

Diplomado en Análisis de Información Geoespacial

Variables regionalizadas

Autor:
M. en G. Alberto Porras Velázquez

Variables regionalizadas

Orígenes de la Geoestadística

El desarrollo de la Geoestadística en la década de los sesenta del siglo pasado fue producto de la necesidad de contar con una metodología para evaluar las reservas que se pueden recobrar en los depósitos de las minas. Se le dio prioridad a la practicidad, elemento que explica su éxito y aplicación en diversos campos.

En un principio la Geoestadística era vista como un medio para describir los patrones espaciales e interpolar los valores de un atributo de interés en sitios no muestreados. Hoy se utiliza cada vez más para modelar la incertidumbre sobre valores desconocidos a través de la generación de imágenes alternativas (realizaciones), las cuales reproducen aspectos de los patrones de dependencia espacial u otras estadísticas consecuencia del problema que se trata.

Daniel Gerhardus Krige (1919-2013) ingeniero sudafricano que trabajó en la explotación de minas de oro, observó que se podían mejorar las estimaciones de las concentraciones de este metal si se consideraba su presencia en los bloques mineros vecinos, es decir, descubrió que existía una autocorrelación.

Al mismo tiempo que Krige trabajó empíricamente con la autocorrelación, Georges François Paul Marie Matheron (1930 - 2000), matemático y geólogo francés, también buscó la forma de estimar con mayor precisión las concentraciones de metales a partir de muestras de datos autocorrelacionados.

Él derivó soluciones al problema de la estimación a partir de la teoría de procesos aleatorios, que llamó teoría de variables regionalizadas.

El éxito de la Geoestadística radica en que varias disciplinas tienen el mismo problema. Se cuenta con un entorno continuo pero, debido al costo o a la dificultad para hacer muestreos, sólo se tienen mediciones en un número finito de lugares. La Geoestadística

se puede aplicar para hacer estimaciones (o predicciones en un sentido espacial) de los valores de un atributo en sitios no muestreados con un mínimo error y sin sesgo.

Variables regionalizadas.

El término variable regionalizada fue acuñado por Matheron a principios de la sexta década del siglo pasado. Una variable de este tipo contempla un aspecto aleatorio que da cuenta de las irregularidades locales y un aspecto estructurado que refleja tendencias de gran escala.

La aleatoriedad se incluye en términos de fluctuaciones alrededor de una superficie fija, llamada “deriva” (drift). Es importante recalcar que las fluctuaciones no son consideradas como errores, sino parte integral del fenómeno.

En muchos casos, cuando se aplican métodos de regresión, se modela mediante una función el comportamiento de una variable. Por lo regular habrá diferencias entre los valores observados y la función que se ajusta, estas diferencias son consideradas como errores aleatorios independientes.

Es importante señalar que para la Geoestadística, las diferencias en torno a la superficie son aleatorias, mas no independientes, ya que hay una correlación entre ellas.

Debemos recordar que la primera tarea de la Geoestadística es buscar estructura, de ahí el nombre de análisis estructural. Posteriormente, el investigador puede estimar o simular las variables.

A continuación abordaremos los supuestos y las hipótesis que constituyen el marco conceptual de la Geoestadística.

Variables aleatorias

En Geoestadística se utilizan modelos probabilísticos, a diferencia de otras disciplinas en las que se emplean modelos deterministas. Es pertinente contrastar estos dos acercamientos al estudio de los problemas.

Un modelo determinista asocia a cada ubicación \mathbf{x} , en donde no se tienen observaciones, un solo valor estimado, digamos $z^*(\mathbf{x})$ para el valor desconocido $z(\mathbf{x})$. En cambio, en un modelo probabilístico el valor estimado para $z(\mathbf{x})$ provee una serie de posibles valores con una probabilidad de ocurrencia (ilustración 1).

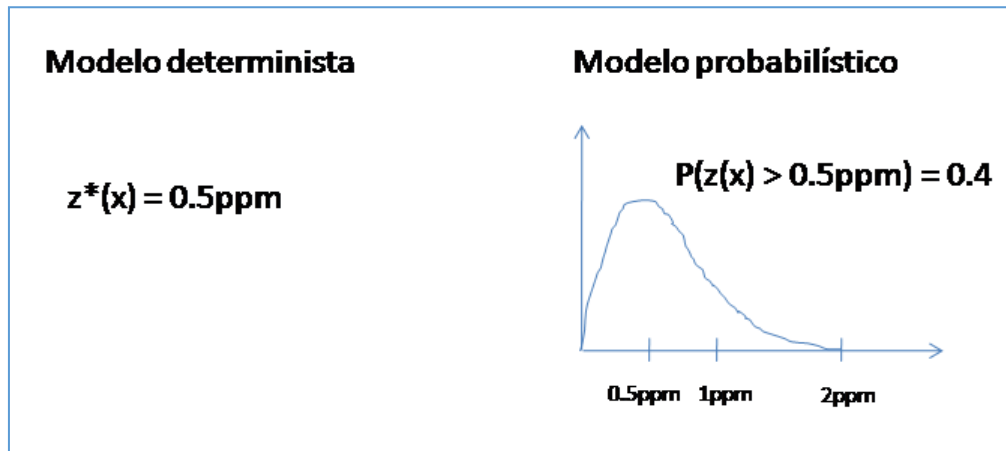


Ilustración 1. Comparación entre un modelo determinista y un modelo probabilístico. Porras, A. (2014).

De esta manera, el valor observado en cada punto \mathbf{x} se considera la realización $z(\mathbf{x})$ de una variable aleatoria $Z(\mathbf{x})$. A la media se le llama deriva (drift), $m(\mathbf{x})$. En los puntos donde no se han realizado mediciones, los valores de $z(\mathbf{x})$ están bien definidos aunque son desconocidos.

Modelo de función aleatoria

La Geoestadística se basa en el concepto de función aleatoria, en el que un conjunto de valores desconocidos se considera como un conjunto de variables aleatorias espacialmente dependientes.

En términos matemáticos, a la familia de todas estas variables aleatorias se le llama función aleatoria, proceso estocástico o campo aleatorio.

Con este modelo no es posible trabajar, a menos que se establezcan algunas suposiciones acerca de las características de estas distribuciones.

Hipótesis estacionaria y estacionariedad débil

En Estadística es común asumir que la variable es estacionaria, es decir, que su distribución es invariante ante traslaciones. De manera análoga, una función aleatoria estacionaria es homogénea y autorepetitiva en el espacio, esto es, para cualquier incremento en una distancia (h), la distribución conjunta de $Z(x_1), Z(x_2), \dots, Z(x_k)$, para las ubicaciones x_1, x_2, \dots, x_k es la misma que la de $Z(x_1 + h), z(x_2 + h), \dots, Z(x_k + h)$ para las ubicaciones $x_1 + h, x_2 + h, \dots, x_k + h$.

Estrictamente hablando, la estacionariedad requiere que todos los momentos sean invariantes ante traslaciones. Debido a que esto no siempre puede ser verificado, en general sólo se requiere que los dos primeros momentos sean constantes, es decir que la media y la covarianza se conserven ante traslaciones.

A esto se le llama estacionariedad débil. Dicho en otras palabras, el valor esperado (media) de $Z(x)$ debe ser constante para todos los puntos x .

$$E(Z(x))=m(x) = m$$

La función de la covarianza entre dos puntos x y $x+h$ depende del vector h (distancia de separación h) mas no de la ubicación x , esto es:

$$E(Z(x)Z(x+h)) - m^2 = C(h)$$

El hecho de que la media sea la misma para todos los lugares implica que no hay una tendencia espacial. La variabilidad en torno a la media (varianza) debe ser igual en todas partes, es decir que la precisión es la misma en todas partes. La covarianza no es una función del lugar, sólo depende de la separación espacial y el ángulo (h).

Hipótesis intrínseca

Para flexibilizar las condiciones anteriores se establece la “**hipótesis intrínseca**”. Ésta asume que los incrementos en la función aleatoria son débilmente estacionarios, lo cual

significa que la media y la varianza de los incrementos $Z(x+h) - Z(x)$ existe y es independiente del punto x :

$$E[Z(x+h) - Z(x)] = 0$$

$$\text{Var}[Z(x+h) - Z(x)] = 2 \gamma(h)$$

La función $\gamma(h)$ se llama semivariograma. Se trata de la herramienta básica para la interpretación estructural del fenómeno y permite realizar estimaciones.

La importancia de la hipótesis intrínseca no radica en el hecho de que el promedio de las diferencias $E[Z(x+h) - Z(x)]$ sea cero, sino en la variabilidad de estas diferencias en términos de la distancia de separación h entre las observaciones.

¿Cómo sabemos si una variable es estacionaria?

La mayoría de los estimadores usados en las Ciencias de la Tierra son combinaciones lineales (promedios ponderados móviles) de los datos. Algunos ejemplos son los métodos IDW y poligonal, así como el Kriging.

Para fines prácticos, el fenómeno que se estudia sólo requiere ser estacionario hasta cierta distancia. El problema consiste en decidir si podemos encontrar ventanas de vecinos dentro de las que el valor esperado y el variograma puedan ser consideradas constantes.

Por ejemplo, en la **ilustración 2** se observa la concentración de sulfuro en un corte lineal de ocho kilómetros. A lo largo de este intervalo el contenido de sulfuro no es estacionario ya que hay un incremento constante de la concentración en los ocho kilómetros; sin embargo, si se toma un corte en la sección central de 3 kilómetros, las fluctuaciones parecen ir en torno una tendencia.

Por ello podemos considerar que en este corte la escala de la variable es localmente estacionaria o, por lo menos, intrínsecamente variable, mientras que claramente no es estacionaria en el rango de los ocho kilómetros.

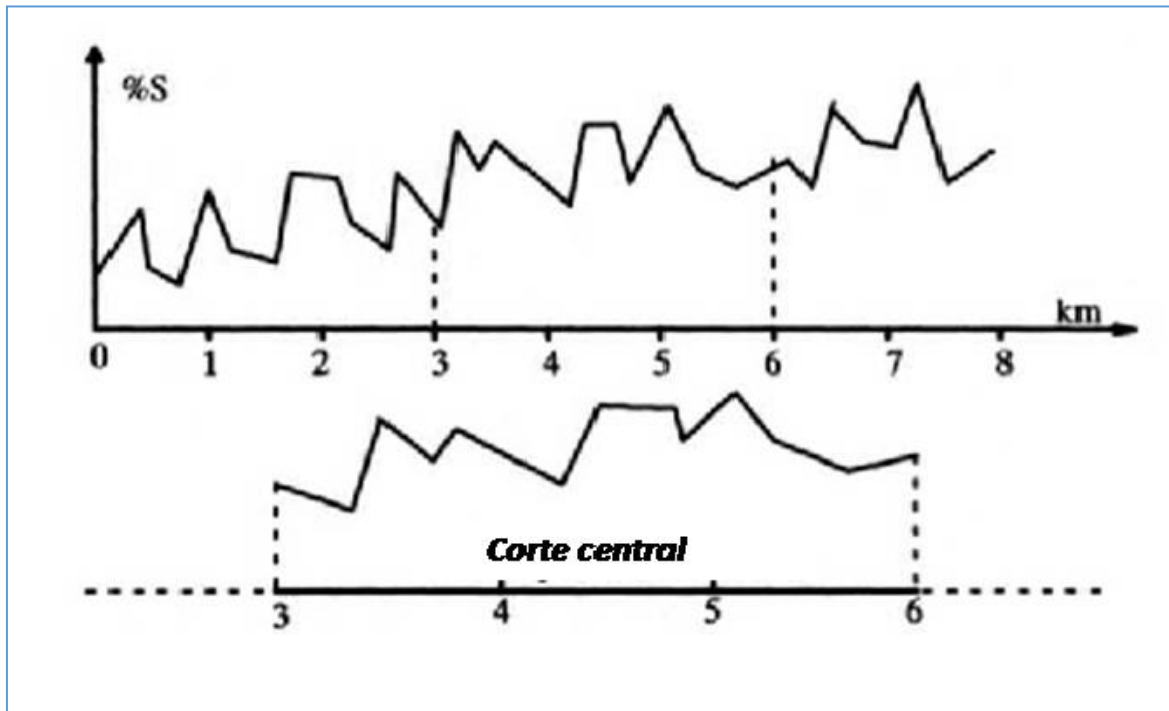


Ilustración 2. Concentración de sulfuro en un corte lineal de 8 km. Imagen editada a partir de la original en Armstrong, M. (1998), p.19.

Referencias

- Armstrong, M. (1998). Basic Linear Geostatistics. Berlin, Germany: Springer-Verlag.
- Bivand R.S., Pebesma, E. & Gómez-Rubio, V. (2013). Applied Spatial Data Analysis with R (2ed). New York, United States of America: Springer.
- Goovaerts, P. (1997). Geostatistics for Natural Resources Evaluation. New York, United States of America: Oxford University Press.