



03061
2
2ej.

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

Unidad Académica de los Ciclos Profesional y de Posgrado del CCH
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas

**ANALISIS DEL CRITERIO DE
VALIDACION CRUZADA COMO
DIAGNOSTICO EN EL AJUSTE DE
MODELOS DE CORRELACION ESPACIAL**

FALLA DE ORIGEN
TESIS

QUE PARA OBTENER EL GRADO DE:

MAESTRA EN CIENCIAS

PRESENTA LA MATEMATICA

Eloísa Díaz Francés Murguía



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

En octubre de 1989, el Dr. Francisco Aranda Ordaz coordinó un seminario de "Análisis de datos con distribución espacial" (perteneciente al programa de la maestría en Estadística e Investigación de Operaciones de la UACPYP del CCH) en el cual tuve la oportunidad de participar. Durante el seminario, se invitó a dos investigadores con gran experiencia en el área, el Dr. Donald E. Myers y el Dr. Fernando Avila Murillo, quienes impartieron minicursos sobre diversos temas de Geoestadística.

La motivación para trabajar en el tema de esta tesis bajo la dirección del Dr. Aranda surgió a raíz de mi participación en este seminario, así como de diversas pláticas sostenidas con él y con el Dr. Myers.

A lo largo del seminario, tras estudiar y analizar diversos temas relativos al análisis de fenómenos con correlación espacial, surgieron muchas interrogantes e inquietudes sobre la forma de estimar y posteriormente de validar el modelo de correlación espacial.

A la fecha, no existe en la literatura geoestadística estudio alguno del comportamiento de los criterios de validación cruzada. Tampoco se conoce algún reporte de la distribución estadística de alguno de ellos con miras a poder realizar un contraste de hipótesis y decidir si el ajuste de un modelo dado es "bueno" o no para un conjunto de datos específico. El presente trabajo pretende abordar estos puntos.

En el primer capítulo, se presentan algunos conceptos básicos de Geoestadística para ayudar a la fácil comprensión del material restante, puesto que muchos de los términos geoestadísticos pueden resultar poco familiares para personas ajenas al área. En el

segundo capítulo, se describen el diseño y la metodología de estudio seguidas para analizar los métodos de estimación del modelo de correlación espacial, así como los criterios para validar al modelo elegido a partir de muestras simuladas. También se describe el algoritmo de simulación que se utilizó para generar muestras con correlación espacial e independientes.

En el tercer capítulo se analizan dos estimadores muy utilizados de la función de correlación espacial. En el cuarto capítulo se describe el comportamiento de cuatro medidas de validación cruzada al ajustar diversos modelos a las muestras simuladas. Finalmente, el quinto capítulo versa sobre la distribución estadística de una de estas cuatro medidas de validación cruzada que se designó como XVAL.

Por una parte, se analizan dos estimadores muy usuales de la función de correlación espacial y, por otra, se estudian cuatro formas de diagnóstico para verificar la bondad del ajuste del modelo elegido a una muestra de datos específica. La idea del presente trabajo es proporcionar lineamientos para llevar a cabo una buena estimación del modelo de correlación espacial incorporando los métodos de validación cruzada y también para realizar un buen diagnóstico. Se deja abierta la puerta para investigaciones futuras que involucren el uso de la distribución estadística de una de las medidas de validación cruzada.

Deseo agradecer a la Dra. Silvia Ruiz Velasco el haber tomado la dirección de este trabajo tras la muerte del Dr. Francisco Aranda, y también el haberme brindado todo tipo de facilidades para llevar a cabo esta tesis. Asimismo, quiero agradecer al Dr. Federico O'Reilly sus comentarios con respecto a la distribución estadística de la medida de validación cruzada que analicé.

Quiero dar las gracias al Dr. Donald Myers por la ayuda, comentarios y sugerencias oportunas que siempre me brindó (inclusive a larga distancia) a lo largo de este trabajo. También agradezco al Dr. Fernando Avila la ayuda para encontrar la mayoría

del material bibliográfico que sirvió de base a esta tesis.

Agradezco al Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas (IIMAS) por las facilidades otorgadas para el desarrollo de este trabajo. En especial doy las gracias por el apoyo para realizar las dos visitas que llevé a cabo a la Universidad de Arizona.

Finalmente, quiero agradecer al Dr. Miguel Nakamura por haberme mostrado las ventajas del lenguaje estadístico de cómputo Gauss y enseñado a usarlo; pero sobre todo, por el enorme apoyo académico y moral a lo largo de la realización de esta tesis.

CONTENIDO

1. Algunos conceptos básicos en Geoestadística	1
1.1. Variables regionalizadas	
1.2. Estacionaridad fuerte o de segundo orden	
1.3. Estacionaridad débil o hipótesis intrínseca	
1.4. Variograma	
1.4.1. Propiedades del variograma	
1.4.2. Ejemplos	
1.5. <i>Kriging</i> ordinario	
1.5.1. Interpretación intuitiva del <i>kriging</i>	
1.5.2. Sensitividad del <i>kriging</i> frente a cambios en los parámetros del variograma	
1.5.3. Comentarios	
1.5.4. Sistema dual del <i>kriging</i>	
1.6. Validación cruzada	
1.7. Procedimiento habitual para estimar datos con correlación espacial	
2. Metodología y diseño	31
3. Análisis del variograma empírico	37
3.1. Estimador de momentos	
3.2. Estimador de Cressie y Hawkins	
3.3. Resultados y comentarios	
4. Análisis de validación cruzada	51
4.1. Comportamiento de la varianza del <i>kriging</i> y de VKM frente a cambios en los parámetros del modelo de correlación espacial	
4.1.1. Cambios en la pepita	
4.1.2. Cambios en el rango	
4.1.3. Cambios en rango y pepita	
4.1.4. Cambios en la rejilla muestral	
4.1.5. Comentarios	
4.2. Metodología para evaluar a XVAL, ECMA y XVAL/VKM	
4.2.1. Comportamiento de XVAL	
4.2.2. Resultados del análisis de XVAL	
4.2.3. Comportamiento de ECMA y de XVAL/VKM	
4.3. Comentarios	

5. Distribución estadística de XVAL	76
5.1. Planteamiento teórico	
5.2. Resultados	
5.3. Comentarios	
CONCLUSIONES	84
BIBLIOGRAFIA	89

Capítulo 1

Algunos conceptos básicos en Geoestadística

La mayoría de los modelos estadísticos suponen que al estudiar un fenómeno se realizan observaciones bajo circunstancias idénticas y, además, que las observaciones son independientes unas de otras. Esta estructura simple intenta explicar una realidad compleja y no sirve para explicar fenómenos en el tiempo o en el espacio, donde no tiene sentido hablar de independencia entre las observaciones. Para este tipo de fenómenos, intuitivamente, se tiene la noción de que dos observaciones cercanas en tiempo o en espacio deben estar correlacionadas.

Para dar respuesta al problema de analizar los fenómenos que ocurren en el tiempo, surgió la teoría de Series de Tiempo durante el siglo pasado. Sin embargo, fueron los trabajos de G. U. Yule, en 1927 en el dominio del tiempo, y de N. Wiener, en 1949 en el dominio de las frecuencias, los que marcaron el arranque del análisis estadístico de las Series de Tiempo.

El estudio de fenómenos con correlación espacial es mucho más reciente, pues comenzó en este siglo, en los años cincuenta. Se intentó dar respuesta a los problemas que surgían: saber si había o no alguna estructura de correlación espacial entre observaciones que se tomaron en distintos puntos en el espacio, predecir un fenómeno con correlación espacial en puntos donde no hubo medición y, finalmente, conocer la variabilidad de este tipo de fenómenos.

Este tipo de problemas se presenta con bastante frecuencia en casi todas las Ciencias de la Tierra, ya que estudian fenómenos naturales con distribución espacial. Ejemplos de estos problemas aparecen en la Geología, la Agronomía, la Meteorología, la Física del Suelo, etc. De aquí que al aplicar metodología estadística a problemas de distribución espacial en su mayoría relativos a las Ciencias de la Tierra, surja lo que hoy se conoce como

Geoestadística.

Al igual que muchos otros conceptos, definir la Geoestadística no es una labor fácil. El primero en usar este término fue Georges Matheron en 1962. A. G. Journel (1986) lo define sintéticamente como sigue: "La Geoestadística es una rama de la Estadística que trata con fenómenos espaciales". Myers (1987), por otra parte, lo define como:

"La Geoestadística es una rama de los procesos estocásticos y de la estadística cuyos intereses primordiales son los problemas de estimación, predicción y simulación en diversas disciplinas. Estas incluyen a la hidrología, ingeniería de minas, física del suelo, ciencias de la tierra y, más recientemente, al monitoreo ambiental y procesamiento de imágenes".

Tras realizar una amplia revisión bibliográfica y analizar muchos de los procedimientos geoestadísticos, considero que la Geoestadística consiste simplemente en aplicar herramienta matemática, probabilística y estadística a situaciones donde existe una correlación espacial. A menudo, la geoestadística provee un enfoque probabilístico para resolver problemas que se venían solucionando en forma determinística, como lo es el problema de interpolación espacial.

En sus orígenes, en la década de los años cincuenta, la aplicación principal fue en el área de la exploración minera pues con un costo relativamente bajo, a través de muestrear un terreno, se podía estimar la distribución espacial del mineral por explotar. Esto permitía una mejor planeación del lugar de perforación de la mina. En la práctica se comprobó que esta metodología daba buenos resultados y fue entonces que se buscó una fundamentación teórica estadística que permitiese generalizar y aumentar los campos y tipos de aplicación.

Más adelante, con el advenimiento de la era computacional, el cálculo numérico requerido para predecir, simular y graficar

fenómenos naturales se facilitó enormemente. Actualmente, han comenzado las aplicaciones en el área de procesamiento de imágenes. Por ejemplo, en la elaboración de mapas cartográficos, muchas veces se parte de una imagen aérea captada vía satélite. Dicha imagen no es continua sino que proporciona información sólo en un número finito de puntos en el espacio. Para obtener un buen mapa, se requiere usar algún método para estimar la imagen en puntos intermedios, y es aquí donde interviene la Geoestadística. A diferencia de los métodos determinísticos, los geoestadísticos proporcionan además una medida del error de predicción.

Otra aplicación importante ha sido en la planeación y explotación de yacimientos petroleros o mineros donde el costo de perforación de un pozo o de una mina es muy elevado. Por eso es importante evitar perforaciones en sitios erróneos prediciendo adecuadamente los lugares con alto contenido petrolífero o mineral. Asimismo, se ha aplicado la Geoestadística a problemas de Ecología donde se quiere estimar la población de una especie determinada en alguna región geográfica. En el caso de que dichas especies sean de interés para el hombre, como la pesca por ejemplo, el estimar la distribución espacial de la especie ayuda a planear la captura y vigilar los niveles de explotación para evitar la extinción de la especie. Estos ejemplos son tan sólo unos cuantos de las posibles aplicaciones de la Geoestadística en áreas de interés.

El desarrollo de la Geoestadística se puede rastrear básicamente en las revistas *Mathematical Geology*, *Computers and Geosciences* y *Water Resources Research*, así como en memorias de congresos del área. Los primeros números de las revistas mencionadas fueron publicados en 1969, 1975 y 1965, respectivamente. En cuanto a libros de texto, a la fecha existen tres generaciones de escritores geoestadísticos. La primera, que incluye a Matheron, se caracterizó por usar un lenguaje matemático muy elevado. En la siguiente década surgió la segunda generación a la cual pertenecen A. Journel, C. Huijbregts, M. David, etc., y publicaron obras más accesibles para los expertos en las áreas de

aplicación. De este tipo es el libro de Journel y Huijbregts (1978) que contiene el material básico, es suficientemente completo, y muy citado por los autores de artículos en Geoestadística. Enfatiza las aplicaciones en ingeniería de minas. La tercera generación de escritores comenzó recientemente, en 1988, con M. Hohn, nuevamente David, Journel, y unos discípulos de él: E. Isaaks y M. Srivastava. El libro de estos últimos es de 1989 y, aunque es introductorio a la Geoestadística, tiene muchas cualidades pues presenta un enfoque más estadístico y es muy didáctico e intuitivo. Otra característica agradable de esta obra es que además de ser una buena guía para un autodidacta, analiza el mismo conjunto de datos reales a lo largo de toda la presentación para ejemplificar todos los métodos que presenta, lo cual permite compararlos fácilmente (ver Isaaks y Srivastava 1989).

A continuación se definen y explican algunos conceptos básicos de Geoestadística que se utilizarán a lo largo del presente trabajo.

1.1. VARIABLES REGIONALIZADAS

A menudo se puede caracterizar un fenómeno natural a través de la distribución espacial y/o temporal de una o más variables aleatorias que presentan cierta estructura de autocorrelación. A dichas variables se les conoce como *variables regionalizadas*. Un enfoque posible es suponer que una variable regionalizada es un proceso estocástico con dominio en \mathbb{R}^m , el espacio euclídeano m -dimensional (en este trabajo se considerará $m=2$). La contribución de la Geoestadística es suponer que una muestra de un fenómeno natural es una realización de una variable regionalizada, y así permitir el uso de herramienta probabilística. Para facilitar el análisis de este tipo de variables, se requiere que se cumplan ciertos supuestos que se describirán a continuación.

1.2. ESTACIONARIDAD FUERTE O DE SEGUNDO ORDEN

Sea $\{ Z(x) : x \in D \subset \mathbb{R}^n \}$ un proceso estocástico real-valuado definido en un dominio D contenido en \mathbb{R}^n . Se dice que $Z(x)$ es estacionario de segundo orden si cumple con

a) $Z(x)$ tiene esperanza finita y es igual a una constante para todo punto en el dominio; es decir, no depende de la ubicación del punto x :

$$E [Z(x)] = k, \quad k \in \mathbb{R} \quad \forall x \in D \subset \mathbb{R}^n.$$

Obsérvese que este supuesto es muy fuerte y es factible que no se cumpla para un fenómeno natural en toda la región de estudio. En estos casos, lo que se hace es dividir ésta en forma excluyente en sub-regiones en donde sí se cumpla este supuesto; entonces se trabaja con cada sub-región por separado.

b) $Z(x)$ tiene covarianza finita y es una función única del vector de separación h entre cada par de puntos x_1, x_2 en el dominio:

$$\text{COV} [Z(x), Z(x+h)] = C(h) < \infty.$$

Generalmente se trabaja con una hipótesis más débil que pide solamente estacionaridad para los incrementos $Z(x+h) - Z(x)$ y además facilita la inferencia sobre la estructura de correlación espacial para una muestra dada. Este supuesto se enuncia a continuación.

1.3. ESTACIONARIDAD DÉBIL O HIPOTESIS INTRINSECA

Se dice que $Z(x)$ es estacionario débil si cumple con:

a) $Z(x)$ tiene esperanza finita constante para todo punto en el dominio. Esto implica que la esperanza del incremento es cero:

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

b) Para cualquier vector h , la varianza del incremento $Z(x+h)-Z(x)$ está definida y es un función única de la distancia. A esta función se le llama *variograma* y se denota por $2\gamma(h)$:

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

Si en los incisos 1.2.b y 1.3.b, la covarianza y varianza dependen tan sólo de la magnitud de h , se dice que el proceso es *isotrópico*. Es decir, la correlación entre dos puntos no depende de la orientación que éstos tengan entre sí, sino que sólo de la distancia que los separa. Esta propiedad otorga un concepto de "homogeneidad" al proceso espacial. En el caso contrario en donde la correlación también depende de la dirección (i.e. hay direcciones privilegiadas de asociación espacial), se le llamará al proceso *anisotrópico*. Este caso es frecuente encontrarlo en fenómenos naturales. En ocasiones lo que se hace es realizar una partición de la región de estudio en zonas en donde el fenómeno sea isotrópico, pues el análisis se simplifica.

Obsérvese que la hipótesis de estacionaridad fuerte implica la débil y no viceversa. Es decir, el conjunto de funciones de covarianza tiene una cardinalidad menor que el conjunto de funciones de variograma. Otra consideración importante es que en el caso de estacionaridad fuerte, para poder estimar la covarianza de $Z(x+h)$ y $Z(x)$, se necesitaría tener muchas realizaciones del proceso estocástico lo cual es imposible en la práctica pues tan sólo contamos con una, la muestra de observaciones. En cambio en la hipótesis débil, al considerar la varianza de incrementos, para n puntos observados se van a tener diferentes "realizaciones" de incrementos para una distancia (o intervalo de distancia) h . Es decir, vamos a tener varias parejas de puntos de las $n(n-1)/2$ parejas posibles que estén separadas por una distancia h y esto permite la inferencia de la varianza de los incrementos.

En este trabajo se considerarán procesos isotrópicos que sean estacionarios fuertes. Uno de los conceptos centrales de la Geostatística es el variograma, que se explica más ampliamente

enseguida.

1.4. VARIOGRAMA

El variograma, denotado por $2\gamma(h)$, es una función que caracteriza las propiedades de dependencia espacial de segundo orden de un proceso estocástico definido en \mathbb{R}^n . Se define como la varianza de los incrementos de la variable regionalizada $Z(x)$:

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

Bajo la hipótesis intrínseca se tiene:

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

Bajo la hipótesis de estacionaridad fuerte se tiene que:

$$\gamma(h) = \text{COV}(0) - \text{COV}(h), \quad (1)$$

donde

$$\text{COV}(0) = \text{COV}(Z(x), Z(x)) = \text{VAR}(Z(x)) = \sigma^2, \quad \forall x \in D \subset \mathbb{R}^n,$$

y

$$\text{COV}(h) = \text{COV}(Z(x+h), Z(x)).$$

A $\gamma(h)$ se le llama el *semivariograma*. Obsérvese que es una función de los incrementos de la variable regionalizada $Z(x)$ para puntos que distan h . En el caso isotrópico y de este trabajo, el semivariograma depende tan sólo de la distancia de separación y no de la dirección del vector h . De aquí en adelante, cuando se escriba h , se aludirá sólo a la magnitud del vector. También se usará la siguiente notación:

$$C_{i,j} = \text{COV}(Z(X_i), Z(X_j)) = \text{COV}(|X_i - X_j|)$$

$$\text{y} \quad \gamma_{i,j} = \gamma(|X_i - X_j|).$$

Matheron le otorgó el nombre de variograma en 1965; sin embargo, el uso de esta función no es privativo de la Geostatística. En 1941, Kolmogorov en Física y Gandin en Meteorología la llamaron "función estructural". En 1952, G. H. Jowett en Series de Tiempo la llamó "diferencia cuadrática media". En seguida se enumeran algunas de las propiedades del semivariograma.

1.4.1. Propiedades del semivariograma

a) El negativo del semivariograma debe ser una función condicionalmente positiva definida. Esta condición está relacionada con la conocida propiedad de la matriz de varianza-covarianza de ser positiva definida y dice que $-\gamma(h)$ es *condicionalmente positiva definida* si (ver Journel y Huijbregts, 1978, p.35)

$\forall x_1, \dots, x_n$ y $\forall \lambda_1, \dots, \lambda_n$ tales que $\sum_{i=1}^n \lambda_i = 0$ se tiene que:

$$-\sum \lambda_i \lambda_j \gamma(|x_i - x_j|) \geq 0.$$

b) Se debe cumplir que

$$\lim_{h \rightarrow \infty} \frac{\gamma(h)}{h^2} = 0.$$

c) Por definición $\gamma(0) = 0$.

d) El semivariograma es una función par: $\gamma(h) = \gamma(-h)$.

e) Cualquier combinación lineal de variogramas con coeficientes positivos es un variograma.

f) Cualquier producto de variogramas es un variograma.

1.4.2. Ejemplos de semivariogramas

De acuerdo a si el semivariograma es acotado o no, se clasifica a las familias de semivariogramas más usadas como sigue:

NO ACOTADAS
Lineal
Logarítmico
Potencia

ACOTADAS
Esférico
Exponencial
Gaussiano
Pepita puro

A continuación se describen los siguientes modelos de semivariogramas: el esférico, el exponencial y el de pepita puro. En este trabajo se decidió analizar estos modelos de semivariograma por el hecho de ser acotados (con lo cual la covarianza de los incrementos es finita), ser ampliamente usados y, finalmente, a manera de simplificación del estudio. En las figuras 1.1.a, b y c, se muestran las gráficas de estos modelos. (En este trabajo, las figuras aparecen al final de cada capítulo).

a) Modelo Esférico

$$\gamma(h) = \begin{cases} 0 & \text{si } h=0 \\ C_1 & \text{si } 0 < h < a \\ C_1 & \text{si } h \geq a \end{cases} \frac{1}{2} \left\{ \frac{3h}{a} - \frac{h^3}{a^3} \right\}$$

donde $C_1 > 0$ y $a > 0$.

Este modelo presenta un comportamiento lineal para distancias cercanas al origen, pero los incrementos marginales van decreciendo para distancias grandes; hasta que para distancias mayores al valor a , dichos incrementos son nulos.

b) Modelo Exponencial

$$\gamma(h) = \begin{cases} 0 & \text{si } h=0 \\ c_1 \left\{ 1 - \exp\left\{ \frac{-2h}{a} \right\} \right\} & \text{si } 0 < h \end{cases}$$

donde $c_1 > 0$ y $a > 0$.

El modelo exponencial también se comporta linealmente cerca del origen donde la pendiente es mayor que en el caso esférico. En cambio, los incrementos marginales del modelo exponencial decrecen de manera más gradual para distancias grandes en comparación al modelo esférico, ya que el exponencial se acerca en forma asintótica a su mínima cota superior.

c) Modelo Pepita Puro

$$\gamma(h) = \begin{cases} 0 & \text{si } h=0 \\ c_0 & \text{si } h > 0 \end{cases}, \text{ donde } c_0 > 0.$$

Es común sumar este modelo a otro variograma valiéndose de la propiedad (e). A esto se le llama *anidar* variogramas. En algunos de los modelos, a veces los parámetros tienen interpretación física. Por ejemplo, en el modelo esférico que se mencionó en el inciso (a) sumado al de pepita puro, a los parámetros c_0 , (c_0+c_1) y a , se les conoce como *efecto pepita*, *meseta* y *rango*, respectivamente. La interpretación de estos términos se proporciona enseguida.

1.4.3. Efecto pepita

Cuando se suma el modelo de pepita puro a algún otro de los modelos de semivariograma, (usando la propiedad (e)), a la constante c_0 se le llama pepita y representa una discontinuidad puntual del variograma en el origen. Esto se puede deber a la escala de medición o a errores en la medición de observaciones. El modelo pepita puro implica una carencia de correlación espacial.

1.4.4. Meseta

La meseta es la mínima cota superior del semivariograma. Es el límite del semivariograma cuando la distancia h tiende a infinito. Puede ser o no finita. Los semivariogramas que tienen meseta finita cumplen con la hipótesis de estacionaridad fuerte; es decir, tienen una función de covarianza asociada. Cuando no es finita la meseta, el modelo de variograma describe a un fenómeno natural que sólo cumple con la hipótesis intrínseca.

En la Figura 1.2 se observa que la meseta es la altura (C_0+C_1) donde C_0 es la pepita. Si se interpreta la pepita como ruido o error en las mediciones, esto explica por qué se sugiere que en un modelo que describa bien a la realidad, la pepita no debe representar más del 50% de la meseta. Es decir, si el ruido en las mediciones explica en mayor proporción que la correlación espacial a la variación del fenómeno, las predicciones que se obtengan a partir de este modelo pueden ser muy imprecisas en cuanto a su descripción de la realidad.

1.4.5. Rango

Es la distancia a partir de la cual $Z(x)$ y $Z(x+h)$ son independientes para cualquier $x \in D \subset \mathbb{R}^m$, de aquí que se interprete al rango como "zona de influencia". En el caso del modelo exponencial, no existe una distancia finita para la cual dos observaciones sean independientes; por ello, se le llama rango efectivo a la distancia para la cual $\gamma(h)$ alcanza el 95% del valor de la meseta, como se indica en la figura 1.1.b. Entre más pequeño sea el rango (o rango efectivo) de un modelo, más cerca se está del modelo de independencia espacial. En el límite, cuando el rango tiende a cero, se tiene el modelo de pepita puro. En cambio, entre más grande sea el rango, la superficie que describa será más suave.

El rango no siempre aparece de manera explícita en la fórmula

del variograma. En el caso del modelo esférico, el rango coincide con el parámetro a . Sin embargo, en el modelo exponencial, el rango efectivo es $a/3$ y en el modelo gaussiano es $a/\sqrt{3}$.

De aquí en adelante, para caracterizar a algún modelo, se tendrá la siguiente notación:

familia del modelo (C0, C1, RANGO EFECTIVO).

En el caso de que el rango no sea finito, se utilizará el rango efectivo para efectos de esta notación. Por ejemplo, un modelo exponencial con pepita unitaria, meseta igual a cinco y rango efectivo igual a ocho sería:

exponencial (1,4,8).

El variograma modela y describe la estructura de correlación espacial; de aquí que el uso más importante de éste sea en la predicción espacial o interpolación. El variograma constituye el ingrediente principal de los métodos estadísticos lineales de predicción como el *kriging*.

En la práctica, generalmente se desconoce el modelo de variograma o de correlación espacial subyacente al fenómeno natural de estudio. Así que hay que estimar esta función. Como se explica con más detalle en el tercer capítulo, lo que generalmente se hace es, a partir de la muestra, obtener un estimador puntual del variograma para algunas distancias preestablecidas. Después, a estos valores estimados, se les ajusta (mediante diversos métodos) uno de los modelos de semivariograma que cumpla con las propiedades que se acaban de mencionar. Este modelo de variograma ajustado es el que se va a usar para simular, o para predecir el valor del fenómeno en los puntos en los que no hubo medición. Para esta predicción se utilizará el *kriging*, que se describe en seguida.

1.5. KRIGING ORDINARIO

El *kriging ordinario* es un método de predicción espacial puntual con la característica de ser el mejor predictor lineal con esperanza igual a la de la variable regionalizada que se quiere predecir. Se observan $Z(x_1), \dots, Z(x_n)$ y se desea predecir $Z(x_0)$ en el punto x_0 donde no hubo medición:

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i), \text{ con } \sum_{i=1}^n \lambda_i = 1.$$

Claramente Z^* es una combinación lineal de las n observaciones muestrales. Bajo la hipótesis de estacionaridad débil $E(Z(x)) = k$, $k \in \mathbb{R}$, Z^* tiene la misma esperanza que $Z(x_0)$, pues gracias a que los pesos λ_i suman uno,

$$E[Z^*(x_0) - Z(x_0)] = 0.$$

Es el "mejor" predictor puesto que los pesos λ_i se obtienen de tal manera que minimicen la varianza del error de predicción,

$$\text{VAR}[Z^*(x_0) - Z(x_0)].$$

Esta última es la característica distintiva del *kriging*, ya que existen otros métodos de interpolación como el de distancias inversas o el poligonal que también son lineales y de media $E(Z(x_0))$, pero no garantizan varianza mínima de predicción. Aquí es donde entra la herramienta probabilística al suponer que $(Z(x))$ es un proceso estocástico. La ventaja de este enfoque es ver a $Z^*(x_0)$ y a $Z(x_i)$ como variables aleatorias ya que entonces se procede a resolver el problema de encontrar los pesos λ_i tal que se cumplan las condiciones que se acaban de mencionar.

Así, se tiene un problema de minimización con restricciones que se resuelve mediante el método de multiplicadores de Lagrange:

$$\underset{(\lambda_1, \dots, \lambda_n)}{\text{minimizar}} \quad \text{VAR} \left[Z^*(x_0) - Z(x_0) \right], \quad \text{s.a.} \quad \sum_{i=1}^n \lambda_i = 1.$$

También se puede expresar a $\text{VAR} \left[Z^*(x_0) - Z(x_0) \right]$ como

$$\text{VAR} \left[\sum_{j=1}^n \lambda_j Z(x_j) - Z(x_0) \right] = \sigma^2 + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C_{ij} - 2 \sum_{i=1}^n \lambda_i C_{i0},$$

o también en términos del variograma:

$$2 \sum_{i=1}^n \lambda_i \gamma_{i0} - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma_{ij}.$$

(Ver Isaaks (1989) para un desarrollo algebraico detallado de este problema). Los pesos óptimos λ_i se encuentran con la solución al siguiente sistema de ecuaciones (donde μ es la mitad del multiplicador de Lagrange):

$$\begin{pmatrix} C_{11} & \dots & C_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C_{n1} & \dots & C_{nn} & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{pmatrix} = \begin{pmatrix} C_{11} \\ \vdots \\ C_{n1} \\ 1 \end{pmatrix}. \quad (2)$$

o el equivalente en términos del variograma:

$$\begin{pmatrix} \gamma_{11} & \dots & \gamma_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma_{n1} & \dots & \gamma_{nn} & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{pmatrix} = \begin{pmatrix} \gamma_{11} \\ \vdots \\ \gamma_{n1} \\ 1 \end{pmatrix}. \quad (3)$$

Por lo cual los pesos que minimizan el error de predicción son:

$$\lambda = C^{-1} d = A^{-1} b \quad \text{en} \quad Z^{\circ}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i).$$

En la práctica, en vez de C_{ij} o γ_{ij} , se usa algún estimador de estas funciones. En el caso de que sólo se cumpla la hipótesis intrínseca se usará la solución $\lambda = A^{-1} b$.

Finalmente, a la varianza mínima del error de predicción que se alcanza con estos pesos óptimos se le llama *varianza del kriging en x_0* y se puede expresar así:

$$\sigma_{E_0}^2 = \sum_{i=1}^n \lambda_i \gamma_{0i} + \mu.$$

Para cada punto x_0 donde se quiera predecir via *kriging* habrá un valor de varianza del *kriging*. (Para un ejemplo numérico muy claro ver Isaaks y Srivastava, 1989 Cap. 12.). Cabe aclarar que en una gran parte de la literatura geoestadística, al predictor de *kriging* se le llama "estimador", la varianza mínima de predicción se denota como "varianza mínima de estimación" y también se habla de un "error de estimación". Esto representa una confusión de términos, pues en estadística clásica, se define un estimador como una función de las observaciones muestrales que se utiliza para representar (o estimar) un parámetro desconocido (o alguna función de éste). Por tanto, queda claro que $Z^{\circ}(x_0)$ no estima a ningún parámetro, sino que "predice" la variable regionalizada en un punto donde no hubo observación. En Geoestadística también se dice que el *estimador de kriging* es insesgado cuando se quiere expresar que su esperanza es igual que la de la variable regionalizada (bajo estacionaridad débil). Nuevamente aquí hay confusión, puesto que el término de insesgamiento en estadística clásica se refiere al hecho de que la esperanza del estimador de un parámetro es igual al parámetro, $E[\hat{\theta}] = \theta$.

1.5.1. INTERPRETACION INTUITIVA DEL KRIGING

En la solución al sistema (2), el vector d proporciona una ponderación similar a los métodos de distancias inversas (que asignan un peso de $|h|^{-p}$ a cada observación donde h es el vector de distancia y es habitual elegir $p=2$). Es decir, a distancias mayores, $COV(h)$ también decrece. La ventaja evidente es que en vez de limitarse a funciones de la forma $|h|^{-p}$, las covarianzas y variogramas pertenecen a una familia mucho más grande de funciones.

Obsérvese que en el caso de estacionaridad fuerte se tiene que $COV(h) = COV(0) - \gamma(h)$. Por ejemplo, en el modelo esférico de semivariograma la covarianza asociada es (recuérdese que $COV(0)=C_1$):

$$COV(h) = \begin{cases} C_1 & \text{si } h = 0 \\ C_1 - \frac{1}{2} \left\{ \frac{3h}{a} - \frac{h^3}{a^3} \right\} & \text{si } 0 < h < a \\ 0 & \text{si } h \geq a \end{cases}$$

donde $C_1, a > 0$.

La figura 1.3 muestra las gráficas de $Cov(h)$ y su variograma asociado para mayor claridad.

Es claro que a las parejas que disten más del rango a se les asigna un peso nulo. Resumiendo, el vector d es un ponderador de "distancias estadísticas" y depende de la función de autocorrelación.

Ahora, en la solución del *Kriging* también interviene la multiplicación por la matriz C^{-1} . La matriz C registra las distancias entre todos los puntos en la muestra (para dos puntos cercanos corresponde una entrada con valor alto en C y viceversa). Por tanto, contiene información del agrupamiento o cúmulos de puntos en la muestra. De aquí que el efecto de multiplicar por C^{-1} es ajustar la ponderación cruda de distancias en d de acuerdo a efectos de cúmulos de observaciones para eliminar redundancias en la muestra y dar más importancia a puntos que no están agrupados.

Como se puede ver, el predictor de *kriging* ordinario toma en cuenta dos aspectos muy importantes en el problema de predicción espacial: la distancia (permitiendo una modelación muy rica) y la manera en que se conglomeran los puntos muestrales.

1.5.2 SENSITIVIDAD DEL KRIGING FRENTE A CAMBIOS EN LOS PARAMETROS DEL VARIOGRAMA

Si se recuerda que en la práctica hay que estimar el tipo de modelo y los parámetros del variograma, queda claro que es muy importante entender cómo afecta a la ponderación del predictor del *kriging* la elección de los parámetros del modelo.

Cambios en C1

Se traducen en cambios de escala; equivale a cambiar las unidades de medición. Esto queda claro al observar que el parámetro C1 se puede factorizar de $\gamma(h)$. Un cambio en C1 no afecta a los pesos del *kriging* (λ_i), pero sí afecta a la varianza del *kriging*. La demostración de esta propiedad es directa al sustituir en el sistema de ecuaciones (3) en lugar de $\gamma(h)$, la expresión $C1\gamma(h)^*$, donde $\gamma(h)^*$ tiene meseta unitaria. Entonces se observa que el parámetro C1 se puede factorizar de ambos lados de las primeras n ecuaciones y se cancela (μ es lo único que se ve afectado, pues queda dividido por C1). En este nuevo sistema de ecuaciones es claro que los pesos del *kriging* seguirán teniendo la misma solución que para otro modelo con C1 diferente pero igual rango y proporción de la pepita, ya que el efecto de C1 se canceló. En cambio, la varianza del *kriging* sí se ve afectada puesto que involucra a C1 en forma multiplicativa:

$$\sigma_E^2 = \sum_i \lambda_i \gamma_{i0} + \mu = C1 \left(\sum_i \lambda_i \gamma_{i0}^* + \gamma_{i0}^* - \sum_i \lambda_i \gamma_{i1}^* \right).$$

Obsérvese que la varianza del *kriging* se puede hacer tan grande o

pequeña como se desee al elegir modelos con un valor de $C1$ mayor o menor.

Cambio en la familia de modelo de semivariograma

Equivale a considerar un cambio en el comportamiento para distancias cercanas al origen. En la figura 1.4 se comparan tres familias de variogramas: esférico, exponencial y gaussiano con un mismo rango a . La diferencia principal es para las distancias pequeñas. El modelo Gaussiano, a diferencia del modelo esférico y el exponencial, tiene un comportamiento parabólico cerca del origen que ocasiona que valores muestrales muy cercanos influyan mucho en la predicción y que los alejados incluso lleguen a tener ponderaciones negativas. Este tipo de modelo describe superficies muy suaves. En cambio, los modelos esféricos y exponenciales describen un comportamiento casi lineal del variograma para distancias pequeñas y así describen superficies que no son tan suaves como las del modelo Gaussiano.

Cambios en la pepita

Aumentar la proporción que representa la pepita con respecto a la meseta ($C0+C1$) ocasiona que el modelo se parezca más al de pepita puro. Por tanto, se homogeneizan los pesos del *kriging* y la predicción se parece más a un promedio aritmético de las observaciones muestrales. En el caso del efecto pepita puro (la pepita representa el 100% de la meseta), se puede demostrar fácilmente que todos los pesos son iguales a n^{-1} y la varianza del *kriging* es igual a $C0(1+n^{-1})$. Es decir, el predictor en x_0 no tiene ningún parecido a los datos más cercanos, sino que toma por igual información contenida en toda la muestra. Este caso es equivalente a cuando no hay correlación espacial.

Cambios en el rango

El efecto de aumentar el rango es disminuir la varianza del *kriging*. En términos de la "distancia estadística" que se mencionaba, es como si las observaciones estuviesen más cercanas. Por el contrario, reducir el rango a un valor muy pequeño equivale a plantear que todos los puntos estuviesen igualmente distantes y se logra un efecto similar al del efecto pepita puro donde hay ausencia de correlación espacial. Entre mayor sea el rango, la superficie de predicción tenderá a ser más suave.

Resumiendo, aumentar el rango equivale a aumentar la zona de influencia de los puntos. En cambio, disminuirlo es tener cada vez zonas más pequeñas de influencia teniendo en el extremo, cuando el rango es cero, el caso de independencia espacial.

1.5.3. COMENTARIOS

1. La varianza del *kriging* y los pesos no dependen de las mediciones muestrales sino solamente del variograma y de la ubicación de los puntos. Esto permite que se puedan evaluar distintos esquemas de muestreo a priori. Ver Russo (1984) y Warrick y Myers (1987).

2. Limitarse a la clase de predictores lineales es conveniente pues sólo se requiere el conocimiento o estimación del momento de segundo orden del proceso estocástico. Es evidente que, si hay información sobre los momentos de orden mayor al segundo, existen métodos de predicción no lineal que tal vez sean mejores que el *kriging*.

3. El *kriging* es un interpolador exacto; i.e. la superficie de predicción pasa por las observaciones utilizadas para predecir. Se puede verificar fácilmente que

$$\hat{Z}(x_i) = Z(x_i), \quad \forall i \in N_1.$$

4. Para una comparación del *kriging* contra otros métodos de "predicción" ver Brooker en Royle et al. (1980) e Isaaks y Srivastava (1989), p.317.

5. En la práctica, no es necesario usar todas las observaciones para la combinación lineal del predictor del *kriging* pues heurísticamente se ha visto que si se eligen tan sólo de ocho a diez observaciones en una vecindad de x_0 , $Z(x_0)$ casi no cambia (ver Myers 1987), aunque esto depende del tipo de modelo de correlación, de sus parámetros y de la muestra. Obsérvese que ésto equivale a usar vecindades móviles. El precio que se paga por el ahorro en cómputo son posibles discontinuidades de la superficie de predicción al usar estas vecindades móviles para el *kriging*. Dubrule (1983) proporciona alternativas para simplificar el cómputo cuando se quieren usar todas las observaciones, que equivale a tener una vecindad única. Su método se basa en el sistema de ecuaciones dual del *kriging*, que se describe a continuación.

1.5.4. SISTEMA DUAL DEL KRIGING

El problema del *kriging* se puede plantear de una manera alternativa mediante otro sistema de ecuaciones que resuelve un problema equivalente. A este sistema de ecuaciones se le conoce como el dual del *kriging*. En notación matricial, el sistema de ecuaciones del *kriging* es:

$$\lambda \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{pmatrix} = \begin{pmatrix} \gamma_{01} \\ \vdots \\ \gamma_{0n} \\ 1 \end{pmatrix},$$

$$A \lambda = b,$$

y el sistema de ecuaciones dual del *kriging*:

$$\lambda \begin{pmatrix} m_1 \\ \vdots \\ m_n \\ \mu \end{pmatrix} = \begin{pmatrix} Z(x_1) \\ \vdots \\ Z(x_n) \\ 1 \end{pmatrix},$$

$$A m = Z.$$

La equivalencia entre ambos sistemas resulta evidente al reescribir la expresión del predictor del *kriging* sustituyendo los pesos λ_i óptimos en términos de la solución al sistema de ecuaciones (3), así (a_{ij}^{-1} es la entrada ij de A^{-1}):

$$\begin{aligned} Z^*(x_0) &= \sum_{i=1}^n \lambda_i Z(x_i) = \sum_{i=1}^n \left(\sum_{j=1}^n a_{ij}^{-1} \gamma_{oj} + a_{i(n+1)}^{-1} \right) Z(x_i) = \\ &= \sum_{j=1}^n \gamma_{oj} \left(\sum_{i=1}^n a_{ij}^{-1} Z(x_i) \right) + \sum_{i=1}^n a_{i(n+1)}^{-1} Z(x_i) = \sum_{j=1}^n \gamma_{oj} m_j + m_{n+1} . \end{aligned}$$

Una interpretación posible para el sistema de ecuaciones dual $\lambda \cdot m = Z$ es que expresa en forma explícita que el *kriging* es un interpolador y por tanto el valor del predictor en un punto muestral debe coincidir con la observación:

$$Z^*(x_i) = \sum_{j=1}^n \gamma_{ij} m_j + m_{n+1} = Z(x_i), \quad \forall i \in N .$$

La última ecuación del sistema dual $\sum_{i=1}^n m_i = 0$ es la contraparte de la restricción $\sum_{i=1}^n \lambda_i = 1$. En particular, se debe a que se desconoce la media $E[Z(x)]$ del proceso estocástico. Además, esta ecuación garantiza que la matriz A sea invertible y en el caso de estacionariedad fuerte ayuda a determinar cómo el *kriging* "extrapola" fuera de la malla de observaciones. (D.E. Myers, junio de 1991, en comunicación personal).

El hecho de que el predictor de *kriging* es un interpolador exacto (la superficie de predicción pasa por los puntos observados) sugiere el siguiente procedimiento, llamado validación cruzada, para evaluar la bondad del ajuste del modelo de semivariograma elegido con respecto a los datos muestrales. Debido a que generalmente se estima la estructura de correlación espacial, la validación cruzada proporciona criterios para elegir entre una función de variograma estimada y otra. La parte central de este trabajo consiste en analizar el comportamiento de algunas

medidas de validación cruzada en la evaluación del "mejor" modelo de variograma. El uso del sistema dual del *kriging* permite simplificar el cálculo de estas medidas de validación, así como reducir significativamente el tiempo de cómputo para obtenerlos. A continuación se explica con mayor detalle en qué consiste el método de validación cruzada.

1.6. VALIDACION CRUZADA

El método de validación cruzada consiste en excluir uno de los n puntos muestrales, y con los $n-1$ puntos restantes y el modelo de variograma elegido, predecir (vía *kriging*) el valor de la variable en estudio en la ubicación del punto que se excluyó. Se piensa que si el modelo de variograma elegido describe bien la estructura de autocorrelación espacial entonces la diferencia entre el valor predicho y la observación muestral excluida debería ser pequeña. Este procedimiento se realiza en forma secuencial con cada uno de los puntos muestrales (uno por cada vez) y así se obtiene un conjunto de n "errores de predicción" (obsérvese que este método es parecido al del *jackknife*). Existe una generalización del método de validación cruzada que consiste en excluir a más de un punto muestral y entonces predecir en estos puntos con los que no se excluyeron. A continuación se cita una definición más formal del método de validación cruzada dada por Samper (1990):

"Sea $Z(x)$ una función aleatoria estacionaria de segundo orden con semivariograma $\gamma(h)$ y función de covarianza $C(h)$. Sean Z_1, Z_2, \dots, Z_N los valores de $Z(x)$ en N puntos de medida (*sic.* puntos en los que se observó). La validación cruzada consiste en suprimir el i -ésimo valor medido Z_i y en estimarlo a partir del resto de los datos. El valor estimado Z_i^* se calcula por *krigado* ordinario mediante la expresión:

$$Z_i^* = \sum_{a \in N_i} \lambda_a Z_a,$$

donde N_i es el conjunto de puntos utilizados para estimar Z_i (N_i no incluye el punto i) ..."

Repetiendo este proceso de predicción en los N puntos, se obtienen los N errores de validación:

$$e_i = Z_i^* - Z_i, \quad i=1,2,\dots,N.$$

Lo usual es calcular medidas que involucren a estos errores de validación para una colección de diversos modelos de variogramas, hasta que se cumplan ciertos criterios. Entonces se dice que el modelo de variograma que mejor describe al fenómeno es aquel que optimiza algún criterio. Puede ocurrir que el modelo que optimice algún criterio, no sea el modelo "óptimo" para otros criterios. Ver Myers (1987) y Samper (1990). En este trabajo se analizarán los siguientes cuatro criterios:

1) Mínimo error cuadrático medio (XVAL). La idea es que si el modelo se eligió en forma adecuada, la diferencia entre el valor "predicho" sin el punto muestral y éste debe ser muy pequeña. Luego, se elige el modelo de semivariograma que proporcione la menor medida:

$$XVAL = \frac{1}{n} \sum_{i=1}^n [Z_i^* - Z_i]^2.$$

Es fácil mostrar que el valor esperado de XVAL es $(1/n) \sum_{i=1}^n \sigma_{E_i}^2$.

2) Mínimo valor medio de las varianzas del kriging (VKM). Se supone que si el modelo describe adecuadamente a la correlación espacial, la varianza del error de predicción debe ser muy pequeña. Se escoge el modelo de semivariograma asociado a la menor medida:

$$VKM = \frac{1}{n} \sum_{i=1}^n \sigma_{E_i}^2.$$

El valor esperado de XVAL coincide con el de VKM. Para una rejilla de muestreo y modelo de variograma fijos, VKM es una constante. Así pues, el valor esperado de esta medida es VKM y coincide con el de XVAL.

3) Error cuadrático medio adimensional (ECMA) próximo a uno.

$$ECMA = \frac{1}{n} \sum_{i=1}^n \frac{[Z_i^* - Z_i]^2}{\sigma_{E_i}^2}$$

Este criterio se debe a que el valor esperado de ECMA es uno,

$$E[ECMA] = \frac{1}{n} \sum_{i=1}^n \frac{1}{\sigma_{E_i}^2} E[Z_i^* - Z_i]^2 = \frac{1}{n} \sum_{i=1}^n \frac{\sigma_{E_i}^2}{\sigma_{E_i}^2} = 1.$$

A la raíz cuadrada de cada sumando de ECMA se le llama "residuo estandarizado" y éstos son muy útiles para detectar puntos extremos y/o aberrantes. Lo que se hace en la práctica es graficar su histograma y se analizan los residuos estandarizados con valores muy grandes en comparación a los otros; entonces se estudian los valores $Z(x_i)$ asociados a los residuos correspondientes. Generalmente estos valores son puntos extremos. Este tipo de análisis forma parte de los llamados "métodos gráficos de validación cruzada".

4) Cociente de XVAL/VKM. Se elige el modelo cuyo cociente asociado esté cercano a uno, por razones similares a las expresadas para el inciso anterior. Existe un método gráfico que es muy utilizado en la práctica (que no se analizó en este trabajo) que consiste en graficar los residuos $(Z_i^* - Z_i)$ contra los valores observados Z_i para $i=1, \dots, n$ (también es usual calcular la correlación de éstos). Para un modelo "adecuado" de variograma, se debe tener que esta correlación sea cercana a uno. La ventaja de este método es que permite identificar fácilmente puntos extremos a partir de la gráfica, pues estos puntos se encontrarán muy alejados de la recta de 45 grados.

Es evidente que para calcular XVAL, ECMA, VKM/XVAL y los métodos gráficos es necesario conocer los n errores de validación. Para calcular estos errores hay que encontrar a los Z_i^* , por lo cual habría que invertir n matrices de dimensión $n \times n$, para encontrar las soluciones a los n respectivos sistemas de ecuaciones del *kriging*. Oliver Dubrule (1983) diseñó un algoritmo que se basa en el sistema dual del *kriging* y, con el cual, basta con invertir una sola vez una matriz de $(n+1) \times (n+1)$ para encontrar los n errores de predicción. Queda entonces clara, la reducción significativa que esto logra en términos de tiempo de cómputo.

A la fecha, no existe en la literatura geostatística un estudio detallado del comportamiento de las medidas de validación cruzada. Una forma de estudiarlo es mediante el uso de muestras simuladas en donde, por construcción de éstas, se conoce la estructura de correlación espacial o modelo subyacente. Entonces se observa si los criterios eligen o no modelos de variogramas que en algún sentido sean parecidos al modelo subyacente que generó las muestras. Es claro que habría que definir el criterio de cercanía o similitud entre modelos que se va a considerar. En el cuarto capítulo se discutirán estas ideas con mayor profundidad.

Los resultados de este trabajo pretenden incidir en el procedimiento habitual de predicción espacial, así como ayudar a que se lleve a cabo de una manera más eficiente. A continuación se da un breve esbozo de los pasos que generalmente se siguen frente a algún problema de correlación espacial.

1.7. PROCEDIMIENTO HABITUAL PARA ANALIZAR DATOS CON CORRELACION ESPACIAL

Un algoritmo muy usado para el análisis de fenómenos con correlación espacial es a grandes rasgos el siguiente:

1. Se realiza un diseño de muestreo de la característica de

estudio.

2. Se obtienen las mediciones en los puntos de observación.
3. Se estima el modelo de variograma.
4. Se utiliza algún método de validación cruzada para verificar la bondad de ajuste del modelo de correlación espacial elegido. Si el modelo satisface algunos criterios se pasa a (5); de otra manera, se regresa a (3) y se elige un modelo alternativo.
5. Se realiza la predicción en un conjunto de puntos diferentes de los muestrales via *kriging*.
6. Se estudia la superficie de predicción via mapas de contornos, mapas de residuos, graficas tridimensionales, etc. y se toman las decisiones pertinentes al problema.

Este trabajo pretende influir en la forma de llevar a cabo el tercer y el quinto incisos; también se espera plantear algunas inquietudes en las personas que apliquen esta herramienta geoestadística para que lo hagan de manera más eficiente y consciente.

Muchos usuarios de la Geoestadística (incluyendo a matemáticos y estadísticos) encuentran que la literatura geoestadística en ocasiones utiliza una terminología un tanto ambigua y diferente de la usual en estadística. Una posible razón de esto es que muchas personas que trabajan en diversas áreas como la Geología, Hidrología e Ingeniería de Minas, y que no están familiarizados con la terminología estadística, contribuyen a la literatura geoestadística. Por otra parte, a la fecha, hay todavía pocos estadísticos que se dediquen a esta área; por lo cual existen pocos libros de texto que "traduzcan" al lenguaje estadístico conocido muchas técnicas que en Geoestadística llevan otro nombre. (A manera de ejemplo, en Geoestadística al

procedimiento que se sigue para transformar alguna variable aleatoria a normalidad se le llama "Anamorfosis Gaussiana"). Esto ocasiona que el lenguaje y metodología de la Geoestadística pueda ser todavía poco conocida para matemáticos y estadísticos.

Por esta razón, en este trabajo se trató de expresar las ideas geoestadísticas en lenguaje simplemente estadístico. También se enfatizó el enfoque intuitivo de los resultados matemáticos y estadísticos a manera de facilitar la comprensión de las ideas y conceptos enunciados.

VARIOGRAMA ESFERICO

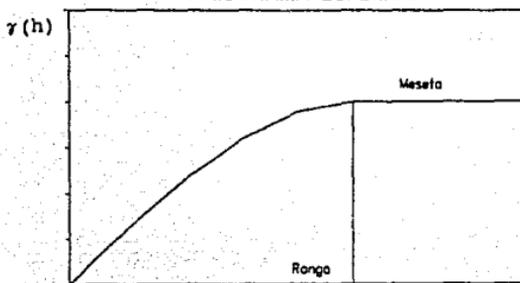


FIGURA 1.1.A

VARIOGRAMA EXPONENCIAL

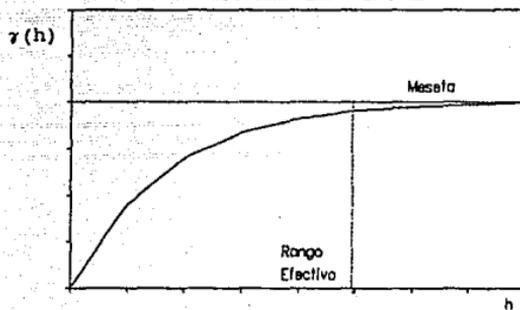


FIGURA 1.1.B

VARIOGRAMA PEPITA PURA

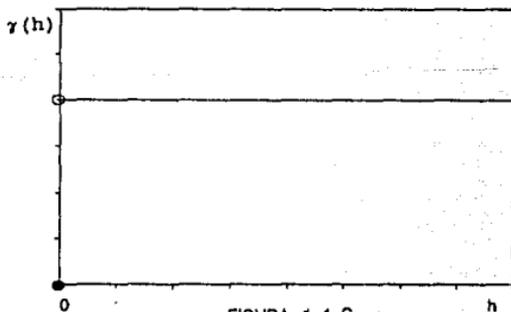


FIGURA 1.1.C

PEPITA, RANGO Y MESETA

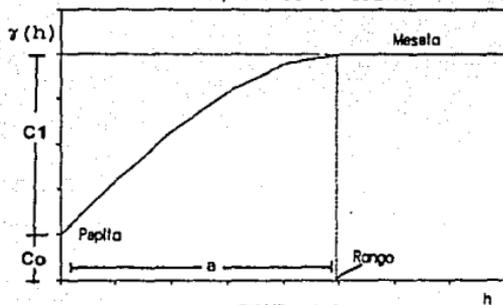


FIGURA 1.2

VARIOGRAMA Y COVARIANZA ASOCIADOS

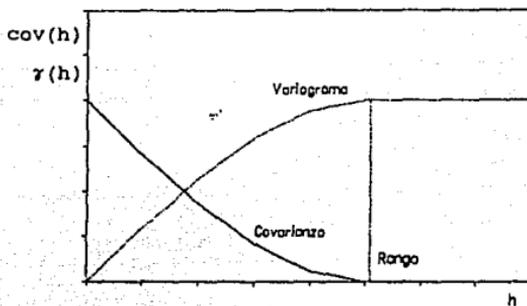


FIGURA 1.3

DIVERSOS VARIOGRAMAS

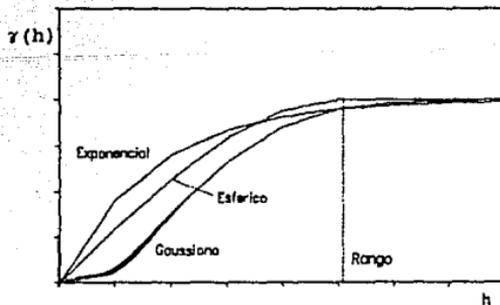


FIGURA 1.4

Capítulo 2

Metodología y diseño

entonces 89. Las implicaciones teóricas de esta decisión son que los resultados de este trabajo se restringen al caso de muestras de tamaño moderado y de rejillas regulares (lo cual se presenta frecuentemente en la práctica).

Se eligió un diseño de muestreo regular para tratar que las 3916 distancias posibles entre los 89 puntos se distribuyesen en forma lo más uniformemente posible, a lo largo de los rezagos elegidos (o clases de distancias):

<u>Intervalo de distancia</u>	<u>Número de parejas</u>
(0,1]	159
(1,2]	283
(2,3]	484
(3,4]	512
(4,5]	652
(5,6]	519
(6,7]	447
(7,8]	415
(8,9]	281
(9,13]	164
<u>TOTAL</u>	<u>3916</u>

Cada rezago representa a más de 150 parejas y el promedio de parejas por rezago es de 391. Para evitar que algún rezago tuviese pocas parejas, el último rezago se construyó cuatro veces más largo que los anteriores e incluyó a 164 parejas. Journel y Huijbregts (1978, p.194) dan como regla que cada rezago contenga un mínimo de 30 parejas, lo cual es cierto en este caso. También comentan que la distancia máxima para la cual se tiene una buena estimación del variograma es en general la que corresponde a la mitad de la máxima distancia entre los puntos muestrales, que en este caso correspondería al séptimo rezago.

Para simular los dos conjuntos de muestras con correlación espacial se usó el algoritmo de descomposición de Cholesky de la matriz de covarianza mencionado en Davis (1987). El algoritmo consiste en lo siguiente:

1. Se construye una matriz de las distancias entre los 89 puntos muestrales. Dicha matriz simétrica es de 89×89 y contiene ceros en la diagonal.

2. Se elige un modelo de semivariograma y se aplica la función del semivariograma a cada una de las entradas de la matriz de distancias. Esta matriz se designa como $\gamma(\text{DIST})$.

3. Bajo el supuesto de estacionaridad fuerte, la matriz de covarianza Σ se construye así:

$$\Sigma = c_1 \mathbf{1} - \gamma(\text{DIST}),$$

donde $\mathbf{1}$ es una matriz de unos de 89×89 y c_1 es el parámetro meseta del modelo de semivariograma.

4. Dado que la matriz de covarianza es positiva definida, se puede descomponer por el método de Cholesky:

$$\Sigma = L U \quad \text{donde} \quad L = U'.$$

L es triangular inferior y U triangular superior. A esta descomposición también se le conoce como la de la "raíz cuadrada".

5. Se genera un vector Y de 89 variables aleatorias normales independientes e idénticamente distribuidas, con media cero y varianza unitaria. Se usó el generador del lenguaje estadístico Gauss que utiliza el algoritmo de Kinderman y Ramage (1976).

6. Se construye la muestra de 89 puntos Z así:

$$Z = L Y.$$

Por construcción se tiene que:

$$E[(Z)] = E[L Y] = L E[Y] = L \mathbf{0} = \mathbf{0},$$

y

$$\text{Var}[(Z)] = \text{Var}[L Y] = L \text{Var}[Y] L' = L \mathbf{I} L' = L L' = \Sigma.$$

Por tanto, la muestra simulada Z tiene la misma estructura de correlación espacial que el modelo de semivariograma elegido para el paso (1).

De esta manera se simuló un conjunto de 100 muestras con un modelo de semivariograma esférico con los siguientes parámetros: pepita= $C_0=0$, meseta= $(C_0+C_1)=4$ y rango= $a=8.5$. Para generar el segundo conjunto de muestras con correlación espacial se usó un semivariograma exponencial con los mismos parámetros. El rango se eligió de tal manera que el 94 % de las parejas posibles de puntos distaran menos de ese valor. Se deseaba que la mayoría de las parejas estuvieran en la zona de influencia del modelo espacial. La meseta se escogió arbitrariamente como un valor aproximadamente igual a la mitad del rango. Dado que la pepita a veces se puede interpretar como "ruido" en las mediciones, se prefirió nulificarla.

Los cinco conjuntos de 100 muestras sin correlación espacial se construyeron a partir del generador de Gauss de normales estándar y correspondían a las siguientes distribuciones:

- * Normal con media 25 y varianza unitaria: $N(25;1)$
- * Normal con media 25 y varianza 10,000: $N(25;10,000)$
- * t de Student con tres grados de libertad $t_{(3)}$
- * Ji-cuadrada con 4 grados de libertad: $X^2_{(4)}$
- * Uniforme en (0,50): $U(0,50)$

Por ejemplo, una de las 100 muestras del tercer conjunto es un vector de 89 variables aleatorias iid que se distribuyen como una $t_{(3)}$. El objeto de incluir las últimas cuatro distribuciones es averiguar el comportamiento de los dos estimadores del variograma que se analizarán, frente a distribuciones con colas pesadas.

En resumen, se tienen 200 muestras con correlación espacial y 500 muestras independientes. Con ellas se analizará el

comportamiento de dos estimadores del variograma así como el de cuatro criterios distintos de validación cruzada como se verá con más detalle en los siguientes dos capítulos.

La ventaja de tener muestras simuladas es que se conoce el modelo de correlación espacial que las generó. Esto permite que se puedan comparar los estimadores del variograma para las muestras contra el modelo teórico que las generó y así verificar en forma empírica propiedades como insesgamiento y varianza de los estimadores, tendencias en los criterios de validación cruzada, etc. Es decir, mediante el uso de muestras simuladas se tiene un marco de referencia contra el cual evaluar la bondad de los estimadores del variograma y la posibilidad de analizar el tipo de modelo que seleccionan los criterios de validación cruzada.

Capítulo 3

Análisis del variograma empírico

La parte medular para lograr una buena predicción o simulación de un fenómeno con distribución espacial consiste en conocer lo mejor posible la estructura de correlación espacial de éste. Una manera de lograrlo es cuando se tiene una buena idea de cómo es la función del variograma, ya que ésta es el ingrediente principal del *kriging* y también de muchos métodos de simulación.

Existen diversos métodos para estimar el variograma: mínimos cuadrados, *jackknife*, máxima verosimilitud, validación cruzada, ajuste visual, validación cruzada de máxima verosimilitud, etc. Samper (1990) presenta y comenta estos métodos; y Cressie (1985) ahonda en el de mínimos cuadrados. De estos métodos el más común es el ajuste visual usado junto con alguna versión de mínimos cuadrados.

La mayoría de los métodos requieren el conocimiento de un variograma empírico, experimental o muestral como punto de partida. Por ejemplo, el método de ajuste visual consiste en elegir un modelo de variograma cuya gráfica se vea que encime o pase visualmente más "cerca" de los puntos del variograma empírico. De igual manera, los métodos de mínimos cuadrados eligen la función de variograma que minimiza la suma de los cuadrados de las diferencias con respecto al variograma empírico. De aquí que cuando se usen estos métodos, sea muy importante escoger un buen estimador del variograma, pues de esto depende la elección del modelo de variograma que posteriormente será la base para predecir el proceso en puntos intermedios o para simularlo.

En el presente capítulo se analizarán dos de los estimadores más usados para el variograma: el de momentos (Matheron, 1971) y el de Cressie-Hawkins (1980).

3.1. ESTIMADOR DE MOMENTOS

En 1963 Matheron propuso este estimador y a la fecha es el más usado. Dado que por definición el variograma es la varianza de los incrementos, un estimador que se ocurre en forma natural es la varianza muestral de los incrementos, es decir, el estimador de momentos:

$$2\hat{\gamma}(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} \left(Z(x+h) - Z(x) \right)^2 .$$

$N(h)$ es el número de parejas diferentes de puntos Z_i, Z_j que distan h , o algún intervalo de distancia preestablecido al cual en Geostatística también se le llama rezago. Obsérvese que se estima en forma puntual la función $\gamma(h)$ que es continua. Por ello, hay que decidir en cuales valores de h se va a estimar el variograma, si estos valores serán equidistantes o no, si conviene imponer alguna condición en cuanto al número de parejas de puntos que influyen en cada estimación puntual, etc. En este trabajo se estimará el variograma en diez valores de h , a partir de $h=1$, con incrementos unitarios hasta $h=9$, y el último rezago (9;13]. Así que en este caso, $N(h)$ es el número de parejas distintas de puntos cuya distancia está en el intervalo $(h-1, h]$ (o rezago h) donde $h \in \{1, 2, 3, \dots, 9\}$ y el décimo rezago es (9;13].

Obsérvese que generalmente se estima la función de variograma puntualmente para un número muy pequeño de distancias h (en este trabajo son sólo diez); además, no se tiene ninguna garantía de que $\hat{\gamma}(h)$ sea condicionalmente positivo definido (véase la sección 1.4.1.a). Así que una solución práctica es estimar de la mejor manera posible puntualmente el variograma y después elegir un modelo de variograma que "ajuste bien" a los valores estimados $\hat{\gamma}(h)$.

Una característica del estimador de momentos del variograma que se puede demostrar fácilmente es que es insesgado. Matheron (1965) señala que el estimador de momentos varía mucho para

distancias grandes. Es decir, se tiene que el estimador de momentos es heteroscedático puesto que la varianza de estimación se incrementa al aumentar la distancia. Más aún, un problema que presenta este estimador es que aunque es consistente, su varianza se reduce muy lentamente al aumentar el tamaño de muestra. De aquí que puede haber una diferencia muy marcada entre el variograma empírico y el variograma "teórico" subyacente al fenómeno natural, lo cual se agrava en el caso de que la muestra no sea muy grande. (Ver Armstrong, 1984). Esto representa un grave problema, pues el ajuste de un modelo válido de variograma muchas veces se realiza a partir del variograma empírico.

La elección de las distancias o rezagos h en donde se va a calcular $\hat{\gamma}(h)$ afecta al valor del estimador ya que cambian las parejas y el número de ellas que intervienen en el cálculo de éste. Esto se empeora en el caso de que la rejilla de muestreo sea irregular, pues puede haber rezagos en donde queden muy pocas parejas o ninguna. Warrick y Myers (1987) proponen algunos criterios para la elección de los rezagos de tal manera que se cumpla con lo siguiente:

1. Para cada rezago, el número de parejas de puntos debe ser lo más grande posible; *en especial para distancias pequeñas.*
2. El promedio de las distancias de las parejas en cada rezago debe ser cercano al valor de h para el que contribuyen en la estimación del variograma.
3. La varianza de las distancias en cada clase o rezago debe ser pequeña.

Cuando se muestrea en una rejilla regular, por lo general se cumplen los últimos dos criterios. En este caso la distribución de las distancias entre los puntos es de forma acampanada y unimodal; la moda se encuentra aproximadamente a la mitad de la máxima

distancia de separación entre las parejas de puntos.

Se ha visto que la región donde es más importante modelar bien al variograma es cerca del origen. De aquí que a la hora de diseñar el muestreo conviene planear que los puntos muestrales estén lo más cercano posible para que haya un buen número de parejas de puntos para calcular el estimador en los rezagos cercanos al origen.

Otro problema del estimador de momentos del variograma es que es muy sensible a valores extremos o aberrantes, como ocurre con los promedios en general. Así que el estimador puede oscilar considerablemente al incluir o no valores extremos en su cálculo. Es por eso que Cressie y Hawkins propusieron como alternativa el siguiente estimador del variograma.

3.2. ESTIMADOR DE CRESSIE Y HAWKINS

En 1980, Cressie y Hawkins presentaron una alternativa robusta al estimador de momentos:

$$\hat{\gamma}(h) = \frac{\left\{ \frac{1}{N(h)} \sum |Z(x+h) - Z(x)|^{1/2} \right\}^4}{.914 + \left\{ \frac{.988}{N(h)} \right\}}$$

Este estimador es mucho menos sensible a puntos aberrantes y se basa en la teoría existente de estimación robusta de P.J. Huber (ver Huber, 1981). Dentro de esta metodología existe herramienta estadística muy buena para estimar centros de simetría. Cressie y Hawkins suponen que el proceso estocástico es gaussiano y, en ese caso, los cuadrados de los incrementos se distribuyen como una ji-cuadrada.

La idea es transformar esta distribución asimétrica a una simétrica y en esta nueva estimar el parámetro que es el centro de

simetría. El usar el promedio de las raíces cuadradas de los incrementos elevado a la cuarta potencia es resultado de esta necesidad de simetría. El denominador del estimador es una corrección para que el estimador sea "casi asintóticamente" insesgado. Cressie y Hawkins también diseñaron un método de mínimos cuadrados para estimar el variograma que utiliza a este estimador.

A continuación presentaremos los resultados obtenidos al calcular los dos estimadores puntuales del variograma que se han mencionado para las 200 muestras correlacionadas y para las 500 independientes. Las muestras simuladas con el modelo esférico y con el modelo exponencial (a las que se llamarán "muestras esféricas" y "muestras exponenciales" para abreviar) son realizaciones de procesos estocásticos gaussianos pues por construcción, el valor en cada punto muestral es una combinación lineal de variables aleatorias con distribución normal estándar. Estas muestras y las independientes no tienen puntos extremos, ni están contaminadas con ruido, pero algunas distribuciones de las muestras independientes son de colas pesadas; es decir, modelan observaciones aberrantes o extremas.

3.3. RESULTADOS Y COMENTARIOS

Tras calcular el estimador de momentos y el de Cressie y Hawkins en diez rezagos para cada una de las muestras correlacionadas y de las independientes, se observó lo siguiente:

1. Se observó la heteroscedasticidad del estimador de momentos para las muestras con correlación espacial, mencionada en Armstrong (1984), y Matheron (1965). En las figuras 3.1.a y 3.1.b están graficados el modelo teórico del variograma que generó a las muestras y la media de los variogramas empíricos para los conjuntos de muestras esféricas y exponenciales, respectivamente. También se ordenaron para cada rezago los valores del estimador y se graficaron los cuantiles 5 y 95 de esta distribución empírica

por rezago. La gráfica de dichos cuantiles da una idea de la dispersión del estimador a través de una "banda de confianza".

2. También se observó empíricamente el inesgamiento del estimador de momentos en todos los tipos de muestras. Como se ve en las figuras 3.1.a y 3.1.b, para los dos conjuntos de 100 muestras con correlación espacial, la media de los variogramas empíricos está muy cerca de la gráfica del modelo teórico. En el caso de las muestras independientes la media de los variogramas empíricos es una constante igual a la varianza de la distribución que generó las muestras. (Se ilustra el caso de las muestras independientes con el conjunto de muestras $X_{(4)}^2$; los resultados de los otros conjuntos de muestras independientes fueron similares. Ver la figura 3.1.c). Además, la media se encontraba aproximadamente a la mitad de la "banda de confianza" en todos los conjuntos de muestras estudiadas (con y sin correlación).

3. El estimador de Cressie y Hawkins (1980) no dio resultados favorables en el caso de las muestras correlacionadas espacialmente. La "banda de confianza" completa, salvo por el primer rezago, cayó muy por abajo del modelo teórico como se ve en las figuras 3.2.a y 3.2.b. No obstante que estos autores afirman en su artículo que el estimador es por construcción asintóticamente "casi" inesgado, si no se cumplen los supuestos que ellos hacen, el estimador puede ser muy malo. También es posible que haya algún error en la construcción del estimador, y que a esto se deba el sesgo en los variogramas empíricos obtenidos. Debido a los resultados que se obtuvieron en este trabajo (en los cuales la banda de los variogramas estaba muy por debajo de la gráfica del modelo teórico) se revisó con mucho cuidado si se cumplían o no los supuestos de este artículo para el caso de las muestras que aquí se analizaron y también se estudió con detenimiento todos los pasos que los autores hicieron para construir el estimador.

Cressie y Hawkins justifican su estimador con base en argumentos que involucran varias aproximaciones asintóticas

(convergencias en distribución y convergencia de momentos) y el hecho básico de que las raíces de los incrementos $[Z(x+h)-Z(x)]$ (las cuales se denotarán como Y_t) son prácticamente independientes y además se distribuyen casi como una normal. En el caso de las muestras correlacionadas de este trabajo algunas de estas aproximaciones no se sustentan. Por ejemplo, las Y_t no son simétricas y poseen colas pesadas; de hecho con un nivel de significancia de .05, la prueba de Anderson-Darling rechaza la normalidad para todas las muestras estudiadas. Sin embargo, lo que se cree que afecta primordialmente al estimador es el hecho de que para una h fija, las Y_t sí tienen una correlación importante que Cressie y Hawkins no consideran y que afecta significativamente a la varianza de la distribución límite que utilizan para construir a su estimador (sobre todo para $h > 1$). Resumiendo, en la construcción del estimador intervienen muchas aproximaciones (independencia y normalidad de Y_t , tener muestras finitas, etc.), las cuales aparentemente están teniendo un efecto agregado sobre la distribución final del estimador.

Por tanto, en el caso de muestras gaussianas con correlación espacial, en donde por consecuencia los incrementos están correlacionados, no es recomendable usarlo pues puede presentar un sesgo inaceptable como se ve en las figuras 3.2.a y 3.2.b. En cambio, para las muestras independientes, el estimador de Cressie y Hawkins presentó un comportamiento más parecido al del estimador de momentos, y la media de los variogramas empíricos fue bastante cercana a la varianza de la distribución correspondiente, como muestra la figura 3.1.d. (En estos casos se cumplía el supuesto de Cressie y Hawkins de independencia de los incrementos).

4. Como muestra la figura 3.3.a, al comparar el ancho de las "bandas de confianza" de los variogramas empíricos de momentos de las muestras exponenciales y esféricas, se observó que la banda de las muestras exponenciales resultó ligeramente más ancha para distancias pequeñas que la banda de las muestras esféricas, pero fue más angosta para distancias grandes. Esto parece indicar que para distancias pequeñas, cuando el modelo subyacente es esférico,

el estimador de momentos presenta una variabilidad menor que cuando la muestra proviene de un modelo exponencial.

En el caso de los variogramas empíricos de Cressie y Hawkins, el ancho de la banda para las muestras esféricas fue menor que el de la banda exponencial para todos los rezagos (véase la figura 3.3.b). Esto parece indicar que la varianza del estimador de Cressie y Hawkins es menor cuando la muestra proviene de un modelo esférico que cuando proviene de un modelo exponencial. Sin embargo, esta información es de nula utilidad en la práctica puesto que para una muestra dada se desconoce el modelo que describe su estructura de correlación espacial. (Ya que si se conociera, se tendría el problema de estimación del modelo resuelto).

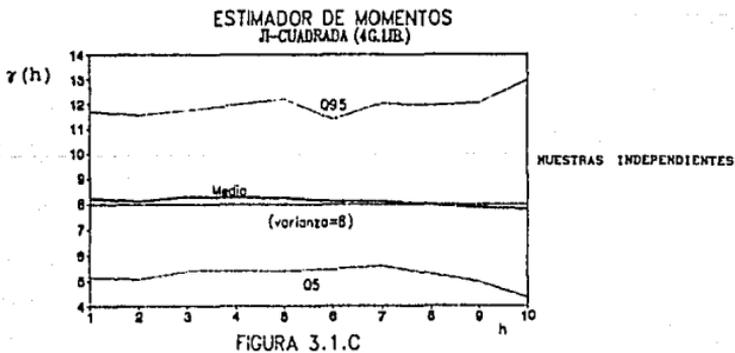
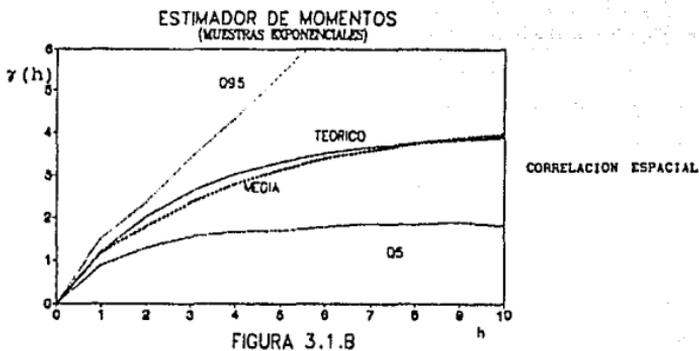
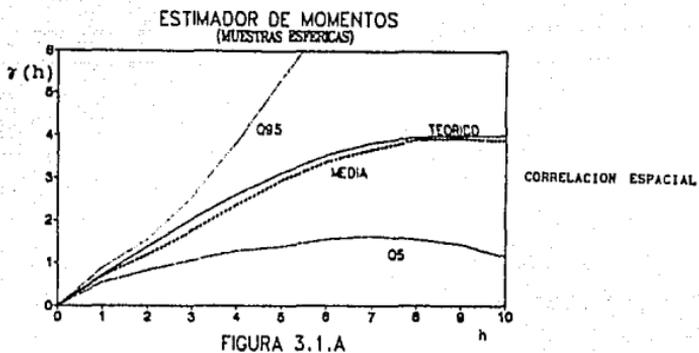
5. En las figuras 3.4 y 3.5, se graficaron los variogramas empíricos de diez muestras esféricas y diez exponenciales. Se observan las distintas formas que toman y que no son necesariamente parecidas a modelos esféricos o exponenciales. Se comprueba la aseveración de Armstrong (1984) antes mencionada: el variograma empírico puede diferir radicalmente en cuanto a su forma del modelo de correlación espacial subyacente. Esto representa un gran problema, sobre todo para el método de ajuste visual que se basa por completo en el estimador empírico del variograma.

Sin embargo, también se puede ver que para los primeros tres rezagos, al ser las bandas de confianza más estrechas, los variogramas empíricos no son tan diferentes unos de otros. La mayoría se comporta en forma lineal con ángulos de inclinación que a lo más difieren en 20 grados.

Si las muestras fuesen independientes, se podría pensar que la media de los variogramas es el variograma empírico de una muestra de tamaño 10,000 con 100 observaciones repetidas en cada uno de los 89 puntos de la rejilla. Aunque en este caso las muestras no son independientes, esto sugiere que el hecho de que

los variogramas empíricos sean de forma diferente al modelo teórico posiblemente se deba al tamaño de muestra, puesto que se observa que la media de los variogramas empíricos se ve muy cercana a la gráfica del modelo teórico. De aquí que se crea que al aumentar el tamaño de muestra se reduzca la diferencia del variograma empírico con el modelo teórico subyacente. De cualquier manera, cuando el tamaño de muestra sea pequeño, es mejor considerar sólo a los estimadores del variograma para distancias pequeñas para ajustar algún modelo de variograma.

6. En el caso de que existan puntos aberrantes se recomienda calcular el estimador de momentos omitiendo dichas observaciones y analizar éstas por separado. La idea es ver qué tanto cambia el variograma empírico al incluir o no a los puntos aberrantes y, después de analizar estos puntos, tomar las decisiones pertinentes con respecto a cuál es el modelo de correlación espacial que mejor describe al fenómeno de estudio.



ESTIMADOR DE CRESSIE-HAWKINS
 J-CUADRADA (4G.LIB.)

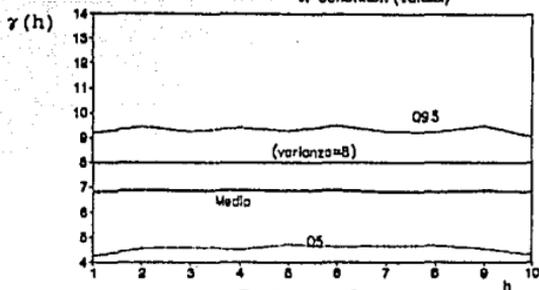


FIGURA 3.1.D

ESTIMADOR DE CRESSIE Y HAWKINS
 (MUESTRAS ESFERICAS)

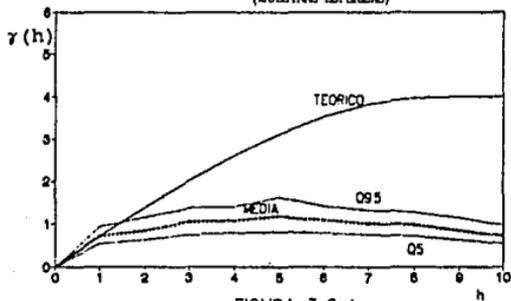


FIGURA 3.2.A

ESTIMADOR DE CRESSIE Y HAWKINS
 (MUESTRAS EXPONENCIALES)

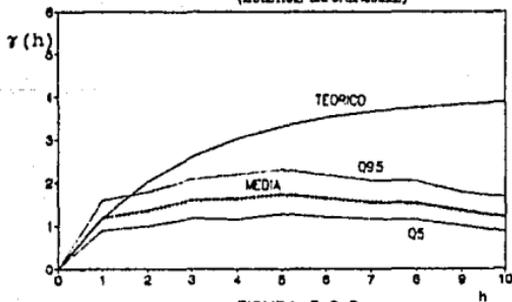


FIGURA 3.2.B

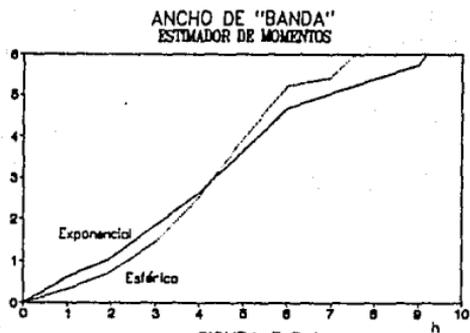


FIGURA 3.3.A

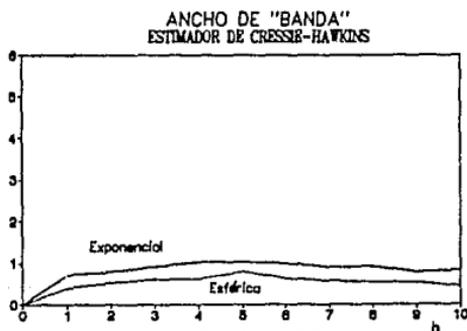


FIGURA 3.3.B

ESTIMADOR DE MOMENTOS
(10 MUESTRAS ESFERICAS)

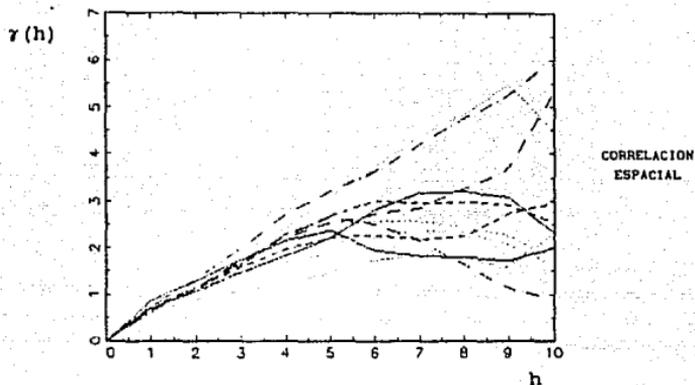


FIGURA 3.4

ESTIMADOR DE MOMENTOS
(10 MUESTRAS EXPONENCIALES)

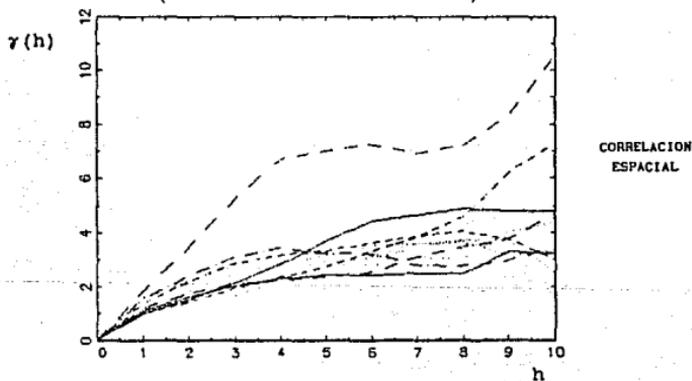


FIGURA 3.5

Capítulo 4

Análisis de validación cruzada

En el procedimiento habitual para analizar datos con correlación espacial, muchas veces se llega mediante algún método a unas cuantas estimaciones diferentes del variograma. Por ejemplo, para una muestra dada, el método de ajuste visual de estimación del variograma podría sugerir los siguientes modelos:

exponencial (0,5,5,8),
exponencial (0,5,7)
y esférico (0,5,9).

Lo que se hace es calcular alguna medida de validación cruzada para cada uno de estos tres modelos, usando los datos muestrales, y se elige el modelo que satisfaga mejor el criterio de validación cruzada. (Por ejemplo, si se usa XVAL, se escoge el modelo que tenga asociado un menor valor de XVAL. En cambio, si se usa ECMA, se eligirá al modelo que tenga un valor asociado de ECMA más cercano a la unidad).

Obsérvese que, en la práctica, tan sólo se calculan una o dos medidas diferentes de validación cruzada y solamente para un número muy pequeño de modelos. Además, se tiene el supuesto de que el modelo que ocasione que el criterio sea "óptimo" será el que mejor describa la correlación espacial del fenómeno. Pero, ¿qué ocurre si existen valores "óptimos" múltiples de algún criterio que tengan modelos asociados que originen superficies de predicción muy diferentes entre sí? O ¿qué tal si se usa más de una medida de validación cruzada y los criterios de elección del mejor modelo se contradicen? Surge entonces la duda de cuál modelo elegir. Por otra parte, ¿qué garantía se tiene de que el valor óptimo de las medidas de validación cruzada determine en forma adecuada al modelo de correlación espacial que mejor

describa al fenómeno en estudio? (Sobre todo si se calculan estas medidas para tan sólo unos cuantos modelos).

Una forma de explorar estos puntos es mediante la simulación de muestras con un modelo de correlación espacial conocido y, después, con cada una de estas muestras calcular las medidas de validación cruzada para una gran variedad de modelos en forma sistemática. De esta manera, se tiene un marco de referencia para evaluar el comportamiento de las medidas de validación cruzada. Es decir, se pueden comparar los modelos para los cuales las medidas alcanzan el valor "óptimo" con el modelo teórico que se usó para generar las muestras.

A la fecha, no existe en la literatura ningún reporte acerca del comportamiento de las medidas de validación cruzada más comúnmente usadas que se mencionaron en el primer capítulo, ni tampoco estudios o aproximaciones de su distribución estadística. La ignorancia del comportamiento de estas medidas es un grave problema pues, por una parte, es posible que conduzcan a elegir un modelo de variación espacial que sea una muy mala representación de la realidad. Por otra parte, el desconocer la distribución estadística de estos criterios impide realizar un contraste de hipótesis para decidir si hay evidencia estadística de que la medida de validación cruzada correspondiente sea significativa de un mal ajuste; esto es, que no sea lo suficientemente pequeña o cercana a la unidad (dependiendo del criterio en consideración).

En este capítulo se describe el comportamiento de los cuatro criterios de validación cruzada mencionados anteriormente y, en el siguiente capítulo, se muestra un método para aproximar la distribución teórica de la medida XVAL cuando se conoce el modelo de correlación espacial. El objetivo del presente capítulo es exhibir los errores que se pueden cometer al usar los criterios de validación cruzada indiscriminadamente y motivar a que se utilicen éstos en forma más sistemática, así como a que se realice un diagnóstico más profundo de los residuos de validación cruzada.

Debido a que las medidas VKM, XVAL/VKM y ECMA involucran a la varianza del *kriging*, a continuación se menciona cómo es que cambiar la combinación de C0 y C1 (siempre que sumen una constante, que en este caso es cuatro) afecta a dicha varianza. La razón por la cual tiene sentido dejar fija la meseta, es que este parámetro se estima en forma natural con la varianza muestral, en el caso de que el fenómeno de estudio sea estacionario fuerte (aunque dicho estimador sea sesgado por el hecho de que la muestra está correlacionada). Recuérdese que el tipo de datos que se analizan en este trabajo cumple con la propiedad de estacionaridad fuerte; luego, no tiene sentido considerar modelos con diferentes mesetas.

En el primer capítulo se comentó que, para un modelo con pepita fija, cambios en el parámetro C1 no afectaban a los pesos del *kriging* (siempre y cuando la proporción pepita a meseta se mantenga igual), pero sí a la varianza del *kriging*. Aquí se va a analizar cómo afectan cambios de C1 en la varianza del *kriging*, pero de tal manera que la meseta (C0+C1) sea la misma para todos los modelos que se estudien. (En este caso, también se verían afectados los pesos del *kriging* puesto que cambia la proporción pepita a meseta).

En particular, una medida que se ve directamente afectada por la varianza del *kriging* es VKM, puesto que es el promedio aritmético de estas varianzas. Recuérdese que la varianza del *kriging* es independiente de los valores muestrales y sólo depende del modelo de correlación espacial y de las distancias entre los puntos muestrales.

4.1. COMPORTAMIENTO DE LA VARIANZA DEL KRIGING Y DE VKM FRENTE A CAMBIOS EN LOS PARAMETROS DEL MODELO DE CORRELACION ESPACIAL

Cuando uno desea analizar qué pasa con el valor de una función real-valuada de más de una variable ($\mathbb{R}^k \rightarrow \mathbb{R}$) al cambiar

alguna de sus variables, lo primero que se ocurre es derivar con respecto a ellas y analizar las primeras y segundas derivadas parciales. En este caso, la función es la varianza del *kriging*, y los parámetros del modelo de variograma serían las variables de interés. Sin embargo, obsérvese que la varianza del *kriging* involucra también a los pesos del *kriging* y esto dificulta que se pueda expresar en forma explícita como función de los parámetros del variograma. Recuérdese que los pesos del *kriging* son la solución al sistema de ecuaciones (3) mencionado en el primer capítulo, de aquí que sea difícil expresarlos en términos de los parámetros del variograma explícitamente. Debido a esto, se decidió evaluar la varianza del *kriging* y VKM puntualmente para algunas combinaciones de los parámetros y entonces analizar el comportamiento de las medidas frente a cambios en los parámetros. A continuación se comentan los resultados al cambiar la pepita, el rango y la rejilla muestral.

4.1.1. CAMBIOS EN LA PEPITA

Para evaluar los comportamientos de VKM y de la varianza del *kriging* frente a cambios en los parámetros, se calcularon éstas para la rejilla regular (comentada en el segundo capítulo) y los siguientes modelos esféricos:

esférico (C_0 , [$4 - C_0$], 8.5),

donde $C_0 = \{ 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4 \}$.

Se analizó el comportamiento de la varianza del *kriging* en cada punto así como la media de estas varianzas (VKM) para cada modelo. En la figura 4.1 se graficó VKM contra el valor de la pepita, a partir de la siguiente tabla (en las gráficas se unieron los puntos con líneas):

CO (Pepita)	C1	VKM
0	4	0.589
0.5	3.5	1.136
1	3	1.634
1.5	2.5	2.105
2	2	2.556
2.5	1.5	2.986
3	1	3.393
3.5	0.5	3.766
4	0	4.045

También se observa en la misma figura que la varianza del *kriging* depende de la pepita en forma monótona creciente. Esto resulta lógico puesto que, cuando la pepita es más grande, significa que se tiene más ruido en las mediciones (o menos información proveniente de los datos). Intuitivamente se esperaría que haya más imprecisión en la predicción; lo que equivale a tener una varianza del *kriging* más grande.

4.1.2. CAMBIOS EN EL RANGO

De igual manera, se realizó un estudio con el fin de determinar cómo influían cambios del rango en el valor de la varianza del *kriging* y en VKM. Para este efecto, lo que se hizo fue calcular VKM para la misma rejilla de puntos muestrales y los siguientes modelos esféricos y exponenciales:

esférico ($C0$, $[4 - C0]$, Rango1) ,

exponencial ($C0$, $[4 - C0]$, Rango2) ,

donde $C0 = 0, 0.4, 2$,

Rango1 = 1, 1.25, 1.5, 1.75, 2, 2.25, ..., 10

y Rango2 = 1, 1.5, 2, 2.5, ..., 18 .

En todos los casos se observó que la varianza del *kriging* y también VKM disminuyen al incrementar el rango a partir de dos (véase la figura 4.2). Es decir, aparentemente VKM depende en

forma monótona decreciente del rango a partir de un valor pequeño, para el cual el modelo es prácticamente igual al de independencia espacial. El hecho de que VKM se incremente al aumentar el rango cuando éste es menor que dos, puede deberse a la marcada pendiente de los variogramas correspondientes cerca del origen.

El comportamiento decreciente de VKM se explica al recordar que un aumento en el rango significa que la "zona de influencia" del proceso es mayor y por tanto van a intervenir más puntos muestrales en cada predicción. Es decir, se tiene más información al predecir el fenómeno en cada punto que sea intermedio a los muestrales. El efecto que se tiene es el de una superficie de predicción más suave, por lo cual es explicable que la varianza de predicción en cada punto disminuya.

4.1.3. CAMBIOS EN RANGO Y PEPITA

Además de este resultado, también se observó el efecto conjunto en VKM de aumentar la proporción que representa la pepita con respecto a la meseta a la vez que se incrementa el rango, lo cual se muestra en la figura 4.3. El efecto es que VKM disminuye al aumentar el rango; pero para un rango fijo, un modelo con mayor pepita ocasiona un valor de VKM y de varianza del *kriging* mayor. El efecto global es como si la gráfica de VKM contra el rango se desplazara hacia la derecha al aumentar la pepita.

4.1.4. CAMBIOS EN LA REJILLA MUESTRAL

Dado que en el cálculo de VKM no intervienen los valores muestrales, sino que tan sólo influyen el modelo y las distancias entre los puntos muestrales, se pensó que sería de interés analizar el efecto en VKM de agrandar la rejilla de los puntos muestrales. Es decir, se calcularon los valores de VKM para dos rejillas regulares similares a la que se tenía pero en vez de que la distancia de separación fuese unitaria, se quintuplicó en un caso y, en un segundo caso, se multiplicó por diez. La figura 4.4

muestra estos resultados que son similares a los de incrementar la pepita: a mayor separación en la rejilla, para un mismo modelo, se tiene que VKM aumenta.

Adicionalmente se observa que VKM es constante hasta que el rango rebasa el valor de la mínima distancia entre los puntos de cada rejilla (5 y 10, respectivamente). Esto es lógico, puesto que como no hay información en la muestra de la relación espacial entre puntos que disten menos que la mínima distancia, todos los modelos cuyo rango sea menor que dicha distancia equivalen al modelo de independencia espacial. Para los modelos cuyo rango sea menor a la mínima distancia entre las observaciones, se tiene que la varianza del error de predicción es la mayor puesto que todos los puntos en la muestra se promedian con pesos iguales para cada punto que se quiere predecir. La superficie de predicción que se obtiene para un modelo así es como un plano a la altura del promedio de todos los puntos, con "brincos" en cada punto muestral. (Véase la figura 4.5).

4.1.5. COMENTARIOS

Resulta claro, entonces, que las dependencias mencionadas de VKM con respecto a los parámetros de los modelos de correlación espacial y de la rejilla van a influir en las medidas de XVAL/VKM y ECMA *independientemente de la muestra en cuestión de que se trate*. El comportamiento de VKM ocasionará que se tienda a elegir modelos con rango lo más grande posible y menor pepita. Obsérvese que esto es completamente independiente del variograma empírico de una muestra dada y del modelo que se sugiera por el método de ajuste visual o algún otro de mínimos cuadrados.

Un resultado interesante, e independiente de la muestra de datos en estudio, es que se observó que la varianza del *kriging* es menor (y por tanto también VKM) cuando se ajustan modelos esféricos que cuando se usan modelos exponenciales. La figura 4.2 muestra las gráficas de VKM contra el rango cuando se ajustan

modelos esféricos y exponenciales con C_0 y C_1 fijos. Es claro que la curva de ajustes exponenciales queda por arriba de la de los ajustes esféricos. Esto indica que los modelos esféricos tienen asociada una varianza del error de predicción menor que los modelos exponenciales, independientemente de la muestra que se tenga.

A continuación se exhibe la metodología seguida para estudiar las tres medidas restantes de validación cruzada.

4.2. METODOLOGIA SEGUIDA PARA EVALUAR A XVAL, ECMA Y XVAL/VKM

Para poder analizar el comportamiento de las tres medidas de validación cruzada restantes (los cuales sí dependen de los valores muestrales) y que son XVAL, ECMA y XVAL/VKM, se realizó lo siguiente para cada una de las muestras de los dos conjuntos de muestras correlacionadas y de los cinco conjuntos de muestras independientes:

1. Se calcularon las tres medidas de validación cruzada (XVAL, XVAL/VKM y ECMA) tras ajustar los siguientes modelos esféricos y exponenciales:

esférico(C_0 , [$4 - C_0$], Rango1),

exponencial(C_0 , [$4 - C_0$], Rango2),

donde $C_0 = 0, 0.4, 2$,

Rango1 = 1, 1.25, 1.5, 1.75, 2, 2.25, ..., 10

y Rango2 = 1, 1.5, 2, 2.5, ..., 18 .

Obsérvese que las tres pepitas estudiadas representan el 0, 10 y 50% de la meseta que es igual a cuatro. En total se ajustaron 37 modelos para cada familia.

2. Se graficó para cada conjunto de muestras correlacionadas e

independientes y para cada rango, la media de la medida de validación cruzada en cuestión. También se graficó una "banda de confianza" conformada por los cuantiles 5 y 95 de los valores de la medida para cada rango de las 100 muestras del conjunto en cuestión.

En seguida se detallan los resultados obtenidos para XVAL, ECMA y XVAL/VKM.

4.2.1. COMPORTAMIENTO DE XVAL

Esta medida es de las más comúnmente usadas para elegir entre dos o más candidatos de modelos de variograma para una muestra de datos dada. Usualmente, se calcula tan sólo para unos pocos modelos, y no es común calcularlo para una amplia colección de ellos. Dado que XVAL depende de Z^* , hereda de éste el ser invariante frente a cambios en C_1 , siempre que se mantenga fija la proporción de la pepita con respecto a la meseta. Esto implica que XVAL será igual para modelos de una cierta familia con valores de C_1 muy diferentes, siempre y cuando tengan el rango y la proporción pepita a meseta iguales.

Las figuras 4.6, y 4.7 muestran las curvas de validación cruzada (que se obtienen al unir los valores de XVAL para cada rango con líneas) al ajustar modelos esféricos y exponenciales a una de las muestras esféricas y a una de las exponenciales, respectivamente, así como el variograma empírico de momentos para cada una de las dos muestras. Como se puede ver, las curvas de ajuste de modelos esféricos oscilan (este comportamiento se observó con todas las muestras correlacionadas). Es decir, se tienen varios mínimos locales, y el mínimo global las más de las veces estaba alejado del valor de rango teórico de 8.5 con el cual se generó a las muestras.

El promedio de los mínimos globales de los ajustes esféricos (sin pepita) para el conjunto de muestras esféricas fue de 5.9 y

para el conjunto de muestras exponenciales fue de 4.25.

Se observó algo muy interesante al promediar para cualesquiera diez o más muestras esféricas los valores de XVAL para cada valor de rango obtenidos al ajustar modelos esféricos: el mínimo del promedio de las curvas se obtuvo en un rango muy cercano a 8.5. Cuando se promediaron los valores de XVAL por rango para todas las muestras esféricas, el mínimo del promedio fue exactamente 8.5.

Obsérvese que si las muestras fuesen independientes considerar el promedio de los valores de XVAL equivale en cierta forma a pensar que se tuvo una muestra enorme, de tamaño 10,000, con cien observaciones que se midieron en cada punto de la rejilla muestral. Aunque en este caso las muestras están correlacionadas, esto sugiere que el comportamiento oscilatorio de la curva de XVAL para ajustes esféricos posiblemente se reduzca al aumentar el tamaño de muestra.

Esta observación sugiere la idea de aplicar técnicas de *bootstrap*, las cuales permiten a partir de una muestra dada, obtener muchas muestras "réplica"; entonces se calculan las curvas de XVAL con alguna familia de modelos para las muestras "réplica", se promedian y con el mínimo de este promedio se puede estimar el rango. En la siguiente sección, en el inciso 5, se menciona con más detalle esta idea.

4.2.2. RESULTADOS DEL ANALISIS DE XVAL

1. Al analizar las curvas de XVAL contra rango al ajustar los modelos esféricos y exponenciales, se observó que un problema del ajuste de modelos esféricos consiste en que tienen varios mínimos locales. Hubo casos en donde estos mínimos tenían casi el mismo valor de XVAL, y sin embargo se alcanzaban para rangos diferentes. En el caso de que los valores muestrales tengan una dispersión grande en comparación a su media, puede ocurrir que modelos cuyos

parámetros difieren muy poco en el rango originen superficies de predicción que sean radicalmente distintas.

En cambio, la curva de XVAL con modelos exponenciales se comportó en forma monótona decreciente. Se observó que en la mayoría de los casos el rango para el cual los decrementos marginales eran casi insignificantes estaba cercano a 8.5. Había ocasiones en donde esta curva sí alcanzaba un "mínimo", y éste parecía ser único y cercano al valor teórico de 8.5.

La figura 4.8 muestra las superficies que se generan con la muestra esférica de la figura 4.6 al usar el modelo esférico de rango igual a cinco, que es en donde alcanza XVAL el mínimo global; también se exhibe la superficie generada con el modelo exponencial de rango 6.75 que es en donde se estabilizan los decrementos marginales de la curva de XVAL de ajuste de modelos exponenciales. Es evidente que para esta muestra, las dos superficies son muy similares y en realidad no hay gran diferencia entre elegir un modelo o el otro.

2. Otro hecho interesante fue que para casi todas las muestras correlacionadas, las diferencias entre la curva de XVAL con modelos esféricos y la de XVAL con modelos exponenciales disminuían en valor absoluto a partir de un valor de rango cercano al teórico (a partir de 6.5 o 7 en casi todas las muestras estudiadas). De aquí que surja la idea de explorar más a fondo el comportamiento de las diferencias entre las curvas de XVAL contra rango para diversas familias de modelos para ver si sirven para estimar el rango del modelo que mejor describa la correlación espacial de la muestra.

3. En cuanto al efecto de incrementar la pepita en XVAL, se observó que un incremento en la proporción de la pepita con respecto a la meseta ocasiona que la curva se desplace hacia la derecha como se ve en la figura 4.9. Si se recuerda que las

muestras simuladas fueron generadas sin pepita, entonces esto sugiere que al usar XVAL se elegirán en forma adecuada modelos sin pepita (puesto que para cada valor de rango, el menor valor de XVAL se alcanza para el modelo con $C_0=0$). Esto parece indicar que XVAL puede ayudar a estimar adecuadamente el valor de la pepita.

4. Una característica muy interesante del comportamiento de XVAL es que parece indicar muy bien cuándo existe independencia espacial en los datos. Al calcular XVAL para las muestras independientes, en algunos casos como el de las muestras normales con media 25 y varianza 10,000, los valores que XVAL tomó fueron exageradamente grandes con respecto a las unidades muestrales (eran del orden de 10^4). En otros casos como el de las muestras t , XVAL tomó un valor mayor a uno para todos los rangos en más del 95% de los casos. Las figuras 4.10 y 4.11 muestran lo que se acaba de describir.

Aparentemente, XVAL es una buena herramienta para indicar independencia, al igual que el variograma empírico como se vio en el capítulo anterior. Existen en la literatura pruebas para probar la hipótesis de independencia espacial, entre otras la de Cliff y Ord, (1971) (mencionada en Johnson y Kotz 1983, V.2, 10-13); sin embargo, el usar XVAL y el variograma empírico podría ser más simple. Tal vez valdría la pena explorar en un futuro el uso de éstos en alguna estadística de prueba para la hipótesis de independencia espacial.

5. El *bootstrap* es una técnica no paramétrica de muestreo repetido que requiere un uso intensivo de la computadora y que se utiliza para obtener las distribuciones de una clase amplia de estadísticas muestrales. Es decir, dada una muestra de tamaño n , se muestrean con reemplazo las n observaciones y se obtienen muchas muestras "réplica" de tamaño n . Entonces se calcula la estadística de interés para cada una de estas muestras; con la distribución empírica de estos n valores, se estima la distribución de muestreo del estimador en cuestión. La teoría

existente para este tipo de técnicas supone que las observaciones son independientes e idénticamente distribuidas; también que la muestra en cuestión representa bien a la población y que el error en las mediciones es mínimo. Solow (1985) explica como adaptar estas técnicas al caso de tener datos con correlación espacial. La idea es transformar las observaciones a otras independientes, con éstas realizar el *bootstrap*, y luego aplicar la transformación inversa a las muestras "réplica" para poder trabajar con ellas.

Sin embargo, un problema para aplicar el *bootstrap* en este caso, consiste en tener que conocer o estimar la estructura de correlación espacial para poder transformar a independencia, lo cual remite al problema inicial de estimar adecuadamente los parámetros de algún modelo de correlación espacial. A manera de resumen a continuación se enuncian los principales resultados que aportó el análisis.

4.2.3. COMPORTAMIENTO DE ECMA Y DE XVAL/VKM

El hecho de que ambas medidas involucren a VKM fue patente en su comportamiento. Las dos medidas resultaron ser extremadamente sensibles a cambios en la proporción de la pepita con respecto a la meseta. Cuando se ajustaron modelos tanto exponenciales como esféricos con proporción de pepita mayor, se observó que al graficar a ECMA y a XVAL/VKM contra el rango, las gráficas de las medias para las cien muestras esféricas y las exponenciales de dichas medidas se recorrían hacia abajo (se ilustra el caso de muestras esféricas; el caso de las exponenciales fue similar; se grafica la media de los valores de las medidas de las 100 muestras en cuestión; véanse las figuras 4.12 y 4.13).

Por una parte, en la figura 4.12 se ve que ECMA estima adecuadamente el rango para los modelos sin pepita (véase el rango para el cual las curvas cortan la línea horizontal a la altura de uno); sin embargo, obsérvese que la medida es indiferente entre un modelo esférico (0,4,8.2) y otros esféricos con proporción de

pepita mayor y un rango muchísimo más grande. Es decir, hay más de un modelo que cumple con el criterio de que ECMA sea igual a uno, lo cual es un gran inconveniente pues las superficies de predicción pueden ser dramáticamente diferentes. Esto se debe a que para un mismo rango, a mayor pepita se tiene mayor varianza del *kriging*; luego, como la varianza del *kriging* aparece en el denominador de estas medidas, ocasiona que sean más chicas y que alcancen el valor de uno (que es el criterio de optimización) para un rango más grande.

También se observó que ECMA y XVAL/VKM no denotan un comportamiento diferente en el caso que exista independencia espacial en comparación a cuando no la hay. Las figuras 4.14 y 4.15 exhiben el comportamiento de estas medidas para el conjunto de muestras independientes t . Las medidas se comportan en forma similar a cuando las muestras tenían correlación espacial. Es decir, si uno se guía por estos criterios, se elegiría un modelo con correlación espacial en forma errónea. Lo que ocurre es que el efecto de VKM se hace muy patente y anula la información que pudiese aportar la muestra.

En resumen, en el caso de muestras con correlación espacial, estas medidas de validación cruzada presentan el problema de elegir por igual a modelos muy diferentes entre sí. Por otra parte, no indican cuando hay independencia espacial en las muestras pues también eligen en estos casos a algún modelo con correlación espacial. De aquí que se recomiende no emplear a ECMA y XVAL/VKM para decidir cual es el modelo de correlación espacial que mejor describe a una muestra dada.

4.3. COMENTARIOS

Es claro que al estudiar una rejilla regular, un tamaño de muestra no tan pequeño y muestras simuladas que son combinaciones lineales de normales, se están analizando casos no "patológicos". De aquí que el escenario sea óptimo para evaluar el comportamiento

de las medidas de validación cruzada en situaciones favorables. Debido a esto, se cree que cualquier anomalía que exhiban dichos criterios se podría extender y/o empeorar en otros casos que fuesen más irregulares en cuanto al diseño de muestreo, distribución de los valores muestrales o menor tamaño de muestra.

En resumen, el problema era averiguar si los cuatro criterios que se mencionaron de validación cruzada eran buenos o no para encontrar al modelo de correlación espacial que mejor describe a una muestra dada y, por tanto, ayudar en la estimación de los parámetros de éste. Los parámetros que son más importantes de estimar mediante el uso de validación cruzada en realidad son la pepita y el rango, puesto que la meseta se estima en forma natural con la varianza muestral, en el caso de que la variable regionalizada sea estacionaria fuerte.

Se observó que tanto ECMA, XVAL/VKM y VKM por depender de la varianza del *kriging*, tienden a elegir modelos con rango mayor y menor pepita, sin importar la muestra de que se trate ni si existe correlación espacial o no. Por otra parte, XVAL aparenta tener mucho potencial para estimar al rango y a la pepita adecuadamente (a la vez que se considera la gráfica del variograma empírico de momentos), así como para indicar si hay independencia espacial o no. Queda abierta la puerta para explorar alguna prueba de independencia espacial que involucre a XVAL.

Usar las medidas de validación cruzada, sin analizarlas más profundamente o sin seguir alguna metodología específica para evaluar la bondad de ajuste de algún modelo de correlación espacial, tiene sus bemoles, como expresan Isaacs y Srivastava (1989):

"Reducing cross validated results to a few summary statistics, often only a global mean and a variance, is a wasteful (but common) practice; there is much to be learned from a thorough spatial analysis of the residuals".

Sintetizando, dada una muestra, se recomienda fuertemente que se calculen y grafiquen las curvas de XVAL contra rango para varias familias de modelos (como los esféricos, exponenciales, lineales y potenciales) y que se analicen los valores mínimos de estas curvas para elegir a modelos "óptimos".

Lo que se propone en este trabajo es realizar con una muestra dada lo siguiente. Primero, obtener el variograma empírico de la manera que se sugirió en el tercer capítulo y calcular la varianza muestral S^2 . En el caso en que el variograma empírico sugiera que el fenómeno de estudio es estacionario fuerte, se puede estimar la meseta (C_0+C_1) con S^2 . El siguiente paso es calcular XVAL para modelos que tengan esta meseta, pero falta decidir que valores de pepita y rango se van a usar. Tras observar al variograma empírico se proponen k posibles valores de pepita C_{0_i} , $i=1,2,\dots,k$. En cuanto a los rangos, una forma de proceder es la siguiente:

Sea $mind$, la mínima distancia entre los puntos muestrales y $maxd$, la máxima distancia. Sea inc , un incremento igual a $(maxd-mind)/n$ donde n es un entero que se elige arbitrariamente, pero se sugiere sea mayor a 20. Entonces se calcula XVAL para los siguientes modelos:

* familia de esféricos:

esférico (C_{0_i} , [$S^2 - C_{0_i}$], rango),

donde $i=1,2,\dots,k$, y $rango=\{mind, mind+inc, mind+2. inc, \dots, maxd\}$;

* familia de exponenciales:

esférico (C_{0_i} , [$S^2 - C_{0_i}$], rango),

donde $i=1,2,\dots,k$, y $rango=\{mind, mind+inc, mind+2. inc, \dots, maxd\}$;

similarmente, se calcula XVAL para familias de modelos potencia, lineales, etc.

Lo que sigue es construir una gráfica de XVAL contra el rango para cada familia de modelos y hacer esto para cada una de las k pepitas elegidas. Así el análisis se reduce a k gráficas de XVAL

contra rango y a la gráfica del variograma empírico. De esta manera, se tienen más bases para estimar los parámetros del "modelo óptimo", en forma visual al elegir la combinación de rango y pepita en donde las gráficas alcancen el "mínimo". Es posible que haya más de un modelo que se proponga a partir de estas gráficas.

En general, se recomienda ser cautos con el comportamiento de la curva de XVAL contra rango al ajustar modelos esféricos debido a las oscilaciones y mínimos locales múltiples que se han observado en XVAL, sobre todo en el caso de que la muestra no sea grande. Se sugiere en especial considerar el punto en que se estabilizan los decrementos marginales de la curva de XVAL para modelos exponenciales, puesto que funcionó bien para el tamaño de muestra y los casos aquí estudiados.

Después, se sugiere graficar los modelos de variograma seleccionados en la misma gráfica que el variograma empírico fijándose en el ajuste para distancias pequeñas y así facilitar la elección del mejor modelo. Hay que tener en cuenta que cuando haya varios modelos entre los cuales elegir que a mayor rango y menor pepita, la superficie de predicción será más suave.

También se recomienda graficar las superficies que se predican con estos modelos a partir de la muestra dada para estar conscientes de las regiones en donde algún modelo predice por arriba o por abajo en comparación con otro. Además se recomienda realizar un diagnóstico gráfico de los residuos de validación cruzada, $(Z_1^* - Z_1)$, para cada uno de los modelos que se hayan elegido, graficando los residuos contra los valores observados o calculando el histograma de residuos estandarizados.

Un camino por explorar es el de analizar el comportamiento de las diferencias entre las curvas de XVAL para ajustes de modelos de diversas familias. Es posible que se pueda establecer un criterio usando estas diferencias que sirva para estimar el rango y también a la pepita.

Otro camino que se ve prometedor es que, a partir de una muestra dada, se estudie en forma visual o analítica la "distancia" entre la superficie de predicción con un modelo de correlación espacial, y alguna otra superficie que se tome como marco de referencia (por ejemplo, la superficie bajo un modelo de rango menor a la mínima distancia muestral, que equivale a independencia espacial). Entonces se podría estudiar la sensibilidad de esta "distancia" frente a cambios en los parámetros del modelo.

Por último, debido al comportamiento de ECMA, VKM y XVAL/VKM que depende fuertemente de los parámetros del variograma, se recomienda no utilizar a estas medidas como criterio para encontrar o decidir cual modelo de correlación espacial describe mejor a una muestra dada. Más aún, esta dependencia parametral es tan fuerte, que incluso en el caso de muestras independientes estas medidas eligen erróneamente modelos con correlación espacial.

Sin embargo, antes de muestrear, se recomienda analizar el comportamiento de VKM para diferentes familias y colecciones de modelos de variogramas así como de diversas rejillas de muestreo. El objetivo es diseñar la rejilla muestral óptima en el sentido de lograr un costo reducido y una menor varianza del error de predicción.

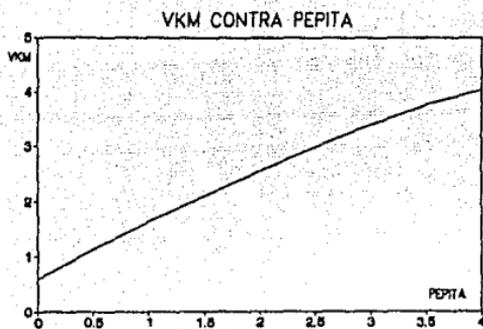


FIGURA 4.1

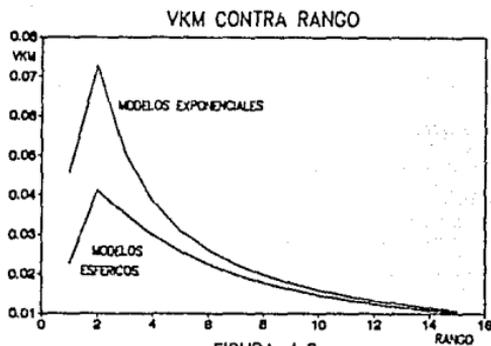


FIGURA 4.2

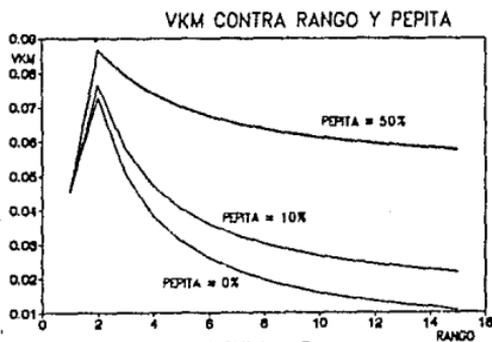
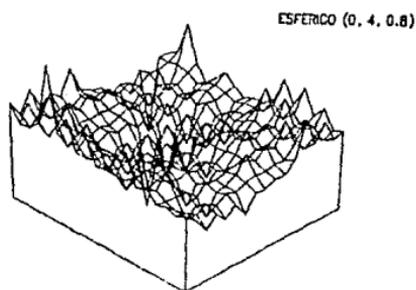
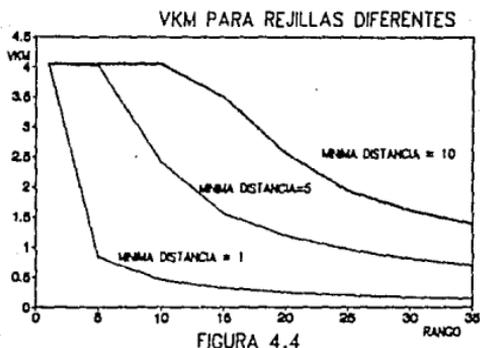


FIGURA 4.3



MUESTRA ESFERICA
(CORRELACION ESPACIAL)

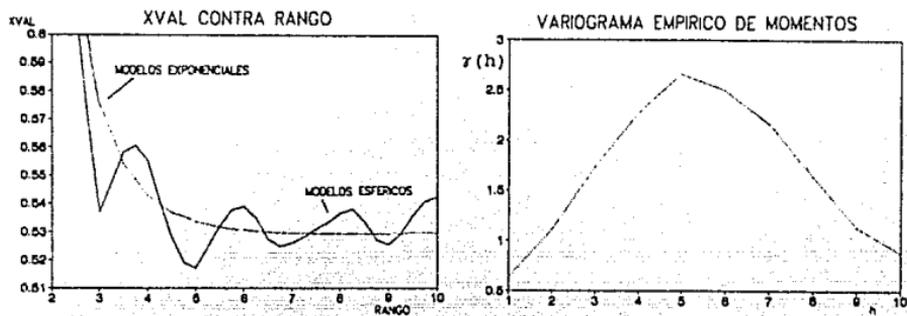


FIGURA 4.6

MUESTRA EXPONENCIAL
(CORRELACION ESPACIAL)

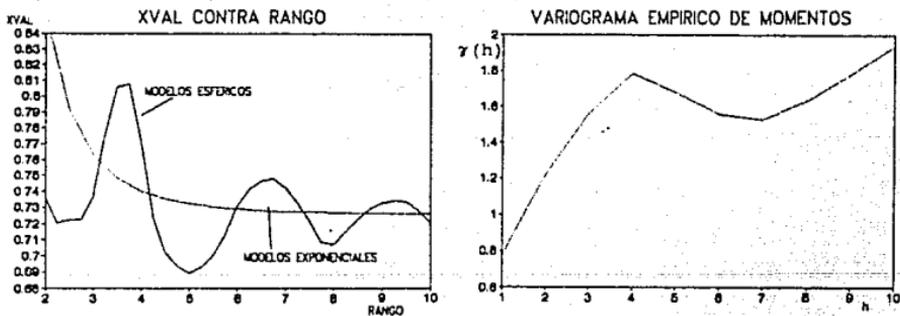


FIGURA 4.7

MUESTRA ESFERICA
(CORRELACION ESPACIAL)

SUPERFICIE DE PREDICION BAJO
ESFERICO (0, 4, 5)

SUPERFICIE DE PREDICION BAJO
EXPONENCIAL (0, 4, 6.75)

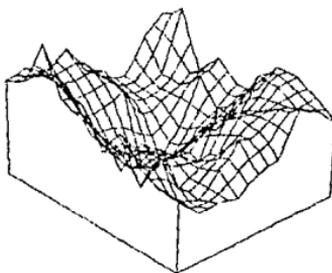
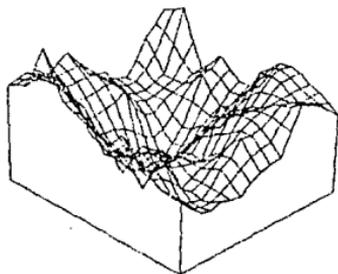


FIGURA 4.8

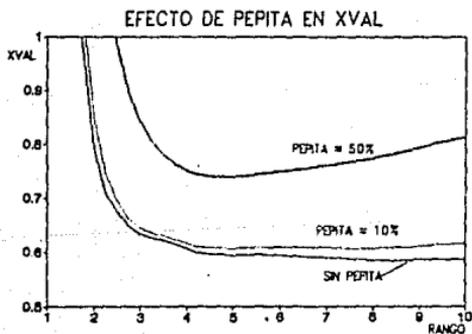


FIGURA 4.9

XVAL

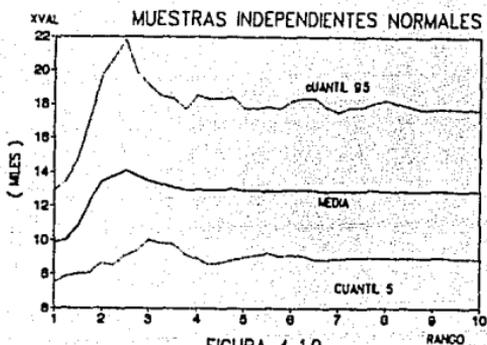


FIGURA 4.10

XVAL

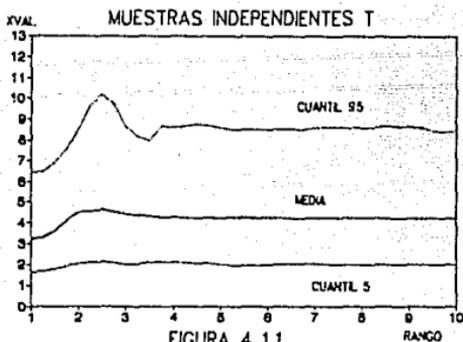
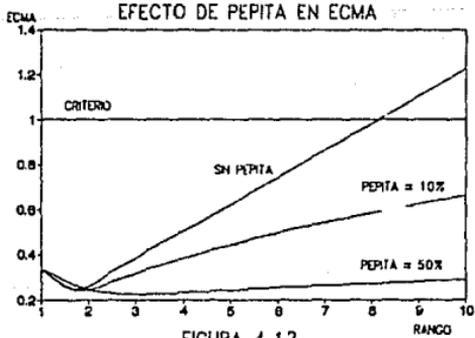


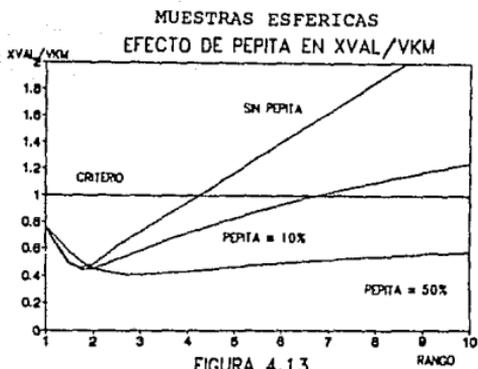
FIGURA 4.11

MUESTRAS ESFERICAS
EFECTO DE PEPITA EN ECMA

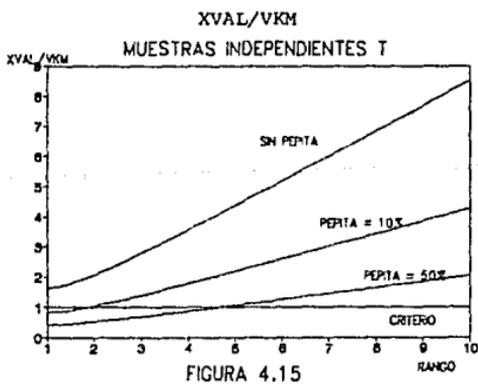
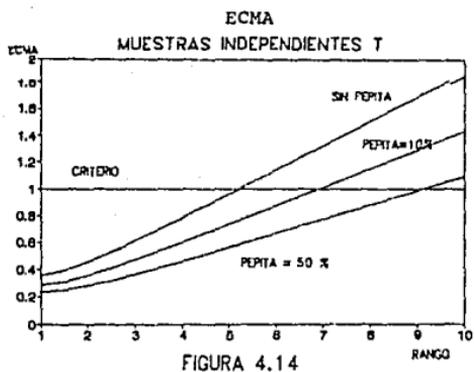


AJUSTE DE
MODELOS
ESFERICOS

FIGURA 4.12



AJUSTE DE
MODELOS
ESFERICOS



Capítulo 5

Distribución estadística de XVAL

La falta de pruebas estadísticas que ayuden a determinar cuándo alguna de las medidas de validación cruzada es lo suficientemente pequeña, o cercana a un cierto valor, motiva a encontrar la distribución estadística de ellas. Así, se podría llevar a cabo algún contraste de hipótesis para evaluar la bondad de ajuste de un modelo de correlación espacial específico. El conocer la distribución de alguna de las medidas de validación cruzada permite usar a dicha medida, calculándola para una muestra dada, como estadística de prueba de un contraste de hipótesis del siguiente estilo:

$$H_0: \gamma(h) = \gamma_0(h; C_0, C_1, a) \quad (\text{con } C_0, C_1, a, \text{ fijos y conocidos}),$$

$$\text{contra} \quad H_a: \gamma(h) \neq \gamma_0(h; C_0, C_1, a).$$

Se eligió estudiar la distribución de la medida XVAL puesto que, en caso de que el proceso $Z(x)$ sea gaussiano, se le puede interpretar fácilmente como una forma cuadrática de variables normales. En la literatura estadística el tema de la distribución de este tipo de formas cuadráticas ha sido discutido ampliamente por diversos autores (ver Johnson y Kotz, 1970).

Es decir, el procedimiento consiste en identificar a XVAL como una forma cuadrática de normales; después, se transforma esta forma cuadrática en otra, que representa una suma ponderada de variables aleatorias independientes χ^2 -cuadradas, para la cual se conoce su función característica. Conociendo la expresión de ésta, mediante el algoritmo numérico de Imhof (1961), se encuentran los cuantiles que se deseen de la función de distribución de la forma cuadrática equivalente. Por construcción, la distribución de XVAL es igual a la de esta forma cuadrática; por tanto, ya se tiene la distribución deseada.

Lo que se exhibe en este capítulo representa un esfuerzo inicial para resolver el problema de encontrar la distribución de XVAL en el caso general. Aquí se obtiene la distribución de esta medida bajo condiciones muy restrictivas, como son el conocer los parámetros y el tipo del modelo de variograma. Obsérvese que tener este conocimiento es poco frecuente en la realidad, así como el hecho de tener un fenómeno gaussiano, pues las variables geoestadísticas son asimétricas frecuentemente. Sin embargo, se espera que este primer paso sea de utilidad para que, más adelante, se encuentre la distribución de XVAL bajo condiciones generales.

5.1. PLANTEAMIENTO TEORICO

Para poder interpretar a XVAL como una forma cuadrática de normales hay que realizar los siguientes supuestos:

1. $Z(x)$, $x \in D \subset \mathbb{R}^2$ es un proceso gaussiano y es estacionario fuerte.
2. $\text{VAR} [Z(x)]$ es conocida.
3. $\gamma_{ij} = \gamma(|x_i - x_j|)$ está bien definida y se conocen los parámetros.

$$\text{Sea } n \cdot (XVAL) = \sum_{i=1}^n (Z_i^* - Z_i)^2 = \sum_{i=1}^n \lambda_i^2, \text{ donde } \lambda_i = (Z_i^* - Z_i).$$

Debido a que el predictor de kriging es insesgado, a que el proceso es gaussiano y que λ_i es una combinación lineal de variables normales, se tiene que λ_i se distribuye como una normal con media cero y cierta varianza conocida a la que se denotará como σ_i^2 , en otras palabras,

$$E [A_i] = E [Z_i^*] - E [Z_i] = k - k = 0;$$

$$\text{VAR} [A_i] = \sigma_i^2;$$

$$A_i \sim N(0, \sigma_i^2) \text{ para toda } i=1, \dots, n.$$

Es claro que dos términos A_i y A_j no son independientes, puesto que en el cálculo de Z_j^* interviene Z_i (y viceversa). A la covarianza entre cualesquiera dos términos se le llamará σ_{ij} , y se calcula de la siguiente manera:

$$\begin{aligned} \sigma_{ij} = & \sum_{k \neq i} \sum_{v \neq j} \lambda_{ik} \lambda_{jv} \text{COV}(|x_k - x_v|) - \sum_{k \neq i} \lambda_{ik} \text{COV}(|x_k - x_j|) - \\ & - \sum_{v \neq j} \lambda_{jv} \text{COV}(|x_v - x_i|) + \text{COV}(|x_i - x_j|). \end{aligned}$$

Si se considera el vector n-dimensional Δ que tiene como componentes a las A_i , es claro que dicho vector se distribuye como una normal n-variada con vector de medias cero y matriz de varianzas V :

$$\Delta = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{pmatrix} \sim N_n(0, V) \text{ donde } V = \begin{pmatrix} \sigma_1^2 & \dots & \sigma_{1n}^2 \\ \vdots & & \vdots \\ \sigma_{n1}^2 & \dots & \sigma_n^2 \end{pmatrix}.$$

La forma cuadrática de Δ asociada a la matriz identidad es

$$Q(\Delta) = \sum_{j=1}^n \sum_{i=1}^n w_{ij} A_i A_j = \sum A_i^2 = n \cdot (XVAL),$$

donde $w_{ij} = I_{ij}$.

La función de densidad de probabilidad de Δ es

$$P_A(\Delta) = (2\pi)^{-n/2} |V|^{-1/2} \exp\left(-\frac{1}{2} \Delta' \cdot V^{-1} \cdot \Delta\right).$$

Se desea encontrar la distribución de $Q(\Delta)$:

$$P [Q(\Delta) \leq y] = \int_{\Delta' \Delta \leq y} P_A(\Delta) d\Delta, \quad y \in \mathbb{R}.$$

Dado que V es positiva definida, se puede diagonalizar. Existe M ortogonal ($M'M = I$), tal que:

$$V = M \begin{pmatrix} \beta_1 & 0 & \dots & 0 \\ 0 & \beta_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \beta_n \end{pmatrix} M' = M B M',$$

donde B es una matriz diagonal con los eigenvalores de V en la diagonal. Si se realiza el siguiente cambio de variable: $Z = M' \cdot \Delta$ se tiene que:

$$E [Z] = M' E [\Delta] = \underline{0} \quad y$$

$$\text{VAR} [Z] = E [Z Z'] = M' E [\Delta \Delta'] M = M' V M = M' [M B M'] M = B.$$

Es decir, Z se distribuye como una normal n variada con vector de medias cero y matriz de varianza B .

Si se hace otro cambio de variable:

$$Y_i = z_i / \sqrt{\beta_i}, \quad \text{para } i=1, \dots, n, \quad \text{se tiene que: } Y \sim N(\underline{0}, I).$$

Por lo tanto, al considerar las formas cuadráticas $Q(Z)$ y $Q(Y)$ asociadas a las matrices identidad y B , respectivamente, se

observa que:

$$Q(\Delta) = \sum \Delta' \Delta = \sum (MZ)' (MZ) = \sum Z' Z = Q(Z) = \sum z_i^2 = \sum \beta_i y_i^2 = Q(Y).$$

Es decir, $Q(\Delta)$ es igual a una suma ponderada por los eigenvalores de V de n normales estándar elevadas al cuadrado (j i-cuadradas con un grado de libertad).

El problema inicial era encontrar la distribución de $Q(\Delta)$, pero basta con encontrar la de $Q(Y)$ puesto que son iguales. La ventaja de centrar la atención en la distribución de $Q(Y)$ es que resulta sencillo encontrar la función característica de esta forma cuadrática. Como se sabe que la función característica de una variable aleatoria X que se distribuye como j i-cuadrada con un grado de libertad y multiplicada por una constante φ es

$$E [\exp (itX)] = (1-2it\varphi)^{-1/2},$$

y también se sabe que la función característica de una suma de variables aleatorias independientes es el producto de sus funciones características, la función característica de $Q(Y)$ es, por tanto,

$$E [\exp (itQ(Y))] = \prod_{j=1}^n (1-2it\beta_j)^{-1/2}.$$

Imhof en 1961 publicó en *Biometrika* un algoritmo para encontrar la función de distribución de formas cuadráticas de normales. Dicho algoritmo se basa en el hecho de que la función de densidad y la función característica constituyen un par transformado de Fourier (ver Brigham, 1974, p.11-15). Por lo cual, tras usar el teorema de inversión de Fourier (ver Breiman,

1968, teorema 8.39) y la aproximación de Imhof, se pueden encontrar los cuantiles teóricos que se desean de la distribución de $n(XVAL)$.

5.2. RESULTADOS

Para el modelo de semivariograma esférico con rango 8.5, meseta 4 y sin pepita, se calculó la matriz V . Se eligieron estos parámetros puesto que con ellos fueron generadas las 100 muestras esféricas que ya se tenían; así se podría observar la distribución empírica de $XVAL$ bajo la hipótesis nula y comparar algunos cuantiles contra los teóricos.

Después de diagonalizar a V y usar sus eigenvalores como los pesos de una suma ponderada de χ^2 -cuadradas con un grado de libertad, se aplicó el algoritmo de Imhof para encontrar algunos cuantiles teóricos de interés. Los resultados se obtuvieron usando un programa en Gauss del algoritmo de Imhof. Los cuantiles que se encontraron fueron el 5, el 90, el 95 y el 97.5.

Por otra parte, se obtuvieron los mismos cuantiles empíricos a partir de los valores que ya se tenían, tras haber calculado $XVAL$ para las 100 muestras esféricas bajo el supuesto del modelo esférico con parámetros conocidos (rango 8.5, meseta 4 y sin pepita). Se observó que los cuantiles empíricos para 100 muestras coincidían prácticamente con los teóricos, como se muestra a continuación:

<u>CUANTIL</u>	<u>EMPIRICO</u>	<u>TEORICO</u>
5	0.438	0.432
90	0.748	0.725
95	0.768	0.773
97.5	0.791	0.807

El cuantil que presenta una diferencia mayor (de 0.023),

entre los valores empírico y teórico, fue el cuantil 90. Esto implica que, para un nivel de 90% de confianza, de acuerdo al cuantil teórico, se rechazaría la hipótesis nula en forma errónea en un poco más del 10% de los casos del conjunto de 100 muestras con correlación espacial esférica estudiadas.

5.3. COMENTARIOS

Es poco frecuente que se tenga una idea *a priori* de la familia de modelo de variograma que describe a un fenómeno, así como de los valores exactos que toman sus parámetros. De aquí que tal vez no tenga mucho sentido la hipótesis nula que se ha planteado. Aún más, faltaría analizar la potencia de la prueba aquí exhibida. Un análisis de este tipo requiere de muchas horas de tiempo máquina puesto que habría que calcular la distribución teórica de XVAL bajo una enorme variedad de combinaciones de parámetros del modelo esférico y de otros modelos alternativos. Debido a que casi nunca se tiene una idea *a priori* del modelo exacto que describe los datos, se decidió no proseguir con el análisis en esta dirección.

Existe un problema adicional y es que la mayoría de los fenómenos naturales no se distribuyen en forma gaussiana, sino que en general tienen una distribución bastante asimétrica. En el caso de tener una muestra pequeña y desconocer la media y varianza de su distribución, el intento de transformar las observaciones a gaussianas puede ser muy inexacto y afectar a la estadística de prueba en el sentido de que no provenga ya de la distribución teórica encontrada.

El mencionar la distribución de XVAL en la tesis obedece a que, aparentemente, no existe en la literatura geoestadística algún intento por encontrar las distribuciones de los criterios de validación cruzada. Es por eso que se consideró importante incluir un capítulo que versara sobre este tema pues representa una línea de investigación muy interesante, abierta y poco explorada.

CONCLUSIONES

Este trabajo pretende ser de utilidad para las personas que se enfrentan a problemas que involucran fenómenos con distribución espacial. Se espera que las ideas que aquí se mencionan contribuyan a que el análisis y estimación de la correlación espacial de un fenómeno se realice más sistemáticamente y con más elementos. Las conclusiones principales de esta tesis se comentan en los siguientes párrafos.

Al igual que en otros tipos de análisis estadístico es importante realizar un buen diseño de muestreo con miras a obtener un buen estimador del variograma, así como del fenómeno espacial, en puntos donde no hubo observación. Para estos fines, resulta útil estudiar al promedio de las varianzas del kriging para distintas rejillas muestrales y varios modelos de variogramas. También conviene analizar el número de las parejas de distancias entre puntos muestrales para cada una de las rejillas de muestreo que se consideren, de manera que el variograma empírico sea más preciso para rezagos pequeños, donde es muy importante modelar bien para llegar a una buena estimación del fenómeno espacial.

En cuanto a la estimación del variograma, el estimador de momentos presentó un comportamiento aceptable, sobre todo si la estimación se basa en los valores del variograma empírico para los primeros rezagos y si la muestra es moderada o grande y no hay puntos discrepantes en ella. Se recomienda no utilizar el estimador del variograma de Cressie y Hawkins puesto que puede presentar un sesgo inaceptable básicamente debido a que no considera la correlación existente de las raíces cuadradas de los incrementos de la variable regionalizada, y además porque involucra varias aproximaciones y supuestos que fácilmente pueden no sustentarse. Cuando haya observaciones extremas en la muestra,

se sugiere calcular el variograma con y sin ellas, estudiando así la sensibilidad del estimador frente a valores aberrantes. Si el variograma empírico muestra evidencia de independencia espacial, una manera de confirmar esta hipótesis es calculando a XVAL de la manera que se hizo en el capítulo cuarto, para la muestra en cuestión.

Para estimar la meseta del modelo de correlación (bajo el supuesto de que los datos son estacionarios fuertes), se recomienda usar la varianza muestral S^2 . En cuanto a la estimación del rango y de la pepita, es aconsejable utilizar a XVAL en la forma sistemática que se llevó a cabo en el capítulo cuarto. Después, se sugiere elegir el modelo de correlación espacial con base en los resultados del análisis de XVAL y de la inspección visual del variograma empírico (sobre todo de los primeros rezagos). Así mismo, conviene graficar las superficies de predicción que resultan con el o los modelos elegidos, para estar conscientes del tipo de superficie que se genera con cada modelo y las implicaciones que esto tiene para el problema en cuestión. Finalmente, se sugiere llevar a cabo un diagnóstico de los errores de predicción en los puntos muestrales (i.e. realizar los métodos gráficos de validación cruzada mencionados en el primer capítulo) para los modelos escogidos y así ayudar en la elección de uno de ellos.

Una ventaja de graficar los residuos estandarizados (las raíces cuadradas de los sumandos de ECMA), así como los residuos de validación cruzada contra los valores observados, es que son muy útiles para detectar puntos aberrantes. En caso de encontrar puntos aberrantes, se recomienda omitirlos, volver a calcular el variograma empírico y proceder a evaluar XVAL en la forma que se hizo en el Capítulo 4.

Como las medidas de validación cruzada ECMA y XVAL/VKM dependen de la varianza del kriging, también lo hacen de los parámetros del modelo de variograma que se está probando. Por esto, hay veces en que la información de la muestra se anula (a

menos que hubiese valores muy discrepantes en la muestra) y no influye en el comportamiento de dichas medidas. El mayor inconveniente reside en que, al seguir el criterio de ECMA y de XVAL/VKM para elegir modelos, resultan igualmente preferibles una infinidad de modelos que difieren bastante entre sí. Por estos motivos se piensa que es mejor no utilizar a ECMA y a XVAL/VKM para elegir el modelo de variograma "óptimo".

En cuanto a VKM, dado que depende tan sólo de la rejilla muestral y del modelo de variograma, mas no de la muestra, es aconsejable usarlo tan sólo con fines de diseño del muestreo.

De las medidas de validación cruzada, la que se encontró que tiene un potencial enorme para la estimación de los parámetros del modelo, así como para indicar si existe o no evidencia de correlación espacial fue XVAL.

Dado que el criterio es elegir al modelo de variograma que tenga asociado un menor valor de XVAL, se pensó que era importante encontrar la distribución estadística de esta medida para llevar a cabo un contraste de hipótesis. Así se podría decidir si hay evidencia estadística de que un modelo dado ajuste bien a los datos muestrales o no. En el caso de conocer los parámetros del modelo de variograma que sustenta la hipótesis nula, se encontró que la distribución de XVAL bajo dicha hipótesis era la de una suma de ji-cuadradas ponderadas. Mediante el algoritmo de Imhof se obtuvo una aproximación de esta distribución tras aplicar el teorema de inversión de Fourier a la función característica. Se espera que este resultado de pie a otros más interesantes y generales en el futuro.

Cabe aclarar que en este trabajo no se estudió el efecto de tener rejillas muestrales con mismo número de puntos pero diferentes formas regulares e irregulares. Por ejemplo, para un mismo tamaño de muestra de cien puntos, una rejilla regular puede ser de 10x10 o de 5x20. En la de 5x20 el perímetro de la rejilla es mayor y esto afecta al predictor de kriging, pues se aumenta lo

que se llama "efecto de borde". Este tipo de análisis de sensibilidad queda por hacer. Otro estudio que falta realizar es analizar el efecto en XVAL y ECMA de considerar vecindades móviles para calcular el predictor de kriging. Recuérdese que en este trabajo se usó una vecindad única.

Quedan muchos caminos más por recorrer en investigaciones ulteriores. Uno de ellos es encontrar la distribución de XVAL para el caso de parámetros desconocidos. Otro camino es el de incorporar a XVAL en alguna prueba para independencia espacial. Uno más es encontrar la manera de aplicar las técnicas de *bootstrap* para estimar los parámetros del variograma. Por último se cree que el análisis de las diferencias entre las curvas de XVAL al ajustar modelos de diversas familias podrían ser de gran ayuda para una buena estimación de los parámetros del variograma.

Sería muy satisfactorio que los conceptos aquí desarrollados en forma empírica sirviesen de base para investigaciones teóricas posteriores.

BIBLIOGRAFIA

- ARMSTRONG, M. (1984). "Common Problems Seen in Variograms". *Mathematical Geology*, V.16, No.3, 305-313.
- BILLINGSLEY, P. (1986). *Probability and Measure* (Segunda edición). Nueva York: John Wiley & Sons, 622 p.
- BREIMAN, L. (1968). *Probability*. Massachussettes: Addison Wesley P.C., 529p.
- BRIGHAM, E.O. (1974). *The Fast Fourier Transform*. Nueva Jersey: Prentice Hall Inc., 252p.
- BROOKER, P.I. (1986). "A Parametric Study of Robustness of Kriging Variance as a Function of Range and Relative Nugget Effect for a Spherical Semivariogram". *Mathematical Geology*, V.18, No.5, 477-488.
- CHICA OLMO, M. (1988). "Análisis geoestadístico en el estudio de la explotación de los recursos minerales". Tesis de doctorado, Universidad de Granada, España.
- CRESSIE, N. (1985). "Fitting Variogram Models by Weighted Least Squares". *Mathematical Geology*, V.17, No.5, 563-586.
- CRESSIE, N. (1989). "Geostatistics". *The American Statistician*, V.43, No.4, 197-202.
- CRESSIE, N. (1990). "The Origins of Kriging". *Mathematical Geology*, V.22, No.3, 239-252.
- CRESSIE, N. y D.M. HAWKINS (1980). "Robust Estimation of the Variogram: I". *Mathematical Geology*, V.12, No.2, 115-125.
- DAVIS, M.W. (1987). "Production of Conditional Simulations via the LU Triangular Decomposition of the Covariance Matrix". *Mathematical Geology*, V.19, No.2, 91-98.

DIAMOND, P. y M. ARMSTRONG (1984). "Robustness of Variograms and Conditioning of Kriging Matrices". *Mathematical Geology*, V.16, No.8, 809-822.

DUBRULE, O. (1983). "Cross Validation of Kriging in a Unique Neighborhood". *Mathematical Geology*, V.15, No.6, 687-699.

HASTINGS, N.A. y J.B. PEACOCK (1975). *Statistical Distributions*. Londres: Butterworth & Co. Publishers, Ltd, 130p.

HENLEY, S. (1987). "Kriging - Blue or Pink?" *Mathematical Geology*, V.19, No.2, 155-158.

HUBER, P.J. (1981). *Robust Statistics*. Nueva York: John Wiley & Sons, 308p.

IMHOF, J.P. (1981). "Computing the Distribution of Quadratic Forms in Normal Variables". *Biometrika*, V.48, No.3 y 4, 419-426.

ISAAKS, E.H. y M. SRIVASTAVA (1989). *An Introduction to Applied Geostatistics*. Oxford: Oxford University Press, 561p.

JOHNSON, N. y S. KOTZ (1970). *Continuous Univariate Distributions-2*. Boston: Houghton Mifflin Co., 306p.

JOURNEL, A.G. y C.J. HUIJBREGTS (1978). *Mining Geostatistics*. Londres: Academic Press, 600p.

JOURNEL, A.G. (1988). "New Distance Measures: The Route Toward Truly Non-Gaussian Geostatistics". *Mathematical Geology*, V.20, No.4, 459-475.

KINDERMAN, A.J. y J.G. RAMAGE (1976). "Computer Generation of Normal Random Numbers". *Journal of the American Statistical Association*, V.71, No.356, 893-896.

KOTZ, S. y N.L. JOHNSON (Editores) (1983). *Encyclopedia of Statistical Sciences*. Nueva York: John Wiley & Sons.

LESTER, J.D. (1980). *Writing Research Papers* (Tercera edición). Glenview: Scott, Foresman and Co., 207p.

MATHERON, G. (1971). *The Theory of Regionalized Variables and its Applications*. Ecole supérieure des Mines de Paris: *Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau*, No.5.

MOOD, A.; F. GRAYBILL y D. BOES (1974). *Introduction to the Theory of Statistics*, (Tercera edición). Auckland: Mc Graw-Hill International Book Co., 564p.

MYERS, D.E. (1988). "Interpolation with Positive Definite Functions". *Etudes Geostatistiques V - Seminaire C.F.S.G. sur la Geostatistique 15-16 Juin 1987, Fontainebleau Sci. de la Terre*, V.28, 251-265.

MYERS, D.E. (1987 a). "On Variogram Estimation". Por publicarse en: *Proceedings of the First International Conference on Statistical Computing, April 29, 1987, Cesme, Izmir, Turquia*.

MYERS, D.E. (1987 b). "Optimization of Sampling Locations for Variogram Calculations". *Water Resources Research*, V.23, No.3, 496-500.

MYERS, D.E. (1989). "Vector Conditional Simulation". En: ARMSTRONG, M. (Ed.), *Geostatistics*, V.1, 283-293.

PARDO IGUZQUIZA, E. (1989). "Simulación condicional geoestadística de parámetros geomíneros en una y dos dimensiones". Tesis de licenciatura, Universidad de Granada, España.

ROYLE, A.; I. CLARK; et al. (1980). *Geostatistics*. Nueva York: Mc Graw Hill, 168p.

RUSSO, D. (1984). "Design of an Optimal Sampling Network for Estimating the Variogram". *Soil Sci. Soc. Am. J.*, V.48, 708-716.

SAMPER CALVETE, F.C. y J.C. RAMIREZ (1990) *Geoestadística. Aplicaciones a la hidrología subterránea*. Barcelona: CIMNE.

SOLOW, A.R. (1985). "Bootstrapping Correlated Data". *Mathematical Geology*, V.17, No.7, 769-775.

SOLOW, A.R. (1990). "Geostatistical Crossvalidation: A Cautionary Note". *Mathematical Geology*, V.22, No.6, 637-639.

WARRICK, A.W.; D.E. MYERS Y D.R. NIELSEN (1986). "Geostatistical Methods Applied to Soil Science". En: *Methods of Soil Analysis, Part I, Physical and Mineralogical Methods - Agronomy Monograph, No.9* (2aEdición), 53-81.

ZIMMERMAN, D.L. y M.B. ZIMMERMAN (1991). "A Comparison of Spatial Semivariogram Estimators and Corresponding Ordinary Kriging Predictors". *Technometrics*, V.33, No.1, 77-91.