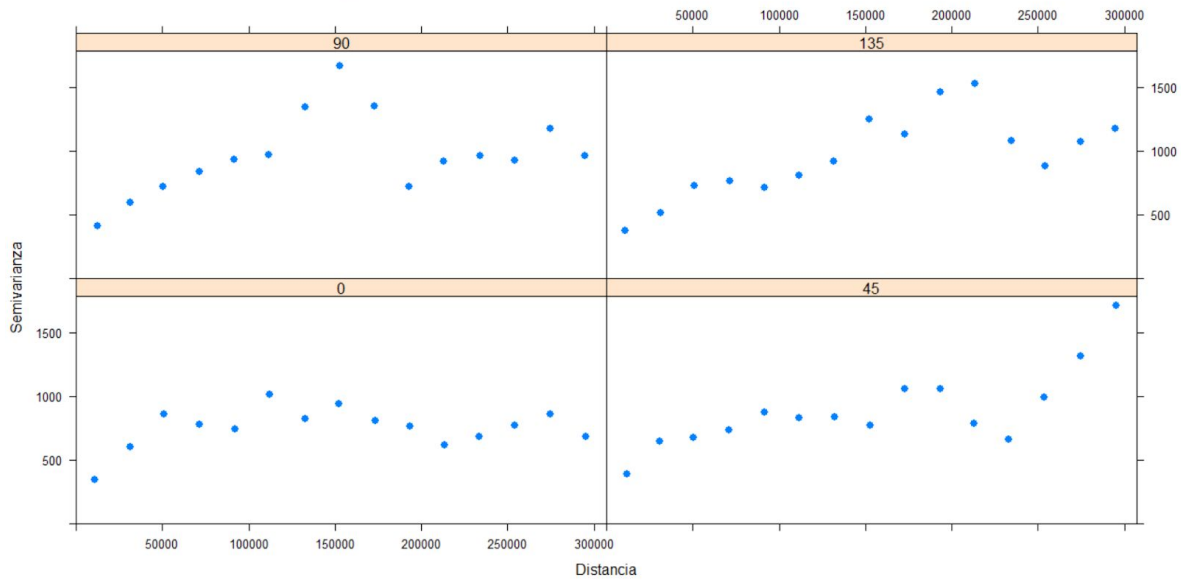
A partir de los datos anteriores, se lleva a cabo el variograma con el fin de analizar si la distribución del zinc posee correlación espacial. La figura 3 muestra el semivariograma para una distancia de 300 kilómetros, y se aprecian oscilaciones, pues en el rango que va de aproximadamente 75 kilómetros hasta los 150 kilómetros se aprecia un comportamiento creciente con concavidad negativa, y después de esta distancia, se observa un comportamiento decreciente; pero este último comportamiento es esperable a mayores distancia.



**Figura 3.** Suroeste de Alaska: Semivariograma de la Concentración del Zinc

El mismo ejercicio se lleva a cabo en dirección a los diferentes puntos cardinales, la figura 4 muestra los resultados obtenidos, y se aprecia que en dirección hacia el norte y el sur el comportamiento es bastante similar al semivariograma de la figura 3. Para el este y el oeste se puede observar que los semivariogramas parecen tender a ser constantes. Pese a esto se semivariograma en las cuatro direcciones es parecido a los otros; por ello, con la figura 3 se asume que los datos poseen estructura espacial y que se pueden modelar con técnicas geoestadísticas. Mientras que a partir de la figura 4 se asume el supuesto de isotropía.



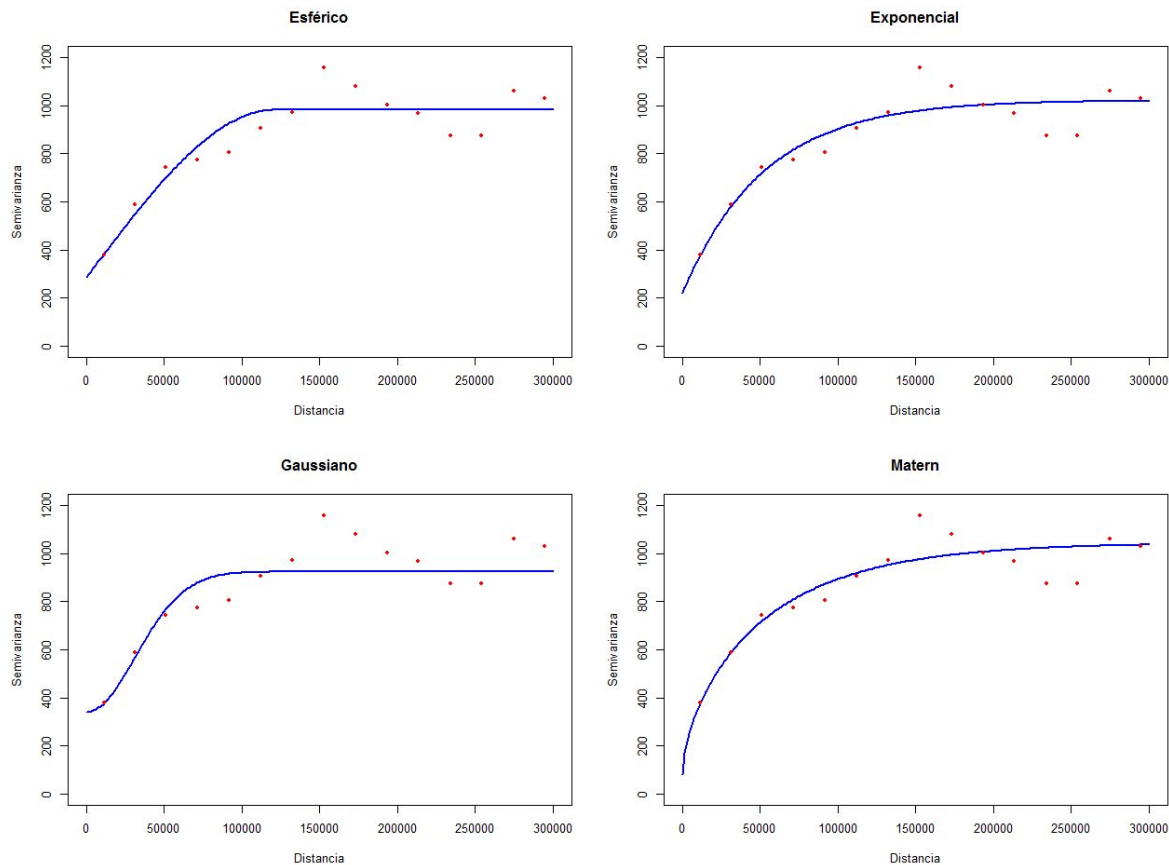


**Figura 4.** Suroeste de Alaska: Semivariograma de la Concentración de Zinc en dirección de los Puntos Cardinales.

Al considerarse la estructura espacial en las medidas del Zinc, se procede a buscar cuál distribución permite ajustar mejor el variograma obtenido en la figura 3. Para ello se comparan las distribuciones Exponencial, Gaussiana, Esférica y la Matern. La figura 5 muestra los resultados obtenidos, y se observan diferencias en los ajustes.

El modelo esférico muestra oscilaciones alrededor de la curva teórica a partir de aproximadamente los 50 kilómetros de distancia. Para este modelo el valor del rango es de 120724,5 metros, o lo que es lo mismo, 120,7 kilómetros aproximadamente. Por su parte el nugget estimado es de 285. El modelo Gaussiano presenta oscilaciones similares a las del modelo esférico; a partir de aproximadamente 50 kilómetros de distancia. El valor del rango es de aproximadamente 44,4 kilómetros, y el del nugget es de 341.

Por otro lado, los modelos Exponencial y Matern muestran bastantes similitudes en sus curvas y el ajuste de los datos. Para ambos casos se puede observar oscilaciones pronunciadas de los datos después de aproximadamente unos 144 kilómetros de distancia; pero de igual manera para ambos casos se ve un buen ajuste de los últimos pares de datos.



**Figura 5.** Suroeste de Alaska: Semivariograma de la Concentración de Zinc utilizando cuatro Distribuciones Exponencial, Gaussiana, Esférica y Matern.

Para el modelo Exponencial, el valor del rango es de 52,2 kilómetros, y valor del nugget es de 219. Para el modelo Matern estos valores son de 73,3 kilómetros y 81,4 respectivamente.

Para determinar cuál de los modelos anteriores presenta el mejor ajuste, se utiliza la función del software estadístico R llamada “fit.variogram”, la cual evalúa los diferentes modelos y permite determinar aquel con el mejor ajuste. Esta función arroja que el modelo Matern es el seleccionado como el mejor de los cuatro, por ende, es el modelo que presenta el mejor ajuste en el estudio de la distribución del Zinc en el suroeste de Alaska.

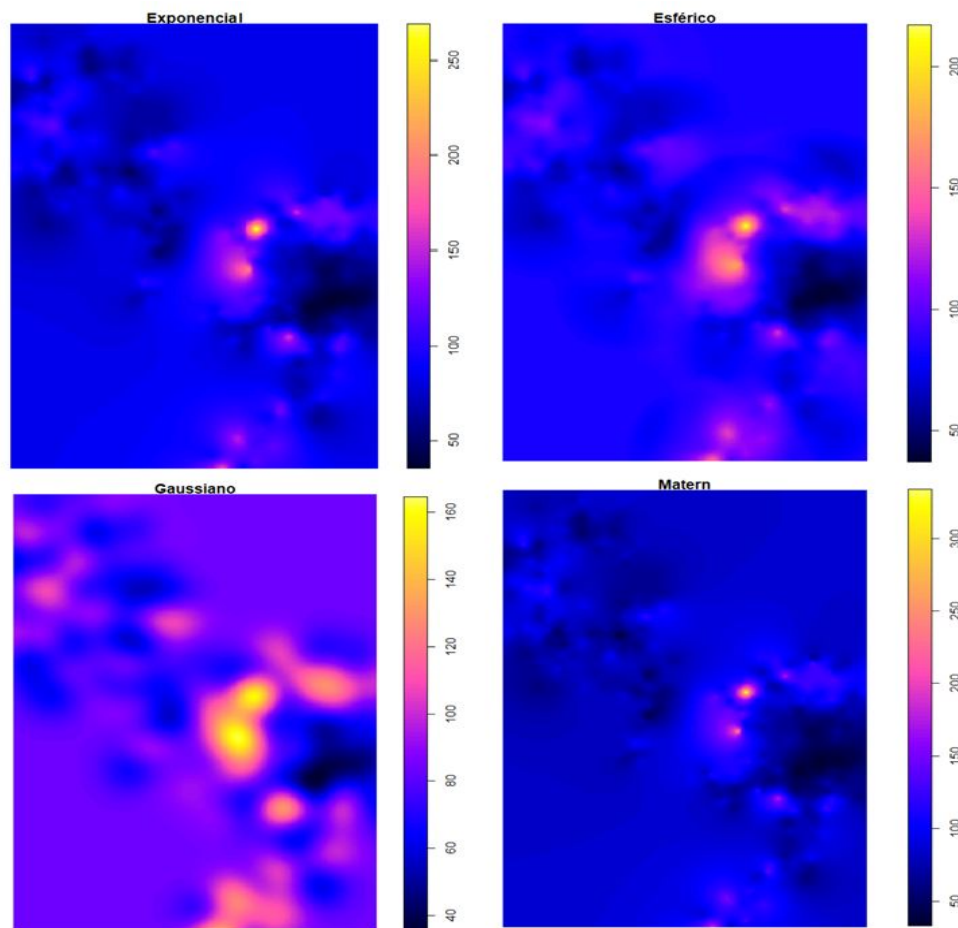
A partir de los 652 puntos muestrales, se construye una grilla de 300 filas y 200 columnas para realizar un kriging ordinario con los cuatro modelos anteriores y poder visibilizar la manera en que estos modelos se aproximan al kriging mediante el modelaje exponencial.

La figura 6 muestra los resultados del kriging ordinario mediante los cuatro modelos en cuestión. Se aprecia que para todos los métodos, aquellos lugares en lo que se no se tiene

muestra, es decir, el noreste y suroeste del área, las estimaciones hechas por todos los modelos asignan valores constantes a estos lugares.

También se aprecia que el modelo Gaussiano tiende a dar estimaciones más bajas de las concentraciones de Zinc en comparación a los otros tres modelos (ver valores de los ejes), además de ser estimaciones más homogéneas, es decir, más parecidas para las diferentes grillas de estudio.

El modelo Esférico tiende a reducir esta homogeneidad de las estimaciones en comparación con las estimaciones hechas por el modelo Gaussiano; pero estas siguen siendo más homogéneas que las estimaciones hechas por el modelo Exponencial y Matern.



**Figura 6.** Suroeste de Alaska: Resultados de kriging ordinario de los Modelos Exponencial, Gaussiana, Esférica y Matern, para la concentración de Zinc.

Estos dos últimos modelos permiten una mayor heterogeneidad en las estimaciones hechas en cada una de las grillas construidas anteriormente; lo que muestra la capacidad

que tienen estos modelos de ajustar los datos para las concentraciones de Zinc en el suroeste de Alaska, lo cual concuerda con la similitud de sus respectivos variogramas; y gracias a la función “fit.variogram” se sabe y concluye que el modelo que presenta el mejor ajuste es el Matern.

## 6 CONCLUSIÓN

---

Al respecto del análisis realizado para este estudio se puede mencionar lo siguiente:

- Con los análisis realizados con los puntos geostadísticos se logró demostrar la existencia una distribución espacial en la concentración de Zinc en el Suroeste de Alaska, con instrumentos como el semivariograma se logra explicar esta situación con respecto a la concentración de zinc en el suelo.
- Con respecto a identificar el modelo lineal espacial que se ajuste mejor a la concentración de Zinc se elaboraron cuatro modelos de los cuales el modelo Matern resultó ser el que mejor se ajustó.

## 7 BIBLIOGRAFÍA

---

- Alfaro M. (2019). Notas clase del curso Tópicos de Estadística Espacial.  
<https://mediacionvirtual.ucr.ac.cr>
- Hijmans R. y Ghosh A. (2019). Spatial Data Analysis with R. Pág. Web  
<https://rspatial.org/raster/analysis/analysis.pdf>
- Bárdossy, A. Haberlandt, U. & Grimm-Strele, J. (2001), Interpolation of groundwater quality parameters using additional information, Technical report, In GeoENV I (Geostatistics for Enviromental Applications), 189-200, Kluwer Academic Publ. Dordrecht.
- Buol, S. W.; Hole, F. D.; McCracken, R. J. (1973). Soil Genesis and Classification (First edición). Ames, IA: Iowa State University Press.
- Oficina Económica y Comercial de la Embajada de España en Los Angeles(OEC) (2018). Guía de Estado Alaska EE.UU.
- Delgado C. (2017) Distribución Espacial De Fe Li, Pb, Mn, V, Y Zn En Suelos Urbanos de Morelia, Michoacán, México.