

DESARROLLO ACTUAL DE LA GEOESTADÍSTICA EN EL MUNDO

Present development of geostatistics in the world

Adrián MARTINEZ-VARGAS¹, Jairo RAMÍREZ-GARCIA²

(1)Instituto Superior Minero Metalúrgico. Email:amvargas@ismm.edu.cu ,(2)Universidad de Antioquia, Colombia

RESUMEN-La geoestadística se emplea en innumerables campos de aplicación para estimar y simular variables regionalizadas. Fue creada en la década de los 60-70, pero su máximo esplendor como ciencia aplicada se alcanza a partir de los 90. Sus técnicas forman un largo listado, pero se pueden sistematizar separándolas en 5 grupos fundamentales: geoestadística lineal, no estacionaria, multivariada, no lineal y las simulaciones. En este trabajo se analiza cada uno de dichos grupos, así como el desarrollo alcanzado en los últimos años en relación con esta ciencia, considerada una de las más recientes ramas de la matemática aplicada..

Palabras clave: Geoestadística, desarrollo, matemática aplicada

ABSTRACT- Geostatistics are applied in many fields in order to simulate and estimate regionalized variables. Was born in 60s – 70s, but reach its actual magnitude in 90s. Exist a long list of techniques, which can be separated in 5 main groups: linear geostatistic, non stationary, multivariate, non linear and simulations. In this work is analyzed each one of those groups. Is also analyzed the development reached by this science, which is considered ones of the recent branches of the applied mathematics.

Key words: Geostatistics, development, applied mathematics

INTRODUCCIÓN

En los últimos 30 años la geoestadística ha probado su superioridad como método de estimación de los recursos de la mayoría de los tipos de minas. La aplicación en la industria del petróleo es más reciente, donde ha demostrado su utilidad para modelar y simular la compleja heterogeneidad interna de muchos reservorios. Su empleo también ha sido extendido a otros campos, tales como, el medioambiente la hidrogeología, la agricultura e incluso la pesca, donde el factor tiempo, al igual que la variabilidad espacial, juegan un papel importante (Armstrong M., 1998).

La herramienta básica, el variograma, no ha cambiado desde entonces; se emplea para cuantificar la correlación entre observaciones espacialmente distribuidas, y para deducir los estimadores de la variables aleatorias en puntos no muestreados, procedimiento que se conoce como krigage -kriging en inglés- en honor al ingeniero sudafricano Danie Krige, quien, junto Herbert Sichel llevaron a cabo los primeros pasos de la naciente ciencia en las minas de oro de Witwatersrand.

En los años 60 y 70 el matemático francés Georges Matheron, partiendo de las primeras publicaciones de Krige, comenzó un intenso trabajo para desarrollar las bases teóricas de la geoestadísticas. Es en el Centro de Investigación de Fontainebleau, creado para el por la escuela de Minas de París, asistido por un equipo de jóvenes investigadores, donde continua su trabajo dando el esplendor que hoy tiene esta ciencia y la Morfología Matemática, pero esa última forma parte de otra historia. Como se muestra más adelante muchas de las técnicas de reciente creación no son más que el perfeccionamiento de sus trabajos iniciales.

En la actualidad el número de técnicas recogidas bajo los términos krigage y simulaciones estocásticas a crecido de forma considerable, muchas de ellas resuelven problemas similares por vías diferentes, por ello quien se introduce en este mundo puede verse parado un una encrucijada, y la decisión de cual camino tomar puede tornarse en un problema más a resolver. Con el objetivo de orientar dicha decisión se analiza el estado actual de salud en que se encuentra esta ciencia y se hace un resumen de los principales grupos de técnicas, considerando aspectos teóricos y aplicabilidad en problemas prácticos diversos.

LA GEOESTADÍSTICA COMO CIENCIA

La geoestadística se encarga del estudio de muestras repartidas en el espacio a partir de modelos aleatorios. Su uso práctico sobrepasa el simple trazado de mapas, pues también tiene en cuenta el sentido espaciotemporal de la probabilidad. El modelo no es el objetivo como tal, este es realmente un útil al servicio de lo real observado, por lo tanto, más que una ciencia es una deontología del tratamiento de muestras, tomado de homenaje a Georges Materno y realizada por el Centro de Geoestadística de Fontainebleau.

Muchos autores se han esforzado para darle una definición, las que giran bajo el concepto de variables regionalizadas introducido por Matheron 1965, el cual enfatiza dos aspectos contradictorios: el aleatorio (que tiene en cuenta las irregularidades locales) y el estructural (que refleja las tendencias a larga escala). Chilès (1999, p. 2) la define como “la aplicación de los métodos probabilísticos a las variables regionalizadas”. Olea (1999 p. 1) la

considera como “un conjunto de técnicas que se auxilian de la caracterización espacial de los atributos, empleando modelos primarios aleatorios, de manera similar a las series temporales”. Quizás, la definición más acertada es “la ciencia que estudia las variables regionalizadas”, dada por Le Loc’h (2005). Una caracterización bastante aceptable es presentada por Dubois (1998); él destaca el papel del análisis espacial o temporal del fenómeno estudiado.

Un aspecto distintivo de esta ciencia es que asume que los valores de las variables regionalizadas están correlacionados unos con otros, el estudio de dicha correlación se conoce como “análisis estructural”. Los pasos básicos de la geoestadística como técnica son:

análisis exploratorio de los datos

análisis estructural (cálculo y modelado de los variogramas)

predicción (kriging o simulación)

La mayoría de los autores resaltan en sus publicaciones los conceptos:

Variable regionalizada

Función aleatoria (FA)

Hipótesis estacionaria e intrínseca

Función de Covarianza espacial

Variograma, experimental y modelos admisibles.

Anisotropía

Drift

Efecto proporcional

Efecto soporte y el soporte de una variable regionalizada.

Efecto información

Regularización

Teoría del krigeage

Otros, como la relación de aditividad de Krige (Armstrong, 1998, p. 77) o el teorema de la aditividad (Armstrong, 1998, p. 94) son solo tratados en algunas publicaciones. En ediciones más específicas aparecen conceptos menos citados aún, por ejemplo el teorema de la microergodisidad (Chilès, 1999, p. 20) y la teoría transitoria o geoestadística transitoria (Chilès, 1999, p. 24) donde la variable $z(x)$ se asume determinística y es idéntica a cero solo fuera de los límites del dominio D donde está definida. La geoestadística transitoria fue desarrollada por Matheron, 1965 y se emplea actualmente para estimar el error geométrico asociado estimaciones globales de los recursos, donde los límites del área mineralizada no se

conoce a priori (Armstrong, 1998, p. 134). Otro aspecto poco trabajado, pero que comienza a ganar en popularidad, es la representación espectral de las funciones aleatorias, las cuales tienen un uso práctico en señales temporales. En el espacio el significado físico de las frecuencias asociadas a las funciones aleatorias no es claro, pero representa una herramienta teórica poderosa, especialmente para simulaciones (Chilès, 1999, p. 17).

De esta se desprenden técnicas como el kriging de variables complejas (Wackernagel, 1998, p. 187) y el modelo bilineal de correogionalización, aplicable a la modelación de funciones de covarianzas cruzadas no simétricas (Wackernagel, 1998, p. 194). Un ejemplo de aplicación de la representación espectral (o fractal) de las funciones aleatorias es la simulación de funciones aleatorias basadas en el movimiento fractal browniano o el ruido gaussiano fractal, de gran utilidad para simular montañas, canales y lagos con características realistas. El poder de dicho modelo radica en que se puede determinar simultáneamente la varianza de una propiedad a diferentes escalas (Dubois, 1998)

PRINCIPALES TÉCNICAS GEOESTADÍSTICAS Y SUS PARTICULARIDADES

Se puede decir que la geoestadística es una ciencia bastante joven, por ello sistematizar su conocimiento a partir de una separación adecuada de sus principales principios y técnicas resulta imprescindible para su mejor estudio y comprensión. Para ello se adopta la secuencia propuesta por el Centro de Geoestadística de la Escuela de Minas de París, en el curso de formación especializada Cycle de Formation Spécialisé en Géostatistique:

Geoestadística lineal

Geoestadística no estacionaria

Geoestadística multivariada

Simulaciones

Geoestadística no lineal

Geoestadística lineal

La geoestadística lineal univariada está relacionada al estudio de variables regionalizadas que satisfacen la hipótesis estacionaria de segundo orden o la intrínseca. Su teoría no se ha enriquecido mucho desde su creación en los años 60 y 70, destacándose dos aspectos claves:

Las funciones de variabilidad espacial.

Los métodos de estimación.

Las funciones de variabilidad espacial están representadas por el variograma y la covarianza espacial centrada. En casos particulares pueden considerarse funciones especiales como: el covariograma transitivo, la covarianza no centrada, la covarianza no ergódica, el correlograma, el correlograma no ergódico, el madograma, el rodograma, el variograma relativo, el variograma relativo (pair wise) (Bleines et al., 2004).

Estas funciones pueden emplearse para orientar el análisis estructural o otro objetivo muy particular, por ejemplo Deutsch y Journel (1998, p. 45) plantean que el variograma general relativo y el relativo pair wise son resistentes a los datos espaciados y a los outliers (valores anómalos) cuando el histograma tiene asimetría positiva, por lo que revelan estructuras ocultas.

En la estimación solamente deben emplearse la función de covarianza y el variograma, pero en la práctica solo el segundo se calcula experimentalmente; para ello existen varios estimadores, el más usual es el mostrado por Chilès (1999, p. 37; Wackernagel, 1998, p. 45; Deutsch & Journel, 1998, p. 44) entre otros.

$$\gamma(h) = \frac{1}{2N} \sum_{x=1}^N [z(x+h) - z(x)]^2$$

Dicho estimador es una alternativa práctica para inferir el variograma regional, definido dentro del dominio D como (Chilès, 1999, p. 38):

$$\gamma_R(h) = \frac{1}{2|D \cap D_{-h}|} \int_{D \cap D_{-h}} [z(x+h) - z(x)]^2 dx$$

La imposibilidad de obtenerlo de forma experimental está dada por la limitada cantidad de muestras disponibles en Chilès (1999, p. 39) también muestra que el variograma experimental es un estimador pobre de la variabilidad espacial si el histograma se extiende mucho hacia sus extremos, como alternativa se pueden emplear estimadores más robustos, tales como el variograma mediana, cuantil, y Huverized, así como variogramas de orden 1/2.

Los modelos de estos variogramas tienen como propiedades más importantes, la continuidad y la diferenciabilidad en el origen, las que controlan el grado de continuidad de las realizaciones de la FA a corta escala. A larga distancia esto se traduce de forma diferente, se dice que una FA es estacionaria si el variograma es finito. Las funciones aleatorias intrínsecas son infinitas, pero debe cumplir la propiedad $\gamma(h)/|h|^2 \rightarrow 0$ cuando $|h| \rightarrow \infty$ (Chilès, 1999, p. 59), de lo contrario estamos en presencia de una FA intrínseca de orden k (IRF-k), la cual se encuentra en el dominio de las geoestadísticas no estacionarias.

Las funciones aleatorias intrínsecas deben ser condicionalmente definida positiva, o definida positiva para el caso de las estacionarias de segundo orden.

Un aspecto de vital importancia es el rango integral, definido como:

$$A = \frac{1}{\sigma^2} \int C(h) dh$$

Si este es finito, o al menos infinito y acotado, entonces la FA que caracteriza es ergódica (Chilès, 1999, p. 74; Bleines, et al., 2004, p. 567) muestran las expresiones de A para diferentes modelos de variogramas y dimensiones. Para ilustrar algunas de sus aplicaciones prácticas consideremos la expresión $P\{|Z_{V-m}| > 0.05\} > 0.9, |V| > 1076\sigma^2 A$, donde V es el volumen de la región simulada (Lantuéjoul, 2005), que permite determinar el tamaño del área, necesario para una precisión de la estimación de la media de una FA simulada, con un 95 % de confianza.

De forma similar se deduce el tamaño de la región simulada para diferentes precisiones de estimación de la varianza y el variograma regional. Otra implicación práctica derivada de estas ecuaciones es que se requiere que $V \gg A$ para que las propiedades del modelo original se mantengan en el modelo simulado y las variaciones ergódicas sean mínimas.

En cuanto a la anisotropía la teoría no ha cambiado mucho, pero la forma práctica de modelarla ha progresado considerablemente, Bleines, et al., 2004 muestran como definir las por estructuras, con rotaciones globales y locales, facilitando el ajuste de variogramas con comportamiento complejo y diferenciado por escalas. También muestran procedimientos de ajuste automático (p. 572), particularmente útiles para el caso de IRF-k y el multivariado.

Con las propiedades de la FA bien definidas, y si estas son adecuadas, se procede a la estimación. En el caso intrínseco se emplea el Krigeage Ordinario, en el estacionario de segundo orden se puede usar además el Krigeage Simple con media conocida a priori. Dada la dificultad de inferir la media, el krigeage simple es raramente utilizado; cuando se estiman datos transformados, por ejemplo por anamorfosis gaussiana, la media es conocida, por ello se considera como el método de base para la geoestadística no lineal (Armstrong, 1998, p. 94) y las simulaciones.

La comparación entre ambos métodos no es lineal a causa de la introducción de un término constante en el sistema de ecuaciones del krigeage ordinario. El krigeage simple también tiene la ventaja de poseer un criterio sólido para evaluar la calidad de la estimación conocido como el peso de la media (weight of the mean).

Dentro de estas técnicas también se encuentra el kriging de la media, el cual es un caso particular del kriging del drift usado en el contexto de universalidad o IRF-k de las funciones aleatorias. El teorema de la aditividad sirve para relacionar los tres estimadores antes mencionados (Armstrong, 1998, p. 94), de donde se desprenden dos conclusiones fundamentales:

La varianza de kriging ordinario puede ser fraccionada en dos partes, la varianza del kriging simple y la varianza del estimador de la media multiplicada por el peso de la media del kriging simple. De esta se deduce que hacer simple kriging con media local estimada a partir del kriging de la media es equivalente al kriging ordinario.

El kriging simple es condicionalmente insesgado, cosa que no se cumple para el kriging ordinario, en términos de probabilidades esto significa que $E[Z_v|Z_v^*] = Z_v^*$ y dicha regresión es lineal con pendiente igual a la unidad (Armstrong, 1998, p. 96). Esta propiedad tiene un impacto importante en la evaluación del error de la estimación de recursos y reservas. No obstante un punto a favor del Kriging Ordinario es que se adapta fácilmente las variaciones locales de la media.

El kriging de los indicadores, los logaritmos o kriging en el contexto multigaussiano no son más que el kriging ordinario o simple de datos no linealmente transformados. El kriging factorial, el cual filtra o estima los factores asociados a cada componente de los variogramas anidados, se puede clasificar como un método lineal y multivariado, pues se deduce a partir de la combinación lineal de los componentes o factores de la FA asociados a cada estructura. Vea en Goovaerts (1997, p. 158 a 198) que estos no son más que casos particulares de Kriging Ordinario.

El Enfoque Dual del Kriging, no es más que otra forma de representar las ecuaciones de kriging, se parte de un estimador expresado como combinación lineal de covarianzas en lugar de datos. Este permite filtrar componentes y disminuir el costo de cómputo con búsqueda global (Goovaerts, 1997, p. 169). Por supuesto existen equivalente de kriging simple, ordinario, kriging de la media y otros para kriging dual de factores.

Otro aspecto a destacar es la incorporación en los software de formas más complejas de soporte, como polígonos, y de mejoras en el cálculo de la varianza del bloque por método numérico, de los cuales el más aceptado es la discretización regular, con desplazamiento aleatorio de los puntos de discretización secundarios (Bleines, et al., 2004, p. 608)

Como aspecto novedoso podemos citar el empleo de derivadas asociadas a las ecuaciones de kriging con fines prácticos (Bleines, et al., 2004, p. 611) para la estimación del gradiente:

$$\begin{cases} \sum_{\beta} \lambda^{\beta} K_{\alpha\beta} + \sum_l \mu_l f_{\alpha}^l = \frac{\partial K_{\alpha 0}}{\partial u} \\ \sum_{\alpha} \lambda^{\alpha} f_{\alpha}^l = \frac{\partial f_0^l}{\partial u} \\ \sigma^2 = -\sum_{\alpha} \lambda^{\alpha} \frac{\partial K_{\alpha 0}}{\partial u} - \sum_l \mu_l \frac{\partial f_0^l}{\partial u} + \text{var} \left[\frac{\partial Z}{\partial u} \right] \end{cases}$$

donde $\left[\frac{\partial Z}{\partial u} \right]$ es igual a $-K''(0)$

Un caso más complejo será discutido en la sección de geoestadística multivariada

Cuando el error y su varianza se conoce en los puntos medidos entonces se puede emplear el Kriging con Error Medido (Bleines et al., 2004, p. 619). Este curioso método, aplicado muy pocas veces en la práctica por falta de información, solo se diferencia del kriging ordinario por tener un sistema de ecuaciones con covarianzas variables en la diagonal, que pueden interpretarse como efecto pepita variable por traslación.

Geoestadística no estacionaria

Las funciones aleatorias no estacionarias son el caso opuesto a las estacionarias, las intrínsecas ocupan una posición intermedia y son además un caso particular las IRF-k con k=0 (Chilès, 1999, p. 231). Una propiedad que caracteriza las funciones aleatorias no estacionarias es que poseen media, varianza y covarianzas variables por traslación, efecto con un impacto nefasto para la aplicación de sistemas lineales de estimación. Para enfrentar este problema existen dos enfoques principales:

El modelo de residuos

Los incrementos de órdenes superiores.

La no estacionaridad se revela en el variograma como un crecimiento parabólico y suele confundirse con modelos gaussianos, un ejemplo de esto puede verse en Martínez et al. (2003). Si el drift aparece a larga distancia es posible evadirla empleando un tamaño de vecindad adecuado en la estimación.

El modelo de residuos es la base del Kriging universal expresado por la dicotomía: $Z(x) = m(x) + Y(x)$, donde $m(x)$ es el drift y $Y(x)$ es el residuo, que puede ser estacionario de segundo orden o intrínseco. En la práctica la función determinística $m(x)$ se define a partir de los polinomios deducidos con las realizaciones disponibles de $Z(x)$, en otras ocasiones es

asumida como una función externa “conocida”; pero raras veces el drift es realmente conocido, por ello se dice que es un concepto elusivo, en ocasiones no claro o muy complejo, en ocasiones claro, pero espurio (Chilès, 1999, p. 233) y esta es la principal debilidad del kriging universal.

La dicotomía se emplea para calcular el variograma de los residuos y se incorpora en las ecuaciones de kriging, pero la combinación lineal que caracteriza el estimador es similar a la del kriging ordinario. El variograma de los residuos es altamente sesgado, excepto para cortas distancias (Chilès, 1999, p. 122), además de esto, la definición del drift en la mayoría de los casos es oscura. Por tales motivos es preferible trabajar en el contexto IRF-k, aunque este tiene algunos detractores como Deutsch y Journel (1998) aludiendo fundamentalmente la complejidad del ajuste de los modelos bajo dicho contexto.

El kriging universal debe ser empleado solamente cuando la definición del drift es clara, o asociada a un fenómeno físico bien conocido y determinado (Renard, 2005). Las funciones intrínsecas de orden k son una generalización de las funciones intrínsecas a órdenes superiores, donde las diferencias de orden 2+k son invariables por traslación. De forma similar el variograma puede extenderse a órdenes superiores, por ejemplo, para orden 1 el variograma generalizado se define como:

$$\Gamma_1(h) = \frac{1}{6} \text{Var}[Z(x+2h) - 2Z(x+h) + Z(x)]$$

Este es capaz de filtrar el drift lineal y mostrar el variograma estacionario asociado al fenómeno (Chilès, 1999, p. 122). Lamentablemente para su construcción se requieren un muestreo regular, por esta razón se emplean las covarianzas generalizadas, que no dependen de la arquitectura del muestreo, pero no tienen representación gráfica.

Una propiedad interesante de las funciones aleatorias de ordenes superiores es que las de ordenes inferiores están definidas en su contexto, por ejemplo, una combinación lineal autorizada (ALC) IRF -6 es también una ALC IRF -5., en el orden inverso esta afirmación no es cierta. La implicación práctica más importante de dicha propiedad es que el número de funciones admisibles es mayor que en el caso estacionario o el intrínseco. Por citar un ejemplo, el variograma power definido como (Armstrong, 1998, p. 134):

$$\gamma(h) = C |h|^\alpha \quad \text{con } 0 < \alpha \leq 2$$

tiene como equivalente de orden superior

$$K(h) = \Gamma\left(-\frac{\alpha}{2}\right) |h|^\alpha \quad \text{con } 0 < \alpha \leq 2k + 2; \text{ con } k \text{ impar}$$

Las ecuaciones de krigeage en ambos modelos (Universal e IRF-k) son similares, la única diferencia es que en el caso del krigeage universal se emplea la covarianza de los residuos y en IRF-k las covarianzas generalizadas. Es posible extender al contexto de la no estacionalidad las técnicas propias del caso intrínseco: estimación del drift (como generalización de la estimación de la media), estimación del gradiente, filtrado de componentes de estructuras, kriging con error medido, etc.

Se introducen nuevas técnicas como el kriging con drift externo definido a partir de la ecuación $E[Z(x)] = a_0 + a_1 S(x)$, donde $S(x)$ es una función que define el drift externo (Bleines et al., 2004, p. 612). Algunas de estas pueden ser generalizadas al caso multivariado, pero aún no han sido implementadas a causa de la dificultad de obtener el modelo de variabilidad espacial. El kriging con drift externo es particularmente útil cuando se tiene una variable que caracteriza el drift, no necesariamente precisa, pero con alta densidad de muestreos en toda el área y una variabilidad inferior que la variable estudiada.

El ejemplo típico es la estimación de la profundidad de una capa a partir de unos pocos puntos medidos en sondeos de exploración y una data auxiliar densa obtenida a partir de perfilaje sísmico 2D (Bleines et al., 2004, p. 286). Como contrapartida de este método se encuentra el cokriging con colocación (en el campo de la geoestadística multivariada). Este es aplicable cuando existe correlación espacial entre las variables, la diferencia entre la variabilidad espacial de la variable secundaria y la primaria no es relevante. Una variante de kriging con drift externo es el Kriging Bayesiano (Deutsch y Journel, 1998, p. 68)

Para que el krigeage de una IRF-k sea una ALC-k debe cumplirse la condición de existencia definida como $\lambda^\alpha f_\alpha^l - f_0^l = 0 \quad \forall l \leq k$, donde f es la función que caracteriza el drift. El sistema de ecuaciones no será regular si no existen duplicados, el modelo que caracteriza la variabilidad espacial es condicionalmente definido positivo y las funciones que caracterizan el drift no son combinación lineal de los puntos de muestreo (Renard, 2005).

Esta última condición es obviada muchas veces y es la causa de muchas indeterminaciones; implica que para un drift lineal los puntos no deben estar ubicados en una línea, para un drift cuadrático no deben estar en una o dos líneas rectas paralelas o cruzadas, ni ubicados sobre una isolinia. Si la configuración de los puntos de muestreo se aproxima a los casos antes mencionados el krigeage se vuelve inestable. Otro aspecto a tener en cuenta es el número de puntos mínimo necesario para cada estimación, el que crece rápidamente con el número de dimensiones del espacio y con el orden del drift.

Geoestadística multivariada

La geoestadística multivariada presupone la existencia de dos o más variables, las que son tenidas en cuenta para estimar y simular las realizaciones de las FA o combinaciones lineales de estas. Las herramientas estructurales son similares, se dispone de matrices de covarianzas centradas y no centradas, directas y cruzadas, también se tienen matrices de variogramas y correlogramas; pero aparecen algunas especificidades, entre las que se destacan:

Los variogramas cruzados pueden ser negativos, con formas complicadas. Esto también ocurre en el caso de las covarianzas cruzadas. La covarianza cruzada puede tener una componente simétrica y otra asimétrica. Los variogramas cruzados siempre son simétricos.

Solo la parte simétrica de la covarianza cruzada se puede relacionar con el variograma. Para el caso de variables con unidades de medidas diferentes los variogramas y las covarianzas cruzadas no son adimensionales. Si la covarianza tiene una parte asimétrica (efecto retraso) los variogramas no son suficientes para determinar la varianza de una combinación lineal admisible (Le Loc'h, 2005)

En este contexto aparece una nueva herramienta: el pseudo-variograma cruzado; para calcularlo de forma robusta las variables deben tener la misma unidad y las diferencias a distancia cero deben ser estacionarias. No es simétrico, pero su parte simétrica está ligada al variograma por la expresión:

$$\frac{\Psi_{12}(h) + \Psi_{12}(-h)}{2} = \frac{1}{2} E\{[Z_2(x) - Z_1(x)]^2\} + \gamma_{12}(h)$$

En la actualidad solo están implementados estimadores multivariados basados en FA cuyo modelo de covarianza es simétrico, debido a la dificultad de determinar la varianza de una ALC multivariada con efecto retardo, solo con variogramas o covarianzas. Wackernagel, 1998, Chilès, Delfiner, 1999, entre otros, muestran dos modelos simétricos de variabilidad espacial como base para el ajuste de los variogramas:

El modelo intrínseco (o proporcional) de correogionalización.

El modelo lineal de correogionalización

El primero es el más simple, presupone que $C_{ij}(h) = b_{ij}\rho(h)$, donde $\rho(h)$ es la estructura de correlación y los coeficientes b_{ij} son las varianzas asociadas a cada componente. Una particularidad de este modelo es que en ausencia de correlación estadística (para $h=0$) las estructuras cruzadas son nulas. En la mayoría de los casos este modelo no es suficiente para explicar las estructuras cruzadas con un carácter complejo, para ello se utiliza el modelo lineal de correogionalización, que puede ser visto como una generalización de los variograma

anidados del caso univariado (Le Loc'h, 2005). En dicho modelo a cada estructura le corresponde una matriz de coeficientes: $C_{ij}(h) = \sum_l b_{ij}^l \rho^l(h)$. Estos coeficientes son usualmente factorizados, considerando la descomposición de las FA en factores, con incrementos no correlacionados. En ambos modelos las matrices B deben ser definidas positivas, para ello se debe cumplir la condición $|b_{ij}^l| \leq \sqrt{b_{ii}^l b_{jj}^l}$, solo suficiente para el caso bivariado; si existen más de dos variables la condición se aplica al sistema factorizado.

El ajuste del sistema cuando el número de variables es muy elevado es complicado, por ello es poco usual encontrar modelos de más de 2 o 4 variables. Frecuentemente se ajustan las estructuras cruzadas con un procedimiento automático (Bleines al. 2004, p 569), luego de haber ajustado las estructuras directas de la forma tradicional. Existe además el modelo bilinear de correogionalización, el que permite el modelado de funciones de covarianzas no simétricas (Wackernagel, 1998, p. 194), este se deduce a partir del modelo complejo lineal de correogionalización.

Los métodos de estimación más utilizados son el cokriging simple y ordinario, para datos heterotópicos se emplea además el cokriging con colocación (Wackernagel, 1998, p. 164) y el cokriging con colocación extendida (o cokrigeage multicolocado) (Bleines al. 2004, p. 616). En el caso isotópico estas técnicas tiene algunas ventajas sobre el modelo univariado, la más importante es que se cumple la condición $Z^{CK}(x) = \sum_{i=1}^n Z_i^{CK}(x)$, si las estructuras cruzadas tienen varianza cero o el modelo de correogionalización es intrínseco (autokrigeable), entonces, $Z^{CK}(x) = \sum_{i=1}^n Z_i^K(x)$, donde $Z_i^{CK}(x)$ y $Z_i^K(x)$ son la funciones aleatorias cokrigeadas y krigeadas respectivamente de las n variables i que forman la combinación lineal $Z(x) = \sum_{i=1}^n Z_i(x)$ (Wackernagel, 1998, p. 164). El concepto de autokrigeabilidad se emplea explícitamente en la formulación del modelo ortogonal de indicadores con residuales (Rivoirar, 1994)

Otra variante de cokrigeage que se pueden considerar es el krigeage con modelo aleatorio del drift (Deutsch y Journel, 1998, p. 68). El cokriging ordinario puede ser tradicional como lo muestra Wackernagel, 1998, p. 170, pero también puede ser con variables estandarizadas (Deutsch y Journel, 1998, p. 74). Wackernagel (1998, p. 181) muestra un enfoque interesante del análisis krigeante (en inglés factorial kriging análisis) el cual implica

diferentes procedimientos de cálculo según el modelo de coregionalización, este es ampliamente utilizado en la exploración geoquímica para detectar anomalías. Wackernagel, (1998, p. 187) también muestra dos variantes para el kriging de variables complejas: kriging complejo y cokriging de la parte imaginaria y la real. Otro método poco visto en la práctica es el kriging de varias variables ligadas por derivadas parciales para estimar combinaciones lineales del tipo $Z^* = \sum_{\alpha} \lambda^{\alpha} Z_{\alpha} + \sum_{\beta} \lambda^{\beta} \left(\frac{\partial z}{\partial x} \right) + \sum_{\delta} \lambda^{\delta} \left(\frac{\partial z}{\partial y} \right)$ (Bleines et al. 2004, p 615).

Simulaciones

Una introducción al buen uso de la simulación se muestra en un ejemplo de Alfaro S. M., 1979, él estima el largo de un cable que cruza el estrecho de Gibraltar a partir de unos pocos puntos medidos de la profundidad del fondo marino. En otra campaña se midió de forma exhaustiva la profundidad, cuyo perímetro se considera como el valor real. Como se muestra en la Figura 1, la realización simulada no es igual a la real, pero tiene la misma variabilidad, por lo que el largo del cable es inferido con bastante precisión; cuando se emplea el kriging el efecto de suavizado conlleva a su subestimación.

Bleines et al. (2004, p 637) define la simulación de una FA en un dominio determinado como una realización del modelo que la describe, que reproduce su variabilidad estadística y geoestadística (histograma y variograma). Contrario a la realidad, la cual puede ser considerada como una realización particular de la FA, la simulación se conoce en todo el dominio, por lo que puede ser empleada para cálculos empíricos, como simulación de fluidos, estudios de sensibilidad y la evaluación de los resultados bajo condiciones restrictivas, como el cutoff. En el caso de que las simulaciones sean iguales a los valores medidos en los puntos de medición, entonces se dice que estas son condicionales. El condicionamiento se puede realizar a partir de la expresión (Bleines et al. 2004, p 639):

$$Z_{sc}(x) = Z_s(x) + [Z(x) - Z_s(x)]^K$$

Cuando se van a construir simulaciones lo primero es determinar las características estadísticas y geoestadísticas del campo simulado y luego se define el algoritmo matemático que garantice reproducir las características requeridas. Generalmente se producen n realizaciones en paralelo, especialmente si se pretenden simular funciones multigaussianas, pues es el modo más sencillo de obtenerlas, siempre y cuando n sea lo suficientemente

grande, de acuerdo con el teorema del límite central (Bleines et al. 2004, p 639). Entre los métodos multigaussianos más conocidos se encuentran:

Método espectral, basado en la representación espectral de las covarianzas.

Método de dilución

Método de tesselación

Método de bandas rotantes

Descomposición LU

Simulación secuencial Gaussiana

Una adaptación a los random set del método secuencial gaussiano es el secuencial indicador. Como alternativa a este modelo, para simular variables categóricas, se tienen el gaussiano y plurigaussiano truncados, una explicación detallada se muestra en Armstrong, et al., 2003.

Un modelo poco usado es el de sustitución, donde la FA es sustituida por otra FA (función directa) y un proceso estocástico (proceso de codificación). Existen varios modelos para estos dos componentes que pueden ser combinados para lograr una morfología prescrita de la FA simulada (Bleines et al. 2004, p 653).

Lantuéjoul (2002) muestra en su libro, con bastante detalle los modelos de funciones aleatorias basados en objetos, los que son controlados por un proceso de Poisson, una familia de random sets y variables aleatorias con la misma ley. El resultado de la simulación puede ser un random set o una FA; en ambos casos con un alto costo computacional si esta es condicionada, debido a la velocidad de convergencia del kernel de transición empleado en el condicionamiento (Lantuéjoul, 2005).

Hasta la actualidad este modelo es uno de los más apropiados para simular objetos como lentes de arena, bloques flotantes (boulders) en una corteza laterítica, entre otros, además, es bastante flexible y brinda un gran número de posibilidades en cuanto a la intensidad del proceso Poisson, que controla la cantidad de objetos generados, la forma de dichos objetos, su rotación y combinación.

El modelo de mosaico, de gran utilidad para comprender aspectos de la geoestadística no lineal y multivariada tiene como principio construir una partición aleatoria de un dominio simulado con valores (constantes) independientes que han sido asociados a cada polígono que constituye la partición. Para establecer la partición aleatoria se emplean procesos de líneas de Poisson (mosaico de Poisson) y de puntos de Poisson (mosaico de Voronoi) (Bleines et al. 2004, p 657)

El modelo Fractal está basado en el movimiento fraccionario Browniano, cuyos incrementos en el tiempo t tienen una distribución gaussiana. Los fractales pueden ser determinísticos o aleatorios, en este caso se emplean varios algoritmos, como: desplazamiento del punto medio, método de interpolación y síntesis espectral.

Además de estos métodos existen otros que modifican las características de la imagen inicialmente simulada para forzar sus propiedades a propiedades determinadas, de ellas las más conocidas son el método autoregresivo determinístico (vea Bleines et al. 2004, p 662) y el método de templado (Annealing) (vea Deutsch y Journel, 1998).

Estos son algunos de los métodos de simulación más conocidos, además existen los equivalentes multivariados, con drift externo, etc. Aún no han sido implementadas bajo el contexto IRF-k, al menos para el modelo multigaussiano, aunque Chilès y Delfiner (1999), 510 muestra la solución para IRF-0 y IRF-k.

Desde el punto de vista teórico existen algunas controversias, por ejemplo Deutsch y Journel (1998), plantea que el método de las bandas rotantes es inaceptable por problemas de artefactos en las simulaciones condicionadas. En versiones modernas de este método el número elevado de bandas y su posición pseudo-aleatoria permitieron eliminar este efecto. Por otra parte Lantuéjoul (2005) y otros autores de la escuela francesa de geoestadística plantean que el método secuencial solo es aceptable bajo condiciones estrictas estacionaridad de segundo orden, de no cumplirse esta condición el método se vuelve rápidamente inestable. Otra controversia está dirigida hacia el empleo de enfoque indicatriz para modelar litofacies defendido a capa y espada por Deutsch y Journel (1998), mientras que Armstrong et al. (2003) muestran sus grandes insuficiencias, por ello desarrollan el plurigaussiano para modelar litofacies.

Geoestadística no lineal

Bajo el término de geoestadística no lineal se agrupan las técnicas que implican una transformación no lineal de los datos. Estas se crearon para modelar variables que luego de transformadas pueden ser tratadas con técnicas lineales, los ejemplos más conocidos son el krigeage lognormal o de logaritmos y el krigeage multigaussiano, donde la estimación se realiza con FA gaussianas obtenidas por anamorfosis.

En otro grupo se encuentra el krigeage de los indicadores de variables categóricas o variables continuas transformadas a indicatrices; el objetivo es determinar la distribución de la FA analizada en un soporte puntual o de bloque y con ella inferir probabilidades o recursos, expresados en porcentajes de toneladas de metales y menas. Un tercer grupo que se emplea

con objetivos similares, es deducido a partir de los indicadores, pero se basa en el enfoque gaussiano, el más popular de estos métodos es el kriging disyuntivo. En el primer grupo la transformación de los datos tiene como objetivo realizar el kriging sobre FA gaussianas, pues estas tienen un grupo de propiedades que facilitan el uso de técnicas geoestadísticas. En ellas se cumple que la estacionaridad de segundo orden es equivalente a la estacionaridad estricta; aún en el caso intrínseco la mayoría de los estimadores lineales funcionan muy bien.

No es recomendable emplear estimadores lineales cuando la distribución es altamente no lineal (Chilès y Delfiner, 1999, p. 191). Un sistema multigaussiano con media cero está completamente caracterizado por la matriz de covarianza, su distribución condicional es gaussiana y relacionada con el kriging simple por la expresión (Chilès y Delfiner, 1999, p. 381): $Y(x_0) | \{Y(x_\alpha) : \alpha = 1, \dots, N\} = Y_{SK}^* + \sigma_{SK} U$, donde U tiene distribución uniforme.

Si la distribución es lognormal es posible efectuar el kriging de los logaritmos normalmente distribuidos. El problema está en la transformación inversa, para garantizar que esta sea insesgada se consideran las expresiones (Bleines et al. 2004, p 622):

$$Z_v \cong \exp \left[a \left(\frac{1}{v} \int_v Y(x) dx \right) + b \right] - \beta$$

$$a^2 \cong \frac{s_Y^2 + \ln \left(\frac{1}{v^2} \int_v \int_v e^{-\gamma^Y(x-y)} dx dy \right)}{s_Y^2 - \gamma_{v,v}^Y}$$

$$b = (1-a)m_Y + \frac{1}{2} \left[(1-a^2)s_Y^2 + a^2\gamma_{v,v}^Y \right]$$

Para el kriging lognormal ordinario con soporte de bloque el estimador resultante y la varianza de estimación son:

$$Z^* = \exp \left[aY^* + \frac{1}{2} a^2 \sigma_Y^2 + b + a^2 \mu \right] - \beta$$

$$\sigma^2 = e^{a^2(s_Y^2 - \gamma_{v,v}^Y)} \left[1 + e^{a^2(s_Y^2 - \gamma_{v,v}^Y)} (e^{a^2 \mu} - 2) \right]$$

Esta técnica tiene muchos detractores a causa de la sensibilidad a los valores extremos. Un método similar es el kriging multigaussiano, donde la transformación se realiza por anamorfosis gaussiana. Generalmente se emplea kriging simple y la transformación inversa no tiene mayores complicaciones (Olea, 1999, p.35), aunque también es posible emplear el kriging ordinario. El modelo lognormal no es más que un caso particular de la anamorfosis gaussiana, pero la forma en que esta se implementa elimina en buena medida el efecto de los valores extremos.

El kriging indicador provee una estimación por mínimos cuadrados de la distribución condicional acumulativa (ccdf) para un valor de corte z_k , según la expresión $[i(x; z_k)]^* = E\{I(x; z_k) | (n)\}^* = \text{Prob}^*\{Z(x) \leq z_k | (n)\}$ (Deutsch y Journel, 1998, p. 76). Es una técnica muy flexible dada las combinaciones posibles, entre las que se encuentra:

Kriging indicador simple

Kriging indicador simple con media conocida a priori

Kriging indicador ordinario

Kriging indicador mediano, una aplicación del modelo de mosaico

Kriging indicador con desigualdades

Cokriging indicador

Kriging probabilístico

Kriging con modelo Markov Bayes

Una particularidad es que necesita corrección de los resultados para que la ccdf sea monótona creciente y en el intervalo $[0,1]$. Una desventaja está dada por la pérdida de resolución asociada a la transformación en el caso de variables continuas. El kriging con desigualdades o Soft kriging (Bleines et al. 2004, p. 618) también puede ser considerada una técnica multivariada, emplea dos tipos de datos a la vez, los primeros definidos en cada posición por un valor (hard data) y los segundos por un intervalo (soft data) que se reemplaza por la probabilidad condicional simulada con el "Gibbs Sampler" o calculada con un modelo Markov-Bayes e indicadores (Deutsch y Journel, 1998, p. 91)

Según Armstrong et al. (2003) la no estacionaridad de los indicadores es usualmente desapercibida pues están acotados por el valor 0.5, note que $\text{Var}[1_F(x)] \leq 0.25$, por lo tanto los variogramas estacionarios no exceden este límite. Los variogramas de una indicatriz y su complemento son iguales, para el caso de más de dos indicatrices el sistema es multivariado y las estructuras cruzadas deben tenerse en cuenta.

Las indicatrices no son independientes y el número de funciones de covarianza está limitado por sus restrictivas propiedades, por ello el modelo gaussiano y los variogramas sin meseta no son admisibles para Random Sets; el esférico, aunque aplicado en la práctica no se ha probado que lo sea. La dificultad está dada en que el teorema de Bochner, empleado para demostrar que las funciones de variogramas son definidas positivas en caso de variables continuas, no es aplicable en el caso de los Random Sets. Por todos estos motivos se han creado métodos basados en funciones gaussianas para evadir el empleo de indicatrices.

Armstrong et al. (2003) también muestran que el krigeage y la simulación de las indicatrices no es capaz, en muchos casos, de reproducir los patrones espaciales reales de variables categóricas; no obstante existen condiciones especiales donde los indicadores ameritan ser empleados, tal es el caso del krigeage disyuntivo.

Este término está asociado al hecho de que la sumatoria de los random sets son disyuntivos, es decir son una codificación en términos de 0,1 y su sumatoria siempre es uno, esto permite desarrollar una FA empleado Random Sets de la forma $f(Y(x)) = \sum_i f_i 1_{Y(x)=i}$. el kriging disyuntivo no es más que $[f(Y(x))]^{DK} = \sum_i f_i [1_{Y(x)=i}]^{CK}$, el kriging disyuntivo de un indicador es $[1_{Y(x)=0}]^{DK} = [1_{Y(x)=0}]^{CK}$.

Rivoirar (1990, p.7) insiste en la necesidad del uso del cokriging, y las simplificaciones que este puede tener. Si el modelo de correogionalización es intrínseco entonces los indicadores pueden ser estimados de forma independiente con kriging (modelo de mozaico). En el caso del modelo isofactorial sin efecto de borde los residuales de los indicadores son ortogonales y el cokrigeage puede simplificarse al krigeage independiente de los residuos de los indicadores.

En el caso del modelo isofactorial con efecto de borde no existe simplificación posible, por ello se emplea el modelo gaussiano en lugar de la indicatrices, donde los factores son los polinomios de hermite, el krigeage disyuntivo se obtiene por estimación con kriging de cada factor por separado.

Otros métodos no lineales basados en el modelo gaussiano son la probabilidad a partir de la esperanza condicional (Bleines et al., 2004, p 626) que emplea directamente la función de anamorfosis para estimar la probabilidad de que una variable $Z(x)$ no gaussiana exceda un cutoff determinado. El Condicionado uniforme (Bleines et al., 2004, p 632) considera conocido el valor gaussiano de cierto metal en el panel y emplea directamente la función de anamorfosis gaussiana para calcular el tonelaje y el tonelaje de metal, de forma condicional al valor gaussiano.

Y finalmente las variables de servicio (Bleines et al. 2004, p 635) que estima directamente variables aditivas a partir de datos transformados a metal por encima de un cutoff en la posición de las muestras, considerando el cambio de soporte, según el modelo gaussiano definido por la función de anamorfosis.

CONCLUSIONES

En los años 90 esta ciencia se ha enriquecido considerablemente y su uso se ha generalizado a innumerables campos de aplicación, gracias al desarrollo de las ciencias informáticas y las mejoras introducidas en los algoritmos, como los de ajustes de variogramas en el contexto multivariado y el IRF-k, la extensión al contexto plurigaussiano del contexto gaussiano truncado, las mejoras introducidas al método de simulación con las bandas rotantes, entre muchos otros más específicos, por ejemplo: el método para ajustar variogramas de sistemas multivariados puramente heterotópicos (Martínez, 2005).

Aún existen muchos problemas por resolver, quizás el de mayor interés para la industria del petróleo sea la modelación de forma aceptable de fracturas. Chilès y Delfiner (1999, p. 554) muestra algunas soluciones preliminares basadas en el empleo del modelo booleano, pero no logra definir la forma y la extensión de las fallas.

Una nueva tendencia es la de fusionar otras técnicas con la geoestadística, un ejemplo es mostrado por Painho y Bação (2000) quienes aplican algoritmos genéticos a problemas de clusterización. También se tiende a remplazarla con algoritmos genéticos y el razonamiento fuzzy (Huang Wong y Gedeon, 1998).

Finalmente se destaca un aspecto aparentemente simple pero que diferencia a un especialista en geoestadística de un buen conocedor y es el análisis inicial del fenómeno. El primero parte del enfoque probabilístico y define un modelo matemático general que le permite deducir el método más adecuado para resolver el problema planteado, el segundo simplemente selecciona uno de los métodos existentes según los objetivos que persigue. La diferencia no tiene mayor trascendencia, mientras el problema o fenómeno que se presente no esté fuera de lo tradicional.

REFERENCIAS BIBLIOGRÁFICAS

- Alfaro Sironvalle M., 1979: Etude de la robustesse des simulations des fonctions aléatoires. Centre de Géostatistique – Ecole des Mines de Paris, Fontainebleau, France (Ph. D).
- Armstrong Margaret, Galli Alain G., Le Loc'h Gaëlle, Geffroy François, Eschard Rémi, 2003: Plurigaussian Simulations in Geosciences. Springer – Verlag Berlin Heidelberg New York, p. 149.
- Armstrong Margaret, 1998: Basic Linear Geostatistics. Springer – Verlag, Berlin, p 153.
- Bleines C., Deraisme J., Geffroy F., Jeannée N., Perseval S., Rambert F., Renard D., Torres O., Touffait Y., 2004: ISATIS Software Manual, Reference Guide, 5th Edition. Geovariances, www.geovariances.com, p. 711.

- Chilès Jean-Paul, Delfiner Pierre, 1999: Geostatistics: Modeling Spatial Uncertainty. Jhon Wiley & Sons Inc., New York, p. 695.
- Deutsch Clayton V., Journel André G., 1998: GSLIB Geostatistical Software Library and User's Guide. Second Edition, Oxford University Press, p 369.
- Dubois Grégoire, 1998: The AI-GEOSTATS FAQ list, www.ai-geostats.org (Dic. 2005).
- Goovaerts Pierre, 1997: Geostatistics for Natural Resources Estimation. Oxford University Press, Oxford, p. 483.
- Huang Yuan-tu, Wong Patrick, Gedeon Tom, 1998: "Spatial Interpolation Using Fuzzy Reasoning and Genetic Algorithms", Journal of Geographic Information and Decision Analysis, 2 (2) :204 -214,.
- Lantuéjoul Christian, 2002: Geostatistical simulation: Models and Algorithms. Springer – Verlag Berlin Heidelberg New York, p. 256.
- Lantuéjoul Christian, 2005: Simulations, CFSG course. Centre de Géostatistique – Ecole des Mines de Paris, Fontainebleau, France (Lectures in Geostatistics).
- Le Loc'h Gaëlle, 2005: Multivariate Geostatistics, CFSG course. Centre de Géostatistique – Centre de Géostatistique – Ecole des Mines de Paris, Fontainebleau, France.
- Martínez Vargas Adrian, 2005: Iron grades estimation in heterogeneous lateritic deposit. Centre de Géostatistique – Ecole des Mines de Paris, Fontainebleau, France (CFSG project report), p. 54.
- Martínez-Vargas Adrián, Legrá-Lobaina Arístides A., 2003, Ferrera-Alba Norberto, Mena-Matos Luis F., "Determinación de un modelo digital de la topografía original en el yacimiento Punta Gorda". Revista Minería y Geología, 18 (3-4) :103-119
- Matheron George, 1965: Les variables regionalisées et leur estimation. Masson, Paris, p. 306.
- Olea Ricardo A., 1999: Geostatistics for engineers and earth scientists. Kluwer Academic Publishers, Boston, p. 303.
- Painho Marco, Bação Fernando, 2000: "Using Genetic Algorithms in Clustering Problems", Geocomputation2000, <http://www.geocomputation.org/2000/GC015/Gc015.htm>..
- Renard Didier, 2005: Non stationary geostatistic, CFSG course. Centre de Géostatistique – Ecole des Mines de Paris, Fontainebleau, France (Lectures in Geostatistics).
- Rivoirar Jacques, 1990: Introduction to disjunctive kriging and nonlinear geostatistics. Centre de Géostatistique – Ecole des Mines de Paris, Fontainebleau, France (Lectures in Geostatistics C-143), p. 89.
- Rivoirar Jacques, 1994, Introduction to disjunctive kriging and nonlinear geostatistics. Oxford University Press, p. 181.
- Wackernagel Hans, 1998: Multivariate Geostatistics, 2nd edition. Springer – Verlag Berlin Heidelberg New York, p. 291.

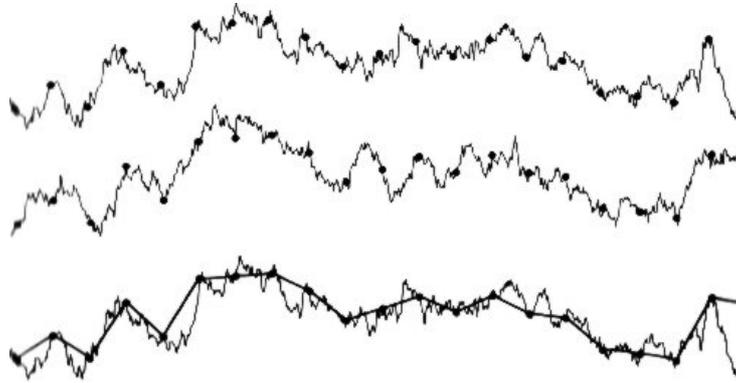
FIGURAS Y TABLAS

Figura 1. Arriba el perfil real con puntos de medición de la profundidad con un largo $L=1182$ m, en el centro una realización de una simulación condicional con $L=1154$ m, abajo estimación con kriging (línea continua) con $L=945$ m