



Apunte de Geoestadística

Preparado por: Xavier Emery
Julio de 2007

Capítulo 1. Introducción

La palabra “geoestadística” fue inventada por Georges Matheron en 1962.

El prefijo “*geo*” alude a las ciencias de la tierra, que ha sido históricamente la disciplina donde la geoestadística se ha desarrollado. Los campos de aplicación actuales alcanzan los dominios más variados, como la evaluación de recursos naturales (mineros, forestales, gas, petróleo, etc.), ciencias del suelo y medio-ambientales, topografía, oceanografía, geofísica, agricultura y análisis de imágenes, por nombrar algunos. El término “*estadística*” se refiere al uso de herramientas estadísticas y probabilísticas. Con respecto a la estadística clásica, la geoestadística busca tomar en cuenta las dependencias entre las observaciones disponibles, considerando que ellas están ubicadas en el espacio.

1. *Noción de variable regionalizada*

La geoestadística se define como el estudio de fenómenos *regionalizados*, es decir, que se extienden en el espacio y presentan una cierta continuidad. Por “espacio”, entenderemos en general el espacio geográfico, pero puede también tratarse del eje temporal o de espacios más abstractos. El objeto sobre el cual trabajaremos será una descripción matemática del fenómeno regionalizado, a saber, una o varias funciones numéricas llamadas **variables regionalizadas**, que miden ciertas propiedades o atributos relacionados con este fenómeno. Por ejemplo:

- la ley de un mineral, la potencia de una veta, la acumulación, la densidad de la roca o la recuperación metalúrgica, describen un fenómeno de mineralización;
- la porosidad y la permeabilidad de la roca en un reservorio de petróleo o en un acuífero;
- la concentración de un elemento contaminante en la atmósfera o en el suelo;
- la altitud topográfica en un punto del espacio geográfico;
- la conductividad eléctrica, el pH y la concentración en nutrientes medidas sobre una muestra de suelo;
- el número de árboles y su diámetro promedio en áreas de observación de un bosque.

Del punto de vista conceptual, una variable regionalizada es una función determinística. En general, esta función presenta dos aspectos complementarios: por una parte, tiene una cierta “continuidad” espacial (zonas de altos valores / zonas de bajos valores), pero por otro lado, varía irregularmente y escapa a toda representación simple (Figura 1).

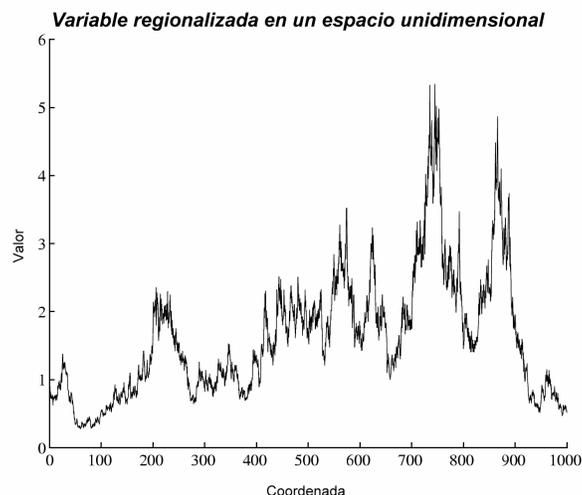


Figura 1. Variable regionalizada en un espacio unidimensional.
 Por ejemplo, podría representar la concentración de un contaminante medida a lo largo del tiempo en una estación de monitoreo.

Dado que un fenómeno regionalizado nunca posee una extensión infinita, estudiaremos la variable regionalizada sólo dentro de un dominio limitado D llamado **campo** de la variable. Este campo puede representar una zona natural, fuera de la cual la variable no está definida. Puede tratarse también de un dominio particular, donde la variable interesa, por ejemplo, los sitios donde no se anula o donde es mayor que un límite de detección.

2. *Noción de soporte*

Una variable regionalizada puede definirse, no sólo en cada punto del espacio, sino que también en una superficie (2D) o en un volumen (3D). La superficie o el volumen sobre el cual se considera la variable regionalizada se denomina **soporte**. En general, el soporte de las mediciones es muy pequeño (asimilado a un “punto”), mientras que el que interesa en la práctica puede ser más voluminoso (por ejemplo, las unidades selectivas de explotación en evaluación minera o las unidades de remediación en contaminación de suelo). Esta noción es esencial debido a la dependencia que existe entre el soporte y la distribución estadística de los valores, conocida como **efecto de soporte**: los soportes voluminosos presentan una menor cantidad de valores extremos y una mayor cantidad de valores intermedios que los soportes puntuales. Así, la distribución de los valores (en especial, su varianza) depende del soporte sobre el cual está definida la variable regionalizada. Este efecto queda ilustrado en la Figura 2, que muestra la distribución de la ley de cobre en un yacimiento medida sobre tres soportes distintos. Se observa que la forma del histograma tiende a simetrizarse cuando aumenta el soporte y que la varianza disminuye, aunque se mantiene constante el valor promedio.

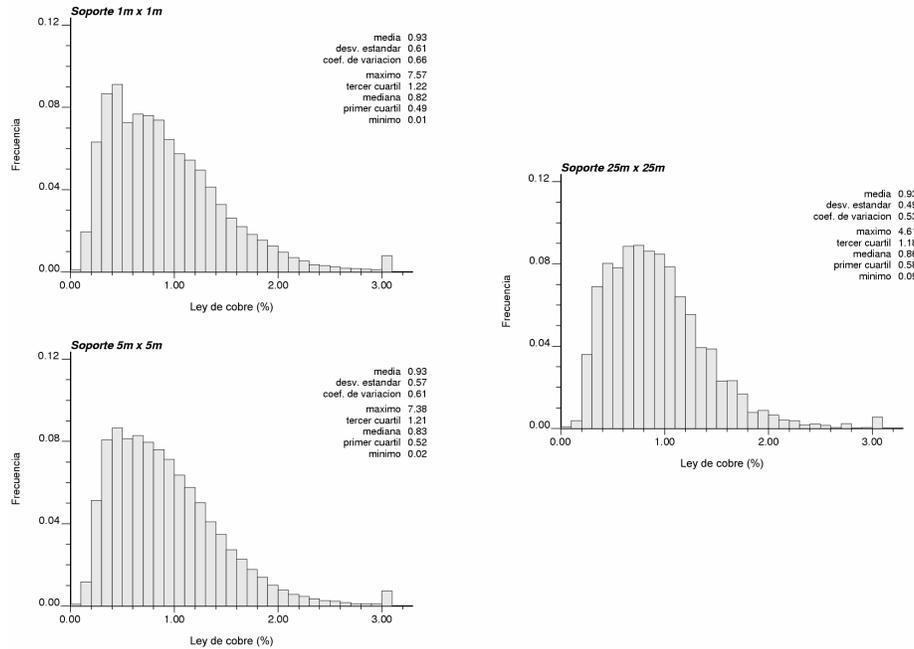


Figura 2. Histogramas de leyes de cobre medidas sobre tres soportes distintos (bloques de 1m×1m, 5m×5m y 25m×25m).

En los problemas que involucran un cambio de soporte, es deseable que la variable regionalizada sea *aditiva*, es decir, que su valor en la unión de varios dominios sea igual a la media de sus valores sobre cada uno de ellos. Esta restricción es necesaria para que el cálculo del valor promedio sobre un soporte más grande que el soporte de las mediciones, tenga un sentido físico.

3. Datos y modelos

Normalmente, no se conoce la variable regionalizada en forma exhaustiva, sino que solamente a través de un conjunto limitado de **datos** (muestras de sondeos en exploración minera, observaciones en terreno en ingeniería forestal, mediciones realizadas en estaciones de monitoreo en contaminación atmosférica, etc.).

Para aprovechar la información disponible, más allá de un simple reporte de los datos, será necesario construir **modelos**. Pero debe tenerse presente que un modelo nunca describe la variable regionalizada en su totalidad, sin simplificación, ni distorsión. El modelamiento y las hipótesis que se requieren aportan mayor información que aquella contenida en los datos experimentales. En consecuencia, conviene reducir al máximo la complejidad del modelo, evitando introducir hipótesis o parámetros arbitrarios, que no se puede controlar experimentalmente. Por el contrario, es necesario encontrar el modelo menos exigente que permite resolver el problema planteado.

4. Los problemas de estimación

La “predicción” o “estimación” constituye un objetivo importante de la geoestadística. Consiste en evaluar, de la manera más precisa posible, un valor que no ha sido medido, a partir de los datos disponibles. Una característica de los métodos geoestadísticos será, no sólo construir estimadores, sino también proporcionar una medida de la precisión de la estimación por medio de herramientas probabilísticas (por ejemplo, varianzas o intervalos de confianza...). Distinguiremos dos tipos de estimaciones: las estimaciones *globales* y las estimaciones *locales*.

La **estimación global** busca caracterizar el campo D completo por un valor único (por ejemplo, la media de la variable regionalizada en estudio) o por una distribución estadística (histograma). Es poco común que una estimación global sea suficiente; frecuentemente, se requiere completarla con estimaciones locales. Por ejemplo en un estudio de contaminación de suelo, no basta con evaluar la concentración promedio en toda la zona de un elemento contaminante, sino que es necesario distinguir los sectores fuertemente contaminados de aquellos que no lo están.

Por el contrario, las **estimaciones locales** se interesan por los diferentes sectores de la zona de estudio. En general, se busca evaluar el valor en un sitio que no ha sido muestreado o el valor promedio de un “bloque” cuyo soporte es mayor que el soporte de los datos. En este problema, debe considerarse la continuidad espacial de la variable regionalizada, las distancias entre el sector a estimar y los sitios con datos, así como la configuración espacial de los mismos sitios con datos puesto que datos agrupados suelen tener valores parecidos, constituyéndose en información redundante.

Otro aspecto de la estimación local es la cuantificación de la **incertidumbre** en los valores de la variable regionalizada y la probabilidad de que estos valores superen o queden debajo de ciertos límites. Por ejemplo, el agrónomo está interesado en conocer los sectores de una parcela donde el pH es demasiado bajo, para poder abonarlos con cal y mejorar la fertilidad. El medio-ambientalista quiere cuantificar el riesgo de que la concentración de un elemento tóxico en el suelo supere un umbral crítico, de modo de poder tomar medidas de remediación. Al ingeniero de minas le interesa saber si las leyes de mineral son mayores que una ley de corte que asegura la rentabilidad de la explotación minera, luego evaluar los tonelajes y cantidades de metal que se puede extraer sobre esta ley de corte.

Los problemas de estimación no son los únicos a los que responde la geoestadística. El análisis de la **dependencia espacial** permite describir la distribución de los valores en el espacio, cuantificar las *correlaciones* o redundancias de información entre valores medidos en sitios diferentes, determinar el tamaño de la “zona de influencia” de una observación, así como detectar *anisotropías* que indican que la variable regionalizada tiene direcciones preferenciales de continuidad en el espacio.

5. Algunas notaciones

En lo que sigue, denotaremos como:

- d la dimensión del espacio de trabajo. En general, $d = 2$ ó 3 , es decir, nos encontramos en el espacio geográfico \mathbb{R}^2 o \mathbb{R}^3 .
- D el campo de la variable regionalizada, que es un dominio limitado del espacio \mathbb{R}^d .
- $\mathbf{x} = (x_1, \dots, x_d)$ un vector de coordenadas espaciales.
- $z = \{z(\mathbf{x}), \mathbf{x} \in D\}$ la variable regionalizada estudiada. A menudo, $z(\mathbf{x}) \in \mathbb{R}$. En el caso multivariable, se podrá tomar $z(\mathbf{x}) \in \mathbb{R}^N$ o, lo que es lo mismo, trabajar con N variables regionalizadas reales.
- n el número de sitios con datos.
- $\{\mathbf{x}_\alpha, \alpha = 1 \dots n\}$ los sitios con datos en D .

Capítulo 2. Estudio exploratorio de datos

La geoestadística busca estudiar una o varias variables regionalizadas conocidas vía una toma de muestra. Previo al uso de métodos geoestadísticos propiamente tal, conviene realizar un estudio exploratorio de los datos disponibles, el cual tiene tres objetivos:

- 1) Analizar (mediante herramientas estadísticas simples) la cantidad, la calidad y la ubicación de los datos disponibles.
- 2) Definir la(s) zona(s) de estudio. Una división del campo en varias sub-zonas puede ser relevante si uno observa cambios abruptos en la distribución espacial de valores, o si la geología del fenómeno lo indica.
- 3) Anticipar dificultades o problemas que puedan surgir en la fase de estimación local (por ejemplo, presencia de valores atípicos que se destacan de aquellos de los datos vecinos).

Se ejemplificará los conceptos y herramientas con el estudio de una base de datos de contaminación de suelo. Estos datos constan de 359 muestras en las cuales se ha medido siete variables (a saber, las concentraciones de cadmio, cobalto, cromo, cobre, níquel, plomo y zinc). A continuación, se da una breve descripción de la zona de estudio (fuente: referencia bibliográfica [44]).

To explore the nature of the problem we chose a region near La Chaux de Fonds in the Swiss Jura (Fig. 1). One reason for this choice was that a previous impact study suggested contamination of soil by cadmium. Manufacturing and agriculture were suspected sources, but geology could not be excluded. The region covers 14.5 km². It is large enough to embrace outcrops of several geological formations (though all of Jurassic limestone) and many farms under similar management for stock raising. It is also large enough to detect spatial variation in atmospheric deposition. It lies between 870 and 1070 m above sea level with smooth slopes (<10%) except for the steep gorge at its eastern margin. The average annual rainfall is 1530 mm. Of the agricultural land (83%) some 80% is permanent grass, which is either grazed (pasture), or cut for hay or silage (meadow), and 3% is sown to barley. The remaining 17% is forest. Manure and commercial fertilizers are spread at similar levels over the whole area. The region had numerous jewellery and watch factories at the beginning of this century, and burning or spreading of town wastes are suspected as sources of metals.

1. Despliegue de datos (mapas)

El despliegue de los datos permite visualizar su ubicación en el espacio. Cuando se trabaja en el espacio tridimensional, el despliegue puede hacerse con proyecciones (plantas y secciones) o vistas en perspectiva. Los mapas de datos pueden incluir información sobre los valores medidos y así dar una idea preliminar de la distribución y continuidad espacial de la variable en estudio, por ejemplo, utilizando símbolos de forma o tamaño variable o escalas de colores / de grises (Figura 1).

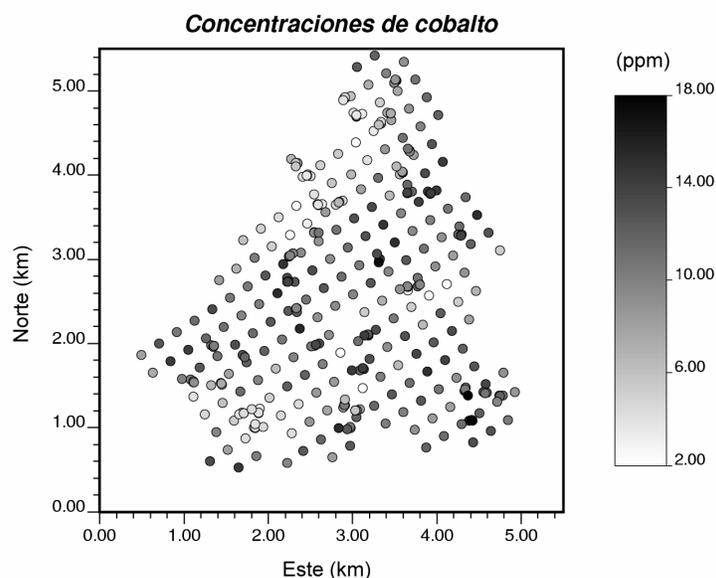


Figura 1. Mapa de datos de contaminación de suelo. La escala de grises se refiere a la concentración de cobalto expresada en ppm.

Se observa que gran parte de los datos están distribuidos según una malla de muestreo regular de tamaño aproximado 0.25×0.25 km. Sin embargo, algunos sectores de la zona de estudio han sido reconocidos con una mayor cantidad de muestras, constituyéndose algunos agrupamientos (*clusters*). El diseño de la malla de muestreo se explica a continuación (ver referencia [44]).

The basic grid was square with an interval of 250 m, and this gave 214 grid nodes. One fifth of the grid points were used as starting points for a nested sampling. In nesting distances from point to point in a geometric progression of approximately 2.5 fold increments were chosen. The steps were 100 m, 40 m, 15 m and 6 m. The distances were fixed, and the directions were chosen at random, as shown in the example in Fig. 2. The scheme covers three orders of magnitude.

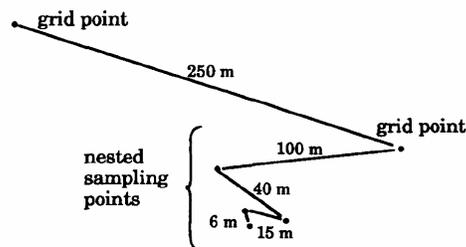


Fig. 2. Example of one spatial configuration of the points for nested sampling, from one grid node.

La presencia de agrupamientos de datos hace que la malla de muestreo sea irregular en la zona de estudio. Al momento del análisis estadístico, será importante tomar precauciones puesto que los sectores más densamente muestreados podrían tener demasiada importancia y distorsionar las estadísticas sobre las concentraciones de cobalto. Por ejemplo, esto sería el caso si los agrupamientos fueran realizados en los sectores de mayores concentraciones de cobalto (*muestreo preferencial*).

El mapa de los datos (Figura 1) da una primera idea de la distribución de los valores en la zona de estudio. Así se puede distinguir sectores con valores altos (datos señalados con colores oscuros) y sectores menos contaminados (datos señalados con colores claros).

2. Distribución estadística de valores

Además de visualizar mapas de los datos, interesa conocer la distribución estadística de sus valores. Para ello, existen varias herramientas, siendo la más conocida el histograma.

2.1. Histograma

El histograma representa gráficamente las frecuencias de ocurrencia en función del valor. Consiste en dividir el rango de los valores en intervalos (generalmente, con el mismo ancho) y visualizar la proporción de datos que caben dentro de cada intervalo (Figura 2).

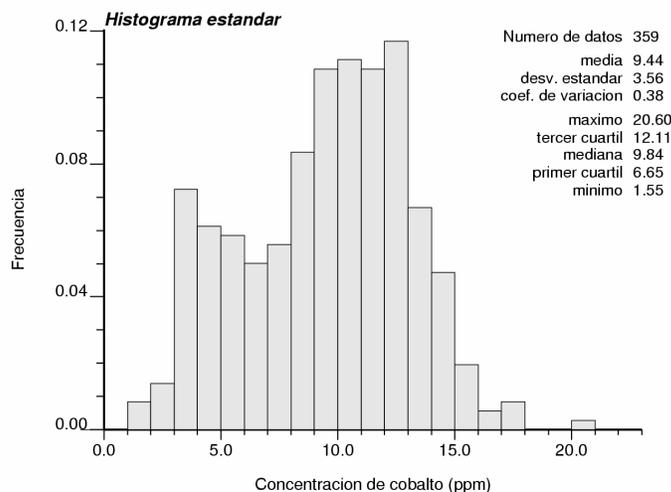


Figura 2. Histograma de las concentraciones de cobalto (en ppm).

El histograma es una herramienta útil para detectar valores atípicos (“outliers”). Ahora, cabe destacar que un dato atípico no es forzosamente falso y nunca debe ser eliminado sin razón (por ejemplo, un error de transcripción, una falla en el protocolo de medición o un valor ausente codificado como -99). Podría reflejar el comportamiento verdadero de la variable regionalizada y eliminarlo impediría prever la ocurrencia de tales valores en las zonas no muestreadas.

La visualización del histograma de los datos también es un primer medio de verificar su homogeneidad. Eventualmente, una división del campo en varias sub-zonas será necesaria. Así, por ejemplo, un histograma multimodal puede conducir a la identificación, entre los datos, de varias “poblaciones” susceptibles de estar geográficamente separadas. A veces, tal separación está impuesta por consideraciones físicas, que impiden mezclar todos los datos: presencia de un obstáculo natural (falla, río...), partición de una zona mineralizada según la caracterización mineralógica o el tipo de roca, etc. En tales casos, un problema que puede plantearse es la delimitación de las sub-zonas “homogéneas”, pues es poco frecuente que sus fronteras puedan ser identificadas con exactitud.

2.2. Estadísticas elementales

Junto con el histograma, es conveniente calcular algunas estadísticas básicas sobre la distribución de valores. Entre ellas, podemos distinguir:

Medidas de posición

- *Media*: promedio aritmético de los valores.
- *Cuantiles o percentiles*: valores que dividen la población en partes de igual número de datos. Por ejemplo, la *mediana* divide la población en dos partes, los *cuantiles* en cuatro partes (la mediana coincide con el segundo cuartil), los *quintiles* en cinco partes y los *deciles* en diez partes. Contrariamente a la media, los cuantiles son parámetros *robustos*, es decir, poco sensibles a la presencia de algunos valores muy altos o muy bajos.
- *Mínimo y máximo*: establecen el rango en el cual se distribuyen los valores.

Medidas de dispersión

- *Varianza*: promedio aritmético de la desviación cuadrática entre cada valor y la media. Esta medida cuantifica la dispersión del histograma y se expresa en el cuadrado de la unidad de la variable en estudio.
- *Desviación estándar*: raíz cuadrada de la varianza; se expresa en la misma unidad que la variable en estudio.
- *Coefficiente de variación* (para variables positivas): razón entre la desviación estándar y la media; es adimensional.
- *Rango intercuartil*: ancho del intervalo entre el primer y el tercer cuartil, que contiene la mitad de los datos.

2.3. Desagrupamiento

Para calcular el histograma y las estadísticas elementales, es frecuente considerar una variable de ponderación o “desagrupamiento” (en inglés, *declustering*) de modo de corregir los efectos provocados por las irregularidades de la malla de muestreo. La idea es asignar a los datos espacialmente agrupados, en parte redundantes, un ponderador más pequeño que a los datos aislados, los cuales son representativos de una porción más extensa del campo.

Para determinar la variable de ponderación, se puede utilizar el llamado *método de las celdas*. Esta técnica consiste en dividir la zona de estudio en celdas rectangulares idénticas y en atribuir a cada dato un ponderador inversamente proporcional al número de datos presentes en la celda a la cual pertenece (Figura 3).

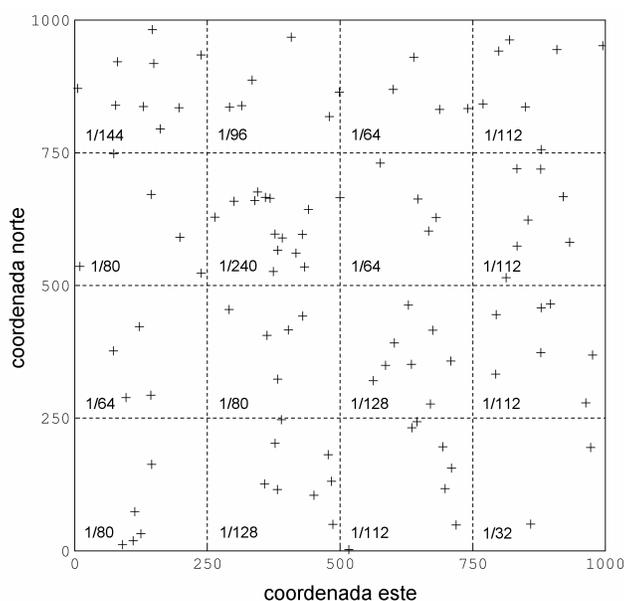


Figura 3. Ubicación de datos (cruces) y ponderación por el método de las celdas. Cada celda tiene una ponderación total de $1/16$, la cual se reparte entre los datos pertenecientes a esta celda.

La puesta en marcha del método requiere definir previamente el tamaño de la celda elemental. Usualmente, se prueba varios tamaños y se examina su impacto en el valor de la media ponderada. La Figura 4 da un ejemplo con los datos de concentración de cobalto, donde se ha escogido utilizar celdas cuadradas.

Si las celdas son infinitamente pequeñas, todos los datos se ubican en celdas distintas y reciben entonces el mismo ponderador; por lo tanto, la media ponderada coincide con la media aritmética de los datos (9.44 ppm). Al aumentar la longitud de celda, la media ponderada aumenta, lo que indica que los ponderadores calculados tienden a privilegiar los valores altos en perjuicio a los valores bajos. Esto es una señal de que la malla de muestreo es más densa en zonas de valores bajos.

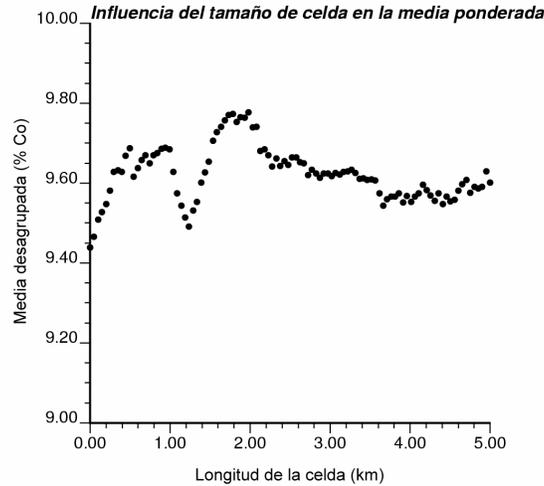


Figura 4. Influencia de la longitud de las celdas sobre la media ponderada (sumatoria de las concentraciones multiplicadas por sus ponderadores)

En este caso, una opción consiste en elegir el tamaño de celda correspondiente al valor máximo de la media, para “compensar” el carácter preferencial de la malla de muestreo. Ahora, dicha decisión es discutible y uno puede considerar elecciones más “neutrales”. Por ejemplo, elegir una celda correspondiente al espaciamiento promedio entre los datos o a la malla del muestreo subyacente (o sea, en el presente caso, de 0.25 km × 0.25km). Esto nos entrega un valor promedio de 9.59 ppm para la concentración de cobalto (Figura 5). En comparación con el histograma de la Figura 2, se observa una disminución importante de las frecuencias en el rango de valores entre 3 y 5 ppm.

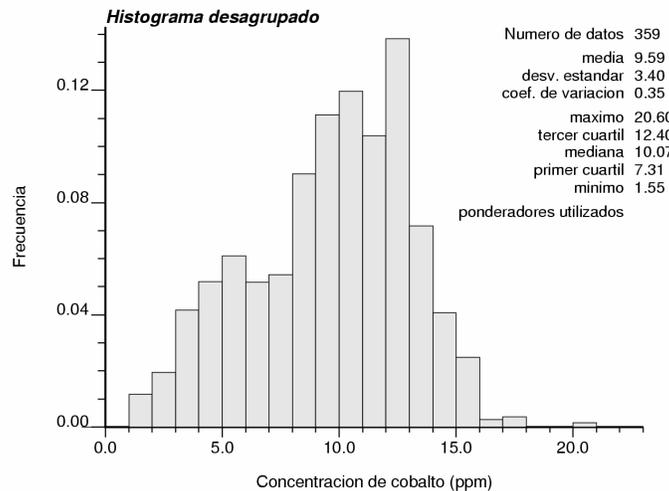


Figura 5. Histograma desagrupado de las concentraciones de cobalto (en ppm). La frecuencia indicada en cada intervalo corresponde a la suma de los ponderadores de los datos asociados a este intervalo.

El método de las celdas realiza la ponderación de los datos según un criterio meramente “geométrico”, es decir, que sólo toma en cuenta la posición espacial de los datos. No toma en cuenta la continuidad espacial de los valores, aunque este criterio también es digno de ser tomado en consideración. Por ejemplo, si los datos tuvieran valores muy erráticos (poca continuidad espacial), existirían pocas redundancias entre datos agrupados y, por lo tanto, no sería necesario ponderarlos. Ahora, incluir el criterio de continuidad espacial requeriría utilizar métodos geoestadísticos (kriging) que, a esta altura del estudio, no se justifican.

2.4. Diagrama de caja

A veces, se acompaña el histograma con un diagrama de caja (“box plot”) que presenta un solo eje en el cual se representan cinco cuantiles: los cuantiles a 2.5% y 97.5%, el primer y el tercer cuartil y la mediana (Figura 6). Entre los dos cuantiles extremos, se observa el 95% de los datos, mientras que entre el primer cuartil y la mediana se observa el 25% de los datos, al igual que entre la mediana y el tercer cuartil. El diagrama de caja permite resumir algunas características de la distribución, tal como su simetría y su dispersión.

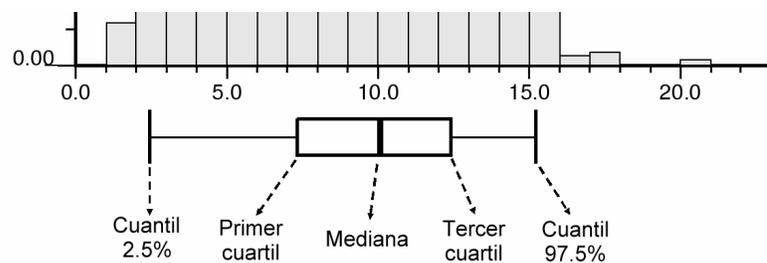


Figura 6. Diagrama de caja para las concentraciones de cobalto.

2.5. Histograma acumulado

En lugar de visualizar la frecuencia de valores para cada intervalo, se puede visualizar la frecuencia acumulada, es decir, la frecuencia de los valores bajo cierto umbral como una función de este umbral, lo que da el histograma acumulado. De la misma forma que para el histograma estándar, las frecuencias se pueden calcular tomando en cuenta ponderadores de desagrupamiento (Figura 7).

El histograma acumulado es una función creciente, desde 0 (para el valor más bajo de la variable) hasta 1 (para el valor más alto). Al contrario, el *histograma acumulado inverso*, que muestra la frecuencia de valores que superan un umbral en función de este umbral, es una función decreciente (desde 1 hasta 0), conocida en geoestadística minera como *curva tonelaje-ley* dado que la frecuencia acumulada es proporcional al tonelaje de material cuya ley supera la ley de corte.

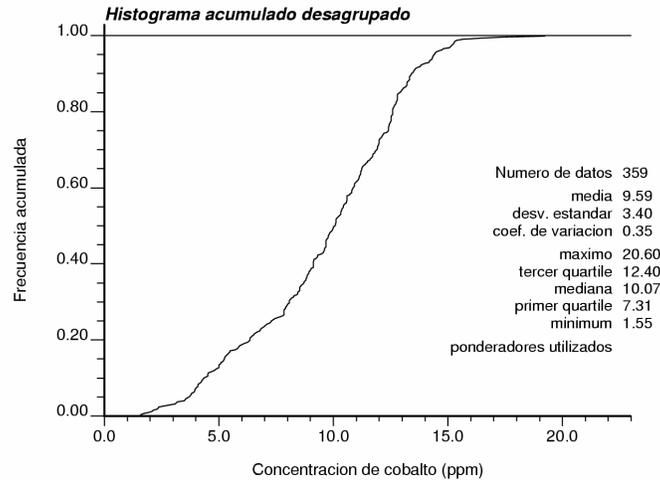


Figura 7. Histograma acumulado desagrupado de las concentraciones de cobalto (en ppm).

2.6. Gráfico de probabilidad

Este gráfico sirve para comparar una distribución experimental con una distribución de referencia (en general, una normal o una lognormal). Consiste en cambiar la escala de los ejes del histograma acumulado de tal modo que, si la distribución experimental coincide con la distribución de referencia, se dibujaría una recta. En el caso de los datos de cobalto, la distribución difiere de una normal (Figura 8).

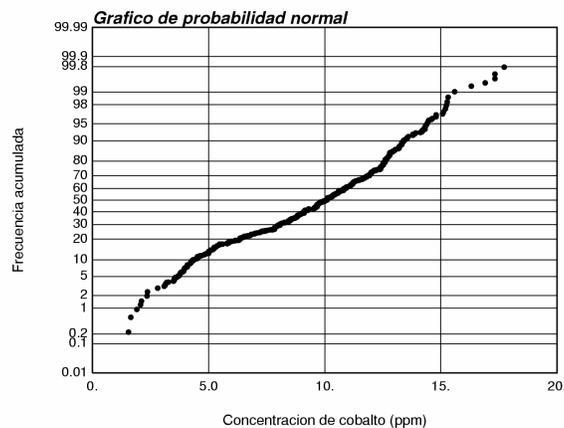


Figura 8. Gráfico de probabilidad normal para las concentraciones de cobalto. Note que la escala del eje de ordenada (frecuencia acumulada) ya no es aritmética

3. Análisis multivariable

3.1. Gráfico cuantil contra cuantil

El gráfico cuantil contra cuantil busca comparar dos distribuciones entre sí, que pueden ser asociadas a variables distintas (por ejemplo, las concentraciones de cobalto y níquel) o a una misma variable medida en zonas distintas o con aparatos distintos. En este último caso, la comparación sirve para decidir si los dos grupos de datos tienen una distribución idéntica y, por ende, provienen de la misma población. El gráfico consiste en visualizar los cuantiles del primer grupo de datos contra los cuantiles correspondientes del segundo grupo (por ejemplo, mediana contra mediana, cuantiles contra cuantiles, deciles contra deciles, etc.).

Un gráfico que da una recta significa que las distribuciones de los dos grupos de datos tienen la misma forma. Si la recta tiene una pendiente igual a 1, ambas distribuciones tienen la misma dispersión, pero posiblemente no tienen la misma media (por ejemplo, un grupo de datos está sesgado con respecto a otro debido a un error sistemático en su medición). Si además la recta coincide con la diagonal, ambas distribuciones son iguales (misma media, misma dispersión y misma forma). Cuando el gráfico no dibuja una recta, entonces los dos grupos de datos tienen distribuciones con formas distintas; es lo que pasa al comparar las distribuciones de las concentraciones de cobalto y níquel (Figura 9).

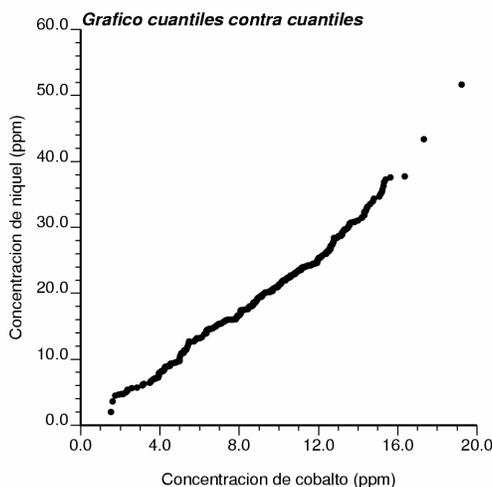


Figura 9. Gráfico cuantil contra cuantil entre concentraciones de cobre y níquel. La pendiente de la curva es distinta entre las bajas y las altas concentraciones.

Es importante señalar que el gráfico cuantil contra cuantil no da ninguna información sobre la correlación que existe entre los dos grupos de datos. Sólo permite comparar sus distribuciones (histogramas).

3.2. Nube de dispersión o de correlación

Esta nube consiste en visualizar los valores de una variable en función de otra, lo cual requiere que ambas variables hayan sido medidas en los mismos sitios. Sirve para ver la relación par a par o “correlación” de ambas variables, así como detectar datos atípicos (puntos que se alejan del resto de la nube). En el ejemplo de los datos de contaminación de suelo, existe una buena correlación entre las concentraciones de cobalto y níquel: cuando aumenta una de estas dos variables, tiende a aumentar la otra, lo que corrobora la pendiente positiva de la recta de regresión (Figura 10).

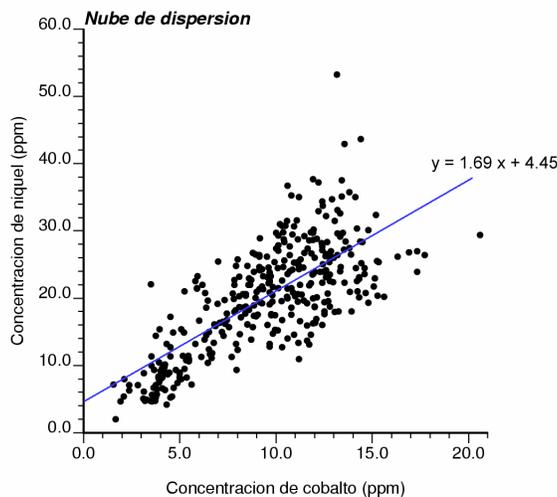


Figura 10. Nube de dispersión entre concentraciones de cobre y níquel.

3.3. Matriz de correlación

La relación entre dos variables puede ser resumida a través de un coeficiente, llamado *coeficiente de correlación lineal*, que mide el grado de “proporcionalidad” que existe entre ambas variables. Este coeficiente tiene un valor entre -1 y 1 (estos extremos corresponden al caso en el cual las variables son exactamente proporcionales entre sí, con un coeficiente de proporcionalidad negativo o positivo, respectivamente).

Cuando existen más de dos variables, se puede construir una matriz de correlación. En la intersección de la fila i con la columna j , se coloca el valor del coeficiente de correlación entre las variables $n^{\circ}i$ y $n^{\circ}j$. En la Tabla 1, se observa correlaciones superiores a 0.7 entre las concentraciones de níquel, cobalto y cromo, así como entre las concentraciones de cobre y plomo. Esto sugiere un origen distinto para estos diferentes grupos de variables. Un análisis detallado indica que las concentraciones de cobalto y níquel están relacionadas con la geología de la zona y que estos contaminantes provienen en gran medida de la roca de caja, mientras que las concentraciones de cobre y plomo son susceptibles de haber sido provocadas por el uso de fertilizantes o abonos o por desechos domésticos.

VARIABLE	Cd	Co	Cr	Cu	Ni	Pb	Zn
Cd	1.00	0.26	0.58	0.15	0.49	0.22	0.62
Co	0.26	1.00	0.48	0.19	0.74	0.16	0.44
Cr	0.58	0.48	1.00	0.21	0.71	0.26	0.61
Cu	0.15	0.19	0.21	1.00	0.22	0.82	0.66
Ni	0.49	0.74	0.71	0.22	1.00	0.27	0.59
Pb	0.22	0.16	0.26	0.82	0.27	1.00	0.67
Zn	0.62	0.44	0.61	0.66	0.59	0.67	1.00

Tabla 1. Matriz de correlación entre concentraciones de contaminantes. Los términos diagonales valen 1, pues corresponden a la correlación entre una variable y ella misma.

3.4. Análisis en componentes principales

Supongamos que se tiene n muestras (*individuos*) en los sitios $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, en las cuales se ha medido N variables $\{z_1, \dots, z_N\}$. Esto se puede representar en una matriz de tamaño $n \times N$, denotada \mathbf{Z} :

$$\mathbf{Z} = \begin{pmatrix} z_1(\mathbf{x}_1) & \cdots & z_N(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ z_1(\mathbf{x}_n) & \cdots & z_N(\mathbf{x}_n) \end{pmatrix}.$$

En adelante, se supone que las variables tienen media 0 y varianza 1 (de lo contrario, basta con restar a cada variable su valor promedio y dividirla por su desviación estándar). Geométricamente, se puede representar la matriz de datos de dos maneras: ya sea n puntos en \mathbb{R}^N (cada coordenada está asociada con una variable), o bien N puntos en \mathbb{R}^n (cada coordenada está asociada con un individuo). Estas representaciones se llaman la *nube de los individuos* y la *nube de las variables*, respectivamente. A continuación, definiremos cómo visualizar estas nubes mediante sus proyecciones en sub-espacios de pequeña dimensión.

Se busca construir nuevas variables, llamadas **factores** y denotadas $\{y_1, \dots, y_N\}$, que sean combinaciones lineales de las variables iniciales y que no tengan correlación entre sí:

$$\text{cov}(y_p, y_q) = \frac{1}{n} \sum_{\alpha=1}^n y_p(\mathbf{x}_\alpha) y_q(\mathbf{x}_\alpha) = 0 \quad \text{si } p \neq q.$$

Sean \mathbf{Y} la matriz $n \times N$ que da los valores de los factores $\{y_1, \dots, y_N\}$ para cada una de los n individuos y \mathbf{V} la matriz $N \times N$ de varianza – covarianza de las variables:

$$\mathbf{V} = \frac{1}{n} \mathbf{Z}' \mathbf{Z}.$$

Se demuestra que \mathbf{Y} se obtiene al plantear $\mathbf{Y} = \mathbf{Z} \mathbf{Q}$, donde \mathbf{Q} es la matriz de vectores propios de \mathbf{V} (también llamados *ejes factoriales*):

$$\mathbf{V} = \mathbf{Q} \mathbf{D} \mathbf{Q}^t \quad \text{con } \mathbf{Q}^t \mathbf{Q} = \mathbf{I} \text{ y } \mathbf{D} = \text{matriz diagonal.}$$

El factor y_p (con $p = 1 \dots N$) está asociado a un valor propio de \mathbf{V} (término d_p en la diagonal de \mathbf{D}), que representa la varianza del factor. De este modo, las varianzas de los factores descomponen la varianza total, ya que la traza de la matriz \mathbf{V} se conserva en el proceso de diagonalización:

$$\text{traza}(\mathbf{V}) = N = \sum_{i=1}^N \text{var}(z_i) = \sum_{p=1}^N \text{var}(y_p).$$

Los valores propios de \mathbf{V} indican la parte de la varianza total asociada a los distintos factores y permiten jerarquizarlos. La razón entre la varianza del factor y la varianza total (o sea, d_p / N) es una medida sin dimensión (entre 0 y 1) de la cantidad de información contenida en el factor y_p . De este modo, el análisis en componentes principales permite resumir la información multivariable contenida en los datos, sólo reteniendo los factores de mayores varianzas, por ejemplo aquellos cuyas varianzas suman el 80% de la varianza total. En adelante, se supone que los valores propios de \mathbf{V} han sido ordenados en orden decreciente: $d_1 \geq \dots \geq d_N$, es decir, que los factores han sido numerados en orden decreciente de sus varianzas.

La **nube de los individuos** está definida en un espacio de dimensión N y no se puede visualizar cuando $N > 3$. Para obtener una representación sintética de los datos, la idea es proyectar esta nube en un sub-espacio de dimensión $k \leq N$ (en la práctica, se suele tomar $k = 2$), eligiendo el sub-espacio más “cercano” a la nube, a saber, el que minimiza su momento de inercia. Se demuestra que éste corresponde al sub-espacio generado por los k primeros ejes factoriales (k primeras columnas de la matriz \mathbf{Q}). En particular, la proyección de la nube de los individuos en el primer plano factorial es el conjunto de puntos cuyas coordenadas corresponden a las dos primeras columnas de la matriz \mathbf{Y} . El examen de esta proyección permite ver la tendencia general de la nube, a veces sin buscar identificar los individuos, ya que su número puede ser muy alto. En particular, es interesante detectar las zonas con alta o baja densidad de puntos, que corresponden a dispersiones no homogéneas de la población.

La diagonalización de la matriz de varianza – covarianza puede escribirse como sigue:

$$\mathbf{V} = \mathbf{A}^t \mathbf{A} \quad \text{con } \mathbf{A} = \sqrt{\mathbf{D}} \mathbf{Q}^t.$$

El vector \mathbf{a}_i , i -ésima columna de \mathbf{A} , contiene los coeficientes de correlación entre la variable z_i y los distintos factores. Además, el producto escalar de los vectores \mathbf{a}_i y \mathbf{a}_j es igual al coeficiente de correlación entre las variables z_i y z_j . En particular, $\langle \mathbf{a}_i, \mathbf{a}_i \rangle = 1$, lo que muestra que la longitud del vector \mathbf{a}_i es igual a 1. En virtud de lo anterior, la matriz \mathbf{A} puede servir para representar geoméricamente la **nube de las variables**: la variable z_i está representada por un punto (\mathbf{a}_i) sobre la esfera unitaria de \mathbf{R}^N , mientras que el coseno entre \mathbf{a}_i y \mathbf{a}_j es igual al coeficiente de correlación entre las variables z_i y z_j . Para facilitar la visualización, se proyecta la posición de las variables sobre los planos creados por los primeros ejes factoriales. Los gráficos que se obtiene se llaman *círculos de correlaciones*,

pues las variables están representadas por puntos dentro de un círculo unitario (proyección de la esfera). Como regla general, la proximidad – o, al contrario, el alejamiento – de las proyecciones de las variables indica su mayor o menor correlación – o antagonismo. Sin embargo, hay que tener cuidado que la proximidad dentro de un círculo no implica necesariamente una proximidad sobre la esfera, salvo si las proyecciones se ubican cerca de la circunferencia del círculo. En caso contrario, para evitar conclusiones erróneas, conviene examinar las proyecciones de la nube de las variables sobre varios planos factoriales.

A continuación, damos una ilustración del análisis en componentes principales con los datos de contaminación de suelo (Figura 11). La primera imagen representa la proyección de la nube de los individuos en el primer plano factorial, con 9 individuos destacados en la nube y en el mapa de los datos (segunda imagen). La tercera imagen muestra el histograma de los valores propios de la matriz de varianza – covarianza (varianzas de los factores), donde se aprecia que las varianzas de los dos primeros factores son mayores que 1 y suman más de 75% de la varianza total. La última imagen representa la nube de las variables en el primer plano factorial, destacando dos grupos de variables (Cu-Pb y Cd-Co-Cr-Ni) con buenas correlaciones intra-grupo y poca correlación inter-grupo.

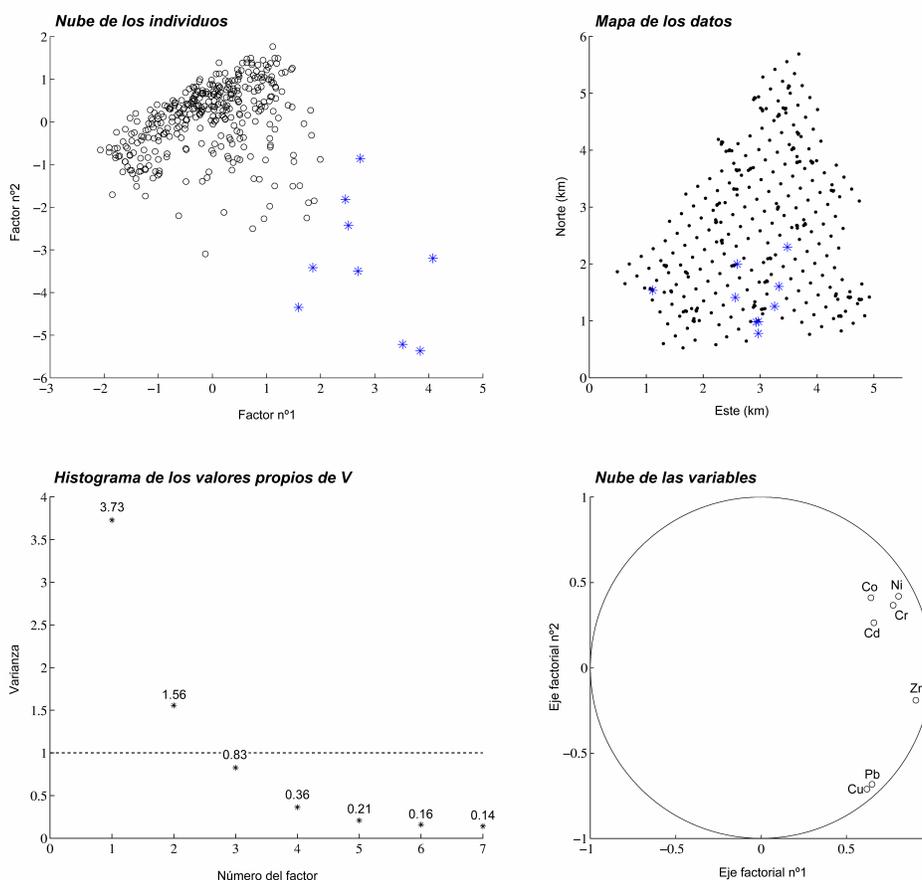


Figura 11. Gráficos de análisis en componentes principales

4. Comportamiento espacial

4.1. Nube direccional

Para hacerse una idea más precisa de la distribución espacial de los valores más allá de un simple despliegue (mapa), es cómodo visualizar las nubes de los valores a lo largo de las direcciones de coordenadas (Figura 12). Estas nubes permiten ver si el comportamiento de los datos experimenta algún cambio en cuanto a su valor promedio, su dispersión o su continuidad, así como detectar los datos más atípicos (aquellos que se destacan del resto de las nubes).

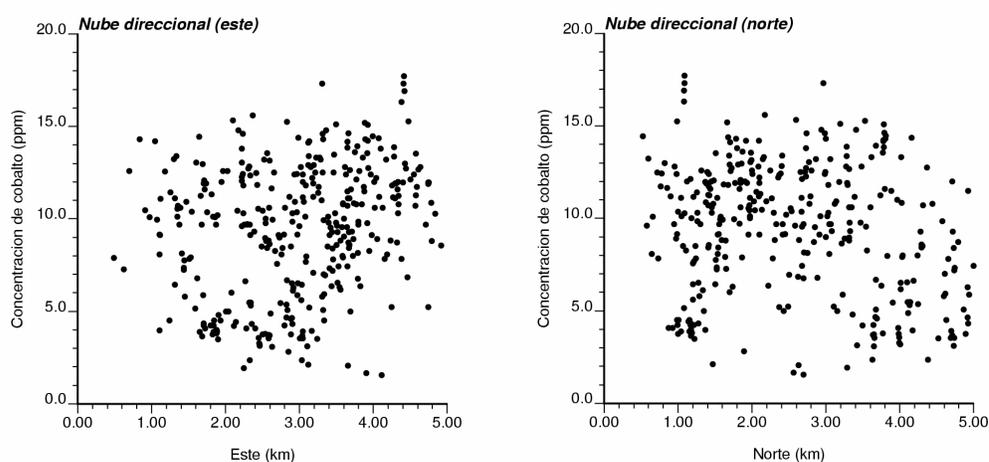


Figura 12. Nubes direccionales, mostrando las concentraciones de cobalto en función de las coordenadas este y norte. No se aprecian tendencias sistemáticas ni datos atípicos.

4.2. Nube de correlación diferida

Para terminar este panorama sobre herramientas de estudio exploratorio, presentaremos la nube de correlación diferida. Se trata de la nube de puntos $(z(\mathbf{x}_\alpha), z(\mathbf{x}_\alpha + \mathbf{h}))$ donde \mathbf{h} es un vector dado, mientras que \mathbf{x}_α y $\mathbf{x}_\alpha + \mathbf{h}$ son sitios con datos. Los pares de datos con valores muy disímiles corresponderán a los puntos de la nube más alejados de la primera bisectriz. Cuando los datos están ubicados en una grilla regular, se toma un vector \mathbf{h} múltiplo del espaciamiento de esta grilla. En caso contrario, se debe introducir tolerancias en la longitud y la orientación de \mathbf{h} , a falta de que la nube se reduciría a muy pocos puntos. La Figura 13 da una ilustración, para un vector \mathbf{h} de longitud 0.25 km (con una tolerancia de 0.01 km) sin importar la orientación. Los puntos más alejados de la bisectriz han sido puestos en relieve y los pares de datos correspondientes han sido destacados en el mapa de ubicación: se trata de datos cercanos cuyos valores son muy diferentes.

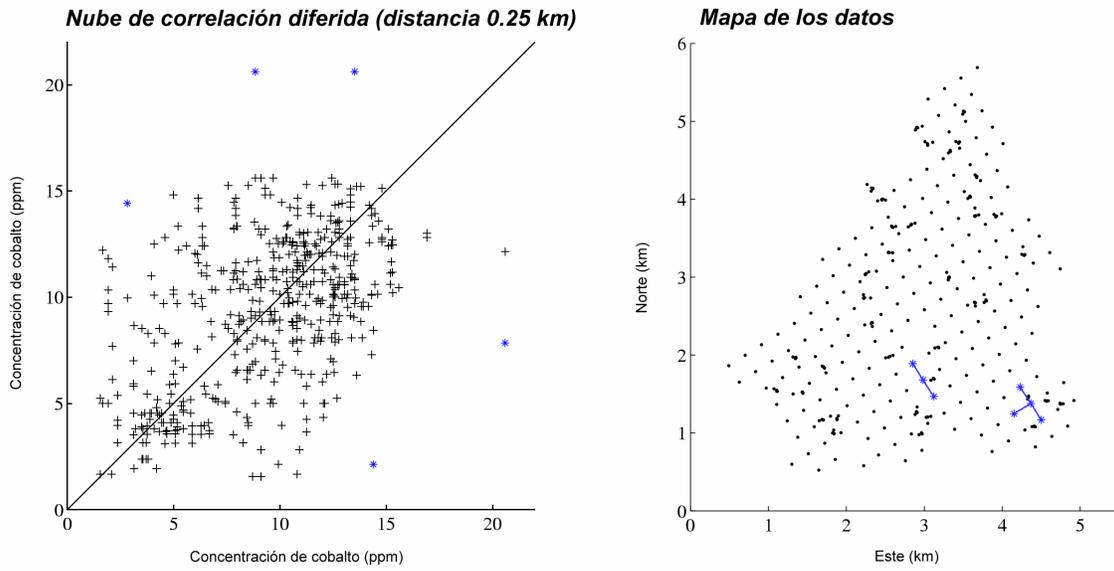


Figura 13. Nube de correlación diferida y mapa de ubicación de los datos.

Capítulo 3. Conceptos generales sobre el modelamiento geoestadístico de datos

Para poder describir y “entender” un fenómeno regionalizado, es necesario elaborar una representación matemática o *modelo*. Una primera solución consiste en utilizar un **modelo determinístico**. En general, este enfoque conduce a una evaluación precisa de la variable regionalizada a partir de un número limitado de observaciones, pero requiere conocer la génesis del fenómeno y las leyes físicas o ecuaciones matemáticas que rigen la distribución de la variable regionalizada. Entre otros dominios de aplicación, citemos

- la meteorología: previsión climática a corto plazo;
- la geofísica: determinación de la intensidad y orientación del campo gravitacional y del campo magnético terrestre en el espacio y el tiempo;
- la teoría de la señal: reconstitución de una señal continua a partir de un conjunto de mediciones, usando propiedades espectrales.

No obstante, en general, los fenómenos regionalizados en estudio son extremadamente complejos y su comprensión puede ser tan parcial que un modelamiento determinístico es imposible o ilusorio. Ejemplos típicos son la evaluación minera, la exploración petrolífera, la caracterización de una zona contaminada o de una parcela agrícola, la estimación de los recursos forestales de una región, o la previsión meteorológica de largo plazo. Estamos entonces obligados a renunciar a una descripción determinística del fenómeno y recurrir a un **modelo probabilístico**. Este proceder resulta operatorio, pues permite formalizar tanto los conocimientos como las incertidumbres que se tiene del fenómeno regionalizado.

1. Límites de la estadística clásica

En estadística clásica, se considera los datos como realizaciones *independientes* de una misma variable aleatoria, es decir, se supone que no tienen relaciones entre sí y que siguen la misma distribución de probabilidad. Se busca estimar los parámetros – en especial, la esperanza y la varianza – de esta distribución, cuya forma a menudo está predeterminada (normal, lognormal, etc.). Sin embargo, cuando los datos están ubicados en el espacio geográfico, las hipótesis de la estadística clásica son raramente aceptables. En particular, si bien simplifica los cálculos estadísticos, la hipótesis de independencia de las observaciones resulta poco realista en el marco espacial. Intuitivamente, observaciones próximas tienen valores cercanos, mientras que aquellas que están más alejadas entre sí tienen una menor relación entre ellas.

Así pues, en general, no puede considerarse modelar las variables regionalizadas por medio de funciones determinísticas, debido a su extrema complejidad. Tampoco se puede asimilar los datos medidos a variables aleatorias independientes. La geoestadística entrega una representación intermedia, a la vez aleatoria y funcional, al basarse en el concepto de *función aleatoria*.

2. Noción de función aleatoria

Los modelos geoestadísticos consideran el valor $z(\mathbf{x})$ de la variable regionalizada en un sitio \mathbf{x} del campo D como una realización de una variable aleatoria $Z(\mathbf{x})$ ¹. Cuando \mathbf{x} recorre D , se obtiene un conjunto de variables aleatorias $Z = \{Z(\mathbf{x}), \mathbf{x} \in D\}$ que constituye una **función aleatoria** (sinónimos: *campo aleatorio*, *proceso aleatorio* o *estocástico*). Así pues, la variable regionalizada $z = \{z(\mathbf{x}), \mathbf{x} \in D\}$ es *una* realización de la función aleatoria Z , pero uno podría imaginar otras realizaciones que presentan características similares en cuanto a cómo se distribuyen los valores en el espacio (Figura 1). Contrariamente al modelo de la estadística clásica, las variables aleatorias así definidas no son independientes; por el contrario, existen interacciones o correlaciones entre ellas, las cuales reflejan la continuidad espacial de la variable regionalizada (Figura 2).

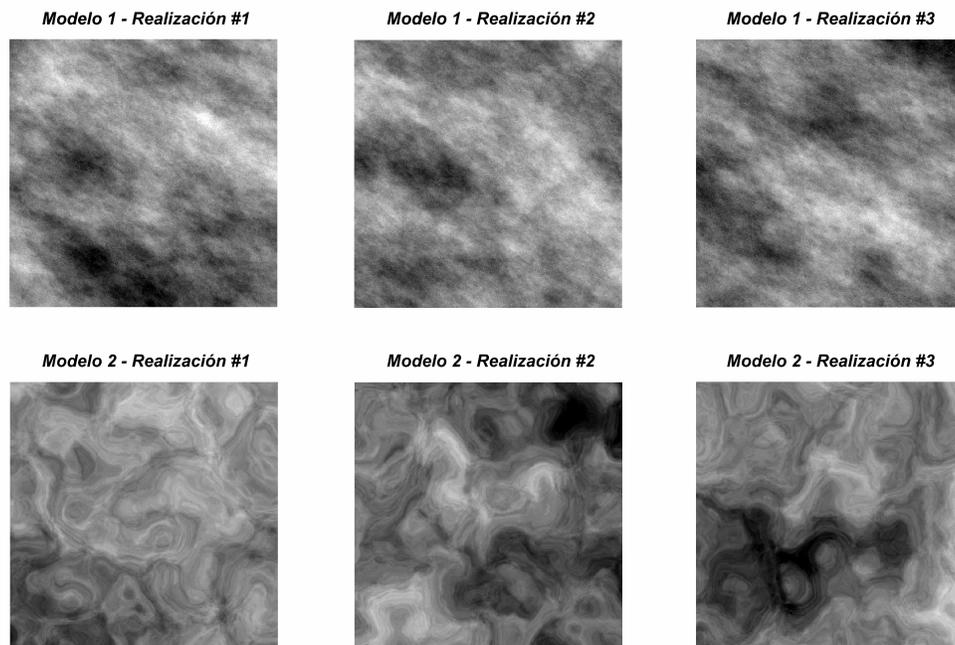


Figura 1. Realizaciones de dos modelos distintos de función aleatoria. Cada modelo define la manera con la cual se distribuyen los valores en el espacio, lo que origina el parentesco que se observa entre las realizaciones de un mismo modelo.

¹ Para distinguir las variables determinísticas de aquellas aleatorias, denotaremos las primeras con *minúscula* y las segundas con *Mayúscula*.

Buena continuidad espacial

1.68	1.31	0.93	0.78	2.57	3.39
2.22	1.07	1.32	1.21	1.71	2.28
3.60	3.24	2.29	1.18	1.09	1.07
2.46	3.39	2.40	2.33	0.50	0.41
2.53	3.43	2.23	1.69	0.72	0.81
1.62	2.62	1.99	0.15	0.01	0.35

Poca continuidad espacial

0.15	2.29	2.46	1.21	0.78	2.53
1.68	2.23	1.07	1.18	2.57	1.32
0.41	0.72	1.62	3.60	0.50	3.39
3.24	0.81	0.35	2.62	2.40	1.71
3.39	2.33	0.93	1.31	2.22	0.01
1.09	2.28	1.69	3.43	1.99	1.07

Figura 2. Noción de correlación para representar la continuidad espacial. Ambos ejemplos presentan los mismos valores, pero distribuidos de forma diferente en el espacio. Las variables aleatorias se modelarán con altas correlaciones en el primer caso y bajas correlaciones en el segundo caso.

El recurrir al concepto de función aleatoria es una *decisión*, ni verdadera ni falsa, pues dicha función aleatoria es un objeto teórico que no existe en la realidad. Asimismo, la determinación de una función aleatoria a partir de una variable regionalizada no es una operación unívoca: varios modelos pueden resultar aceptables, en cuanto sean compatibles con la información disponible sobre la variable regionalizada.

3. Caracterización de una función aleatoria

3.1. Distribución espacial

Consideremos una función aleatoria $Z = \{Z(\mathbf{x}), \mathbf{x} \in D\}$ y una serie de sitios $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ en D . El grupo de variables aleatorias $\{Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k)\}$ está caracterizado por una *función de distribución* multivariable que depende de k argumentos:

$$F_{\mathbf{x}_1, \dots, \mathbf{x}_k}(z_1, \dots, z_k) = \text{Prob}\{Z(\mathbf{x}_1) < z_1, \dots, Z(\mathbf{x}_k) < z_k\} \quad \forall z_1, \dots, z_k \in \mathbb{R}.$$

El conjunto de funciones de distribución, para todos los enteros k y todas las elecciones posibles de $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ en D , constituye la **distribución espacial** de la función aleatoria. Ahora, siendo finito el número de datos disponibles sobre la variable regionalizada, es ilusorio querer inferir la distribución espacial completa de la función aleatoria, que contiene un número infinito de distribuciones de probabilidad. Por ende, la puesta en marcha de las herramientas probabilísticas requiere simplificaciones.

3.2. Distribuciones univariable y bivariadas

Estas distribuciones corresponden a los casos particulares donde $k = 1$ y $k = 2$. En general, los datos disponibles permiten inferir estas distribuciones. Es la razón por la cual la determinación de un modelo de distribución espacial suele basarse en dichas distribuciones, aunque las distribuciones de orden superior (trivariadas, cuadrivariadas...) del modelo no se respaldan en la información proporcionada por los datos.

- **Distribución univariable:**

$$F_{x_1}(z_1) = \text{Prob}\{Z(\mathbf{x}_1) < z_1\}.$$

F_{x_1} corresponde a la función de distribución de la variable aleatoria $Z(\mathbf{x}_1)$ ubicada en \mathbf{x}_1 (Figura 3, izquierda).

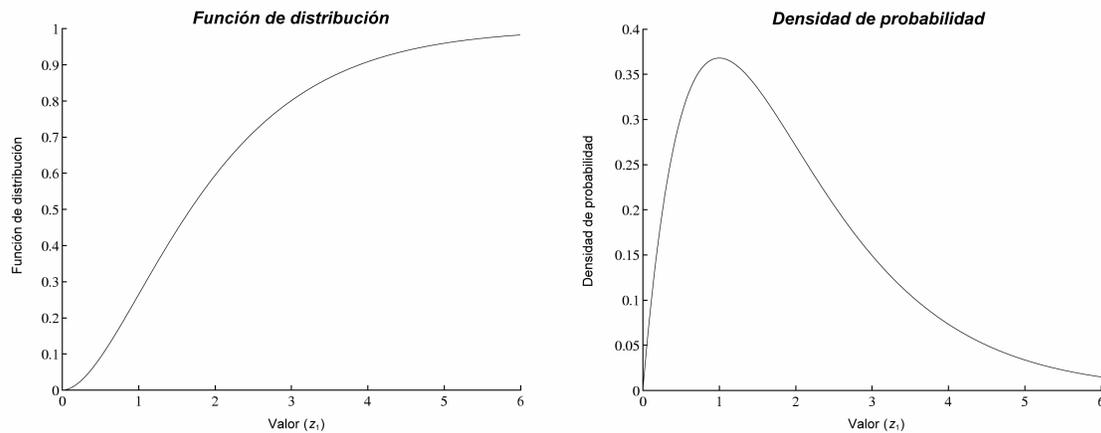


Figura 3. Ejemplo de función de distribución y densidad de probabilidad univariable. La densidad de probabilidad se obtiene al derivar la función de distribución.

- **Distribución bivariada:**

$$F_{x_1, x_2}(z_1, z_2) = \text{Prob}\{Z(\mathbf{x}_1) < z_1, Z(\mathbf{x}_2) < z_2\}.$$

F_{x_1, x_2} corresponde a la función de distribución conjunta del par de variables aleatorias $\{Z(\mathbf{x}_1), Z(\mathbf{x}_2)\}$ (Figura 4, izquierda).

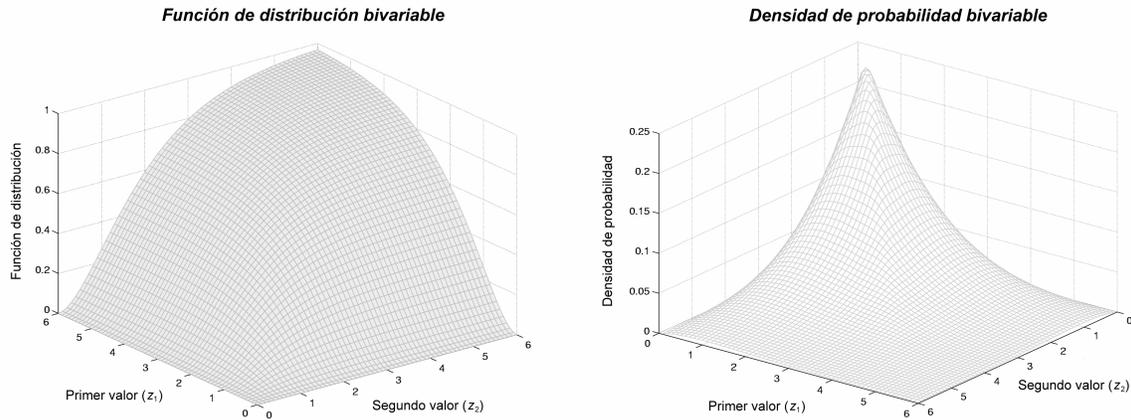


Figura 4. Ejemplo de función de distribución y densidad de probabilidad bivariable. La densidad de probabilidad se obtiene al tomar las derivadas parciales de la función de distribución con respecto a cada uno de sus argumentos.

3.3. Momentos

En muchos problemas (como el de interpolación por kriging), se puede simplificar aun más la caracterización de la función aleatoria, al considerar solamente algunos parámetros descriptivos o *momentos* de las distribuciones univariadas y bivariadas, que “resumen” la información más relevante.

Estos son:

- El **valor esperado** (esperanza, o momento de primer orden):

$$m(\mathbf{x}) = E[Z(\mathbf{x})].$$

En cada sitio \mathbf{x} dado, $m(\mathbf{x})$ representa la “media” alrededor de la cual se distribuyen los valores tomados por las realizaciones de la función aleatoria.

- La **varianza**, o varianza *a priori*, definida por:

$$\begin{aligned} \sigma^2(\mathbf{x}) &= \text{var}[Z(\mathbf{x})] \\ &= E\{[Z(\mathbf{x}) - m(\mathbf{x})]^2\} \\ &= E[Z(\mathbf{x})^2] - m(\mathbf{x})^2 \end{aligned}$$

La varianza es una cantidad positiva. Su raíz cuadrada se llama **desviación estándar**. La varianza y la desviación estándar constituyen medidas de la dispersión de $Z(\mathbf{x})$ en torno a su valor medio $m(\mathbf{x})$ y cuantifican, de esta forma, su carácter “aleatorio”.

- La **covarianza** centrada entre dos variables aleatorias $Z(\mathbf{x}_1)$ y $Z(\mathbf{x}_2)$:

$$\begin{aligned} C(\mathbf{x}_1, \mathbf{x}_2) &= \text{cov}[Z(\mathbf{x}_1), Z(\mathbf{x}_2)] \\ &= E\{[Z(\mathbf{x}_1) - m(\mathbf{x}_1)][Z(\mathbf{x}_2) - m(\mathbf{x}_2)]\} \\ &= E[Z(\mathbf{x}_1)Z(\mathbf{x}_2)] - m(\mathbf{x}_1)m(\mathbf{x}_2) \end{aligned}$$

La covarianza da una visión elemental del vínculo o “interacción” que existe entre $Z(\mathbf{x}_1)$ y $Z(\mathbf{x}_2)$. La desigualdad de Cauchy-Schwarz relaciona la covarianza entre $Z(\mathbf{x}_1)$ y $Z(\mathbf{x}_2)$ con las varianzas de $Z(\mathbf{x}_1)$ y $Z(\mathbf{x}_2)$:

$$|\text{cov}[Z(\mathbf{x}_1), Z(\mathbf{x}_2)]| \leq \sqrt{\text{var}[Z(\mathbf{x}_1)] \text{var}[Z(\mathbf{x}_2)]}.$$

- El **correlograma** (coeficiente de correlación lineal) entre dos variables aleatorias $Z(\mathbf{x}_1)$ y $Z(\mathbf{x}_2)$:

$$\begin{aligned} \rho(\mathbf{x}_1, \mathbf{x}_2) &= \text{corr}[Z(\mathbf{x}_1), Z(\mathbf{x}_2)] \\ &= \frac{\text{cov}[Z(\mathbf{x}_1), Z(\mathbf{x}_2)]}{\sqrt{\text{var}[Z(\mathbf{x}_1)] \text{var}[Z(\mathbf{x}_2)]}}. \end{aligned}$$

Al contrario de la covarianza, el correlograma es adimensional y toma sus valores en el intervalo $[-1, 1]$. Un coeficiente ρ nulo indica que las variables $Z(\mathbf{x}_1)$ y $Z(\mathbf{x}_2)$ no están correlacionadas (condición necesaria para que sean independientes), mientras que un coeficiente igual a 1 ó -1 indica que son proporcionales.

- El **semi-variograma** entre dos variables aleatorias $Z(\mathbf{x}_1)$ y $Z(\mathbf{x}_2)$:

$$\gamma(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{2} \text{var}[Z(\mathbf{x}_1) - Z(\mathbf{x}_2)].$$

En adelante, para aliviar la escritura, se omitirá sistemáticamente el prefijo “semi” y se hablará solamente de “variograma”.

3.4. Inferencia estadística – Hipótesis de estacionaridad

Para poner en marcha el formalismo probabilístico, es necesario poder determinar, por lo menos parcialmente, la distribución espacial de la función aleatoria a partir de los datos disponibles sobre la variable regionalizada (etapa de *inferencia estadística*). Dos razones impiden poder realizar la inferencia estadística en su forma más general: por una parte, la variable regionalizada sólo es *una* realización de la función aleatoria; por otra parte, esta realización se conoce de manera fragmentaria, en algunos sitios de muestreo.

Para salir de este problema algunas restricciones son necesarias. Recurren a la noción de **estacionaridad**. La idea es permitir la inferencia estadística, reemplazando la repetición sobre las realizaciones de la función aleatoria (inaccesibles, por disponer solamente de una realización) por una repetición en el espacio: se supone que los valores que se encuentran en las diferentes regiones del campo presentan las mismas características y, por ende, pueden considerarse como distintas realizaciones del mismo proceso aleatorio.

Del punto de vista matemático, la hipótesis de estacionaridad consiste en postular que la distribución espacial de la función aleatoria es invariante por traslación, es decir, que las propiedades de un conjunto de datos no dependen de su posición absoluta en el espacio, sino que solamente de sus posiciones relativas. Esto implica las siguientes simplificaciones:

- La **distribución univariable** no depende del sitio considerado

$$F(z_1) = \text{Prob}\{Z(\mathbf{x}_1) < z_1\}.$$

En particular, la **esperanza** y la **varianza** son constantes en el espacio:

$$m = E[Z(\mathbf{x})]$$

$$\sigma^2 = \text{var}[Z(\mathbf{x})]$$

- La **distribución bivariante** sólo depende de la separación entre los sitios considerados:

$$F_{\mathbf{h}}(z_1, z_2) = \text{Prob}\{Z(\mathbf{x} + \mathbf{h}) < z_1, Z(\mathbf{x}) < z_2\}.$$

Lo mismo ocurre para los momentos de esta distribución (**covarianza**, **correlograma** y **variograma**):

$$C(\mathbf{h}) = \text{cov}[Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})]$$

$$\rho(\mathbf{h}) = \text{corr}[Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})]$$

$$\gamma(\mathbf{h}) = \frac{1}{2} \text{var}[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})]$$

Estos parámetros son más importantes que la esperanza y la varianza, pues muestran la “interacción” existente entre dos valores, luego dan una descripción sintética de la continuidad espacial de la variable regionalizada. La covarianza y el correlograma indican qué tan *semejantes* son los valores entre dos sitios, mientras que el variograma indica qué tan *desemejantes* son. Siendo constante la esperanza, se tiene también:

$$\gamma(\mathbf{h}) = \frac{1}{2} E\{[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})]^2\}.$$

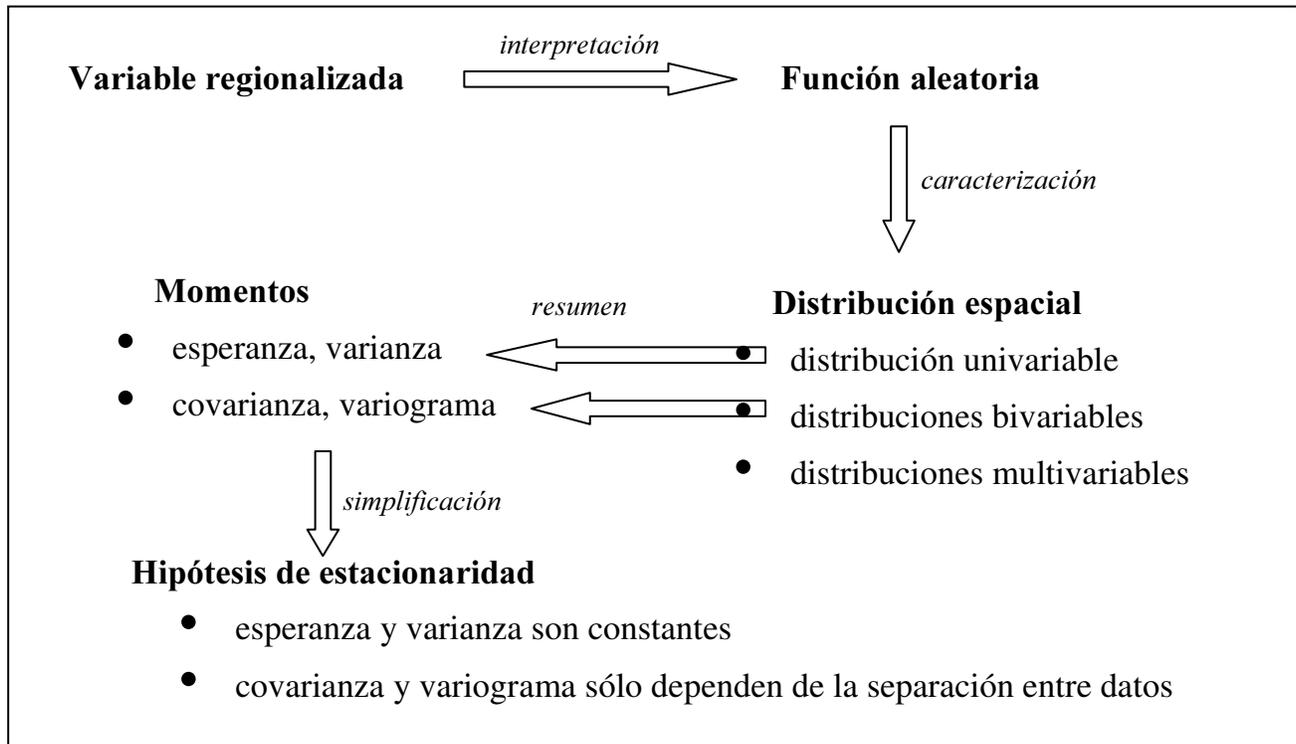


Figura 5. Esquema sintético de los conceptos e hipótesis que sustentan el modelo geoestadístico

3.5. Relaciones entre momentos

Bajo la hipótesis de estacionaridad, se tiene las siguientes relaciones:

- La varianza es igual a la función de covarianza evaluada para el vector $\mathbf{h} = \mathbf{0}$:

$$\sigma^2 = C(\mathbf{0})$$

- El correlograma es igual a la covarianza dividida por la varianza:

$$\rho(\mathbf{h}) = C(\mathbf{h}) / C(\mathbf{0})$$

- El variograma es igual a la varianza menos la covarianza:

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}) .$$

Cuando la norma del vector de separación \mathbf{h} se vuelve infinita, la covarianza tiende a 0 y el variograma es igual a la varianza:

$$\gamma(\infty) = C(\mathbf{0}) = \sigma^2 .$$

Capítulo 4. Análisis variográfico:

1. Variograma experimental

Los valores de una variable regionalizada no son independientes, en el sentido que un valor observado en un sitio proporciona información sobre los valores de los sitios vecinos. En la interpretación probabilística de la variable regionalizada, esta noción intuitiva de dependencia está descrita por la *distribución espacial* de la función aleatoria, que modela la manera como se relacionan los valores observados en distintos sitios por una distribución de probabilidad multivariable.

En muchos problemas la descripción de la distribución espacial se limita a los primeros momentos. El momento de orden 1 (esperanza) hace intervenir un solo sitio a la vez y no entrega realmente información sobre dependencia espacial. En cambio, los momentos de orden 2 (covarianza, correlograma y variograma) están definidos con la ayuda de dos sitios, es decir del más pequeño conjunto que se puede considerar para describir la “interacción” entre valores. Son estos momentos los que entregan una descripción elemental y operatoria de la continuidad espacial de la variable regionalizada.

En este capítulo, abordamos la primera etapa del análisis variográfico, que consiste en la inferencia del variograma, es decir, el cálculo de un variograma experimental a partir de los datos disponibles. Posteriormente (capítulo siguiente), se verá cómo ajustar un modelo de variograma en torno al variograma experimental.

1. El variograma experimental tradicional

1.1. Definición e interpretación

Consideremos una variable regionalizada z conocida en n sitios $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$. El estimador tradicional del variograma para un vector de separación \mathbf{h} dado, se define de la siguiente manera:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{N(\mathbf{h})} [z(\mathbf{x}_\alpha) - z(\mathbf{x}_\beta)]^2$$

donde $N(\mathbf{h}) = \{(\alpha, \beta) \text{ tal que } \mathbf{x}_\alpha - \mathbf{x}_\beta = \mathbf{h}\}$;
 $|N(\mathbf{h})|$ es el número de pares contenidos en el conjunto $N(\mathbf{h})$.

Se ve que el estimador anterior consiste en reemplazar la esperanza en la expresión del variograma teórico

$$\gamma(\mathbf{h}) = \frac{1}{2} E\{[Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})]^2\}$$

por la media aritmética sobre los pares de datos separados por el vector \mathbf{h} .

El estimador así definido lleva el nombre de **variograma experimental**. No se trata de una función propiamente tal, sino que de una serie de valores, pues sólo se puede calcular para vectores \mathbf{h} tales que $N(\mathbf{h})$ no es vacío. El variograma experimental para un vector \mathbf{h} puede interpretarse como el momento de inercia de la nube de correlación diferida (nube de los puntos $(z(\mathbf{x}_\alpha), z(\mathbf{x}_\beta))$ con $\mathbf{x}_\alpha - \mathbf{x}_\beta = \mathbf{h}$), que mide la distancia cuadrática promedio entre los puntos de la nube y la línea diagonal. Mientras más apretada la nube de correlación diferida en torno a la diagonal, más pequeña su inercia.

1.2. Ejemplo

Consideremos los siguientes datos espaciados cada 100 m:

5 3 6 4 2 1 1 2 4 3 2

El variograma experimental se puede calcular para distancias múltiplos de 100m, esto es:

$$\hat{\gamma}(100m) = \frac{1}{2 \times 10} (2^2 + 3^2 + 2^2 + 2^2 + 1^2 + 0^2 + 1^2 + 2^2 + 1^2 + 1^2) = 1.45$$

$$\hat{\gamma}(200m) = \frac{1}{2 \times 9} (1^2 + 1^2 + 4^2 + 3^2 + 1^2 + 1^2 + 3^2 + 1^2 + 2^2) = 2.39$$

$$\hat{\gamma}(300m) = \frac{1}{2 \times 8} (1^2 + 1^2 + 5^2 + 3^2 + 0^2 + 3^2 + 2^2 + 0^2) = 3.06$$

...

$$\hat{\gamma}(1000m) = \frac{1}{2 \times 1} (3^2) = 4.5$$

1.3. Tolerancias en los parámetros de cálculo

Si los datos están irregularmente distribuidos en el campo D , el número de pares $|N(\mathbf{h})|$ que interviene en el cálculo de $\hat{\gamma}(\mathbf{h})$ para un vector \mathbf{h} dado, es generalmente muy pequeño (incluso igual a 0 ó 1). El variograma experimental tiene entonces un aspecto muy errático y resulta imposible interpretarlo y modelarlo. Para que sea más robusto, se suele permitir algunas tolerancias de cálculo, sobre las distancias y las direcciones:

$$\hat{\gamma}^+(\mathbf{h}) = \frac{1}{2|N^+(\mathbf{h})|} \sum_{N^+(\mathbf{h})} [z(\mathbf{x}_\alpha) - z(\mathbf{x}_\beta)]^2$$

donde $N^+(\mathbf{h}) = \{(\alpha, \beta) \text{ tal que } \mathbf{x}_\alpha - \mathbf{x}_\beta \in T(\mathbf{h})\} = \bigcup_{\mathbf{h}' \in T(\mathbf{h})} N(\mathbf{h}')$;

$T(\mathbf{h})$ es una región de tolerancia alrededor de \mathbf{h} , de la forma $[\mathbf{h} - \Delta\mathbf{h}, \mathbf{h} + \Delta\mathbf{h}]$ en el caso unidimensional. En el caso bi- o tridimensional, existen tolerancias tanto sobre la longitud de \mathbf{h} como sobre su orientación, tal como se ilustra en la Figura 1.

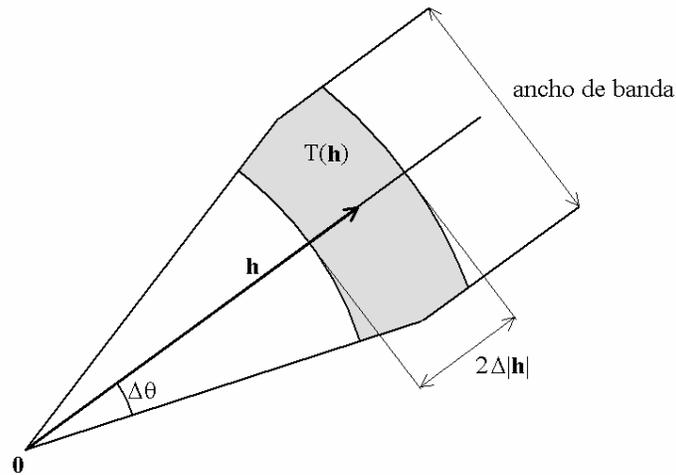


Figura 1. Región de tolerancia $T(\mathbf{h})$ alrededor del vector \mathbf{h} (caso bidimensional)

El ancho de banda limita la separación del cono de tolerancia a una extensión máxima. En el espacio de tres dimensiones, se introduce dos anchos de banda: uno horizontal y otro vertical.

En resumen, los parámetros a especificar para calcular un variograma experimental son los siguientes:

- dirección de interés: acimut, inclinación
- distancias de interés, en general múltiplos de una distancia elemental llamada *paso* (“lag”)
- tolerancia en la dirección: tolerancia angular, ancho(s) de banda
- tolerancia en las distancias.

La Figura 2 muestra un variograma experimental de las concentraciones de cobalto procedentes de los datos de contaminación de suelo. Este variograma ha sido calculado para distancias múltiplos de la malla de muestreo (0.25 km), con una tolerancia en la distancia de 0.125 km y una tolerancia angular de 90°, es decir, sin importar la orientación del vector \mathbf{h} . En este caso, se habla de variograma “omnidireccional”.

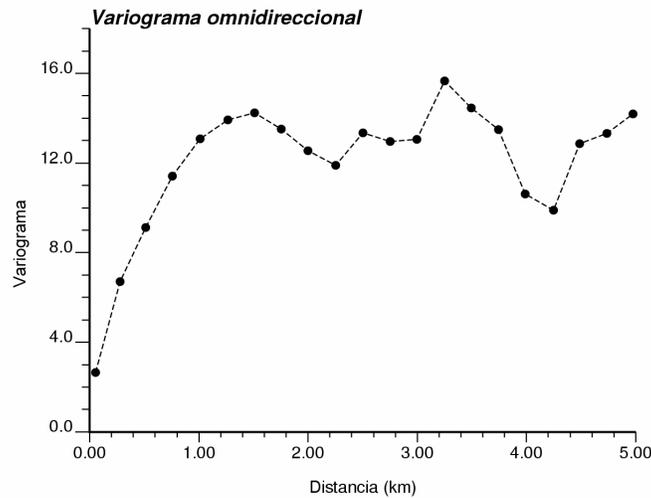


Figura 2. Variograma experimental omnidireccional de las concentraciones de cobalto. Las líneas punteadas sólo unen a los distintos puntos del variograma experimental.

1.4. Propiedades del variograma experimental

El variograma experimental $\hat{\gamma}(\mathbf{h})$ es un estimador **insesgado** del variograma teórico:

$$E[\hat{\gamma}(\mathbf{h})] = \gamma(\mathbf{h}) .$$

Un indicador de la **robustez** de $\hat{\gamma}(\mathbf{h})$ es su varianza relativa

$$\text{var}[\hat{\gamma}(\mathbf{h})] / [\gamma(\mathbf{h})]^2 .$$

Mientras más elevada dicha varianza, más susceptible es el variograma experimental de fluctuar en torno a su valor esperado (el variograma teórico $\gamma(\mathbf{h})$) y más difícil se vuelve la inferencia estadística. Aunque esta varianza relativa sólo puede ser expresada en algunos casos particulares, puesto que requiere conocer la función aleatoria hasta sus distribuciones cuadrivariadas, los principales factores que la influyen son:

- La distancia considerada (norma del vector \mathbf{h}): la varianza relativa de $\hat{\gamma}(\mathbf{h})$ suele tomar valores considerables para las grandes distancias (para fijar las ideas, aquellas distancias mayores que la mitad del diámetro del campo). Este efecto es perceptible en la Figura 2, donde se observa que las fluctuaciones aumentan cuando la distancia aumenta.

- La irregularidad o el carácter preferencial de la malla de muestreo, que pueden provocar grandes fluctuaciones en el variograma experimental, incluso para pequeñas distancias.
- El número de pares de datos: mientras más bajo, mayores son las fluctuaciones.
- La presencia de datos extremos (*outliers*), los cuales tienen un impacto considerable en el cálculo del variograma experimental, pues este último eleva los valores al cuadrado.

1.5. Nube variográfica

Para estudiar la estabilidad numérica del variograma experimental, es útil visualizar la *nube variográfica*, es decir, la nube de las diferencias cuadráticas $\{[z(\mathbf{x}_\alpha) - z(\mathbf{x}_\beta)]^2 / 2, \text{ con } (\alpha, \beta) \in N(\mathbf{h})\}$ en función del vector \mathbf{h} o de su módulo $|\mathbf{h}|$. El variograma experimental se obtiene al tomar, para cada vector \mathbf{h} , la media de los puntos correspondientes de la nube variográfica. El análisis de esta nube permite localizar los pares de datos responsables de los valores altos del variograma experimental y poner en evidencia los datos notablemente diferentes de sus vecinos.

En lugar de visualizar la nube entera de las diferencias cuadráticas, se puede también sólo localizar sus extremos, su media (que no es otra cosa que el valor del variograma experimental para el vector \mathbf{h}) y sus cuartiles. Si la media está fuera del rango intercuartil (intervalo entre el primer y tercer cuartil), esto puede indicar que el valor del variograma experimental para el vector considerado está fuertemente influenciado por algunos valores extremos, luego es poco robusto.

La Figura 3 presenta la nube variográfica (calculada de forma omnidireccional) de los datos de concentración de cobalto, donde se ha resaltado los puntos de la nube que exceden un valor de 130. La línea punteada indica la media de los puntos de la nube, o sea, el variograma experimental. La parte derecha de la figura muestra el mapa de ubicación de los datos, donde se indica los pares de datos asociados a los puntos resaltados de la nube variográfica. Todos estos pares se originan en un mismo dato, que corresponde al dato de mayor concentración de cobalto (20.6 ppm).

Este dato de 20.6 ppm tiene mucho impacto en el variograma experimental. La Figura 4 muestra el variograma experimental que se obtendría si este dato no existiera (lado izquierdo) y el variograma experimental que se obtendría si este dato tuviera un valor de 35 ppm en lugar de 20.6 ppm (lado derecho). Se aprecia que la presencia o ausencia de un solo dato extremo puede afectar considerablemente el variograma experimental, en particular para las distancias pequeñas. Para paliar esta situación, en la práctica se puede recurrir a las siguientes aproximaciones:

- Utilizar una herramienta alternativa al variograma (ver sección siguiente).
- No considerar los “outliers” al momento de calcular el variograma.
- Bajar arbitrariamente el valor de los outliers (“capping”), por ejemplo asignando un valor convencional de 15 ppm a todos los datos que superan este valor.

- Transformar los datos, por ejemplo pasándolos a logaritmo. Ahora, si bien el logaritmo atenúa los valores extremos y facilita el modelamiento del variograma, la dificultad de este procedimiento radica en que, en general, se desconoce la relación entre el variograma de la variable original y el variograma de su logaritmo, por lo que una hipótesis adicional es necesaria (ver sección siguiente).

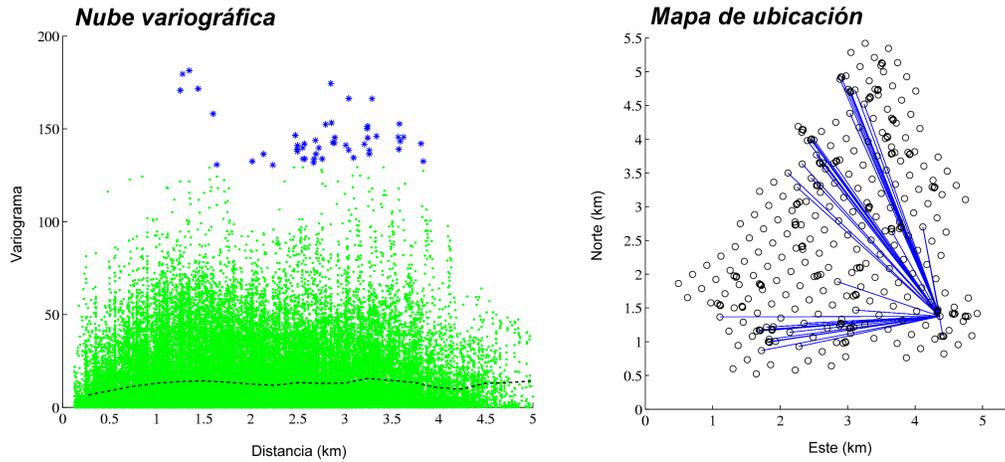


Figura 3. Izquierda: nube variográfica para los datos de concentración de cobalto. Derecha: mapa de ubicación.

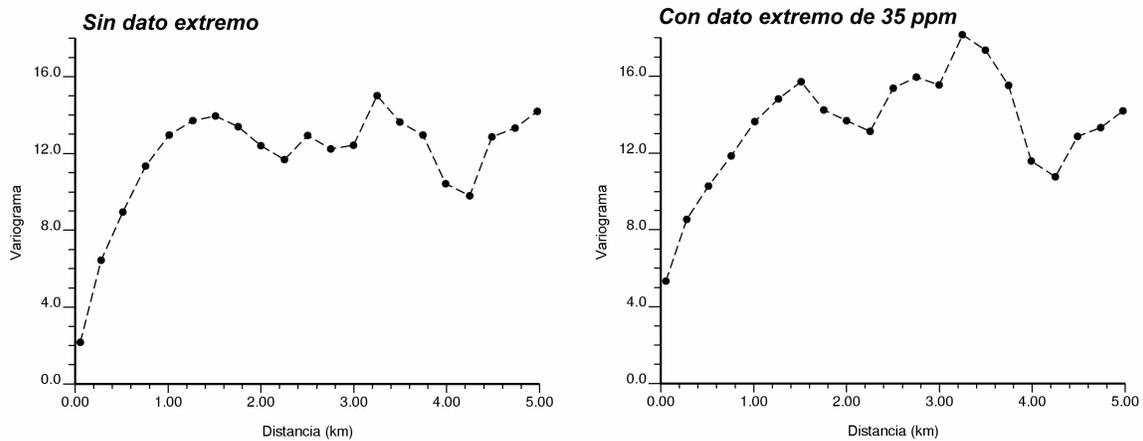


Figura 4. Variograma experimental calculado sin el dato extremo (izquierda) y con un dato extremo de 35 ppm en lugar de 20.6 ppm (derecha).

1.6. Mapa variográfico

El mapa variográfico consiste en visualizar el variograma experimental en todas las direcciones del espacio, bajo la forma de un mapa con escala de color o de grises. Ayuda a distinguir si existe **anisotropía**, para luego calcular el variograma experimental a lo largo de las direcciones principales de anisotropía.

A modo de ilustración, el mapa variográfico de los datos de concentración de cobalto (Figura 5, izquierda) dibuja una dirección de mayor continuidad con una orientación de 60° aproximadamente desde el eje norte-sur (los colores claros indican valores pequeños para el variograma experimental, mientras que los colores oscuros indican valores más grandes). Esta observación se confirma al calcular los variogramas experimentales direccionales (Figura 5, derecha), donde se aprecia un crecimiento más lento en la dirección N60°E y más rápido en la dirección ortogonal N30°W. Para el modelamiento, se preferirá utilizar estos variogramas direccionales en lugar del variograma omnidireccional (Figura 2), puesto que este último no captura el cambio de continuidad espacial con la dirección (anisotropía).

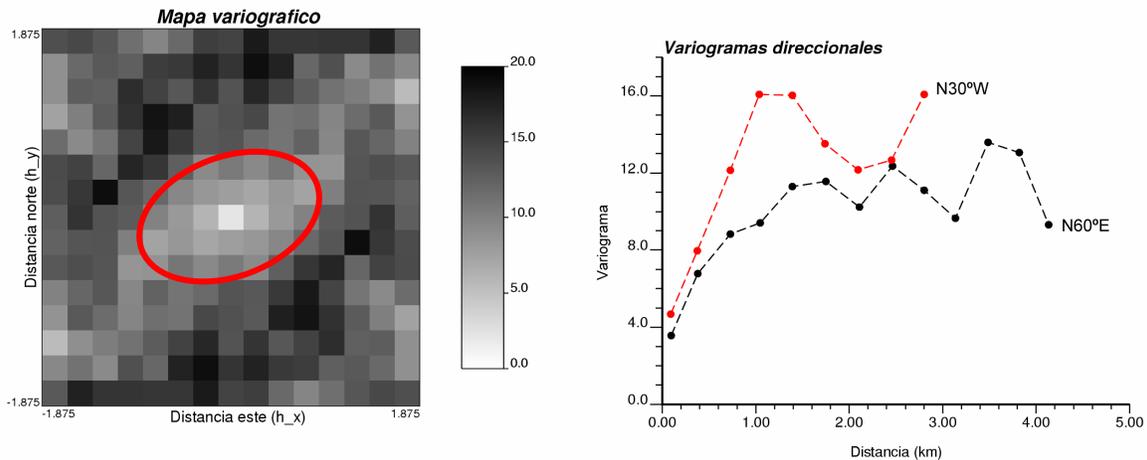


Figura 5. Mapa variográfico (izquierda) y variograma experimental calculado a lo largo de las direcciones de anisotropía reconocidas (N60°E y N30°W) (derecha). Los parámetros de cálculo son los siguientes: paso = 0.35 km, tolerancia en la distancia = 0.175 km, tolerancia angular = 20°, ancho de banda = 1km.

2. Consideraciones prácticas

- 1) Las direcciones de cálculo del variograma experimental deben considerar la anisotropía de la variable regionalizada. Tal como en el ejemplo anterior, su elección se puede hacer al examinar el mapa variográfico. En el caso isótropo (es decir, si los variogramas direccionales se superponen salvo por pequeñas fluctuaciones estadísticas), se podrá considerar un variograma “omnidireccional”:

$$\bar{\gamma}^+(r) = \frac{1}{2|N^+(r)|} \sum_{N^+(r)} [z(\mathbf{x}_\alpha) - z(\mathbf{x}_\beta)]^2$$

donde $N^+(r) = \{ (\alpha, \beta) \text{ tal que } |\mathbf{x}_\alpha - \mathbf{x}_\beta| \approx r \}$.

- 2) En general, cada variograma direccional se calcula para distancias múltiplos de una distancia elemental, llamada “paso”. La elección del paso puede tener repercusiones en el cálculo. Un paso demasiado pequeño conducirá a un variograma experimental errático, poco estable, mientras que un paso demasiado grande no permitirá apreciar los detalles de la continuidad espacial, en especial el comportamiento en el origen del variograma. Cuando la malla de muestreo es regular², el paso está prácticamente impuesto. En caso contrario, se podrá utilizar la “nube variográfica” para encontrar un paso tal que el número de pares de datos que intervienen en el cálculo del variograma experimental sea suficiente (para fijar ideas, superior a la mitad del número de datos), de manera que estos puntos sean representativos.
- 3) El uso de tolerancias en las distancias y ángulos busca suavizar y hacer más robusto el variograma experimental cuando la malla de muestreo no es regular. Ahora, conviene advertir el usuario de la utilización de tolerancias excesivas, que pueden suavizar artificialmente el variograma experimental; aunque un variograma suave es más fácil de modelar, esto no garantiza que sea más representativo de los datos.
- 4) En general, se utiliza una tolerancia en las distancias igual a la mitad del paso: en la dirección considerada, el i -ésimo punto del variograma experimental se calcula con los pares de datos cuya separación cae en el intervalo $[(i - 0.5) \times \text{paso}, (i + 0.5) \times \text{paso}]$. De esta manera, todas las distancias son utilizadas una vez y una sola. Se puede también considerar una tolerancia menor, en cuyo caso las clases de distancias involucradas en el cálculo del variograma experimental ya no son contiguas y algunos pares de datos no son tomados en cuenta. Esta situación, aparentemente poco favorable (pues se pierde información) no es siempre la peor, sobre todo cuando se trabaja en una malla regular o casi-regular.
- 5) Las tolerancias angulares tienden a disipar la eventual anisotropía del variograma (las direcciones con mayor continuidad espacial se mezclan con direcciones de continuidad intermedia, de manera que la anisotropía aparece menos marcada de lo que está en realidad). Por ende, hay que buscar un justo equilibrio en la elección de las tolerancias. Es recomendable completar la definición de las tolerancias angulares por aquella de las separaciones máximas: *ancho de banda* y, en el espacio de tres dimensiones, *alto de banda*; este último parámetro resulta importante en la práctica, pues evita mezclar las direcciones horizontales con direcciones inclinadas, cuyo comportamiento es a menudo bastante distinto.
- 6) Es conveniente tener cuidado con la representatividad del variograma experimental. Para cada punto de este variograma, se podrá determinar el número de pares de datos utilizados para su cálculo; un número de pares bajo indica un punto poco confiable. Igualmente, resulta de utilidad visualizar la nube variográfica, para identificar los pares de datos responsables de una inestabilidad numérica del variograma experimental.

² Por ejemplo, para una malla cuadrada de lado a , se elegirá un paso de cálculo igual a a en las direcciones principales de la malla, y $a\sqrt{2}$ en las direcciones diagonales.

3. Complemento: otras herramientas para el análisis variográfico

3.1. Covarianza experimental

Bajo la hipótesis de estacionaridad, la función de covarianza se define como:

$$C(\mathbf{h}) = E\{[Z(\mathbf{x} + \mathbf{h}) - m][Z(\mathbf{x}) - m]\}.$$

Un primer estimador de esta covarianza se obtiene al reemplazar el valor esperado m por una media aritmética sobre los pares de datos cuya separación coincide con el vector \mathbf{h} :

$$\hat{C}_1(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} [z(\mathbf{x}_\alpha) - \bar{z}][z(\mathbf{x}_\beta) - \bar{z}],$$

donde $N(\mathbf{h}) = \{(\alpha, \beta) \text{ tal que } \mathbf{x}_\alpha - \mathbf{x}_\beta = \mathbf{h}\}$, mientras que \bar{z} es un estimador de la esperanza, a saber la media aritmética de los n datos disponibles:

$$\bar{z} = \frac{1}{n} \sum_{\alpha=1}^n z(\mathbf{x}_\alpha).$$

El estimador \hat{C}_1 se conoce como *covarianza ergódica*. Un segundo estimador, llamado *covarianza no ergódica* consiste en plantear:

$$\hat{C}_2(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} [z(\mathbf{x}_\alpha) - \bar{z}_{cabeza}(\mathbf{h})][z(\mathbf{x}_\beta) - \bar{z}_{cola}(\mathbf{h})],$$

$$\text{con } \bar{z}_{cabeza}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{(\alpha, \beta) \in N(\mathbf{h})} z(\mathbf{x}_\alpha) \text{ y } \bar{z}_{cola}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{(\alpha, \beta) \in N(\mathbf{h})} z(\mathbf{x}_\beta).$$

Se demuestra que estos estimadores de la covarianza son más robustos que el estimador del variograma (los datos no están elevados al cuadrado), pero presentan un **sesgo**³:

³ Por ejemplo, para $\mathbf{h} = \mathbf{0}$, ambos estimadores coinciden con el estimador clásico de la varianza *a priori* $C(\mathbf{0})$:

$$\hat{C}(\mathbf{0}) = \frac{1}{n} \sum_{\alpha=1}^n [z(\mathbf{x}_\alpha) - \bar{z}]^2$$

cuya esperanza es, bajo la hipótesis simplificadora que los datos no están correlacionados

$$E[\hat{C}(\mathbf{0})] = \frac{n-1}{n} C(\mathbf{0}) = (1 - \frac{1}{n}) C(\mathbf{0}).$$

Bajo la hipótesis de no correlación entre los datos, se tiene un sesgo igual a $C(\mathbf{0})/n$ para el estimador de la varianza *a priori* $C(\mathbf{0})$. En realidad, como la variable regionalizada manifiesta cierta continuidad espacial, los datos están correlacionados y el sesgo es todavía más importante.

$$E[\hat{C}_1(\mathbf{h})] \neq C(\mathbf{h})$$

$$E[\hat{C}_2(\mathbf{h})] \neq C(\mathbf{h})$$

El sesgo es más importante cuando el número de datos es pequeño. Proviene de haber recurrido, en el cálculo de la covarianza experimental, a una estimación de la esperanza m por una media aritmética de datos. El variograma experimental no tiene este inconveniente, pues evita estimar la esperanza de la función aleatoria.

De la misma manera que para el variograma, se introduce tolerancias en las distancias y los ángulos cuando la malla de muestreo es irregular.

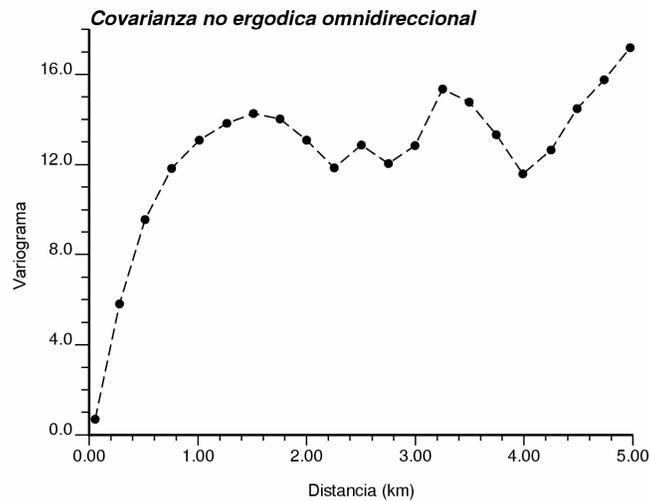


Figura 6. Variograma omnidireccional calculado a partir del estimador de la covarianza no ergódica, planteando $\hat{\gamma}(\mathbf{h}) = \hat{C}_2(\mathbf{h}) - \hat{C}_2(\mathbf{0})$.

3.2. Covarianza no centrada

Una alternativa interesante es recurrir a la *covarianza no centrada*, denotada C_{nc} , que, bajo la hipótesis de estacionaridad, sólo depende de la separación entre datos:

$$C_{nc}(\mathbf{h}) = E[Z(\mathbf{x} + \mathbf{h})Z(\mathbf{x})] = C(\mathbf{h}) + m^2.$$

La covarianza no centrada está relacionada con el variograma por la relación:

$$\gamma(\mathbf{h}) = C_{nc}(\mathbf{h}) - C_{nc}(\mathbf{0})$$

y constituye, en consecuencia, una herramienta equivalente a la covarianza centrada o al variograma. Puede estimarse sin sesgo con:

$$\hat{C}_{nc}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} z(\mathbf{x}_\alpha) z(\mathbf{x}_\beta).$$

De este modo, $\hat{C}_{nc}(\mathbf{0}) - \hat{C}_{nc}(\mathbf{h})$ constituye un estimador insesgado del variograma $\gamma(\mathbf{h})$. Este estimador atenúa la influencia de los valores altos, en comparación con el variograma experimental clásico, puesto que los valores de los datos ya no intervienen en un cuadrado. Esto es particularmente interesante en el caso de datos cuya distribución es muy asimétrica y presenta algunos valores extremos que hacen inestable el variograma experimental.

3.3. Correlograma experimental

El correlograma se define como el cociente entre la covarianza y la varianza:

$$\rho(\mathbf{h}) = \text{corr}[Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})] = C(\mathbf{h})/C(\mathbf{0}).$$

Tal como para la covarianza, se puede definir un estimador *ergódico*, que utiliza todos los datos al momento de estimar la esperanza y la varianza:

$$\rho_1(\mathbf{h}) = \frac{\hat{C}_1(\mathbf{h})}{\hat{C}_1(\mathbf{0})},$$

y un estimador *no ergódico*, que sólo utiliza los datos que están apareados para el vector \mathbf{h} :

$$\hat{\rho}_2(\mathbf{h}) = \frac{\hat{C}_2(\mathbf{h})}{\hat{\sigma}_{cabeza}(\mathbf{h}) \hat{\sigma}_{cola}(\mathbf{h})},$$

$$\text{con } \hat{\sigma}_{cabeza}(\mathbf{h}) = \sqrt{\frac{1}{|N(\mathbf{h})|} \sum_{(\alpha, \beta) \in N(\mathbf{h})} [z(\mathbf{x}_\alpha) - \bar{z}_{cabeza}(\mathbf{h})]^2}$$

$$\hat{\sigma}_{cola}(\mathbf{h}) = \sqrt{\frac{1}{|N(\mathbf{h})|} \sum_{(\alpha, \beta) \in N(\mathbf{h})} [z(\mathbf{x}_\beta) - \bar{z}_{cola}(\mathbf{h})]^2}.$$

Los estimadores así definidos son más robustos que el estimador del variograma, pero sesgados debido a que utilizan estimadores de la esperanza y de la varianza:

$$E[\hat{\rho}_1(\mathbf{h})] \neq \rho(\mathbf{h})$$

$$E[\hat{\rho}_2(\mathbf{h})] \neq \rho(\mathbf{h})$$

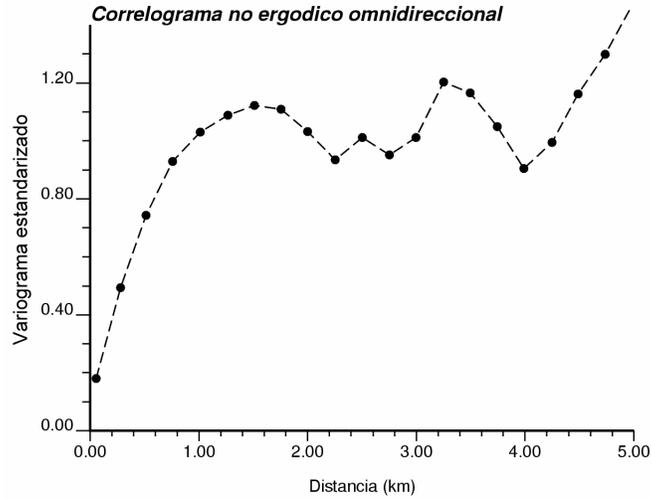


Figura 6. Variograma estandarizado omnidireccional calculado a partir del estimador del correlograma no ergódico, planteando $\hat{\gamma}(\mathbf{h}) = \hat{\rho}_2(\mathbf{0}) - \hat{\rho}_2(\mathbf{h})$.

3.4. Variograma desagrupado

Para corregir los efectos provocados por una malla irregular de muestreo, hemos visto que era usual ponderar los datos (operación de desagrupamiento) al momento de calcular el histograma. Este procedimiento se puede extender al cálculo del variograma experimental, planteando:

$$\tilde{\gamma}(\mathbf{h}) = \frac{1}{2} \sum_{N(\mathbf{h})} \omega_{\alpha\beta} [z(\mathbf{x}_\alpha) - z(\mathbf{x}_\beta)]^2,$$

donde los ponderadores $\{\omega_{\alpha\beta}, (\alpha, \beta) \in N(\mathbf{h})\}$ suman 1.

Varias opciones han sido planteadas para definir estos ponderadores, por ejemplo $\omega_{\alpha\beta}$ podría ser el promedio de los ponderadores de desagrupamiento asignados a los datos $n^\circ\alpha$ y $n^\circ\beta$ por el método de las celdas. Aunque parece interesante para desagrupar el variograma experimental, esta alternativa todavía no es de uso común.

3.5. Otras herramientas

Los estimadores definidos hasta ahora (variograma tradicional, covarianza, covarianza no centrada, correlograma, variograma desagrupado) pueden ser utilizados para calcular un variograma experimental. Como recomendación general, es preferible utilizar estimadores insesgados (luego, evitar la covarianza centrada o el correlograma).

Varias otras herramientas variográficas han sido propuestas en la literatura, pero sus propiedades son mal conocidas y no sustituyen al variograma experimental tradicional. Citemos:

- **Variograma relativo general**

Esta herramienta consiste en normalizar el valor del variograma en \mathbf{h} por el cuadrado de la media de los datos separados por \mathbf{h} :

$$\hat{\gamma}_{RG}(\mathbf{h}) = \frac{\hat{\gamma}(\mathbf{h})}{\hat{m}(\mathbf{h})^2} \quad \text{con} \quad \hat{m}(\mathbf{h}) = \frac{\bar{z}_{cabeza}(\mathbf{h}) + \bar{z}_{cola}(\mathbf{h})}{2}.$$

- **Variograma relativo por pares**

La normalización interviene a nivel de cada par de datos:

$$\hat{\gamma}_{RP}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{N(\mathbf{h})} \frac{[z(\mathbf{x}_\alpha) - z(\mathbf{x}_\beta)]^2}{[z(\mathbf{x}_\alpha) + z(\mathbf{x}_\beta)]^2 / 4}$$

Se plantea un problema en la definición anterior cuando la variable regionalizada puede anularse, caso en el cual el denominador puede ser nulo.

Los variogramas relativos son más robustos que el variograma clásico a la presencia de valores extremos, cuya influencia se reduce a causa de la normalización efectuada. El principal inconveniente de estos variogramas es que no pueden ser utilizados tal cual en el formalismo geoestadístico: las ecuaciones que permiten resolver los problemas de estimación recurren al variograma o la covarianza, pero no a los variogramas relativos. Estos últimos sólo entregan un suplemento de información para mejorar la descripción de la continuidad espacial de los datos (alcance y anisotropía en especial).

- **Variograma logarítmico**

En lugar de interesarse por la variable regionalizada misma, puede ser ventajoso hacer el análisis variográfico de una transformada de ésta. Una transformación frecuente es el paso a logaritmos, planteando $Y = \ln(Z)$ o, más generalmente, $Y = \ln(Z + \mu)$ donde μ es un escalar; esta transformación atenúa las diferencias entre los valores altos.

El problema principal concierne la vuelta a la variable original Z . Para poder deducir el variograma γ_Z de esta última a partir del variograma logarítmico γ_Y , se requiere hipótesis sobre las distribuciones bivariantes de la función aleatoria Y . La hipótesis más frecuente es la de distribuciones Gaussianas, en cuyo caso se tiene:

$$\gamma_Z(\mathbf{h}) = m_Z^2 \exp(\sigma_Y^2) \{1 - \exp[-\gamma_Y(\mathbf{h})]\}$$

donde $m_Z = E(Z)$ y $\sigma_Y^2 = \text{var}(Y) = \ln[1 + \text{var}(Z)/m_Z^2]$.

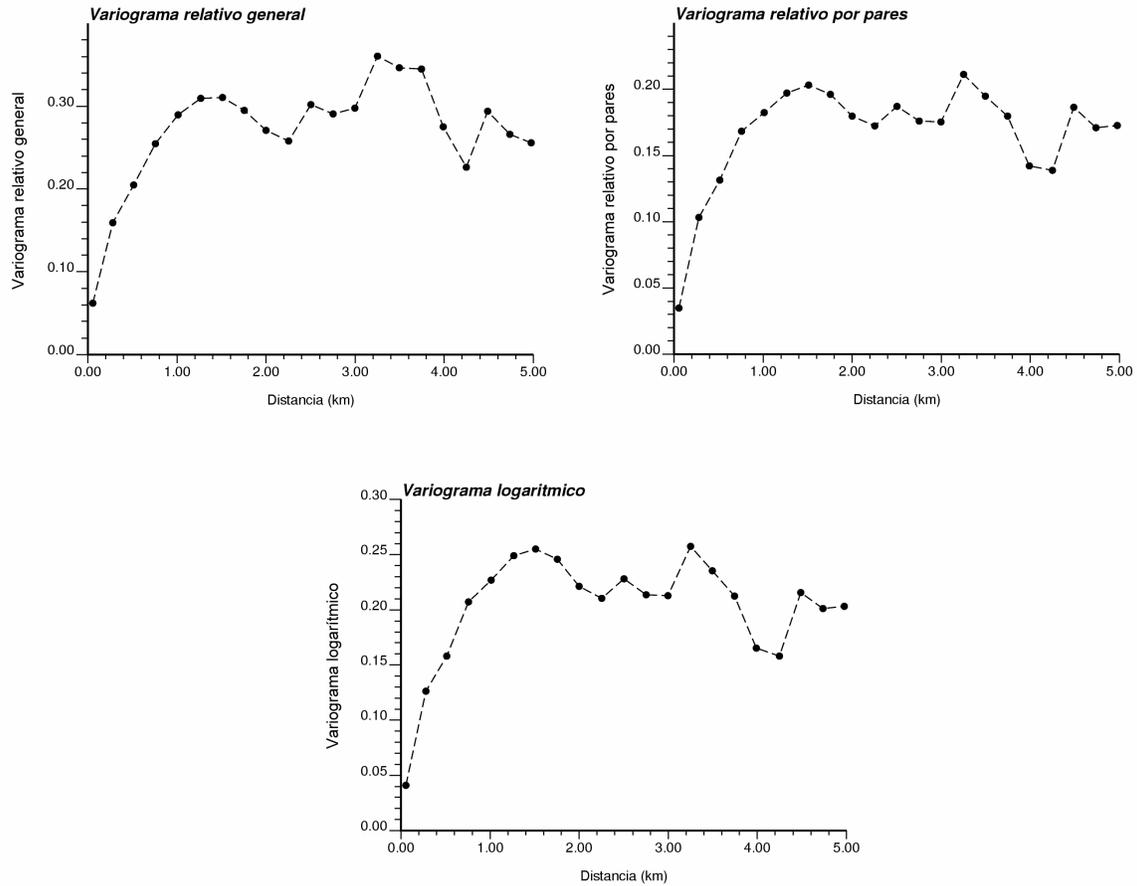


Figura 7. Variogramas relativos y variograma logarítmico para los datos de concentración de cobalto. Nótese que los valores en los cuales se estabilizan estos variogramas difieren de aquellos del variograma tradicional (Figura 2)

Capítulo 5. Análisis variográfico:

2. Variograma modelado

El variograma experimental no puede utilizarse directamente. Por una parte, sólo está definido para ciertas distancias y direcciones, por lo que es incompleto. Por otra parte, está sujeto a ciertas aproximaciones, debido al número limitado de datos y a los parámetros de tolerancia utilizado en el cálculo. Para subsanar esta situación, la idea es ajustar un modelo teórico de variograma en torno al variograma experimental. Esta etapa es la fase esencial de todo estudio geoestadístico, pues aquí es donde uno “interpreta” la continuidad espacial de la variable en estudio.

1. Propiedades de un variograma teórico

1.1. Propiedades matemáticas

Para que una función sea un variograma, debe cumplir con varias propiedades:

- Paridad: $\gamma(\mathbf{h}) = \gamma(-\mathbf{h})$.
- Nulidad en el origen: $\gamma(\mathbf{0}) = 0$.
- Positividad: $\gamma(\mathbf{h}) \geq 0$.
- Comportamiento al infinito: $\lim_{|\mathbf{h}| \rightarrow +\infty} \gamma(\mathbf{h})/|\mathbf{h}|^2 = 0$.
- Función de *tipo negativo condicional*:

$$\forall k \in \mathbb{N}^*, \forall \lambda_1, \dots, \lambda_k \in \mathbb{R} \text{ tales que } \sum_{i=1}^k \lambda_i = 0, \forall \mathbf{x}_1, \dots, \mathbf{x}_k \in \mathbb{D}, \sum_{i=1}^k \sum_{j=1}^k \lambda_i \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j) \leq 0.$$

Esta propiedad es una condición necesaria y suficiente para que γ sea el variograma de una función aleatoria. El adjetivo *condicional* se refiere al hecho de que la desigualdad es válida sólo para una clase restringida de ponderadores $(\lambda_1, \dots, \lambda_k)$, a saber, los de suma total nula.

1.2. Comportamiento en el origen

Mientras más regular el variograma en el origen (distancias cercanas a 0), más regular es la variable regionalizada en el espacio. Se suele distinguir tres tipos de comportamiento para el variograma en el origen (Figura 1):

- *parabólico*: corresponde a una variable regionalizada muy regular en el espacio
- *lineal*: corresponde a una variable regionalizada continua, pero no tan regular
- *discontinuo*: corresponde a una variable regionalizada más errática, es decir, con discontinuidades en la distribución espacial de sus valores. La desemejanza entre dos datos muy cercanos no es despreciable: los valores medidos varían a una escala muy pequeña y su continuidad no es perceptible. Este fenómeno se llama **efecto pepita**, por referencia a las leyes de oro en los yacimientos auríferos, que cambian repentinamente cuando hay pepitas de oro.

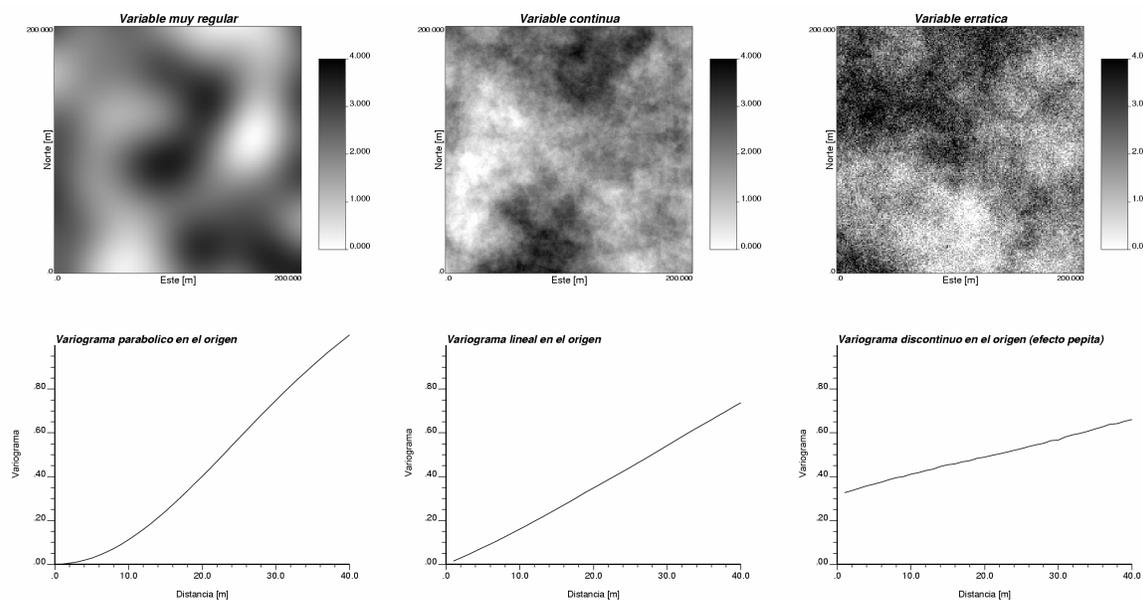


Figura 1. Relación entre la regularidad espacial de una variable regionalizada (arriba) y el comportamiento en el origen del variograma (abajo)

1.3. Comportamiento para distancias muy grandes

Frecuentemente, el variograma crece a partir del origen y se estabiliza, a partir de una distancia a , en torno a una **meseta**. En este caso, se demuestra que esta meseta es igual a la varianza a priori:

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h})$$

$\downarrow_{|\mathbf{h}| \rightarrow \infty}$ $\downarrow_{|\mathbf{h}| \rightarrow \infty}$

$$\gamma(\infty) = C(\mathbf{0}) = \sigma^2: \text{varianza } a \text{ priori.}$$

Las dos variables aleatorias $Z(\mathbf{x})$ y $Z(\mathbf{x} + \mathbf{h})$ estarán correlacionadas si la longitud del vector de separación \mathbf{h} es inferior a la distancia a , que se denomina el **alcance**. El alcance corresponde a la noción intuitiva de *zona de influencia* : más allá de $|\mathbf{h}| = a$, el variograma es constante e igual a su meseta y las variables $Z(\mathbf{x})$ y $Z(\mathbf{x} + \mathbf{h})$ son independientes (en rigor, no están correlacionadas). Los variogramas con meseta se denominan también *modelos de transición*.

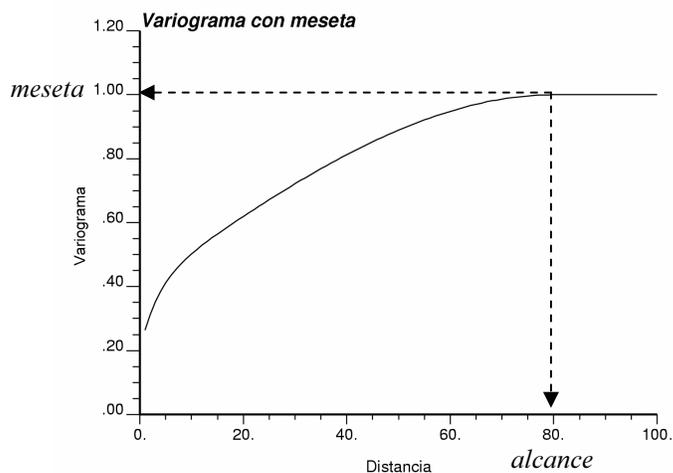


Figura 2. Variograma con meseta y alcance

A veces, el variograma crece infinitamente cuando la distancia aumenta y no presenta meseta ni alcance. En este caso, la varianza *a priori* es infinita y no existe la función de covarianza ni el correlograma.

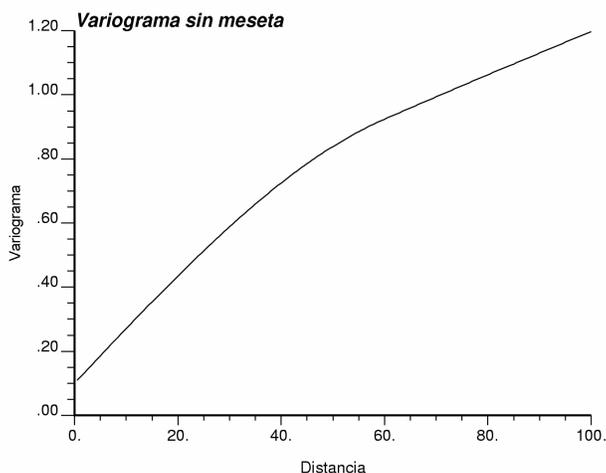


Figura 3. Variograma sin meseta

La ausencia de meseta puede ser también un efecto de la escala de observación, cuando el variograma sí alcanza una meseta, pero para una distancia mayor a la máxima distancia de cálculo.

1.4. Comportamiento direccional

Un variograma $\gamma(\mathbf{h})$ es isótropo si es idéntico en todas las direcciones del espacio, es decir, si no depende de la orientación del vector \mathbf{h} , sino sólo de su módulo $|\mathbf{h}|$. En caso contrario, hay anisotropía; tal propiedad caracteriza un fenómeno que se extiende de preferencia en ciertas direcciones.

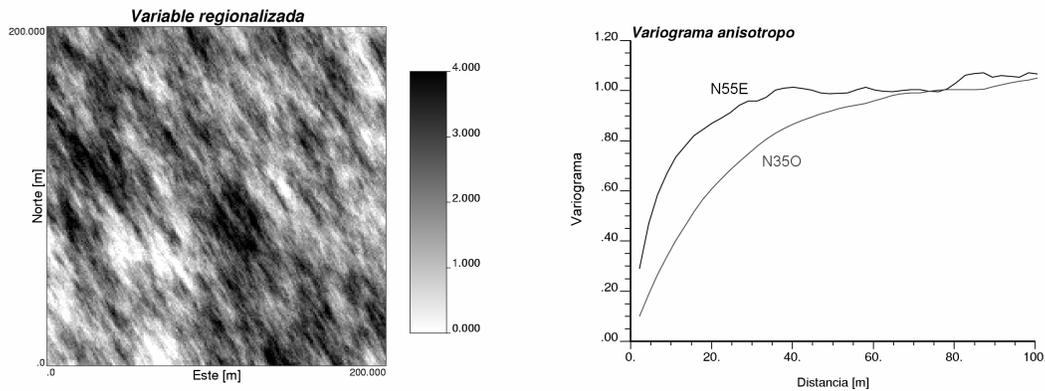


Figura 4. Variable regionalizada con dirección preferencial de continuidad (izquierda) y variograma en las direcciones de mayor y menor continuidad (derecha)

1.5. Otras propiedades

Periodicidad: suele ocurrir con fenómenos temporales, menos con fenómenos espaciales.

Efecto de hoyo: el variograma no es monótono.

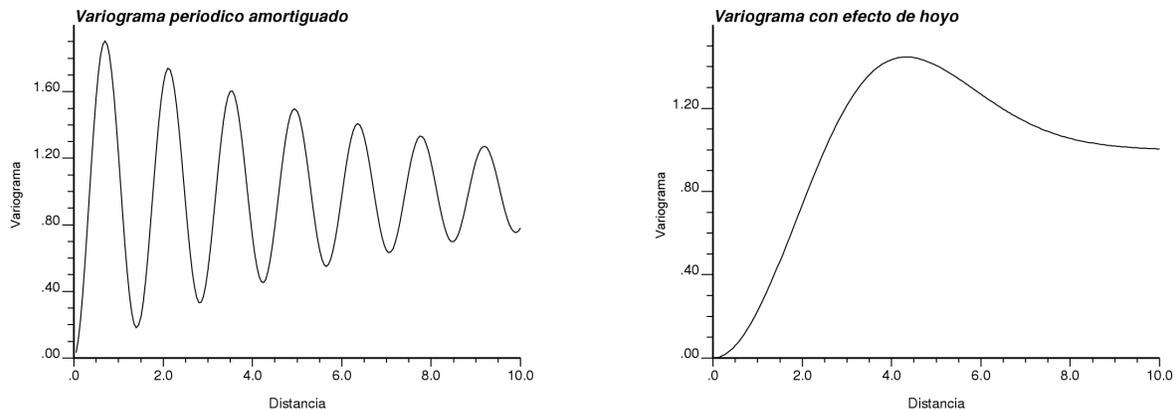


Figura 5. Variograma con periodicidad (izquierda) y efecto de hoyo (derecha)

El variograma sólo provee una caracterización parcial de la función aleatoria. Varias características de la distribución espacial no están descritas por esta herramienta, como la **conectividad espacial de valores**. Esta característica sólo se puede investigar al examinar distribuciones multivariadas (que involucren a más de dos sitios a la vez).

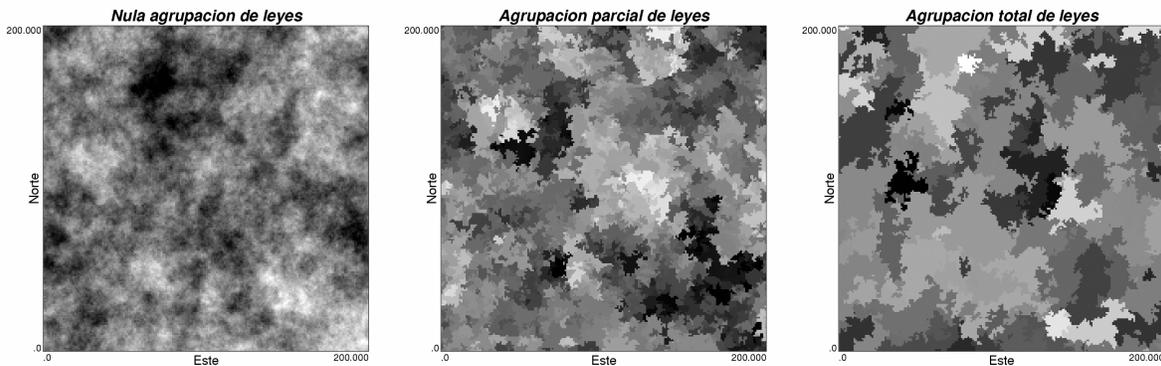


Figura 6. Ejemplos de realizaciones con diferentes características de conectividad espacial. A pesar de esto, todas tienen el mismo variograma

2. Modelos elementales de variograma

Una función es un variograma si y sólo si es de tipo negativo condicional. Es una propiedad muy restrictiva y difícil de controlar, incluso si se trabaja en un espacio de una dimensión. Por esto, en general, se elige un modelo de variograma entre las funciones de las cuales sabemos que son de tipo negativo condicional. A continuación, daremos algunos ejemplos de funciones que tienen esta propiedad y que constituyen los modelos más usados en la práctica.

2.1. Efecto pepita

El variograma pepítico de meseta C se define como:

$$\gamma(\mathbf{h}) = \begin{cases} 0 & \text{si } \mathbf{h} = \mathbf{0} \\ C & \text{en caso contrario} \end{cases}$$

Este modelo de variograma alcanza inmediatamente su meseta: hay ausencia total de correlación espacial, de modo que la geoestadística encuentra todos los resultados de la estadística clásica. Es un caso poco frecuente en las aplicaciones, a menos que los errores de medición sean muy grandes o que las distancias entre datos sean mayores que el alcance real, en cuyo caso la continuidad espacial es imperceptible.

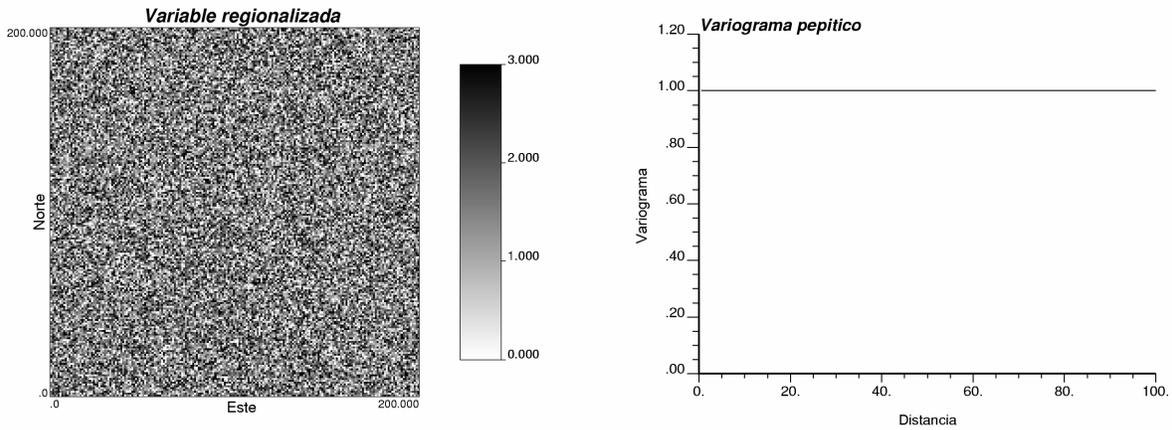


Figura 7. Variograma pepítico (derecha) y ejemplo de variable regionalizada asociada (izquierda)

2.2. Modelo esférico

El variograma esférico de alcance a y meseta C se define como:

$$\gamma(\mathbf{h}) = \begin{cases} C \left\{ \frac{3}{2} \frac{|\mathbf{h}|}{a} - \frac{1}{2} \left(\frac{|\mathbf{h}|}{a} \right)^3 \right\} & \text{si } |\mathbf{h}| \leq a \\ C & \text{en caso contrario} \end{cases}$$

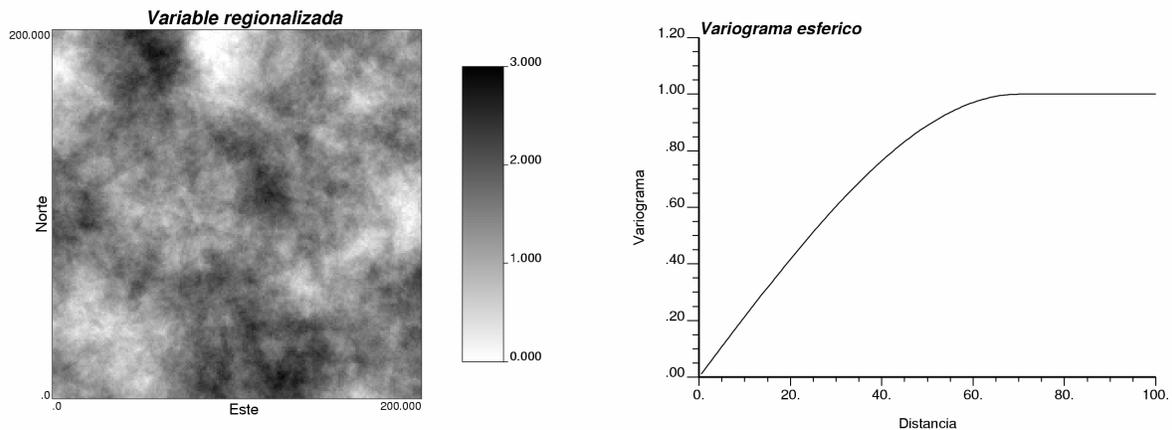


Figura 8. Variograma esférico (derecha) y ejemplo de variable regionalizada asociada (izquierda)

El efecto pepita podría verse como un modelo esférico de alcance infinitamente corto. Sin embargo, desde el punto de vista físico, existe una diferencia fundamental entre los modelos pepítico y esférico: el primero representa una variable regionalizada discontinua a

la escala de observación, para la cual los valores cambian repentinamente de un punto a otro, mientras que el segundo describe una variable regionalizada continua.

2.3. Modelo exponencial

El variograma exponencial de parámetro a y meseta C se define como:

$$\gamma(\mathbf{h}) = C \left\{ 1 - \exp\left(-\frac{|\mathbf{h}|}{a}\right) \right\}$$

Contrariamente al modelo esférico que llega a la meseta exacta para $|\mathbf{h}| = a$, el modelo exponencial sólo alcanza su meseta asintóticamente. En todo caso, se puede considerar un alcance práctico igual a $3a$ para el cual el variograma llega al 95% del valor de su meseta

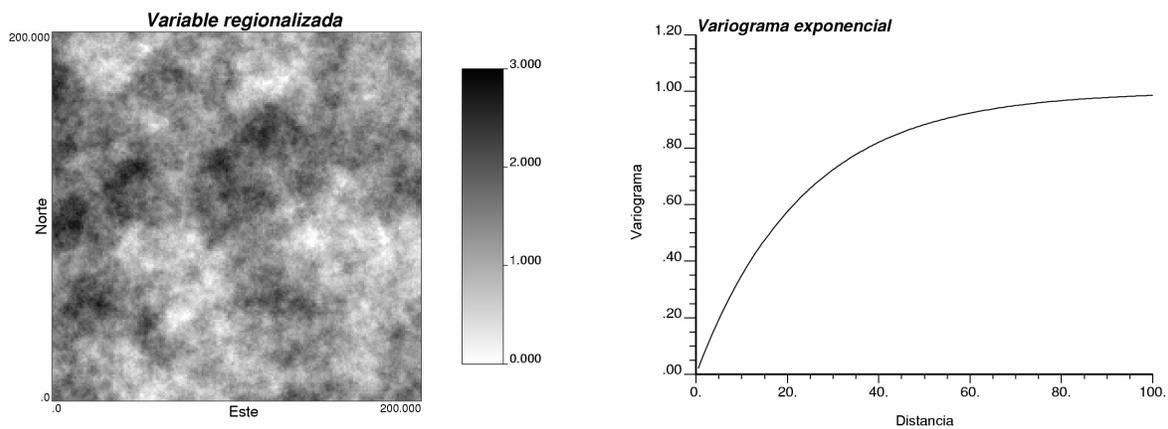


Figura 9. Variograma exponencial (derecha) y ejemplo de variable regionalizada asociada (izquierda)

2.4. Modelo Gaussiano

El variograma Gaussiano de parámetro a y meseta C se define como:

$$\gamma(\mathbf{h}) = C \left\{ 1 - \exp\left(-\frac{|\mathbf{h}|^2}{a^2}\right) \right\}$$

La meseta se alcanza asintóticamente y el alcance práctico puede considerarse igual a $a\sqrt{3}$.

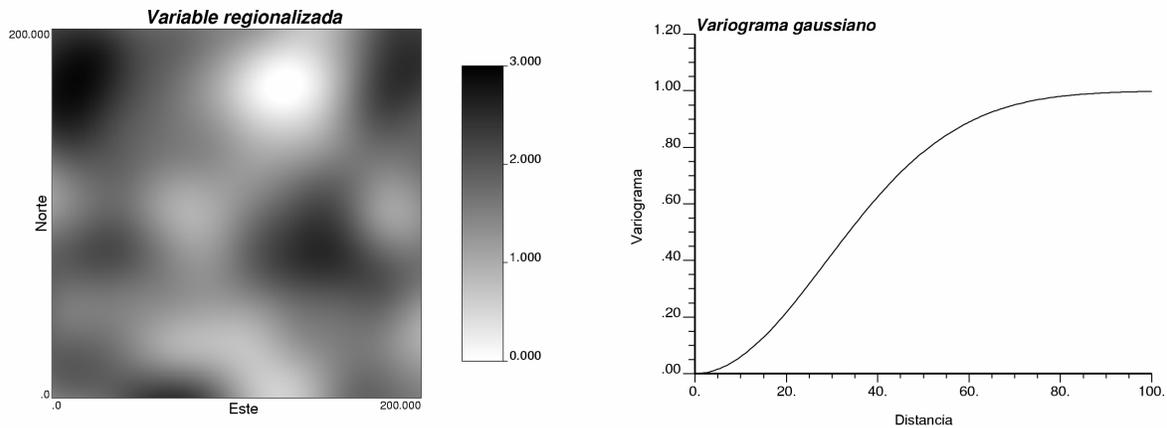


Figura 10. Variograma Gaussiano (derecha) y ejemplo de variable regionalizada asociada (izquierda)

2.5. Modelo seno cardinal

El variograma seno cardinal de parámetro a y meseta C se define como:

$$\gamma(\mathbf{h}) = C \left\{ 1 - \frac{a}{|\mathbf{h}|} \operatorname{sen} \left(\frac{|\mathbf{h}|}{a} \right) \right\}$$

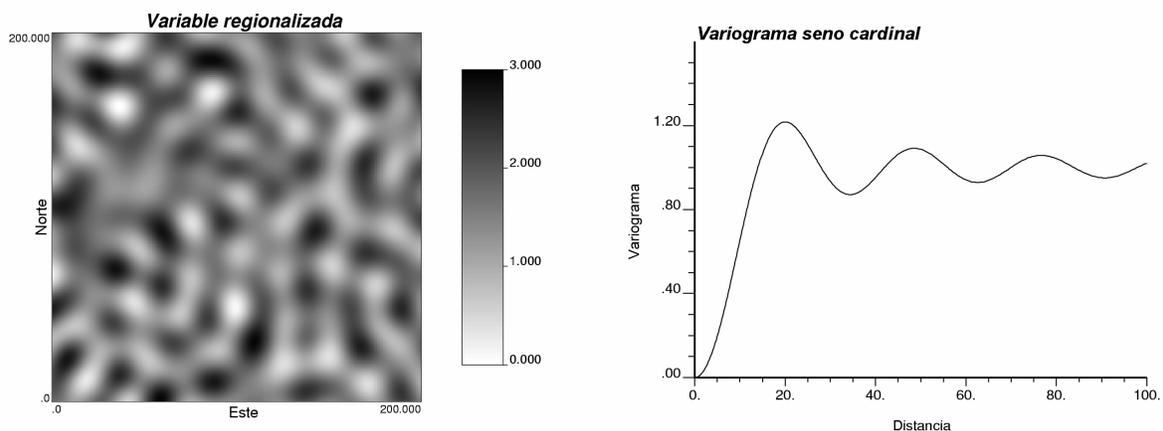


Figura 11. Variograma seno cardinal (derecha) y ejemplo de variable regionalizada asociada (izquierda)

Este modelo presenta oscilaciones, las cuales tienen generalmente una interpretación física (fenómeno periódico amortiguado, por ejemplo). El alcance práctico vale $20.37 a$ y la mitad del seudo-período es igual a $4.49 a$, distancia para la cual el variograma vale $1.21 C$.

2.6. Modelo potencia

El variograma potencia de pendiente ω y exponente θ se define como:

$$\gamma(\mathbf{h}) = \omega |\mathbf{h}|^\theta$$

Este variograma no posee ni meseta ni alcance, sino que crece en forma indefinida. El exponente θ puede variar entre 0 (variograma péptico) y 2 (variograma parabólico). El modelo se llama *lineal* cuando el exponente vale 1. A modo de ejemplo, el movimiento browniano fraccionario de parámetro $\eta \in]0,1[$ es un proceso aleatorio cuyo variograma es el modelo potencia $\gamma(\mathbf{h}) = |\mathbf{h}|^{2\eta}$.

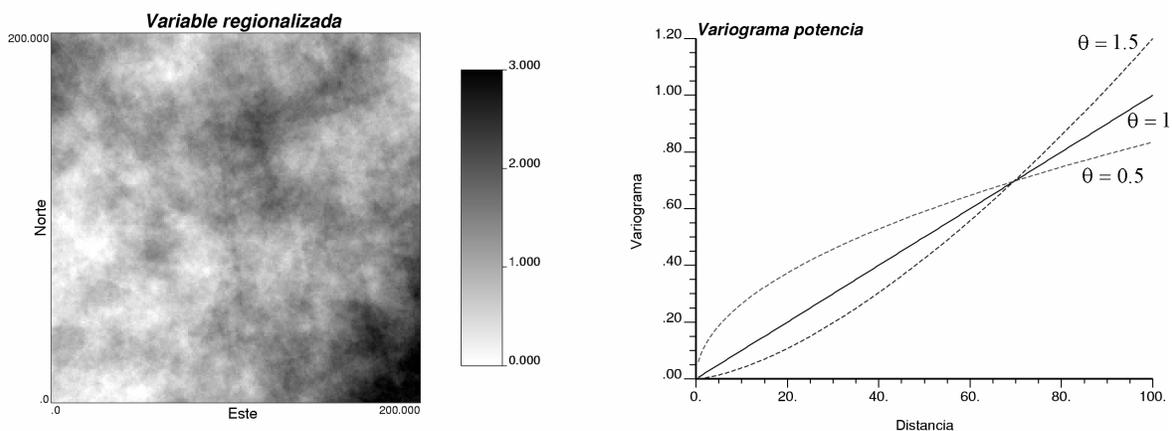


Figura 12. Variograma potencia (derecha) y ejemplo de variable regionalizada asociada a un variograma lineal (izquierda)

3. Modelamiento de un variograma experimental

3.1. Modelos anidados

El variograma experimental mide la desemejanza promedio entre dos datos en función de su separación. A menudo, presenta cambios de pendiente, que indican un cambio en la continuidad espacial a partir de ciertas distancias. Se pone así en evidencia distintas escalas de variación en la variable regionalizada. El variograma puede modelarse como la suma de varios modelos elementales denominados *modelos anidados* o *estructuras anidadas* (Figura 13):

$$\gamma(\mathbf{h}) = \gamma_1(\mathbf{h}) + \gamma_2(\mathbf{h}) + \dots + \gamma_S(\mathbf{h}) .$$

Cada escala de observación integra todas las estructuras de los niveles inferiores. Por ejemplo, si se trabaja a la escala métrica, el variograma incorporará la variabilidad presente

a escalas de milímetros, centímetros... pero la variabilidad kilométrica no se manifestará. Por lo tanto, hay una jerarquía entre las estructuras, que se encajan unas dentro de las otras, de donde viene el calificativo de *anidadas*.

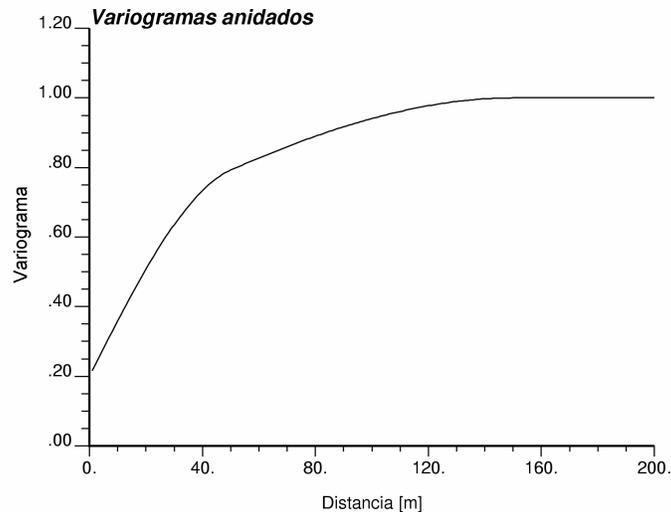


Figura 13. Variograma anidado obtenido por la suma de un efecto pepita y dos modelos esféricos

En la práctica, la descomposición anterior no es única, pues las escalas de variabilidad de una variable regionalizada son raramente tan diferenciadas. Además, en general, el uso de estructuras anidadas sólo representa una comodidad para el ajuste, sin que exista una interpretación de cada estructura. Por otro lado, es claro que una estructura de pequeña escala sólo podrá identificarse si la malla de muestreo es suficientemente fina. De la misma manera, una estructura de gran escala sólo aparecerá en el variograma si el diámetro del dominio muestreado es bastante grande.

3.2. Efecto pepita

El efecto pepita puede deberse a varios factores, entre los cuales destacan:

- **Presencia de una “micro-estructura”**, o sea, un modelo anidado de alcance pequeño con respecto a la escala de observación. Por ejemplo, consideremos un modelo anidado compuesto de dos estructuras: la primera de alcance a centimétrico y de meseta C , y la segunda de alcance kilométrico. En la vecindad del origen, el variograma presentará una zona de crecimiento que alcanzará rápidamente (después de algunos centímetros) la meseta C . A la escala kilométrica, este crecimiento se confundirá con una discontinuidad en el origen, es decir, con un efecto pepita, de meseta C (Figura 14).

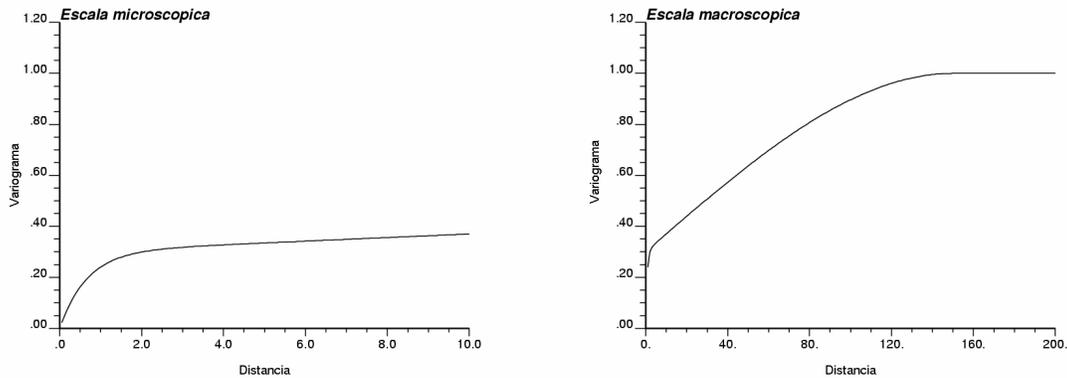


Figura 14. Génesis de un efecto pepita

Así, un efecto pepita puede ser la manifestación de una escala de variabilidad cuyas dimensiones características son muy inferiores a la escala de trabajo: sólo la meseta de esta “micro-estructura” es mensurable (está dada por la amplitud del efecto pepita), sus otros rasgos son imperceptibles.

- **Soporte de las mediciones.** Debido al efecto de soporte, se sabe que la variabilidad de una variable regionalizada aumenta cuando disminuye el soporte en el cual se mide esta variable. Así, un soporte de mediciones pequeño puede significar una importante componente péptica que oculta la continuidad subyacente de la función aleatoria. Un soporte más voluminoso conduciría a una disminución de la constante péptica y una mejor apreciación de la continuidad de los valores. La teoría establece que la amplitud (meseta) del efecto pepita es inversamente proporcional al volumen de la medición.
- **Errores de medición.** En el caso de que las mediciones estén afectadas por errores, el variograma suele presentar un efecto pepita. Si los errores son independientes entre sí, de misma varianza y se suman a los valores reales, entonces el variograma de los datos erróneos es igual al variograma de los datos reales (sin error) más un efecto pepita igual a la varianza del error.
- **Errores de ubicación de los datos.** Del mismo modo que los errores de medición, el tener errores en la posición de los datos introduce un efecto pepita en el variograma.

3.3. Anisotropía

Las anisotropías se manifiestan cuando el variograma difiere según las direcciones del espacio. En ausencia de anisotropía, $\gamma(\mathbf{h})$ sólo depende del módulo del vector de separación \mathbf{h} , no de su orientación. Como los modelos elementales de variograma están definidos para el caso isótropo, debemos examinar las transformaciones que permiten el paso a modelos anisótropos.

En la práctica, las anisotropías se pueden identificar al comparar los variogramas experimentales calculados a lo largo de varias direcciones del espacio, por ejemplo, en el

caso bidimensional, a lo largo de las direcciones orientadas 0° , 45° , 90° y 135° con respecto al eje de las abscisas. A menudo, se completa este examen dibujando el mapa variográfico. Cuando hay isotropía, los variogramas experimentales en las diferentes direcciones se superponen y el mapa variográfico dibuja círculos (caso bidimensional) o esferas (caso tridimensional).

En el caso contrario, se está en presencia de una anisotropía. Se distingue varios tipos de anisotropía, en especial, la anisotropía geométrica y la anisotropía zonal.

Anisotropía geométrica

Una anisotropía se dice *geométrica* cuando el mapa variográfico dibuja elipses concéntricas (caso 2D) o elipsoides (caso 3D). Los variogramas direccionales tienen la misma forma, pero alcances diferentes.

En el caso bidimensional, la anisotropía geométrica queda completamente definida por su *coeficiente de anisotropía* a/b , razón entre el eje menor a y el eje mayor b de cualquiera de las elipses de isovalores, y el *ángulo de anisotropía* θ , formado por el eje mayor y el eje de las abscisas. El primero indica la “intensidad” de la anisotropía (más marcada cuando se aleja de 1), mientras que el segundo indica su orientación. En el caso tridimensional, se debe definir dos coeficientes de anisotropía y dos ángulos.

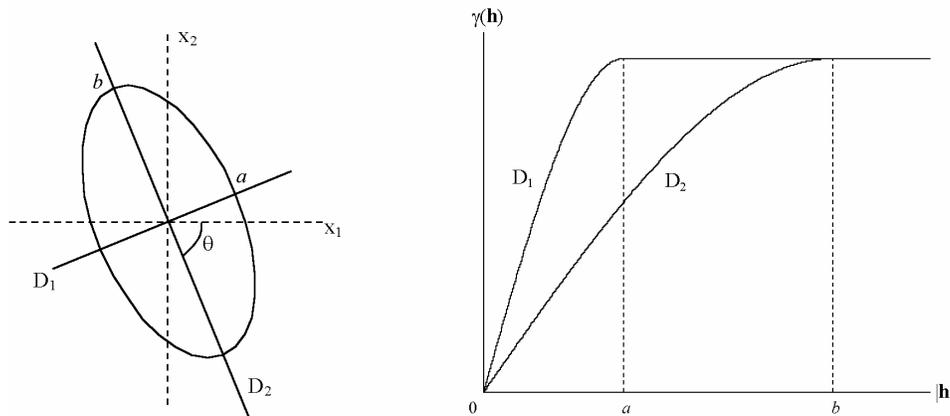


Figura 15. Ejemplo (2D) de anisotropía geométrica de razón a/b y ángulo θ . Izquierda: elipse de alcances. Derecha: variograma en las direcciones principales de anisotropía

Se puede escribir:

$$\gamma(\mathbf{h}) = \gamma_0(|\mathbf{A}\mathbf{h}|)$$

donde γ_0 es un variograma isótropo y \mathbf{A} una matriz (2×2) ó (3×3) que depende de los ángulo(s) y coeficiente(s) de anisotropía. Es decir, una transformación lineal de las coordenadas espaciales (rotación seguida de una homotecia) basta para volver a la situación de isotropía.

Anisotropía zonal

Existen variogramas cuya anisotropía no puede corregirse con una transformación lineal de coordenadas, especialmente aquellos que presentan una meseta variable según la dirección del espacio.

En un sentido estricto, la anisotropía *zonal* es un modelo en el cual la función aleatoria no depende de una (o varias) coordenada(s). Esto sucede a veces cuando se trabaja en el espacio de tres dimensiones. Por ejemplo, en el estudio de un fenómeno sedimentario, la variabilidad suele ser más importante a lo largo de la dirección vertical que atraviesa los estratos, que en los estratos mismos. Como una primera aproximación, la función aleatoria es constante en el plano horizontal, de manera que su variograma $\gamma(\mathbf{h})$ sólo depende de la componente vertical de \mathbf{h} . Denotando como θ el ángulo entre \mathbf{h} y la dirección vertical y \mathbf{u} un vector unitario orientado verticalmente, se puede escribir:

$$\gamma(\mathbf{h}) = \gamma(|\mathbf{h}| \cos \theta \mathbf{u}).$$

Este variograma tiene la misma meseta en todas las direcciones del espacio excepto en el plano horizontal para el cual $\cos \theta = 0$ y $\gamma(\mathbf{h}) = 0$.

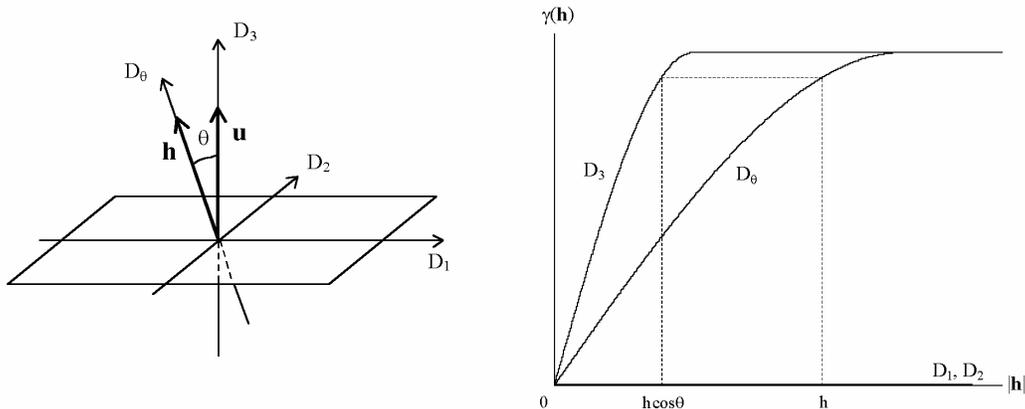


Figura 16. Ejemplo (3D) de anisotropía zonal en dirección vertical. Derecha: variogramas en las direcciones horizontales (D_1, D_2), vertical (D_3) y una dirección oblicua (D_θ).

Una forma equivalente de definir la anisotropía zonal consiste en verla como una anisotropía geométrica donde el eje mayor de la elipse (elipsoide) de alcances se vuelve infinitamente grande. En el mapa variográfico, la elipse se convierte entonces en una banda.

Anisotropías complejas

La suma de varios variogramas isotrópos o con anisotropías zonales o geométricas con características diferentes (ángulos y coeficientes de anisotropía) permite hacer frente a la mayor parte de las situaciones encontradas en la práctica.

3.4. Reglas de ajuste

Consideremos el siguiente ejemplo de variograma experimental, calculado a lo largo de las direcciones principales de anisotropía en el espacio de tres dimensiones.

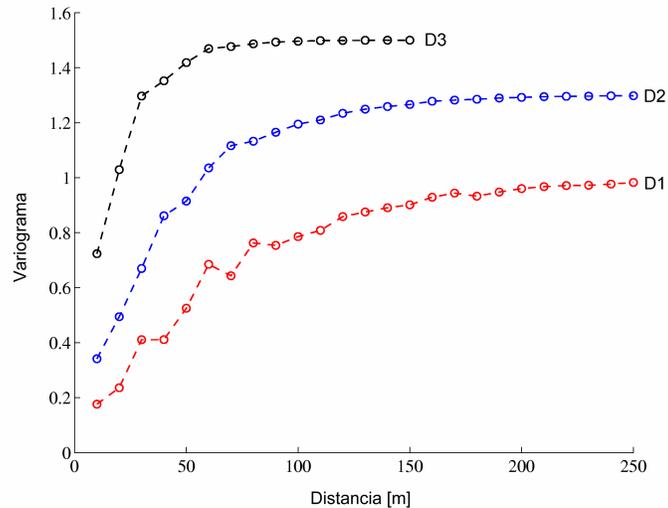


Figura 17. Variograma experimental a lo largo de tres direcciones principales de anisotropía (D1, D2, D3). ¿Cuál puede ser el modelo de variograma?

Antes de ajustar un modelo, empecemos con determinar (visualmente) las principales características del variograma experimental:

- el variograma tiene un efecto pepita de meseta 0.1 aproximadamente;
- en la primera dirección (D1), el variograma tiene un alcance cercano a 200m y una meseta cercana a 1.0;
- en la segunda dirección (D2), el variograma tiene un alcance cercano a 120m y una meseta cercana a 1.3;
- en la tercera dirección (D3), el variograma tiene un alcance cercano a 50m y una meseta cercana a 1.5;
- en cada dirección, aparte del efecto pepita, el comportamiento en el origen es lineal, lo que sugiere utilizar modelos anidados de tipo esférico o exponencial.

Dado que el modelo será la suma de modelos básicos (variograma anidado), se va a ir sumando contribuciones positivas a medida que se agregan los modelos anidados. Por esta razón, el modelamiento se realizará yendo “desde abajo hacia arriba” en el eje de ordenada correspondiente al variograma.

El primer “hito” en el eje de ordenada corresponde al efecto pepita, cuya amplitud es de 0.1. Por lo tanto, usaremos como primer modelo básico dicho efecto pepita (Figura 18A). El segundo hito corresponde a la meseta (1.0) en la primera dirección. Como ya tenemos un efecto pepita de meseta 0.1, sólo falta agregar un modelo básico (digamos, un exponencial) de meseta 0.9 y cuyos alcances en las tres direcciones D1, D2 y D3 son 200m, 120m y

50m, respectivamente (Figura 18B). El siguiente hito es la meseta (1.3) en la dirección D2. Estando en la meseta 1.0, falta agregar un modelo básico (exponencial) de meseta 0.3; sus alcances en las direcciones D2 y D3 son 120m y 50m respectivamente, mientras que el alcance en la dirección D1 se pone infinito de modo de no alterar más el variograma en esta dirección (esto equivale a considerar una anisotropía zonal a lo largo de la dirección D1) (Figura 18C). De manera análoga, para llegar al último hito (meseta 1.5 en la dirección D3), basta con agregar un último modelo básico de meseta 0.2 y alcances infinitos en las direcciones D1 y D2 y 50m en la dirección D3 (Figura 18D).

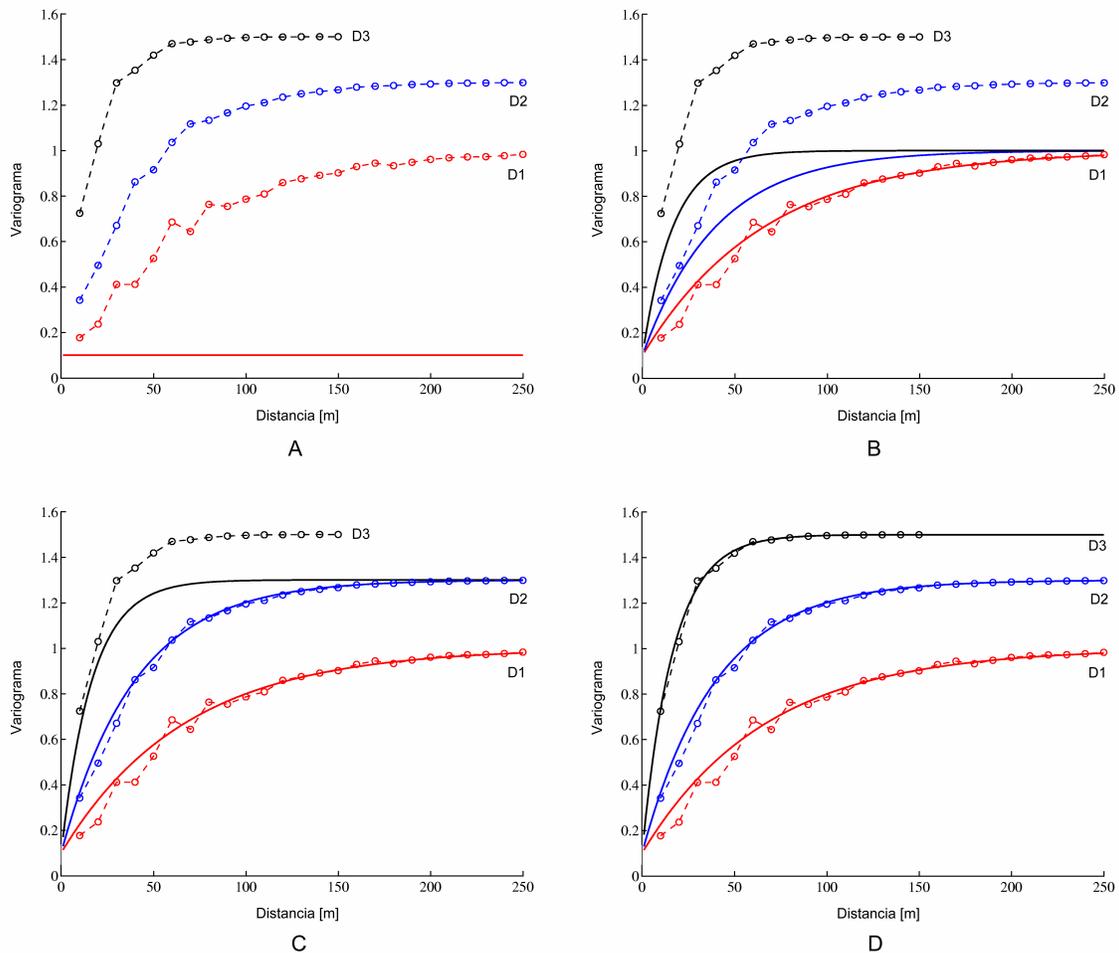


Figura 18. Pasos para el modelamiento del variograma experimental, sumando modelos anidados.

El modelo final queda como:

$$\gamma(\mathbf{h}) = 0.1 \text{ pepa} + 0.9 \exp(200\text{m}, 120\text{m}, 50\text{m}) + 0.3 \exp(\infty, 120\text{m}, 50\text{m}) + 0.2 \exp(\infty, \infty, 50\text{m}).$$

donde las distancias entre paréntesis indican los alcances prácticos en las direcciones D1, D2 y D3, respectivamente.

3.5. Aplicación a los datos de contaminación de suelo

Reanudamos con el variograma experimental de los datos de concentración de cobalto, calculado a lo largo de las direcciones reconocidas como de mayor y menor continuidad (a saber, N60°E y N30°W) (Figura 5, capítulo anterior). Aplicando el procedimiento descrito previamente, se propone el siguiente modelo compuesto de un efecto pepita y tres modelos esféricos:

$$\gamma(\mathbf{h}) = 2.5 \text{ pepa} + 5.5 \text{ esf}(0.85\text{km}, 0.85\text{km}) + 3.5 \text{ esf}(3\text{km}, 1.8\text{km}) + 4.0 \text{ esf}(\infty, 1.8\text{km}).$$

En esta ecuación, el primer alcance corresponde a la dirección N60°E, mientras que el segundo alcance se refiere a la dirección ortogonal N30°W.

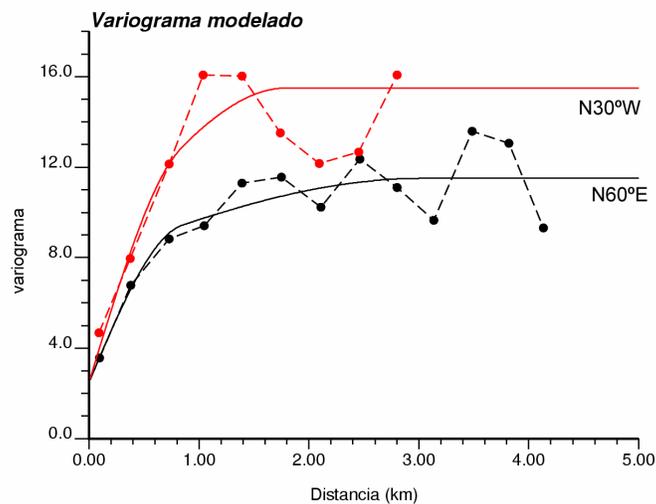


Figura 19. Modelamiento del variograma experimental de las concentraciones de cobalto.

3.6. Consideraciones prácticas

La representación de una variable regionalizada por un variograma es una operación subjetiva: no existe un modelo subyacente “verdadero”. En la práctica, conviene asegurarse que el modelo elegido respete las principales características del variograma experimental (efecto pepita, comportamiento en el origen, existencia o no de una meseta, anisotropía) y que no sea inútilmente complejo.

En general, se agregan varios modelos de base, que idealmente modelan componentes espaciales de escalas diferentes. Inicialmente, se selecciona la familia de modelos básicos que se quiere utilizar, luego se ajusta sus parámetros (meseta, alcance...). Se procurará evitar el uso exclusivo de métodos de ajuste automáticos. Por el contrario, el modelamiento de un variograma debe ser un trabajo interactivo, donde el usuario tiene la última palabra.

Se puede añadir las siguientes observaciones.

- La elección de un modelo con o sin **meseta** se decide en función del comportamiento del variograma experimental para distancias grandes. Sin embargo este comportamiento puede ser errático, por intervenir pocos pares de datos cuando la distancia se acerca a la dimensión máxima del campo. En general, se considera que el variograma experimental en una dirección es poco confiable para distancias superiores a la mitad del diámetro del campo en esta dirección. El ajuste de la meseta se hace en torno a las fluctuaciones del variograma y/o de la varianza empírica de los datos. Cabe señalar que la meseta del variograma puede diferir de la varianza del histograma (varianza empírica de los datos), pues representa el valor que tendría la varianza empírica cuando el dominio donde se distribuyen los datos se vuelve infinitamente grande.
- El **comportamiento del variograma en el origen** puede no ser accesible si la malla de muestreo es demasiado grande. Así, una micro-estructura puede estar oculta y aparecer como un efecto pepita si la distancia más pequeña para la cual se calcula el variograma es mayor que el alcance práctico. Frecuentemente, se adopta un modelo lineal en el origen (esférico o exponencial) con una constante pepítica que se obtiene extrapolando el variograma experimental hasta el eje de las ordenadas. Si el efecto pepita aparente es negativo, una solución consiste en utilizar un modelo con comportamiento parabólico en el origen. Pero esta situación debe cuestionarnos acerca de los parámetros de cálculo utilizados: el efecto pepita negativo (siempre artificial) puede verse exagerado, debido al uso impropio de la tolerancia; puede deberse también al soporte de los datos, cuando no es puntual.
- El ajuste a un modelo no se hace considerando solamente el variograma experimental, sino que se **debe tomar en cuenta toda la información disponible sobre la variable regionalizada**. Una anisotropía o una periodicidad deben modelarse si son obvias en el variograma experimental o si son explicables por consideraciones físicas. Asimismo, es esperable observar un efecto pepita cuando las mediciones están afectadas por errores.
- Finalmente, la mayoría de las aplicaciones (problemas de estimación) no requieren el conocimiento del variograma para todas las distancias. Conviene ajustar el variograma hasta la **escala de trabajo** considerada. Más allá, el ajuste no tiene importancia. Por ejemplo, esta escala representa la “vecindad” donde se tomará la información para estimar un valor en un sitio no muestreado (trabajo en una *vecindad móvil*, ver capítulo siguiente).

En conclusión el análisis variográfico hace uso tanto de la información disponible sobre la variable en estudio, como del tacto en la elección y el uso de las herramientas, así como también de un sentido de la aproximación. Se trata de la etapa esencial de todo estudio geoestadístico y seguramente de aquella que menos se presta al automatismo.

Capítulo 6. La estimación local

La estimación local busca *estimar* o *predecir* el valor de la variable regionalizada en un sector del campo estudiado, a partir de los datos disponibles (problema de “interpolación” espacial). En este capítulo, estudiaremos algunos métodos tradicionales de interpolación e introduciremos el método de *kriging* (bautizado así en honor de uno de los precursores de la geoestadística, Danie Krige).

1. Métodos tradicionales

El problema se formula de la siguiente manera: estimar el valor $z(\mathbf{x}_0)$ (desconocido) de la variable regionalizada estudiada en un sitio \mathbf{x}_0 a partir de los valores conocidos en sitios con datos. Para ello, se buscará asignar un ponderador a cada dato.

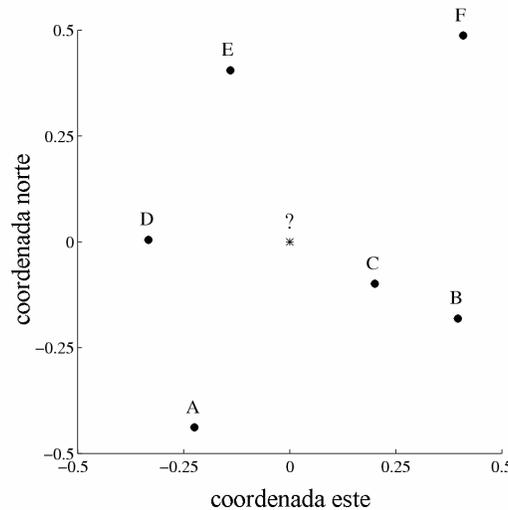


Figura 1. ¿Cómo estimar el valor en el sitio “?” a partir de los valores en los sitios con datos A, B, C, D, E, F?

1.1. Interpolación por el más cercano vecino

Este estimador atribuye toda la ponderación al dato más cercano al sitio a estimar. En el caso mostrado en la Figura 1, se trata del dato ubicado en C. El estimador apantalla a todos los datos salvo el más cercano, luego omite gran parte de la información y probablemente carece de precisión.

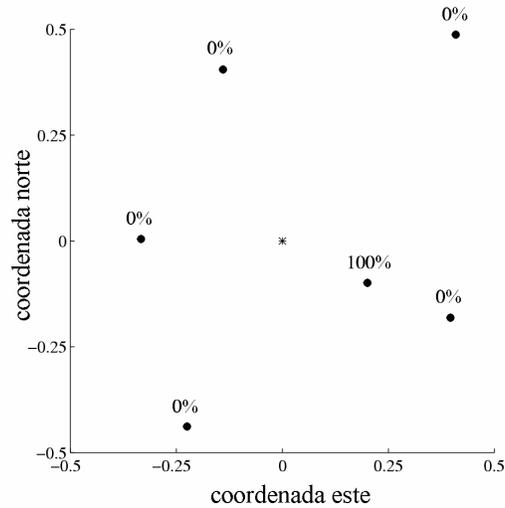


Figura 2. Ponderación obtenida con la interpolación por el más cercano vecino.

1.2. Inverso de la distancia

Este segundo estimador asigna a cada dato una ponderación inversamente proporcional a (una potencia de) su distancia al sitio a estimar.

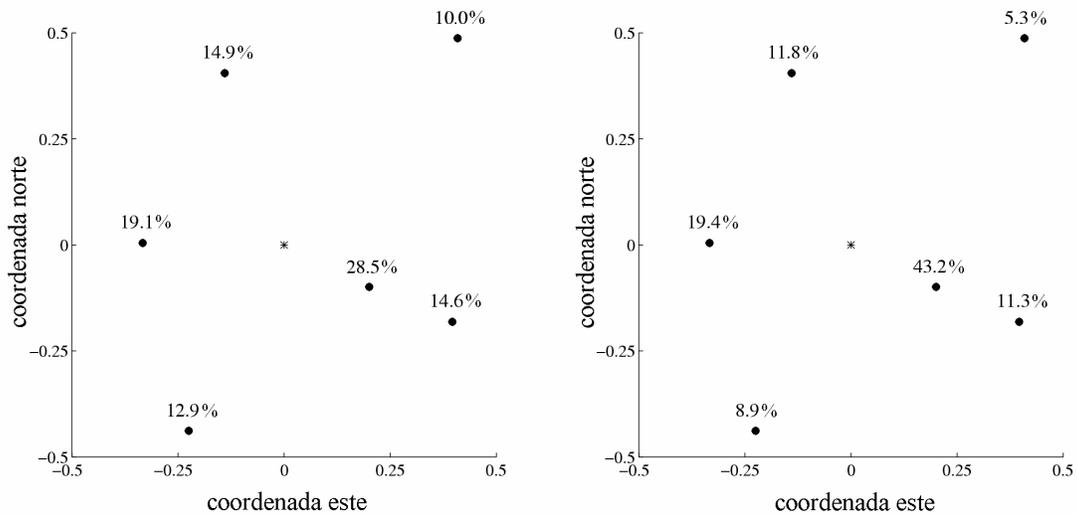


Figura 3. Ponderación obtenida con la interpolación por inverso de la distancia (izquierda) e inverso del cuadrado de la distancia (derecha).

Si la potencia es baja (cercana a cero), la distancia elevada a esta potencia es cercana a 1, por lo que el estimador asigna la misma ponderación a todos los datos (media aritmética de los datos). Al contrario, si la potencia es muy alta, el inverso de la distancia más pequeña se vuelve preponderante sobre todos los otros términos, y el estimador se convierte en el del más cercano vecino.

1.3. Otros métodos

Muchos otros métodos de interpolación han sido propuestos en la literatura, entre los cuales podemos citar:

- El **método de Sibson** (*interpolación por vecindad natural*), que busca una función de interpolación continua y derivable en el espacio. La ponderación se concentra en los sitios vecinos del sitio a estimar, que hacen de pantalla a los sitios más lejanos.
- Las interpolaciones basadas en una **triangulación** del espacio. Consisten en dividir el espacio en triángulos disjuntos cuyos vértices son los sitios con datos, luego en interpolar dentro de cada triángulo. Existen varias opciones para la triangulación, siendo la más conocida la triangulación de Delaunay (los vértices de cada triángulo son los sitios con datos cuyos polígonos de influencia tienen un lado en común). En cuanto a la interpolación dentro de cada triángulo, puede hacerse con funciones lineales (*interpolación lineal*) o polinomios (*método de Akima*). Tal como el método de Sibson, la interpolación por triangulación sólo considera los datos vecinos del sitio a estimar, que apantallan a los otros datos.
- Los **métodos baricéntricos**, que realizan la interpolación por una suma ponderada de los datos (tal como el inverso de la distancia). Como otro ejemplo, mencionemos el método de *interpolación bilineal*, que se utiliza cuando los datos están ubicados en una grilla regular.
- Los métodos de **superficies de tendencia**, que consisten en aplicar una regresión polinomial sobre los datos.
- Los **splines de interpolación** que recurren a una familia de funciones muy regulares en el espacio (splines cúbicos, splines laplacianos...), tal que se minimice un cierto criterio matemático bajo la restricción de interpolar los datos.
- Los **modelos numéricos de terreno**, para el caso específico de la interpolación de la altitud topográfica, que buscan incorporar la información que estructura el paisaje (líneas características como valles y curvas de nivel).

1.4. Propiedades

La pregunta que cabe plantearse es saber si, entre todos los métodos anteriores, existe uno que sea *mejor* que los otros. El término “mejor” podría ser tomado en el sentido de que la desviación cuadrática promedio entre los valores estimados y los valores verdaderos sea la más pequeña posible. En la práctica, ningún método es siempre mejor que otro, de modo que la elección de uno u otro método no es trivial.

Para saber a qué método recurrir, uno puede implementar la técnica de **jack-knife**. Consiste en dividir el conjunto de datos en dos sub-conjuntos. Luego, los datos del primer

sub-conjunto son utilizados para interpolar los valores del segundo sub-conjunto. El análisis de los errores cometidos permite entonces decidir cuál es el método de interpolación que tiene el mejor desempeño. Una variante del método de jack-knife consiste en considerar un segundo sub-conjunto reducido a un solo dato, siendo todos los otros datos parte del primer sub-conjunto, y repitiendo la interpolación para cada dato a la vez. Esta variante lleva el nombre de **validación cruzada**. Su puesta en marcha y los criterios que merecen atención para juzgar el desempeño de un método serán detallados más adelante.

Todos los métodos presentados con anterioridad toman en cuenta la información de naturaleza geométrica al momento de realizar la interpolación, principalmente las distancias que existen entre los sitios con datos y el sitio a estimar. Sin embargo, ignoran otra fuente de información, que es la continuidad espacial de la variable en estudio, resumida a través del modelo de variograma. El *kriging* que presentaremos a continuación busca mejorar la interpolación de los datos al tomar en cuenta:

- 1) sus distancias al sitio a estimar
- 2) las redundancias entre los datos debidas a posibles agrupamientos
- 3) la continuidad espacial de la variable regionalizada (variograma):
 - a. privilegiar los datos cercanos si el variograma es muy regular en el origen
 - b. repartir la ponderación entre los datos si existe un importante efecto pepita
 - c. en caso de anisotropía, privilegiar los datos ubicados en las direcciones de mayor continuidad con respecto al sitio a estimar.

Asimismo, el kriging permitirá cuantificar la precisión de la estimación mediante una varianza que mide la dispersión del error potencial cometido en la estimación.

2. Construcción del kriging

La resolución de un problema de estimación por kriging se articula siempre en torno a las mismas etapas. Las diferentes variantes sólo radican en las hipótesis realizadas sobre la función aleatoria $Z = \{Z(\mathbf{x}) \in \mathbf{D}\}$ que representa la variable regionalizada.

2.1. Restricción de linealidad

El estimador tiene que ser una combinación lineal ponderada (promedio ponderado) de los datos, que denotaremos de la siguiente forma:

$$Z^*(\mathbf{x}_0) = a + \sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{x}_{\alpha}),$$

donde \mathbf{x}_0 es el sitio donde se busca tener una estimación, $\{\mathbf{x}_{\alpha}, \alpha = 1 \dots n\}$ son los sitios con datos, mientras que los ponderadores $\{\lambda_{\alpha}, \alpha = 1 \dots n\}$ y el coeficiente a son las incógnitas

del problema de kriging. En rigor, se debería escribir $\lambda_\alpha(\mathbf{x}_0)$ en lugar de λ_α , puesto que los ponderadores de kriging dependerán del sitio \mathbf{x}_0 a estimar.

Esta restricción se debe a la decisión de considerar sólo los primeros momentos de las distribuciones de probabilidad (esperanza y covarianza / variograma). La construcción de estimadores más sofisticados que no sean combinaciones lineales de los datos, requeriría la especificación de la distribución espacial de la función aleatoria más allá de su variograma. Esto se puede realizar con métodos de geoestadística no lineal (kriging de indicadores, kriging disyuntivo, kriging lognormal, kriging multi-Gaussiano).

Nota: el estimador $Z^*(\mathbf{x}_0)$ así definido es una combinación de variables aleatorias y, por lo tanto, es una cantidad aleatoria. Para obtener una estimación numérica, basta con aplicar la fórmula anterior a la realización particular que constituyen los datos experimentales:

$$z^*(\mathbf{x}_0) = a + \sum_{\alpha=1}^n \lambda_\alpha z(\mathbf{x}_\alpha).$$

2.2. Restricción de insesgo

Esta etapa consiste en expresar que el error de estimación tiene esperanza nula, es decir

$$E[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] = 0.$$

Se puede interpretar esta restricción, reemplazando la esperanza matemática por una media en el espacio: *si se calcula sobre numerosas configuraciones de kriging idénticas, la media de los errores de estimación cometidos se acerca a cero*. La ausencia de sesgo no garantiza que los errores sean bajos, sino sólo que su media global es aproximadamente nula.

2.3. Restricción de optimalidad

Al superar las etapas anteriores, el estimador está sometido a una o varias restricciones pero no está totalmente especificado. La última etapa consiste en buscar los ponderadores que minimizan la varianza del error de estimación:

$$\text{var}[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] \text{ es mínima.}$$

En términos intuitivos, esta restricción significa que *si se calcula sobre numerosas configuraciones de kriging idénticas, la varianza estadística de los errores de estimación cometidos es la más baja posible*. Este criterio de precisión equivale a la minimización del error cuadrático promedio.

3. Plan de kriging

En general, los problemas de estimación local no involucran la totalidad del campo y tampoco utilizan todos los datos disponibles. Se define la **vecindad de kriging**, como el dominio del espacio que contiene el sitio a estimar y los datos utilizados en la estimación. El usuario puede considerar varias posibilidades.

3.1. Vecindad única

Se habla de vecindad única cuando se efectúa el kriging conservando todos los datos. En este caso, incluso los datos muy alejados intervendrán en la estimación. Sin embargo, salvo excepciones, su influencia será muy baja (intuitivamente, un sitio alejado no aporta demasiada información al sitio a estimar y se verá afectado por un ponderador de kriging bajo).

Cuando los datos son muy numerosos, es inútil conservarlos todos para una estimación local, puesto que se corre el riesgo de aumentar considerablemente los tiempos de cálculo. Por lo tanto, es necesario reducir el tamaño de la vecindad de kriging.

3.2. Vecindad móvil

El kriging se realiza en una vecindad móvil cuando sólo utiliza los datos “cercaños” al sitio a estimar. Ahora, en general, uno no se limita a una sola estimación local, sino que busca estimaciones en los nodos de una grilla regular que cubre la zona estudiada. Falta definir el tamaño y la forma de la vecindad, que se centra en el sitio a estimar y que se desplaza a través del campo, a medida que se realiza las estimaciones (de donde viene el adjetivo *móvil*).

Tamaño de la vecindad

El tamaño de la vecindad debe permitir un equilibrio entre varios factores:

- precisión de las estimaciones: aumenta cuando la vecindad es más grande;
- tiempos de cálculo, confiabilidad del modelo de variograma para grandes distancias, cambios en la continuidad espacial de la variable regionalizada: debido a todos estos factores, se tiende a elegir una vecindad de tamaño limitado.

Un criterio de decisión son las llamadas técnicas de *validación cruzada* o *jackknife* que veremos más adelante: se prueba varios tamaños de vecindad y se elige aquel que entrega los resultados más satisfactorios.

Cabe notar que no hay justificación particular para limitar el tamaño de la vecindad al alcance del modelo variográfico, bajo el pretexto que los datos localizados más allá de este

alcance no tienen correlación con el sitio a estimar. De hecho, en la mayoría de los casos, estos datos intervienen indirectamente en la estimación del valor promedio y mejoran la precisión, a veces de manera no despreciable. El factor a considerar en la elección del tamaño de la vecindad es la cantidad de datos disponibles en la vecindad más que el alcance del variograma.

Forma de la vecindad

En la medida de lo posible, la forma de la vecindad debe tomar en cuenta la anisotropía de la variable regionalizada, revelada por el análisis variográfico. Así, en el caso de una anisotropía geométrica, se considerará una vecindad en forma de elipse (o elipsoide) cuyas características – orientación y excentricidad – sean idénticas a las de la elipse (elipsoide) de anisotropía. A menudo, también, se divide esta elipse en varios sectores (en general, en cuadrantes u octantes), en cada uno de los cuales se trata de buscar un número fijo de datos, con el fin de repartir de mejor manera en torno al sitio que se quiere estimar, la información que se va a conservar. La Figura 4 presenta un ejemplo de vecindad móvil en el espacio de dos dimensiones, en forma de elipse centrada en el sitio a estimar. Los datos retenidos, tres por cuadrante al máximo, están indicados.

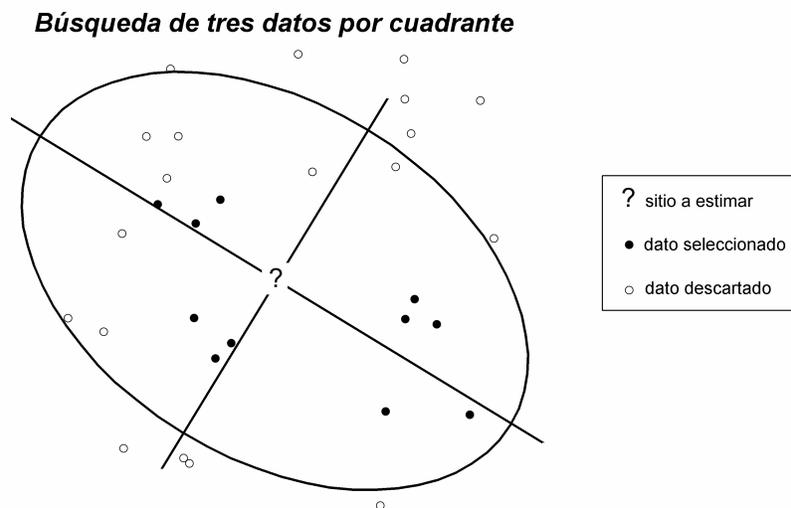


Figura 4. vecindad elíptica, dividida en cuadrantes
“●” = datos retenidos ; “○” = datos descartados

En caso de anisotropía más compleja que la anisotropía geométrica, se suele conservar una vecindad en forma de elipse o elipsoide, aunque idealmente se debería escoger una forma más sofisticada (por ejemplo, una vecindad en forma de banda en caso de anisotropía zonal pura). Hay que buscar entonces una elipse que se acerca lo mejor posible a las curvas de isovalores del modelo variográfico, que indican el nivel de correlación en función de la distancia geográfica.

4. Kriging de media conocida (kriging simple)

4.1. Hipótesis

Se supone que la variable regionalizada z es la realización de una función aleatoria Z estacionaria tal que

$$\begin{cases} \forall \mathbf{x} \in V, E[Z(\mathbf{x})] = m \text{ conocida} \\ \forall \mathbf{x}, \mathbf{x} + \mathbf{h} \in V, \text{cov}[Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})] = C(\mathbf{h}) \end{cases}$$

donde V representa la vecindad de kriging (para un kriging en una vecindad única, V es el campo completo de la variable regionalizada).

4.2. Determinación del estimador

Examinemos una a una las distintas etapas del kriging.

- **Linealidad:** se asegura esta restricción al tomar como estimador en \mathbf{x}_0

$$Z^*(\mathbf{x}_0) = a + \sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{x}_{\alpha})$$

- **Insesgo:** el valor esperado del error de estimación es

$$E[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] = a + \sum_{\alpha=1}^n \lambda_{\alpha} \underbrace{E[Z(\mathbf{x}_{\alpha})]}_{=m} - \underbrace{E[Z(\mathbf{x}_0)]}_{=m} = a + \left(\sum_{\alpha=1}^n \lambda_{\alpha} - 1 \right) m$$

Este valor esperado es nulo si:

$$a = \left(1 - \sum_{\alpha=1}^n \lambda_{\alpha} \right) m$$

- **Optimalidad:** debemos calcular la varianza del error de estimación:

$$\text{var}[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] = \text{var}\left\{ \sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{x}_{\alpha}) - Z(\mathbf{x}_0) \right\}$$

El término constante a no influye en la varianza, por lo cual se omite de la expresión. Ahora, se tiene la siguiente regla de cálculo:

$$\text{var}\{\lambda_1 Z_1\} = \lambda_1^2 \text{cov}\{Z_1, Z_1\}$$

$$\text{var}\{\lambda_1 Z_1 + \lambda_2 Z_2\} = \lambda_1^2 \text{cov}\{Z_1, Z_1\} + \lambda_2^2 \text{cov}\{Z_2, Z_2\} + 2\lambda_1 \lambda_2 \text{cov}\{Z_1, Z_2\}$$

Generalizando:

$$\text{var}\left\{\sum_{k=1}^K \lambda_k Z_k\right\} = \sum_{k=1}^K \sum_{k'=1}^K \lambda_k \lambda_{k'} \text{cov}\{Z_k, Z_{k'}\}$$

Aplicando esta fórmula, se obtiene:

$$\begin{aligned} \text{var}[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] &= \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} \text{cov}\{Z(\mathbf{x}_{\alpha}), Z(\mathbf{x}_{\beta})\} + \text{cov}\{Z(\mathbf{x}_0), Z(\mathbf{x}_0)\} - 2 \sum_{\alpha=1}^n \lambda_{\alpha} \text{cov}\{Z(\mathbf{x}_{\alpha}), Z(\mathbf{x}_0)\} \\ &= \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) + C(\mathbf{0}) - 2 \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{x}_{\alpha} - \mathbf{x}_0) \end{aligned}$$

donde $C(\cdot)$ es la covarianza de la función aleatoria Z . El mínimo de esta expresión se obtiene anulando sus derivadas parciales con respecto a las incógnitas $\{\lambda_{\alpha}, \alpha = 1 \dots n\}$. Se obtiene finalmente el sistema de ecuaciones:

$$\sum_{\beta=1}^n \lambda_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) = C(\mathbf{x}_{\alpha} - \mathbf{x}_0) \quad \forall \alpha = 1 \dots n.$$

Es un sistema lineal, en el cual el número de ecuaciones y de incógnitas es igual a la cantidad de datos utilizados. En escritura matricial, este sistema es:

$$\begin{pmatrix} C(\mathbf{x}_1 - \mathbf{x}_1) & \cdots & C(\mathbf{x}_1 - \mathbf{x}_n) \\ \vdots & & \vdots \\ C(\mathbf{x}_n - \mathbf{x}_1) & \cdots & C(\mathbf{x}_n - \mathbf{x}_n) \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = \begin{pmatrix} C(\mathbf{x}_1 - \mathbf{x}_0) \\ \vdots \\ C(\mathbf{x}_n - \mathbf{x}_0) \end{pmatrix},$$

lo que permite determinar los ponderadores de kriging $\{\lambda_{\alpha}, \alpha = 1 \dots n\}$.

Es interesante notar que, debido a la condición de insesgo, el estimador se pone bajo la forma:

$$Z^*(\mathbf{x}_0) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{x}_{\alpha}) + (1 - \sum_{\alpha=1}^n \lambda_{\alpha}) m,$$

de modo que el valor de la media aparece como si fuera un dato adicional, al cual se asigna una ponderación igual al complemento de la ponderación acumulada de los otros datos.

Mientras menos ponderación le damos a los datos (en la práctica, esto ocurre cuando uno se aleja de estos datos), más ponderación recibe la media. El rol de la media es de compensar la falta de información cuando los datos son escasos o alejados. Este kriging que supone conocida la media fue bautizado **kriging simple**.

4.3. Varianza de kriging

La varianza mínima del error de estimación en el sitio \mathbf{x}_0 , llamada *varianza de kriging*, se simplifica de la siguiente forma:

$$\sigma_{KS}^2(\mathbf{x}_0) = \sigma^2 - \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{x}_{\alpha} - \mathbf{x}_0),$$

donde $\sigma^2 = C(\mathbf{0})$ es la varianza *a priori* de la función aleatoria Z . Se puede mostrar que la varianza de kriging simple siempre es menor o igual a la varianza *a priori*:

$$\sigma_{KS}^2(\mathbf{x}_0) \leq \sigma^2.$$

5. Kriging de media desconocida (kriging ordinario)

5.1. Hipótesis

Se supone ahora que la variable regionalizada es la realización de una función aleatoria Z estacionaria tal que

$$\begin{cases} \forall \mathbf{x} \in V, E[Z(\mathbf{x})] = m \text{ desconocida} \\ \forall \mathbf{x}, \mathbf{x} + \mathbf{h} \in V, \text{cov}[Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})] = C(\mathbf{h}) \end{cases}$$

donde V representa la vecindad de kriging.

5.2. Determinación del estimador

Las etapas del kriging dan:

- **Linealidad:** se asegura esta restricción al tomar como estimador en \mathbf{x}_0

$$Z^*(\mathbf{x}_0) = a + \sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{x}_{\alpha})$$

- **Insesgo:** el valor esperado del error de estimación es

$$E[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] = a + \sum_{\alpha=1}^n \lambda_{\alpha} \underbrace{E[Z(\mathbf{x}_{\alpha})]}_{=m} - \underbrace{E[Z(\mathbf{x}_0)]}_{=m} = a + \left(\sum_{\alpha=1}^n \lambda_{\alpha} - 1 \right) m$$

Como se desconoce el valor de la media m , este valor esperado es nulo si:

$$a=0 \text{ y } \sum_{\alpha=1}^n \lambda_{\alpha} = 1.$$

La igualdad sobre la suma de los ponderadores asegura que, en el caso en que todos los datos son iguales a una misma constante, el valor estimado restituirá esta constante.

- **Optimalidad:** como en el caso del kriging simple, la varianza del error de estimación es:

$$\text{var}[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] = \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) + C(\mathbf{0}) - 2 \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{x}_{\alpha} - \mathbf{x}_0)$$

Se necesita minimizar esta expresión *bajo la condición de insesgo*, que impone que la suma de las incógnitas es igual a 1. Esto se logra introduciendo una incógnita adicional llamada multiplicador de Lagrange, que denotaremos como μ . Se escribe:

$$\begin{aligned} & \text{var}[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] \\ & = C(\mathbf{0}) + \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) - 2 \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{x}_{\alpha} - \mathbf{x}_0) + 2\mu \underbrace{\left(\sum_{\alpha=1}^n \lambda_{\alpha} - 1 \right)}_{=0} \end{aligned}$$

y se minimiza la función de las $n+1$ variables $\lambda_1, \dots, \lambda_n, \mu$. Calculando las $n+1$ derivadas parciales de esta función y luego anulándolas, se obtiene el sistema:

$$\begin{cases} \frac{\partial}{\partial \lambda_{\alpha}} = 0 : \sum_{\beta=1}^n \lambda_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) + \mu = C(\mathbf{x}_{\alpha} - \mathbf{x}_0) \quad \forall \alpha = 1 \dots n \\ \frac{\partial}{\partial \mu} = 0 : \sum_{\alpha=1}^n \lambda_{\alpha} = 1 \quad (\text{condición de insesgo}) \end{cases}$$

Este sistema contiene una incógnita y una ecuación más que el sistema de kriging simple. Se puede escribir en notación matricial:

$$\begin{pmatrix} C(\mathbf{x}_1 - \mathbf{x}_1) & \cdots & C(\mathbf{x}_1 - \mathbf{x}_n) & 1 \\ \vdots & & \vdots & \vdots \\ C(\mathbf{x}_n - \mathbf{x}_1) & \cdots & C(\mathbf{x}_n - \mathbf{x}_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{pmatrix} = \begin{pmatrix} C(\mathbf{x}_1 - \mathbf{x}_0) \\ \vdots \\ C(\mathbf{x}_n - \mathbf{x}_0) \\ 1 \end{pmatrix}$$

Este kriging se denomina “**kriging ordinario**”. Siendo el variograma una herramienta equivalente a la covarianza, a partir de la relación $\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h})$, se puede elegir utilizarlo en lugar de la función de covarianza. Las ecuaciones de kriging pasan a ser:

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) - \mu = \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_0) \quad \forall \alpha = 1 \dots n \\ \sum_{\alpha=1}^n \lambda_{\alpha} = 1 \end{cases}$$

Esto es:

$$\begin{pmatrix} \gamma(\mathbf{x}_1 - \mathbf{x}_1) & \cdots & \gamma(\mathbf{x}_1 - \mathbf{x}_n) & 1 \\ \vdots & & \vdots & \vdots \\ \gamma(\mathbf{x}_n - \mathbf{x}_1) & \cdots & \gamma(\mathbf{x}_n - \mathbf{x}_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ -\mu \end{pmatrix} = \begin{pmatrix} \gamma(\mathbf{x}_1 - \mathbf{x}_0) \\ \vdots \\ \gamma(\mathbf{x}_n - \mathbf{x}_0) \\ 1 \end{pmatrix}$$

5.3. Varianza de kriging

La varianza de kriging ordinario (varianza del error cometido en el sitio \mathbf{x}_0) se expresa de la siguiente forma:

$$\begin{aligned} \sigma_{KO}^2(\mathbf{x}_0) &= \sigma^2 - \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{x}_{\alpha} - \mathbf{x}_0) - \mu \\ &= \sum_{\alpha=1}^n \lambda_{\alpha} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_0) - \mu \end{aligned}$$

donde $\sigma^2 = C(\mathbf{0})$ es la varianza *a priori* de la función aleatoria Z , o sea, la meseta de su variograma. Ahora, la segunda igualdad muestra que la varianza de kriging no depende de este valor σ^2 , por lo cual el kriging ordinario sigue aplicable incluso cuando el variograma no presenta meseta (por ejemplo, cuando es un modelo potencia).

6. Otros tipos de kriging

6.1. Kriging con derivas

Este kriging se sustenta en la hipótesis que la función aleatoria ya no es estacionaria y su valor esperado (llamada “deriva” y denotada como $m(\mathbf{x})$) varía en el espacio, reflejando una tendencia sistemática en la distribución espacial de los valores. Existen tres variantes, que dependen de la forma de esta deriva:

- Kriging *universal*, que supone que la deriva es un polinomio de las coordenadas.
- Kriging *trigonométrico*, donde la deriva es combinación de funciones senos y cosenos.
- Kriging *con deriva externa*, que supone que la deriva es proporcional a una variable externa conocida en forma exhaustiva. Por ejemplo, para estimar la temperatura a nivel del suelo, se podría considerar la cota topográfica (modelo digital de elevación) como deriva externa. Para evaluar la profundidad del techo de un horizonte en un reservorio de petróleo, se puede utilizar el mapa obtenido por la interpretación de datos sísmicos (tiempos de reflexión). En un inventario forestal, la deriva externa podría provenir de la interpretación de fotografías aéreas o de imágenes satelitales.

La determinación del sistema de kriging sigue las mismas etapas que aquellas vistas en el kriging simple y kriging ordinario (linealidad, insesgo y optimalidad). Sin embargo, cabe señalar que la dificultad de estos métodos de kriging con deriva radica en la obtención del modelo variográfico: la presencia de la deriva hace que el variograma experimental de los datos está sesgado con respecto al variograma teórico:

$$E[\hat{\gamma}(\mathbf{h})] \geq \gamma(\mathbf{h}).$$

De alguna manera, la deriva oculta el variograma verdadero y dificulta la inferencia del variograma. El único caso en que la inferencia es posible es aquel en que la deriva es constante en una dirección del espacio: para todo vector \mathbf{h} orientado en esta dirección, se tiene una esperanza constante y la inferencia del variograma $\gamma(\mathbf{h})$ es posible. Por ejemplo, es el caso de la profundidad de los fondos marinos, donde puede considerarse que la deriva es invariante en la dirección paralela a la costa. Se puede entonces identificar el variograma en esta dirección, y luego, emprender un modelamiento suponiendo que este variograma es isótropo.

6.2. Kriging de bloques

El kriging de bloques consiste en estimar directamente el valor promedio de la variable sobre un soporte mayor que el soporte de los datos (bloque). Este kriging tiene interés en varios dominios de aplicación, como la evaluación de recursos minerales (estimar leyes de

mineral sobre unidades selectivas de explotación) o las ciencias medio-ambientales (estimar la concentración de un contaminante sobre unidades de remediación). Para que los cálculos tengan un sentido físico, es necesario que la variable estudiada sea aditiva; por ejemplo, la estimación en un bloque de una variable como el pH no se podrá realizar, pues el pH de un bloque no es igual a la media aritmética de los pH puntuales en ese bloque.

A modo de ejemplo, supongamos que se desea estimar el valor en un bloque V a partir de datos puntuales ubicados en los sitios $\mathbf{x}_1 \dots \mathbf{x}_n$. El sistema de kriging ordinario de bloques es muy similar al de kriging puntual, salvo por el segundo miembro:

$$\begin{pmatrix} C(\mathbf{x}_1 - \mathbf{x}_1) & \cdots & C(\mathbf{x}_1 - \mathbf{x}_n) & 1 \\ \vdots & & \vdots & \vdots \\ C(\mathbf{x}_n - \mathbf{x}_1) & \cdots & C(\mathbf{x}_n - \mathbf{x}_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{pmatrix} = \begin{pmatrix} \bar{C}(\mathbf{x}_1, V) \\ \vdots \\ \bar{C}(\mathbf{x}_n, V) \\ 1 \end{pmatrix}$$

con

$$\bar{C}(\mathbf{x}_\alpha, V) = \frac{1}{|V|} \int_V C(\mathbf{u} - \mathbf{x}_\alpha) d\mathbf{u} \approx \frac{1}{M} \sum_{m=1}^M C(\mathbf{u}_m - \mathbf{x}_\alpha),$$

siendo $\{\mathbf{u}_1 \dots \mathbf{u}_M\}$ un conjunto de puntos que discretizan el bloque V .

6.3. Co-kriging

Se trata de la versión multivariable del kriging, donde se busca estimar el valor de una variable tomando en cuenta los datos de esta variable y de otras variables correlacionadas. Por ejemplo, estimar la concentración de cobalto en un sitio determinado, con ayuda de los datos de concentración de cobalto y de níquel. La puesta en marcha del co-kriging requiere tener los modelos variográficos de cada variable (Co y Ni), así como *variogramas cruzados* entre las distintas variables, para medir la correlación existente entre estas variables.

El variograma cruzado entre dos variables (Z_1 y Z_2) se define como:

$$\gamma_{12}(\mathbf{h}) = \frac{1}{2} \text{cov}\{Z_1(\mathbf{x} + \mathbf{h}) - Z_1(\mathbf{x}), Z_2(\mathbf{x} + \mathbf{h}) - Z_2(\mathbf{x})\}$$

y se puede inferir a partir de los datos disponibles al plantear:

$$\hat{\gamma}_{12}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{N(\mathbf{h})} [z_1(\mathbf{x}_\alpha) - z_1(\mathbf{x}_\beta)][z_2(\mathbf{x}_\alpha) - z_2(\mathbf{x}_\beta)]$$

donde $N(\mathbf{h}) = \{(\alpha, \beta) \text{ tal que } \mathbf{x}_\alpha - \mathbf{x}_\beta = \mathbf{h}, \text{ siendo ambas variables } z_1 \text{ y } z_2 \text{ medidas en } \mathbf{x}_\alpha \text{ y } \mathbf{x}_\beta\}$. Para poder determinar el variograma cruzado experimental, se necesita tener los datos de

las diferentes variables en los mismos sitios; por lo tanto, no se puede calcular en caso de *heterotopía* (es decir, cuando las variables están medidas en sitios distintos, en cuyo caso el conjunto $N(\mathbf{h})$ es vacío para todo vector \mathbf{h}). Más detalles se exponen en el Anexo A.

6.4. Kriging no lineal

Estos métodos consisten en aplicar kriging a una transformada (no lineal) de la variable Z , luego en volver a esta variable. Esta etapa de transformación de vuelta no es trivial, pues requiere introducir correcciones para que el estimador final no tenga sesgo. Por ejemplo, si la transformada logarítmica tiene una distribución Gaussiana, se obtiene el llamado *kriging lognormal*:

$$Z^*(\mathbf{x}_0) = \exp\left\{ \sum_{\alpha=1}^n \lambda_{\alpha} \ln[Z(\mathbf{x}_{\alpha})] + \frac{\sigma_{KO}^2(\mathbf{x}_0)}{2} + \mu \right\}$$

donde $\sigma_{KO}^2(\mathbf{x}_0)$ es la varianza de kriging ordinario de $\ln[Z(\mathbf{x}_0)]$ y μ es el multiplicador de Lagrange introducido en el sistema de kriging ordinario. Esta expresión muestra que el kriging lognormal de la variable Z no coincide con la exponencial del kriging de $\ln(Z)$. El regreso a la variable inicial (Z) necesita introducir un factor correctivo $\sigma_{KO}^2(\mathbf{x}_0)/2 + \mu$, sin el cual la estimación Z^* sería sesgada.

Otros métodos de kriging no lineal buscan caracterizar el valor desconocido $Z(\mathbf{x}_0)$ no por un valor estimado, sino que por una distribución de probabilidad. Entre otros métodos, citemos:

- El **kriging de indicadores**, basado en una codificación de la variable Z en un conjunto de variables binarias o indicadores, según si Z sobrepasa o no determinados umbrales.
- El **kriging disyuntivo** (co-kriging de indicadores)
- El **kriging multi-Gaussiano**, basado en una transformación de la variable original Z en una variable de distribución Gaussiana.

7. Observaciones sobre el sistema de kriging

Se puede hacer los siguientes comentarios sobre el sistema de kriging.

- Los ponderadores y la varianza de kriging toman en cuenta:
 - 1) las distancias entre el sitio a estimar y los sitios con datos, mediante los términos $C(\mathbf{x}_{\alpha} - \mathbf{x}_0)$ o $\gamma(\mathbf{x}_{\alpha} - \mathbf{x}_0)$;
 - 2) la configuración geométrica de los sitios con datos y la posible redundancia de la información que contienen, por medio de los términos $C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta})$ o $\gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta})$;

3) la continuidad espacial de la variable, descrita por la función de covarianza C o por el variograma γ .

Sin embargo no toman en cuenta los valores de los datos mismos $\{z(\mathbf{x}_\alpha), \alpha = 1 \dots n\}$. Por lo tanto, conociendo el modelo variográfico, se puede anticipar cual será la precisión de la estimación a partir de una configuración dada de los sitios con datos. Esta propiedad no obstante es una limitación del kriging lineal, como contrapartida a la simplicidad del método. Intuitivamente, uno presiente que la precisión de una estimación es menor en las zonas cuyos valores tienen mayor variabilidad (a menudo, corresponden a las zonas de valores altos, debido a lo que se denomina el *efecto proporcional*) que en aquellas de baja variabilidad.

- En general, el ponderador asignado a un sitio con dato es mayor cuando este sitio es más cercano al sitio a estimar. Pero varias situaciones pueden perturbar esta constatación “intuitiva”:
 - **Existencia de un fuerte efecto pepita** en el modelo variográfico, que tiende a dar el mismo ponderador a todos los sitios con datos (como hay ausencia de continuidad espacial, un dato próximo aporta tanta información como uno lejano).
 - **Presencia de una anisotropía**: existe una dirección del espacio donde la variable tiene menos continuidad espacial, es decir, donde la correlación es menor. Así, un dato cercano en esta dirección puede aportar “menos información” que un dato más alejado en otra, y tener de esta forma un ponderador de kriging menor.
 - **Redundancias entre datos**: cuando varios datos están agrupados y cercanos unos a otros, se vuelven redundantes. Su ponderador acumulado es prácticamente el mismo que el ponderador que recibiría un solo dato en lugar del grupo. Dicho en otras palabras, el kriging corrige los efectos debidos a las irregularidades de la malla de muestreo y no sobre-pondera los datos agrupados en perjuicio de los datos aislados (contrariamente a estimadores clásicos como el inverso de la distancia).
 - **Efecto pantalla**: puede ocurrir que un sitio con dato haga pantalla a otro respecto al sitio a estimar. El sitio “cubierto”, aunque cercano al sitio a estimar, puede tener un ponderador bajo, incluso nulo (efecto pantalla total⁴). La estimación ignora el valor observado en el sitio cubierto. Cuando este efecto es indeseable, se puede atenuar, introduciendo en el modelo un efecto pepita. Otra solución consiste en agrupar los dos sitios y asignarles el mismo ponderador, con el costo de una pérdida de precisión.

⁴ En el espacio de una dimensión, se tiene como ejemplos de efecto pantalla total, el caso de una covarianza exponencial (en kriging simple) y el de un variograma lineal (en kriging ordinario). Para variogramas lineales en el origen, como el modelo esférico, se tiene un efecto pantalla casi total, en kriging ordinario, si todas las distancias consideradas quedan en la parte lineal del variograma. En dos dimensiones, el kriging simple produce un efecto pantalla entre direcciones perpendiculares cuando la covarianza puede factorizarse, como por ejemplo, la covarianza Gaussiana: si se plantea $\mathbf{h} = (h_1, h_2)$, entonces $C(\mathbf{h}) = C(h_1) \times C(h_2)$.

- **Efecto pantalla inverso:** en el kriging ordinario, puede ocurrir que sitios alejados tengan ponderadores mayores que sitios cercanos. Este fenómeno se explica porque los sitios alejados participan en la estimación de la *media* (desconocida) de la variable, la que interviene indirectamente en la estimación del valor desconocido.
- **Efecto de relevo:** la presencia de un sitio con dato cercano al sitio a estimar permite a un sitio con dato “lejano”, o sea, ubicado a una distancia superior al alcance del variograma, tener un ponderador no despreciable en la estimación, mientras que en ausencia del sitio cercano, su ponderador habría sido nulo (kriging simple) o muy bajo (kriging ordinario).
- La mayoría de los modelos variográficos⁵ permiten encontrar ponderadores negativos o mayores que 1, incluso con la condición que la suma de los ponderadores valga 1 (kriging ordinario). La ventaja de este proceder es que puede entregar estimaciones que salen de los límites dados por los datos. Por el contrario, las técnicas de interpolación por combinaciones lineales ponderadas que imponen pesos comprendidos entre 0 y 1 entregan estimaciones que siempre están entre el mínimo y el máximo valor observado. Ahora bien, en la mayoría de los casos, no hay razón para que los valores de los datos alcancen los valores extremos potenciales de la zona de estudio, por lo que resulta interesante tener ponderadores fuera del intervalo [0,1]. Como contrapartida, en el caso en que la variable es positiva (por ejemplo, la concentración de un contaminante), existe el riesgo de encontrar estimaciones negativas. Este riesgo es más importante cuando la variable presenta una distribución fuertemente asimétrica: un ponderador negativo, incluso si es bajo, asignado a un valor alto puede conducir a una estimación negativa si los otros valores de los datos no son muy altos.

La Figura 5 ilustra esta propiedad, con un ejemplo unidimensional de ocho sitios con datos, donde se compara el estimador del kriging ordinario con aquel del inverso de la distancia. El primero usa un modelo de variograma parabólico en el origen, propicio a la obtención de ponderadores negativos, mientras que el segundo atribuye a cada dato un ponderador siempre positivo, inversamente proporcional a su distancia al sitio a estimar, y no permite que el valor estimado supere el mayor dato.

- La matriz del primer miembro del sistema de kriging depende solamente de la posición relativa de los sitios con datos. Por consiguiente, cuando varias estimaciones utilizan configuraciones de sitios idénticas, basta con invertir esta matriz una sola vez. Esto ocurre cuando se trabaja en una vecindad única, o en una vecindad móvil con datos ubicados sobre una grilla regular.
- Si los sitios con datos son distintos, se demuestra que el sistema de kriging es regular, es decir, entrega una solución única. Por ende, cuando el sistema de kriging es singular, se está en presencia de datos duplicados (repetidos en la base de datos): la matriz del primer miembro del sistema de kriging posee dos filas idénticas y no es invertible.

⁵ Una excepción notable la constituye el modelo pepítico puro, que entrega ponderadores nulos (kriging simple) o iguales a $1/n$ (kriging ordinario).

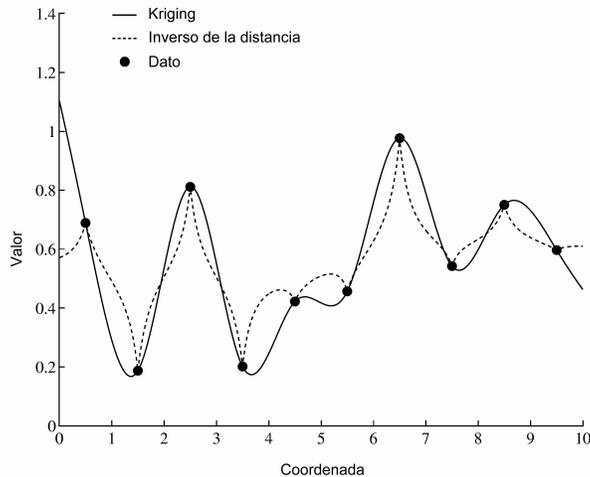


Figura 5. Comparación de las estimaciones obtenidas por kriging ordinario y por el método del inverso de la distancia

8. Propiedades del kriging

8.1. Interpolación exacta

El kriging es un interpolador exacto, es decir, la estimación en un sitio con dato vuelve a dar el valor de este dato, mientras que la varianza de kriging en ese sitio es nula. Para demostrarlo, basta con verificar que, si el sitio a estimar \mathbf{x}_0 es el mismo donde se midió un dato, el estimador definido por $Z^*(\mathbf{x}_0) = Z(\mathbf{x}_0)$ cumple las tres restricciones del kriging. En especial, el error de estimación asociado es nulo, luego obviamente insesgado y de varianza mínima (nula).

8.2. Propiedad de suavizamiento (alisamiento)

El mapa de los valores estimados por kriging es más suave que el mapa de los valores reales, es decir, que presenta menos fluctuaciones. La búsqueda de una estimación precisa se acompaña inevitablemente de este efecto de *suavizamiento*, pues no se puede inventar los detalles que no aparecen en los datos observados (Figura 6).

La propiedad de suavizamiento tiene importantes consecuencias en la práctica cuando se busca determinar los valores de la variable regionalizada en relación a un umbral. Por ejemplo, en evaluación de recursos minerales, se desea saber si la ley de mineral supera o no una ley de corte que asegura la rentabilidad del negocio. En contaminación ambiental, uno está interesado en determinar si la concentración de contaminantes es mayor que un umbral crítico de toxicidad, para decidir si vale la pena o no tomar medidas de remediación. En agricultura, es preciso que las concentraciones de nutrientes o que el pH sean mayores

que ciertos límites para asegurar que el terreno sea fértil. En todas estas problemáticas donde interesa la distribución de los valores en relación a un valor umbral, el kriging aporta una respuesta sesgada a causa de la propiedad de suavizamiento. Para obtener estimaciones insesgadas, se debe recurrir a métodos de kriging no lineal (kriging de indicadores, kriging disyuntivo, kriging multi-Gaussiano) o métodos de simulación cuyo objetivo es reproducir la variabilidad espacial de la variable regionalizada.

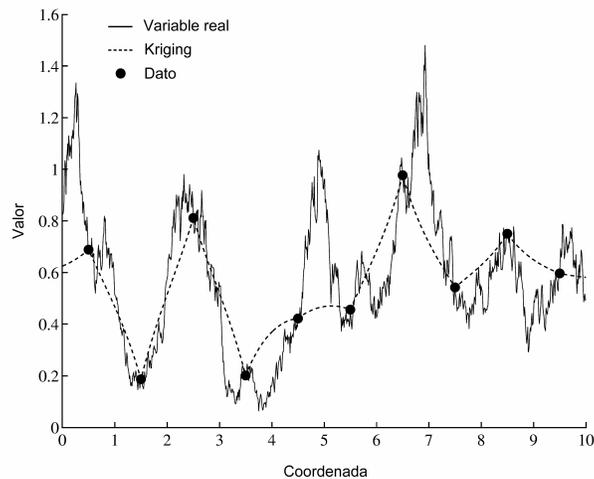


Figura 6. Ilustración de la propiedad de suavizamiento

9. Validación cruzada

Se puede verificar la adecuación entre los datos y los parámetros adoptados (modelo de variograma, vecindad de kriging), utilizando la llamada técnica de la **validación cruzada**. El principio es estimar sucesivamente, mediante kriging, cada dato, considerando sólo los datos restantes. Se puede calcular entonces el error de estimación (diferencia entre el valor estimado y el valor verdadero) en cada sitio con dato y realizar un análisis estadístico de los errores cometidos en todos los sitios con datos.

La validación cruzada es presentada usualmente bajo la forma de pruebas gráficas, en especial:

- La nube de correlación entre los valores de los datos $\{z(\mathbf{x}_\alpha), \alpha = 1 \dots n\}$ y los valores estimados $\{z^*(\mathbf{x}_\alpha), \alpha = 1 \dots n\}$.
- El histograma de los errores estandarizados

$$\frac{z^*(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha)}{\sigma^*(\mathbf{x}_\alpha)},$$

donde $\sigma^*(\mathbf{x}_\alpha)$ es la desviación estándar de kriging (o sea, la raíz cuadrada de la varianza de kriging) en el sitio \mathbf{x}_α . Una estimación se considera buena si el error estandarizado asociado pertenece al intervalo $[-\alpha, \alpha]$, por ejemplo con $\alpha = 2.5$ (este límite es elegido arbitrariamente). El análisis de los errores estandarizados, en lugar de los errores absolutos, permite liberarse de ciertas limitaciones de naturaleza geométrica. En efecto, independientemente del modelo variográfico, el error absoluto tiende a ser alto para un dato aislado, y bajo para un dato cercano a otros datos. Esto se corrige al considerar el error estandarizado, pues la desviación estándar de kriging cuantifica la precisión que se espera de la estimación. Por construcción, el error estandarizado tiene una varianza igual a 1, independientemente de si existen datos cercanos del sitio a estimar.

- La nube de correlación entre los errores estandarizados y los valores estimados⁶.
- El mapa de ubicación de los datos, donde se localiza los datos “mal” estimados, es decir aquellos cuyos errores estandarizados salen del intervalo $[-\alpha, \alpha]$.

Se puede procurar satisfacer lo mejor posible los siguientes criterios estadísticos.

- 1) Las medias de los errores y de los errores estandarizados miden el *sesgo* del estimador, y deben ser cercanas a cero. De hecho, este criterio resulta secundario con respecto al modelamiento del variograma y la elección de la vecindad de kriging pues los valores obtenidos por kriging son insesgados por construcción. Así, en la práctica, la media de los errores tenderá hacia cero.
- 2) La varianza de los errores, que mide la *precisión* del estimador, debe ser mínima.
- 3) La varianza de los errores *estandarizados* debe ser cercana a 1:

$$\frac{1}{n} \sum_{\alpha=1}^n \left(\frac{z^*(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha)}{\sigma^*(\mathbf{x}_\alpha)} \right)^2 \approx 1.$$

Este criterio indica que el modelo de variograma cuantifica adecuadamente la amplitud de los errores cometidos, es decir, que el variograma elegido no es demasiado optimista ni tampoco pesimista.

- 4) El coeficiente de correlación entre los valores estimados y los valores de los datos debe ser lo más cercano posible a 1.
- 5) El número de datos mal estimados debe ser el menor posible. Para fijar las ideas, este número se considera satisfactorio si representa menos del 5% del total de los datos.

⁶ Esta prueba permite verificar si la media de los errores cometidos (*sesgo*) depende de los valores estimados, caso en el cual se dice que existe un *sesgo condicional*. En la práctica, para evitar tal situación, se debe elegir una vecindad de kriging que contenga suficientes datos (> 15).

Estos criterios permiten comprobar el desempeño del kriging y comparar la calidad de diferentes ajustes posibles para el variograma.

Una técnica similar a la validación cruzada es el **jack-knife**, el cual no considera una reposición de los datos: se divide los datos en dos grupos y se estima los datos de un grupo a partir de los datos del otro grupo.

10. Aplicación a los datos de concentración de cobalto

10.1. Elección del plan de kriging

Reanudamos ahora con el ejemplo de los datos de contaminación de suelo presentados en los capítulos anteriores. Nos interesamos en estimar las concentraciones de cobalto en toda la zona de interés mediante kriging ordinario (de media desconocida). Primero, se debe elegir un plan de kriging, para lo cual se propone las tres opciones siguientes:

- **Plan 1**, que utiliza los 2 datos más cercanos del sitio a estimar.
- **Plan 2**, que utiliza 24 datos (6 por cuadrante).
- **Plan 3**, que utiliza 48 datos (12 por cuadrante).

Todos los planes utilizan una vecindad en forma de elipse, con el eje mayor orientado en la dirección de mayor continuidad (N60°E). Para decidir cuál es el plan más adecuado, se realiza una validación cruzada. La Figura 7 presenta los principales resultados obtenidos: histogramas de los errores cometidos en los sitios con datos y nubes de correlación entre valores verdaderos y valores estimados. Se observa que el plan 1 conduce a los errores con mayor dispersión y a la menor correlación entre valores verdaderos y estimados, mientras que los planes 2 y 3 entregan mejores resultados (con una leve superioridad para el plan 2). Acorde a estos resultados, se decide proseguir la estimación local utilizando el plan 2.

	errores		errores estandarizados	
	media	varianza	media	varianza
Plan 1	0.08	5.34	0.03	0.83
Plan 2	0.02	4.49	0.00	0.91
Plan 3	0.03	4.62	0.01	0.93

Tabla 1. Estadísticas básicas sobre errores y errores estandarizados de validación cruzada.

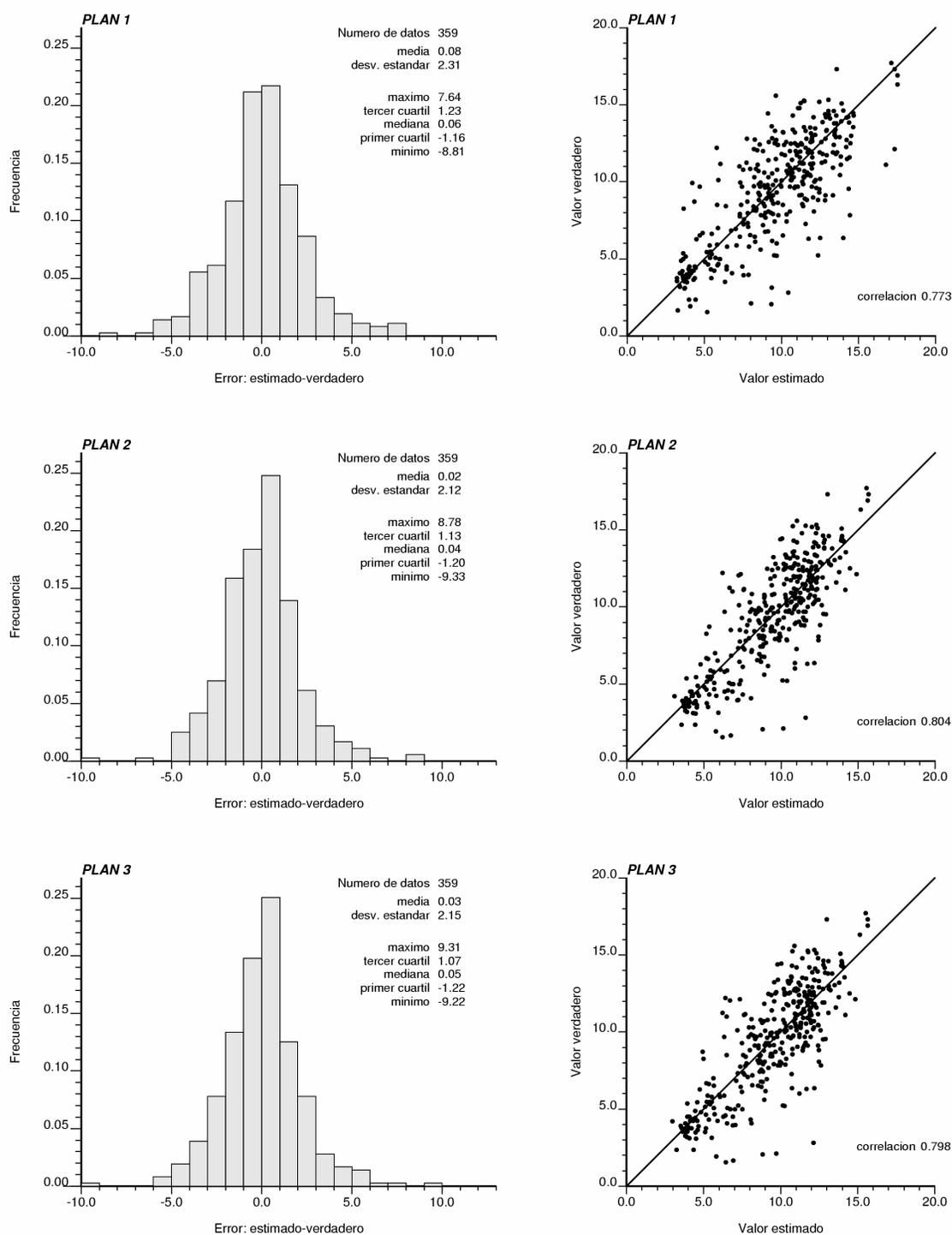


Figura 7. Validación cruzada con los tres planes de kriging propuestos. Izquierda: histogramas de errores de estimación. Derecha: nubes de correlación entre valores verdaderos y valores estimados

10.2. Estimación de las concentraciones de cobalto sobre soporte puntual

Provistos del modelo de variograma y de la vecindad de kriging, podemos estimar las concentraciones de cobalto en la zona de estudio. Las Figuras 8 y 9 presentan los mapas de los valores estimados y de las desviaciones estándar de estimación (desviación estándar de los errores de kriging), respectivamente. En cada caso, se muestra los resultados obtenidos con kriging simple (tomando la media desagrupada como valor de la media) y con kriging ordinario.

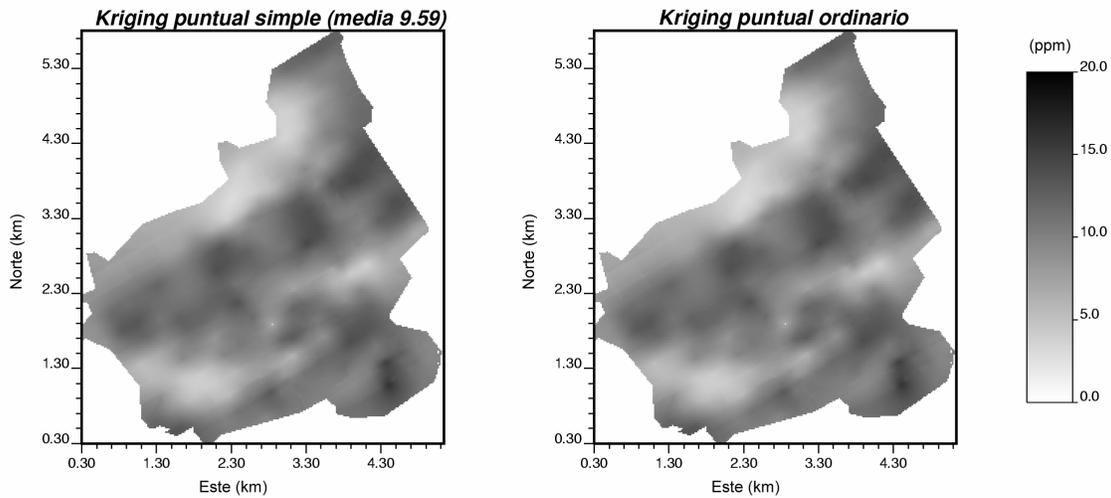


Figura 8. Estimaciones de las concentraciones de cobalto por kriging simple (izquierda) y kriging ordinario (derecha)

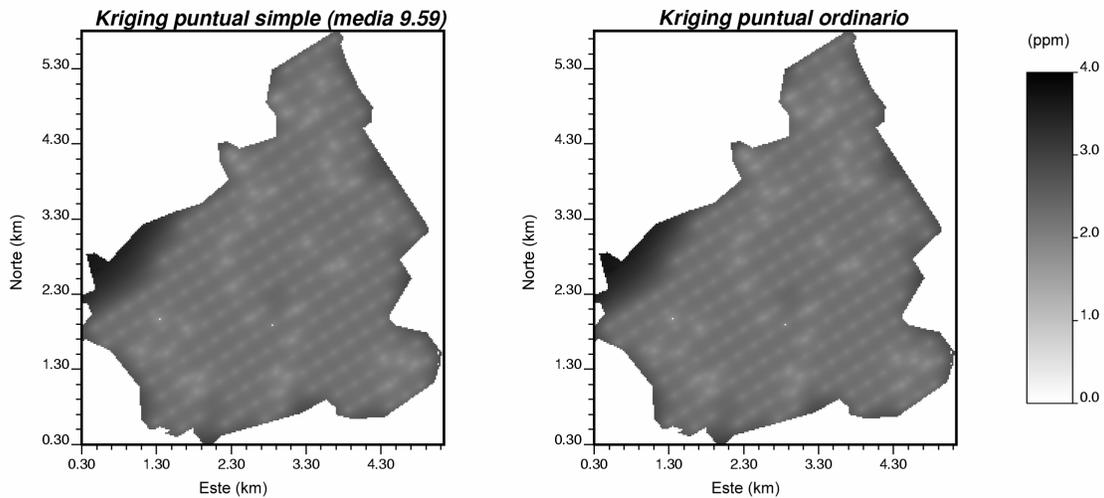


Figura 9. Desviación estándar de kriging para las concentraciones de cobalto (raíz cuadrada de la varianza de kriging)

Los mapas de estimaciones ponen en evidencia las áreas más contaminadas de la zona, que aparecen con tonos oscuros. Las variaciones de los valores estimados son muy suaves, ilustrando la propiedad de alisamiento del kriging. Esto se puede corroborar al comparar las estadísticas básicas de los datos originales y de los valores estimados (Tabla 2): se observa que la dispersión de los últimos es mucho menor que aquella de los primeros, aunque los valores promedios son muy similares (propiedad de insesgo). En cuanto a las desviaciones estándar de kriging, éstas son bajas cerca de los sitios con datos y aumentan al acercarse de los bordes de la zona, donde uno trabaja cada vez más extrapolando. Además, no dependen de los valores estimados, sino que solamente de la configuración geométrica de los datos.

	mínimo	máximo	media	varianza
datos desagrupados	1.55	20.6	9.59	11.6
kriging simple	2.81	15.9	9.53	6.45
kriging ordinario	2.81	15.9	9.52	6.57

Tabla 2. Estadísticas básicas sobre datos y estimaciones por kriging.

Los mapas de estimación obtenidos por kriging simple y kriging ordinario son muy parecidos. Esto indica que los datos son suficientemente abundantes como para no necesitar el valor de la media. En otras palabras, la suma de los ponderadores de kriging simple es cercana a 1, de modo que la media recibe una ponderación muy pequeña y, finalmente, no importa si esta media se conoce (kriging simple) o no se conoce (kriging ordinario). Las diferencias entre kriging simple y kriging ordinario suelen producirse en zonas que poseen escasos datos y en situaciones de extrapolación, en cuyos casos el kriging simple acerca el valor estimado a la media, mientras que el kriging ordinario da una media local calculada con los datos contenidos en la vecindad.

Resulta interesante comparar los mapas de kriging con aquellos que se obtendría por un método de interpolación tradicional, como el del inverso de la distancia, que pondera cada dato por una potencia del inverso de la distancia entre su posición y la posición del sitio a estimar. La Figura 10 presenta los mapas de estimaciones obtenidas para potencias 1, 2 y 3, utilizando la misma vecindad que en el caso del kriging.

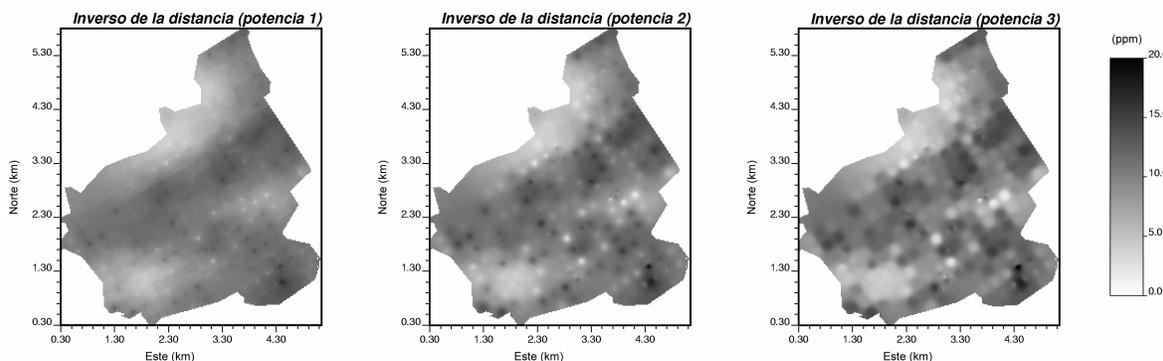


Figura 10. Estimaciones de las concentraciones de cobalto por inverso de la distancia, con potencias 1, 2 y 3

Como el kriging, el método del inverso de la distancia interpola exactamente los datos, pero produce artefactos en los mapas de valores estimados. En especial, las estimaciones tienden a coincidir con extremos locales en los sitios con datos, de modo que las líneas de isovalores tienen forma de círculos alrededor de estos sitios; este efecto es muy visible en los mapas correspondientes a las potencias 2 y 3. Algunos algoritmos intentan disminuir estos artefactos al introducir coeficientes de suavizamiento, pero la interpolación ya no es exacta. La principal crítica que se puede hacer al método del inverso de la distancia, así como a los otros métodos de interpolación determinísticos (más cercano vecino, splines, interpolación por triangulación, etc.), es que no toma en cuenta la continuidad espacial de la variable regionalizada en estudio: efecto pepita, anisotropía, alcance. Se trata de un método automático que se aplica “ciegamente” al conjunto de datos y cuyo fin es dar un mapa más estético que preciso. Por el contrario, el kriging hace preceder a la interpolación un análisis variográfico, que permite elaborar un modelo coherente con las observaciones, construir una estimación insesgada de la variable en los sitios sin datos y medir la precisión de esta estimación.

10.3. Estimación de las concentraciones de cobalto sobre bloques

Para el tratamiento de las áreas contaminadas, se considera “unidades de remediación” que corresponden a bloques de tamaño $20\text{m} \times 20\text{m}$. Estas unidades deben ser extraídas si la concentración promedio de cobalto es demasiado alta (superior a un umbral de toxicidad definido por las normas ambientales). Por lo tanto, interesa la estimación sobre un soporte de bloques, más que las estimaciones sobre un soporte puntual (soporte de los datos). La Figura 11 presenta los mapas de kriging de bloques y de desviaciones estándar de kriging asociadas. Las estimaciones han sido realizadas por kriging ordinario, con el mismo modelo de variograma y la misma vecindad que aquellos definidos en el caso puntual.

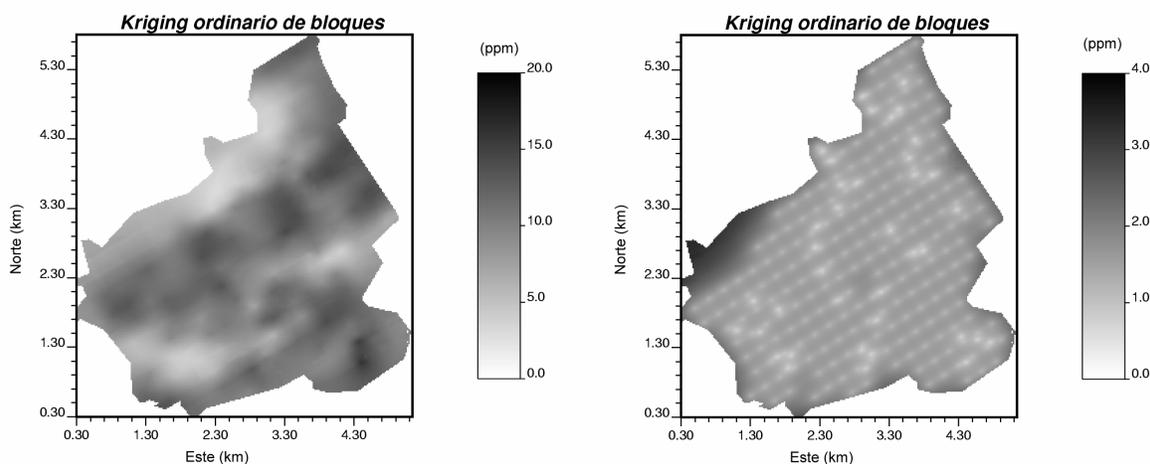


Figura 11. Kriging ordinario de las concentraciones de cobalto en bloques de $20\text{m} \times 20\text{m}$. Izquierda: concentración estimada. Derecha: desviación estándar del error.

El mapa del kriging de bloques es muy parecido al mapa del kriging puntual (Figura 8), o sea, la estimación de la concentración promedio en un bloque difiere poco de aquella de su punto central. Esta similitud se debe a que el tamaño de los bloques es inferior a la malla de muestreo: la estimación puntual no permite diferenciar los puntos de un mismo bloque y es prácticamente idéntica a la estimación del bloque mismo. Por otro lado, el mapa de las desviaciones estándar de kriging presenta diferencias notables con respecto al caso puntual (Figura 9), en particular sus valores son menores. Esto es una consecuencia del *efecto de soporte*, según el cual los valores de soporte de bloques tienen menos variabilidad que los valores de soporte puntual; en este contexto, es lógico que las incertidumbres de estimación sean menores en el caso del kriging de bloques.

Ahora, hay que tener presente que la estimación, por más precisa que sea, nunca es idéntica a la realidad. En particular, debido a la propiedad de suavizamiento del kriging, se debe esperar que las concentraciones reales de cobalto en la zona tengan más variabilidad que las concentraciones estimadas por kriging. Esta diferencia de variabilidad entre valores reales y estimados puede originar un sesgo en las previsiones si se efectúa una selección sobre el mapa de los valores estimados. En particular, se puede subdimensionar el volumen de terreno que se debe remediar (unidades cuyas concentraciones de cobalto son mayores que un umbral dado). Para evitar este sesgo, deberemos recurrir a métodos de simulación condicional.

11. Efectos de los parámetros en los resultados del kriging

A continuación, consideraremos la configuración de la Figura 1, donde se tiene 6 sitios con datos (denotados A, B, C, D, E y F) y un sitio en donde se desea estimar el valor (?). Se busca estudiar los efectos de los parámetros del modelo de variograma (meseta, alcance, anisotropía, efecto pepita...) y del tipo de kriging en los ponderadores asignados a los sitios con datos y en la varianza de kriging. En cuanto no se diga lo contrario, las estimaciones se realizan con kriging ordinario, es decir, de media desconocida.

11.1. Tipo de modelo

Examinaremos las repercusiones del tipo de modelo en la estimación del sitio "?", al comparar los resultados obtenidos con cuatro modelos isótropos:

- Gaussiano de alcance 1 y meseta 1
- esférico de alcance 1 y meseta 1
- esférico de alcance 1 y meseta 0.5 + pepita de meseta 0.5
- pepita pura de meseta 1.

La Figura 12 muestra que los ponderadores de kriging y la varianza del error difieren notablemente según el modelo⁷. En el caso del modelo Gaussiano, los datos situados cerca del sitio a estimar (en particular el dato C) tienen ponderadores positivos bastante mayores que en el caso del modelo esférico y la varianza del error es 12 veces menor. La explicación considera el hecho de que el comportamiento parabólico en el origen del modelo Gaussiano representa una variable regionalizada extremadamente regular en el espacio⁸, por lo que los datos cercanos tienen mayor importancia en la estimación; la varianza de kriging es menor, puesto que la estimación se supone más precisa que en el caso del variograma esférico, que modela una variable regionalizada más errática.

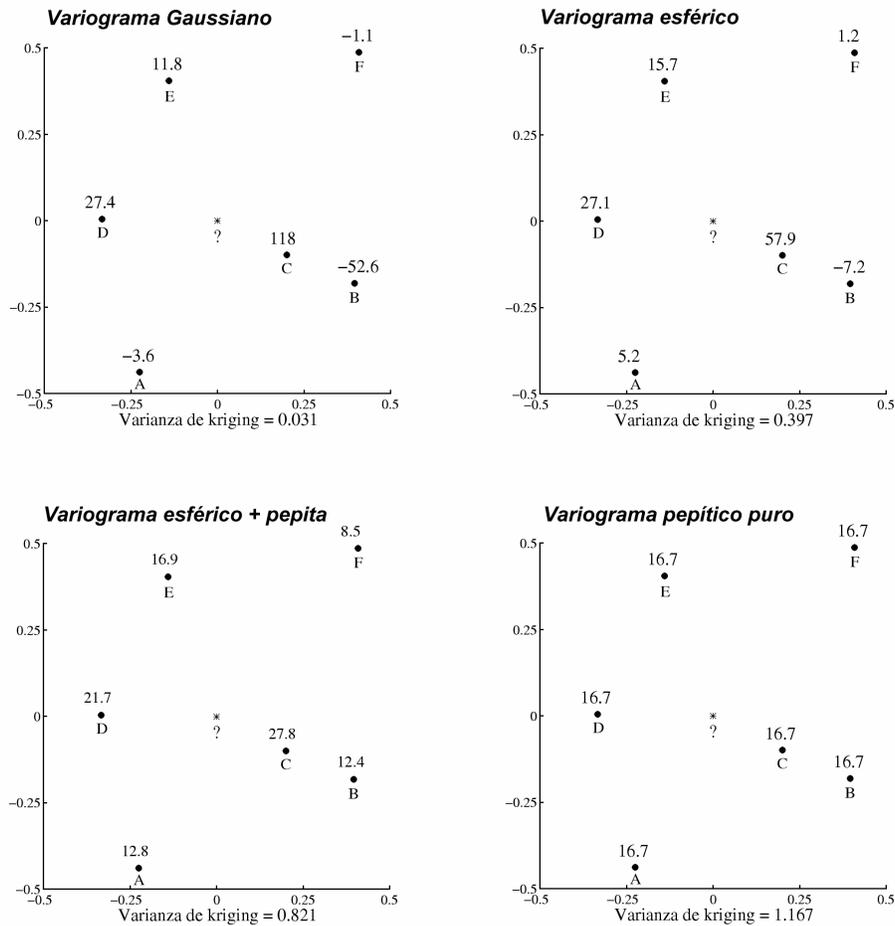


Figura 12. Influencia del tipo de modelo en los resultados del kriging

⁷ Por razones de visualización, los ponderadores de kriging están expresados en porcentajes; se verificará que su suma es igual a $100\% = 1$.

⁸ La experiencia muestra que el variograma Gaussiano es susceptible de conducir a inestabilidades numéricas cuando algunos datos son cercanos a otros. La introducción de un efecto pepita, aunque pequeño, mejora significativamente la robustez de los resultados: los ponderadores de kriging son entonces menos sensibles a leves modificaciones de la configuración de kriging.

Tanto en el caso del modelo esférico como en el del modelo Gaussiano, el sitio B se ve afectado por un ponderador negativo. Se puede notar que con el modelo Gaussiano, este efecto es más importante que en el caso esférico. La explicación está en la estructura muy continua del modelo Gaussiano. En efecto, el sitio B está oculto por el sitio C con respecto al sitio a estimar. Podemos imaginar que, si el valor medido en el sitio B fuera alto y si el del sitio C que lo oculta resultara bajo, una evolución regular proporcionaría en el sitio a estimar un valor todavía más bajo, por lo cual se requiere ponderar negativamente el sitio más externo B.

La introducción de un efecto pepita modifica a la vez los ponderadores y la varianza de kriging, que aumenta cuando la constante de pepita aumenta. En presencia de efecto pepita, la diferencia entre el ponderador mayor y el menor se reduce con respecto al caso de un variograma sin efecto pepita; también desaparecen los ponderadores negativos. En el caso del efecto pepita puro, deja de existir correlación espacial entre los valores de los diferentes sitios. Los ponderadores asignados son todos iguales a $1/6$ (o sea, 16.7%) y la estimación se transforma en la media aritmética de los datos, sin importar su posición.

11.2. Meseta

Veamos ahora cómo se traduce, en la estimación del sitio “?”, el valor de la meseta del variograma. Se considera dos modelos esféricos isótropos de mismo alcance (1), que sólo se diferencian por su meseta, igual a 1 ó 2. La Figura 13 muestra que los ponderadores de kriging no cambian y, por consiguiente, tampoco el valor estimado. En otros términos, multiplicar el variograma por una constante positiva (en este caso, por 2) no modifica la estimación. En cambio, se puede verificar que la varianza de kriging se multiplicó por 2, es decir, que fue multiplicada por el mismo factor que la meseta del variograma.

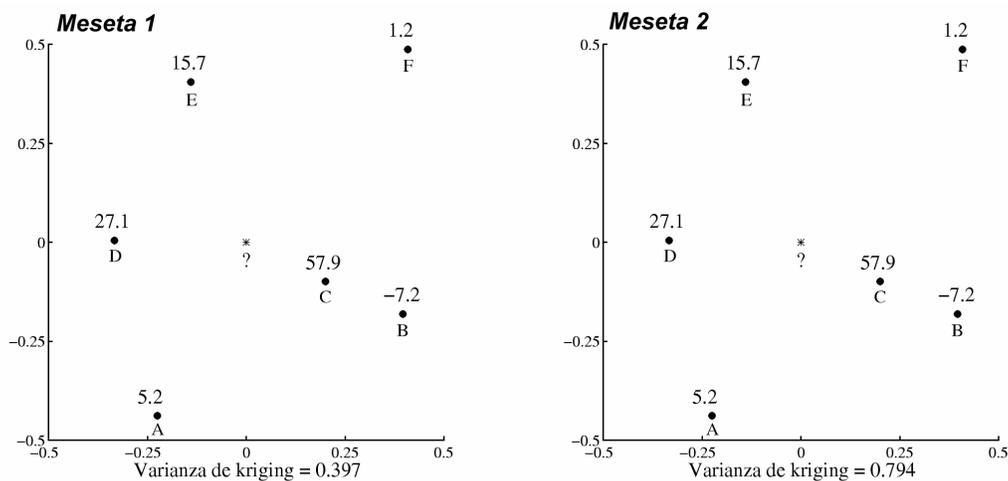


Figura 13. Influencia de la meseta

11.3. Alcance

Supongamos que tenemos tres variogramas esféricos isótropos con la misma meseta (1) y con alcances de 0.5, 1 y 2 respectivamente. La Figura 14 visualiza los resultados de las estimaciones.

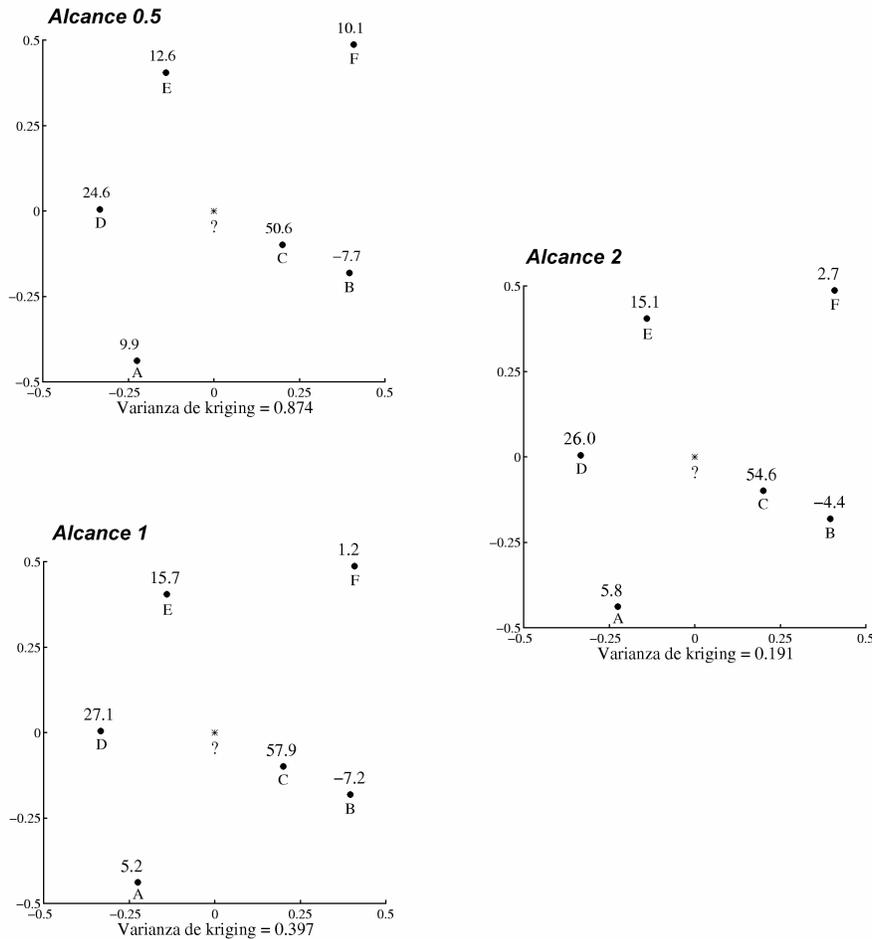


Figura 14. Influencia del alcance

En el primer caso, el alcance es bajo (0.5), de modo que la varianza del error es más alta. Se observa que los datos alejados (en especial, A y F) tienen ponderadores cercanos al 10%, aunque su distancia al sitio a estimar es mayor que el alcance. Veremos la explicación a este fenómeno al momento de comparar los resultados de kriging ordinario con aquellos de kriging simple. De este caso, se destaca una conclusión importante para la práctica: los radios de búsqueda utilizados para definir la vecindad de kriging no necesariamente tienen que ser iguales a los alcances del variograma. Si la cantidad de datos es pequeña (como el caso analizado aquí), conviene aumentar estos radios y buscar datos más allá del alcance, de modo de mejorar la precisión de la estimación. En el caso extremo de un variograma con alcance 0 (efecto pepita puro), el kriging sólo depende de la cantidad de datos encontrados en la vecindad, no de su alejamiento al sitio a estimar.

En los otros dos casos (alcances de 1 y 2), se encuentran valores similares para los ponderadores de kriging. La explicación considera que, *a la escala de trabajo* (distancias comprendidas entre 0 y 1 unidad), ambos modelos esféricos son prácticamente lineales y difieren sólo en un factor multiplicativo, lo que provoca que los ponderadores de kriging no cambien (ver sección previa acerca de la influencia de la meseta). En cambio, la varianza de kriging depende fuertemente del alcance. Al aumentar el alcance, la correlación entre los datos y el valor a estimar aumenta: los datos aportan mayor información y, por ende, la estimación se supone más precisa.

11.4. Efecto de hoyo

Examinaremos ahora las repercusiones de utilizar un variograma con efecto de hoyo (es decir, una función no monótona). Elegimos un modelo pseudo-periódico (seno cardinal) de período 0.2 y de meseta 1. El valor del período será un parámetro importante para explicar el signo y la amplitud de los ponderadores de kriging.

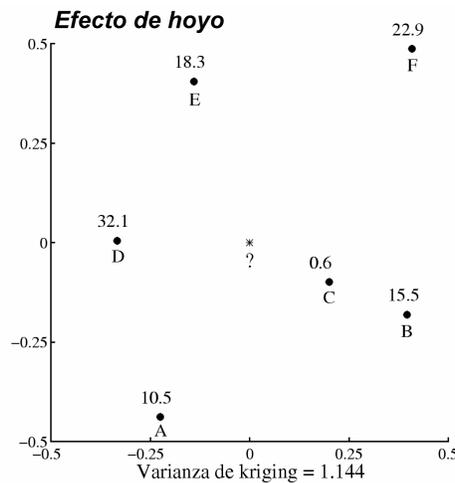


Figura 15. Influencia del efecto de hoyo

Con respecto al modelo esférico, el modelo de efecto de hoyo modifica radicalmente los ponderadores de kriging y conduce a una situación curiosa en que el sitio más cercano (C) tiene un ponderador muy bajo. Esto se debe a que la distancia entre este dato y el sitio a estimar es cercana a la distancia de mínima correlación (alrededor de 0.2). En cambio, el dato apantallado (B) recibe una ponderación mayor, por estar más correlacionado con el valor del sitio a estimar. La varianza del error es mayor que aquella obtenida con el modelo esférico, pues el modelo seno cardinal crece con más velocidad hasta sobrepasar la meseta, por lo que la continuidad espacial se deteriora más rápidamente.

11.5. Anisotropía

Hasta ahora, hemos supuesto que la continuidad de la variable estudiada era idéntica en todas las direcciones. A continuación, compararemos los resultados del kriging ordinario en el caso de un variograma esférico isótropo de meseta 1 y alcance 1 con aquellos obtenidos en el caso de un variograma esférico de meseta 1 y anisotropía geométrica, con alcance 2 en la dirección N45°E y 0.5 en la dirección ortogonal N45°W.

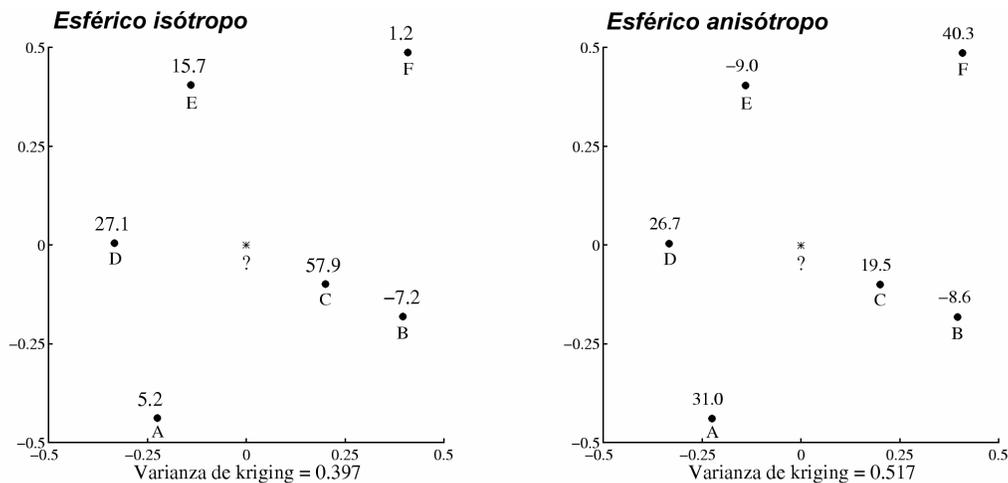


Figura 16. Influencia de la anisotropía

Los ponderadores de kriging calculados para sitios situados en la dirección principal de anisotropía (N45°E) han aumentado (sitios A y F), lo que refleja que la correlación en esta dirección es mayor que en la dirección ortogonal (sitios C y E).

11.6. Tipo de kriging: simple / ordinario

Todas las pruebas hechas anteriormente hicieron uso del kriging ordinario (es decir, el kriging de media desconocida). Nos proponemos aquí examinar las diferencias en los ponderadores de kriging que conlleva el conocimiento de la media. Se utiliza dos modelos esféricos, con alcances de 1 y 0.5 respectivamente y con la misma meseta unitaria (Figura 17).

Se tiene el siguiente resultado (denominado *teorema de aditividad*): el kriging ordinario coincide con el kriging simple si se reemplaza la media desconocida por su estimación óptima. El hecho de reemplazar la media por su estimación explica que el kriging ordinario siempre da una varianza de error mayor que el kriging simple, pues utiliza una información menos certera.

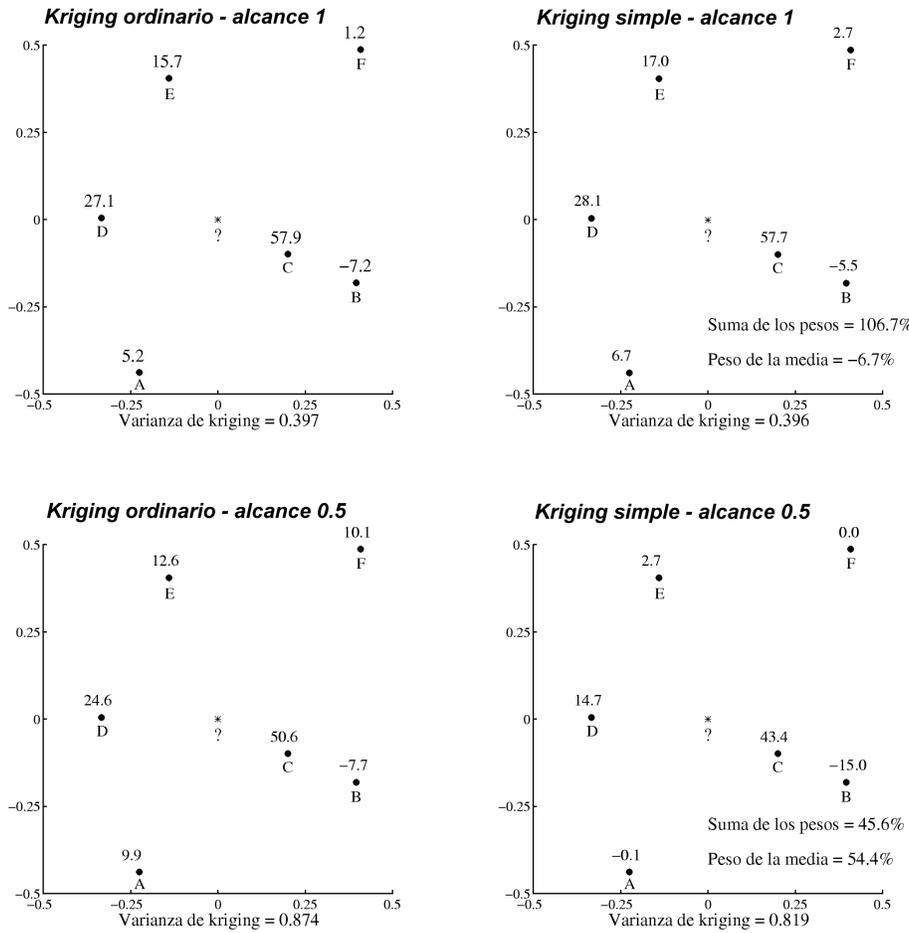


Figura 17. Influencia del tipo de kriging (ordinario / simple)

En el primer caso (variograma de alcance 1), el kriging simple y el kriging ordinario difieren poco. Esto significa que la información es suficientemente abundante para que la media, verdadera o estimada, tenga un ponderador pequeño. La pérdida de precisión entre usar un kriging simple y uno ordinario es baja. Esta situación es similar a aquella observada en el caso de estudio con los datos de contaminación de suelo.

En el segundo caso, donde el alcance es dos veces más pequeño, todo ocurre como si los datos disponibles estuvieran dos veces más lejos del sitio a estimar, de donde se produce una imprecisión mayor en la estimación. El kriging simple compensa la pérdida de información acercando el valor estimado a la media conocida (cuyo ponderador crece de -6.7% a 54.4%), mientras que el kriging ordinario estima esta media y, con respecto al kriging simple, reparte su ponderador entre los datos, a razón de alrededor de un 9% por dato.

11.7. Tipo de kriging: puntual / de bloque

Para terminar este panorama, examinemos las consecuencias de estimar el valor en un soporte mayor que el soporte de los datos, utilizando kriging ordinario con un variograma esférico isótropo de alcance 1 y meseta 1. Además del soporte puntual, se considera tres soportes, correspondientes a bloques de tamaño 0.25×0.25 , 0.5×0.5 y 1×1 , respectivamente.

Mientras el bloque queda pequeño con respecto al espaciamiento entre los datos, no se aprecia grandes cambios en la ponderación de éstos. Esto se explica porque el kriging de bloque es equivalente a promediar las estimaciones obtenidas en cada punto del bloque: al ser pequeño el bloque, la ponderación de kriging no varía mucho al estimar un punto u otro. En contraste, cuando el bloque se vuelve muy grande (1×1), todos los datos situados dentro de este bloque reciben una ponderación importante: aunque estén lejos del centro, los datos periféricos (A y F) aportan información para estimar las esquinas del bloque. En cuanto a la varianza del error, se aprecia que disminuye fuertemente al aumentar el soporte del bloque, debido al efecto de soporte (los valores de bloques tienen menos variabilidad, por lo que su estimación es más confiable).

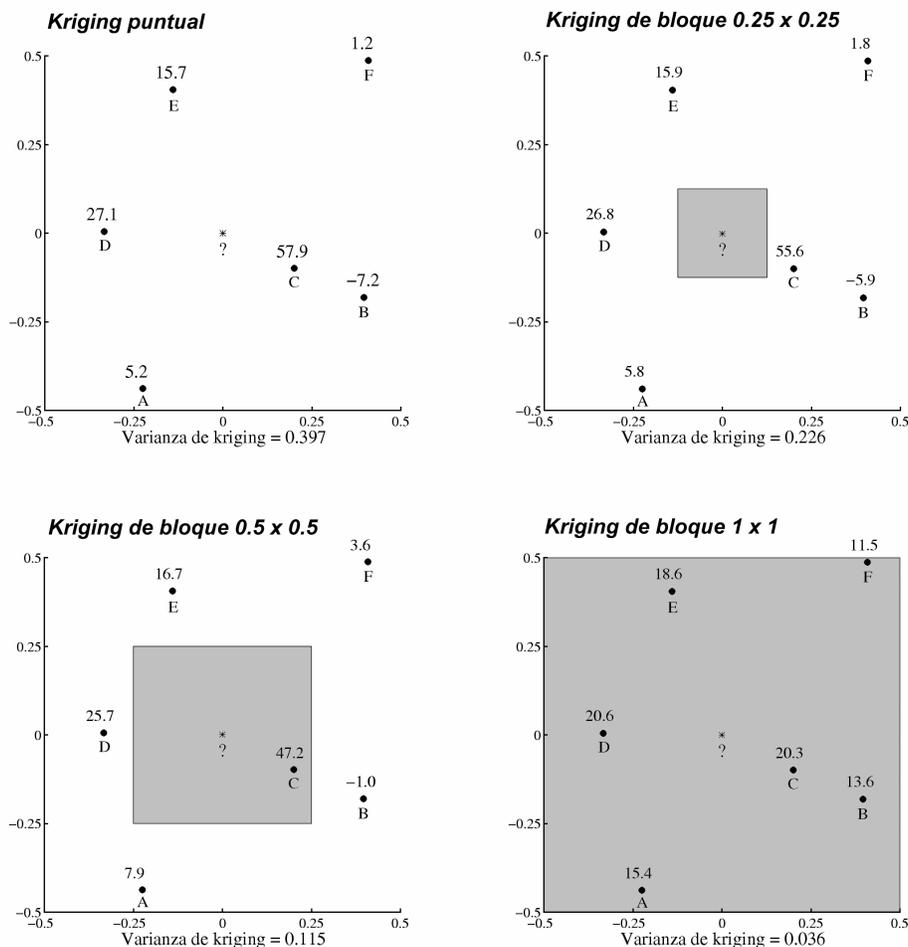


Figura 18. Influencia del tipo de kriging (puntual / de bloque)

12. Complemento: el kriging como método de interpolación

La propiedad de interpolación exacta asemeja el kriging a una técnica de cartografía. Sin embargo, tal como se ha presentado el kriging, cada sitio a estimar requiere resolver un sistema de ecuaciones lineales. Ahora bien, muchos de los métodos de cartografía (mínimos cuadrados, splines...) no se construyen al ponderar los datos vecinos del sitio a estimar, sino al ajustar una función de forma predefinida al conjunto de datos. El objetivo de este párrafo es revisar en qué medida el kriging puede considerarse como una función de interpolación.

Consideremos el sistema de kriging ordinario en un sitio \mathbf{x} :

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) + \mu = C(\mathbf{x}_{\alpha} - \mathbf{x}) & \forall \alpha = 1 \dots n \\ \sum_{\alpha=1}^n \lambda_{\alpha} = 1 \end{cases}$$

Los ponderadores $\{\lambda_{\alpha}, \alpha = 1 \dots n\}$ y el multiplicador de Lagrange μ que satisfacen este sistema son combinaciones lineales de los términos $\{C(\mathbf{x}_{\alpha} - \mathbf{x}), \alpha = 1 \dots n\}$ que aparecen en el segundo miembro. Por lo tanto, el estimador

$$z^*(\mathbf{x}) = \sum_{\alpha=1}^n \lambda_{\alpha} z(\mathbf{x}_{\alpha})$$

también es una combinación lineal de estos términos y se puede escribir:

$$z^*(\mathbf{x}) = b_0 + \sum_{\alpha=1}^n b_{\alpha} C(\mathbf{x}_{\alpha} - \mathbf{x}).$$

Esta expresión, escrita en términos de la variable regionalizada (z) y no de la función aleatoria (Z), constituye la llamada formulación *dual* del kriging. Puede servir de definición del kriging, que aparece entonces desprovisto de connotación probabilística; en especial, la noción de varianza de estimación es abandonada. Ahora, falta especificar los coeficientes $\{b_{\alpha}, \alpha = 0 \dots n\}$. Reemplazando, en la formulación dual, la función $C(\mathbf{x}_{\alpha} - \mathbf{x})$ por su expresión proveniente del sistema tradicional, y procediendo por identificación, se obtiene el siguiente sistema de ecuaciones:

$$\begin{cases} b_0 + \sum_{\beta=1}^n b_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) = z(\mathbf{x}_{\alpha}) & \forall \alpha = 1 \dots n \\ \sum_{\beta=1}^n b_{\beta} = 0 \end{cases}$$

Las primeras ecuaciones de este sistema corresponden a la restricción de interpolación exacta de los datos:

$$\forall \alpha = 1 \dots n, z^*(\mathbf{x}_\alpha) = z(\mathbf{x}_\alpha).$$

Se nota además que el sistema de ecuaciones, luego los coeficientes $\{b_\alpha, \alpha = 0 \dots n\}$, no dependen del sitio \mathbf{x} a estimar. El kriging dual aparece así bajo la forma de una función de interpolación, combinación lineal de las funciones $\{C(\mathbf{x}_\alpha - \mathbf{x}), \alpha = 1 \dots n\}$. Por lo tanto, el estimador z^* presentará las mismas propiedades analíticas que la función de covarianza C . Una consecuencia es la propiedad de suavizamiento del kriging, puesto que los modelos usuales de covarianza sólo poseen singularidades analíticas en el origen y en la vecindad del alcance y son regulares para todas las otras distancias. Por ejemplo, si la covarianza es un efecto pepita o un modelo exponencial, los únicos puntos singulares de z^* serán los sitios con datos.

Lo interesante de la formulación dual del kriging es que basta con resolver un solo sistema para obtener los coeficientes $\{b_\alpha, \alpha = 0 \dots n\}$ y deducir la estimación de todos los sitios del espacio, mientras que el enfoque directo necesita resolver tantos sistemas como hay sitios a estimar. En contrapartida, se renuncia al cálculo de la varianza de estimación. Al plantear $z^*(\mathbf{x}) = \text{constante}$, se dispone además de una ecuación implícita de las curvas de nivel. Algunas complicaciones surgen cuando se trabaja en una vecindad móvil: la función de interpolación cambia según la vecindad utilizada y puede crear discontinuidades en el mapa final (estos problemas no se plantean cuando se trabaja en una vecindad única).

Capítulo 7. Principios de simulación geoestadística

Una simulación o *modelo numérico* consiste en construir una variable regionalizada que “se parece” a aquella que se estudia, es decir que reproduce su continuidad espacial y también coincide con los valores medidos en los sitios con datos. Se recurre a las técnicas de simulación porque el conocimiento exhaustivo de la variable real es imposible, por falta de tiempo y presupuesto o a causa de obstáculos naturales. Estas técnicas se apoyan en la interpretación de la variable regionalizada como una realización de una función aleatoria y en el modelamiento de su distribución espacial.

1. Motivación: la alternativa entre kriging y simulación

El kriging permite estimar una variable regionalizada por medio de una combinación lineal ponderada de los datos. Este estimador verifica las siguientes propiedades:

- 1) **Interpolación exacta:** el kriging devuelve el valor medido en cada sitio con dato.
- 2) **Insesgo:** el error de estimación tiene una esperanza nula.
- 3) **Precisión:** el error de estimación tiene una varianza mínima.
- 4) **Suavizamiento (alisamiento):** la dispersión de los valores estimados es menor que la dispersión de los valores verdaderos. Esta propiedad tiene varias consecuencias:
 - El variograma de los valores estimados no tiene efecto pepita, aunque no sea el caso con los valores reales, y su meseta es menor que el variograma de los datos.
 - La variabilidad de los valores estimados no es uniforme en el campo: disminuye al alejarse de los sitios con datos (es decir, el mapa los valores estimados muestra más dispersión en las zonas donde el muestreo es más denso y es más suave en las zonas con escasos datos).
 - El rango de los valores estimados suele ser menor que el rango de los valores reales; por lo tanto, el kriging no permite predecir adecuadamente la ocurrencia de valores extremos.
 - No se puede manipular los valores estimados como si fueran los valores verdaderos. Por ejemplo, la cantidad de valores estimados que superan un umbral determinado no es la misma que la cantidad de valores reales sobre este mismo umbral. Dicho de otra manera, existe un *sesgo* en la estimación de funciones que involucran a valores umbrales (sólo hay insesgo cuando se estima la variable misma).

- 5) La **varianza de kriging** no depende de los valores de los datos, sino que solamente de su posición en el espacio y del modelo de variograma. Por lo tanto, no mide todas las fuentes de incertidumbre: intuitivamente, la incertidumbre es mayor cuando los datos tienen valores más dispersos que cuando tienen valores parecidos.

El principio de una simulación consiste en construir una variable ficticia que reproduce la continuidad/variabilidad espacial de la variable regionalizada $z = \{z(\mathbf{x}), \mathbf{x} \in D\}$. Esta construcción se basa en la interpretación de z como *una* realización particular de una función aleatoria $Z = \{Z(\mathbf{x}), \mathbf{x} \in D\}$. La idea es generar *otras* realizaciones de esta función aleatoria. Mientras que el kriging y otros métodos de interpolación conducen a una imagen suavizada de la realidad, las realizaciones de Z presentan la misma variabilidad espacial que la variable regionalizada real, en particular, el mismo histograma y el mismo variograma. En consecuencia, la “respuesta” de una simulación a una operación (por ejemplo, selección sobre un valor umbral) es similar a la respuesta que daría la variable real.

La simulación tendrá las siguientes propiedades:

- 1) **Interpolación exacta** cuando es “condicional” (ver sección siguiente).
- 2) **Insesgo**: la variable regionalizada o, más generalmente, una función de esta variable puede ser estimada sin sesgo (es decir, con un error de esperanza nula) por la misma función aplicada a los valores simulados.
- 3) **No suaviza**: la dispersión de los valores simulados es la misma que la dispersión de los valores verdaderos.
- 4) **No es precisa**: el error entre valor real y valor simulado no tiene una varianza mínima.

Para remediar a esta última limitación, la idea es no utilizar una simulación única, sino que un conjunto grande de simulaciones, cada una de las cuales representa un “escenario” plausible. Así, para evaluar una magnitud cualquiera, basta con efectuar los cálculos sobre un gran número de simulaciones, *como si* se tratara de los valores reales, y posteriormente tomar la media como estimación de la magnitud buscada. Mejor aun: la distribución de los resultados obtenidos sobre el conjunto de simulaciones da una imagen de la incertidumbre asociada a la magnitud estudiada, lo que permite medir la precisión de la estimación (por una varianza o por intervalos de probabilidad).

Ejemplo: datos de contaminación de suelo

Supongamos que el umbral de toxicidad para la concentración de cobalto es de 12 ppm (según la norma regulatoria) y que el valor estimado por kriging en una determinada unidad de remediación es de 10 ppm. El medio-ambientalista quiere saber si debe extraer esta unidad para tratamiento o dejarla intacta in situ. Ahora, como el kriging tiende a suavizar los valores, existe la posibilidad de que la verdadera concentración de cobalto sea mayor que el umbral de toxicidad, por lo cual la decisión de extraer o no la unidad de remediación debe considerar no sólo el valor estimado por kriging (10 ppm), sino que la incertidumbre

en torno a este valor. Sería entonces interesante caracterizar esta incertidumbre por la *probabilidad* de que la concentración de cobalto supere el umbral de 12 ppm.

Una primera opción consiste en suponer que el valor verdadero tiene una distribución normal (Gaussiana) con media igual al valor estimado por kriging y varianza igual a la varianza de kriging, lo que permite determinar la probabilidad buscada. Sin embargo, las hipótesis en las cuales se basa este cálculo son difícilmente controlables: ¿por qué el valor verdadero debería de tener una distribución Gaussiana? También se sabe que la varianza de kriging no depende de los valores de los datos, por lo cual puede no ser una medida realista de la dispersión del valor verdadero en torno al valor estimado.

El uso de simulaciones entrega una alternativa más confiable. Podríamos proceder de la siguiente forma:

- 1) Discretizar la unidad de remediación en varios puntos (por ejemplo, 10×10).
- 2) Simular 100 veces la concentración de cobalto en estos puntos.
- 3) En cada realización, rebloquear (promediar) los valores simulados para obtener la concentración promedio en la unidad de remediación.
- 4) Construir el histograma de los 100 valores rebloqueados y determinar la frecuencia con la cual se está sobre el umbral de 12 ppm.
- 5) Identificar esta frecuencia con la probabilidad buscada.

Además de calcular el riesgo de sobrepasar cualquier umbral, el histograma construido en la etapa 4) permite también darse una idea del valor máximo que se puede encontrar en la unidad de remediación. A veces, interesa más conocer este máximo para tomar la decisión de remediación, por ejemplo, si el elemento analizado es muy tóxico para la salud humana.

Situaciones que ameritan recurrir a simulaciones

- *Problemas de estimación*: se resuelven al promediar los escenarios (para estimar la variable misma) o al calcular la frecuencia de ocurrencia de un evento (para estimar la probabilidad de este evento, por ejemplo, la probabilidad de que la variable tenga un valor mayor que un umbral dado).
- *Medición de incertidumbre*: se analiza las variaciones observadas entre un escenario y otro, pudiendo definir intervalos de probabilidad para los valores desconocidos de la variable regionalizada.
- *Análisis de riesgo*: proponer diversos escenarios (uno optimista, uno pesimista, uno intermedio) para analizar el impacto en los resultados y guiar la toma de decisión.

2. Simulación condicional y no condicional

2.1. Simulación no condicional

La simulación **no condicional** busca construir realizaciones de la función aleatoria que representa la variable regionalizada, pero *sin tomar en cuenta los valores de los datos*. Por lo tanto, aunque reproduce la variabilidad de la variable regionalizada (mismo histograma, mismo variograma...), no interpola los datos.

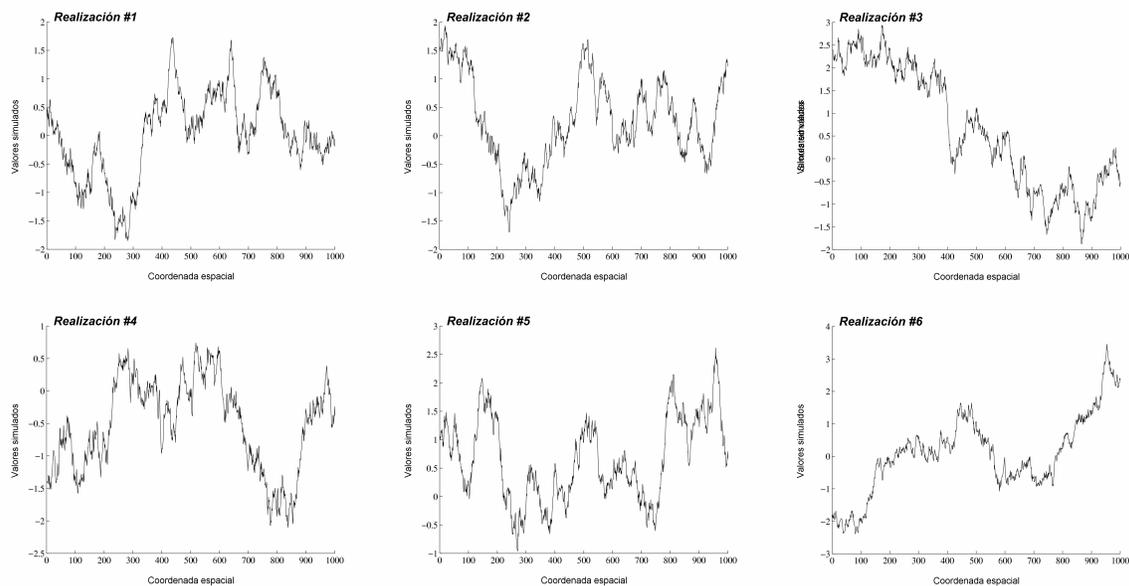


Figura 1. Seis realizaciones no condicionales de una función aleatoria en una dimensión

2.2. Simulación condicional

Para que una simulación sea realista, se desea que restituya los valores medidos en los sitios con datos. Se habla entonces de simulación *condicional*. En el modelo probabilístico, esta restricción adicional se formaliza por la noción de distribución condicionada: se busca construir una función aleatoria con la misma distribución espacial que Z y *conociendo los valores que debe tomar en los sitios con datos*.

Las distribuciones de probabilidad condicional describen la incertidumbre que se tiene localmente sobre los valores de la variable regionalizada, tomando en cuenta los valores de los datos circundantes. Por ejemplo, en un sitio con dato, no hay incertidumbre y se obtiene el valor del dato en todas las realizaciones. Al contrario, lejos de los datos, la distribución condicional se parece a la distribución *a priori* (no condicional), de modo que los datos son “inactivos”. Entre medio, las distribuciones condicionales suelen tener menos dispersión

que las distribuciones *a priori*, traduciendo que los datos aportan algún conocimiento o información sobre el valor que se busca caracterizar.

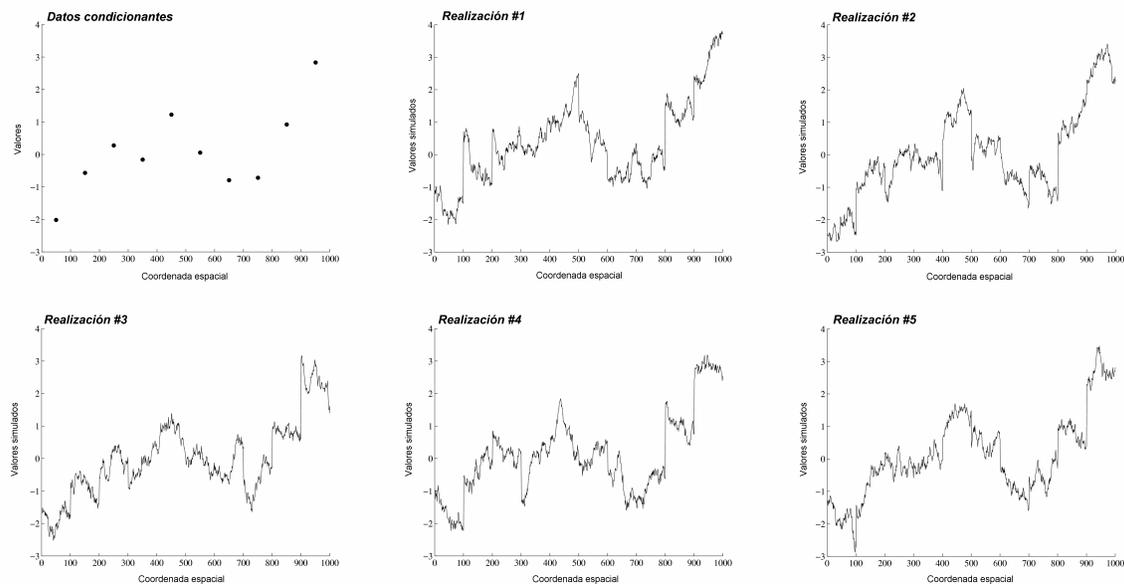


Figura 2. Datos condicionantes y cinco realizaciones condicionales de una función aleatoria en una dimensión

En el transcurso de un estudio práctico, la evaluación de un resultado se efectúa vía simulaciones condicionales. A menudo, las simulaciones no condicionales constituyen una etapa preliminar en la construcción de las simulaciones condicionales. También, cuando los datos disponibles son muy pocos y no permiten especificar convenientemente el modelo variográfico, se utiliza simulaciones no condicionales para estudiar la influencia de los parámetros de este modelo (alcance, comportamiento en el origen, meseta...) en el resultado buscado.

3. Requisitos para realizar una simulación

A diferencia del kriging que sólo requiere conocer la covarianza o el variograma, la simulación geoestadística necesita definir completamente la función aleatoria que se desea simular. Esto es, conocer la *distribución espacial* de esta función aleatoria, es decir, todas sus distribuciones de probabilidad multivariantes (ver capítulo 3 y Figura 3):

- distribución univariable (histograma)
- distribuciones bivariantes
- distribuciones trivariantes, ...

En particular, es insuficiente conocer la media y el variograma, como lo demuestran los ejemplos siguientes (Figuras 4 y 5).

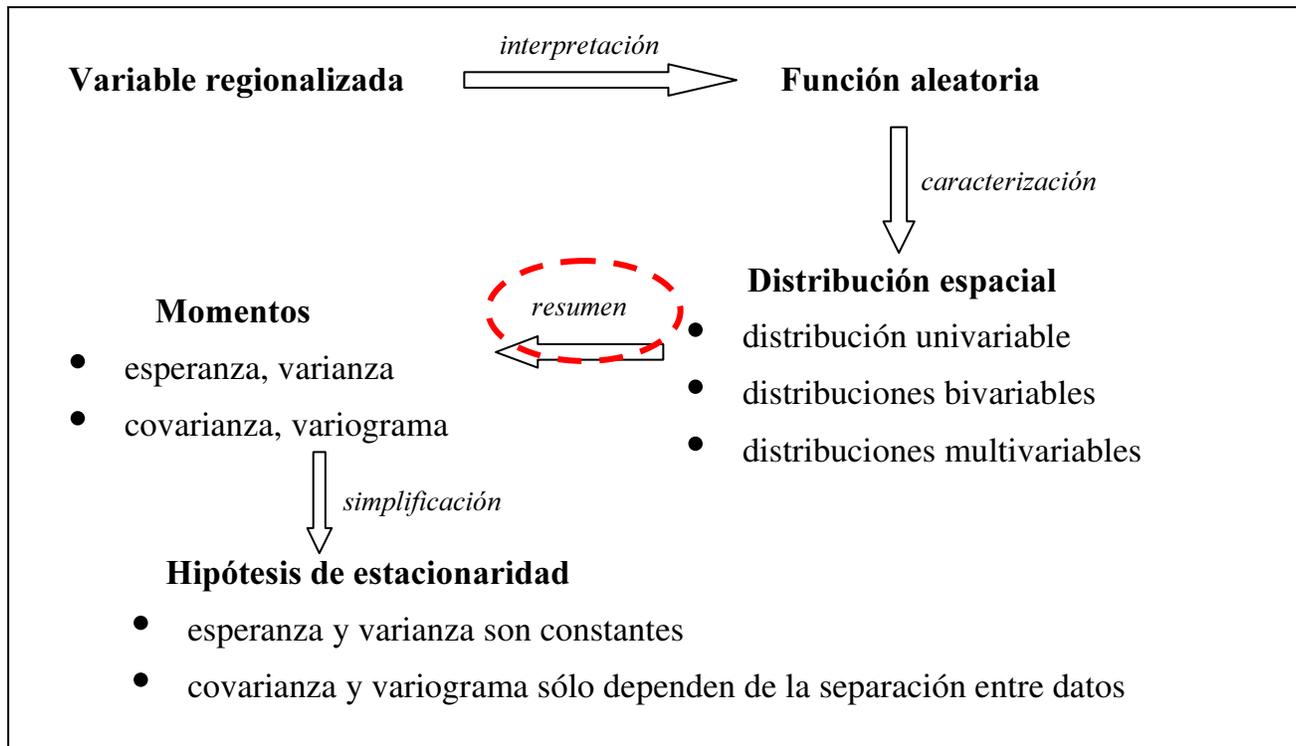


Figura 3. Esquema sintético del modelamiento geoestadístico. Esperanza, varianza, covarianza y variograma sólo proporcionan un resumen de las características de la función aleatoria

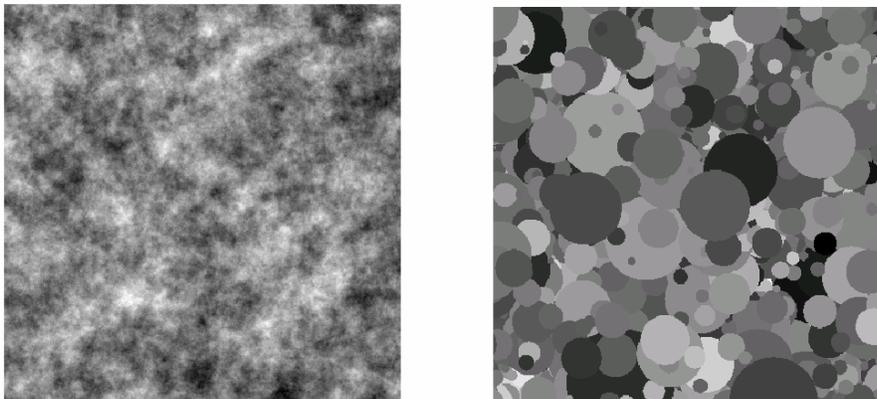


Figura 4. Realizaciones de dos funciones aleatorias con la misma distribución univariable y el mismo variograma

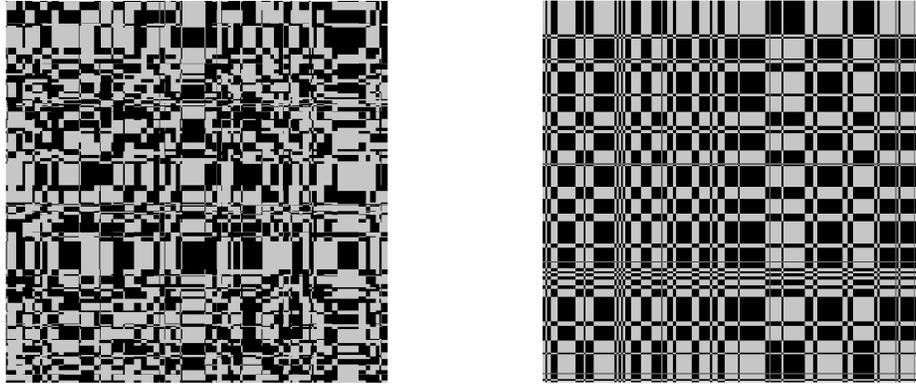


Figura 5. Realizaciones de funciones aleatorias con mismas distribuciones univariadas, bivariadas (luego, mismos variogramas) y trivariadas

4. Ejemplos de funciones aleatorias

Existe una gran diversidad de modelos de función aleatoria que pueden ser simulados por técnicas geoestadísticas. A continuación, mostramos algunos ejemplos, que difieren en la naturaleza de la variable en estudio.

- **Variables continuas**, como las concentraciones de elementos contaminantes, leyes de mineral, pH, conductividad eléctrica, permeabilidad de una roca, carga hidráulica, etc.

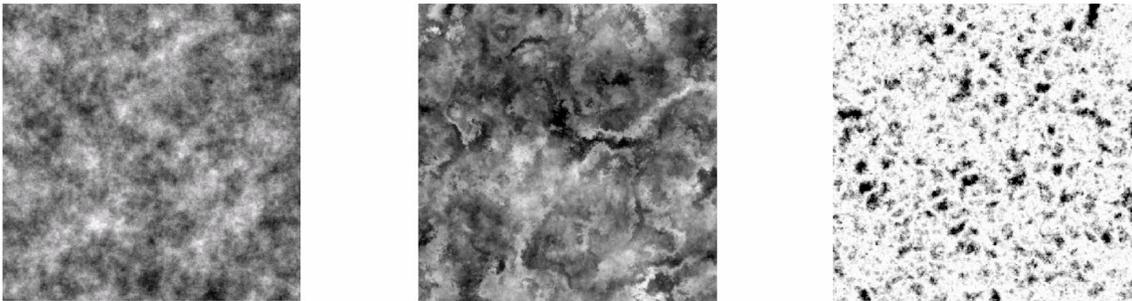


Figura 6. Realizaciones de tres modelos de variables continuas

- **Variables discretas o categóricas**, que permiten codificar un conjunto de dominios que subdividen el espacio (por ejemplo, dominios definidos por tipos de roca) o representar variables con un número limitado de valores, por ejemplo, las variables de conteo cuyos valores son enteros.



Figura 7. Realizaciones de tres modelos de variables categóricas

- **Conjuntos aleatorios**, que pueden ser representados por *indicadores* (variables binarias que valen 1 cuando se está en el conjunto, 0 cuando se está fuera). Pueden representar variables que cuentan con solamente dos categorías, por ejemplo roca permeable versus roca impermeable.

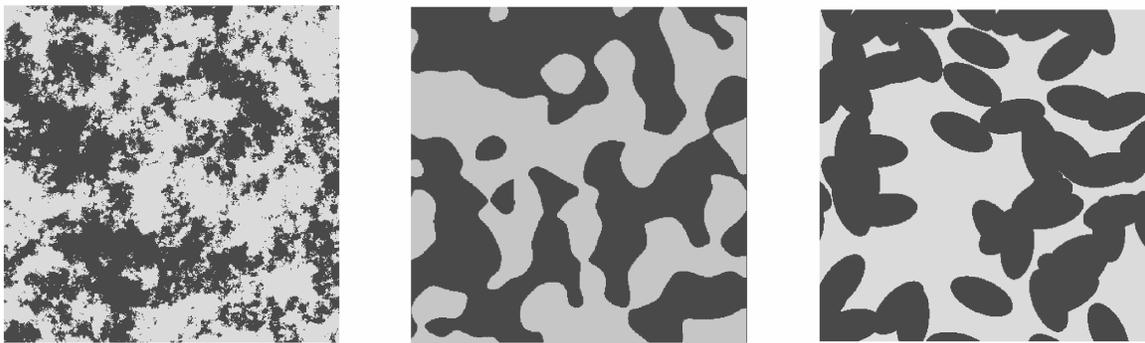


Figura 8. Realizaciones de tres modelos de conjuntos aleatorios

- **Redes de fracturas**, que son casos particulares de conjuntos aleatorios (porciones de plano o de recta, con dimensiones y orientación aleatorias en el espacio)

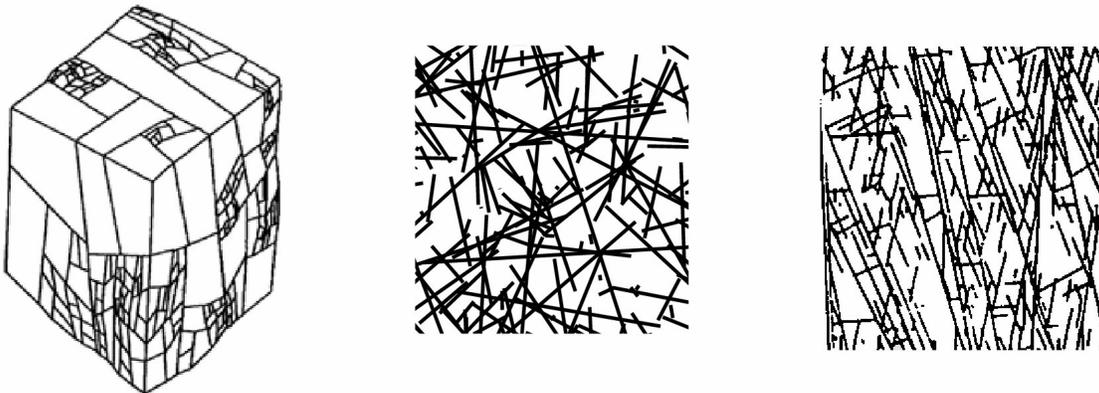


Figura 9. Realizaciones de tres modelos de redes de fracturas

- **Procesos puntuales**, que son otros casos particulares de conjuntos aleatorios. Permiten modelar fenómenos naturales como: distribución de árboles en un bosque, de piedras preciosas en un yacimiento, o de epicentros de temblores. También sirven para construir algunos modelos de funciones aleatorias, de conjuntos aleatorios y de redes de fractura.

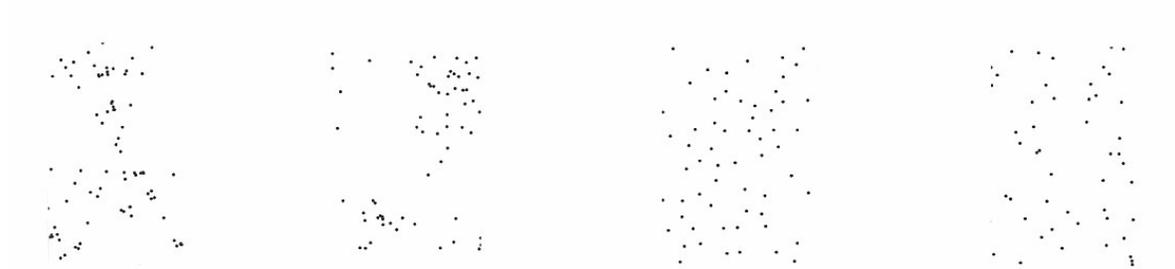


Figura 10. Realizaciones de cuatro modelos de procesos puntuales

5. Aspectos del problema de simulación

Las etapas y desafíos planteados para simular una función aleatoria son los siguientes:

- 1) Definir un modelo de función aleatoria (distribución espacial).
- 2) Inferir los parámetros del modelo (variograma, etc.) y validarlos.
- 3) Proponer algoritmos para construir realizaciones de la función aleatoria definida.
- 4) Condicionar las realizaciones a los datos disponibles.

Estos pasos convierten la simulación geoestadística en un enfoque mucho más complejo que el kriging. En el próximo capítulo, examinaremos un modelo muy sencillo denominado *modelo multi-Gaussiano*, que permite simular variables continuas. La simulación de otros modelos de función aleatoria es un tema que sobrepasa los alcances de este curso.

Capítulo 8. Simulación de variables continuas: modelo multi-Gaussiano

En este capítulo, nos interesaremos por un modelo de función aleatoria cuya definición requiere modelar sólo una distribución univariable (o sea, un histograma) y un variograma. Este modelo se basa en las distribuciones de probabilidad multivariantes Gaussianas (*multi-Gaussianas* o *multinormales*).

1. Transformación Gaussiana (*anamorfosis*)

Es poco frecuente que la variable estudiada pueda ser considerada como Gaussiana: a menudo, la distribución univariable (histograma de los valores medidos) es asimétrico y no es compatible con un modelo Gaussiano. Una transformación – llamada *anamorfosis* – es necesaria para convertirla en una distribución Gaussiana. Gráficamente, la transformación consiste en deformar el histograma de los datos en un histograma Gaussiano estándar, es decir, de media 0 y varianza 1 (Figura 1).

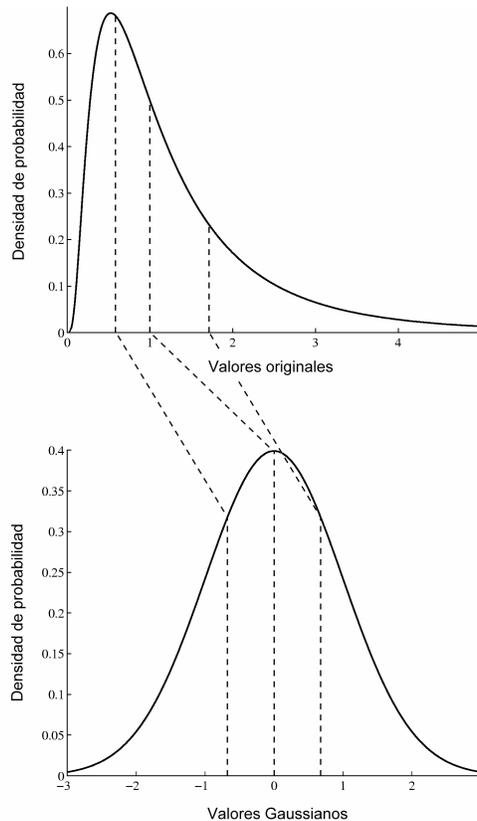


Figura 1. Construcción gráfica de la anamorfosis Gaussiana

A continuación, denotaremos como

- $Z = \{Z(\mathbf{x}), \mathbf{x} \in D\}$ la función aleatoria que representa a la variable original
- $Y = \{Y(\mathbf{x}), \mathbf{x} \in D\}$ la función aleatoria que representa a la variable transformada
- $F(z)$ la función de distribución (histograma acumulado) de Z
- $G(y)$ la función de distribución de Y (Gaussiana de media 0 y varianza 1).

En los histogramas acumulados, la transformación consiste en asociar a cada valor z de la variable original el valor Gaussiano y que tiene la misma frecuencia acumulada, es decir, se plantea que $F(z) = G(y)$ (Figura 2).

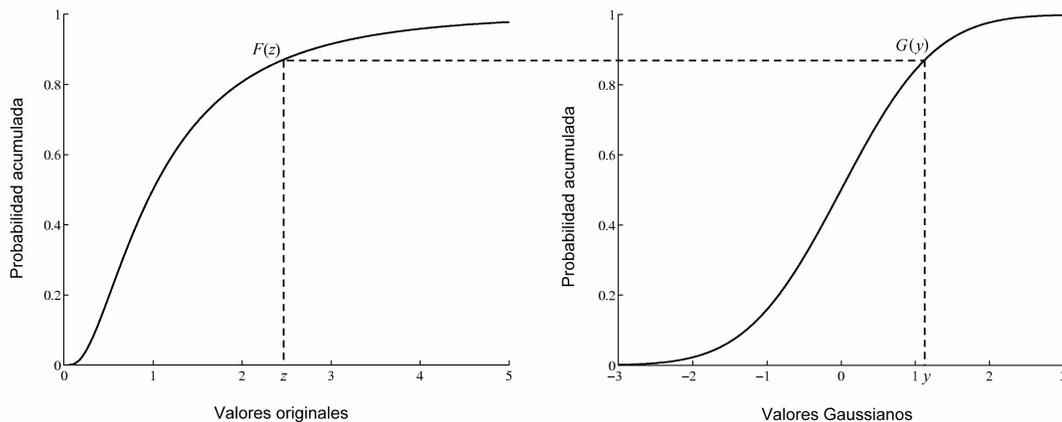


Figura 2. Construcción de la anamorfosis con ayuda de las funciones de distribución F (variable original) y G (variable Gaussiana)

Se denomina *función de anamorfosis Gaussiana* a la función que relaciona los valores Gaussianos con los valores originales (Figura 3). Conforme a lo anterior, esta función se escribe como: $\phi = F^{-1} \circ G$ y se puede plantear:

$$\forall \mathbf{x} \in D, Z(\mathbf{x}) = \phi[Y(\mathbf{x})]$$

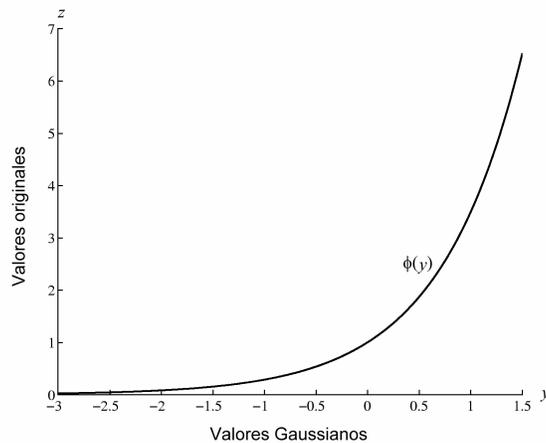


Figura 3. Función de anamorfosis. En este ejemplo (variable Z con una distribución lognormal), se trata de una función exponencial

1.1. Observaciones

- 1) La construcción de la anamorfosis no es posible cuando la función de distribución F no es invertible. Esto es el caso cuando la variable en estudio es discreta o categórica, o cuando los datos presentan una proporción importante de valores iguales. Por ejemplo, si una variable positiva cuenta con 50% de valores nulos – que deberían corresponder al 50% de valores Gaussianos negativos – no se puede asignar de manera unívoca un valor Gaussiano a cada uno de los valores nulos.
- 2) Conocer la función de anamorfosis es equivalente a conocer la función de distribución F (la función de distribución Gaussiana G ha sido tabulada), es decir, el histograma de la variable original Z .

1.2. Complemento: determinación práctica

Para determinar numéricamente la función de anamorfosis, se empieza por ordenar los datos en orden creciente:

$$z_1 < z_2 < \dots < z_n$$

con probabilidades de ocurrencia $p_1, p_2 \dots p_n$. En primera aproximación, se puede identificar estas probabilidades con los ponderadores de desagrupamiento (normalizados de modo que su suma sea igual a 1):

$$\forall i \in \{1, \dots, n\}, p_i = \text{ponderador de desagrupamiento del } i\text{-ésimo dato.}$$

Los valores de los datos están asociados a las frecuencias acumuladas empíricas:

$$\begin{cases} \hat{F}(z_1) = \text{Prob}(Z < z_1) = 0 \\ \forall i \in \{2, \dots, n\}, \hat{F}(z_i) = \text{Prob}(Z < z_i) = \sum_{j=1}^{i-1} p_j \end{cases}$$

Se asocia a cada valor z_i el valor Gaussiano y_i de misma frecuencia acumulada, es decir tal que (Figura 4)

$$\hat{F}(z_i) = G(y_i),$$

lo que define la función de anamorfosis empírica, denotada como $\hat{\phi}$ (Figura 5). Si Y es una variable Gaussiana estándar, entonces $\hat{\phi}(Y)$ tiene exactamente la distribución empírica de los datos originales.

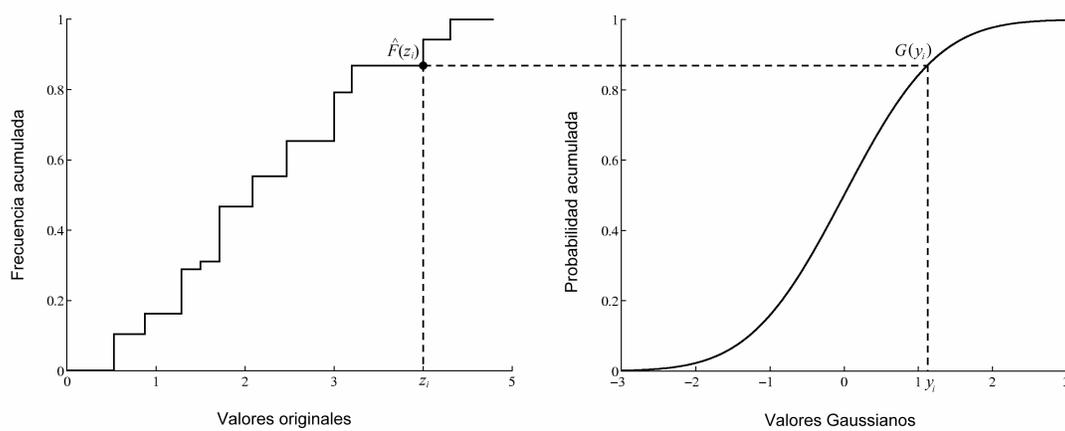


Figura 4. Determinación de la anamorfosis empírica

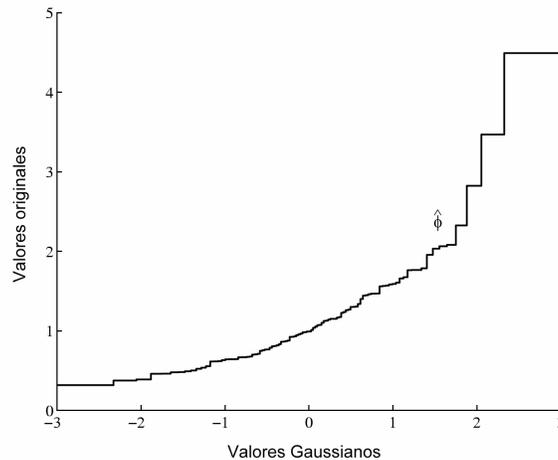


Figura 5. Función de anamorfosis empírica

Varias razones justifican la necesidad de efectuar un suavizamiento de $\hat{\phi}$:

- Obtener una función de anamorfosis invertible, que permite asociar un valor Gaussiano a cualquier valor de la variable Z y vice-versa (la anamorfosis empírica no es invertible, pues es una función escaloneada).
- Suavizar la anamorfosis empírica equivale a suavizar el histograma experimental de Z , luego permite modelar su distribución univariable de manera más aceptable que por su distribución empírica.
- El valor Gaussiano asociado con el dato mínimo es infinito: $y_1 = -\infty$, pues $G(y_1) = F(z_1) = 0$. Esta situación desaparece después del suavizamiento de la función de anamorfosis.

Existen varios métodos de suavizamiento de la anamorfosis empírica, en particular con ayuda de polinomios (Figura 6).

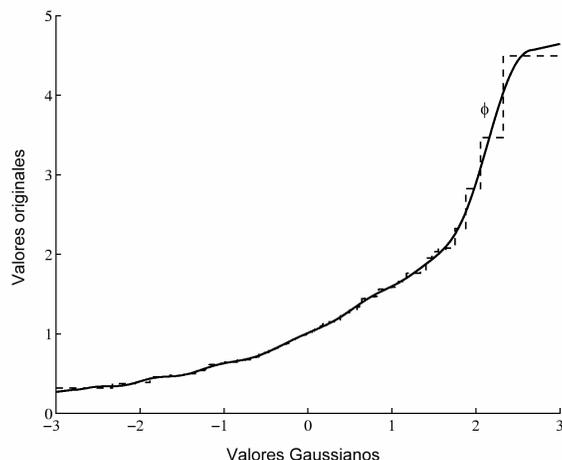


Figura 6. Anamorfosis empírica (línea punteada) y modelada (línea continua)

2. Modelo multi-Gaussiano

2.1. Hipótesis

Se supone que el conjunto de valores Gaussianos tiene una distribución multivariable Gaussiana (multi-Gaussiana), es decir, la densidad de probabilidad de los valores ubicados en los sitios $\{x_1, \dots, x_n\}$ es

$$g(y_1, \dots, y_n) = \frac{1}{(\sqrt{2\pi})^n \sqrt{\det(\mathbf{C})}} \exp\left\{-\frac{1}{2} \mathbf{y} \mathbf{C}^{-1} \mathbf{y}'\right\},$$

donde \mathbf{y} es el vector (y_1, \dots, y_n) , mientras que \mathbf{C} representa la matriz de varianza-covarianza de los datos Gaussianos. Se ve que, en este caso muy particular, la función de covarianza (o, en forma equivalente, el variograma) permite obtener las distribuciones de probabilidad multivariadas, es decir, caracteriza completamente la función aleatoria Gaussiana. Se trata, por lo tanto, de un modelo extremadamente sencillo, puesto que bastará con realizar el análisis variográfico de los datos Gaussianos para determinar el modelo.

2.2. Validación

Por construcción, el histograma de los datos transformados es Gaussiano, por lo cual su distribución univariable es consistente con el modelo. Sin embargo, hace falta verificar que las distribuciones de orden superior (bivariadas, trivariadas...) sean también compatibles

con la hipótesis multi-Gaussiana. Desde un punto de vista teórico, tal verificación es muy exigente y difícil de confirmar con los datos experimentales disponibles. Por esta razón, en la práctica, uno se contenta con validar que la hipótesis multi-Gaussiana se cumple a nivel de las distribuciones bivariantes (es decir, aquellas que involucran a dos valores Gaussianos a la vez).

Primer test: nubes de correlación diferida

Consideremos un par de valores Gaussianos ($Y(\mathbf{x}), Y(\mathbf{x} + \mathbf{h})$) en dos sitios separados por un vector \mathbf{h} . La densidad de probabilidad conjunta de estos valores es:

$$g(y_1, y_2) = \frac{1}{2\pi\sqrt{1-\rho(\mathbf{h})^2}} \exp\left\{-\frac{y_1^2 - 2\rho(\mathbf{h})y_1y_2 + y_2^2}{2(1-\rho(\mathbf{h})^2)}\right\}$$

Las curvas de isodensidad $g(y_1, y_2) = \text{constante}$ son elipses concéntricas:

$$y_1^2 - 2\rho(\mathbf{h})y_1y_2 + y_2^2 = \text{constante}.$$

En la práctica, esto significa que la nube de correlación diferida calculada con los datos Gaussianos para el vector \mathbf{h} , debe tener una forma elíptica. Además, cuando la norma de \mathbf{h} es muy grande, la nube de correlación diferida se vuelve circular (señal de que ya no existe correlación entre $Y(\mathbf{x})$ e $Y(\mathbf{x} + \mathbf{h})$ para las distancias grandes). Al contrario, cuando $|\mathbf{h}|$ tiende a 0, la nube se restringe en torno a la primera bisectriz, ya que $Y(\mathbf{x} + \mathbf{h})$ está cada vez más correlacionado con $Y(\mathbf{x})$ (Figura 7).

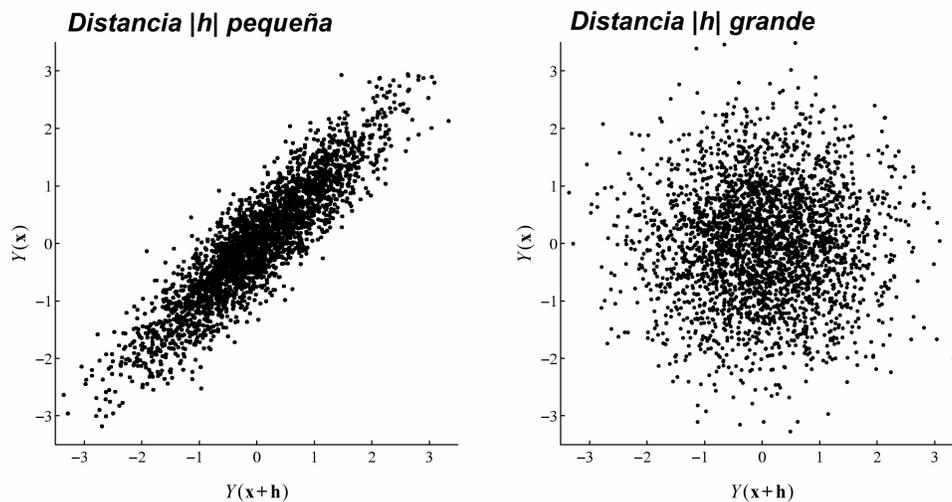


Figura 7. Nubes de correlación diferida en forma de elipses

Segundo test: comparación entre variograma y madograma

El *madograma* (o *variograma de orden 1*) de la variable Gaussiana para un vector \mathbf{h} se define como:

$$\gamma_1(\mathbf{h}) = \frac{1}{2} E\{|Y(\mathbf{x}+\mathbf{h}) - Y(\mathbf{x})|\}$$

Dado que la distribución multi-Gaussiana está enteramente caracterizada por la función de covarianza (o por el variograma $\gamma(\mathbf{h})$), se puede expresar este madograma en función del variograma. Se tiene la siguiente relación:

$$\frac{\sqrt{\gamma(\mathbf{h})}}{\gamma_1(\mathbf{h})} = \sqrt{\pi} \approx 1.77 \quad (\text{independiente de } \mathbf{h}).$$

Basta con verificar esta relación con los variograma y madograma experimentales de los datos Gaussianos.

Tercer test: variogramas de indicadores

Un indicador es una función binaria definida por referencia a un umbral y :

$$I_Y(\mathbf{x}; y) = \begin{cases} 1 & \text{si } Y(\mathbf{x}) < y \\ 0 & \text{en caso contrario} \end{cases}$$

Al igual que para el test anterior, existe una relación entre el variograma de la variable Gaussiana y el variograma $\gamma_{I,y}(\mathbf{h})$ de la variable indicador:

$$\gamma_{I,y}(\mathbf{h}) = G(y)[1 - G(y)] - \frac{1}{2\pi} \int_0^{\arcsen[1-\gamma(\mathbf{h})]} \exp\left[-\frac{y^2}{1+\sen\theta}\right] d\theta$$

donde $G(\cdot)$ es la función de distribución Gaussiana estándar.

El test propuesto consiste en los siguientes pasos:

- 1) Modelar el variograma $\gamma(\mathbf{h})$ de los datos Gaussianos.
- 2) Deducir el variograma $\gamma_{I,y}(\mathbf{h})$ del indicador asociado a un determinado umbral y .
- 3) Codificar los datos Gaussianos en indicador (0 ó 1) según si superan o no el umbral y .
- 4) Calcular el variograma experimental de los datos de indicador.
- 5) Comparar este variograma experimental con la expresión teórica obtenida en el paso 2).

- 6) Repetir el procedimiento (pasos 2 a 5) para varios valores del umbral y , por ejemplo, para los cuartiles de la distribución.
- 7) Concluir si los variogramas teóricos de indicador se ajustan razonablemente bien a los variogramas experimentales.

2.3. Propiedades características del modelo multi-Gaussiano

- 1) Si el umbral y tiende a infinito, el variograma $\gamma_{I,y}(\mathbf{h})$ tiende a su meseta $G(y) [1 - G(y)]$ cualesquiera \mathbf{h} , es decir, los variogramas de indicadores asociados a umbrales extremos se vuelven pepíticos. En otras palabras, la ocurrencia de valores extremos es puramente aleatoria, de modo que el modelo multi-Gaussiano es inapropiado cuando los valores extremos están espacialmente correlacionados, en particular cuando están agrupados en ciertas zonas del campo (Figura 8).

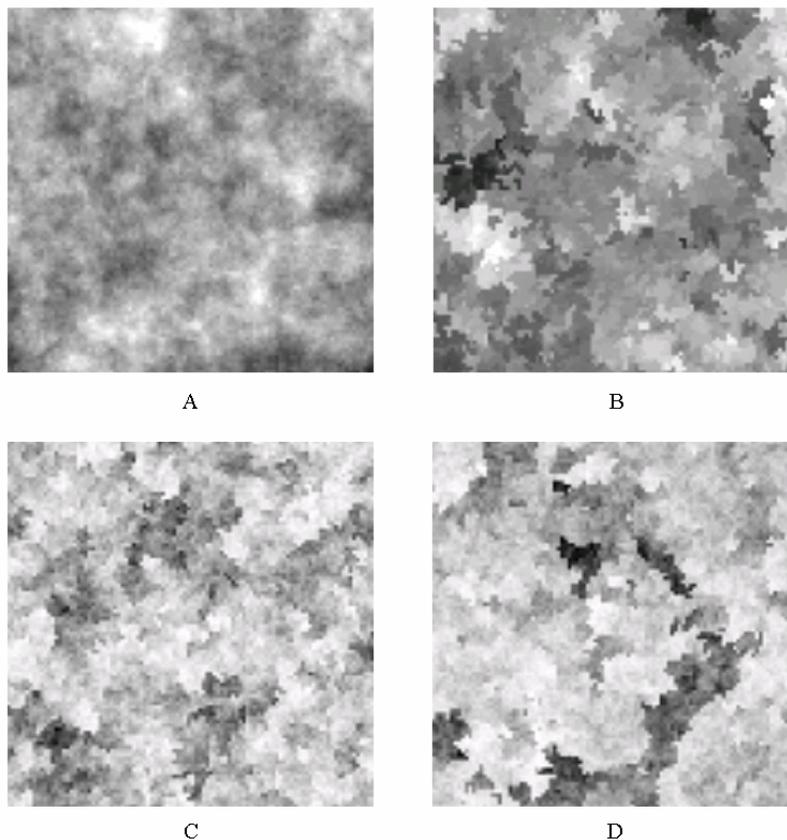


Figura 8. Realizaciones de cuatro modelos de función aleatoria. Sólo el modelo en la lámina A es compatible con la distribución multi-Gaussiana, pues no presenta agrupación espacial de los valores extremos

- 2) La expresión del variograma de indicador $\gamma_{I,y}$ es invariante cuando se cambia y en $-y$: la continuidad espacial de los indicadores es simétrica con respecto al umbral mediano (0). Esto se traduce por una igualdad entre los variogramas de indicadores asociados con cuantiles simétricos con respecto a la mediana. En términos más intuitivos, las características (anisotropía, continuidad a pequeña escala...) que se manifiestan para los valores bajos también caracterizan los valores altos.

3. Simulación condicional

Una vez que se ha comprobado la validez de la hipótesis multi-Gaussiana, el modelo queda enteramente caracterizado por la función de anamorfosis y por el variograma de los datos Gaussianos. Falta ahora definir un algoritmo para poder construir realizaciones de este modelo y condicionarlas a los datos disponibles. Entre los algoritmos que responden a este problema, el más sencillo es el llamado “método secuencial Gaussiano”.

3.1. Algoritmo secuencial

Supongamos que se busca simular una función aleatoria multi-Gaussiana Y de media 0 y variograma $\gamma(\mathbf{h})$ en los sitios $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ del espacio, condicionada a los datos Gaussianos disponibles en los sitios $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$. El algoritmo secuencial procede de la siguiente manera. Para cada sitio \mathbf{u}_i ($i = 1, \dots, m$):

- 1) Realizar el kriging de $Y(\mathbf{u}_i)$ a partir de los datos condicionantes $\{Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)\}$ y de los valores previamente simulados $\{Y(\mathbf{u}_1), \dots, Y(\mathbf{u}_{i-1})\}$. Como resultado, se obtiene un valor estimado $Y^*(\mathbf{u}_i)$ y una desviación estándar $\sigma^*(\mathbf{u}_i)$.
- 2) Simular un valor Gaussiano, cuya media es igual a $Y^*(\mathbf{u}_i)$ y cuya desviación estándar es igual a $\sigma^*(\mathbf{u}_i)$:

$$Y(\mathbf{u}_i) = Y^*(\mathbf{u}_i) + \sigma^*(\mathbf{u}_i)N_i,$$

donde N_i es una variable aleatoria Gaussiana de media 0 y varianza 1, independiente de N_1, \dots, N_{i-1} y de los datos originales $Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)$.

Así pues, en la i -ésima etapa, se simula el valor Gaussiano en el sitio \mathbf{u}_i y se agrega el valor simulado a los datos condicionantes para simular los sitios siguientes, de donde viene el nombre “secuencial”. En teoría, se requiere usar un **kriging simple** (de media conocida = media de la variable Gaussiana = 0).

Las ventajas del algoritmo secuencial radican en su sencillez y facilidad de ejecución, así como en el hecho de que proporciona directamente simulaciones condicionales a datos. En contrapartida, su principal desventaja es su lentitud, debido a que el sistema de kriging se vuelve cada vez más grande a medida que se desarrolla la simulación. En la práctica, se debe restringir los datos condicionantes (datos iniciales + valores previamente simulados) a

los más cercanos del sitio a simular, es decir, se utiliza una *vecindad móvil* en lugar de la vecindad única requerida en teoría. A su vez, esta restricción introduce aproximaciones / imprecisiones en los valores simulados, lo que requiere ciertos “trucos” para minimizar sus efectos:

- Uso de *grillas múltiples* (cuando los sitios a simular están ubicados en una grilla): se hace la simulación primero en una malla amplia, luego en mallas más pequeñas. Por ejemplo, se simula uno de cada 10 sitios en cada dirección de la grilla, luego uno de cada 5 sitios y finalmente los sitios restantes. La idea es que la simulación en mallas amplias puede hacerse con una vecindad única (por ende, sin aproximación) puesto que estas mallas contienen pocos sitios, luego la simulación en las mallas pequeñas se apoyará en estos valores “bien” simulados.
- *Aleatorización de la secuencia de simulación*: el orden según el cual se simula los distintos sitios es elegido al azar. Esto se debe a que el uso de una secuencia regular (por ejemplo, simular los nodos de una grilla en 2D fila por fila) suele provocar la aparición de artefactos cuando se utiliza una vecindad móvil. Para cada realización que se quiere construir, es recomendable definir una secuencia distinta.

A modo de ejemplo, la Figura 9 muestra los variogramas experimentales calculados en 100 realizaciones de una función aleatoria Gaussiana de variograma esférico (alcance 50) + pepita, realizadas con el algoritmo secuencial en un dominio de 400×400 nodos. No se han considerado datos iniciales, por lo que las realizaciones obtenidas no son condicionales. La lámina de la izquierda corresponde a una vecindad de kriging que contiene hasta 20 valores previamente simulados, mientras que la lámina de la derecha corresponde a una vecindad con 100 valores previamente simulados. En ambos casos existe un sesgo en la reproducción del variograma, ya que el promedio de los variogramas experimentales no coincide con el modelo teórico de variograma, aunque el sesgo disminuye notablemente cuando se utiliza la vecindad más exigente.

Es interesante notar que siempre existe una discrepancia entre el variograma simulado (o sea, el variograma experimental calculado sobre una realización de la función aleatoria) y el modelo de variograma. Esta discrepancia, llamada *fluctuación estadística* o *fluctuación ergódica*, se debe a que la realización no posee un campo infinito. La teoría indica que la fluctuación esperada es pequeña cerca del origen y aumenta con la distancia (esto explica la falta de robustez del variograma experimental para grandes distancias, ver capítulo 4).

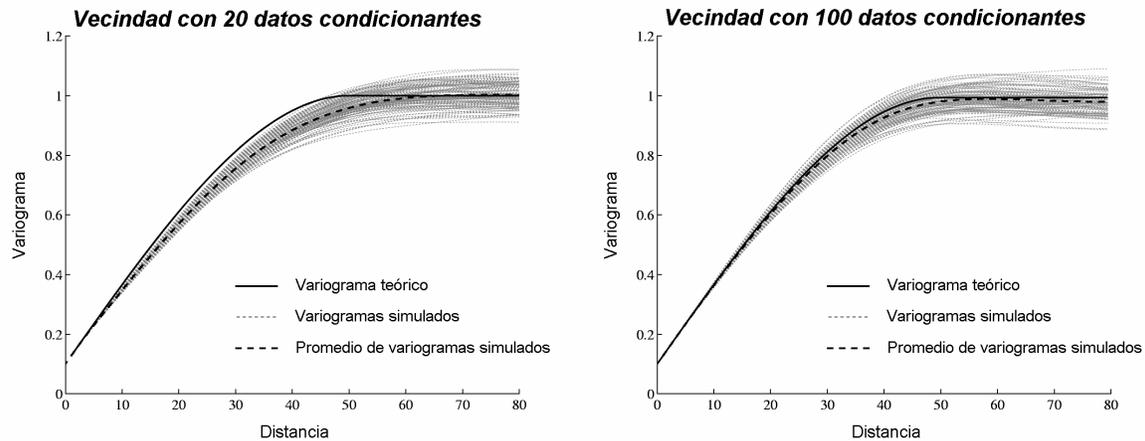


Figura 9. Sesgo en la reproducción del variograma producido por el algoritmo secuencial implementado con una vecindad móvil

3.2. Otros algoritmos

Existen numerosos algoritmos alternativos al método secuencial, por ejemplo:

- Método de descomposición matricial
- Métodos de convolución (medias móviles, métodos auto-regresivos)
- Método espectral discreto
- Método espectral continuo
- Método de las bandas rotantes.

Los cuatro últimos algoritmos generan realizaciones no condicionales (es decir, que no reproducen los valores de los datos). El condicionamiento de las realizaciones requiere una etapa adicional basada en un kriging simple. Aunque son matemáticamente más complejos, estos algoritmos son más rápidos que el algoritmo secuencial por dos razones: 1) el sistema de kriging sólo involucra a los datos originales (no a los valores simulados) y 2) se puede condicionar múltiples realizaciones con un solo kriging.

La Figura 10 retoma la situación vista en la Figura 9 (100 realizaciones de una función aleatoria Gaussiana de variograma esférico + pepita en un dominio de 400×400 nodos), pero utilizando el llamado algoritmo de bandas rotantes en lugar del algoritmo secuencial. Se observa que el sesgo en la reproducción del variograma ha desaparecido por completo: en promedio, el variograma experimental de las realizaciones es igual al modelo teórico de variograma.

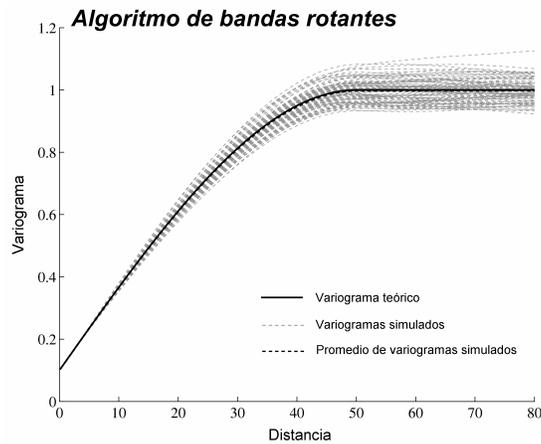


Figura 10. Reproducción del variograma con el algoritmo de bandas rotantes

3.3. Resumen: pasos a seguir para la simulación

Los pasos a seguir para la simulación condicional son finalmente los siguientes:

- 1) Desagrupar los datos originales.
- 2) Transformar estos datos en datos Gaussianos, tomando en cuenta los ponderadores de desagrupamiento.
- 3) Realizar el análisis variográfico de los datos Gaussianos.
- 4) Validar la hipótesis multi-Gaussiana (nubes de correlación diferida, comparación entre variograma y madograma, variogramas de indicadores).
- 5) Simular la función aleatoria Gaussiana
 - Elegir un algoritmo de simulación
 - Construir varias realizaciones
 - Condicionar a los datos Gaussianos disponibles si el algoritmo escogido no lo hace directamente.
- 6) Transformación Gaussiana inversa, para volver a la variable original.
- 7) Procesar los resultados.

4. Aplicación a los datos de contaminación de suelo

Aplicaremos ahora los pasos anteriores para simular las concentraciones de cobalto en la zona contaminada.

4.1. Anamorfosis Gaussiana

El primer paso para aplicar el modelo multi-Gaussiano corresponde a la transformación (anamorfosis) Gaussiana de los datos de concentración de cobalto. La anamorfosis empírica se realiza al deformar el histograma desagrupado de los datos de cobalto en un histograma Gaussiano; sobre esta anamorfosis empírica, se ajusta una función polinomial para obtener el modelo de anamorfosis. Ahora, como el polinomio diverge para los valores extremos de la Gaussiana, este modelo ha sido corregido para dar valores entre 0 y 22 ppm, que serán los valores extremos considerados para la concentración de cobalto en toda la zona (Figura 11). Si bien el histograma de los datos transformados no es perfectamente Gaussiano, tiene una forma muy similar a la campana de Gauss (Figura 12), por lo que se puede suponer sin problema que los datos transformados sí tienen una distribución Gaussiana.

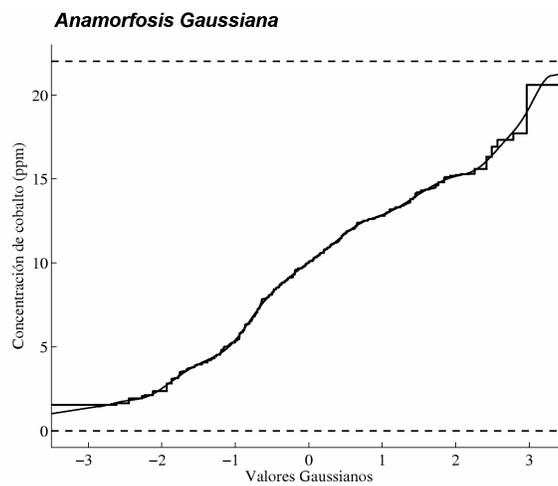


Figura 11. Anamorfosis empírica (línea escalonada) y modelada (línea continua) para las concentraciones de cobalto. Las líneas punteadas indican los extremos considerados para las concentraciones de cobalto (0 y 22 ppm)

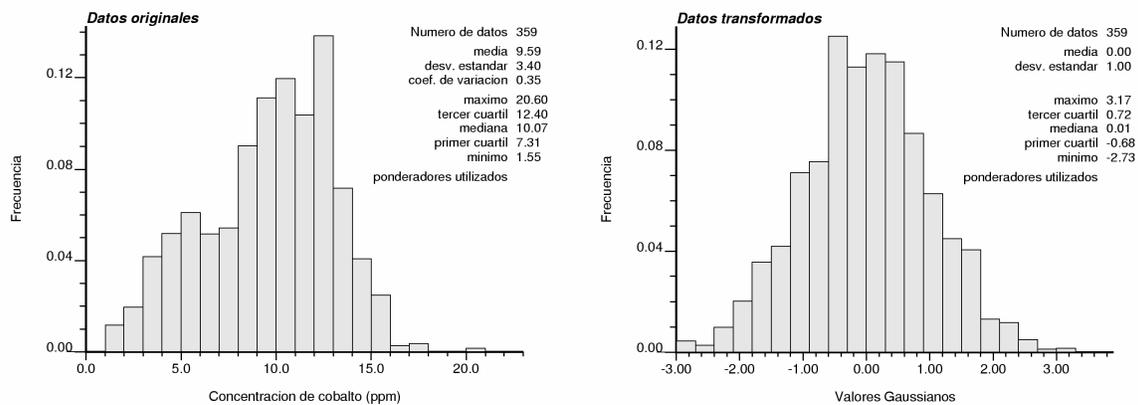


Figura 12. Histograma de los datos originales (izquierda) y de los datos transformados (derecha)

4.2. Validación de la hipótesis multi-Gaussiana

A continuación, verificaremos si los datos Gaussianos obtenidos son compatibles con una distribución multi-Gaussiana. Un primer método consiste en dibujar algunas nubes de correlación diferida y comprobar (visualmente) que tienen una forma de elipse (Figura 13).

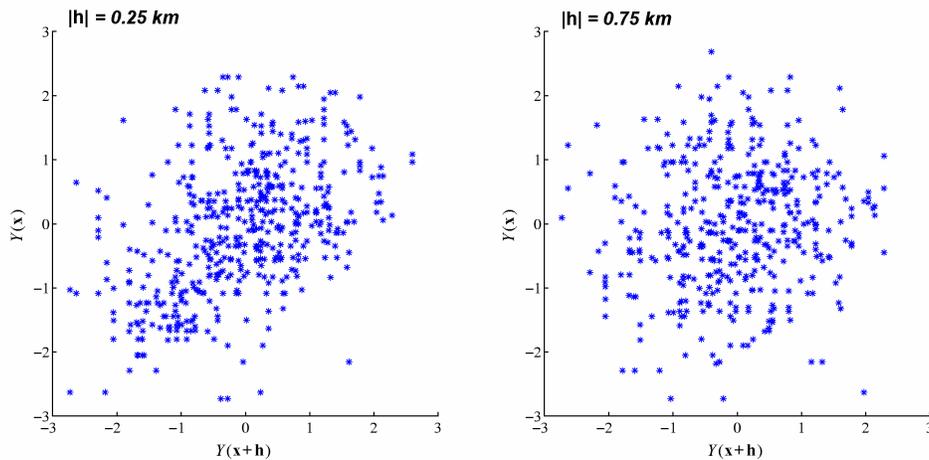


Figura 13. Nubes de correlación diferida para los datos Gaussianos

Un segundo test consiste en verificar que la raíz cuadrada del variograma dividida por el madograma es constante e igual a $\sqrt{\pi}$, o sea, 1.77. El test fue realizado con los siguientes parámetros de cálculo: paso = 0.25 km, tolerancia en el paso = 0.125 km, tolerancia angular = 90° (omnidireccional), número de pasos = 14. Los resultados (Figura 14) indican que se cumple relativamente bien con el valor esperado, salvo quizás para el primer punto. Ahora, este punto fue calculado con sólo 356 pares de datos y podría no ser confiable, mientras que los puntos siguientes se basan en 2000 a 5000 pares de datos.

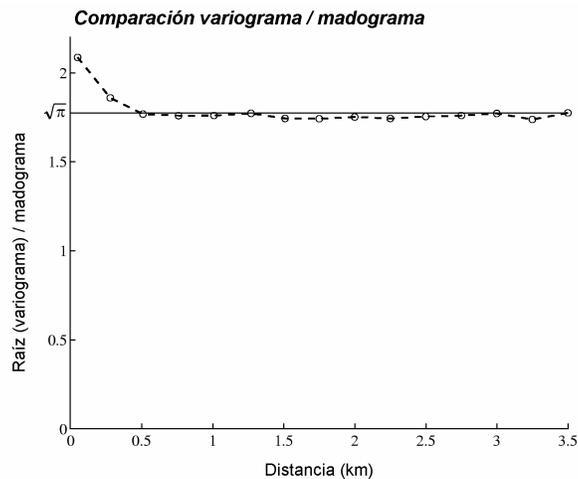


Figura 14. Razón entre la raíz cuadrada del variograma y el madograma de los datos Gaussianos, en función de la distancia

4.3. Análisis variográfico de los datos Gaussianos

Si bien los tests previos (nubes de correlación diferida; comparación entre variograma y madograma) han sido realizados en forma omnidireccional, es decir, sin tomar en cuenta la orientación del vector de separación \mathbf{h} en el espacio, el variograma de los datos Gaussianos puede no ser isótropo. De hecho, un análisis del mapa variográfico (Figura 15, izquierda) indica que este variograma presenta una anisotropía similar a aquella del variograma de los datos originales, con direcciones principales N60°E y N30°W (ver capítulo 4). Para calcular el variograma experimental, se elige los siguientes parámetros:

- paso = 0.35 km
- tolerancia en la distancia = 0.175 km
- tolerancia angular = 20°
- ancho de banda = 1km.

En cuanto al ajuste, se propone el siguiente modelo (Figura 15, derecha):

$$\gamma(\mathbf{h}) = 0.3 \text{pepa} + 0.2 \text{esf}(0.5\text{km}, 0.5\text{km}) + 0.5 \text{esf}(2\text{km}, 1.5\text{km}) + 0.2 \text{esf}(\infty, 1.5\text{km}),$$

en donde los alcances corresponden a las direcciones N60°E y N30°W, respectivamente.

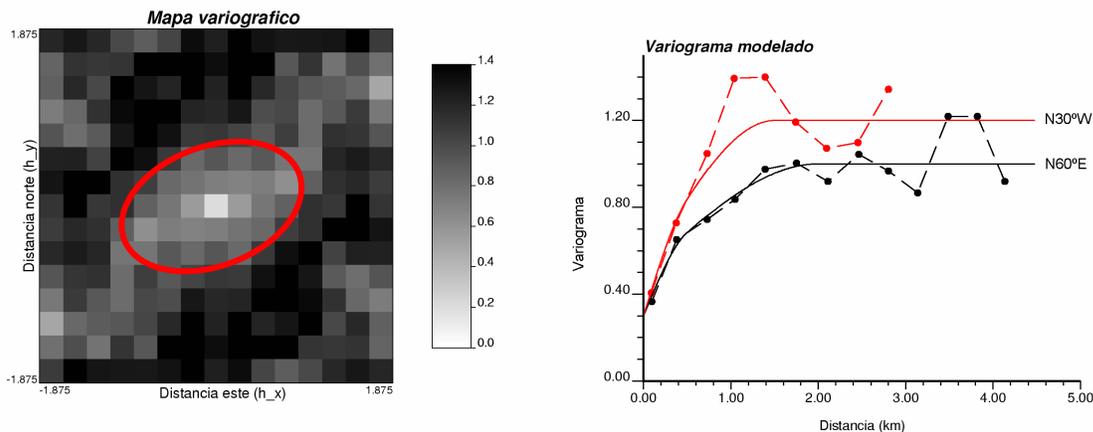


Figura 15. Mapa variográfico (izquierda) y variograma a lo largo de las direcciones de anisotropía (N60°E y N30°W) (derecha) para los datos Gaussianos.

4.4. Simulación condicional

Procedemos ahora a simular las concentraciones de cobalto en una grilla densa (con una malla de 2m × 2m) que cubre la zona de estudio. Se construye 100 realizaciones con el método de las bandas rotantes, utilizando la vecindad de kriging definida en el capítulo 6 para condicionar estas realizaciones a los 359 datos disponibles. Luego, se rebloquea cada realización al soporte de las unidades de remediación (20m × 20m), promediando las 100 concentraciones simuladas en cada unidad. La Figura 16 presenta cuatro realizaciones de las concentraciones al soporte de las unidades de remediación.

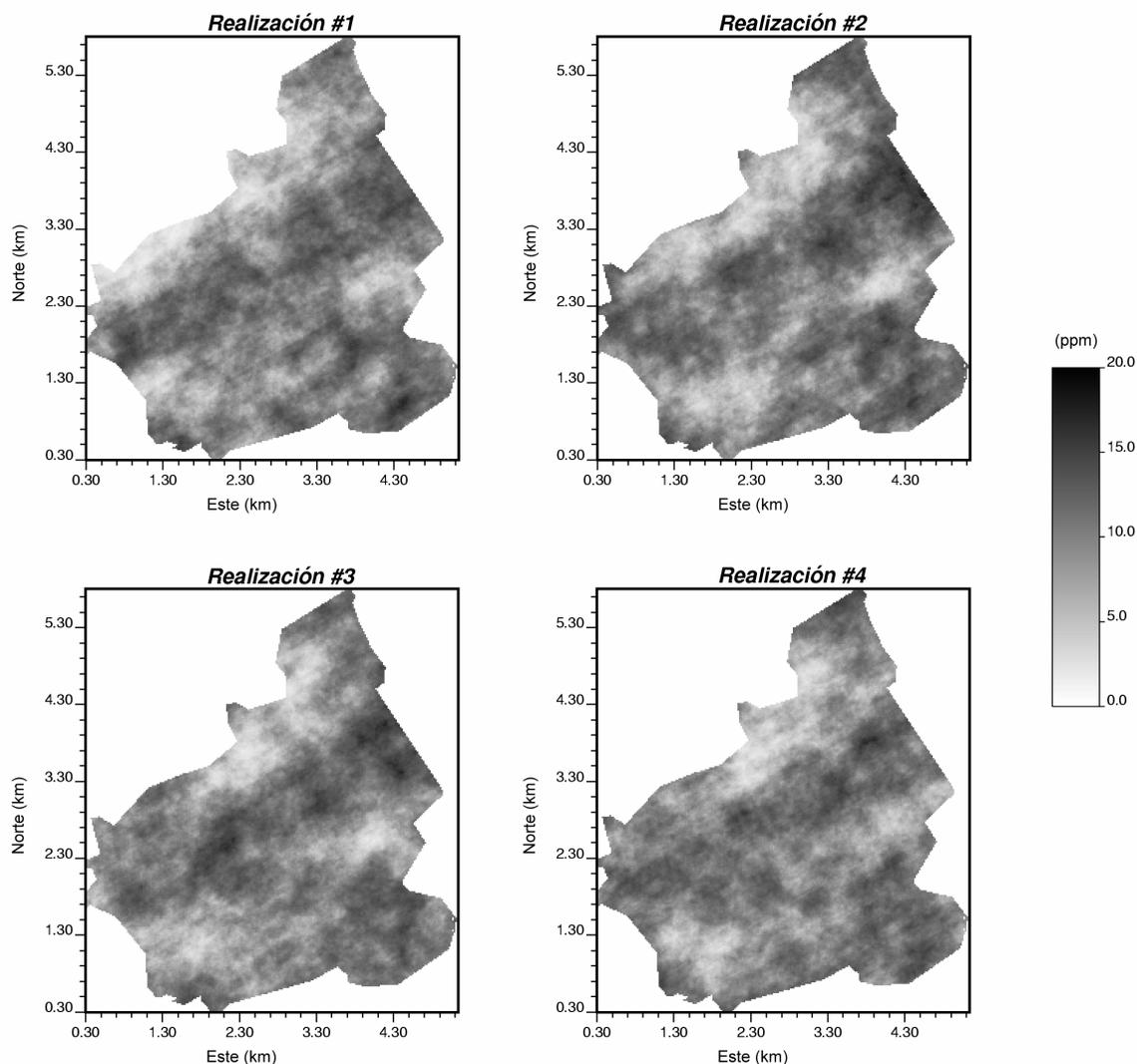


Figura 16. Cuatro realizaciones condicionales de la concentración promedio de cobalto en unidades de remediación de 20m × 20m.

4.5. Procesamiento de los resultados

A partir de las realizaciones generadas, se puede buscar estimar las concentraciones de cobalto, por ejemplo, tomando el promedio de las realizaciones (Figura 17, izquierda). Este estimador entrega el valor esperado de la concentración de cobalto y se asemeja mucho al estimador de kriging. También se puede caracterizar la incertidumbre en la concentración de cobalto al calcular la desviación estándar de las realizaciones (Figura 17, derecha). Se observa que, contrariamente a la desviación estándar de kriging, esta desviación no depende solamente de la configuración geométrica de los datos, sino que también de sus valores, por lo cual constituye una medida de incertidumbre más completa.

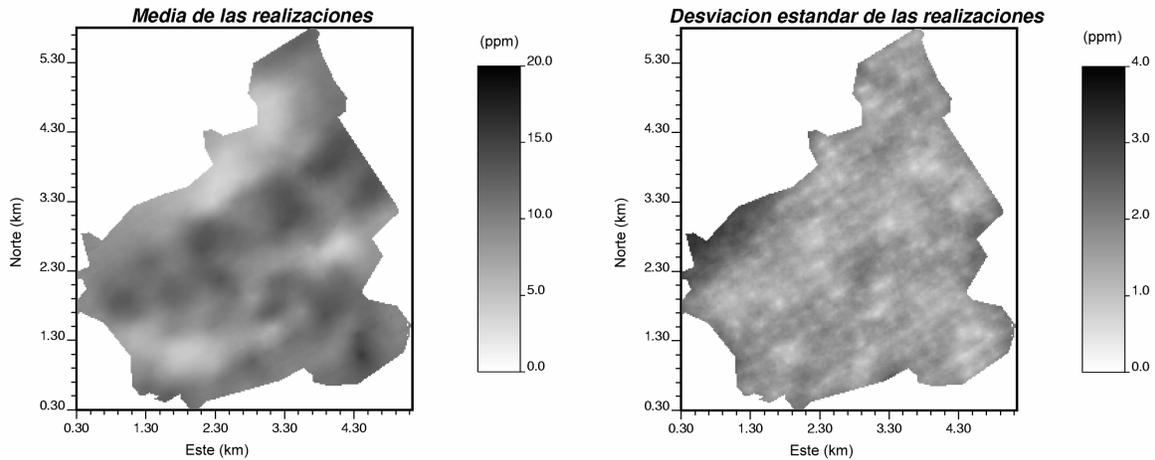


Figura 17. Promedio y desviación estándar de 100 realizaciones condicionales de la concentración promedio de cobalto en unidades de remediación de $20\text{m} \times 20\text{m}$.

También se puede estimar la probabilidad de que la concentración de cobalto supere un determinado umbral de toxicidad (por ejemplo, 12 ppm), al determinar la frecuencia con la cual las concentraciones simuladas superan este umbral. Obtenemos así un mapa del riesgo de que la concentración de cobalto sea perjudicial a la salud humana (Figura 18, izquierda). Este mapa es una herramienta útil para la toma de decisión. Por ejemplo, uno puede definir tres áreas (Figura 18, derecha): un *área de remediación* que contiene las unidades de mayor riesgo (por ejemplo, aquellas cuya probabilidad de superar el umbral 12 ppm es mayor que 0.5) y requiere tratamiento; un *área segura* que contiene las unidades de menor riesgo (por ejemplo, aquellas cuya probabilidad de superar el umbral 12 ppm es menor que 0.2), tales que se pueden dejar in situ; finalmente, un *área incierta* (unidades con una probabilidad de superar el umbral 12 ppm entre 0.2 y 0.5) para la cual el riesgo es demasiado pequeño para ser considerado en el área de remediación, pero demasiado grande para ser dejado in situ sin mayor información (muestreo adicional).

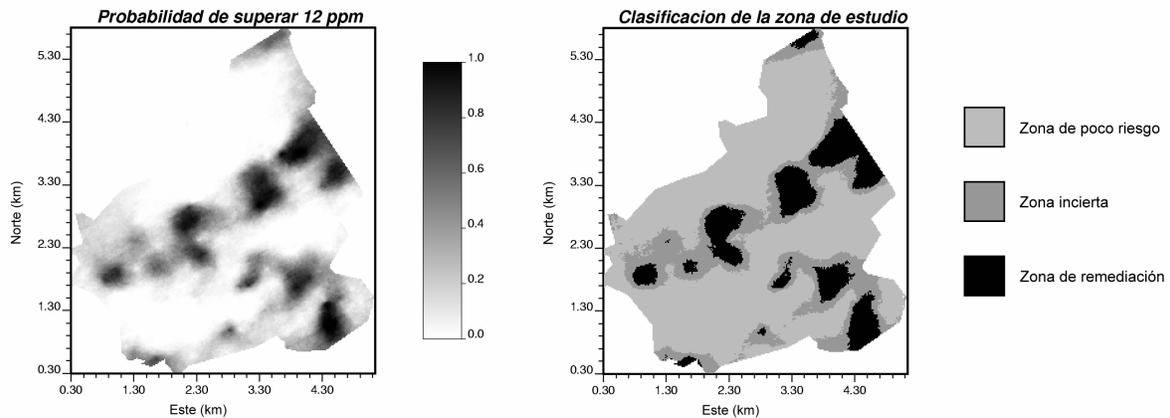


Figura 18. Izquierda: probabilidad de encontrar una concentración de cobalto mayor que 12 ppm. Derecha: clasificación de la zona de estudio en tres áreas.

La simulación condicional también se presta al análisis de riesgo cuando el cambio de soporte (desde los datos hasta las unidades de remediación) no corresponde a un promedio aritmético de las concentraciones contenidas en estas unidades. Por ejemplo, el medio-ambientalista puede decidir que la unidad de remediación sea tratada si la concentración *máxima* de cobalto dentro de esta unidad supera el umbral de 12 ppm. Aunque presente en un soporte pequeño, una concentración alta de cobalto puede ser dañina y requerir medidas de remediación. Este criterio lleva a una clasificación de la zona mucho más conservadora que aquella definida anteriormente (Figura 19).

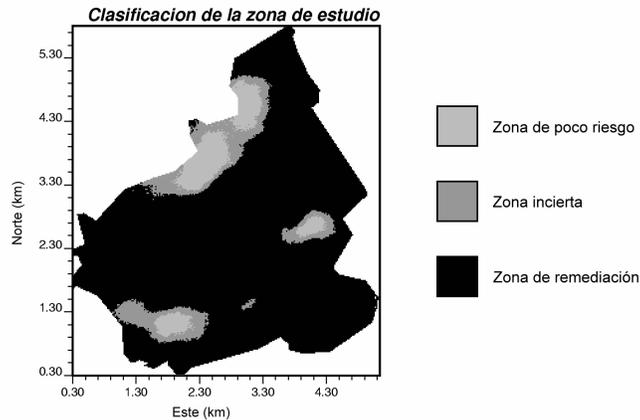


Figura 19. Clasificación de la zona de estudio, considerando la concentración máxima de cobalto en cada unidad de remediación en lugar de su concentración promedio.

Según el criterio de la concentración promedio en las unidades de remediación, sólo el 21.3 % de estas unidades tienen una concentración de cobalto superior a 12 ppm, mientras que este porcentaje sube a 81.3 % al considerar la concentración máxima en cada unidad (Figura 20). Estos porcentajes dan una estimación de la proporción de la zona de estudio que requiere un tratamiento, según se elige uno u otro criterio.

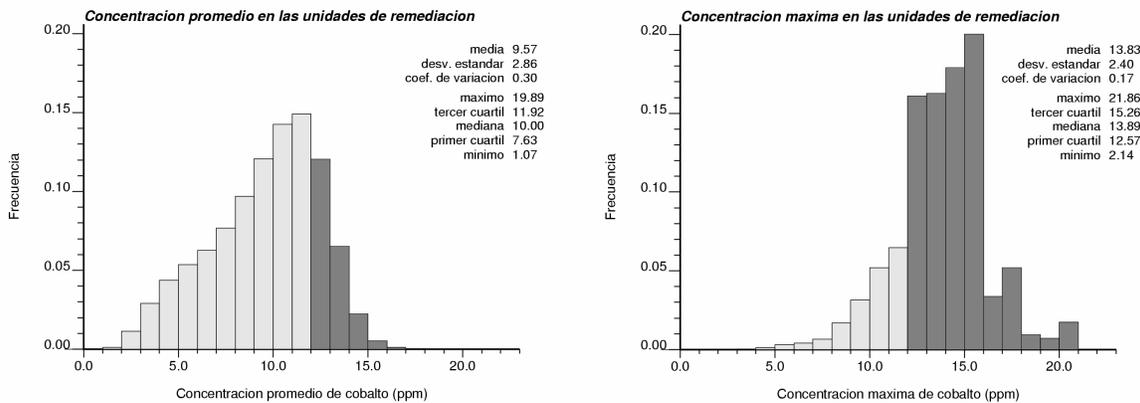


Figura 20. Histogramas de concentraciones de cobalto. Izquierda: concentración promedio en las unidades de remediación; derecha: concentración máxima en estas unidades.

Anexo A. Geoestadística multivariable

Es frecuente que varias variables regionalizadas se refieran a mismo fenómeno; a modo de ejemplos, citemos la evaluación de los recursos minerales en yacimientos polimetálicos, la caracterización de propiedades petrofísicas (porosidad, permeabilidad...) en un acuífero o el análisis de las concentraciones de elementos contaminantes o de parámetros pedológicos medidos sobre muestras de suelo. Se tendrá interés en realizar un estudio *conjunto* de todas las variables regionalizadas, de manera de poder tomar en cuenta los vínculos entre ellas y la información aportada por variables auxiliares sobre la variable de interés principal.

Desde el punto de vista teórico, las dificultades del análisis multivariable no están ni en los conceptos, ni en los métodos matemáticos, sino que más bien en las notaciones, más complejas. Para evitar una proliferación de índices, es conveniente adoptar una escritura vectorial o matricial.

Precisemos el vocabulario y las notaciones empleadas. Se llama *corregionalización* al conjunto de variables regionalizadas, que denotaremos como z_1, \dots, z_N . Estas variables están definidas sobre un mismo dominio acotado de \mathbb{R}^d , llamado *campo* de la corregionalización y denotado D . Se habla de *homotopía* o *isotopía*, cuando todas las variables son medidas en todos los sitios de muestreo, de *heterotopía total*, cuando son medidas sobre conjuntos de sitios disjuntos, y de *heterotopía parcial*, cuando sólo una parte de los sitios de medición son comunes a todas las variables (a menudo, ocurre que el conjunto de sitios de medición de una variable está incluido en el de otra). Finalmente, las funciones aleatorias asociadas a las variables regionalizadas se denotarán con letras mayúsculas, a saber Z_1, \dots, Z_N .

1. Análisis variográfico

Tal como en el marco univariable, se define herramientas variográficas destinadas a analizar la continuidad espacial de las variables. Se supondrá que las funciones aleatorias asociadas son conjuntamente estacionarias, de modo que dichas herramientas variográficas serán funciones del vector de separación entre datos, no de su posición absoluta.

1.1. Funciones de covarianza simple y cruzada

La covarianza *cruzada* entre dos funciones aleatorias Z_i y Z_j para un vector \mathbf{h} se define como:

$$\begin{aligned} C_{ij}(\mathbf{h}) &= \text{cov}\{Z_i(\mathbf{x}+\mathbf{h}), Z_j(\mathbf{x})\} \\ &= E\{Z_i(\mathbf{x}+\mathbf{h}) Z_j(\mathbf{x})\} - E\{Z_i(\mathbf{x}+\mathbf{h})\} \times E\{Z_j(\mathbf{x})\} \end{aligned}$$

Cuando $i = j$, se encuentra la función de covarianza de Z_i , llamada covarianza *simple* o *directa*. También se puede adoptar una presentación matricial, definiendo la matriz $\mathbf{C}(\mathbf{h})$ de funciones de covarianzas simples y cruzadas:

$$\mathbf{C}(\mathbf{h}) = [C_{ij}(\mathbf{h})]_{i,j=1\dots N}.$$

Las covarianzas simples y cruzadas tienen las siguientes propiedades:

- **Simetría:** en general, para $i \neq j$, $C_{ij}(\mathbf{h}) \neq C_{ji}(\mathbf{h})$ y $C_{ij}(\mathbf{h}) \neq C_{ij}(-\mathbf{h})$, pero siempre se tiene $C_{ij}(\mathbf{h}) = C_{ji}(-\mathbf{h})$.
- **Signo y extremos.** La función de covarianza simple $C_{ii}(\mathbf{h})$ tiene su valor máximo en el origen $\mathbf{h} = \mathbf{0}$, donde es igual a la varianza *a priori* de Z_i , luego es positiva. Ahora, una función de covarianza cruzada puede no ser positiva o cambiar de signo. Su valor en el origen puede incluso ser negativo, lo que indica una correlación negativa o antagonismo entre las variables correspondientes. En particular, el máximo de la covarianza cruzada puede estar desplazado del origen. Tal desplazamiento es frecuente en los fenómenos temporales, cuando una variable tiene un efecto no instantáneo sobre otra, produciendo un desfase de la correlación entre ambas variables. Este fenómeno se conoce bajo el nombre de *efecto de retardo*.
- **Desigualdad de Cauchy-Schwarz:** $\forall i, j \in \{1, \dots, N\}$, $C_{ii}(\mathbf{0})C_{jj}(\mathbf{0}) \geq |C_{ij}(\mathbf{h})|^2$.
- **Carácter de tipo positivo.** Para todo conjunto de sitios $\{\mathbf{x}_\alpha, \alpha = 1 \dots p\}$ y ponderadores $\{\lambda_\alpha^i, i=1 \dots N, \alpha=1 \dots p\}$, se tiene:

$$\sum_{i=1}^N \sum_{j=1}^N \sum_{\alpha=1}^p \sum_{\beta=1}^p \lambda_\alpha^i C_{ij}(\mathbf{x}_\alpha - \mathbf{x}_\beta) \lambda_\beta^j \geq 0.$$

1.2. Variogramas simples y cruzados

El variograma *cruzado* entre las funciones aleatorias Z_i y Z_j para un vector \mathbf{h} se define como:

$$\begin{aligned} \gamma_{ij}(\mathbf{h}) &= \frac{1}{2} \text{cov}\{Z_i(\mathbf{x}+\mathbf{h}) - Z_i(\mathbf{x}), Z_j(\mathbf{x}+\mathbf{h}) - Z_j(\mathbf{x})\} \\ &= \frac{1}{2} E\{[Z_i(\mathbf{x}+\mathbf{h}) - Z_i(\mathbf{x})][Z_j(\mathbf{x}+\mathbf{h}) - Z_j(\mathbf{x})]\} \end{aligned}$$

El variograma *simple* o *directo* corresponde al caso $i = j$. Se puede definir la matriz de variogramas de la siguiente manera:

$$\Gamma(\mathbf{h}) = [\gamma_{ij}(\mathbf{h})]_{i,j=1\dots N}$$

Los variogramas simples y cruzados tienen las siguientes propiedades:

- **Simetría:** $\gamma_{ij}(\mathbf{h}) = \gamma_{ji}(\mathbf{h})$ y $\gamma_{ij}(\mathbf{h}) = \gamma_{ij}(-\mathbf{h})$.
- **Nulidad en el origen:** $\gamma_{ij}(\mathbf{0}) = 0$.
- **Signo.** Los variogramas simples son funciones positivas, pero los variogramas cruzados no lo son necesariamente.
- **Desigualdad de Cauchy-Schwarz:** $\forall i, j \in \{1, \dots, N\}, \gamma_{ii}(\mathbf{h})\gamma_{jj}(\mathbf{h}) \geq |\gamma_{ij}(\mathbf{h})|^2$.
- **Carácter de tipo positivo.** Para un vector \mathbf{h} dado, la matriz $\Gamma(\mathbf{h})$ es simétrica y de tipo positivo, es decir que

$$\forall \lambda_1 \dots \lambda_N \in \mathbb{R}, \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j \gamma_{ij}(\mathbf{h}) \geq 0.$$

- **Relación con la función de covarianza cruzada:**

$$\gamma_{ij}(\mathbf{h}) = C_{ij}(\mathbf{0}) - \frac{1}{2}[C_{ij}(\mathbf{h}) + C_{ij}(-\mathbf{h})]$$

Matricialmente:

$$\Gamma(\mathbf{h}) = \mathbf{C}(\mathbf{0}) - \frac{1}{2}[\mathbf{C}(\mathbf{h}) + \mathbf{C}(-\mathbf{h})].$$

Esta relación muestra que el variograma cruzado toma el promedio de la función de covarianza cruzada en los valores $+\mathbf{h}$ y $-\mathbf{h}$. Descomponiendo esta función en su parte par y su parte impar, se ve que el variograma cruzado no contiene más que la parte par de la función de covarianza cruzada y, por ende, “pierde” información:

$$C_{ij}(\mathbf{h}) = \underbrace{\frac{1}{2}[C_{ij}(\mathbf{h}) + C_{ij}(-\mathbf{h})]}_{\text{término par}} + \underbrace{\frac{1}{2}[C_{ij}(\mathbf{h}) - C_{ij}(-\mathbf{h})]}_{\text{término impar}}.$$

En consecuencia, el variograma cruzado no es una herramienta adecuada para modelar datos cuando el término impar de la covarianza cruzada juega un papel significativo (frecuentemente, este término es el responsable de un efecto de retardo, es decir, de un desfase del máximo de las funciones de covarianza cruzada con respecto al origen). No obstante, a menudo, el número restringido de datos impide medir este desfase. Además,

los modelos usuales suelen considerar solamente funciones pares, por lo que uno omite la parte impar, lo que equivale a trabajar con los variogramas simples y cruzados.

1.3. Seudo-variograma cruzado

Otra herramienta variográfica ha sido introducida bajo el nombre de *seudo-variograma cruzado*, denotado $\pi_{ij}(\mathbf{h})$:

$$\pi_{ij}(\mathbf{h}) = \frac{1}{2} \text{var}[Z_i(\mathbf{x} + \mathbf{h}) - Z_j(\mathbf{x})].$$

Aunque se trata de una generalización del variograma, puesto que encontramos $\pi_{ii}(\mathbf{h}) = \gamma_{ii}(\mathbf{h})$ cuando $i = j$, aparece una complicación ligada a la naturaleza diferente de las variables regionalizadas: ¿qué significa la diferencia $Z_i(\mathbf{x} + \mathbf{h}) - Z_j(\mathbf{x})$ si las variables Z_i y Z_j no se expresan en unidades idénticas?

Se tiene la siguiente relación con la covarianza cruzada:

$$\pi_{ij}(\mathbf{h}) = \frac{1}{2} [C_{ii}(\mathbf{0}) + C_{jj}(\mathbf{0})] - C_{ij}(\mathbf{h}).$$

Así, el seudo-variograma cruzado puede servir para identificar y modelar los efectos de retardo, de la misma forma que la covarianza cruzada.

1.4. Inferencia estadística

Como en el caso univariable, se puede calcular las covarianzas y variogramas simples y cruzados experimentales para un determinado vector \mathbf{h} . En la práctica, se suele introducir tolerancias sobre la norma y la orientación de \mathbf{h} , con el fin de dar más robustez a los estimadores. En adelante, denotaremos como $\{\mathbf{x}_\alpha^i, \alpha = 1 \dots n_i\}$ los sitios de medición de la variable z_i ; su número y posición pueden diferir de una variable a otra (caso de *heterotopía*, total o parcial).

- **Estimador de la covarianza cruzada:**

$$\hat{C}_{ij}(\mathbf{h}) = \frac{1}{|N_{ij}(\mathbf{h})|} \sum_{N_{ij}(\mathbf{h})} [z_i(\mathbf{x}_\alpha^i) - \bar{z}_i][z_j(\mathbf{x}_\beta^j) - \bar{z}_j]$$

con $\bar{z}_i = \frac{1}{n_i} \sum_{\alpha=1}^{n_i} z_i(\mathbf{x}_\alpha^i)$ y $\bar{z}_j = \frac{1}{n_j} \sum_{\beta=1}^{n_j} z_j(\mathbf{x}_\beta^j)$

$$N_{ij}(\mathbf{h}) = \{(\alpha, \beta) \text{ tal que } \mathbf{x}_\alpha^i - \mathbf{x}_\beta^j = \mathbf{h}\}$$

$|N_{ij}(\mathbf{h})|$ es el número de pares distintos del conjunto $N_{ij}(\mathbf{h})$.

▪ **Estimador del variograma cruzado:**

$$\hat{\gamma}_{ij}(\mathbf{h}) = \frac{1}{2|N_{ij}(\mathbf{h})|} \sum_{N_{ij}(\mathbf{h})} [z_i(\mathbf{x}_\alpha) - z_i(\mathbf{x}_\beta)][z_j(\mathbf{x}_\alpha) - z_j(\mathbf{x}_\beta)]$$

donde, esta vez, $N_{ij}(\mathbf{h}) = \{(\alpha, \beta) \text{ tal que } \mathbf{x}_\alpha - \mathbf{x}_\beta = \mathbf{h}, \text{ siendo a la vez } z_i \text{ y } z_j \text{ medidas en } \mathbf{x}_\alpha \text{ y } \mathbf{x}_\beta\}$. El variograma cruzado necesita tener los datos de las diferentes variables en los mismos sitios. Por lo tanto, no se puede calcular en el caso de heterotopía total (caso en el cual el conjunto $N_{ij}(\mathbf{h})$ es vacío para todo vector \mathbf{h}), contrariamente a la covarianza cruzada.

1.5. Modelo lineal de correogionalización

Los variogramas simples y cruzados de un conjunto de variables no pueden modelarse independientemente, pues entre ellos existen restricciones matemáticas (en particular, la matriz $\Gamma(\mathbf{h})$ debe ser simétrica y de tipo positivo para todo vector \mathbf{h}). En la práctica, para satisfacer estas restricciones, se recurre al llamado *modelo lineal de correogionalización*, que generaliza el concepto de modelo anidado al caso multivariable.

Se supone que los variogramas simples y cruzados son combinaciones de un mismo conjunto de modelos de base:

$$\forall i, j \in [1, N], \gamma_{ij}(\mathbf{h}) = \sum_{u=1}^S b_{ij}^u g_u(\mathbf{h})$$

o sea, matricialmente

$$\Gamma(\mathbf{h}) = \sum_{u=1}^S \mathbf{B}_u g_u(\mathbf{h})$$

donde $\Gamma(\mathbf{h}) = [\gamma_{ij}(\mathbf{h})]_{i,j=1\dots N}$ es la matriz de los variogramas simples y cruzados

$g_u(\mathbf{h})$ es un modelo básico de variograma

$\mathbf{B}_u = [b_{ij}^u]_{i,j=1\dots N}$ se llama *matriz de correogionalización*.

Una condición suficiente para que el modelo así definido sea matemáticamente válido es que *cada una* de las matrices de correogionalización $\{\mathbf{B}_u, u = 1 \dots S\}$ sea simétrica de tipo positivo. Esto es muy fácil de controlar: todos los valores propios de \mathbf{B}_u deben ser positivos o nulos.

En el caso de dos variables (Z_1 y Z_2), la condición de positividad se cumple si

$$\forall u \in [1, S], |b_{12}^u| \leq \sqrt{b_{11}^u b_{22}^u}$$

Esta desigualdad implica que un modelo básico que aparece en el variograma cruzado con una meseta no nula, presenta forzosamente una meseta no nula en los dos variogramas simples. En la práctica, el usuario tiene que definir los número y tipos de modelos anidados que desea utilizar para el ajuste de los variogramas experimentales. Luego, tiene que buscar los parámetros (alcances, matrices de corregeionalización) que conducen a un buen ajuste, verificando que se cumple la desigualdad anterior para cada modelo anidado.

En el caso de tres o más variables, el chequeo es más difícil, puesto que la desigualdad anterior ya no es suficiente para asegurar que las matrices de corregeionalización sean de tipo positivo, sino que se debe calcular sus valores propios para comprobar si son positivos o no. Para ayudar a determinar estas matrices de corregeionalización, existen algoritmos de ajuste semi-automáticos que buscan minimizar el error cuadrático entre los variogramas experimentales y modelados, bajo las restricciones de positividad de los valores propios. Para ello, el usuario debe proponer los modelos básicos, en particular sus número (S), tipos y alcances.

1.6. Otros modelos

Existen otros procedimientos para modelar los variogramas simples y cruzados de una corregeionalización:

- Correlación intrínseca. Se trata de un modelo particular de corregeionalización, donde todos los variogramas simples y cruzados son proporcionales entre sí.
- Modelos basados en una hipótesis de Markov. Se postula que, conociendo el valor de la variable Z_i en un sitio \mathbf{x} , los valores de las otras variables en el mismo sitio no dependen más de los valores de Z_i en sitios distintos a \mathbf{x} . Esto trae como consecuencia que los variogramas cruzados $\{\gamma_{ij}(\mathbf{h}), j = 1 \dots N\}$ son proporcionales al variograma simple $\gamma_{ii}(\mathbf{h})$.
- Modelo bilineal de corregeionalización. Permite obtener covarianzas cruzadas que no son funciones pares. Resulta muy interesante para el estudio de fenómenos temporales en los cuales las variables evolucionan de manera asincrónica (efecto de retardo).
- Modelos basados en ecuaciones lineales o diferenciales entre las variables (por ejemplo, en hidrogeología, la transmisividad y la carga hidráulica).

2. La estimación local: el co-kriging

El objetivo del co-kriging es estimar el valor de una variable en un sitio o un bloque, a partir no sólo de las mediciones de ésta, sino también de otras variables correlacionadas con ella. Se trata de una extensión de la técnica de kriging al caso multivariable: la estimación

será una combinación lineal ponderada de los datos, sin sesgo y con una varianza del error mínima.

2.1. Co-kriging simple (medias conocidas)

Supongamos que se quiera estimar la variable Z_1 en el sitio \mathbf{x}_0 y denotemos como m_i el valor esperado (conocido) de la variable Z_i ($i = 1 \dots N$). El estimador se plantea como una combinación lineal de los datos disponibles en la vecindad de \mathbf{x}_0 , o sea:

$$Z_1^*(\mathbf{x}_0) = a + \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i Z_i(\mathbf{x}_{\alpha}^i),$$

donde el coeficiente a y los ponderadores $\{\lambda_{\alpha}^i, i = 1 \dots N, \alpha = 1 \dots n_i\}$ son las incógnitas del problema.

Se exige que el error de estimación tenga una esperanza nula:

$$\begin{aligned} E[Z_1^*(\mathbf{x}_0) - Z_1(\mathbf{x}_0)] &= a + \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i E[Z_i(\mathbf{x}_{\alpha}^i)] - E[Z_1(\mathbf{x}_0)] \\ &= a + \left[\sum_{\alpha=1}^{n_1} \lambda_{\alpha}^1 - 1 \right] m_1 + \sum_{i=2}^N \left[m_i \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i \right] \end{aligned}$$

lo que conduce a la siguiente condición:

$$a = \left[1 - \sum_{\alpha=1}^{n_1} \lambda_{\alpha}^1 \right] m_1 - \sum_{i=2}^N \left[m_i \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i \right].$$

Finalmente, se busca que la varianza del error de estimación sea mínima. Esta varianza se expresa en función de las covarianzas simples y cruzadas:

$$\text{var}[Z_1^*(\mathbf{x}_0) - Z_1(\mathbf{x}_0)] = \sum_{i=1}^N \sum_{j=1}^N \sum_{\alpha=1}^{n_i} \sum_{\beta=1}^{n_j} \lambda_{\alpha}^i \lambda_{\beta}^j C_{ij}(\mathbf{x}_{\alpha}^i - \mathbf{x}_{\beta}^j) - 2 \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i C_{i1}(\mathbf{x}_{\alpha}^i - \mathbf{x}_0) + C_{11}(\mathbf{0}).$$

La minimización conduce al siguiente sistema de ecuaciones lineales:

$$\sum_{j=1}^N \sum_{\beta=1}^{n_j} \lambda_{\beta}^j C_{ij}(\mathbf{x}_{\alpha}^i - \mathbf{x}_{\beta}^j) = C_{i1}(\mathbf{x}_{\alpha}^i - \mathbf{x}_0) \quad \forall i=1 \dots N, \forall \alpha=1 \dots n_i,$$

cuya resolución entrega los ponderadores de co-kriging. Este sistema de ecuaciones puede escribirse en forma matricial y resolverse con pivote de Gauss o con inversión matricial. Cabe señalar que el miembro de la izquierda no depende de la variable que se está estimando (Z_1). Por consiguiente, basta con invertir una sola matriz para obtener la

estimación de *todas* las variables en el sitio \mathbf{x}_0 . Sin embargo, se debe observar el importante tamaño de la matriz a invertir (en comparación con el caso univariable). Por ejemplo, este tamaño se multiplica por dos cuando se mide dos variables en los mismos sitios. Así, los tiempos de cálculo aumentan rápidamente al aumentar el número de variables.

La varianza del error de estimación (*varianza de co-kriging simple*) vale:

$$\sigma_{CKS-Z_1}^2(\mathbf{x}_0) = C_{11}(\mathbf{0}) - \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i C_{i1}(\mathbf{x}_{\alpha}^i - \mathbf{x}_0).$$

2.2. Co-kriging ordinario (medias desconocidas)

El co-kriging ordinario supone que los valores esperados m_1, \dots, m_N son desconocidos. La aplicación de las etapas clásicas (linealidad, insesgo, optimalidad) conducen al siguiente sistema de ecuaciones:

$$\left\{ \begin{array}{l} \sum_{j=1}^N \sum_{\beta=1}^{n_j} \lambda_{\beta}^j C_{ij}(\mathbf{x}_{\alpha}^i - \mathbf{x}_{\beta}^j) + \mu_i = C_{i1}(\mathbf{x}_{\alpha}^i - \mathbf{x}_0) \quad \forall i=1 \dots N, \forall \alpha=1 \dots n_i \\ \sum_{\alpha=1}^{n_1} \lambda_{\alpha}^1 = 1 \\ \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i = 0 \quad \forall i=2 \dots N \end{array} \right.$$

y varianza del error

$$\sigma_{CKO-Z_1}^2(\mathbf{x}_0) = C_{11}(\mathbf{0}) - \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i C_{i1}(\mathbf{x}_{\alpha}^i - \mathbf{x}_0) - \mu_1,$$

en donde μ_1, \dots, μ_N son incógnitas adicionales (multiplicadores de Lagrange).

Se puede escribir las ecuaciones anteriores en términos de variogramas en lugar de covarianzas: basta con reemplazar las covarianzas simples y cruzadas por el opuesto de los variogramas correspondientes. Como en el caso del co-kriging simple, el miembro de la izquierda no depende de la variable de interés. Por esto, basta con invertir una sola matriz para obtener la estimación de todas las variables en el sitio \mathbf{x}_0 .

Se propone a veces una variante del cokriging ordinario (llamada “cokriging ordinario estandarizado”), donde cambia la condición sobre la suma de los ponderadores

$$\sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i = 1$$

en lugar de las condiciones tradicionales

$$\begin{cases} \sum_{\alpha=1}^{n_1} \lambda_{\alpha}^1 = 1 \\ \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i = 0 \quad \forall i=2 \dots N \end{cases}$$

Los promotores de esta variante afirman que la nueva condición da más influencia a las variables secundarias, puesto que en el sistema tradicional los ponderadores asignados a cada variable secundaria suman cero. Sin embargo, esta nueva condición corresponde a una restricción de insesgo sólo si las variables poseen la *misma media desconocida*. Es el caso de mismo atributo medido sobre soportes distintos o con aparatos distintos; de lo contrario, se tiene que re-escalar cada variable en torno a una misma media.

2.3. Co-kriging co-localizado

Cuando se dispone de una variable auxiliar más muestreada que la variable de interés, incluso exhaustivamente conocida⁹, las mediciones situadas cerca del sitio a estimar tienden a hacer de pantalla a las mediciones alejadas. Además, la abundancia de datos de la variable auxiliar puede provocar inestabilidades numéricas al resolver el sistema de co-kriging.

Una simplificación consiste en utilizar para la estimación, sólo el dato de cada variable auxiliar más cercano al sitio a estimar, además de todos los datos asociados a la variable de interés. Si se supone que las variables auxiliares son conocidas en forma exhaustiva, esto equivale a considerar solamente su valor en el sitio a estimar (de donde viene el adjetivo *co-localizado*). En esta situación, ni siquiera se requiere conocer las covarianzas simples de las variables auxiliares, sino que solamente su valor en el origen, lo que reduce el esfuerzo de inferencia y modelamiento.

Entre los distintos tipos de co-kriging presentados anteriormente, se puede considerar el co-kriging simple o ordinario estandarizado (es decir, con una sola restricción de insesgo). El co-kriging ordinario clásico impone que la suma de los ponderadores asignados a las variables auxiliares sea nula, lo que conduce a un ponderador nulo para el único dato de cada variable auxiliar; por ende, sólo se tomaría en cuenta los datos de la variable de interés y se ignoraría las variables auxiliares.

Aunque simplifica el sistema de ecuaciones, el co-kriging co-localizado presenta varias desventajas:

⁹ Esta situación es análoga a aquella vista en el caso de kriging con una deriva externa. La variable conocida en forma exhaustiva puede provenir de imágenes satelitales, fotografías aéreas, modelos digitales de elevación o levantamientos geofísicos basados en datos sísmicos, según la aplicación considerada.

- Es necesario conocer las variables auxiliares en todos los sitios donde se busca estimar la variable de interés o a proximidad inmediata de estos.
- No se toma en cuenta toda la información disponible: se olvida casi todos los valores de las variables auxiliares, lo que puede traducirse por una subestimación de la precisión de la estimación (varianza de co-kriging demasiado alta).
- No es posible realizar un co-kriging ordinario clásico. Una solución para este problema consiste en utilizar como datos auxiliares aquellos ubicados en los mismos sitios que los datos de la variable de interés, además de los datos auxiliares co-localizados con el sitio a estimar.

2.4. Otras variantes

- **Co-kriging con tendencias:** esta variante sirve cuando los valores esperados m_1, \dots, m_N ya no son constantes, sino que variables en el espacio (*derivadas*). Podemos distinguir el co-kriging *universal* (derivadas polinomiales), *trigonométrico* (derivadas periódicas) y *con derivadas externas* (derivadas proporcionales a variables conocidas exhaustivamente).
- **Co-kriging mixto:** se conoce la media de algunas variables, mientras que se desconoce la media de las otras.
- **Co-kriging de bloque:** esta variante permite estimar directamente el valor promedio de las variables sobre un soporte mayor que el soporte de los datos.
- **Co-kriging factorial:** consiste en descomponer las funciones aleatorias (representadas por un modelo lineal de correogionalización) en un conjunto de *factores* mutuamente no correlacionados, jerarquizados en función de la cantidad de información que contienen, y estimar estos factores a partir de los datos disponibles.
- **Co-kriging de indicadores:** aplicación del co-kriging a variables binarias (indicadores) en vista a estimar una distribución de probabilidad.

2.5. Propiedades del co-kriging

El co-kriging comparte las mismas propiedades que el kriging, en particular:

- **Interpolación exacta:** en un sitio con dato, el valor estimado vuelve a dar el valor del dato, mientras que la varianza del error es nula.
- **Suavizamiento:** las estimaciones presentan menos variabilidad que los valores reales desconocidos. En consecuencia, el co-kriging tiende a sobrestimar las zonas de valores bajos y subestimar las zonas de valores altos, aunque en promedio los errores tienden a compensarse debido a la restricción de insesgo.

Como en el caso del kriging, los ponderadores y la varianza del error sólo dependen de la configuración geométrica formada por los sitios con datos y el sitio a estimar, así como de los variogramas (simples y cruzados) de las distintas variables, pero no dependen de los valores de los datos.

En lo que concierne a la regularidad del sistema de cokriging, es decir, la existencia y unicidad de la solución, es necesario que no existan datos duplicados y que las variables sean linealmente independientes en los sitios con datos, es decir, que no estén ligadas por una relación lineal. Por ejemplo, si las N variables representan proporciones y suman 100%, no se puede utilizarlas todas en el co-kriging: habrá que remover una variable (por ejemplo, Z_N) y deducir su estimación como 100% menos la suma de las estimaciones de Z_1, \dots, Z_{N-1} . El resultado es independiente de la elección de la variable que se deja de lado.

2.6. La alternativa entre kriging y co-kriging

En teoría, dos razones justifican que siempre es preferible estimar por co-kriging varias variables en lugar de hacer un kriging por separado de cada una de ellas:

- Se aprovecha, para estimar una variable, la información aportada por las demás. En especial, el co-kriging da siempre una varianza de estimación menor o igual que el kriging.
- Se mejora la coherencia de los resultados de estimación, puesto que se toma en cuenta las relaciones lineales entre las variables. Por ejemplo, cuando las variables representan proporciones, la suma de sus estimaciones es igual a 100%, situación que no se cumple necesariamente al realizar el kriging de cada variable por separado.

Ahora, existen dos situaciones donde el co-kriging coincide con el kriging por separado de cada variable:

- 1) Las variables son independientes entre sí (sus variogramas cruzados son nulos).
- 2) Los variogramas simples y cruzados son proporcionales entre sí (modelo conocido como *correlación intrínseca*) y el muestreo es homotópico, o sea, todas las variables han sido medidas en todos los sitios de muestreo. En este caso, las variables poseen la misma continuidad espacial y se vuelven redundantes entre sí.

El co-kriging mejora notablemente los resultados del kriging en caso de *heterotopía*, cuando las variables poseen una buena correlación espacial. Este método es particularmente ventajoso cuando una variable adicional es más accesible que la variable de interés. Esta ventaja se difumina si la correlación entre las variables es pequeña, caso en el cual se corre el riesgo de aumentar los tiempos de cálculo sin mejorar la precisión.

3. Simulación multivariable

Se puede extender los métodos de simulación condicional al ámbito multivariable (se habla entonces de *co-simulación*). En este capítulo, nos interesaremos por el modelo multi-Gaussiano, que sirve para simular variables de distribuciones continuas.

Se supone que cada variable ha sido transformada a una Gaussiana (anamorfosis) y que se ha verificado el carácter bi-Gaussiano para cada variable y cada par de variables, vía el análisis de nubes de correlación diferida, variogramas de indicadores o comparación entre variograma y madograma. A continuación, presentaremos distintas opciones para la co-simulación.

3.1. Simulación secuencial conjunta

Se define una secuencia aleatoria sobre todos los sitios a simular. Luego, en cada sitio de esta secuencia:

- Buscar los datos cercanos, tanto de la variable primaria como de la(s) variable(s) secundaria(s).
- Hacer un co-kriging simple de las distintas variables en el sitio a simular, a partir de los datos encontrados en la vecindad.
- Simular las variables como un vector Gaussiano, cuyos primeros momentos son el co-kriging simple y la matriz de varianza-covarianza de los errores de cokriging simple (la covarianza de los errores cometidos para dos variables distintas se expresa de una forma análoga a la varianza del error cometido para una sola variable).
- Incorporar los valores simulados a la base de datos, los cuales servirán de información condicionante para todos los sitios a simular a continuación.

3.2. Simulación secuencial jerárquica

En esta variante, se define una jerarquía entre las variables a simular (empezando con la variable más relevante o con aquella que tiene la mejor continuidad espacial) y, para cada variable, una secuencia aleatoria sobre los sitios a simular. En la primera corrida, se simula la variable primaria. Luego, en cada una de las corridas siguientes, se simula una variable secundaria según la jerarquía definida.

En cada caso, la distribución condicional del valor a simular es una Gaussiana de media igual al estimador de co-kriging simple y varianza igual a la varianza de co-kriging simple. Se puede utilizar:

- un co-kriging completo, que considera los datos Gaussianos de todas las variables y los valores previamente simulados en la corrida en cuestión y en las corridas anteriores;
- un co-kriging co-localizado, que considera sólo los datos Gaussianos de la variable que se está simulando, los valores ya simulados de esta variable y los valores co-localizados de las variables simuladas en las corridas anteriores. Esta opción simplificada permite disminuir la cantidad de cálculos, los cuales aumentan considerablemente a medida que se desarrolla la simulación secuencial.

3.3. Otros algoritmos

A pesar de su sencillez, el algoritmo secuencial requiere aproximaciones, debido a que el tamaño del sistema de co-kriging aumenta a medida que se desarrolla la simulación (en la práctica, esto se traduce por la definición de una vecindad móvil donde buscar los valores condicionantes: datos originales + valores previamente simulados).

Para evitar estas aproximaciones, se puede recurrir a otros algoritmos de co-simulación:

- Método de descomposición matricial
- Métodos de convolución
- Método espectral discreto
- Método espectral continuo
- Método de bandas rotantes.

En cada uno de estos algoritmos, es preciso utilizar un co-kriging para condicionar las realizaciones a los datos disponibles. Aquí el enfoque multivariable (co-kriging, en lugar de kriging) es indispensable para poder reproducir las correlaciones entre las variables, incluso en caso de homotopía. Ahora, estos métodos son más rápidos que el algoritmo secuencial porque el sistema de co-kriging sólo involucra a los datos originales y se puede condicionar varias realizaciones con un solo co-kriging (los ponderadores de co-kriging son los mismos para todas las realizaciones, luego pueden ser calculados de una vez por todas).

3.4. Aplicación a los datos de contaminación de suelo

A continuación, reanudamos con los datos de contaminación de suelo y aplicaremos los conceptos expuestos para co-simular las concentraciones de cobalto y níquel en la zona de estudio.

Primero, se transforma los datos de concentraciones de cobalto y níquel a datos con una distribución Gaussiana estándar (etapa de *anamorfosis Gaussiana*). Luego, se calcula los

variogramas experimentales simples y cruzados de las variables Gaussianas, a lo largo de las direcciones principales de anisotropía, a saber, las direcciones N60°E y N30°W. Para el ajuste, se propone el siguiente modelo, compuesto de un efecto pepita y tres estructuras esféricas anidadas (Figura 1):

$$\Gamma(\mathbf{h}) = \begin{pmatrix} \gamma_{Co}(\mathbf{h}) & \gamma_{Co-Ni}(\mathbf{h}) \\ \gamma_{Co-Ni}(\mathbf{h}) & \gamma_{Ni}(\mathbf{h}) \end{pmatrix} = \begin{pmatrix} 0.3 & 0.2 \\ 0.2 & 0.3 \end{pmatrix} \text{pepa} + \begin{pmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{pmatrix} \text{esf}(0.5\text{km}, 0.5\text{km}) \\ + \begin{pmatrix} 0.5 & 0.45 \\ 0.45 & 0.55 \end{pmatrix} \text{esf}(2\text{km}, 1.5\text{km}) + \begin{pmatrix} 0.2 & 0.08 \\ 0.08 & 0.15 \end{pmatrix} \text{esf}(\infty, 1.5\text{km}),$$

en donde los alcances corresponden a las direcciones N60°E y N30°W, respectivamente. Se puede verificar que todas las matrices de mesetas (matrices de correogionalización) son de tipo positivo (por ejemplo, por tratarse de matrices 2×2 , verificando que sus determinantes y trazas son positivos).

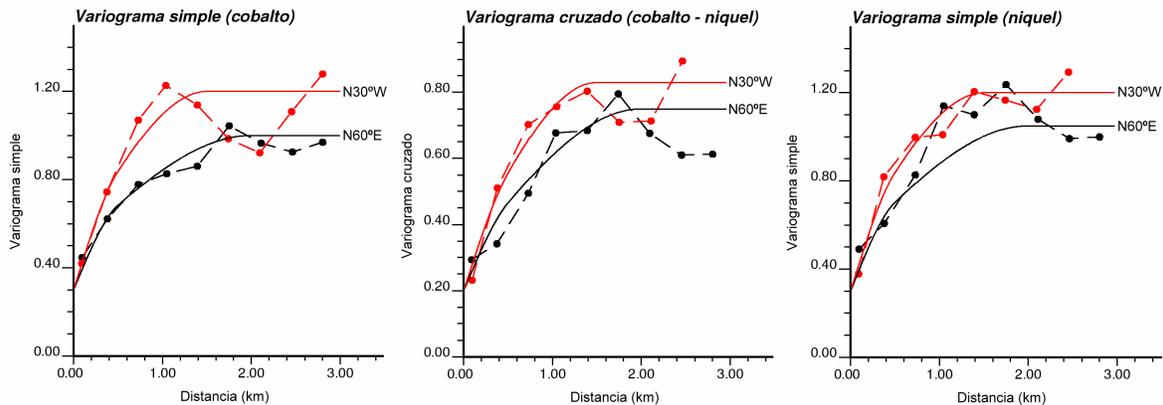


Figura 1. Variogramas simples y cruzados de los datos Gaussianos de cobalto y níquel. Líneas punteadas: variogramas experimentales. Líneas continuas: modelos ajustados.

Provistos del modelo variográfico bivariable, procedemos a simular las concentraciones de cobalto y níquel en una grilla densa (de malla $2\text{m} \times 2\text{m}$) que cubre la zona de estudio. La simulación se realiza mediante el algoritmo de bandas rotantes. Para el condicionamiento, se utiliza una vecindad de co-kriging que contiene hasta 48 datos (6 datos de cada variable en cada cuadrante del espacio). Se construye 100 realizaciones, las cuales son rebloqueadas al soporte de las unidades de remediación, promediando las concentraciones simuladas en cada unidad de $20\text{m} \times 20\text{m}$. La Figura 2 presenta dos de estas realizaciones.

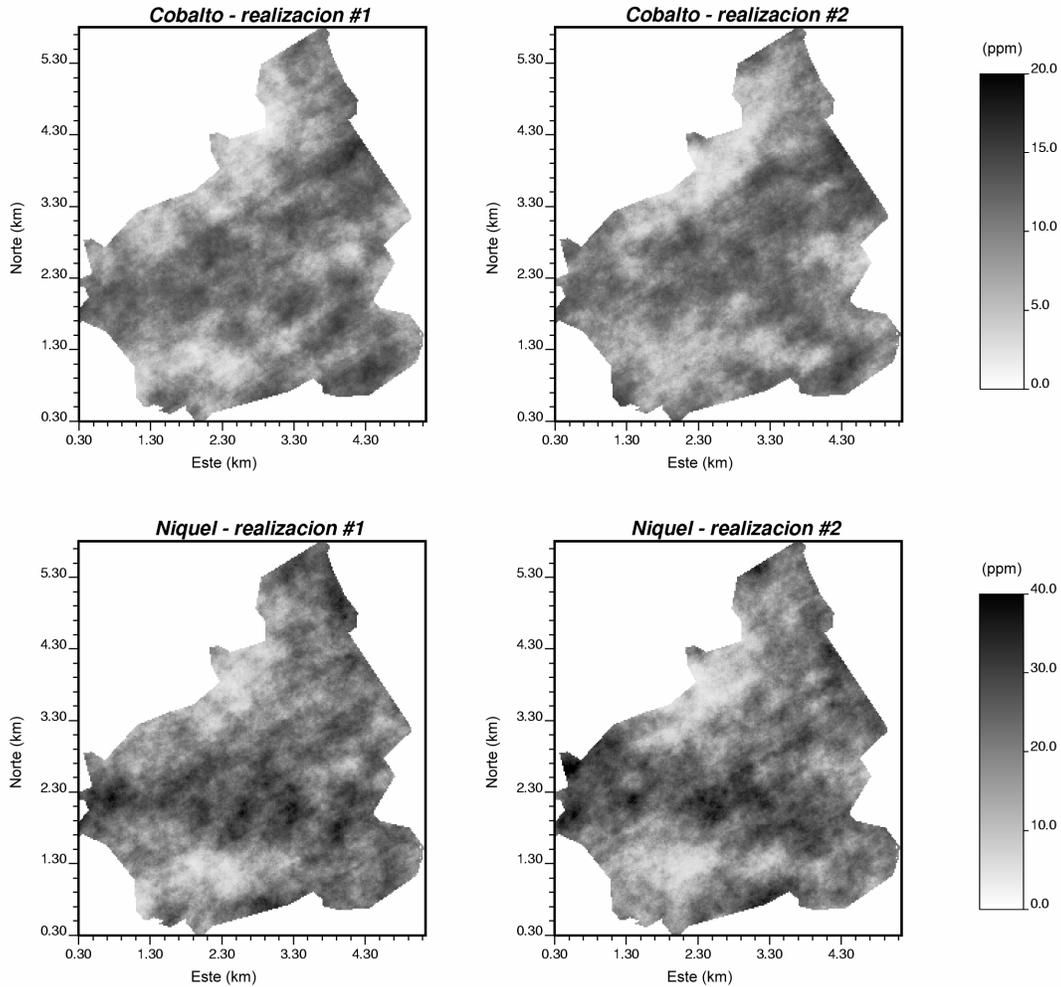


Figura 2. Dos realizaciones de las concentraciones promedio de cobalto (arriba) y níquel (abajo) en las unidades de remediación.

Supongamos que las normas ambientales indican que las concentraciones de cobalto y níquel no deben sobrepasar límites de 12 ppm y 30 ppm, respectivamente. Para cada unidad de remediación, utilizando los 100 escenarios simulados, se puede calcular la probabilidad de encontrar concentraciones mayores que estos valores límites (Figura 3, izquierda):

$$\begin{aligned} \text{Prob}\{\text{toxicidad}\}_{\text{análisis bivariable}} &= 1 - \text{Prob}\{\text{Co} < 12 \text{ ppm}, \text{Ni} < 30 \text{ ppm}\} \\ &= 1 - \text{Prob}\{\text{Co} < 12 \text{ ppm}\} \times \text{Prob}\{\text{Ni} < 30 \text{ ppm} \mid \text{Co} < 12 \text{ ppm}\} \end{aligned}$$

A modo de comparación, la parte derecha de la Figura 3 muestra las probabilidades que se obtendrían al suponer equivocadamente que las concentraciones de cobalto y níquel son independientes y al tratarlas en forma separada:

$$\text{Prob}\{\text{toxicidad}\}_{\text{análisis univariable}} = 1 - \text{Prob}\{\text{Co} < 12 \text{ ppm}\} \times \text{Prob}\{\text{Ni} < 30 \text{ ppm}\}.$$

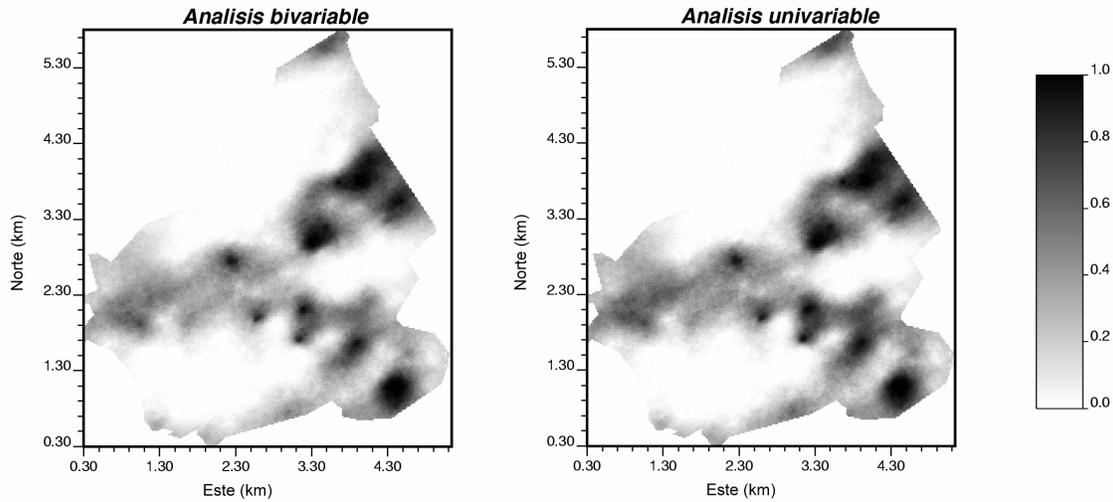


Figura 3. Mapas de las probabilidades de encontrar concentraciones tóxicas de cobalto o de níquel. Izquierda: análisis bivariable, que considera las correlaciones entre ambas variables. Derecha: análisis univariable, que ignora estas correlaciones.

Las probabilidades calculadas al tratar las concentraciones de cobalto y níquel en forma separada tiende a sobre-estimar las probabilidades obtenidas al considerar las correlaciones entre ambas variables (Figura 4). Este sesgo se explica por la correlación positiva entre las variables, lo que implica, en general:

$$\text{Prob}\{\text{Ni} < 30\text{ppm} \mid \text{Co} < 12\text{ppm}\} \geq \text{Prob}\{\text{Ni} < 30\text{ppm}\}.$$

En otras palabras, las unidades donde la concentración de níquel es menor que 30 ppm tienden a ser las mismas que las unidades donde la concentración de cobalto es menor que 12 ppm.

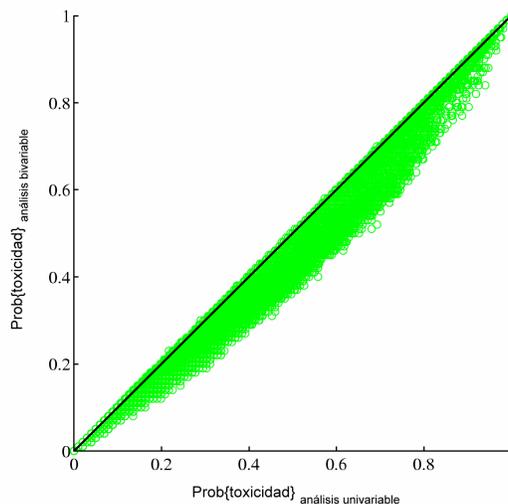


Figura 4. Comparación de las probabilidades obtenidas en la Figura 3.

Bibliografía

1. Obras generales

- [1] Chilès J.P. and Delfiner P., 1999. *Geostatistics: Modeling Spatial Uncertainty*, Wiley, New York, 695 p.
- [2] Cressie N.A.C, 1993. *Statistics for Spatial Data*, Wiley, New York, 928 p.
- [3] Deutsch C.V., 2002. *Geostatistical Reservoir Modeling*, Oxford University Press, New York, 376 p.
- [4] Deutsch C.V. and Journel A.G., 1998. *GSLIB: Geostatistical Software Library and User's Guide*, Oxford University Press, New York, 369 p.
- [5] Goovaerts P., 1997. *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York, 480 p.
- [6] Isaaks E.H. and Srivastava R.M., 1989. *An Introduction to Applied Geostatistics*, Oxford University Press, New York, 561 p.
- [7] Journel A.G., 1987. *Geostatistics for the Environmental Sciences*, Department of Applied Earth Sciences, Stanford University, California, 135 p.
- [8] Journel A.G. and Huijbregts C.J., 1978. *Mining Geostatistics*, Academic Press, London, 600 p.
- [9] Kitanidis P.K., 1997. *Introduction to Geostatistics: Applications to Hydrogeology*, Cambridge University Press, Cambridge, 271 p.
- [10] Matheron G., 1971. *The Theory of Regionalized Variables and its Applications*, Ecole Nationale Supérieure des Mines de Paris, Fontainebleau, 212 p.
- [11] Wackernagel H., 2003. *Multivariate Geostatistics: an Introduction with Applications*, Springer, Berlin, 387 p.
- [12] Webster R. and Oliver M.A., 2001. *Geostatistics for Environmental Scientists*, Wiley, New York, 286 p.

2. Variogramas, kriging y co-kriging

- [13] Emery X., 2005. Variograms of order ω : a tool to validate a bivariate distribution model, *Mathematical Geology*, Vol. 37, No. 2, in press.
- [14] Emery X. and Ortiz J.M., 2007. Weighted sample variograms as a tool to better assess the spatial variability of soil properties, *Geoderma*, Vol. 140, No. 1-2, p. 81-89.
- [15] Goovaerts P., 1992. Factorial kriging analysis: a useful tool for exploring the structure of multivariate spatial information, *Journal of Soil Science*, Vol. 43, No. 4, p. 567-619.
- [16] Goulard M. and Voltz M., 1992. Linear coregionalization model: tools for estimation and choice of cross-variogram matrix, *Mathematical Geology*, Vol. 24, No. 3, p. 269-286.
- [17] Rivoirard J., 1987. Two key parameters when choosing the kriging neighborhood, *Mathematical Geology*, Vol. 19, No. 8, p. 851-858.
- [18] Rivoirard J., 2001. Which models for collocated cokriging?, *Mathematical Geology*, Vol. 33, No. 2, p. 117-131.
- [19] Rivoirard J., 2002. On the structural link between variables in kriging with external drift, *Mathematical Geology*, Vol. 34, No. 7, p. 797-808.
- [20] Rivoirard J., 2004. On some simplifications of cokriging neighborhood, *Mathematical Geology*, Vol. 36, No. 8, p. 899-915.
- [21] Sandjivy L., 1984. The factorial kriging analysis of regionalized data. In: Verly G., David M., Journel A.G. and Maréchal A. (eds.), *Geostatistics for Natural Resources Characterization*: Reidel, Dordrecht, Vol. 1, p. 559-571
- [22] Wackernagel H., 1994. Cokriging versus kriging in regionalized multivariate data analysis, *Geoderma*, Vol. 62, No. 1-3, p. 83-92.
- [23] Wackernagel H., Bertino L. and Sierra J.P., 2002. Multivariate kriging for interpolating with data from different sources. In: Anderson C.W., Barnett V., Chatwin P.C. and El-Shaarawi A.H. (eds.), *Quantitative Methods for Current Environmental Issues*, Springer, p. 57-75.

3. Geoestadística no lineal

- [24] Emery X., 2006. Ordinary multigaussian kriging for mapping conditional probabilities of soil properties, *Geoderma*, Vol. 132, No. 1-2, p. 75-88.

- [25] Journel A.G., 1983. Non parametric estimation of spatial distributions, *Mathematical Geology*, Vol. 15, No. 3, p. 445-468.
- [26] Journel A.G., 1984. The place of non-parametric geostatistics. In: Verly G., David M., Journel A.G. and Maréchal A. (eds.), *Geostatistics for Natural Resources Characterization*: Reidel, Dordrecht, Vol. 1, p. 307-335.
- [27] Rivoirard J., 1994. *Introduction to Disjunctive Kriging and Nonlinear Geostatistics*, Oxford University Press, Oxford, 181 p.
- [28] Vann J. and Guibal D., 1998. Beyond ordinary kriging: an overview of non-linear estimation. In: Vann J. and Humphries M. (eds.), *Symposium on Beyond Ordinary Kriging*: Geostatistical Association of Australasia, Perth, p. 6-25.
- [29] Webster R. and Oliver M.A., 1989. Optimal interpolation and isarithmic mapping of soil properties. VI. Disjunctive kriging and mapping the conditional probability, *Journal of Soil Science*, Vol. 40, No. 3, p. 497-512.
- [30] Wood G., Oliver M.A. and Webster R., 1990. Estimating soil salinity by disjunctive kriging, *Soil Use and Management*, Vol. 6, No. 3, p. 97-104.
- [31] Yates S.R. and Yates M.V., 1988. Disjunctive kriging as an approach to management decision making, *Soil Science Society of America Journal*, Vol. 52, No. 6, p. 1554-1558.

4. Simulación

- [32] Armstrong M., Galli A., Le Loc'h G., Geffroy F. and Eschard R., 2003. *Plurigaussian Simulations in Geosciences*, Springer, Berlin, 160 p.
- [33] Chilès J.P. and Delfiner P., 1997. Discrete exact simulation by the Fourier method. In: Baafi E.Y. and Schofield N.A. (eds.), *Geostatistics Wollongong'96*: Kluwer Academic, Dordrecht, Vol. 1, p. 258-269.
- [34] Davis M., 1987. Production of conditional simulations via the LU triangular decomposition of the covariance matrix, *Mathematical Geology*, Vol. 19, No. 2, p. 91-98.
- [35] Emery X., 2007. Using the Gibbs sampler for conditional simulation of Gaussian-based random fields, *Computers & Geosciences*, Vol. 33, No. 4, p. 522-537.
- [36] Emery X. and Lantuéjoul C., 2006. TBSIM: A computer program for conditional simulation of three-dimensional Gaussian random fields via the turning bands method, *Computers & Geosciences*, Vol. 32, No. 10, p. 1615-1628.

- [37] Gómez-Hernández J.J. and Cassiraga E.F., 1994. Theory and practice of sequential simulation. In: Armstrong M. and Dowd P.A. (eds.), *Geostatistical Simulations*: Kluwer Academic, Dordrecht, p. 111-124.
- [38] Journel A.G., 1974. Geostatistics for conditional simulation of orebodies, *Economic Geology*, Vol. 69, No. 5, p. 673-687.
- [39] Lantuéjoul C., 1994. Non conditional simulation of stationary isotropic multigaussian random functions. In: Armstrong M. and Dowd P.A. (eds.), *Geostatistical Simulations*: Kluwer Academic, Dordrecht, p. 147-177.
- [40] Lantuéjoul C., 2002. *Geostatistical Simulation: Models and Algorithms*, Springer, Berlin, 256 p.
- [41] Matheron G., 1973. The intrinsic random functions and their applications, *Advances in Applied Probability*, Vol. 5, p. 439-468.
- [42] Shinozuka M.J. and Jan C.M., 1972. Digital simulation of random processes and its applications, *Journal of Sound and Vibrations*, Vol. 25, No. 1, p. 111-128.

5. Aplicaciones

- [43] Ahmed S. and de Marsily G., 1987. Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity, *Water Resources Research*, Vol. 23, No 9, p. 1717-1737.
- [44] Atteia O., Dubois J.P. and Webster R., 1994. Geostatistical analysis of soil contamination in the Swiss Jura, *Environmental Pollution*, Vol. 86, p. 315-327.
- [45] Balabdaoui F., Bocquet-Appel J.P., Lajaunie C. and Irudaya Rajan S., 2001. Space-time evolution of the fertility transition in India, 1961-1991, *International Journal of Population Geography*, Vol. 7, No. 2, p. 129-148.
- [46] Cassiraga E.F. and Gómez-Hernández J.J., 1993. Improved rainfall estimation by integration of radar data: a geostatistical approach. In: Soares A. (ed.), *Geostatistics Tróia '92*: Kluwer Academic, Dordrecht, Vol. 1, p. 363-374.
- [47] Chica-Olmo M. and Abarca-Hernández F., 1998. Radiometric coregionalization of Landsat TM and SPOT HRV images, *International Journal of Remote Sensing*, Vol. 19, No. 5, p. 997-1005.
- [48] Daly C., Jeulin D. and Lajaunie C., 1989. Application of multivariate kriging to the processing of noisy images. In: Armstrong M. (ed.), *Geostatistics*: Kluwer Academic, Dordrecht, Vol. 2, p. 749-760.

- [49] Delhomme J.P., 1978. Kriging in the hydrosociences, *Advances in Water Resources*, Vol. 1, No. 5, p. 251-266.
- [50] Doyen P.M., 1988. Porosity from seismic data: a geostatistical approach, *Geophysics*, Vol. 53, No. 10, p. 1263-1275.
- [51] de Fouquet C. and Mandallaz D., 1993. Using geostatistics for forest inventory with air cover: an example. In: Soares A. (ed.), *Geostatistics Tróia'92*: Kluwer Academic, Dordrecht, Vol. 2, p. 875-886.
- [52] Goovaerts P., 1999. Geostatistics in soil science: state-of-the-art and perspectives, *Geoderma*, Vol. 89, No. 1-2, p. 1-45.
- [53] Goovaerts P., 2001. Geostatistical modelling of uncertainty in soil science, *Geoderma*, Vol. 103, No. 1-2, p. 3-26.
- [54] Hudson G. and Wackernagel H., 1994. Mapping temperature using kriging with external drift: theory and an example from Scotland, *International Journal of Climatology*, Vol. 14, No. 1, p. 77-91.
- [55] Lajaunie C., 1984. A geostatistical approach to air pollution modeling. In: Verly G., David M., Journel A.G. and Maréchal A. (eds.), *Geostatistics for Natural Resources Characterization*: Reidel, Dordrecht, Vol. 2, p. 877-891.
- [56] Ma Y.Z. and Royer J.J., 1988. Local geostatistical filtering: application to remote sensing, *Sciences de la Terre, Série Informatique*, Vol. 27, p. 17-36.
- [57] Oliver M.A., Lajaunie C., Webster R., Muir K.R. and Mann J.R., 1993. Estimating the risk of childhood cancer. In: Soares A. (ed.), *Geostatistics Tróia'92*: Kluwer Academic, Dordrecht, Vol. 2, p. 899-910.
- [58] Petitgas P., 2001. Geostatistics in fisheries survey design and stock assessment: models, variances and applications, *Fish and Fisheries*, Vol. 2, p. 231-249.
- [59] Raspa G., Bruno R., Dosi P., Filippi N. and Patrizi G., 1993. Multivariate geostatistics for soil classification. In: Soares A. (ed.), *Geostatistics Tróia'92*: Kluwer Academic, Dordrecht, Vol. 2, p. 793-804.
- [60] Raspa G., Tucci M. and Bruno R., 1997. Reconstruction of rainfall fields by combining ground rain gauges data with radar maps using external drift method. In: Baafi E.Y. and Schofield N.A. (eds.), *Geostatistics Wollongong'96*: Kluwer Academic, Dordrecht, Vol. 2, p. 1306-1315.
- [61] Rivoirard J., Simmonds J., Foote K.G., Fernandes P. and Bez N., 2000. *Geostatistics for Estimating Fish Abundance*, Blackwell Science, London, 216 p.

- [62] Rouhani S. and Wackernagel H., 1990. Multivariate geostatistical approach to space-time data analysis, *Water Resources Research*, Vol. 26, No. 4, p. 585-591.
- [63] Séguret S. and Huchon P., 1990. Trigonometric kriging: a new method for removing the diurnal variation from geomagnetic data, *Journal of Geophysical Research – Solid Earth and Planets*, Vol. 95, No. B13, p. 21383-21397.
- [64] Sousa A.J., 1989. Geostatistical data analysis – an application to ore typology. In: Armstrong M. (ed.), *Geostatistics*: Kluwer Academic, Dordrecht, Vol. 2, p. 851-860.
- [65] von Steiger B., Webster R., Schulin R. and Lehmann R., 1996. Mapping heavy metals in polluted soil by disjunctive kriging, *Environmental Pollution*, Vol. 94, No. 2, p. 205-215.
- [66] Wackernagel H., Lajaunie C., Thiéry L., Vincent R. and Grzebyk M., 1993. Applying geostatistics to exposure monitoring data in industrial hygiene. In: Soares A. (ed.), *Geostatistics Tróia'92*: Kluwer Academic, Dordrecht, Vol. 1, p. 463-476.
- [67] Wartenberg D., 1993. Some epidemiologic applications of kriging. In: Soares A. (ed.), *Geostatistics Tróia'92*: Kluwer Academic, Dordrecht, Vol. 2, p. 911-910.
- [68] Webster R., Oliver M.A., Muir K.R. and Mann J.R., 1990. Kriging the local risk of a rare disease from a register of diagnoses, *Geographical Analysis*, Vol. 26, No. 2, p. 168-185.
- [69] Yarus J.M. and Chambers R.L. (eds.), 1994. *Stochastic Modeling and Geostatistics: Principles, Methods and Case Studies*, American Association of Petroleum Geologists (AAPG), Tulsa, 379 p.