

Una introducción a la geoestadística y sus posibles aplicaciones a las ciencias del agua^(*)

Por JOSE JAIME GOMEZ HERNANDEZ

Ingeniero de Caminos, Canales y Puertos,
Departamento de Hidráulica, Escuela Técnica Superior de Ingenieros de Caminos,
Universidad Politécnica de Valencia, Applied Earth Sciences Department,
Stanford University, California

La incertidumbre en los datos, la existencia de una estructura espacial, la necesidad de disponer de una medida del error cometido en la estimación del fenómeno en los puntos en los que no se disponía de datos con la intención de minimizarlo, fueron algunas de las razones que dieron lugar al nacimiento de la Geoestadística para su uso en la minería. Pero, y aunque hasta el momento no ha captado la atención del Ingeniero de Caminos, la Geoestadística se ha mostrado como un potente instrumento para ser usado en el campo de las ciencias del agua. Algunas de las aplicaciones han sido: dibujo automático de curvas de nivel, evaluación de recursos medios, calibrado de modelos, diseño y optimización de redes de medida, y simulación.

INTRODUCCION

Los datos utilizados en ciencias como la hidrología o la hidrogeología (pluviometría, transmisividad, coeficiente de almacenamiento, cotas piezométricas...) son utilizados, en la mayoría de los casos, en modelos deterministas. Algún método de interpolación que implícitamente admitirá una variación regular del fenómeno entre dos puntos conocidos tiene que usarse para obtener valores en los puntos desconocidos. Pero los datos ni varían regularmente ni son suficientes en número para darnos una visión global de su variabilidad ni siquiera son conocidos con certeza: o existen errores de medida o se han obtenido de manera indirecta (como ocurre con la transmisividad obtenida a partir de la capacidad específica). Para tener en cuenta, tanto la incertidumbre como la variabilidad de los datos, es necesario recurrir a una aproximación probabilista, más exactamente a una aproximación estocástica cuyo objetivo será modelar teniendo en cuenta dicha incertidumbre.

Esta incertidumbre en los datos, el conocimiento de una estructura espacial en los mismos y la necesidad de poder estimar la cantidad de oro extraíble en las minas de Africa del

Sur a partir de los datos proporcionados por los sondeos, con la máxima fiabilidad posible dio lugar al nacimiento de una nueva rama de la Estadística que primeramente fue formulada en base a resultados experimentales y que terminó formalizándose en la teoría de la variable regionalizada.

La teoría de la variable regionalizada desarrollada por Matheron (1967 y 1971) y el método de estimación por él propuesto y que dio en llamar kriging se han mostrado como herramientas propicias para el tipo de aproximación perseguido. Esta teoría que inicialmente se utilizó en la evaluación de recursos mineros y que dio lugar al nacimiento de la Geoestadística pronto encontró aplicación en otros campos como la hidrología y la hidrogeología.

En el resto del artículo haremos una revisión de los conceptos básicos de la Geoestadística y presentaremos algunas de las aplicaciones que en el campo de las ciencias del agua han sido llevadas a cabo desde 1974.

VARIABLE REGIONALIZADA Y VARIABLE ALEATORIA

Las variables utilizadas en los modelos matemáticos usados tanto en hidrología como en

(*) Se admiten comentarios sobre el presente artículo, que podrán remitirse a la Redacción de esta Revista hasta el 31 de enero de 1987.

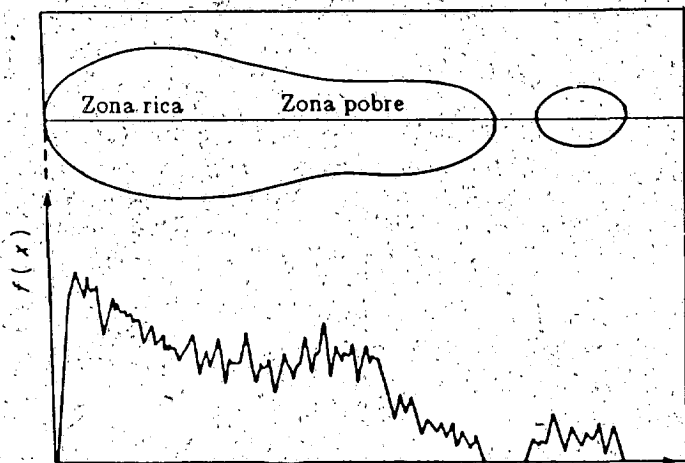


Figura 1.—Aspecto estructural y aleatorio de una variable regionalizada.

hidrogeología varían de un punto a otro, ya sea en el espacio o en el tiempo. Al mismo tiempo incluyen un cierto grado de incertidumbre debido a errores de medida o a la mera extrapolación de un valor donde no existe dato. Y, por otra parte, mantienen una cierta correlación espacial o temporal.

Desde el punto de vista matemático tal tipo de variables son conocidas como variables regionalizadas, siguiendo la terminología de Matheron (1971), y pueden representarse por una función $z(x)$ que proporciona un valor de la característica z en cada punto x , donde x es un vector con 1, 2 ó 3 componentes. En esta función pueden apreciarse dos componentes: a pequeña escala presenta un comportamiento errático mientras que a una escala mayor se observa claramente una estructura espacial como puede verse en la Figura 1. La teoría de la variable regionalizada tiene en cuenta este doble aspecto aleatorio y estructurado, al utilizar una aproximación probabilista para poder representar el comportamiento errático a nivel local junto con el carácter estructural a nivel global.

Una variable aleatoria es una variable que toma sus valores según una cierta distribución de probabilidad (seguiremos el convenio de representar con letras mayúsculas la variable aleatoria y con minúsculas cada posible valor de la misma). Desde este punto de vista, una variable regionalizada puede entenderse como una realización de un conjunto de variables aleatorias, que a partir de aquí llamaremos función alea-

toria y que representaremos por $Z(x)$. Quedan recogidos los dos aspectos antes citados, pues, en cada punto x_i , $Z(x_i)$ es una variable aleatoria con lo que se recoge esa variabilidad a pequeña escala, pero la función aleatoria es tal que $Z(x_i)$ y $Z(x_i + h)$ no son independientes sino que guardan una correlación espacial que refleje la estructura de la variable regionalizada $z(x)$.

Esta interpretación probabilista de la variable regionalizada no tendría sentido si no pudiéramos inferir las distribuciones de las distintas variable aleatorias a partir de los datos disponibles. Ahora bien, para inferir la distribución de una variable aleatoria necesitamos conocer un número suficientemente extenso de posibles realizaciones y, de las variables regionalizadas, sólo disponemos de una realización (por ejemplo, los valores de transmisividad medidos en un determinado acuífero) de la que es imposible deducir las distribuciones de cada una de las variables aleatorias que asignamos a cada punto del espacio. Se necesita, por tanto, establecer la hipótesis de estacionaridad, que no debe verse como una limitación del alcance de la geoestadística puesto que no es otra cosa sino una hipótesis más de las que constantemente se hacen en Estadística y que, por ejemplo, también se hace en el estudio de series temporales.

ESTACIONARIDAD

Si el fenómeno en estudio se considera homogéneo en el espacio de manera que la función aleatoria puede repetirse a sí misma de un punto a otro, entonces puede establecerse la hipótesis de estacionaridad según la cual, dos valores experimentales $z(x_0)$ y $z(x_0 + h)$ aunque son dos realizaciones distintas de dos variables aleatorias distintas, puede considerarse que poseen la misma distribución de probabilidad. De este modo, los momentos de la única función de distribución a tener en cuenta, la misma para todos los puntos del espacio, podrían ser inferidos del conjunto experimental de datos $\{z(x_1), z(x_2), \dots, z(x_n)\}$. En otras palabras, pasamos de considerar $\{z(x_1), z(x_2), \dots, z(x_n)\}$ como n variables aleatorias distintas, a considerarlas como n realizaciones de la misma distribución.

Esta hipótesis, que formalmente equivale a

decir que la función de distribución de probabilidad conjunta de n variables aleatorias es independiente de la posición absoluta de las mismas y sólo depende de su posición relativa, es demasiado estricta y suele relajarse para los propósitos perseguidos en geoestadística. La estacionariedad se limita, al primer momento de la variable y , al segundo momento de las diferencias, esto es, el valor medio de las variables aleatorias es independiente de la posición y la varianza de las diferencias entre dos variables aleatorias separadas por un vector h sólo depende de dicho vector:

$$E\{Z(x)\} = m, \quad \forall x \quad (1)$$

$$\text{Var}\{Z(x+h) - Z(x)\} = E\{[Z(x+h) - Z(x)]^2\} = 2\tau(h), \quad \forall x \quad (2)$$

una variable regionalizada que cumple estos requisitos se dice débilmente estacionaria de segundo orden.

La condición (2) es el soporte del instrumento más importante de que dispone la Geoestadística: el semivariograma. El semivariograma, que denotaremos τ , se define como la mitad de la varianza de los incrementos y es una función exclusiva del vector h . Conviene hacer notar, con el fin de evitar malinterpretaciones, que algunos autores se refieren a la misma cantidad τ con el nombre de variograma, aquí por variograma se entenderá $2\tau(h)$ y su valor es:

$$2\tau(h) = E\{[Z(x+h) - Z(x)]^2\} \quad (3)$$

En la práctica la hipótesis de débil estacionariedad de segundo orden todavía se relaja más exigiendo, únicamente, que se verifique sólo para un determinado rango del vector h , que corresponde a la escala de trabajo. Así, la variable regionalizada será localmente débilmente estacionaria de segundo orden. Consideremos, por ejemplo, el conjunto de alturas piezométricas en un acuífero. A nivel regional es claro que existe una deriva en la dirección del gradiente pero que generalmente es lo suficientemente pequeña como para que en un estudio local podamos considerar que las alturas piezométricas son estacionarias. (Figura 2). De todos modos

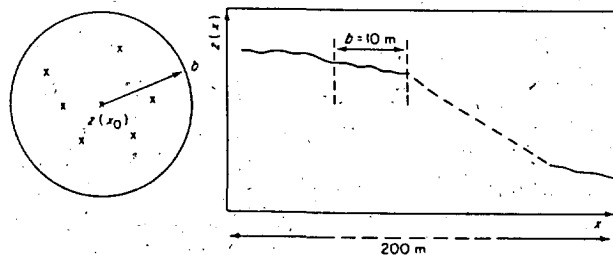


Figura 2. — Entorno en el que una variable no estacionaria puede considerarse localmente estacionaria.

en el caso en que la variable regionalizada presente una deriva apreciable incluso a pequeña escala, es posible abordar el problema recurriendo a la descomposición $z(x) = m(x) + y(x)$ como se verá más adelante, donde $m(x)$ es una variable regionalizada que representa la deriva e $y(x)$ es otra variable regionalizada que representa las fluctuaciones respecto a esa deriva y que podemos asumir tiene valor medio nulo.

SEMIVARIOGRAMA

El semivariograma teórico quedó definido en (3) como un medio de la varianza de los incrementos y que bajo la hipótesis de estacionariedad de segundo orden puede ser inferido fácilmente del conjunto de datos experimentales.

Estimaremos

$$E\{[Z(x+h) - z(x)]^2\}$$

con

$$\frac{1}{2N} \sum [z(x_i) - z(x_i + h)]^2 \quad [4]$$

donde N = número de parejas separadas por el vector h que se pueden encontrar en el conjunto de datos. Puesto que, generalmente, los puntos están irregularmente distribuidos en el espacio, éstos tienen que ser reagrupados en clases para poder calcular la varianza de las diferencias. Así, en lugar de fijar una dirección y un módulo del vector h , se fija un ángulo (que generalmente no será inferior a $1/8$ de círculo) y un rango para el módulo, de tal manera que todas las parejas que quedan dentro de ese án-

gulo y cuya separación esté dentro del rango fijado se consideran en el cálculo del valor del semivariograma para la clase así definida.

El semivariograma debe valer cero en el origen por definición, e ir incrementando en valor a medida que el vector h aumenta en módulo y la correlación espacial se pierde. Es una función simétrica, esto es, toma el mismo valor en h que en $-h$. El semivariograma contiene en sí mismo gran cantidad de información y posee unas características que pasamos a analizar ya que pueden ayudar en la identificación del fenómeno en estudio.

Anisotropía

Puesto que el semivariograma es función de un vector, podemos calcular para diferentes direcciones diferentes semivariogramas función del módulo. La comparación entre ellos puede revelar la existencia de anisotropía en el fenómeno en estudio. (Figura 3).

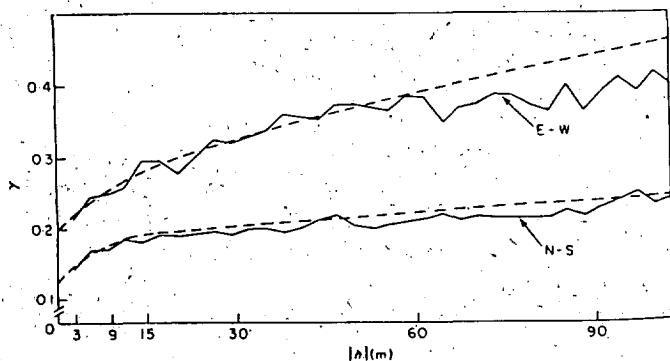


Figura 3. — Reflejo de la anisotropía en el semivariograma. En este caso la variable regionalizada es la cantidad de Cobre en cierto depósito.

Comportamiento en el origen

Pueden presentarse cuatro diferentes comportamientos cuando el módulo del vector h tiende a cero: parabólico, lineal, discontinuo y efecto pepita puro. En la figura 4, tomada de Journel y Huijbregts (1978) puede verse la evolución de 4 piezómetros en Korgohò (Costa de Marfil) a lo largo de 6 meses y los respectivos semivariogramas que se obtienen. (En este ejemplo el fenómeno es unidimensional y estudia variabilidad temporal).

Parabólico. El comportamiento parabólico, es, $\gamma(h) \propto h^2$, representa un fenómeno ex-

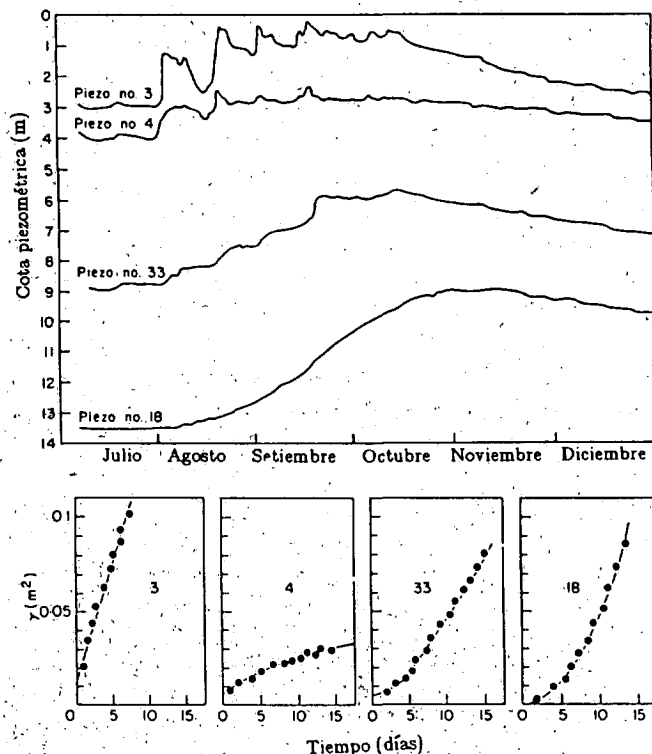


Figura 4. — Distintos comportamientos del semivariograma en el origen. Piezómetros 3 y 4, efecto pepita; piezómetro 33, lineal; piezómetro 18, parabólico.

tremadamente continuo donde la primera de las derivadas existe en el origen. Es el caso del piezómetro número 18 donde la cota piezométrica está muy profunda y el efecto de las recargas discontinuas por lluvias no se aprecian. (Nótese que un fenómeno cuya variable regionalizada estuviera definida por una función lineal tendría exactamente un semivariograma parabólico).

Lineal. El comportamiento lineal $\gamma(h) \propto h$ corresponde a un fenómeno menos regular que el anterior, el semivariograma es continuo pero no diferenciable. Es el caso del piezómetro 33 donde la cota piezométrica está lo suficientemente profunda para que el efecto de las lluvias quede casi completamente amortiguado aunque presenta algunas discontinuidades debido a grandes descargas.

Discontinuidad en el origen. Conocido como «efecto pepita», este comportamiento indica que la variabilidad entre dos puntos aunque estén muy próximos puede ser bastante grande, tanto más grande cuanto mayor sea la discontinuidad de $\gamma(h)$ en el origen. la denominación

proviene de los primeros tiempos en que la Geoestadística se utilizaba en la predicción de la cantidad de oro que se podía esperar en las explotaciones mineras de Suráfrica y del hecho de que esta cantidad de oro podía tener grandes variaciones entre dos muestras tomadas una junta a la otra si una de ellas tenía una pepita y la otra no. Esto es representativo de un fenómeno irregular y el efecto pepita es una medida del carácter aleatorio de la variable regionalizada que no puede explicarse por la distancia que separa las muestras. Este carácter aleatorio puede ser debido a errores de medida o a microvariabilidad que, aunque continua, no puede detectarse a la escala de trabajo. Es el caso de los piezómetros 3 y 4 en los que la cota piezométrica esta próxima a la superficie y la respuesta a las acciones exteriores (lluvia en este caso) es casi instantánea.

Efecto pepita puro. Corresponde a un fenómeno puramente aleatorio: la variabilidad esperada entre dos puntos es independiente de la situación de los mismos, es decir $\tau(h) = C$, para todo $h > 0$. Este es un efecto que pocas veces puede presentarse en el campo experimental y que, en general, no es deseado. Imaginemos que la variable regionalizada, transmisividad, presentara un efecto pepita puro con un determinado valor C que habría sido estimado igual a la varianza muestral, en este caso el hecho de que conociéramos los valores de la transmisividad en un determinado número de puntos, no nos aportaría ninguna información a la hora de estimar los valores en localizaciones desconocidas, por muy próximas que fueran a las conocidas. Solamente tendríamos una idea del rango de variabilidad esperado, medido éste por la varianza muestral.

Comportamiento en el infinito

En cuanto al comportamiento en el infinito hay dos posibilidades, que el límite del semivariograma cuando h tiende a infinito exista o que no exista. En el caso de que exista, diremos que el modelo presenta meseta y representa un fenómeno cuya capacidad de variación entre dos puntos aumenta con la distancia entre los mismos hasta que alcanza un máximo que corres-

ponderá a la varianza del fenómeno. El valor de h para el que el semivariograma se estabiliza se conoce como rango o alcance. Si no existe meseta quiere decir que el fenómeno tiene una capacidad infinita de variación, y por tanto la varianza aumenta con el tamaño del área considerada.

El hecho de utilizar el semivariograma en lugar de la covarianza permite trabajar con estos fenómenos que presentan una capacidad infinita de variación espacial. Hay que señalar que semivariograma y covarianza son términos unívocamente equivalentes, siempre que el fenómeno tenga meseta, relacionados por la expresión:

$$\tau(h) = C(0) - C(h) \quad [5]$$

donde $C(h)$ es la covarianza entre puntos separados por el vector h , y, por tanto, $C(0)$ representa la varianza del fenómeno en sí.

SEMIVARIOGRAMAS TEORICOS

Para las aplicaciones que veremos es importante ajustar un semivariograma teórico a nuestro semivariograma experimental con diferentes propósitos, como poder automatizar los procesos de cálculo o, más importante, poder caracterizar la estructura espacial de nuestro fenómeno con pocos parámetros. Pensemos, por ejemplo, en la necesidad de estimar el campo de transmisividades de un acuífero a partir de unos pocos datos con la mayor confianza posible. Un método consiste en establecer una función objetivo con tantas incógnitas como puntos a estimar con la intención de minimizar las desviaciones entre los valores simulados y los observados durante cierto período de tiempo (lo que se conoce como problema inverso). Otro método más simple consistiría, no en estimar puntualmente los valores de la transmisividad en cada punto sino estimar los parámetros del semivariograma que caracterizaría el modelo. Los semivariogramas teóricos más utilizados sólo cuentan con dos parámetros. La combinación lineal de éstos puede llevarnos a modelos más sofisticados que en ningún caso suelen sobrepasar los cinco parámetros y que

son capaces de representar casi cualquier tipo de fenómeno.

Cualquier función no puede utilizarse para representar un semivariograma, se requiere que la función sea condicionalmente definida positiva para satisfacer que la varianza de cualquier combinación lineal de las variables aleatorias representadas sea no negativa. Puesto que la varianza de una combinación lineal de variables aleatorias que cumplan la condición de local débil estacionaridad de segundo orden puede expresarse como:

$$\begin{aligned} \text{Var} \left[\sum \alpha_i Z(x_i) \right] &= \\ &= \sum_i \sum_j \alpha_i \alpha_j C(x_i - x_j) \geq 0 \end{aligned} \quad (6)$$

si se hace uso de la relación (5), y se añade la condición de que

$$\sum \alpha_i = 0 \quad (7)$$

se llega a que

$$\begin{aligned} \text{Var} \left[\sum \alpha_i Z(x_i) \right] &= \\ &= - \sum_i \sum_j \alpha_i \alpha_j \tau(x_i - x_j) \geq 0 \end{aligned} \quad (8)$$

es decir, $-\tau$ debe ser una función definida positiva bajo la condición de que $\sum \alpha_i = 0$.

Las expresiones analíticas de los semivariogramas que satisfacen la condición (8) y son de uso más común puede verse en la tabla 1 y su representación gráfica en la figura 5. La existencia del efecto pepita mencionado se introduce sumando una constante al semivariograma salvo en el origen donde por definición $\tau(0) = 0$.

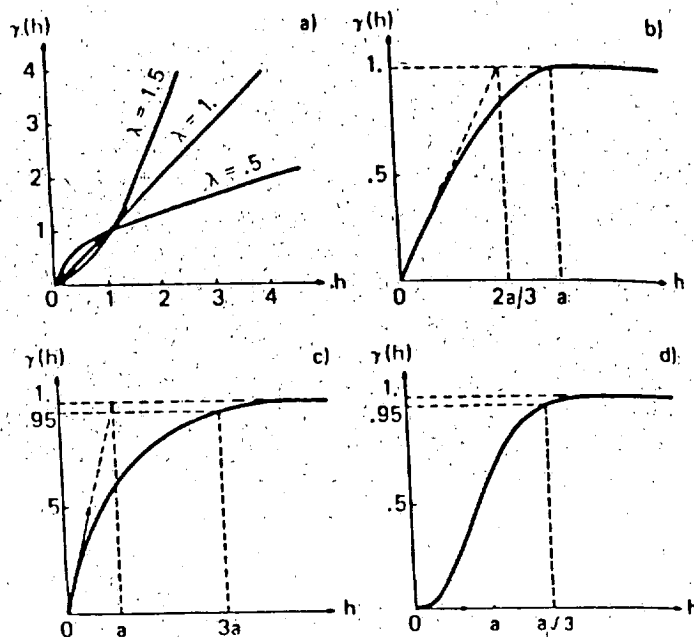


Figura 5.—Semivariogramas teóricos. a) Monómic, b) Esférico, c) Exponencial, d) Gaussiano.

TABLA 1

EXPRESIONES MATEMATICAS DE LOS SEMIVARIOGRAMAS TEORICOS MAS UTILIZADOS

Tipo	Fórmula	Pendiente en el origen	Rango	Meseta
Efecto pepita	$\gamma(h) = C(1 - \delta)$	∞	0	C
Monómic	$\gamma(h) = \omega h ^\lambda$	$\begin{cases} \infty & 0 < \lambda < 1 \\ \omega & \lambda = 1 \\ 0 & 1 < \lambda < 2 \end{cases}$	∞	∞
Esférico	$\gamma(h) = \begin{cases} \omega \left[\frac{3}{2} \frac{ h }{a} - \frac{1}{2} \left(\frac{ h }{a} \right)^3 \right] & h \leq a \\ \omega & h > a \end{cases}$	$3\omega/2a$	a	ω
Exponencial	$\gamma(h) = \omega [1 - \exp(- h /a)]$	ω/a	$3a$	ω
Gaussiano	$\gamma(h) = \omega [1 - \exp(- h ^2/a^2)]$	0	$a\sqrt{3}$	ω

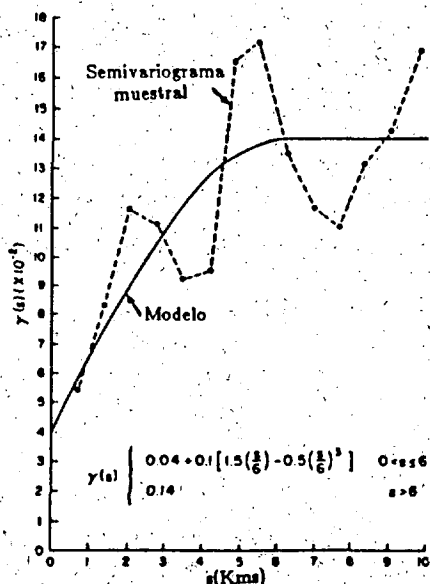


Figura 6. — Ejemplo de semivariograma muestral y ajuste de un semivariograma esférico. La variable regionalizada considerada es el logaritmo de la trasmisividad.

El hecho de que la varianza de los estimadores de los valores del semivariograma sea grande hace que el ajuste de un semivariograma teórico a uno experimental requiera una gran dosis de sentimiento por parte del ingeniero, sentimiento éste que provendrá de un exhaustivo conocimiento de la información accesible que permita intuir parte de la información encerrada en el semivariograma como por ejemplo, si, para la escala de trabajo, el fenómeno tendrá meseta o no, si se espera un comportamiento en el origen de tipo parabólico o discontinuo, si se espera un fenómeno isotrópico o anisotrópico o cuál será el alcance en cada dirección de anisotropía. Un ejemplo de semivariograma experimental y del «sentimiento» que se necesita para ajustar un semivariograma teórico puede verse en la Figura 6.

VARIANZA DE LA ESTIMACION

Una de las aplicaciones de la Geoestadística es estimar los valores de los parámetros que definen el fenómeno a la vez que proporcionar una medida de la exactitud de dicha estimación. Este último punto no es contemplado por ningún otro método de estimación y el semivariograma es la herramienta que puede ser utilizada pa-

ra evaluar la validez de cualquier estimación lineal a través de la varianza de la estimación.

Consideremos la estimación del valor medio de nuestra variable regionalizada sobre un dominio o soporte V que representaremos por $z_v(x_i)$ (x_i puede ser el centro de gravedad del soporte y para aligerar la notación se omitirá salvo que sea imprescindible). Para esta estimación utilizaremos los valores medios $\{z_v(x_i), i = 1, \dots, n\}$ en n soportes v_i . Un ejemplo podría ser la estimación de la precipitación media en una cuenca a partir de los valores medidos en n estaciones donde el soporte V sería la totalidad de la cuenca y el soporte v_i sería el área del pluviómetro que, en este caso, es tan pequeña respecto a V que podría considerarse como puntual. El valor estimado, que denotaremos con un asterisco como superíndice, será una cierta función de los datos dependiendo del método de estimación escogido (media aritmética, media geométrica, polígonos de Thiessen o cualquier otro) $z_v^* = f(\{z_v(x_i), i = 1, \dots, n\})$. El error cometido en esta estimación vendrá dado por la diferencia $z_v - z_v^*$ y se mantendrá desconocido. Ahora bien, puesto que z_v^* es una función de z_v podemos considerar que es una realización de la variable aleatoria z_v^* definida por la misma función anterior en términos de z_v . Y, a partir de ésta, definir una nueva función aleatoria $R(x) = Z_v(x) - Z_v^*(x)$ de la que podremos inferir su distribución de probabilidad o al menos sus dos primeros momentos.

Consideremos el caso en que la función f que proporciona el valor en el punto a estimar a partir de los datos medidos cumple los siguientes requisitos:

- a) Es insesgada, esto es

$$E\{Z_v^*\} = E\{Z_v\} \quad [9]$$

- b) Es lineal, esto es

$$Z_v^* = \sum_{i=1}^n \alpha_i Z_v(x_i) \quad [10]$$

calculando el valor esperado en (10) y recordando que los subíndices v y V sólo hacen referencia al valor medio de la variable Z en el corres-

pendiente soporte, se llega a que para que la condición de sesgo nulo (9) requiera.

$$\sum \alpha_i = 1 \quad [11]$$

La condición de sesgo nulo se traduce en que la esperanza matemática de los residuos, $R(x)$, es nula y el hecho de que escojamos funciones lineales hace que sea sencillo obtener una expresión analítica de la varianza de dichos residuos que llamaremos varianza de la estimación (σ_E^2). Tras algunos cálculos se llega a:

$$\begin{aligned} \sigma_E^2 &= E \{ [Z_v - Z_v^*]^2 \} = \\ &= E \{ [Z_v - \sum_i \alpha_i Z_v(x_i)]^2 \} = \\ &= 2 \sum_i \alpha_i \bar{\tau}(v_i, V) - \bar{\tau}(V, V) - \\ &\quad - \sum_i \sum_j \alpha_i \alpha_j \bar{\tau}(v_i, v_j) \end{aligned} \quad [12]$$

donde $\bar{\tau}(v, V)$ representa el valor medio del semivariograma $\tau(h)$ cuando un extremo del vector h se mueve sobre v y el otro extremo se mueve sobre V .

Esta expresión, independiente del método de estimación usado siempre que ésta sea lineal y cumpla la condición (11), nos da una medida de la confianza que podemos depositar en nuestra estimación. Asimismo, incluye cuatro esenciales factores que condicionan la calidad de cualquier estimación:

- Se tiene en cuenta la geometría del soporte a estimar V , a través del segundo término: ya que el semivariograma es una función creciente con h , el valor medio sobre un soporte V será creciente con el tamaño del soporte. Intuitivamente cuanto mayor sea el soporte V , más fácil debe ser la estimación de su valor medio puesto que las irregularidades habrán quedado promediadas y tenderá a estar próximo al valor medio de Z .
- Se tiene en cuenta la posible redundancia del conjunto de datos que se utilizan en la estimación a través del último término. Cuanto más próximos se encuentren entre

sí los datos menos información pueden proporcionar lo cual corresponde a pequeños valores de $\tau(v_i, v_j)$, mientras que si se encuentran repartidos, la variabilidad global del fenómeno queda mejor representada.

- Igualmente, se tiene en cuenta la posición relativa de los datos respecto al soporte a estimar, es claro que cuanto más cerca estén los datos del punto que queremos estimar más confianza debemos depositar en la estimación. Esto queda reflejado en el primer término que puede interpretarse como la distancia estructural entre los datos y el punto a estimar.
- Y finalmente también se tiene en cuenta características estructurales tales como anisotropía o grado de regularidad que hace que de dos datos a la misma distancia del punto a estimar, el que se encuentre en la dirección de menor variación aportará mayor información.

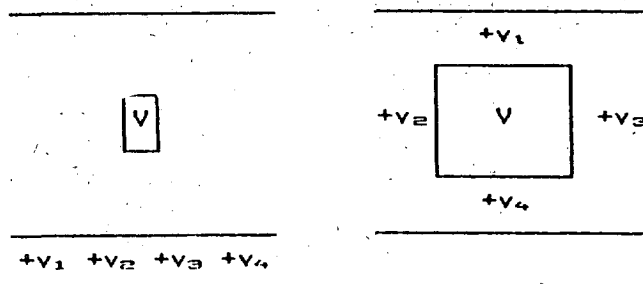


Figura 7. — Varianza de la estimación. En la estimación del bloque V a partir de los bloques v , la varianza de la estimación será mucho menor en el caso de la derecha puesto que su cálculo incluye factores como tamaño de bloque a estimar, posición relativa de los datos entre ellos mismos, posición relativa entre los datos y el bloque a estimar y anisotropía.

En la Figura 7 tenemos dos posibles conjuntos de datos para estimar el valor medio en el bloque V , suponiendo que se escojan los mismos coeficientes en la combinación lineal que da el valor del estimador en V , el conjunto de la derecha proporcionará una estimación con más confianza por las siguientes razones (1) al ser el bloque V de mayor tamaño la variabilidad de su valor medio será menor, (2) al estar los datos esparcidos alrededor del bloque proporcionan más información en cuanto a la variabilidad del fenómeno, (3) al estar los datos más

próximos al bloque V la información es de mayor calidad y finalmente, (4) si admitimos que la variabilidad en sentido horizontal es mucho menor que en sentido vertical los datos situados en aquella dirección de nuevo proporcionan mayor información.

VARIANZA DE LA DISPERSION

Otro concepto importante introducido por la Geoestadística es el de la varianza de la dispersión. Supongamos un acuífero discretizado en n celdas, donde en cada una de las celdas tenemos un número de medidas puntuales y a cada celda le hemos asignado el valor medio de los puntos que se encuentran en ella. Intuitivamente es claro que los valores puntuales $z(x_i)$ van a presentar una mayor dispersión respecto al valor medio del acuífero z_v , que la que tendrán los valores medios obtenidos en cada celda $z_v(x_i)$. Precisamente, el valor esperado de esta dispersión, directamente relacionada con el tamaño de los soportes, es lo que mide la varianza de la dispersión.

Por definición, varianza de la dispersión de las unidades v en el soporte V es:

$$D^2(v/V) = E \left\{ \frac{1}{V} \int_v [Z_v - Z_v(x_i)]^2 dx \right\} \quad [13]$$

y que bajo la hipótesis de débil escionaridad de segundo orden puede expresarse en función del semivariograma medio como:

$$D^2(v/V) = \bar{\tau}(V, V) - \bar{\tau}(v, v) \quad [14]$$

donde recordamos que $\bar{\tau}(v, V)$ es el valor medio del semivariograma $\tau(h)$ cuando un extremo del vector h recorre v y el otro recorre V . Esta expresión recoge lo que intuitivamente esperábamos, cuanto mayor sea el área V , la dispersión de los valores que encontraremos en él será mayor, y cuanto mayor sean los soportes v los valores medios obtenidos sobre éstos se aproximarán más al valor medio sobre todo el área.

Es inmediato demostrar, si admitimos la hipótesis de ergodicidad, esto es, que el valor medio de cada una de nuestras variables aleato-

rias puntuales puede obtenerse como el valor medio de una sola realización de nuestra función aleatoria sobre un soporte infinito, que la varianza de la dispersión de los valores puntuales ($v=0$) sobre un soporte infinito ($V=\omega$) coincide con la varianza de las variables aleatorias puntuales, si existe:

$$D^2(0/\omega) = \text{Var} \{ Z(x) \}$$

(si el límite de la expresión (13) cuando $V \rightarrow \infty$ existe) y utilizando (14), puesto que por definición $\bar{\tau}(0/0) = \tau(0) = 0$, deducimos que

$$\bar{\tau}(\infty, \infty) = \text{Var} \{ Z(x) \}$$

REGULARIZACION

En todas las expresiones anteriores, el semivariograma se supone calibrado a partir de valores puntuales de las variables. En la realidad nunca disponemos de valores puntuales, la transmisividad obtenida de un ensayo de bombeo representa el valor medio de la zona de influencia del ensayo, o la porosidad obtenida de una muestra es un valor medio de la del testigo, incluso la altura de lluvia medida en un pluviómetro es el valor medio de la lluvia recogida sobre el área que ocupa dicho pluviómetro. A estos valores medios se les denomina valores regularizados de la variable Z sobre el soporte v . Por lo general, los soportes de nuestros datos son tan pequeños respecto al área de estudio que pueden considerarse puntuales pero no siempre es así y conviene establecer la relación existente entre el valor del semivariograma obtenido a partir de datos puntuales y el obtenido a partir de valores medios, regularizadores, sobre un soporte v : (para mayor detalle vs. Journel y Huijbregts, (1978)).

$$\bar{\tau}_v(h) = \bar{\tau}(v, v_h) - \bar{\tau}(v, v)$$

donde τ_v es el semivariograma calculado cuando se utilizan los valores medios sobre un soporte v , v_h representa el soporte v trasladado respecto a sí mismo el vector h y $\bar{\tau}$ el valor medio del semivariograma puntual cuando cada extremo del vector describe v y v_h respectivamente. Para valores de h mucho más grandes

que v , $\bar{\tau}(v, v_h)$ puede aproximarse por $\tau(h)$ quedando la expresión anterior:

$$\tau_v(h) = \tau(h) - \tau(v, v) \text{ para } h \gg v$$

KRIGEAGE

La posibilidad de poder calcular la varianza de la estimación (12) en función de los parámetros utilizados en la combinación lineal (10) nos permite optimizar el valor de estos parámetros de tal manera que dicha varianza sea mínima. El estimador así obtenido ha sido llamado el mejor estimador lineal insesgado (si los parámetros cumplen la condición (11)), pero hay que matizar que es el mejor estimador cuando penalizamos los cuadrados de las diferencias. Esto supone que nuestra función de pérdidas por desviaciones del valor real es simétrica, puesto que bajo este criterio, tan mala es una estimación que quede por debajo, como una que quede por encima. En general esto no es así, no tiene igual efecto una sobreestimación de recursos que una infraevaluación de los mismos y por ello es posible que fuera más interesante tratar de minimizar las diferencias pero con una función asimétrica según estemos más interesados en sobreestimar o al contrario. A este respecto puede consultarse una reciente publicación de Journel (1984) en la que propone otro criterio, distinto de la varianza de la estimación, para valorar la validez del estimador.

La minimización de la ecuación (12) bajo la condición (11) respecto a los parámetros α , utilizando el método de los multiplicadores de Lagrange, lleva al siguiente sistema de $n + 1$ ecuaciones:

$$\left[\begin{array}{l} \sum_{j=1}^n \alpha_j \tau(v_i, v_j) + \mu = \tau(v_i, V) \quad i = 1, \dots \\ \sum_{j=1}^n \alpha_j = 1 \end{array} \right] \quad [15]$$

y el valor mínimo de la varianza de la estimación que llamaremos varianza del krigage, σ_k^2 vale,

$$\sigma_k^2 = \sum_{i=1}^n \alpha_i \bar{\tau}(v_i, V) + \mu - \bar{\tau}(V, V) \quad [16]$$

donde μ es el multiplicador de Lagrange utilizado en la optimización.

El sistema de ecuaciones así obtenido tiene solución y es única si el menos semivariograma puntual $-\tau(h)$ es condicionalmente definido positivo, esto es, cumple las condiciones (7) y (8), y no hay dos datos que tengan exactamente el mismo soporte, aunque sí puede ocurrir que los soportes presenten solapamiento. Esta condición implica además que la varianza del krigage sea no negativa.

El estimador proporcionado por (10) utilizando las soluciones del sistema (15), además de ser insesgado, es exacto, esto es si el soporte V coincide con alguno de los soportes de los datos el valor estimado resulta exactamente el valor conocido de dicho dato y la varianza del krigage es nula.

El sistema (15) y la varianza (16) dependen sólo de la geometría relativas de los diferentes soportes v_i , v_j y V , siendo independiente de los valores actuales de los datos. Este punto es de particular interés en el diseño de redes de control puesto que antes de instalar un nuevo punto de medida podemos determinar cuál será la disminución de la varianza de la estimación, pudiendo escoger entonces, aquel punto que proporcione la mayor disminución. Por contra, el hecho de no considerar los valores de los datos puede suponer que se le dé un peso inapropiado a los datos singulares que corresponden a valores extremos de la función de distribución y que por tanto no son tan representativos del comportamiento general del fenómeno.

KRIGEAGE UNIVERSAL

No siempre es posible considerar que el fenómeno que tratamos estudiar está representado por una función local débilmente estacionaria sino que hay que considerar el hecho de que exista una deriva, es decir que el valor medio de la función aleatoria sea a su vez una variable regionalizada dependiente de las coordenadas x , como es el caso de las alturas piezométricas en un acuífero considerado a nivel regional o en una zona escarpada. Si

$$E\{Z(x)\} = m(x)$$

se recurre a descomponer la función aleatoria en dos términos, la deriva $m(x)$ propiamente dicha y un residuo $Y(x)$ que tendrá media nula.

$$Z(x) = m(x) + Y(x) \text{ con } E\{Y(x)\} = 0$$

Si la deriva es conocida el problema se reduce a efectuar el kriging sobre los residuos y posteriormente añadirle el valor de $m(x)$. Pero generalmente la deriva es desconocida, en este caso supondremos conocida la forma de la misma y nos limitaremos a estudiar el caso en que esta deriva sea polinómica es decir

$$E\{Z(x)\} = m(x) = \sum_{i=1}^k a_i f_i(x) \quad [17]$$

donde $f_i(x)$ son monomios conocidos expresados en términos de las coordenadas del vector x , y los coeficientes a_i son desconocidos.

En este caso, el semivariograma no puede ser estimado mediante la ecuación (4) puesto que la varianza de las diferencias ya no es la esperanza del cuadrado de las diferencias como se obtuvo en (2) al no cumplirse (1), sino que es la esperanza del cuadrado de las diferencias de los residuos:

$$\begin{aligned} 2\tau(h) &= \text{Var}\{Z(x+h) - Z(x)\} = \\ &= E\{[Z(x+h) - Z(x)]^2\} - [m(x+h) - \\ &\quad m(x)]^2 = \\ &= E\{[Z(x+h) - Z(x) - (m(x+h) - \\ &\quad - m(x))]\}^2\} = \\ &= E\{[Y(x+h) - Y(x)]^2\} = \\ &= \text{Var}\{Y(x+h) - Y(x)\} \end{aligned}$$

Entonces la expresión (4) del semivariograma experimental puede utilizarse con los residuos $Y(x)$, pero puesto que la deriva es desconocida los residuos lo son también y el semivariograma experimental no puede ser calculado. Una aproximación consiste en extraer el semivariograma a partir de datos próximos que pue-

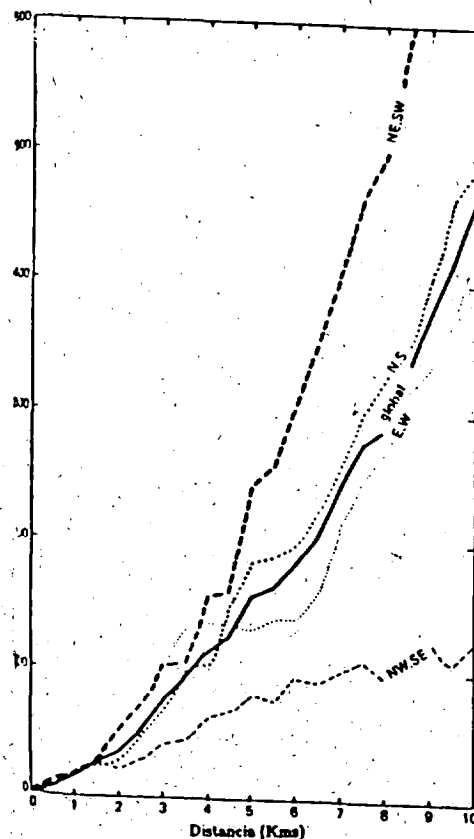


Figura 8. — Cálculo del semivariograma experimental en un fenómeno con deriva. El aspecto parabólico del semivariograma en la dirección noreste-suroeste denota deriva lineal en esa dirección por lo que el semivariograma noreste-suroeste puede considerarse representativo de los residuos.

dan considerarse localmente estacionarios o si la deriva tiene una dirección preferencial (como sería el caso de una deriva limitada a una única dirección) utilizar parejas de datos escogidas en la dirección perpendicular a la misma para las cuales la media es constante o, utilizar datos fuera de nuestra zona de interés pero que sabemos con certeza que poseen la misma estructura y presentan estacionaridad. Un ejemplo puede verse en la figura 8 tomada de Delhomme (1978), donde el semivariograma experimental obtenido en el acuífero de Crau en Francia para la dirección noreste-suroeste presenta una forma parabólica que revela la existencia de deriva correspondiendo a la dirección del gradiente, por ello, el semivariograma en la dirección perpendicular noreste-suroeste fue utilizado como semivariograma de los residuos al no existir deriva en esa dirección. Por último, añadir que no es recomendable aproximar la de-

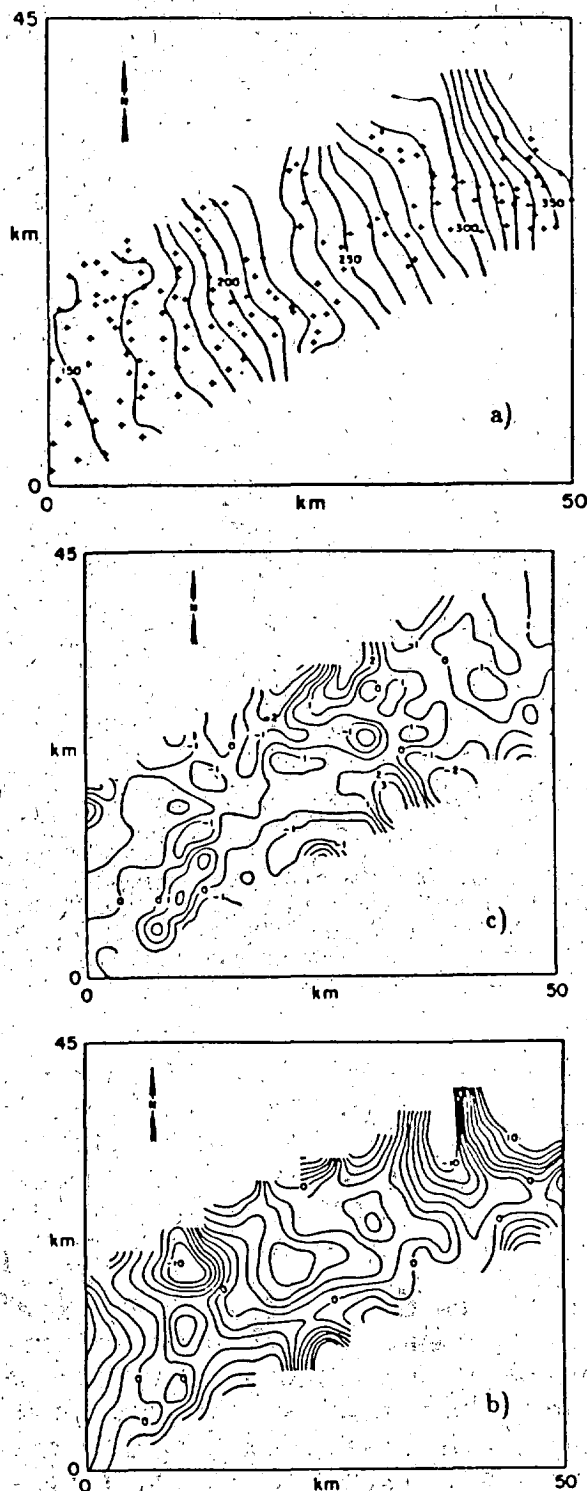


Figura 9. — Comparación entre los residuos tras remover la deriva calculada por mínimos cuadrados y tras removerla calculándola por krigage considerando como variograma experimental el obtenido con las parejas en la dirección perpendicular a la deriva. Ejemplo tomado de Aboufirassi y Mariño (1983) aplicado al acuífero de El Souss en Marruecos. a) Superficie piezométrica dibujada manualmente. b) Residuos tras restar la deriva obtenida por mínimos cuadrados. c) Residuos tras restar a la deriva krigéada.

riva por el resultado provisto por un ajuste por mínimos cuadrados porque los residuos así obtenidos quedan correlacionados con los valores experimentales y, por tanto, también presentan deriva, como ha sido comprobado experimentalmente por Aboufirassi y Mariño (1983) en el acuífero de El Souss en Marruecos y como puede verse en la figura 9 donde se muestran los mapas de residuos correspondientes a remover la deriva $m(x)$ obteniéndola por mínimos cuadrados u obteniéndola por krigage como se explicará más adelante haciendo uso del semivariograma experimental que se obtiene usando parejas de datos en la dirección perpendicular a la deriva.

Supuesto que conocemos nuestro semivariograma experimental al que habremos ajustado un semivariograma teórico y supuesto que conocemos la forma de la deriva, es decir cuáles son las funciones f_l que aparecen en la expresión (17), de nuevo se plantea el problema de estimar la combinación lineal de los datos que minimice la varianza y que sea insesgada.

La condición de no-sesgo que para el krigage ordinario era (11), ahora se transforma en el siguiente conjunto de condiciones que desarrollaremos para el caso en que los soportes v y V sean los puntos x_i y x , con el fin de reducir la complejidad de las notaciones.

$$\sum_{i=1}^n \alpha_i f_l(x_i) = f_l(x) \quad l = 1, \dots, k$$

y así, la minimización de la varianza de la estimación sujeta a las k anteriores condiciones lleva al siguiente sistema de ecuaciones:

$$\left. \begin{aligned} \sum_{j=1}^n \alpha_j \tau(x_i - x_j) + \sum_{l=1}^k \mu_l f_l(x_i) &= \tau(x_i - x) \\ i &= 1, \dots, n \\ \sum_{i=1}^n \alpha_i f_l(x_i) &= f_l(x) \\ l &= 1, \dots, k \end{aligned} \right\} [18]$$

y la varianza del krigage vale,

$$\sigma_k^2 = \sum_{i=1}^n \alpha_i \tau(x_i - x) + \sum_{l=1}^k \mu_l f_l(x)$$

Cabe añadir que el krigeage ordinario proporciona resultados igualmente satisfactorios cuando se pretende evaluar recursos globales sobre una zona o cuando los datos están homogéneamente distribuidos en torno al punto que se desea interpolar, aun cuando el fenómeno presente deriva. Estudios experimentales a este respecto han sido llevados a cabo por Buxton (1982).

La deriva puede ser estimada siguiendo el mismo método: si suponemos que la deriva en un punto puede ser expresada como combinación lineal de los datos experimentales

$$m^*(x) = \sum \alpha_i z(x_i)$$

puede obtenerse la expresión de la varianza de la estimación y las condiciones que deben cumplirse para que el estimador m^* sea insesgado. Después, la minimización de dicha varianza lleva al siguiente sistema de ecuaciones:

$$\left[\begin{array}{l} \sum_{j=1}^n \alpha_j \tau(x_i - x_j) + \sum_{l=1}^k \mu_l f_l(x_i) = 0 \\ \sum_{j=1}^n \alpha_j f_l(x_j) = f_l(x) \end{array} \right. \quad \left[\begin{array}{l} i = 1, \dots, n \\ l = 1, \dots, k \end{array} \right]$$

y la varianza en este caso vale,

$$\sigma_{km}^2 = \sum_{l=1}^k \mu_l f_l(x)$$

El lector interesado en la obtención de las anteriores ecuaciones y en el desarrollo del caso general en que los soportes v_i y V no son puntuales puede consultar Journel y Huijbregts (1978).

VALIDACION DEL SEMIVARIOGRAMA

Después de haber ajustado un semivariograma teórico a nuestro semivariograma experimental y de haber decidido cuál es la expresión polinómica que representa la deriva hay que comprobar que ambas suposiciones son válidas. Uno de los métodos utilizados es el de valida-

ción cruzada que consiste en estimar el valor predicho por krigeage en cada uno de los puntos conocidos utilizando el resto de los mismos en la estimación. El análisis de los residuos entre el valor predicho por krigeage y el valor conocido, en busca de desviaciones sistemáticas y comprobando la consistencia de estas desviaciones con la varianza asociada a cada una de las estimaciones, nos llevará a decidir si el modelo es válido o es necesario reajustar alguno de los parámetros.

SIMULACION CONDICIONAL

El krigeage no es una herramienta que tenga como objetivo simular la realidad sino que es un estimador local que pretende obtener aquel valor que minimiza la varianza de la estimación y que resulta ser el valor esperado de la variable en el punto a estimar dados los datos de que se dispone. Por ello cuando, a partir de un conjunto de datos, se obtienen por krigeage los valores que minimizan la varianza de la estimación sobre un mallado que cubre el área de estudio, el resultado es un mapa que presenta pocas irregularidades, suave, y que, de hecho, puede diferir de la realidad en una cantidad directamente relacionada con la varianza de la estimación.

Quando el interés no es visualizar el valor medio esperado sino la variabilidad general del fenómeno hemos de recurrir a las simulaciones condicionadas (Journel, 1974).

Al comienzo admitimos que el conocimiento actual de nuestro fenómeno correspondía a una de las posibles realizaciones de una función aleatoria definida en cada punto del espacio de trabajo. El objetivo de la simulación condicional es encontrar posibles realizaciones de la misma función aleatoria condicionadas a que en los puntos en los que disponemos de datos la simulación coincida con la realidad. Cada una de estas simulaciones mantendrá la misma estructura que los datos y coincidirá con ellos en los puntos en que se disponga de los mismos. Por esto, cada una de las simulaciones condicionales es una posible versión de la realidad desconocida. El valor medio de todas las posibles simulaciones tendrá el valor proporcionado por el krigeage y su varianza a la varianza del krigeage.

Antes de explicar los fundamentos teóricos de la simulación condicional constataremos que la importante diferencia entre una simulación condicional y una incondicional es que ambas reproducen un fenómeno con la misma estructura espacial pero la primera, además, coincide con los datos experimentales en los puntos en que se obtuvieron. La simulación condicional puede reproducir no sólo la estructura espacial representada por el semivariograma teórico sino que puede reproducir cualquier otro parámetro, por el hecho de coincidir con los datos experimentales.

El principio básico de la simulación condicional es que el error cometido al estimar un valor por su valor krigado es ortogonal al propio valor krigado y puede obtenerse a partir de cualquier realización de la función aleatoria, en particular de cualquier simulación incondicional que presente la estructura definida por dicha función aleatoria. Formalmente esto puede expresarse en los siguientes términos:

Descompongamos el valor real en la suma del valor estimado por krigeage y el error cometido:

$$z(x) = z^*(x) + (z(x) - z^*(x))$$

sea z_s una realización simulada incondicionalmente sobre el área de estudio y sea $z_{s,k}^*$ el valor obtenido por krigeage a partir de los valores simulados condicionalmente en los mismos puntos en que se dispone de datos. Podremos efectuar la misma descomposición:

$$z_s(x) = z_{s,k}^*(x) + (z_s(x) - z_{s,k}^*(x))$$

puesto que los errores tienen la misma distribución (proviene de la misma configuración de los datos), y del hecho de que el error $Z - Z_k^*$ es ortogonal a Z_k^* , el error real desconocido puede ser reemplazado por el simulado $z_s - z_{s,k}^*$ (Journal (1974)) quedando

$$z_c(x) = z_k^*(x) + (z_s(x) - z_{s,k}^*(x))$$

El procedimiento, por tanto, será obtener los valores krigados a partir de los datos reales, realizar una simulación incondicional usando uno de los métodos propuestos por Journal (1974) o por Clifton y Neuman (1982), obtener los valores krigados en la simulación incondi-

cional a partir de las posiciones de los datos experimentales, obtener la desviación entre estos valores krigados y los valores incondicionalmente simulados y esta desviación añadirla a los valores krigados obtenidos a partir de los datos experimentales.

ALGUNAS APLICACIONES DE LA GEOESTADISTICA A LAS CIENCIAS DEL AGUA

Aunque la Geoestadística nació para ser directamente aplicada en minería, pronto encontró aplicaciones en otros campos en los que la naturaleza de los fenómenos en estudio aunque determinista es tan compleja que la aproximación estocástica es apropiada, como meteorología, hidrología, ingeniería forestal, batimetría o hidrogeología. Aquí examinaremos algunas de las situaciones en que se ha aplicado la geoestadística en hidrología y en hidrogeología.

Cartografía

La primera aplicación y más directa del krigeage es el dibujo de curvas de nivel aunque no es éste el propósito para el cual fue concebido. A partir de una red de observación se pueden obtener por krigeage los valores de la variable en estudio sobre un mallado tan denso como se desee así como la varianza de la estimación. Un ejemplo de lo que se puede obtener de este modo puede verse en la figura 10, tomada de Aboufirassi (1982) donde se muestran la piezometría obtenida por Krigeage y la varianza de la estimación en el acuífero del congado de Yolo en California. El mapa de varianza de la estimación no es más que una medida de la confianza que puede depositarse en los valores interpolados y puesto que el krigeage es exacto, en los puntos donde se encuentran los piezómetros la varianza de la estimación vale cero. Podemos ver que la máxima incertidumbre se encuentra en la zona oeste donde existen pocos piezómetros mientras que la acumulación de piezómetros en la zona central reduce la varianza de la estimación considerablemente. Otra característica que debe ser notada es el aspecto irrealmente exento de irregularidades del mapa krigado consecuencia de

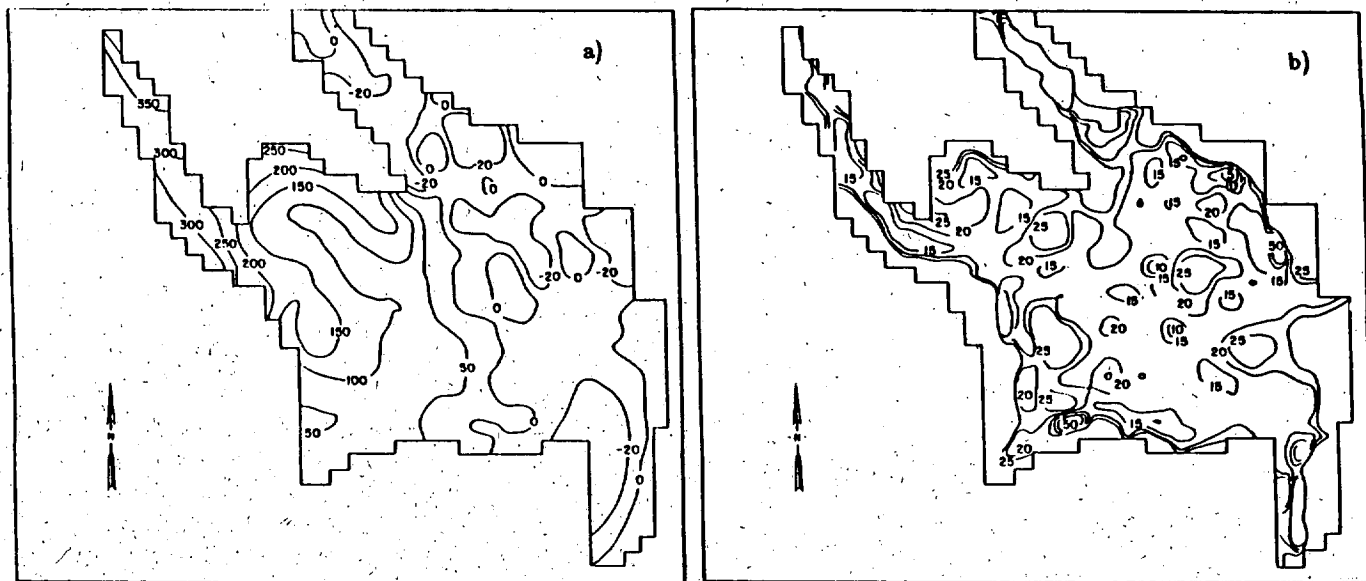


Figura 10. — Ejemplo de krigeage en el trazado de curvas de nivel tomado de Aboufirassi (1982) aplicado al acuífero del condado de Yolo en California. a) Mapa de alturas piezométricas krigeado. b) Mapa de varianzas de la estimación.

que, como ya se dijo, el krigeage no fue concebido con el propósito de reproducir el fenómeno a nivel general sino para obtener estimaciones locales, por ello todo mapa obtenido por krigeage debe ir acompañado del correspondiente mapa de varianza de la estimación que nos da una medida del rango de variación por-

sible que pueden presentar los valores reales sin perder la consistencia con los datos experimentales.

Evaluación de recursos medios

En lugar de interpolar los valores de la variable sobre una malla que cubra el área de estudio puede interesarnos obtener el valor medio sobre toda ella, como en el estudio de Delfiner (1975) en el que obtenía el valor medio de la pluviometría sobre la cuenca del ouadi Kadjemour, El Chad, 245 Km² a partir de los valores de 33 pluviómetros irregularmente distribuidos. En la figura 11 pueden verse los valores de los coeficientes de la combinación lineal obtenidos usando el método de los polígonos de Thiessen y krigeage resultando valores bastante similares por lo que en este caso la ventaja del krigeage no es evidente.

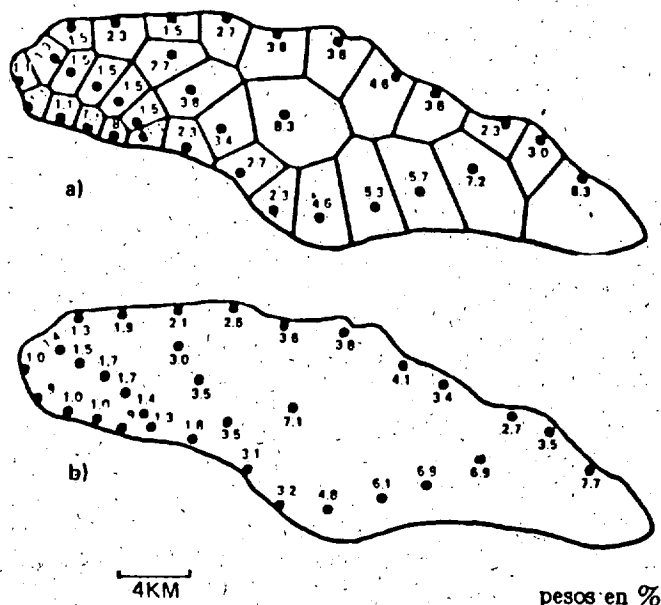


Figura 11. — Ejemplo de krigeage en la evaluación de recursos totales. Comparación de los pesos provistos por krigeage (b) y los provistos por polígonos de Thiessen (a) tomada de Delhomme (1978) en la cuenca de ouadi Kadjemour, El Chad.

Calibrado de modelos

Si asumimos que las desviaciones entre los valores obtenidos por krigeage y los valores reales están distribuidas normalmente, entonces el intervalo con el 95 por 100 de confianza será $[z_k - 2\sigma_k, z_k + 2\sigma_k]$. Así, a la hora de calibrar un

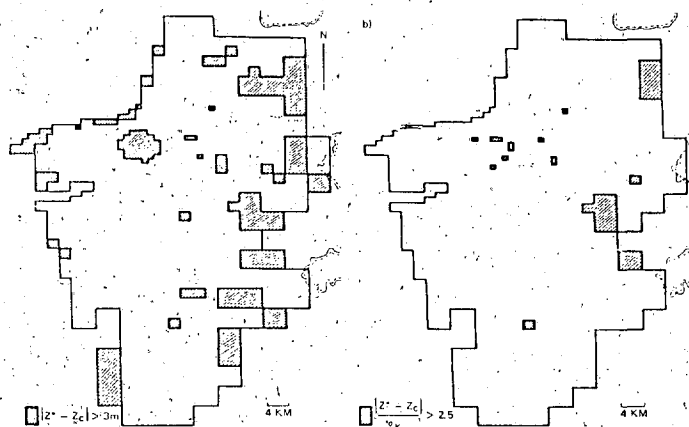


Figura 12. — Ejemplo de uso de krigage en la calibración de un modelo de flujo tomada de Delhomme (1978) en el acuífero de Kairouan. (Túnez). a) Errores en la calibración. b) Errores estandarizados (en las alturas piezométricas).

modelo podemos obtener el mapa de alturas piezométricas por krigage a partir de los puntos de medida con su correspondiente mapa de desviaciones típicas, y asimismo, obtener el mapa de alturas piezométricas a partir del modelo a calibrar. El resultado del modelo no debería, con un 95 por 100 de confianza, presentar desviaciones de los valores krigeados en más de dos veces la desviación típica del krigage, lo que equivale a intentar conseguir una simulación más perfecta junto a los puntos de medida y aceptable donde la información es poca. La figura 12, tomada de Delhomme (1978) y correspondiente al acuífero de Kairouan, Túnez, muestra las zonas donde los errores de calibración son mayores de tres metros y aquellos puntos donde estos errores superan en más de dos veces y media la desviación típica provista por el krigage. Es en estos puntos donde hay que tratar de ajustar el modelo, aun cuando existen puntos en el Este con errores absolutos superiores pero donde el desconocimiento de la realidad hace que el modelo pueda considerarse satisfactorio.

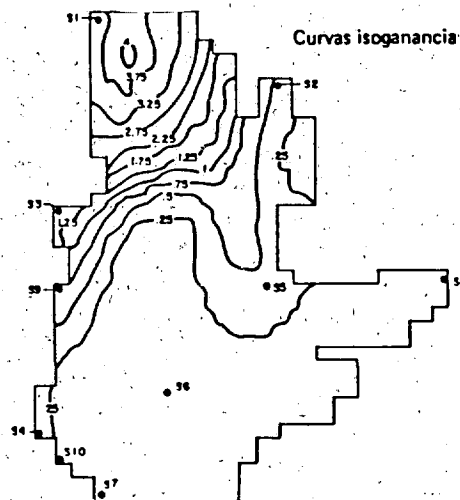
Diseño y optimización de redes de medida

Dada una red de medida podemos estar interesados en reforzarla añadiendo nuevos puntos. El criterio a utilizar dependerá de los objetivos perseguidos y en grandes líneas podemos aproximarnos al problema desde dos perspectivas: intento de minimizar la varianza del error

cometido en una estimación global, como podría ser el caso de la evaluación de la precipitación total en una cuenca o minimizar la varianza del error en la estimación local de los valores de la variable sobre un mallado como podría ser el caso de obtener los mejores valores de la transmisividad en cada una de las celdas del modelo de un acuífero.

A continuación se presentan los resultados obtenidos por Villeneuve et al (1979) en el estudio que llevaron a cabo para el refuerzo de la red de medida de caudales en la provincia de Quebec, Canadá. La variable regionalizada que usaron fue el caudal específico (caudal medido dividido por el área de la cuenca de la que procede) con 100 años de periodo de retorno. El problema puede abordarse de dos modos, dada una red de medida, localizar el punto que produjera la mayor reducción en la varianza de la estimación del valor medio de la variable regionalizada sobre toda la provincia o localizar aquellos puntos que produjeran las mayores reducciones en el mapa de desviaciones típicas asociadas a la estimación por krigage de la variable regionalizada en cada uno de los puntos de la provincia.

Para conseguir el primer objetivo se utiliza lo que se conoce como el método del punto ficticio que consiste en recorrer el área de estudio



suponiendo que en cada punto se localizará la nueva estación y obteniendo la reducción de la varianza respecto a la situación anterior. Esto da lugar a unas curvas de isorreducción de la varianza como pueden verse en la figura 13 donde la reducción está expresada en tanto por ciento respecto a la varianza de la red actual. Como ya se dijo, no necesitamos conocer el valor actual de la variable en los puntos de medida para obtener la varianza, ya que ésta sólo depende de la geometría relativa lo cual es una ventaja siempre que los datos no presenten grandes irregularidades o valores extremos; en otro caso debería calcularse la varianza condicionada a los datos para lo que sí necesitaríamos conocer los datos en los puntos de medida propuestos. La localización de la estación que proporciona mayor reducción en la varianza se encuentra en la zona norte mientras que cualquier estación que se añadiera en la zona sur no aportaría información relevante en la estimación del valor medio de la variable regionalizada. Estos resultados tienden a reducir los pesos de las estaciones que más contribuyen a la estimación intentando equilibrar las informaciones aportadas por cada una. Un último comentario que debe hacerse es que, si en lugar de uno, se desea introducir dos nuevos puntos de medida la mejor localización no necesariamente debe tener un punto coincidente con el máximo antes calculado. El problema de obtener la situación de un determinado número de

estaciones con el criterio de minimizar la varianza de la estimación en un solo paso ha sido resuelto satisfactoriamente por Carrera et al (1984) y Chou (1986) quien integrando geostatística con programación no lineal ha desarrollado un algoritmo para la óptima localización conjunta de N puntos de medida.

En el caso de estimación local, el método del punto ficticio no puede aplicarse, puesto que cada punto añadido modifica las varianzas de la estimación de los puntos que se encuentran alrededor. La búsqueda, en tal caso, no puede ser sistemática, Villeneuve et al. (1979) a la vista del mapa krigado y de las desviaciones típicas asociadas al mismo (figura 14), deciden añadir 10 estaciones en los puntos S1-S10 y observar el mapa de desviaciones típicas que se obtendrían en tal caso. La conclusión es que las mayores reducciones se obtienen en las zonas donde se añaden las estaciones S1 y S3 y que la mejora introducida por S4, S6, S7, S8 y S10 es prácticamente despreciable. Un último cálculo hecho añadiendo S2, S5 y S9 tras haber hecho los cálculos con S1 y S3 incorporados a la red denotó que tampoco representan una gran mejora en el mapa de desviaciones típicas.

Simulación

En la cuenca del ouadi Kadjemour donde las lluvias son escasas se llevaron a cabo simula-

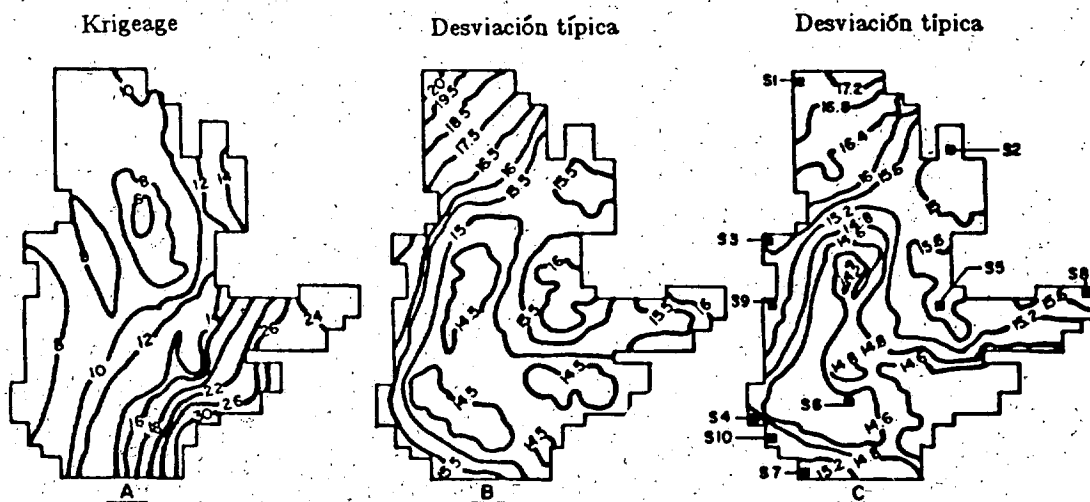


Figura 14. — Ejemplo del uso del krigage en la optimización de redes de optimización tomado de Villeneuve (1979) aplicada al mismo caso que la figura anterior. a) Mapa de caudales específicos con período de retorno de 100 años obtenido por el krigage. b) Mapa de desviaciones típicas asociado con el anterior. c) Mapa de desviaciones típicas si se añaden las estaciones S1-S10.

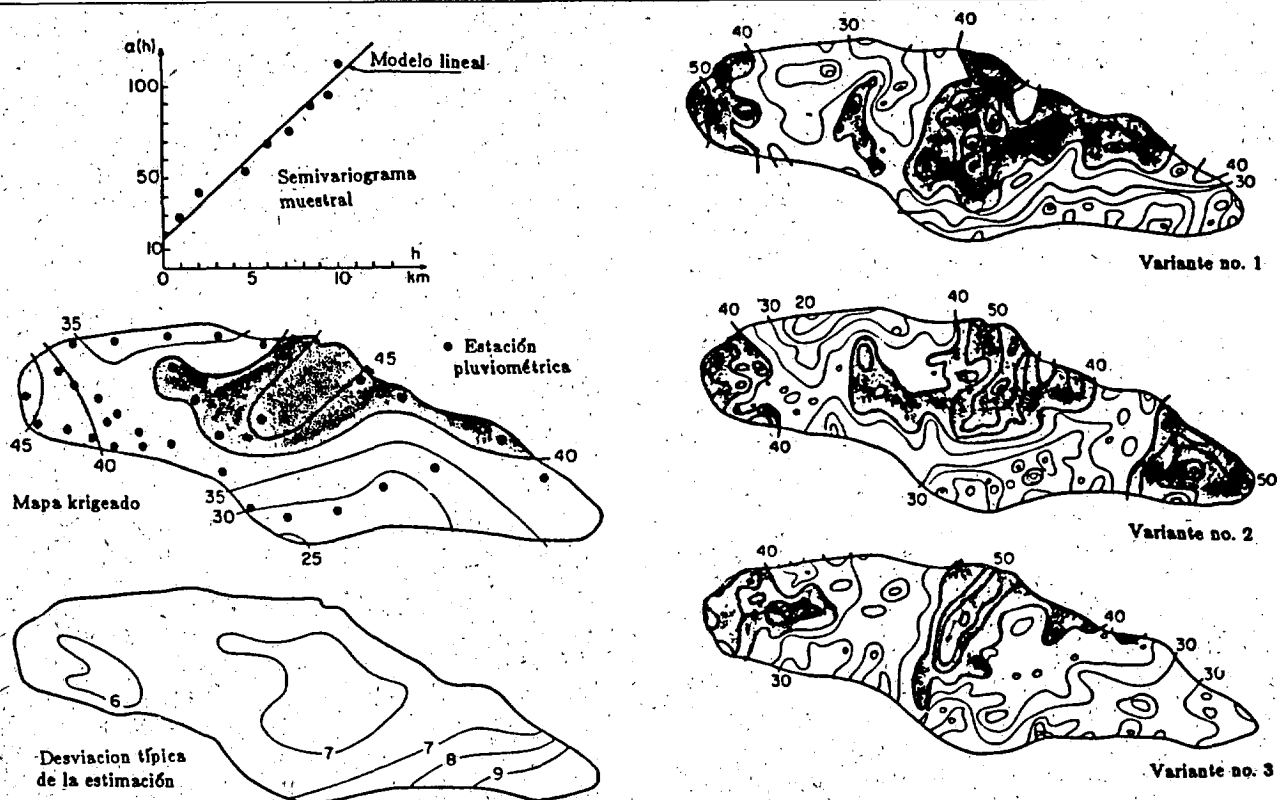


Figura 15. — Simulación de tres aguaceros tomado de Journel (1974) aplicado a la cuenca del ouadi Kadjemeur, El Chad (las zonas sombreadas corresponden a lluvia superior a 40 mm).

ciones condicionadas con el interés de conocer posibles distribuciones de aguaceros similares a uno de los recogidos en dicha cuenca con el fin de evaluar la que podría ser la verdadera distribución de la lluvia (Journel, 1974). En la figura 15 pueden verse el variograma ajustado a los 33 valores de pluviometría medidos en una lluvia, el mapa krigado y el mapa de desviaciones típicas, donde la zona sombreada representa lluvia superior a 40 mm, y el resultado de tres simulaciones condicionadas que corresponderían a aguaceros con la misma correlación espacial que el experimental y que tienen los mismos valores en los pluviómetros. Lo primero que debe notarse es el contraste entre el mapa krigado y las simulaciones, cualquiera de las cuales podría corresponder a la verdadera distribución del aguacero, o en otras palabras, la varianza de la dispersión de los valores krigados es mucho más pequeña que la varianza de la dispersión de los valores simulados. También debe notarse la gran dispersión que puede existir entre los valores de las distintas simulaciones en aquellas zonas donde la varianza del krigage es grande, como es el caso del

extremo este de la cuenca. Los estudios de recursos fueron llevados a cabo sobre las pluviometría simuladas en vez de sobre el mapa obtenido por krigage.

Otro estudio llevado a cabo por Delhomme (1979) utiliza las simulaciones condicionales para evaluar el efecto que, la variabilidad espacial y la incertidumbre en los parámetros que gobiernan el flujo en un acuífero, tienen en las simulaciones. El estudio se desarrolló sobre el acuífero Bathonian en la Baja Normandía en Francia, y consistió en obtener por krigage los valores de los logaritmos base 10 de la transmisividad (Neuman [1980] expone las razones por las que es preferible trabajar con los logaritmos de la transmisividad en lugar de con la transmisividad misma y que pueden resumirse en una mejor estructuración espacial), tras esto obtener el mapa de desviaciones típicas correspondiente que puede verse en la figura 16, a partir del mapa krigado realizar 50 simulaciones condicionales del campo de transmisividades y, finalmente, obtener las alturas piezométricas correspondientes a cada una de las simulaciones condicionales por medio de un

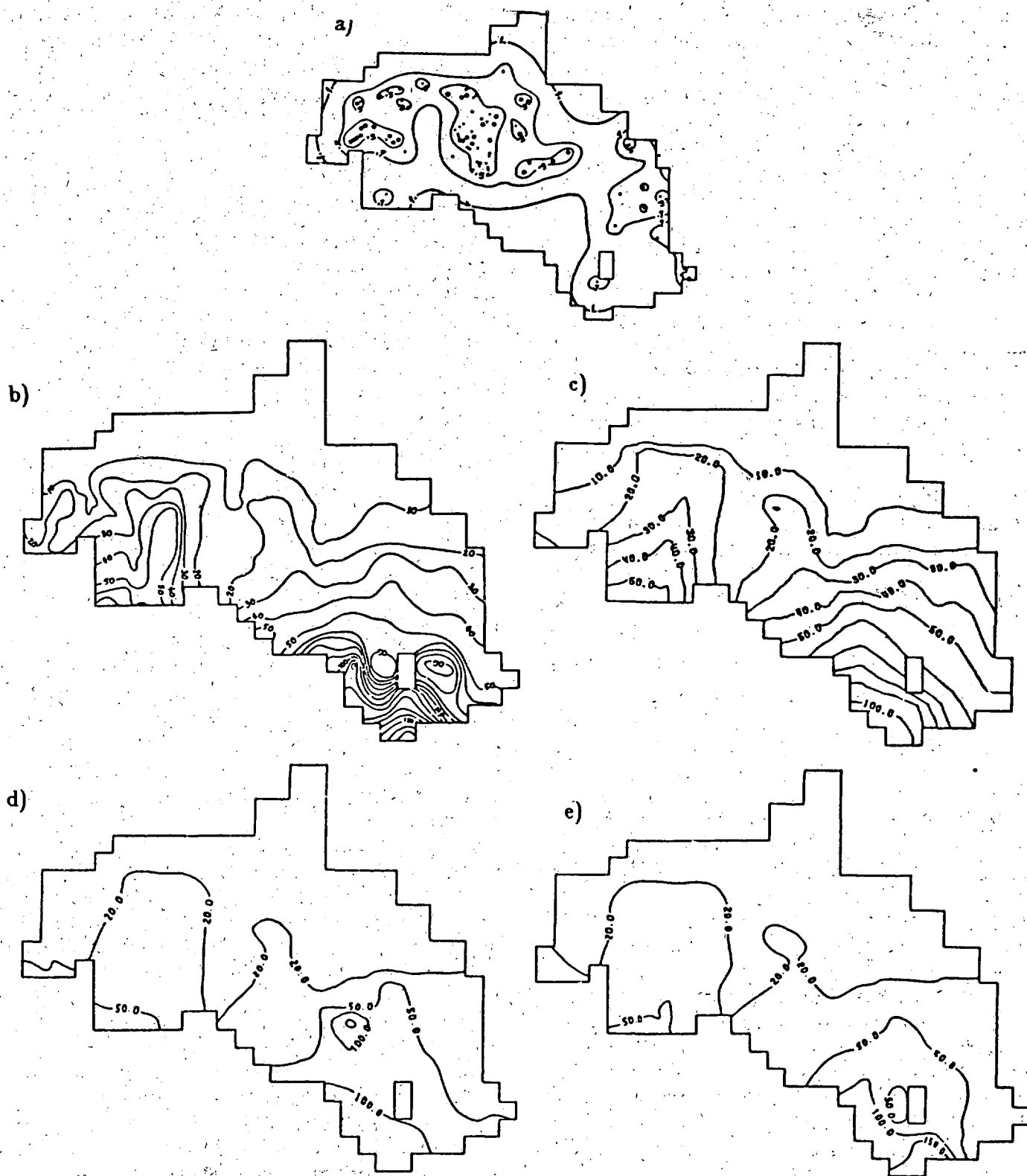


Figura 16.—Simulación condicional y modelos de flujo tomada de Delhomme (1979) aplicada al acuífero de Bathonian, Francia. a) Mapa de desviaciones típicas asociadas a la estimación por krigage de los logaritmos base 10 de la transmisividad. b) Mapa de alturas piezométricas reales. c) Mapa de alturas piezométricas obtenidas a partir de los valores de la trasmisividad krigeados. d) y e) Mapas de alturas piezométricas obtenidas a partir de dos de las simulaciones de las transmisividades.

modelo de flujo. En la figura 16 pueden también verse las alturas piezométricas reales, las calculadas a partir de los valores interpolados por krigage de los logaritmos de las transmisividades y las calculadas a partir de dos de las simulaciones de los logaritmos de las transmisividades. Las conclusiones obtenidas por Delhomme y que quedan de manifiesto en esta figura son, que los valores krigeados de $\log_{10} T$ son incapaces de reproducir los fuertes gradientes observados en la zona sur del acuífero, que algunas de las simulaciones reproducen con bastante aproximación la situación real probando la consistencia de los datos experimentales con la superficie piezométrica existente, mientras que, por el contrario, otras simulaciones presentan elevaciones o depresiones inaceptables lo que es consecuencia de la falta de datos suficientes y finalmente que la parte norte es siempre bien simulada debido a la imposición de potencial constante a lo largo del límite norte del acuífero, lo que indica que en las zonas donde existen restricciones a la cota piezométrica, es más importante calibrar estas restricciones que el campo de transmisividades en sí.

Simulación y el problema inverso

La sugerencia hecha por Delhomme (1979) al final del artículo citado anteriormente de que podrían obtenerse mejores resultados si las simulaciones estuvieran condicionadas no sólo a los valores experimentales de $\log_{10} T$ sino también a los datos de piezometría a través del problema inverso fue desarrollada por Neuman (1982) y aplicada por Clifton y Neuman (1982) en el acuífero del valle de Avra en el sur de Arizona donde consiguieron reducir la varianza de las alturas piezométricas obtenidas a partir de simulación incondicional del campo de transmisividades en un factor de 46 mientras que la máxima varianza fue disminuida en un factor de 68. Las reducciones obtenidas (siempre respecto a simulación incondicional) por el krigage y por el método inverso independientemente fueron en factores de 3.2 y 14.3 respectivamente. La combinación de krigage y el problema inverso aparece como una herramienta importante en la confección de modelos de flujo puesto que proporciona estimaciones del campo de trans-

misividades muy próximas a la realidad teniendo la ventaja de que no resulta tan afectado por la escasez de valores experimentales de T como le ocurre al krigage.

CONCLUSIONES

La Geoestadística ha entrado en el mundo del hidrólogo y principalmente en el del hidro geólogo como una herramienta más cuyo uso potencial todavía no ha encontrado límite. Algunas de las aplicaciones directas que se han llevado a cabo, con interesantes resultados, han sido en el trazado de mapas de curvas de nivel acompañados por mapas de isodesviación típica que dan una medida de la fiabilidad de los primeros; en la estimación de recursos medios o globales a partir de valores puntuales; en el calibrado de modelos; en el diseño y optimización de redes de control; y, finalmente, la simulación condicional ha permitido obtener representaciones del fenómeno en estudio que puedan reflejar la variabilidad real manteniendo la estructura espacial de los datos experimentales con múltiples propósitos como analizar la influencia de la densidad de datos en ciertas zonas o proveer una herramienta que chequee los resultados en los modelos de flujo. Las últimas aplicaciones llevadas a cabo consiguen obtener los campos de transmisividad con suficiente aproximación como para reducir al mínimo la incertidumbre en las alturas piezométricas simuladas. Queda así abierta una nueva división de la hidrogeología que contribuye y contribuirá a aumentar la confianza de los resultados obtenidos al conseguir aproximar más la realidad a los todavía hoy vistos con recelo modelos matemáticos.

AGRADECIMIENTOS

Este trabajo ha sido posible gracias a la Beca de Estancias Anuales en el Extranjero que fue concedida al autor por la Generalitat Valenciana en la convocatoria de 1985-86. El autor quiere agradecer asimismo los comentarios y sugerencias llevados a cabo por el profesor D. André Journel del Applied Earth Sciences Department de la Universidad de Stanford.

BIBLIOGRAFIA

ABOUFIRASSI, M. Identification of Transmissivities in Yolo Basin, California: A Geostatistical Approach. *Ph. D. Dissertation*. 286 p. Universidad de California, Davis, California, 1982.

ABOUFIRASSI, M.; MARIÑO, M. Kriging of water levels in the Souss Aquifer, Morocco. *Mathematical Geology*, 15(4), 1983.

BUXTON, B. E.; Coal Reserve Assesment: a geostatistical case studi. *Ms. Sc. Dissertation*. 84 p. Universidad de Stanford, California, 1982.

CHOUM, D.: Selecting optimum exploratory drill locationsby group. *Enviado a Mathematical Geology*, 1986.

CARRERA, Jesús; USUNOFF, E.; SZIDAROVSKY, F.: «A metod for optimal observation network design for groundwater management», *Journal of Hydrology*, Vol 73, p. 147-163, 1984.

CLIFTON, P. M.; NEUMAN, S. P. Effects of Kriging and Inverse Modelling on Conditional Simulation of the Avra Valley in Southern Arizona. *Water Resources Research*, 18(4), 1215-1234, Agosto 1982.

DELFINER, P. Recursos de agua y explotación de datos pluviométricos poco numerosos. *Aplicaciones de la Teoría de la Variable Regionalizada*. Cátedra de Matemáticas de la Escuela Técnica de Ingenieros de Minas. Madrid, 1975.

DELHOMME, J. P. Kriging in the Hydrosiences. *Advances in Water Resources*, 1(5), 251-266. 1978.

DELHOMME, J. P. Spatial Variability an Uncertainty in Groundwater Flow Parameters: A Geostatistical Approach. *Water Resources Research*, 15(2), 269-280, Abril, 1979.

JOURNEL, A. G. Geostatistics for Conditional Simulations of Ore Bodies. *Economic Geology*, 69, 673-687. 1974.

JOURNEL, A. G. mAD and Conditional Quantile Estimators, *Geostatistics for Natural Resources Characterization, Part. 1*, 261-270. D. Reidel Publishing Company, 1984.

JOURNEL, A. G.; HIJBREGTS, CH. J. *Mining Geostatistics*. Academic Press. London, 1978.

MATHERON, G. *Les Variables Régionalisées el Leur Estimation*. Masson. Paris, 1967.

MATHERON, G. Theory of Regionalizad Variables and its Applications. *Cahiers du Centre de Morphologie Mathématique 5*. Ecole de Mines. Fontanebleau, Francia, 1971.

NEUMAN, S. P. A Statistical Approach to the Inverse Problem or Aquifer Hydrology, 3, Improved Solution Method and Added Perspective. *Water Resources Research*, 16(2), 331-346, 1980

NEUMAN, S. P. Statistical Characterization of Aquifer Heterogeneties: An Overview, in Recent Trends in Hydrogeology. *Biological Society of America Special Paper*, 184, 81-102, 1982.

VILLENEUVE, J. P. ; BOBEE, B.; LEBLANC, D.; DELHOMME, J. Kriging in the Design of Streamflow Sampling Networks. *Water Resources Research*, 15(6), 1833-1840, 1979.

José Jaime Gómez Hernández



Es Ingeniero de Caminos por la Escuela de Ingenieros de Caminos de Valencia, y cursó toda la carrera como becario del Colegio Mayor San Juan de Ribera de Burjassot (Valencia). Número uno de su promoción y Premio Nacional de Terminación de Estudios, ha colaborado, siempre en temas directamente nacionales e internacionales entre la Universidad Politécnica de Valencia y, entre otros, las universidades estadounidenses de Fort Collins (Colorado) y Stanford (California), la Confederación Hidrográfica del Júcar, el Servicio Geológico de Obras Públicas y el Ayuntamiento de Valencia. En la actualidad, becado por la Conselleria de Cultura de la Generalitat Valenciana, finaliza sus estudios de *Master of Sciences* en Hidrogeología Aplicada en la Universidad de Stanford.

